



Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors

Sruti Das Choudhury, Tardi Tjahjadi *

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

ARTICLE INFO

Article history:

Received 6 September 2011
Received in revised form
15 December 2011
Accepted 21 February 2012
Available online 5 March 2012

Keywords:

Gait recognition
Human identification
Procrustes shape analysis
Elliptic Fourier descriptor
Silhouette
Nearest neighbour classifier
Classifier combination
Hu moments

ABSTRACT

This paper presents a gait recognition method which combines spatio-temporal motion characteristics, statistical and physical parameters (referred to as STM–SPP) of a human subject for its classification by analysing shape of the subject's silhouette contours using Procrustes shape analysis (PSA) and elliptic Fourier descriptors (EFDs). STM–SPP uses spatio-temporal gait characteristics and physical parameters of human body to resolve similar dissimilarity scores between probe and gallery sequences obtained by PSA. A part-based shape analysis using EFDs is also introduced to achieve robustness against carrying conditions. The classification results by PSA and EFDs are combined, resolving tie in ranking using contour matching based on Hu moments. Experimental results show STM–SPP outperforms several silhouette-based gait recognition methods.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Biometrics has emerged as a reliable means of identifying a human subject based on the subject's distinctive biological features. Physiological biometrics such as face, fingerprints or iris pattern generally require cooperation from the subject for a particular view, physical contact or proximity [1]. Behavioural biometrics examines human behaviour and the most promising example is gait which exploits a subject's distinctive way of walking to perform identification without interfering with the subject's activity [2,3]. Gait has the potential to identify a human subject at a distance using low resolution video sequences, when physiological biometrics are not perceivable [4]. Furthermore it is difficult for a subject to conceal and disguise his/her gait characteristics [5,3]. However, variations of the subject's clothing, footwear and hair styles over different days bring challenges to gait recognition, and the subject's physical and mental conditions, e.g., leg injury, carrying conditions, drunkenness, illness, fatigue, pregnancy, etc., distort the walking patterns [1]. Gait recognition methods tend to be not robust due to the presence of occlusions in the scene, and like most other biometrics a subject's gait characteristics change with his/her age. Despite these challenges,

gait has contributed to many potential applications in the field of visual surveillance, access control, forensics, biometric authentication and criminology (e.g. [6,7]). Approaches to gait recognition can be broadly classified into two categories: model-based and silhouette-based.

Model-based gait recognition methods (e.g. [7–9,5]) characterise a walking subject by a structural model and a motion model. The structural model usually represents the subject by a 2-dimensional (2D) contour, a stick figure or a volumetric model based on the proportions of the human body parts, and measures time-varying gait parameters such as gait period, stance width and stride length to obtain the gait signatures. The motion model incorporates the kinematical and dynamical motion parameters of the subject, e.g., rotation patterns of hip and thigh, joint angle trajectories and orientation changes of limbs. These methods can reliably deal with self-occlusions and occlusions caused by the presence of other objects in the scene. They are invariant to scale changes, rotational effects and slight variations in viewpoint. Also, they are robust to noise. However, they are characterised by complex searching and mapping processes which increase the size of their feature space and computational cost.

Silhouette-based gait recognition methods do not assume an explicit model of the human body, but analyse the spatio-temporal shape and motion characteristics of silhouettes. Spatio-temporal motion-based methods (e.g. [1,3,10–14]) capture both the spatial structural and temporal transitional characteristics of

* Corresponding author. Tel.: +44 24 76523126; fax: +44 24 76418922.
E-mail addresses: t.tjahjadi@warwick.ac.uk, tt@eng.warwick.ac.uk (T. Tjahjadi).

gait. These methods are easy to implement with low computational complexity and simplified feature space. However, they are susceptible to variation in camera view and walking speed. Statistical methods (e.g. [2,15–17]) usually describe silhouettes using shape and motion descriptors such as velocity moments [15], Zernike velocity moments [16] and Procrustes mean shape distance [2]. Eigenspace transformation and canonical space analysis are widely used in these methods to reduce the dimensionality of input feature space and optimise class discrimination. Statistical methods are more resilient to noise [6]. The physical parameter based methods (e.g. [18,19]) estimate the subject's geometrical and structural properties, e.g., step length, cadence and height. These methods are robust against lighting variations and segmentation imperfections. However, they require improved techniques for camera calibration, body-part labelling and depth compensation [6].

Since the subject's shape provides more significant information than its kinematics in gait analysis for human identification [20], we propose STM-SPP which analyses the shape of silhouette contours using Procrustes shape analysis (PSA) and elliptic Fourier descriptors (EFDs) [21]. Since the performance of gait recognition algorithm increases with the number of appropriate gait signatures considered, we aim to effectively combine spatio-temporal gait characteristics, statistical and physical parameters of human body for improved classification rate with reduced computational complexity using simplified feature space. The method is defined on the lateral (i.e., profile) view of silhouettes of a subject walking parallel to the image plane, as most of the significant gait characteristics are captured in this view. Instead of analysing the silhouette, only its boundary (i.e., contour) is considered to reduce the computational complexity of STM-SPP.

