Robust View-Invariant Multiscale Gait Recognition

Sruti Das Choudhury, Tardi Tjahjadi

School of Engineering, University of Warwick Gibbet Hill Road, Coventry, CV4 7AL, United Kingdom.

Abstract

The paper proposes a two-phase view-invariant multiscale gait recognition method (VI-MGR) which is robust to variation in clothing and presence of a carried item. In phase 1, VI-MGR uses the entropy of the limb region of a gait energy image (GEI) to determine the matching gallery view of the probe using 2-dimensional principal component analysis and Euclidean distance classifier. In phase 2, the probe subject is compared with the matching view of the gallery subjects using multiscale shape analysis. In this phase, VI-MGR applies Gaussian filter to a GEI to generate a multiscale gait image for gradually highlighting the subject's inner shape characteristics to achieve insensitiveness to boundary shape alterations due to carrying conditions and clothing variation. A weighted random subspace learning based classification is used to exploit the high dimensionality of the feature space for improved identification by avoiding overlearning. Experimental analyses on public datasets demonstrate the efficacy of VI-MGR.

Keywords: Gait recognition, entropy, Gaussian filter, focus value, weighted random subspace learning.

1. Introduction

Compared with physiological biometrics, e.g., fingerprint, face, iris and earlobe geometry, the behavioural biometrics of gait has the advantage of being able to identify a human subject unobtrusively using low resolution video sequences [1]. The markerless gait recognition methods can be classified into appearance-based and model-based. Appearance-based methods (e.g., [2, 3, 4, 5, 6, 7]) analyse the spatio-temporal shape and dynamic motion characteristics of silhouettes in a gait sequence without using a human body model. Model-based methods (e.g., [8, 9, 10, 11]) characterise a human subject using a structural model to measure time-varying gait parameters, e.g., gait period, stance width and stride length, and a motion model to analyse the kinematic and dynamical motion parameters of the subject, e.g., rotation patterns of hip and thigh, joint angle trajectories and orientation change of limbs.

The main challenges to successful gait recognition are variation in view, variation in subject's clothing and presence of a carried item. Motivated by the unavailability of a gait recognition method that addresses all these challenges, this paper proposes the view-invariant multiscale gait recognition method (VI-MGR) which is robust to variation in clothing and presence of a carried item. VI-MGR is based on integrative scientific principles with a consideration of low computational complexity that enable it to robustly identify a human subject in the presence of numerous challenging factors of realistic scenario for effective visual surveillance.

A gait energy image (GEI) [3], which is formed by averaging the silhouettes of a gait period, is widely used in appearance-based methods as it facilitates noise-resilient robust gait feature extraction with reduced storage space and

computation time. The proposed VI-MGR is an appearance-based method based on GEIs. It comprises two phases: (1) to determine the matching gallery view with the probe; and (2) to identify the probe subject. Compared to other parts of a subject's body, the limb region of a GEI better captures the discriminative information due to variation in view, and is least affected by most carrying conditions and clothing variation. Thus, the phase 1 of VI-MGR computes entropy of the limb region of the GEIs based on anthropometry to determine the matching view of the probe in the gallery using 2-dimensional principal component analysis (2D PCA) and 2D Euclidean distance classifier.

The Gaussian filter is the only filter which can generate a scale-space representation of an image parameterized by the size of a smoothing kernel. Since the Gaussian filter suppresses the fine details in an image by attenuating high frequency components, it blurs an image. As the scale of the Gaussian filter increases, the blurriness increases which results in a gradual loss of boundary and exterior region, thus highlighting the inner shape characteristics. The method in [12] demonstrated the superiority of multiscale shape analysis using Gaussian filter for shape classification compared to conventional shape classification methods, e.g., elliptic Fourier descriptor, Zernike moments, and wavelet transform. Multiscale shape analysis using Gaussian filter is more effective in gait recognition as it results in the loss of boundary characteristics of a subject which are usually affected by the variation in clothing and carrying conditions. Although successful attempts have been made, e.g., [5, 6, 7, 13, 14, 15] in using GEI to outperform the original GEI method [3], multiscale analysis of a GEI has not been exploited despite its high discriminatory power. Hence, the main motivation for VI-MGR is to demonstrate the efficacy of multiscale shape analysis in gait recognition to achieve high identification rate in the presence of clothing variation and carrying conditions. Thus, in phase 2, VI-MGR analyses shape characteristics of GEIs in the image space using Gaussian filter to exploit the discriminatory shape characteristics of the subject at 3 scales. The blurred GEIs obtained by the application of Gaussian filter at different scales, are combined to generate a multiscale gait image (MGI) for use as the gait signature. The probe gait signatures are matched with the gait signatures of the same view of the gallery subjects for identification. Since only dominant features persist across scales, the method is noise-resilient.

If the gallery subjects for training are recorded under similar physical conditions, the learned features in the presence of covariates are likely to cause overfitting that decreases the subject identification rate. Thus, the methods in [3, 5] manually compute synthetic gait templates following a distortion model based on anthropometry to take into account of lower body part distortions due to variation in walking surface, footwear and clothes. The use of the templates enables these methods to be insensitive to the lower body-part distortions, but not upper body-part due to clothing variation and carrying conditions. The dynamics normalisation based gait recognition (DNGR) method [4] uses eigenstance reconstruction model to improve the silhouettes by reducing the effect of shadows and segmentation errors. Since there are numerous covariates, it is challenging to create the appropriate gait template for robust gait recognition. The method in [16] applies part-based strategy to adaptively assign more weight to the unaffected body parts and less weight to the affected body parts to achieve insensitiveness to clothing variation. However, it is unrealistic to train the model with all known clothing types as attempted in [16]. VI-MGR avoids overfitting, and achieves invariance to clothing variation and carrying conditions by introducing MGI with weighted random subspace

learning (WRSL) for classification. WRSL is an ensemble classifier which randomly selects multiple subspaces from the feature space, with a classifier for each subspace. It addresses the apparent dilemma of accuracy optimization and over-adaptation by exploiting high dimensionality.

The novelties of VI-MGR are: (1) It achieves robustness to variation in view, clothing and presence of a carried item (i.e., the three main challenges of gait recognition), as well as several other covariate factors, e.g., segmentation noise, missing body parts, change in ground surface, shadows under feet and occlusions; (2) It introduces reflected GEI (R-GEI) to create variation of the reference gallery views to address any unknown probe view in the range [0°, 360°]; (3) It provides a new approach to achieve view invariance by comparing the probe with all the reference gallery views based on entropy of the limb region of a GEI; (4) It introduces multiscale shape analysis in gait recognition, and achieves invariance to clothing variation and carrying conditions by introducing MGI; (5) It uses focus value as a measure of blurriness of the filtered GEIs to determine the ideal range of scales. A minimum number of three scales are selected from this range to make a trade-off between computational complexity and identification rate; (5) By using WRSL, VI-MGR exploits high dimensionality to avoid overfitting and achieve high identification rate.

The rest of the paper is organized as follows. Section 2 discusses related works and Section 3 presents VI-MGR. Section 4 presents the experimental results and Section 5 concludes the paper.

2. Related Work

Gait recognition methods that address variation in view either depend on (a) extraction of gait features which are view invariant; (b) learning mapping or projection relationship between the gait characteristics of one view and another based on view transformation; or (c) construction of a 3-dimensional (3D) model of a subject from 2D images captured from different views using multiple calibrated cameras.

A statistical feature extraction strategy is used in [17] to extract view-invariant features from parts of a GEI which overlap between different views of a probe and a gallery subject. However, the performance of the method is not satisfactory if there is little overlap due to extreme variation in view. The method based on joint's position estimation and viewpoint rectification (JPE-VR) in [18] determines motion of a subject's lower limb based on anatomical positions of hip, shin and ankle for view-invariant gait recognition using a viewpoint rectification. However, the ankle is most likely to be occluded by the presence of shadows under feet. Since it is impossible to estimate the positions of hip and shin in the case of a subject either wearing a skirt or a long coat, and carrying an item in an upright position, the method is also not robust against variation in clothing and carrying conditions. The method in [19] projects a gait texture image formed by averaging binary gait images of a gait period of a certain view onto the canonical view based on domain transformation using transform invariant low-rank textures. The method in [20] computes view-normalised trajectories of the subject's head and feet. The normalisation involves the decomposition of walking trajectory into piece-wise linear segments to transform the head and feet trajectories from different views into fronto-parallel view based on homography. Since the feet trajectory is affected by self-occlusions, the method is applicable only to a

limited variation in view.

View-invariant methods based on view transformation aim to learn a mapping relationship among gait features of a subject perceived across views. To identify a subject based on gait sequences of different views, the gait features in the probe view are transformed to that of the gallery view before a distance measure is computed [21]. The method in [22] uses discrete Fourier transform (DFT) to obtain gait features from a spatio-temporal gait silhouette volume, and applies a view transformation model on the extracted gait features. The method in [23] creates a view transformation model using support vector regression based on local dynamic feature extraction to transform gait characteristics of one view into the probe view. The method in [24] uses joint subspace to learn a subject's prototype of different views, and represents the subject as a linear combination of these prototypes. Although these methods can cope with large variation in view without relying on camera calibration, they suffer from degeneracies and singularities caused by gait features which are perceived in one view but not in the other view usually due to large view angle difference with the former. Similar to the methods based on view transformation, the method based on canonical correlation analysis (CCA) in [21] also captures the mapping relationship between gait features of different views. However, instead of reconstructing gait features to the required gallery view, this method uses CCA to project the gait sequences of two views onto two maximally correlated subspaces, and uses the correlation strength as a similarity measure between the two gait sequences to overcome the problems associated with the view transformation model. But the effect of variation in clothing and carrying conditions are not considered in this method. The methods based on mapping and projection relationships [21, 22, 23, 24] rely on supervised learning, i.e., require the availability of the gait characteristics of all views to establish a relationship among them during training.

The method in [25] constructs the subject's 3D model from 2D images captured from multiple calibrated views. 2D gait features of the probe view are then obtained from the model for view-invariant gait recognition. The method uses a stick model to simulate a subject's gait. It combines static gait characteristics obtained by anthropometric measurements of different body parts with the dynamic gait characteristics obtained by analysing the joint angle trajectories of lower limbs for identifying a subject based on linear time normalisation. In addition to variation in view, the method is also robust to self-occlusions and change in ground surface. However, the method is only suitable for a fully-controlled environment.

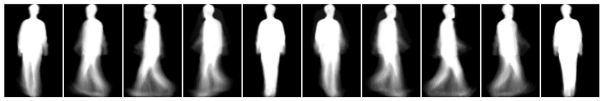
The methods in [26, 27, 28, 11] aim to achieve invariance to carrying conditions. The method based on spatio-temporal motion characteristics, statistical and physical parameters (STM-SPP) [26] analyses the shape of a silhouette contour using Procrustes shape analysis at the double support phase and elliptic Fourier descriptors (EFDs) at ten phases of a gait period. The method in [27] combines model-based and model-free approaches to analyse the spatio-temporal shape and dynamic motion (STS-DM) characteristics of a subject's contour. A part-based EFD analysis and a component-based FD analysis based on anthropometry are respectively used in STM-SPP and STS-DM to address shape distortions due to a subject carrying small items. An iterative local curve embedding algorithm is used in [28] to extract double helical signatures from the subject's limb to take into account of shape distortion due to a specific carrying condition, e.g., a briefcase in upright position. The method in [11] uses models to obtain

skeleton parameters by wavelet decomposition of a GEI and extract invariant moments for combining anatomical and behavioural gait characteristics. The use of thermal imaging enables it to achieve invariance to carrying conditions and lighting variation.