Shape analysis by comparing landmarks has been applied in diverse fields [22], e.g., study of shape differences of brains to identify schizophrenic patients, handwriting recognition, fish recognition, robotic harvesting of mushrooms, and study of shape and size variability of microfossils, using traditional or geometrical methods. The traditional shape analysis methods either examine the ratios of distances or the angles between landmarks, whereas the geometrical methods analyse coordinates of the landmarks. One of the major motivations for STM-SPP is to demonstrate the potential of traditional shape analysis for identifying human subjects based on their gait signature.

With regards to a subject's identity, the temporal changes of the subject's shape in a gait sequence provides better discriminative power than the discrete snapshots of images, but it increases the computational complexity. The performance of any shape-based gait recognition method degrades due to variations in hair styles, clothing and footwear over different days, as these factors distort the silhouette shape. The shape of a silhouette is also significantly altered due to carrying conditions. If either of the gallery or probe subject carries any item, certain parts of the silhouette shape are likely to change and the discriminative ability of the shape-based gait recognition algorithm decreases with respect to these parts.

STM-SPP extends the application of traditional shape analysis method in gait recognition by comparing distances of specific landmarks from the centre of mass of the subject's silhouette contour (COM-SC) using PSA. The purpose of using these distances rather than coordinates of the landmarks is to reduce the dimensionality of the feature space from 2D to 1-dimensional (1D). STM-SPP provides an efficient means of obtaining dissimilarity score between gallery and probe sequences using PSA and EFDs for subject classification. The method validates similar dissimilarity scores obtained by PSA using spatio-temporal gait characteristics and physical parameters of human body to enhance classification rate in the presence of across-day gait

variations (e.g., walking speed, different types of clothes and footwear, and change of hair styles). The method also provides an experimentally supported insight into the detection procedure of small carried items based on anatomical studies of human body and introduces a part-based EFD analysis to achieve robustness against shape variations due to carrying conditions among the probe and gallery subjects (as detailed in Section 3.3). To utilise the benefit of shape sequence processing with reduced processing time, STM-SPP characterises the subject's contours using EFDs at specific phases of a gait period to considerably reduce computational complexity as well as to achieve robustness against walking speed variations and missing or distorted frames (as detailed in Section 3.3). The output of the two classifiers (PSA and EFDs) combined using rank-summation based combination rule is used for identifying the subject, with the application of hierarchical contour matching based on Hu moments [23] to resolve the case of two classes with the same combined rank.

The paper is organised as follows. Section 2 discusses related work and Section 3 presents STM-SPP. The experimental results are analysed in Section 4 and Section 5 concludes the paper.

2. Related work

The silhouette based methods have advanced gait analysis for human identification. The correspondence-free, view-dependent method in [24] obtains 2D gait signatures from the 3-dimensional (3D) volume which encloses the walking subject. The signature consists of image self-similarity plot (SSP) which is defined as the absolute correlation of each pair of images in two gait sequences of a walking subject. The SSPs of two different gait sequences are normalised following the determination of frequency and phase of gait, and compared using pattern classification techniques for human identification. Spectral partitioning is performed in [25] and human identification is achieved using weighted correlation and median weighted distance.

The correspondence-free method in [18] computes the gait period of a subject by analysing width of the bounding box which encloses the subject's moving silhouette and uses Bayesian classifier to confirm the subject's identity. However, silhouette width is not effective for computing gait period of the frontal view of a moving subject. Hence, temporal change of silhouette height is also considered in [19] to achieve robustness against different views. This method performs template matching between key frames of the gallery and probe gait sequences using normalised correlation to obtain correlation scores, and uses nearest neighbour classifier (NNC) for gait recognition. The method in [14] converts a binary silhouette into 1D normalised distance signal by contour unwrapping with respect to the silhouette centroid. Principal Component Analysis (PCA) is then used to reduce the dimensionality of the feature space and to obtain projection centroids corresponding to each gallery sequence in the eigenspace. Finally, NNC and NNC with respect to class exemplars are used for identification. The identification is validated based on the subject's physical parameters for increased accuracy. The method in [26] uses 3D radial silhouette distribution transform and 3D geodesic silhouette distribution transform for assigning depth information to the sequence of silhouettes. A genetic algorithm is used to combine features extracted using three different feature extractors, namely, radial integration transform, circular integration transform and weighted Krawtchouk moments for gait recognition.

Gait Energy Image (GEI) [1] which contains spatio-temporal motion information of a gait period, is computed from binary silhouettes. Real gait templates are computed from each gait period and distorted to generate synthetic gait templates.

Component and discriminant analyses are then performed on the templates for dimensionality reduction. The real and synthetic gait features thus obtained are combined using a feature fusion strategy for improved identification performance. This approach is not only computationally efficient and consumes less storage space, but also robust against noise. The gait recognition method in [27] computes Gait Entropy Image from a sequence of images of a subject's gait period to identify noncooperating individuals in unconstrained environment with varying covariate conditions for gallery and probe sequences. The baseline method [13] computes gait period by counting the number of foreground pixels mostly from the legs and performs subject identification using spatio-temporal correlation of silhouettes. The method in [28] captures temporal information of the gait sequence into a single colour image called Chrono-Gait Image, and the method in [29] computes gait flow image by determining the optical flow field from sequence of silhouettes for gait recognition. The feature extraction process in [12] involves the computation of angular distance between the foreground pixels and the centre of the silhouette. The method uses linear time normalisation to determine the subject's identity. The different clothes worn by a subject in the gallery and probe sequences change specific parts of silhouette's shape and thus complicate the identification of the subject. Therefore, the method in [30] applies a part-based feature extraction strategy to achieve substantial clothing invariance and uses an adaptive weight control mechanism to identify the subject.

The use of general tensor discriminant analysis for gait recognition in [31] overcomes the undersampling problem of conventional linear discriminant analysis (LDA), while preserving discriminative information of the gallery tensors. The use of GaborD, GaborS and GaborSD based averaged gait image representations in [31] considerably reduces the computational complexity of Gabor based image representations. The method in [32] normalises gait dynamics using population Hidden Markov Model, whose states are specific stance phases of a gait period determined by Viterbi algorithm, and computes shape distances between stance phases using LDA to maximise inter-class and minimise intra-class variations of the subjects for human identification. The method in [33] considers gait sequences as a third-order tensor and uses an EigenTensorGait obtained by multilinear PCA. The method in [34] recognises the periodic movements of a human subject using motion power spectral analysis of the Fourier coefficients of unstructured feature-point kinematic data acquired from a marker-based 3D optical motion capture system.

The method in [35] represents deforming shape sequences of human subjects over a gait period by 2D discrete Fourier series and uses the resulting magnitude spectra to form the gait signatures. It uses Bhattacharya distance metric for maximising inter-class separation to achieve improved classification rate and finally uses k -nearest neighbour classifier (k -NNC) for human identification. Motivated by the encouraging identification rate of this method, STM-SPP analyses 10 specific image frames of a gait period with EFDs to form gait signatures for subject classification.

STM-SPP compares favourably with the method in [20] (referred to as VCR-C) and the method in [2] that also use PSA for gait recognition. VCR-C uses both parametric and non-parametric methods to compare deforming silhouettes for human identification and activity recognition. Human shapes are described as k -dimensional complex vectors, and Procrustes distance metrics are used to compute distances between the shape sequences. Although we consider the distances of specific landmarks from COM-SC to compare human shapes at specific phases of a gait period to simplify the feature space, a specific frame-wise comparison may not always produce satisfactory identification due to the distortions of silhouettes, e.g., caused by partial occlusions. We thus characterise the contour of a

silhouette at specific phases of a gait period by translation-rotation-scale invariant EFDs and perform subject classification based on a dissimilarity score to exploit the benefit of shape sequence processing without compromising the simplicity in implementation and simplification of feature space.

The PSA in [2] describes the boundary points of the silhouette as a vector of k complex numbers in 2D shape space called a configuration and compares two such configurations for measuring similarity using Procrustes mean shape distance. To ensure the same set of boundary points in different images are used for comparison, an interpolation technique with point correspondence analysis is used. We consider 28 landmarks based on anatomical and geometric properties of human body on the silhouette contour. Instead of considering the coordinates of landmarks for comparing shapes, we use the distances of the landmarks from COM-SC which correspond to the different rows of the uni-columnar configuration matrix. Identification is achieved by analysing the average dissimilarity score obtained by comparing the probe configuration matrix with the gallery configuration matrices of multiple sequences of same subjects from the gallery data set.

The method in [2] employs classifier combination rules to combine the static and dynamic gait characteristics obtained, respectively, by analysing silhouettes' shapes and joint angle trajectories of lower limbs. STM-SPP combines the classification results obtained by PSA and EFDs of silhouette contours using rank-summation based classifier combination rule for its simplicity and effectiveness.

3. STM-SPP

The proposed method, STM-SPP, comprises three modules: (1) extraction and postprocessing of a subject's silhouettes; (2) classification of the subject using PSA (phase 1) and using EFDs (phase 2); and (3) combining the two classification results.

3.1. Module 1: silhouette extraction and postprocessing

Silhouette extraction involves segmenting region/s that correspond to a walking subject in a cluttered scene. STM-SPP employs background modelling and moving object segmentation in [36,37], where background is considered to be any static or periodically moving parts of a scene that remains static or periodic over the period of interest. The segmented regions are smoothed using Gaussian filter and subjected to connected-component analysis involving morphological operation of dilation to remove noisy pixels and followed by erosion to fill up any small holes inside the silhouette to give a single connected region. The smoothed segmented region is then tracked based on the overlap of the centroid of the bounding rectangle which encloses the region in the subsequent frames as in [37]. The process is illustrated in Fig. 1(a)–(d). The tracked segmented region is binarised using 2D Otsu automatic thresholding technique [38], which utilises both the grey level information of each pixel and its spatial correlation information within the 2D neighbourhood to outperform the Otsu method [39] in the presence of noise for extracting the subject's silhouette as illustrated in Fig. 1(e). The extreme outer boundary of the largest connected component, i.e., silhouette contour as shown in Fig. 1(f) is obtained using the sequence of vertices traversal algorithm based on connectivity [37].