Most gait recognition methods (e.g., [5, 6, 7, 15, 16, 26, 27]) aim to address different covariates but assume same views, i.e., lateral view of the gallery and probe gait sequences. The method in [15] enhances the dynamics of a GEI by computing gait entropy image to identify a subject with varying covariates. The method in [5] uses chrono-gait image (CGI) to generate real and synthetic CGI templates to address lower body part distortions due to carrying a briefcase, variation in ground surface, clothing and footwear, but not the distortions in the upper body due to carrying conditions and variation in clothing. The method in [6] computes gait flow image (GFI) from binary silhouettes, which uses optical flow field to determine relative motion in a gait sequence. The method based on general tensor discriminant analysis and Gabor features (GTDA-GF) [14] uses sum of Gabor filter responses over directions, sum of Gabor filter responses over scales, and sum of Gabor filter responses over scales and directions as gait features. The methods GFI [6] and GTDA-GF [14] outperform the method GEI for the cases of carrying a briefcase, variation in view and footwear of the HumanID gait challenge dataset. The Gabor wavelet and patch distribution feature (PDF) based method (GPDF) in [7] uses locality-constrained group sparse representation (LGSR) to classify local augmented Gabor features extracted at different scales and orientations of GEIs to obtain high identification rate at the expense of high computational complexity.

Since different clothes worn by the same subject in gallery and probe gait sequences reduce the identification rate, the method in [16] introduces part-based feature extraction based on adaptive weight control to achieve clothing invariance. The method in [29] uses genetic algorithm to fuse the features extracted by radial integration transform, circular integration transform and weighted Krawtchouk moments (RCK-G). It assigns depth information to binary silhouettes using 3D radial silhouette distribution transform and 3D geodesic silhouette distribution transform. RCK-G is robust to limited clothing variation, but not insensitive to carrying conditions.

Since VI-MGR does not aim to learn a mapping relationship between gait features across views to reconstruct gait features of a probe view to that of a gallery view, it does not suffer from degeneracies/singularities problems associated with view transformation models due to large variation in view angles. Instead, it detects the matching gallery view of the unknown view of the probe subject, and compares the probe subject with the matching view of the gallery subjects for identification using multiscale shape analysis. Hence, VI-MGR achieves robustness against large variation in view angles without requiring any mapping or projection relationship, and does not rely on a fully controlled environment fitted with calibrated cameras. While the trend of existing gait recognition methods is to address one or a few covariates, VI-MGR aims to address three main challenges of gait recognition in addition to other covariates, i.e., change in ground surface, missing body parts due to occlusions and segmentation errors.The aim is achieved by multiscale shape analysis and WRSL.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Figure 1: (a)-(e): A subject's GEI from CASIA B gait dataset for views 18°, 54°, 90°, 126°, and 180°, respectively; (f)-(j) the corresponding R-GEIs.

3. Proposed method: VI-MGR

The algorithm of the proposed two-phase VI-MGR is summarised in Algorithm 1.

Algorithm 1 VI-MGR

Phase 1

Input: The different known views of a GEI in the gallery and a probe GEI of an unknown view.

Output: The matching gallery view of the probe GEI.

1: Create synthetic gallery views, i.e., R-GEIs, by computing mirror reflection of the available views of the gallery GEIs.
2: Compute segmented GEIs, i.e., SGEIs, by cropping the leg region of the gallery GEIs (reference GEIs (Rf-GEIs)).
3: Compute target SGEIs, i.e., Tr-SGEIs by cropping the leg region of the probe GEIs.
4: Create reference segmented gait entropy image, i.e., Rf-SGEnI, and target segmented gait entropy image, i.e., Tr-SGEnI, by computing entropy of Rf-SGEI and Tr-SGEI, respectively.
5: Perform 2D PCA of Rf-SGEnIs for data decorrelation and dimensionality reduction.
6: The matching gallery view of the probe is detected based on the Euclidean distance classifier.

Phase 2

Input: The gallery classes of GEIs of detected matching view with the probe and the probe GEIs.

Output: Classification of the probe subject.

1: Apply Gaussian filter to a GEI at three scales to generate the blurred GEIs.
2: Concatenate the blurred GEIs to form the MGI.
3: Apply 2D PCA on MGIs for dimensionality reduction and data decorrelation in the eigenspace.
4: Classify the probe subject using WRSL.

3.1. Phase 1: Detect matching gallery view of the probe

To automatically determine the matching gallery view of the probe subject, all the probe views are required to be available in the gallery. However, since phase 2 of VI-MGR is robust against about 30° variation in view due to Gaussian blurring (see Section 4.1), the subject identification rate is not significantly affected for the following cases: (1) if phase 1 incorrectly matches the probe view to a gallery view close to the matching view; and (2) if

phase 1 matches the closest gallery view due to absence of the exact matching view in the gallery. In addition, to take into account of a probe subject moving freely in different directions between 0° to 360°, VI-MGR computes mirror reflection of the available views of the gallery GEIs to create R-GEIs using

$$f(x,y) = f(-x,y), \tag{1}$$

where $f(x,y)$ is the original GEI. Fig. 1 shows reflected GEIs corresponding to the 11 different views of CASIA B gait dataset in [0°,180°] to create additional gallery views in [180°,360°]. Thus, VI-MGR does not require the availability of all probe views in the gallery as in the method CCA [21].

When there is a variation in view, better distinguishable shape variation is manifested in the limb region compared to the head and torso of the GEIs. Also, the shape of the head and/or torso of a GEI are significantly affected by the carried items on head and back, and using folded arms. The torso is also affected by carrying a briefcase in upright position, and most clothing types. Thus, VI-MGR obtains segmented GEIs, i.e., SGEIs, by cropping the region enclosed between the bottom of the GEIs and up to the anatomical positions of knee, i.e., $0.285H$ [30], where H is the height of GEI. Fig. 2 shows SGEIs of the GEIs of a normal walking sequence from CASIA B gait dataset at 11 views. Fig. 3 shows SGEIs of the GEIs of a subject carrying a bag from CASIA B gait dataset at 11 views. The similarity between SGEIs of the normal walking sequence and walking with a bag for a particular view shows that SGEIs are not affected by the carrying conditions. Although SGEIs are likely to be partially occluded by the presence of long skirts and long coats, they remain unaffected by the most clothing which supports the use of SGEIs to achieve invariance to view. The method in [21] does not consider the occlusion of a limb region due to carried items or clothing while using the truncated GEIs for probe view detection. Thus, a SGEI is more effective for view detection due to its ability to overcome the adverse effect of clothing and carried items based on anthropometric analysis. Additionally, matching gallery view detection of the probe based on entropy analysis of SGEIs is not affected by self-occlusions. This is a major challenge for the method in [20] as it requires to determine the feet trajectories of the silhouettes on a frame-by-frame basis. VI-MGR computes 22 SGEIs corresponding to 22 views of a subject's GEI created by mirror reflection of its 11 views from CASIA gait dataset B walking without wearing coat or carrying a bag. These 22 SGEIs are set as the Reference SGEIs (Rf-SGEIs) for use as the matching probe view detection.

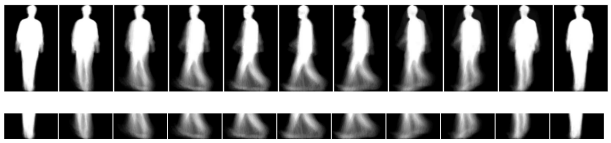


Figure 2: Row 1: GEIs of a same subject from CASIA B gait dataset for normal walking at 11 views. Row 2: the corresponding SGEIs.

The Gabor wavelets are useful for local feature extraction due to their following properties: (a) their kernels are similar to the receptive field profiles of the simple cells of the mammalian visual cortex, which add to their suitability for biology related pattern recognition applications; (b) Gabor wavelets are optimally localized in time and

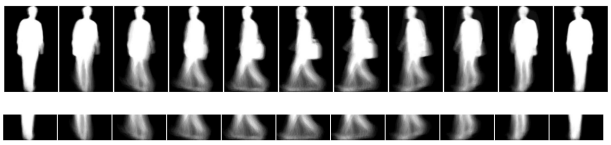


Figure 3: Row 1: GEIs of a subject carrying a bag from CASIA B gait dataset at 11 views. Row 2: the corresponding SGEIs.

Figure 4: SGEnIs of a subject from CASIA B gait dataset for 11 views.

frequency domains, and exhibit useful characteristics of frequency and orientation selectivity; and (c) they facilitate the extraction of desired local features due to their spatial localization property [31]. Since the aim of VI-MGR is to demonstrate the effectiveness of multiscale approach for subject identification when the matching gallery view of the probe subject is not detected by keeping computational complexity as low as possible, the effectiveness of Gabor wavelets for directionally selective local feature extraction at the expense of very high computational complexity is not desirable in VI-MGR (see Section 4.4). The local binary patterns are effective for analysing homogeneous textures at low computational complexity, but lack directional selectivity. Although they can distinguish between images in terms of first derivative information, they cannot adequately determine the velocity of local variation [32]. To overcome the limitation of local binary patterns which only deal with the static information of an image for homogeneous texture analysis, volume local binary patterns are introduced in [33] to take into account of spatio-temporal information of an image for dynamic texture analysis. However, volume local binary patterns consider the co-occurrences of all neighbouring points from three parallel frames in 3D space, which not only add complexity, but increase the size of the feature vector [33]. Compared to Gabor wavelets, entropy analysis also lacks directional selectivity. But, unlike local binary patterns, 2D entropy enhances the dynamic information content of an image as the pixel values comprising the dynamic regions of an image (i.e., SGEIs in VI-MGR) are more uncertain and thus more informative leading to higher entropy values. Hence, local binary patterns are not used in VI-MGR to detect matching gallery view of the probe based on local dynamic information of the SGEIs, but entropy has been used instead.

The entropy of an image is a measure of its randomness used to characterise its texture. To automatically detect the view of an unknown probe subject, the entropy of Rf-SGEIs and the SGEI of the unknown view of the probe, i.e., target SGEI (Tr-SGEI) are computed using [15]

$$\text{SGEnI} = -\sum_{b=1}^{B} P_b(x,y)\log_2 P_b(x,y), \tag{2}$$

where $P_b(x,y)$ denotes the probability that the grey level of pixel (x,y) of a SGEI is b, and B denotes the total number of grey levels in that SGEI. Fig. 4 shows the segmented gait entropy images (SGEnI) obtained from the SGEIs shown in Fig. 3. It shows that the SGEIs are characterised by high intensity values in their corresponding SGEnIs, thus enhancing the dynamic characteristics of the limb region for probe view detection.

Let $\{\mathbf{En}_1, \mathbf{En}_2, ..., \mathbf{En}_M\}$ be the set of M number of reference SGEnIs, i.e., Rf-SGEnIs of size $m \times n$, where $M =$

22 corresponds to c=22 views. The scatter matrix is

$$\mathbf{S} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{En}_i - \mathbf{M})^T \times (\mathbf{En}_i - \mathbf{M}), \tag{3}$$

where $\overline{\mathbf{M}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{En}_i$. For a given $\mathbf{En}$, let

$$\mathbf{Y}_k = \mathbf{EnX}_k, \tag{4}$$

where $\mathbf{X_k}$ for k=1, 2,..., d are the orthonormal eigenvectors of $\mathbf{S}$ corresponding to the first d largest eigenvalues. Thus, $m \times d$ feature images are represented by $\mathbf{V}_i = [\mathbf{Y}_1^i, \mathbf{Y}_2^i, ..., \mathbf{Y}_d^i]$ for i=1,2,..., M. To detect the view of a probe subject, we define

$$D(\mathbf{V}_i, \mathbf{V}_j) = \sum_{k=1}^{d} \|\mathbf{Y}_k^i - \mathbf{Y}_k^j\|_2. \tag{5}$$

The probe sample $\mathbf{V}$ is assigned to class W_k if

$$D(\mathbf{V}, \mathbf{V}_l) = \min_{j=1}^{c} D(\mathbf{V}, \mathbf{V}_j), \qquad j = 1, ..., c, \tag{6}$$

where $c = 22$ denotes the gallery class views, and $\mathbf{V}_l \in W_k$.

3.2. Phase 2: subject identification

The probe subject and its matching views of the gallery subjects are subjected to multiscale analysis. The motivation for analysing scale-space representation of a GEI originates from the fact that a subject in the real world shows different discriminatory shape characteristics at different scales, e.g., in a world map, the use of large scale shows the shape of the continents, while the use of appropriately small scale reveals the shape characteristics of cities. Multi-scale shape analysis is also more noise resilient as it can selectively utilise the dominant features that persist across scales.

The transfer function of 2D Gaussian distribution in spatial domain is given by [34]

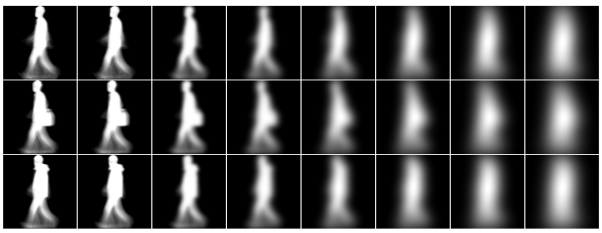
$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}, \tag{7}$$

where σ denotes the standard deviation, i.e., the scale of the Gaussian distribution. To highlight the characteristics of inner region of a GEI gradually towards its centre while removing its boundary shape characteristics that are more likely to be distorted due to variation in clothing and carrying conditions, the Gaussian filter is applied to the GEI using selected scales s, i.e.,

$$GEI_L(u, v)^s = GEI(u, v)G(u, v), \tag{8}$$

where $GEI_L(u, v)$, $GEI(u, v)$ and $G(u, v)$ are respectively the DFT of the filtered GEI, GEI and Gaussian filter, and u and v are frequency variables. The blurred GEI at scale s in the image space is the inverse DFT of $GEI_L(u, v)$.

Fig. 5 shows the GEIs and filtered GEIs of a subject for three types of walking sequences, i.e., normal walking, walking with variation in clothing, and walking with carrying conditions of CASIA B gait dataset. Since Gaussian



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5: Original and Gaussian-blurred GEIs of CASIA B gait dataset for gait sequences: row 1- normal walking; row 2- walking with a carried item; and row 3- walking wearing a coat. (a) GEIs; and (b)-(h) - GEIs filtered by Gaussian filter using increasing scales: (b) $\sigma_1 = 1$; (c) $\sigma_2 = 5$; (d) $\sigma_3 = 10$; (e) $\sigma_4 = 15$; (f) $\sigma_5 = 20$; (g) $\sigma_6 = 24$; and (h) $\sigma_7 = 30$.

filter attenuates high frequency components, it blurs the GEIs. As the scales of the Gaussian filter increases, the blurriness increases resulting in gradual loss of exterior regions of a GEI with its inner shape characteristics highlighted. The variation in clothing and carrying conditions cause alterations to the boundary of a GEI as evident from first column of Fig. 5. The shape distortions due to clothing variation and presence of a carried item are gradually reduced as the blurriness increases as evident from the other columns of Fig. 5, thus enabling VI-MGR to achieve robustness against these covariates. In addition, VI-MGR is also insensitive to other covariates that also cause distortions to the shape of a boundary, i.e., variation in footwear and hair style, shadows under feet and segmentation imperfections.

3.2.1. Classification using WRSL

While the performance of other well-known classification methods like nearest neighbour (NN), support vector machine and Bayesian classifier degrades with increasing dimensionality [35], WRSL as an ensemble classifier benefits from high dimensionality of the feature space to achieve improved identification rate. Experimental results in [35] show that random subspace ensemble classifier performs better than bootstrapping [36] and AdaBoost [37] in the case of the high dimensionality of the feature space for a small number of gallery samples. In gait recognition, the dimension of feature space is very large compared to the available number of gallery samples which lead to overfitting. It has been demonstrated experimentally in [38] that the ensemble classifier called random subspace learning, plays an important role for the avoidance of overfitting and brings significant performance improvements compared to a single classifier, e.g., NN, in gait recognition. Random subspace learning combines the identification rates of the component classifiers associated with the randomly selected independent feature subsets of smaller dimensions than the original feature space using majority voting policy, and thus avoids overfitting due to gallery subjects for training often recorded at a particular walking condition. Due to the random nature of the classification method, it produces different identification rates (a few of them are produced more than once) when run several times. Thus, WRSL is used in VI-MGR which uses a weighted average approach to increase the significance of the identification rates that are repeatedly produced at different runs to compute the final identification rate.

2D PCA is used in VI-MGR as a preprocessing step for WRSL to satisfy its initial requirement of eigenspace

construction. It also enables to reduce the dimensionality of the feature space via data decorrelation without loss of discriminatory information by projecting the highly redundant GEIs onto fewer reconstruction coefficients in an uncorrelated eigenspace. Since 2D PCA operates directly on 2D matrices instead of 1-dimensional (1D) vectors, the 2D image matrix does not need to be transformed into a vector prior to feature extraction. Hence, an image covariance matrix is constructed directly from the original image matrices, i.e., MGIs in VI-MGR, to derive the eigenvectors for extracting gait features. 2D PCA is superior to PCA in terms of more accurate estimation of covariance matrices and reduced computational complexity for feature extraction [39]. Given n MGIs, i.e., {$\mathbf{G}_1$, $\mathbf{G}_2$,..., $\mathbf{G}_n$} in the gallery, the scatter matrix $\mathbf{S}$ is

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{G}_i - \mathbf{M})^T \times (\mathbf{G}_i - \mathbf{M}), \tag{9}$$

where $\mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{G}_i$. Since there are at most $n-1$ eigenvectors of $\mathbf{S}$ with nonzero eigenvalues, N eigenvectors (where N $< n-1$) are randomly chosen from the set of $n-1$ eigenvectors, i.e., $\{\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_{n-1}}\}$, with the largest eigenvalues to construct L subspaces $\{R_k\}_{k=1}^{L}$. The n-th eigenvector with zero eigenvalue is discarded in order to reduce the dimensionality of the feature space while preserving discriminatory information.

For each randomly chosen subspace R_k, k = 1,..., L, the projected image is

$$\mathbf{Y}_i^k = \mathbf{M}_{\mathrm{2DPCA}}\mathbf{G}_i = [\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_N}]^T\mathbf{G}_i, \qquad i = 1, ..., n. \tag{10}$$

where $\{\mathbf{Y}_1^k, ..., \mathbf{Y}_n^k\}$ belong to C gallery classes. VI-MGR uses LDA to determine the projection directions that maximise the inter-class separability, while minimising the distances between the samples of the same class. The key difference between 1D LDA and 2D LDA lies in the mode of data representation. While 1D LDA relies on vectorised representation of data, 2D LDA operates on the data represented in the form of a matrix. To achieve optimal class separability and overcome singularity problem associated with 1D LDA which seeks the discriminant vectors to maximise the ratio of the between-class distance to the within-class distance [40], 2D LDA is used in VI-MGR to seek a transformation matrix $\mathbf{W^k}$ that maximizes the ratio of the between-class scatter matrix $\mathbf{S_B}^{\mathbf{k}}$ to the within-class scatter matrix $\mathbf{S_W}^{\mathbf{k}}$ for each subspace R_k, k = 1,..., L, i.e.,

$$J(\mathbf{W^k}) = \frac{|\mathbf{W^T}\mathbf{S_B^k}\mathbf{W}|}{|\mathbf{W^T}\mathbf{S_W^k}\mathbf{W}|}, \tag{11}$$

where

$$\mathbf{S_W}^{\mathbf{k}} = \sum_{i=1}^{C}\sum_{\mathbf{Y}\in D_i}(\mathbf{Y} - \mathbf{m}_i)(\mathbf{Y} - \mathbf{m}_i)^T, \tag{12}$$

$$\mathbf{S_B}^{\mathbf{k}} = \sum_{i=1}^{C} n_i(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T, \tag{13}$$

$\mathbf{m_i} = \frac{1}{n_i}\sum_{\mathbf{Y}\in D_i}\mathbf{Y}$, D_i is the gallery template set that belongs to the i-th class and n_i is the number of templates in D_i. $J(W^k)$ is maximised when the columns of W are the generalised eigenvectors that correspond to the largest eigenvalues in

$$\mathbf{S_B}^{\mathbf{k}}\mathbf{w}_i = \lambda_i\mathbf{S_W}^{\mathbf{k}}\mathbf{w}_i, \qquad i = 1, ..., C. \tag{14}$$

Since there are at most C-1 nonzero eigenvalues, the corresponding $\mathbf{v_1}$,..., $\mathbf{v_{C-1}}$ eigenvectors are used to form the training gait feature matrix $\mathbf{Z}_i^k$ for each of L subspaces using

$$\mathbf{Z}_i^k = \mathbf{M}_{\text{2DLDA}}\mathbf{Y}_i^k = [\mathbf{v_1}, ..., \mathbf{v_{C-1}}]^T\mathbf{Y}_i^k, \qquad i = 1, ..., n. \tag{15}$$

Let $\{\mathbf{r}\}$ be the set of gallery gait feature matrices belonging to C classes. Thus, each of the C classes of the gallery with n_j feature matrices is represented by its centroid, i.e., $G_j = \frac{1}{n_j}\sum_{r\in R_j}\mathbf{r}$, where j = 1,..., C and R_j is the set of feature matrices belonging to the j-th class.