3.2. Phase 1 of module 2: PSA

The successful identification of a subject should not depend on how far the subject is from the camera and the direction of

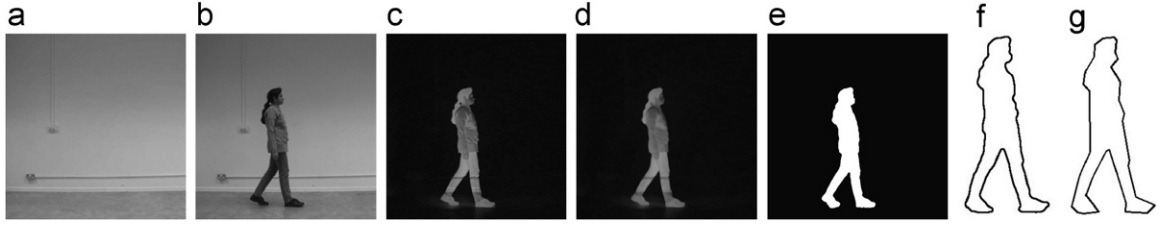


Fig. 1. Extraction of a moving silhouette: (a) background image; (b) original image; (c) segmented regions; (d) smoothed segmented region; (e) binary silhouette; (f) silhouette contour; and (g) contour after polygon approximation.

walking. Thus, the shape feature vector used for identification must be invariant to scale, translation and rotation. One means of achieving this is through PSA, which involves analysing the distribution of a set of shapes by matching configurations (where each configuration is a set of geometric locations of landmarks of a shape) to calculate the best shape-preserving Euclidean transformations, using least squares techniques. The 2D Cartesian moment of order p and q of a contour $I(x,y)$ is

$$m_{p,q} = \sum_{i=1}^N I(x_i,y_i) x_i^p y_i^q, \quad (1)$$

where N is the total number of pixels in the contour. The coordinates of COM-SC, (x_c, y_c) , is given by the ratio of first-order to zero-order contour moments [40], i.e.,

$$x_c = \frac{m_{10}}{m_{00}}, \quad y_c = \frac{m_{01}}{m_{00}}. \quad (2)$$

We assign anatomical landmarks by considering the different positions of the human body joints as a fraction of subject's height (H), obtained by measuring the height of the bounding rectangle which encloses the silhouette contour. The vertical positions of ankle (A), knee (K), hip (HI), chest (C), shoulder (SH) and head (HD) are then estimated as a fraction of the body height following anatomical studies in [41] as $0.039H$, $0.285H$, $0.530H$, $0.720H$, $0.818H$ and $0.870H$ measured from the bottom of the bounding rectangle, respectively. The boundary points of the contour that correspond to A , K , HI , C , SH and HD are located (labelled as 11, 12, 17, 19, 9, 14, 15, 21, 7, 23, 5, 25, 4, 26, 3 and 27 in Fig. 2) and are treated as anatomical landmarks.

Two mathematical landmarks are also defined as the end points of the contour diameter (Diam) joining two farthest boundary points (labelled as 1 and 18 in Fig. 2), i.e. [42]

$$\text{Diam}(\text{contour}) = \max [\text{Dist}(q_i, q_j)], \quad (3)$$

where $\text{Dist}(\cdot, \cdot)$ computes the distance between two boundary points q_i and q_j . For better results, 10 additional pseudo landmarks are also considered (labelled as 2, 28, 6, 24, 8, 22, 10, 13, 16 and 20 in Fig. 2), each of which is equi-distantly spaced between the anatomical landmarks. The 28 landmarks are traversed in anticlockwise direction starting from landmark 1 with respect to COM-SC, in the double support phase of the gait period when both feet are almost flat on the ground and farthest from each other as shown in Fig. 2, resulting in the maximum width of the bounding rectangle of the contour. Therefore, the width of the bounding rectangle in each frame of a gait period is measured in terms of pixel units and the frame which corresponds to the maximum width of the bounding rectangle is considered as the subject's double support phase.

The presence of shadows under feet distort the contours near the feet and thus bring challenges to the estimation of landmarks at the toe and ankle. Therefore, in the case of data sets containing shadows under feet, STM-SPP encloses the silhouettes using a bounding rectangle and estimate the region of interest (ROI) from the bounding rectangle having identical width but slightly

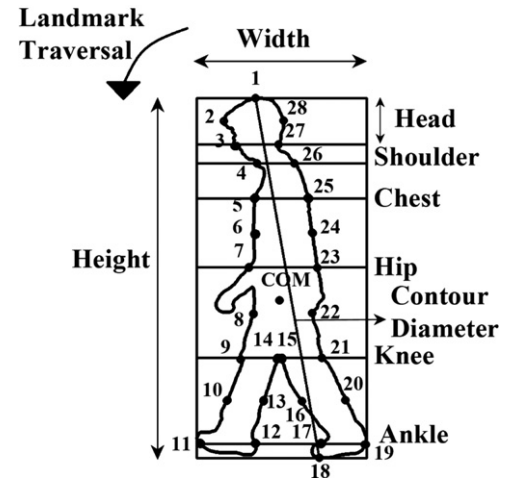


Fig. 2. Vertical positions of ankle (A), knee (K), hip (HI), chest (C), shoulder (SH) and head (HD) as a fraction of body height. Positions of COM-SC, landmarks 1–28, contour diameter and direction of landmark traversal for one subject from CMU MoBo data set.

reduced height αH to discard the shadows under feet, where H is the height of the bounding rectangle and α is a fraction. The value of α is experimentally set to be 0.9375 for USF data set [28] and $0.990H$ for CMU MoBo data set. The estimated ROI is then copied to a destination image of fixed height, i.e., height-normalised, for all the subjects of CMU data set to remove camera depth variations. However, the silhouettes provided by the USF Human1D data set are already cropped, centre-aligned and normalised to a fixed size 128×88 . Thus, for the USF data set, we do not need to perform any normalisation of a silhouette after cropping its height to remove the shadows.