A gait period is defined as the time interval between successive heel strikes of the same limb (known as the initial contact phase of the gait period [27]). To estimate the gait period, the number of foreground pixels enclosed by the region bounded by the bottom of the bounding rectangle and the anatomical position of just before the subject's hand measured from the bottom (i.e., $0.377H$ [30] where H is height of the bounding rectangle) is counted in each frame of the gait sequence. Since this foreground region is not distorted by self-occlusions due to arm-swing [27], and the number of foreground pixels reaches its maximum when the two feet are farthest apart (i.e., at the initial contact phase), a gait period is determined as the sequence of successive frames enclosed by two frames of a gait sequence with the maximum number of foreground pixels. The distance between a probe sequence P with m_p gait periods and a gallery class centroid G_j is [3]

$$D(P, G_j) = \frac{1}{m_p}\sum_{i=1}^{m_p}||\mathbf{s} - G_j||, \qquad j = 1, ..., C. \tag{16}$$

where $\mathbf{s}$ is the set of a probe gait feature matrices. We assign

$$P \in w_k, \quad \text{if } D(P, G_k) = \min_{j=1}^{C} D(P, G_j), \tag{17}$$

where w_k is the k-th gallery class. A probe sequence P is assigned a class label by each component classifier $A_{\text{LH-GF}} \in \{A_{\text{LH-GF}}\}_{k=1}^{L}$, where $\{A_{\text{LH-GF}}\}_{k=1}^{L}$ is the set of classifiers constructed from L subspaces. The final class of P is determined based on majority voting, and the classification at rank-r implies that the number of votes received by a probe belong to the top r ranks.

Unlike most of the state-of-the-art gait recognition methods which use single classifiers, e.g., NN, Bayesian classifier as in [26, 3], VI-MGR uses an ensemble classifier, i.e., WRSL, as WRSL has the ability to provide better identification rates than the single classifiers by exploiting high dimensionality. Unlike single classifiers, WRSL is based on ensemble of classifiers on randomly selected independent feature subsets. Hence, to take into consideration of the different subspaces which produce different identification rates, WRSL is run ten times for each experiment with N fixed and a randomly chosen value of L. It is observed that some values for rank-1 identification rate are produced more than once during ten runs and similarly for rank-5 identification rate as well. Thus, to increase the significance of the identification rates that are produced more than once in different runs while minimising the effects of the outliers in the final identification rate, VI-MGR computes weighted average of the rank-1 and rank-5 identification rates

produced in different runs using

$$\frac{X_i W_i + X_j W_j + ...}{W_i + W_j + ...}, \tag{18}$$

where W_i (W_j) denotes the number of times the identification rate X_i (X_j) is produced.

3.3. Scale selection

For multiscale shape analysis, appropriate choice of scales enhances the shape recognition significantly. VI-MGR computes the focus value of a silhouette of CASIA B gait dataset filtered by Gaussian filter using different scales to determine the ideal range of scales. The focus value used to measure the degree of sharpness of an image is maximum for the most focused, i.e., the original silhouette. It is inversely proportional to the image blurriness caused by the Gaussian filtering at different scales. Common methods for computing focus values of an image include spatial domain based methods, e.g., Tenengrad [41] and sum modified Laplacian [42], and wavelet based methods [43]. The first level 2D Daubechies-6 wavelet decomposition of a silhouette image $f(x, y)$ of size $M \times N$ results in four subband images, W_{LL}, W_{HL}, W_{LH} and W_{HH}, where L and H respectively denote lowpass filtered and highpass filtered, and their order denotes the order of the filtering applied, e.g., W_{HL} is a subband image obtained by highpass filtering followed by lowpass filtering. The focus value of a silhouette is [43]

$$\begin{aligned} FV &= (\frac{1}{MN}\sum_{y=0}^{N}\sum_{x=0}^{M}(\text{var}(|W_{HL}(x,y)|) + \\ &\quad \text{var}(|W_{LH}(x,y)|) + \text{var}(|W_{HH}(x,y)|)))^{\frac{1}{2}} \end{aligned} \tag{19}$$

where var(.) computes variance.

It has been graphically demonstrated in [43] that the 2D Daubechies-6 wavelet based method of computing focus value has the sharpest focus measure profile compared to Tenengrad and sum modified Laplacian methods due to the localised support property of wavelet basis. It has also been reported in [43] that the depth resolution of 2D Daubechies-6 wavelet based focus measure is approximately 29% higher than that of sum modified Laplacian and 36% higher than that of Tenengrad. Since 2D Daubechies-6 wavelet based focus value computation method has better focus measure profile and higher depth resolution, it is very effective for low resolution images. Thus, this method is chosen for computing the focus value of the low resolution images in VI-MGR. Fig. 6 shows that normalised focus value decreases with increasing scales in the range [1, 30]. The focus values are normalised by dividing them with the maximum focus value in the range, i.e., the focus value of the silhouette filtered using σ=1. The normalised focus value decreases rapidly in the low scale range, i.e., up to scale 14, and then steadily up to σ = 24. If the scale is increased above σ = 24, the GEIs fail to preserve any inter-subject discriminatory information due to excessive blurriness. Thus, the ideal range of scales is chosen as [1,24]. However, the computational complexity increases as the number of scales increases. Hence, to make a trade-off between the identification rate and the computational complexity, a set of minimum number of scales are chosen based on analysing the entropy of the filtered GEIs in the range [1,24].

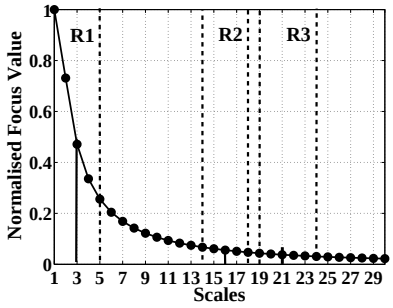


Figure 6: Normalised focus value w.r.t. increasing scales of the Gaussian filter.

The identification rate increases if the discriminability between different subjects is high but the same subjects show similar shape characteristics despite the variation in clothing, and presence of a carried item in one situation while absence in the other. Therefore, to choose the scales effectively, the following three cases are separately considered: case 1- different subjects with no variation in clothing and carrying conditions; case 2- same subjects with and without a carried item; and case 3- same subjects with variation in clothing. The criteria of scale selection for each case are: (1) to choose a scale from a range at which the shape difference between the different subjects is the maximum (case 1); (2) to choose a scale from a range at which the shape difference between the same subjects at different situations, i.e., walking normally vs walking with a bag, is the minimum; and (3) to choose a scale from a range at which the shape difference between the same subjects with considerable variation in clothing and multiple carrying conditions is the minimum.

Fig. 7 shows normalised entropy of filtered GEIs of two subjects walking normally in three range of scales, i.e., R1: [1,5], R2: [14,18] and R3: [19,24] (case 1). Fig. 8 shows normalised entropy of filtered GEIs of a normal walking subject and the same subject carrying a bag in three range of scales, i.e., R1: [1,5], R2: [14,18] and R3: [19,24] (case 2). Fig. 9 shows normalised entropy of filtered GEIs of a normal walking subject and the same subject walking with clothing variation in three range of scales, i.e., R1: [1,5], R2: [14,18] and R3: [19,24] (case 3). Note that in the high scale range, i.e., R3 [19,24], the pair of normalised entropy value curves behave differently for three cases, but the actual difference in the normalised entropy value between the filtered GEIs in each case is very small. Therefore, the behavioural characteristics of the normalised entropy of the filtered GEIs in the high scale range are clearly indicated using small range of values along y-axis, i.e., 0.975 to 1 in Fig. 7 (c), Fig. 8 (c) and Fig. 9 (c).

Case 1- Different subjects with no variation in clothing and carrying conditions. Fig. 7, Fig. 8 and Fig. 9 show that the difference in entropy between two normal walking sequences of different subjects are higher at three scale ranges compared to the same subjects walking with clothing variation and with/without carrying conditions. However, Fig. 7 shows that the discriminability is the highest in the low range of scales, i.e., [1,5]. Thus, this range is most informative for inter-subject discriminability without variation in clothing and presence of a carried item.

Case 2 - Same subjects with and without a carried item. The entropy difference between a normal walking subject

and the same subject walking with a carried item is moderate in the low range of scales, i.e., [1,5], but decreases when the entropy values of two sequences become similar in the intermediate range of scales, i.e., [14,18] (see Fig. 8).

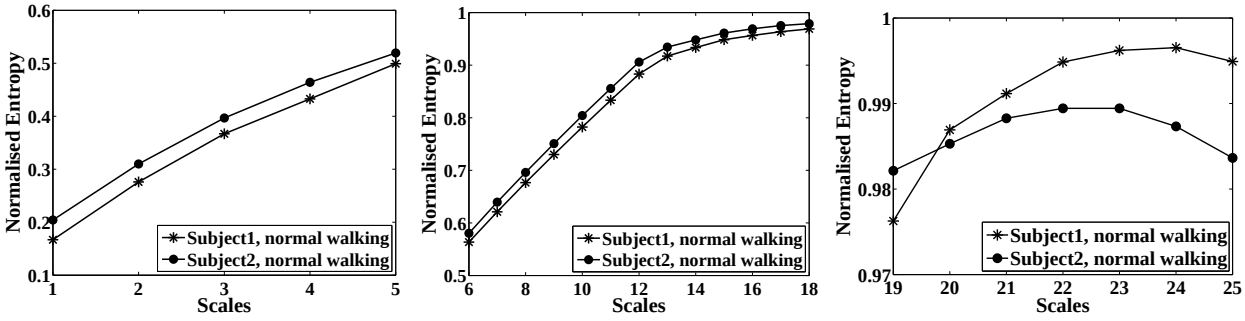
Case 3 - Same subjects with variation in clothing. The entropy difference between a normal walking subject and the same subject walking with clothing variation is high in the range [1,5], but low in the range [14,18] ensuring that Gaussian blur reduces the shape distortions due to clothing variation in this range (see Fig. 9). However, for excessive clothing variation and presence of more than one carried items (at the back, folded arms or is upright position), more blurriness resulted by the high range of scales, i.e., [19,24], is desirable. VI-MGR thus considers three ranges of scales - R1: [1,5], R2: [14,18] and R3: [19,24] (see Fig. 6) for multiscale analysis of GEIs. For computational simplicity, the mid scale from each range, i.e., 3 (from R1), 16 (from R2) and 21 (from R3) are selected. The filtered GEIs using scales 3, 16 and 21 are combined to form MGI for use as a gait signature.

4. Experiments

To demonstrate the ability of VI-MGR to achieve combined robustness against the most challenging factors of gait recognition, it is extensively compared with several related methods that individually address one or more covariate factors. Therefore, VI-MGR is evaluated using several experimental setups to match with the experimental setup of different methods for uniform comparison with their reported results on three publicly available gait datasets: CASIA B gait dataset [44], USF HumanID gait challenge dataset [2] and OU-ISIR treadmill gait dataset B [45]. VI-MGR is compared with the following methods: canonical correlation analysis (CCA) based method [21], joint's position estimation and viewpoint rectification (JPE-VR) based method [18], chrono-gait image (CGI) based method [5], gait energy image (GEI) based method [3], radial integration transform, circular integration transform and weighted Krawtchouk moments (RCK-G) based method [29], gait flow image (GFI) based method [6], method using matrix-based marginal Fisher analysis (MMFA) [13], general tensor discriminant analysis and Gabor features (GTDA-GF) based method [14], dynamics normalisation based gait recognition (DNGR) method [4], spatio-temporal motion characteristics, statistical and physical parameters (STM-SPP) based method [26], spatio-temporal shape and dynamic motion (STS-DM) analysis based method [27] and Gabor wavelet and patch distribution feature (PDF) based method (GPDF) [7]. VI-MGR shows both rank-1 and rank-5 identification rates to make a fair comparison with all the related methods listed in Table 8 and Table 1 which also present identification rates in terms of rank-1 and rank-5.

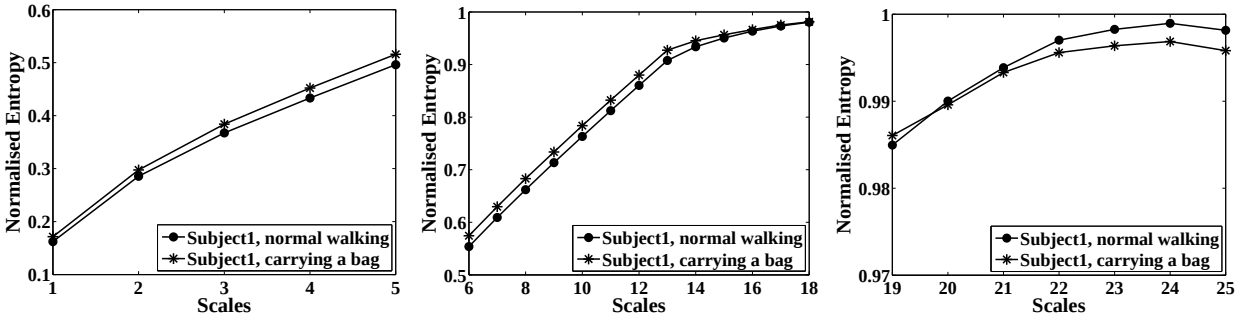
4.1. Experiments on USF HumanID gait challenge dataset

The small version of USF HumanID gait challenge dataset comprises 452 sequences of 74 subjects and the full version comprises 1870 sequences of 122 subjects walking along an elliptical path in front of two cameras. VI-MGR is evaluated on the GEIs of the full version of this dataset which are downloaded from http://www.GaitChallenge.org. The dataset provides up to thirty-two possible testing conditions by combining the following five covariates: (a) walking surface (grass (G) or concrete (C)); (b) shoe type (A or B); (c) viewpoint (right (R) or left (L)); (d) carrying



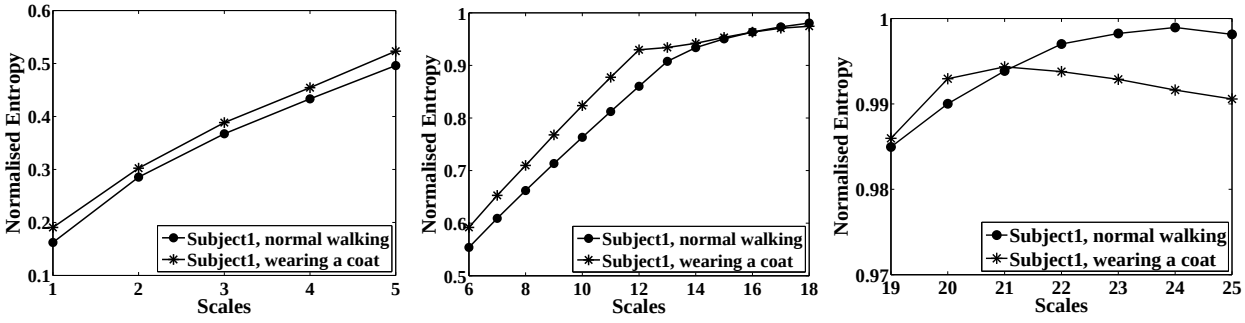
(a) (b) (c)

Figure 7: Normalised entropy of filtered GEIs of normal walking sequences of two subjects: (a) σ = [1,5]; (b) σ = [6,18] and (c) σ = [19,25].



(a) (b) (c)

Figure 8: Normalised entropy of filtered GEIs of a normal walking subject and the same subject carrying a bag: (a) σ = [1,5]; (b) σ = [6,18] and (c) σ = [19,25].



(a) (b) (c)

Figure 9: Normalised entropy of filtered GEIs of a normal walking subject and the same subject walking with clothing variation: (a) σ = [1,5]; (b) σ = [6,18] and (c) σ = [19,25].

conditions (carrying a briefcase (BF) or not carrying a briefcase (NB)); and (e) elapsed time between the acquisition of the sequences (May (M) or November (N))[2]. There are thirty-three common subjects whose gait sequences are recorded in both May and November to account for time covariate. The GEIs provided by the USF HumanID dataset are centre-aligned and normalised to a fixed size 128×88. There is no common sequence between the gallery set and any of the probe sets, and all subjects did not participate in all gait challenge experiments [3, 2].

Since there are different number of probe subjects in the gait challenge experiments, the weighted average identification rate (W-AvgI) is obtained using

$$\text{W-AvgI} = \frac{\sum_{i=1}^{g} w_i x_i}{\sum_{i=1}^{g} w_i}, \tag{20}$$

where g denotes the number of gait challenge experiments whose value is 12 for Exp. A-L, x_i denotes the rank-1 identification rate (%) of the ith gait challenge experiment and w_i denotes the number of probe subjects participating in that experiment. Table 1 shows the final W-AvgI of VI-MGR computed by averaging the identification rates obtained by WRSL for eighty randomly chosen values of L in the range [100, 500] on the large version of the USF dataset at ranks 1 and 5 for comparison with the identification rates of the methods that outperform Baseline, i.e., RCK-G, GFI, GEI, MMFA, GTDA-GF, CGI, DNGR, STM-SPP, STS-DM, GPDF-NN and GPDF-LGSR. All the methods listed in Table 1 use the same gallery set (G, A, R, NB, M/N) consisting of 122 subjects to report the identification rates for the 12 gait challenge experiments as specified by the USF dataset (see the first three rows of Table 1).

The method MMFA [13] applies marginal Fisher analysis on GEIs for gait representation to reduce the dimensionality of the feature space and extends marginal Fisher analysis to marginal based analysis for content-based image retrieval. Table 1 shows the identification rates of GEI obtained by fusing real and synthetic gait features based on statistical feature fusion [3, 46]. GTDA-GF reports the identification rates obtained by applying general tensor discriminant analysis (GTDA) as a preprocessing step of LDA on the magnitude of convolving a GEI with sum of Gabor functions over scales with direction fixed. With the exception of GPDF-LGSR [7], VI-MGR outperforms all other methods in terms of W-AvgI at rank-1 and rank-5 including GPDF-NN [7]. When the method GPDF uses NN classifier, i.e., GPDF-NN, it is significantly outperformed by VI-MGR at ranks 1 and 5, which demonstrates the effectiveness of multiscale shape analysis and WRSL classification method in gait recognition. However, VI-MGR is outperformed by GPDF when it uses a sophisticated classifier called LGSR, i.e., GPDF-LGSR, for improved W-AvgI but this is achieved at the expense of high computational complexity (see Section 4.4).

In gait recognition, while same covariate conditions contribute positively to inter-subject discrimination, intra-class variations due to clothing, shadows under feet or carrying items decrease the recognition rate. Thus, if the gallery subjects for training are recorded at a particular walking condition, and the size of the training set is small compared to the high dimensionality of the feature space, the performance of a single classifier, e.g., NN and Bayesian, to identify a probe subject in the presence of unpredictable covariate factors decreases due to overfitting [38]. A single classifier, e.g., NN, is also very sensitive to the sparsity in the high dimensional space, and performs poorly in the absence of adequate number of training samples compared to the high dimensionality of the feature space [47].

WRSL randomly creates subspaces of smaller dimensions than the original feature space, while the number of training samples remains the same. It uses majority voting policy to combine the identification rates of the classifiers associated with each random subspace, thus it reduces the effects of the weak classifiers for improved final identification. WRSL is characterised by two main parameters, i.e., the number of subspaces (L) and the dimension of the subspace (N), which play a significant role on the recognition accuracy. Since a very small value of N will cause underlearning while a large value will cause overlearning, it is important to choose a suitable value of N for optimal performance. Based on experimental analysis, the method in [35] introducing the random subspace learning, concludes that the recognition accuracy does not decrease with the increase in the number of decision trees. Therefore, we choose the value of N based on experimental analysis on USF HumanID Gait Challenge dataset by keeping L fixed. Fig. 10 shows that if L=50, the value of N in the range [13,20] enables to achieve the optimal rank-1 W-AvgI. Since the aim of the paper is to demonstrate the efficacy of multiscale approach in gait recognition rather than achieve higher W-AvgI through intensive parameter calibration, we fix N=16 for all values of L used in VI-MGR, as very high value of N causes overlearning.

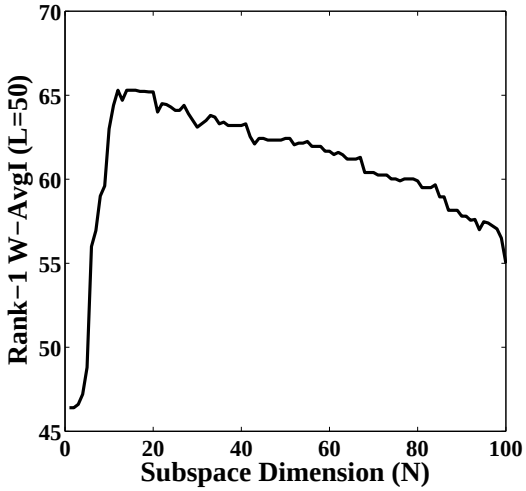


Figure 10: Rank-1 WAvgI using L=50 w.r.t. dimension of the subspace.

Experiments have been conducted (see Table 2) to obtain the identification rates of VI-MGR using L=5, L=15, L=25, L=50, L=100, L=300, L=500 on the large version of the USF dataset at ranks 1 and 5 to determine the ideal range for randomly choosing values of L for optimal performance. The table shows that the identification rates obtained using smaller values of L, i.e., L =5, L=15, L=25 are significantly low due to underlearning. Also, the W-AvgIs for different values of L in Table 2 show that the recognition accuracy gradually increases with the increase in the number of subspaces, which verifies the conclusion in [35] regarding the relation between the number of random subspaces and recognition accuracy. Therefore, the ideal range of VI-MGR is chosen as [100, 500] to report W-AvgIs in Table 1, while smaller values of L, i.e., L =5, L=15, L=25 are discarded to avoid low identification rate, and values higher than 500 are not considered to reduce the computational complexity and avoid overlearning. To avoid

random values of L close to 100 which are likely to generate lower identification rates, 80 random integer values uniformly distributed in the range [100, 500] are selected for L, and W-AvgI is obtained using WRSL for each of these values. Fig. 11 shows that W-AvgI increases with increasing L values. The 80 W-AvgIs obtained are averaged to give the final W-AvgI, which lies between the W-AvgI obtained using L=300 and L=500 (see Table 2). Thus, for reduced computational complexity, VI-MGR uses L=300 for experimental analysis on CASIA B dataset and OU-ISIR treadmill gait dataset B to demonstrate the robustness of VI-MGR against variation in view and clothing of a subject.