The distance dl_i between each landmark (x_i, y_i) and COM-SC is given by

$$dl_i = [(x - x_i)^2 + (y - y_i)^2]^{1/2}. \quad (4)$$

These distance values as a function of equally spaced monotonically increasing positions along contours form the gallery and probe shape signals, and are labelled with integers that correspond to landmarks (denoted by solid circles) in Fig. 3(b) and (c). The dimension of the configuration matrices corresponding to gallery and probe sequences is $k \times m$, where $k=28$ is the number of landmarks and $m=1$ is the number of dimensions of the landmarks. PSA is used to measure the dissimilarity between two such configuration matrices to achieve view, rotation and scale invariance. Firstly, the gallery and probe configuration matrices (S_1 and S_2 , respectively) are centred using [22]

$$S_{1c} = CS_1, \quad S_{2c} = CS_2, \quad (5)$$

where centring matrix $C = I_k - (1/k)\mathbf{1}_k\mathbf{1}_k^T$, I_k is a $k \times k$ identity matrix, $\mathbf{1}_k$ is a k -dimensional vector of ones, and T is the transpose operator. The centred probe configuration matrix S_{2c} is then

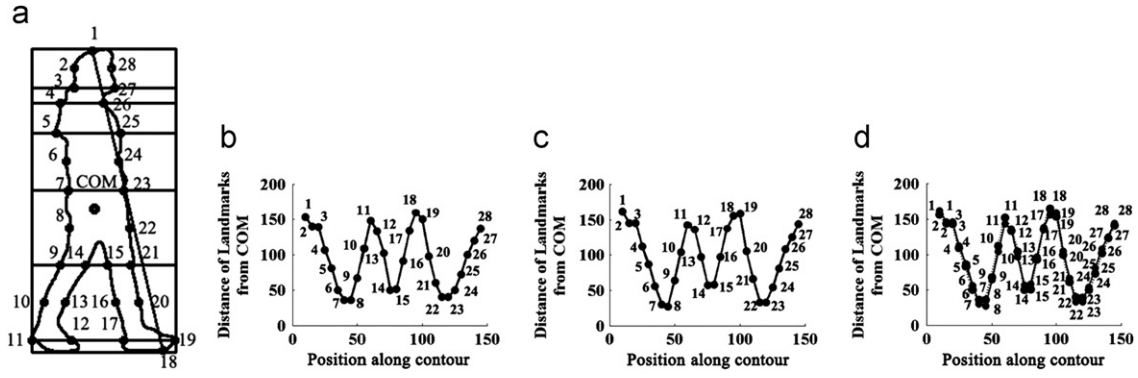


Fig. 3. Silhouette representation: (a) positions of COM-SC and landmarks 1–28; (b) and (c) are, respectively, probe and gallery shape signals consisting of distances of landmarks from COM-SC in anti-clockwise direction; (d) superimposition of the transformed probe shape signal (dashed line) on the gallery shape signal (bold line) for visualising differences between landmark distances from COMs of the contours of probe and gallery sequences of the same subject.

subjected to PSA to be transformed using a combination of translation, scaling and rotation operations to give the transformed probe configuration matrix, i.e. [22]

$$\mathbf{Y} = \{\beta \mathbf{S}_{2c} \Gamma + \mathbf{1}_k \gamma^T : \beta \in \mathbb{R}^+, \Gamma \in \mathbf{SO}(m), \gamma \in \mathbb{R}^m\}, \quad (6)$$

where $\beta \in \mathbb{R}^+$ is scale, Γ is an $(m \times m)$ rotation matrix, $\mathbf{SO}(m)$ is the special orthogonal group of $(m \times m)$ rotation matrices and γ is an $(m \times 1)$ translation m -vector. The similarity parameters γ , Γ and β are estimated by minimising the squared Euclidean distance [22], i.e.,

$$D_{\text{OPA}}^2(\mathbf{S}_{1c}, \mathbf{S}_{2c}) = \|\mathbf{S}_{1c} - \beta \mathbf{S}_{2c} \Gamma - \mathbf{1}_k \gamma^T\|^2, \quad (7)$$

where OPA stands for Ordinary Procrustes Analysis, $\|\mathbf{S}\| = [\text{trace}(\mathbf{S}^T \mathbf{S})]^{1/2}$ is the Euclidean norm. The rows of the transformed matrix contain the transformed values of landmark distances from COM-SC. These distance values are used to form the transformed probe shape signal (represented by dashed line in Fig. 3(d)) and is superimposed on the gallery shape signal to visualise the differences between corresponding positions of landmark-distances from COM-SC in the two shape signals. The transformed probe configuration matrix \mathbf{Y} is then compared with the centred gallery configuration matrix \mathbf{S}_{1c} to obtain dissimilarity score D_{PSA} between them using

$$D_{\text{PSA}} = \frac{\sum_{i=1}^k \sum_{j=1}^m (\mathbf{S}_{1cij} - \mathbf{Y}_{ij})^2}{\sum_{i=1}^k \sum_{j=1}^m (\mathbf{S}_{1cij} - \mathbf{A}_j)^2}, \quad (8)$$

where \mathbf{A} is a row vector whose each element is the mean value of the elements of the corresponding columns of \mathbf{S}_{1c} . The range of D_{PSA} is $[0, 1]$ and denotes the difference between the gallery shape and probe shape, the larger the value the more dissimilar are the two shapes.