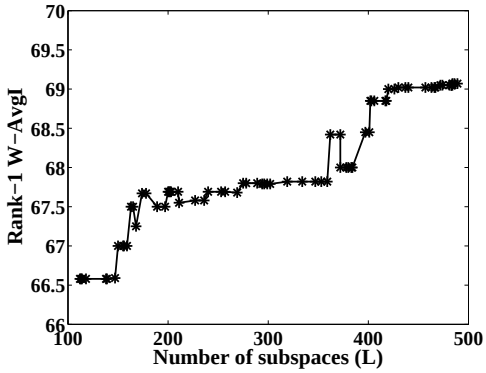


Figure 11: W-AvgI for 80 randomly selected number of subspaces (L), denoted by asterisks.

The rank-1 identification rates obtained using single scale (σ=0, i.e., without blurring of the GEIs) and multiple scales (σ=3,16,21) for all 12 gait challenge experiments (exp. A-L) are shown in Fig. 12 to demonstrate the effectiveness of the proposed multiscale approach. The dashed line and the straight line in the figure respectively denote the W-AvgI using single scale, i.e., 62.90% and multiple scales, i.e., 67.80%. Note that L=300 for WRSL is used to obtain the rank-1 identification rates in the figure for both single scale and multiple scales. The figure shows that the use of multiple scales enables to achieve better identification rates than using single scale for all gait challenge experiments.

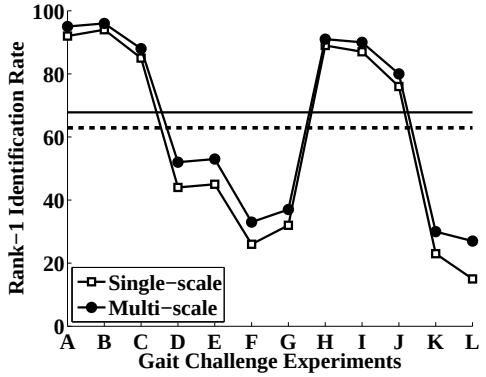


Figure 12: Rank-1 identification rates of VI-MGR using single scale (σ = 0) and multiple scales (σ=3,16,21) using L = 300 for 12 gait challenge experiments of HumanID gait challenge dataset. Dashed line: W-AvgI using single scale and straight line: W-AvgI using multiscale.

Table 1: Identification rates (%) at rank-1 and rank-5 of the gait recognition methods on full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects. Keys for covariates: V - view; H - shoe; S - surface; B - briefcase; T - time; and C - clothes.

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	W-AvgI
Probe Size	122	54	54	121	60	121	60	120	60	120	33	33	-
Covariate	V	H	VH	S	SH	SV	SHV	B	BH	BV	THC	STHC	-
Rank-1 Identification Rate													
RCK-G [29]	83	86	78	39	34	20	21	43	40	40	16	5	44.34
GFI [6]	89	93	70	19	23	7	8	78	67	48	3	9	46.14
GEI [3]	90	91	81	56	64	25	36	64	60	60	6	15	57.66
MMFA [13]	89	94	80	44	47	25	33	85	83	60	27	21	59.90
GTDA-GF [14]	91	93	86	32	47	21	32	95	90	68	16	19	60.58
CGI [5]	91	93	78	51	53	35	38	84	78	64	3	9	61.69
DNGR [4]	85	89	72	57	66	46	41	83	79	52	15	24	62.81
STM-SPP [26]	92	95	84	72	68	29	40	69	60	64	20	18	63.05
STS-DM [27]	93	96	86	70	69	39	37	78	71	66	27	22	66.68
GPDF-NN [7]	90	91	85	53	52	32	28	92	86	64	12	15	62.99
GPDF-LGSR [7]	95	93	89	62	62	39	38	94	91	78	21	21	70.07
VI-MGR	95	96	86	54	57	34	36	91	90	78	31	28	68.13
Rank-5 Identification Rate													
RCK-G [29]	96	94	88	66	63	51	46	66	68	65	44	22	67.03
GFI [6]	98	94	93	40	47	26	25	94	85	74	24	24	63.89
GEI [3]	94	94	93	78	81	56	53	90	83	82	27	21	76.23
MMFA [13]	98	98	94	76	76	57	60	95	93	84	48	39	79.90
GTDA-GF [14]	98	99	97	68	68	50	56	95	99	84	40	40	77.58
CGI [5]	97	96	94	77	77	56	58	98	97	86	27	24	79.12
DNGR [4]	96	94	89	85	81	68	69	96	95	79	46	39	82.05
STM-SPP [26]	96	98	95	80	84	59	61	92	84	85	30	27	79.13
STS-DM [27]	97	98	96	82	83	61	60	95	89	83	39	28	80.48
GPDF-NN [7]	98	94	94	82	79	57	53	99	98	88	33	36	80.84
GPDF-LGSR [7]	99	94	96	89	91	64	64	99	98	92	39	45	85.31
VI-MGR	100	98	96	80	79	66	65	97	95	89	50	48	83.75

Table 2: Identification rates (%) at rank-1 and rank-5 of VI-MGR on full version of USF HumanID gait challenge dataset using the gallery set (G, A, R, NB, M/N) of 122 subjects for L=5, L=15, L=25, L=50, L=100, L=300, L=500. Keys for covariates: V - view; H - shoe; S - surface; B - briefcase; T - time; and C - clothes.

Probe Set	A	B	C	D	E	F	G	H	I	J	K	L	W-AvgI
Probe Size	122	54	54	121	60	121	60	120	60	120	33	33	-
Covariate	V	H	VH	S	SH	SV	SHV	B	BH	BV	THC	STHC	-
Rank-1 Identification Rate													
VI-MGR (L=5)	88	93	78	35	36	20	25	80	72	61	21	16	55.05
VI-MGR (L=15)	91	94	86	36	38	23	27	83	80	65	24	19	58.29
VI-MGR (L=25)	94	95	85	39	41	25	32	88	84	73	27	26	62.03
VI-MGR (L=50)	94	95	85	48	50	29	37	90	86	75	28	26	65.21
VI-MGR (L=100)	94	95	85	50	51	29	37	91	86	79	29	26	66.18
VI-MGR (L=300)	95	96	88	51	52	33	37	92	90	80	30	27	67.79
VI-MGR (L=500)	95	96	89	55	58	35	38	92	90	80	32	29	69.07
Rank-5 Identification Rate													
VI-MGR (L=5)	97	97	93	69	64	46	57	94	89	84	40	37	75.69
VI-MGR (L=10)	97	97	93	75	69	50	60	95	90	86	41	38	77.96
VI-MGR (L=25)	97	97	94	78	72	62	60	95	90	86	41	38	80.10
VI-MGR (L=50)	97	98	95	80	76	62	60	95	90	86	42	38	80.75
VI-MGR (L=100)	97	98	95	81	80	63	60	95	90	87	44	39	81.36
VI-MGR (L=300)	100	99	96	80	78	66	64	98	95	92	46	43	83.87
VI-MGR (L=500)	100	99	96	81	80	67	65	99	95	92	50	48	84.75

4.2. Experiments on CASIA B dataset

CASIA B gait dataset comprises video sequences of 124 subjects (31 females and 93 males) captured by 11 USB cameras (Fametech 318SC) from 11 views in the range [0°,180°] with a difference of 18° between two adjacent views. There are 10 video sequences for each view of a subject: 6 sequences for normal walking, i.e., without wearing a coat or carrying a bag; 2 sequences for walking wearing a coat; and 2 sequences for walking with either a knapsack, a satchel or a handbag. The video sequences are recorded indoor at a rate of 25 frames per second and the resolution of each frame is 320×240. The first 4 normal walking sequences of all subjects are considered as gallery, while the remaining 2 normal walking sequences, 2 walking sequences wearing a coat and 2 walking sequences with carrying conditions are considered as probe. The GEIs of CASIA B gait dataset used in VI-MGR are obtained from http://www.cbsr.ia.ac.cn/users/szheng/. Each GEI is 240×240.

Table 3: Correct view matching rate (CVR) of VI-MGR on CASIA B gait dataset. The available rank-1 view classification results of the method in [21] obtained using Gaussian Process on CASIA B gait dataset are enclosed in parenthesis.

Probe View	CVR (%)			
	Normal	Bag	Coat	AvgA
0°	83	80	79	80.67
18°	94	87	85	88.67
36°	88 (84)	85 (83.4)	80 (84.0)	84.33
54°	92 (91.2)	90 (88.7)	89 (91.2)	90.33
72°	81 (85.3)	80 (84.9)	78 (85.3)	79.67
90°	89 (74.0)	79 (68.6)	72 (74.0)	80
108°	79 (86.0)	75 (83.0)	70 (86.0)	74.67
126°	90 (91.2)	88 (92.7)	85 (91.2)	87.67
144°	83 (93.5)	81 (93.5)	79 (93.5)	81
162°	89	86	84	86.33
180°	82	80	75	79
Mean	86.4	82.8	79.6	-

Table 3 shows the correct view matching rate (CVR) of VI-MGR on CASIA B gait dataset for the subjects walking normally, walking with a bag and walking wearing a coat for 11 views. The mean CVR is the highest, i.e., 86.4%, for the normal walking sequences, followed by walking with a bag, and walking wearing a coat as the SGEIs are partially occluded by the long coats for several subjects. The average CVRs of a particular angle for all types of walking (AvgAs) are comparatively low for views 0° and 180°, as the shape characteristics of a subject remain almost same in these cases as evident from Fig. 3. Also, the degradation in performance for views 72° and 108° is attributed

to the similar shape characteristics with the subjects at 90°. However, the degradation in performance does not affect the identification rate, as the subject identification rate is invariant to slight variation in view. In support of this observation, three confusion matrices for CVR of VI-MGR corresponding to three cases of probe subjects are presented, i.e., normal walking (Table 5), walking with a bag (Table 6) and walking wearing a coat (Table 7). The columns of a confusion matrix represents instances in a predicted class, while the rows represent the instances in an actual class [48].

Table 4: Rank-1 identification rates (%) of VI-MGR, GEI and CGI on CASIA gait dataset B, with rates of GEI and CGI obtained from [5] for lateral views.

Gallery/Probe	GEI [3]	CGI [5]	VI-MGR
Normal/Normal	91.57	88.06	100
Normal/Bag	31.71	43.67	89
Normal/Coat	24.07	42.98	76

Table 5: Confusion matrix for CVR of the normal walking subjects of CASIA B gait dataset.

Actual Class	Predicted class										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
0°	102	0	0	0	0	0	0	0	0	0	22
18°	0	116	6	0	0	0	0	0	1	1	0
36°	0	0	109	10	3	0	0	2	0	0	0
54°	0	0	2	114	7	1	0	0	0	0	0
72°	0	0	0	4	100	15	5	0	0	0	0
90°	0	0	0	0	8	110	6	0	0	0	0
108°	0	0	0	0	6	16	98	4	0	0	0
126°	0	0	0	0	0	0	10	112	2	0	0
144°	0	0	0	0	0	0	0	9	103	12	0
162°	0	0	4	0	0	0	0	2	8	110	0
180°	22	0	0	0	0	0	0	0	0	0	102

The method CCA is not suitable for classifying the front and back views (i.e., 0° and 180°) and views close to them (i.e., 18° and 162°), as these views provide very little gait information [21]. Thus, only rank-1 view classification results using Gaussian process for 7 views (i.e., 36° to 144°) are presented which are enclosed in parentheses in Table 3. Unlike the method CCA [21] which assumes the availability of all probe views in the training dataset, VI-

Table 6: Confusion matrix for CVR of the subjects of CASIA B gait dataset walking with a bag.