We obtain dissimilarity scores by comparing the transformed probe configuration matrix with the centred gallery configuration matrix of every sequence of a particular subject at the double support phase of a gait period, and computing the average of the dissimilarity scores. This average dissimilarity score is then used for classification. The gallery class with whose sequences the probe sequence obtains the lowest average dissimilarity score is assigned rank 1, the second lowest average dissimilarity score is assigned rank 2, and so on. In this way each class receives a ranking based on the dissimilarity score obtained by PSA. In cases where a probe subject generates similar dissimilarity scores with two different gallery subjects, STM-SPP validates the two scores with the spatio-temporal gait characteristics (gait period and step length) and physical parameters of human body (build and compactness) to determine the class with the lower rank for classifying the subject (see Sections 3.2.1–3.2.3).

3.2.1. Estimation of gait period and step length

A step is the motion between successive heel strikes of a subject's opposite feet and that a gait period consists of two steps. The method in [19] computes gait period by considering the change in width and height of the bounding rectangle which encloses a moving silhouette contour in the lateral and frontal views, respectively, and the method in [18] performs autocorrelation of the bounding box width of a series of consecutive silhouettes for view-invariant gait recognition. We use the method similar to that used in [28] to compute gait period as it is robust to shadows under feet and carrying conditions, thereby outperforming the gait period estimation method used in the baseline method [13] for near fronto-parallel view of subjects for USF HumanID gait challenge data set. The average width W of the leg region enclosed between aH and bH in frame I is given by

$$W = \frac{1}{aH - bH} \sum_{i=aH}^{bH} (d_i), \quad 0 \leq a \leq b \leq 1, \quad (9)$$

where the values of a and b are, respectively, chosen to be 0.750 and 0.9375 for USF data set to reduce the effects of shadows under feet and carrying condition, and d_i is the Euclidean distance between the leftmost foreground pixel (l_x, l_y) and the rightmost foreground pixel (r_x, r_y) on the i th line, i.e.,

$$d_i = [(l_x - r_x)^2 + (l_y - r_y)^2]^{1/2}. \quad (10)$$

It results in a periodic signal with distinct peaks and valleys that, respectively, correspond to the expansion and contraction of the bounding rectangle as the subject's legs extend and come back together during a gait period. The gait period is estimated as the average distance (in terms of number of frames) between each pair of consecutive valleys or peaks.

Differences in walking speed of the same subject in different gait sequences result in variations in the gait period, i.e., if a subject walks slowly in a certain situation and more quickly in another, the gait period will comprise different number of constituent frames. Thus, to detect similarities in walking patterns of the same subject in different video sequences collected over different days, Dynamic Time Warping (DTW) [43,4] is applied to account for variation in human movement. DTW uses dynamic programming to compute a warping function that optimally aligns two time-dependent sequences of variable lengths for measuring similarity under certain restrictions. Given two sequences of the same subject $E = (W_{E1}, W_{E2}, \dots, W_{EM})$ and $F = (W_{F1}, W_{F2}, \dots, W_{FN})$ of respective lengths $M \in \mathbb{N}$ and $N \in \mathbb{N}$, and W_{Ei} and W_{Fj} are the respective average width of leg region (as given by Eq. (9)) of their elements, DTW constructs an $M \times N$

matrix of Euclidean distances between corresponding widths, i.e.,

$$d(W_{Ei}, W_{Fj}) = (W_{Ei} - W_{Fj})^2. \quad (11)$$

An $M \times N$ warping path is a sequence $p = (p_1, p_2, \dots, p_L)$ with $p_l = (m_l, n_l) \in [1 : M] \times [1 : N]$ for $l \in [1 : L]$ for mapping two sequences which satisfies the followings: (a) boundary condition: $p_1 = (1, 1)$ and $p_L = (M, N)$; (b) monotonicity condition: $m_1 \leq m_2 \leq m_3 \leq \dots \leq m_L$ and $n_1 \leq n_2 \leq n_3 \leq \dots \leq n_L$; and (c) step size condition: $p_{l+1} - p_l \in (1, 0), (0, 1), (1, 1)$ for $l \in [1 : L-1]$. DTW minimises the cost of warping E and F together, i.e.,

$$\text{DTW}(E, F) = \min \left(\frac{(\sum_{l=1}^L p_l)^{1/2}}{L} \right). \quad (12)$$

The similarity between sequences E and F is measured using Eq. (12) to determine gait period of the same subject.

The step length (SL) is the longitudinal distance between two feet when they are maximally apart in a gait period. It is measured as the width of the bounding rectangle enclosing the silhouette contour at the double support phase of a gait period only for the lateral view of the silhouettes to essentially remove the foreshortening effects due to different views.

3.2.2. Estimation of physical parameters

We estimate the physical parameters of the height-normalised and centre-aligned silhouettes of the CMU MoBo [44] and USF HumanID [13] data sets in lateral views and at the double support phase of the gait period. The compactness (CM) of the contour is estimated in terms of its *perimeter* and *area* of the silhouette as $(\text{perimeter})^2 / \text{area}$. We use CM because it is dimensionless and is thus scale invariant. It is also invariant to orientation and thus acts as a good region descriptor [42]. Build (B), being a ratio of subject's chest width to subject's height is used to differentiate between thin and fat subjects, and between tall and short subjects. These parameters are combined with the gait period (G) and subject's step length (SL) to form a 4-dimensional vector $\langle G, SL, CM, B \rangle$ for each sequence of the same subject. Note that this is unlike the method in [14] which uses $\langle G, SL, H, B \rangle$ to validate the similarity scores between projection centroids of two gait sequences obtained by using normalised Euclidean distance. The measurement values of the physical parameters and spatio-temporal gait characteristics for different sequences of all subjects are stored to form the gallery database.