Actual Class	Predicted class										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
0°	99	0	0	0	0	0	0	0	0	0	25
18°	0	108	12	0	0	0	0	0	2	2	0
36°	0	2	105	12	5	0	0	0	0	0	0
54°	0	0	4	112	7	1	0	0	0	0	0
72°	0	0	0	6	99	15	4	0	0	0	0
90°	0	0	0	3	10	98	10	3	0	0	0
108°	0	0	0	0	10	16	93	5	0	0	0
126°	0	0	0	0	0	0	12	109	3	0	0
144°	0	0	0	0	0	0	0	10	100	14	0
162°	0	0	6	0	0	0	0	1	10	107	0
180°	25	0	0	0	0	0	0	0	0	0	99

MGR does not require the presence of exact matching probe view in the gallery, hence it outperforms the method CCA in terms of robustness. While the method in [21] uses 60% of the total subjects of CASIA B gait dataset as training and the remaining 40% for testing using a gallery and a probe set, VI-MGR only uses 22 Rf-SGEIs to determine the matching view of a Tr-SGEI based on entropy analysis. The process of matching view detection of an unknown probe subject used in VI-MGR is much simpler compared to the method CCA which uses Gaussian process involving computationally expensive squared exponential covariance function.

Since VI-MGR compares the probe subject with the view of the gallery subjects among all available views in the gallery that matches most closely to it, it is justified to compare VI-MGR with the rank-1 identification rates of CCA and JPE-VR for the same views of the probe and the gallery subjects. Note that the identification rates for the same views of the probe and the gallery subjects are not available in the method in [23, 49]. The method in [49] combines multiple gallery views for comparison with the known probe view based on Radon transform, while the method in [23] uses a view transformation model to transform the known probe view into another gallery view to provide rank-1 identification rate. Fig. 13 shows the rank-1 identification rate of VI-MGR for the view angles 0° to 180° of normal walking subjects to compare with the available rank-1 identification rate of JPE-VR for the same gallery and probe views in the range 36° to 126° only, as JPE-VR is not suitable for other extreme viewpoints. Fig. 14 shows rank-1 identification rates of VI-MGR for subjects carrying a bag and wearing a coat for the view angles 0° to 180° and the rank-1 identification rates of the same probe and gallery view for the view angles 36°, 54°, 90°, 108°, 126° and 144° of CCA for subjects carrying a bag and wearing a coat using GFI as a gait signature. Although probe view is unknown

Table 7: Confusion matrix for CVR of the subjects of CASIA B gait dataset walking wearing a coat.

Actual Class	Predicted class										
	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°
0°	98	0	0	0	0	0	0	0	0	0	26
18°	0	105	10	9	0	0	0	0	0	0	0
36°	0	4	99	10	8	0	0	3	0	0	0
54°	0	0	4	110	8	2	0	0	0	0	0
72°	0	0	4	10	97	12	1	0	0	0	0
90°	0	0	0	3	13	89	17	2	0	0	0
108°	0	0	0	0	10	18	87	6	3	0	0
126°	0	0	0	0	0	0	7	105	12	0	0
144°	0	0	10	0	0	0	0	6	98	10	0
162°	0	4	8	0	0	0	0	0	8	104	0
180°	31	0	0	0	0	0	0	0	0	0	93

in VI-MGR, and VI-MGR does not assume the availability of the exact matching view of the probe in the gallery, it significantly outperforms CCA and JPE-VR even for exactly same views of the probe and gallery subjects, which demonstrates superb efficacy of VI-MGR compared to the state-of-the-art view invariant gait recognition methods.

Table 3 shows that CVR of VI-MGR is outperformed by the method CCA in [21] for most views. It is worth pointing that although CCA benefits from the favourable condition of using the exact matching probe and gallery view for identification and that VI-MGR does not detect the exact matching gallery view of the probe for several cases in phase 1, VI-MGR significantly outperforms CCA in terms of identification rates (see Fig. 14). Thus, the results demonstrate the benefit of phase 2 of VI-MGR which includes multiscale shape analysis and classification using WRSL to achieve invariance to slight variation in view for improved identification rate.

Table 8 shows the rank-1 and rank-5 identification rates produced by VI-MGR using L=300 on CASIA B gait dataset. The better subject identification rate even when the CVR is low is attributed to the robustness of the phase 2 of VI-MGR against limited variation in view. The table shows that VI-MGR provides excellent identification rates, i.e.,100% at rank-1 and rank-5 for most views of the normal walking sequences. However, the identification rate is decreased in the presence of carrying conditions and clothing variation.

Since CGI is evaluated on CASIA B gait dataset using gait sequences captured from 90° only, Table 4 shows the rank-1 identification rates of VI-MGR using L=300 evaluated on 90° view of the gait sequences using same experimental set up as that of CGI for a fair comparison. Note that GEI performs better than CGI when the training and testing conditions are the same, and CGI performs better than GEI when the training and testing conditions are differ-

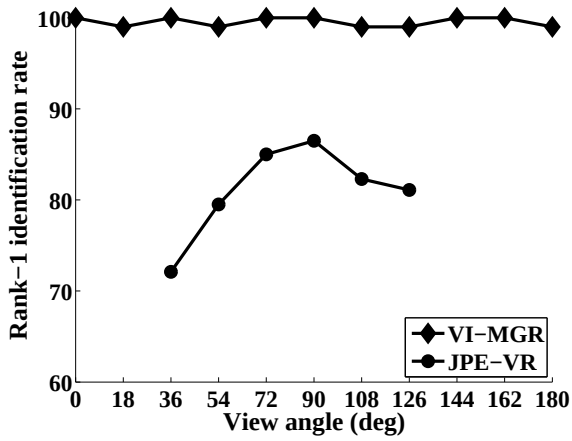


Figure 13: Rank-1 identification rates of VI-MGR and JPE-VR for normal walking subjects of CASIA B gait dataset at different views.

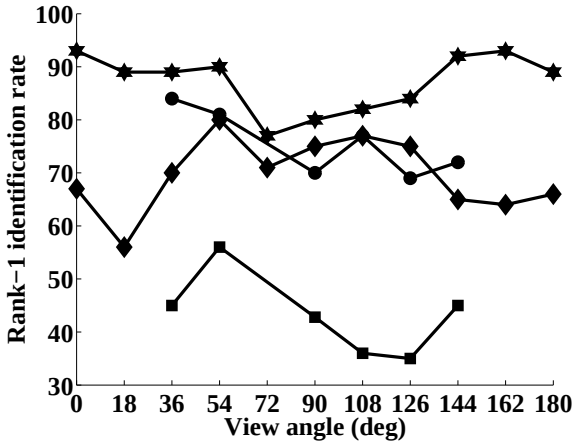


Figure 14: Rank-1 identification rates of VI-MGR and CCA for walking subjects carrying a bag and wearing a coat of CASIA B gait dataset at different views. Keys: '★'-VI-MGR (bag); '●'-CCA (bag); '◆'-VI-MGR (coat); '■'-CCA (coat).

ent [5]. This shows that CGI is more robust to covariates than GEI. However, VI-MGR significantly outperforms CGI and GEI for all experiments, thus showing it is substantially invariant to clothing variation and carrying conditions. Additionally, VI-MGR is also robust to variation in view.

4.3. Experiments on OU-ISIR treadmill gait dataset B

VI-MGR is evaluated on the size-normalised, lateral-viewed GEIs of OU-ISIR treadmill gait dataset B [16] comprising 68 subjects with up to 32 combinations of different types of clothing. The dataset is divided into three subsets, i.e., a training set comprising 446 sequences of 20 subjects with all types of clothes; a gallery set comprising sequences of the remaining 48 subjects with standard clothes (type 9/ regular pants+full shirt); and a probe set comprising 856 sequences for these 48 subjects with other types of clothes excluding the standard clothes. Unlike the method in [16], we do not use the training dataset as it is unrealistic to train a gait recognition system with all possible types

Table 8: Rank-1 and rank-5 identification rates (%) of VI-MGR on CASIA B gait dataset.

View	Identification Rate (%)					
	Normal/Normal		Normal/Coat		Normal/Bag	
	Rank		Rank		Rank	
	1	5	1	5	1	5
0°	100	100	67	80	93	95
18°	99	100	56	77	89	94
36°	100	100	70	83	89	95
54°	99	99	80	95	90	97
72°	100	100	71	95	77	91
90°	100	100	75	93	80	94
108°	99	100	77	94	82	91
126°	99	100	75	88	84	91
144°	100	100	65	76	92	97
162°	100	100	64	75	93	96
180°	99	100	66	73	89	93
Mean	99.5	99.9	69.6	84.5	87.1	94

of clothing combinations. Since no experimental results of clothing invariant gait recognition methods, e.g., [16], are available based on OU-ISIR dataset B in terms of identification rates, we evaluated GEI on OU-ISIR dataset B using gallery and probe sets each comprising 48 subjects to compare with VI-MGR. Fig. 15 shows that VI-MGR significantly outperforms GEI at rank-1 identification rate for all 32 probe items of varying clothing types.

4.4. Computational complexity

The average processing time required to detect the matching gallery view with the probe, extract multiscale shape features of the gallery subjects (124 for CASIA B dataset) of matching probe view (training), extract multiscale shape features of the probe subject and to identify the probe subject using WRSL (L=500) (testing) is 52.82 secs using Matlab 7.11.0 (R2010b) on an Intel (R) Core (TM) i7 processor with 3 GB RAM working at 2.93-GHz for CASIA B gait dataset. Note that the running time of global Gaussian mixture model learning and the average running time of maximum a posteriori adaptation used for two-stage Gabor-PDF feature extraction in GPDF-LGSR (in which 40 Gabor kernel functions from five scales and eight orientations are employed) of a single probe/gallery image are 36702.0 seconds and 3.2 seconds, respectively, when the number of Gaussian components used is 500 [7]. These reported times are based on experiments conducted on an IBM workstation (3.33-GHz CPU with 16-GB RAM) using Matlab. Hence, VI-MGR is significantly less computationally expensive than GPDF- LGSR. Note that the average

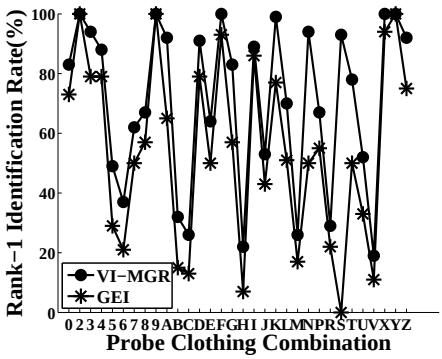


Figure 15: Identification rate at rank 1 for 32 probe items of OU-ISIR dataset B with different clothing combinations using gallery set of subjects with RP+FS.

processing time to obtain only the reconstruction coefficient of a single probe subject for LGSR classifier in GPDF-LGSR using $\lambda = 1/16$ and $\sigma = 1/2$ is 9.22 secs, which is higher than the total time required for matching view detection, feature extraction and classification using WRSL in VI-MGR (note VI-MGR uses a slower processor than GPDF-LGSR). Although LGSR classification method provides better W-AvgI using Gabor-PDF as the input feature for USF HumanID gait challenge dataset, it has not been used in VI-MGR due to its higher computational complexity. Like STS-DM, VI-MGR also uses 2D FFT to compute the DFT of a GEI and the Gaussian filter to reduce the computational complexity. Note that the training time for constructing a view transformation model in [23] requires approximately 10-20 minutes using Quad Processor 2.66 GHz and 4 GB RAM. The computational complexity of VI-MGR increases with the number of scales. Hence, to make a trade-off between identification rate and the computational complexity, a minimum number of three scales are selected for multiscale shape feature extraction.