3.2.3. Validation using NNC

We use NNC due to its simplicity and ease of implementation [40] to validate similar dissimilarity scores generated by a probe sequence with two different gallery sequences to determine the lower rank of class to classify the probe sequence. We compute the sum of the Euclidean distances for all four measurements of $\langle G, SL, CM, B \rangle$ between the probe sequence (K_i) and each of the two gallery sequences (T_i), i.e.,

$$d = \left(\sum_{i=1}^4 (T_i - K_i)^2 \right)^{1/2}. \quad (13)$$

The gallery class whose sequence is the nearest neighbour of the probe, i.e., gives the smaller d is selected as the correct class for the probe and is assigned the lower of the two rankings determined by PSA.

3.3. Phase 2 of module 2: shape characterisation using EFDs

We characterise the subject's contours at specific phases of its gait period using EFDs, as EFDs are capable of describing complex contours (i.e., straight lines emanating from the geometrical centre of the contour intersect the contour more than once) with

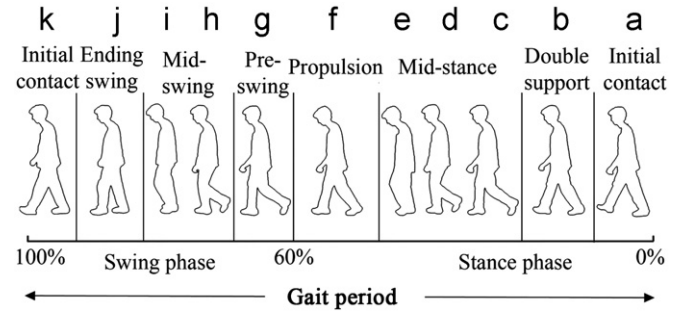


Fig. 4. The specific phases of gait period (a)–(k): stance phase (a)–(f) comprising about 60% of the gait period; and swing phase (g)–(j), comprising remaining 40% of the gait period for a subject from CMU MoBo data set.

any degree of irregularity more accurately but using fewer coefficients than the classical Fourier descriptors [45]. Thus, EFDs serve as a data compression tool which reduces space complexity.

A gait period comprises periodic alternating movement of the lower limbs resulting in a forward movement of the body. It starts with the heel strike of either foot and continues until the heel of the same foot touches the ground again. Each foot in a gait period alternates between two phases, a stance phase and a swing phase which, respectively, constitute 60% and 40% of the gait period as illustrated in Fig. 4. In the stance phase the foot remains in contact with the ground, while in the swing phase the foot is in the air. The stance phase begins with initial contact of the heel with the ground and ends with the toe lift off the same foot from the ground. This phase has the following components: (a) initial contact when the heel of the forward foot (i.e., the foot making a forward movement) touches the ground; (b) mid-stance when the foot is positioned flat on the ground carrying the weight of the body while the other foot is in swing phase; and (c) propulsion which begins with lifting of the heel from the ground and ends with the toe lifting off the same foot indicating the termination of the stance phase. The swing phase begins with lifting of the foot from the ground and continues until the heel of the same foot touches the ground. This phase has the following three components: (a) pre-swing which begins with the toe off the ground and continues until the occurrence of maximum knee flexion; (b) mid-swing, i.e., the motion between maximum knee flexion and when the tibia is vertical to the ground; and (c) ending swing which starts from the vertical position of the tibia and continues until just prior to the forward foot making initial contact with the ground.

We captured the video sequence of a subject (Fig. 1(a)) walking laterally to the image plane in a stationary indoor background using a digital camera (Nikon Coolpix S3000) fixed on a tripod at a rate of 30 frames/s. After estimating the subject's gait period, we obtain its 10 specific phases by visually analysing the constituent frames of the gait period and extract the corresponding contours. The portion of the contours enclosed in the region between the bottom of the bounding rectangle and up to the anatomical position of just before the hand of an upright human subject (i.e., $0.377H$ measured from the bottom of the bounding rectangle [41]) at the 10 specific phases are set as the Reference Region-Of-Interests (Rf-ROI's), as this portion of contour remain unaffected by the self-occlusions caused by arm-swing. To obtain the 10 specific phases of any gait sequence automatically, the Rf-ROI's are compared one at a time with the same portion of contours of all the frames of a subject's gait period (referred to as the Target Region-Of-Interests (Tr-ROI's)) using contour matching based on Hu moments to obtain similarity scores (S_{score}) using [37]

$$S_{score} = \sum_{i=1}^7 \left| \frac{1}{m_i^{Rf}} - \frac{1}{m_i^{Tr}} \right|, \quad (14)$$

where $m_i^{Rf} = \text{sign}(h_i^{Rf}) \cdot \log|h_i^{Rf}|$ and $m_i^{Tr} = \text{sign}(h_i^{Tr}) \cdot \log|h_i^{Tr}|$. h_i^{Rf} and h_i^{Tr} denote the Hu moments of the Rf-ROI and Tr-ROI, respectively. The frame whose Tr-ROI results in the lowest S_{score} with the Rf-ROI, is extracted as one of the 10 specific phases of the gait period and the process is continued by comparing the next Rf-ROI with the remaining Tr-ROI's until all the 10 specific phases are obtained.