5. Conclusion

A gait recognition method is mainly affected by the variation in view, clothing types, and presence of a carried item as well as other covariates, e.g., segmentation imperfection, shadow under feet, change in ground surface and occlusion. This paper proposes a two-phase view-invariant multiscale gait recognition method, i.e., VI-MGR, to achieve invariance to all these covariates for identifying a subject in an unconstrained environment. Phase 1 determines which of the available views of the gallery matches most closely with the probe view; and phase 2 compares the probe with the matching view of the gallery subjects for identification.

In phase 1, VI-MGR computes the entropy of the limb region of the GEIs to obtain SGEnIs to enhance dynamic gait characteristics for better discriminability. To determine the matching gallery view of the probe, the SGEnIs of the probe are compared with the SGEnIs of all available views in the gallery using 2D PCA and Euclidean distance classifier. It introduces a R-GEI to create synthetic gallery views to take into account of any unknown probe view in the range 0° and 360°. In phase 2, VI-MGR applies Gaussian filter to the GEIs at three scales to generate MGI as a gait

feature. Focus value is used as a measure of blurriness of the filtered GEIs to determine the ideal range of scales. From this range, three scales are selected based on entropy value analysis of the filtered GEIs which are effective to provide improved identification rate in presence of clothing variation and carrying conditions with reduced computational complexity. A probe subject is classified using WRSL for overfitting avoidance exploiting high dimensionality of the feature space. Excellent identification rates on publicly available datasets demonstrate the efficacy of VI-MGR.

VI-MGR is robust against a wide range of views, but requires the availability of the matching or closely matching probe view in the gallery. Therefore, future work will consider the augmentation of VI-MGR with the extraction of view-invariant gait features or construction of a view transformation model to provide improved identification rate in the absence of matching or closely matching probe view in the gallery. Furthermore, instead of choosing the subspaces randomly for WRSL, future work will also consider to intelligently select the feature subsets in order to capture the most inter-subject discriminatory information by avoiding the adverse impact of the covariate factors.

Acknowledgment

The authors would like to thank Warwick Postgraduate Research Scholarship and the Institute of Advanced Study Early Career Fellowship of the University of Warwick for providing the funds for this research.

References

[1] M. S. Nixon, T. N. Tan, R. Chellappa, Human identification based on gait, Springer Science+Business Media, Inc., New York, USA, 2006.

[2] S. Sarkar, P. J. Philips, Z. Liu, I. Vega, P. Grother, K. Bowyer, The HumanID gait challenge problem: data sets, performance, and analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, 27 (2) (2006) 162-177.

[3] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (2) (2006) 316-322.

[4] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28 (6) (2006) 863-876.

[5] C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan, Human identification using temporal information preserving gait template, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (11) (2012) 2164-2176.

[6] T. H. W. Lam, K. H. Cheung, J. N. K. Liu, Gait flow image: A silhouette-based gait representation for human identification, Pattern Recognition, 44 (2011) 973-987.

[7] D. Xu, Y. Huang, Z. Zeng, X. Xu, Human gait recognition using patch distribution feature and locality-constrained group sparse representation, IEEE Transactions on Image Processing, 21 (1) (2012) 316-326.

[8] S. L. Dockstader, M. J. Berg, A. M. Tekalp, Stochastic kinematic modeling and feature extraction for gait analysis, IEEE Transactions on Image Processing, 12 (8) (2003) 962-976.

[9] H. Lu, K. N. Plataniotis, A. N. Venetsanopoulos, A full-body layered deformable model for automatic model-based gait recognition, EURASIP Journal on Advances in Signal Processing, (2008) 1-13.

[10] F. Tafazzoli, R. Safabakhsh, Model-based human gait recognition using leg and arm movements, Engineering Applications of Artificial Intelligence, 23 (2010) 1237-1246.

[11] Z. Xue, D. Ming, W. Song, B. Wan, S. Jin, Infrared gait recognition based on wavelet transform and support vector machine, Pattern Recognition, 43 (8) (2010) 2904-2910.

[12] C. Direkoglu, M.S. Nixon, Shape classification via image-based multiscale description, Pattern Recognition, 44 (2011) 2134-2146.

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

[14] D. Tao, X. Li, X. Wu, S. J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (10) (2007) 1700-1715.

[15] K. Bashir, T. Xiang, S. Gong, Gait recognition without subject cooperation, Pattern Recognition Letters, 31 (2010) 2052-2060.

[16] M. A. Hossain, Y. Makihara, J. Wang, Y. Yagi, Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control, Pattern Recognition, 43 (2010) 2281-2291.

[17] J. Han, B. Bhanu, A. K. Roy-Chowdhury, A study on view-insensitive gait recognition, in: Proceeding of the International Conference on Image Processing, Genoa, Italy, 2005, pp. 297-300.

[18] M. Goffredo, I. Bouchrika, J. N. Carter, M. S. Nixon, Self-calibrating view-invariant gait biometrics, IEEE Transactions on Systems, Man, and Cybernetics, Part B 40 (4) (2010) 997-1008.

[19] W. Kusakunniran, Q. Wu, J. Zhang, Y. Ma, H. Li, A New View-Invariant Feature for Cross-View Gait Recognition, IEEE Transactions on Information Forensics and Security, (8) (10) (2013) 1642-1653.

[20] F. Jean, R. Bergevin, A.B. Albu, Computing and evaluating view-normalized body part trajectories, Image and Vision Computing, 27 (9) (2009) 1272-1284.

[21] K. Bashir, T. Xiang, S. Gong, Cross-view gait recognition using correlation strength, in: Proceedings of the 21st British Machine Vision Conference, United Kingdom, 2010, pp. 109.1-109.11.

[22] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Gait recognition using a view transformation model in the frequency domain, in: Proceedings of the 9th European Conference on Computer Vision, Graz, Austria, 2006, pp. 151-163.

[23] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Support vector regression for multi-view gait recognition based on local motion feature selection, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, San Francisco, United States, June 2010, pp. 974 -981.

[24] N. Liu, J. Lu, Y. Tan, Joint subspace learning for view-invariant gait recognition, IEEE Signal Processing Letters, 18 (7) (2011) 431-434.

[25] G. Zhao, G. Liu, H. Li, M. Pietikainen, 3D gait recognition using multiple cameras, in: Proceeding of the 7th International Conference on Automatic Face and Gesture Recognition, Southampton, United Kingdom, 2006, pp. 529-534.

[26] S. D. Choudhury, T. Tjahjadi, Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors, Pattern Recognition, 45 (2012) 3414-3426.

[27] S. D. Choudhury, T. Tjahjadi, Gait recognition based on shape and motion analysis of silhouette contours, Computer Vision and Image Understanding, 117 (2013) 1770-1785.

[28] Y. Ran, Q. Zheng, R. Chellappa, T. M. Strat, Applications of a simple characterization of human gait in surveillance, IEEE Transactions on Systems, Man, and Cybernetics, Part B 40 (4) (2010) 1009-1019.

[29] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, K. Moustakas, Gait recognition using compact feature extraction transforms and depth information, IEEE Transactions on Information Forensics and security, 2 (3) (2007) 623-630.

[30] D.A. Winter, Biomechanics and motor control of human movement, 3rd ed., John Wiley & Sons, New Jersey, 2004.

[31] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE Transactions on Image Processing, 11 (4) (2002) 467-476.

[32] X. Huang, S.Z. Li, Y. Wang, Shape localization based on statistical method using extended local binary pattern, in: Proceedings of the IEEE First Symposium on Multi-Agent Security and Survivability, 2004, pp. 184-187.

[33] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (6) (2007) 915-928.

[34] T. Lindeberg, Feature Detection with Automatic Scale Selection, International Journal of Computer Vision, 30 (2) (1998) 79-116.

[35] T. K. Ho, The random subspace method for constructing decision forests, IEEE Transactions on Pattern Analysis and Machine Intelligence,

20 (8) (1998) 832-844.

[36] L. Breiman, Bagging predictors, Machine Learning, 24 (1996) 123-140.

[37] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, 55 (1) (1997) 119-139.

[38] Y. Guan, C.-T. Li, Y. Hu, Random subspace method for gait recognition, in: Proceedings of the IEEE International Conference on Multimedia and Expo Workshop (ICMEW'12), Melbourne, Australia, 2012, pp. 9-13.

[39] J. Yang, D. Zhang, A. F. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26 (1) (2004) 131-137.

[40] Z. Liang, Y. Li, P. Shi, A note on two-dimensional linear discriminant analysis, Patter Recognition Letters, 29 (2008) 2122-2128.

[41] J. M. Tenenbaum, Accommodation in computer vision, Ph.D. Thesis, Stanford University, 1970.

[42] S. K. Nayar, Y. Nakagawa, Shape from Focus, IEEE Transactions on Pattern Analysis and Machine Intelligence, 16 (1994) 824-831.

[43] G. Yang, B.J. Nelson, Wavelet-based autofocusing and unsupervised segmentation of microscopic images, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, Nevada, USA, vol. 3, 2003, pp. 2143-2148.

[44] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, in: Proceedings of the IEEE International Conference on Image Processing, Brussels, Belgium, 2011, pp. 2073-2076.

[45] Y. Makihara, H. Mannami, A. Tsuji, M. A. Hossain, K. Sugiura, A. Mori, Y. Yagi, The OU-ISIR gait database comprising the treadmill dataset, IPSJ Transactions on Computer Vision and Applications, Technical Note, 4 (2012) 53-62.

[46] J. Han, B. Bhanu, Statistical feature fusion for gait-based human recognition, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 842-847.

[47] P. Sanguansat, W. Asdornwised, S. Marukatat, S. Jitapunkul, Two-dimensional random subspace analysis for face recognition, in: Proceedings of the IEEE International Symposium on Communications and Information Technologies, Sydney, Australia, 2007, pp. 628-631.

[48] S. V. Stehman, Selecting and interpreting measures of thematic classification accuracy, Remote Sensing of Environment, 62 (1) (1997) 77-89.

[49] N. Liu, Y-P Tan, View invariant gait recognition, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, USA, 2010, pp. 1410-1413.