Analysing a subject's shape at the specific phases of a gait period enables STM-SPP not only to reduce computational complexity, but also to considerably overcome the adverse effect of walking speed variations under different circumstances, e.g., due to the subject's mood changes. Furthermore, the extracted contours can be considerably distorted by the presence of occlusions in the scene, severe shadows under feet and extreme lighting variations. If these distorted contours are not part of any of the 10 specific phases of the gait period, they can be discarded in the case of shape analysis using EFDs, without having any effect on the classification rate. However, if any of the 10 specific phases is missing including the double support phase, its immediate adjacent frame is considered. In this way, STM-SPP achieves robustness to missing or distorted frames to some extent.

Since a silhouette contour is a closed curve, it can be expressed by a periodic signal $c(t)$ of period T , i.e.,

$$c(t+T) = c(t), \quad (15)$$

where T is the perimeter of the contour. We consider 10 contours corresponding to the 10 specific phases of the gait period as shown in Fig. 4 to obtain EFDs for the gait signatures. To ensure similar set of equal number of points along the selected 10 contours, each contour is approximated by $m = 2^7$ i.e., 128 points using interpolation based on point correspondence analysis [35].

Each contour with points of coordinates $(x(t), y(t))$ is defined in a complex plane as

$$c(t) = x(t) + jy(t). \quad (16)$$

The elliptical Fourier representation of a contour is [40]

$$c(t) = \frac{a_{x0}}{2} + \sum_{k=1}^{m/2} (a_{xk} \cos(k\omega t) + b_{xk} \sin(k\omega t)) + j \left[\frac{a_{y0}}{2} + \sum_{k=1}^{m/2} (a_{yk} \cos(k\omega t) + b_{yk} \sin(k\omega t)) \right], \quad (17)$$

where

$$a_{xk} = \frac{2}{m} \sum_{i=1}^m x_i \cos(k\omega i\tau), \quad b_{xk} = \frac{2}{m} \sum_{i=1}^m x_i \sin(k\omega i\tau), \quad (18)$$

$$a_{yk} = \frac{2}{m} \sum_{i=1}^m y_i \cos(k\omega i\tau), \quad b_{yk} = \frac{2}{m} \sum_{i=1}^m y_i \sin(k\omega i\tau), \quad (19)$$

fundamental frequency $\omega = T/2\pi$ and sampled period $\tau = T/m$. The contour represented in matrix form is [40]

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}. \quad (20)$$

This matrix resembles that of an ellipse, with a_{xk} and b_{yk} representing its major and minor axes, respectively. The scale, rotation and translation invariant EFD is

$$\text{EFD} = \frac{\sqrt{a_{xk}^2 + a_{yk}^2}}{\sqrt{a_{x1}^2 + a_{y1}^2}} + \frac{\sqrt{b_{xk}^2 + b_{yk}^2}}{\sqrt{b_{x1}^2 + b_{y1}^2}}. \quad (21)$$

We represent EFDs in the form of a matrix of dimension $a_r \times b_c$ (with a_r representing the 10 specific phases of the gait period and b_c representing the number of elliptic Fourier coefficients considered). Since the number of elliptic Fourier coefficients equaling one half of the total number of contour points is capable of

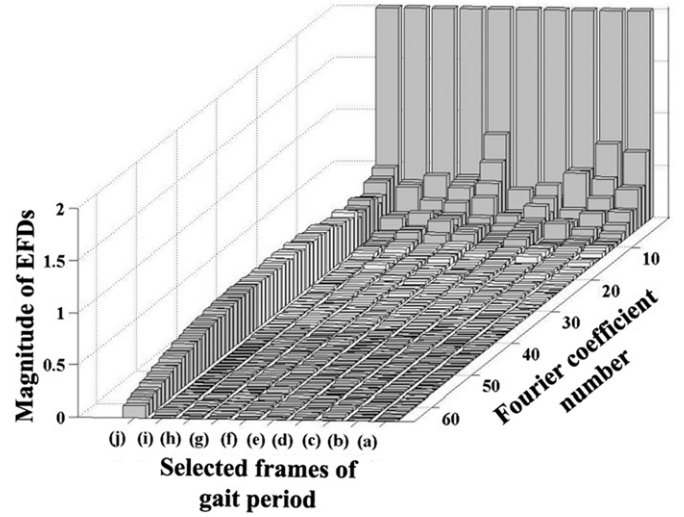


Fig. 5. The 3D bar graph with bars representing the magnitude of EFDs corresponding to 64 Fourier coefficients grouped together for each of 10 specific phases of gait period.

reconstructing a very good approximation of the original contour [40], we use $b_c = m/2$, i.e., 64, for STM-SPP. Each EFD is represented by a bar in the 3D bar graph shown in Fig. 5, where elements of the rows of the matrix are grouped together. Let P and Q be two such matrices for a gallery and a probe gait sequence, respectively. The dissimilarity score between them is

$$D_{\text{EFD}} = \frac{\sum_{i=1}^{a_r} \sum_{j=1}^{b_c} (\mathbf{P}_{ij} - \mathbf{Q}_{ij})^2}{\sum_{j=1}^{b_c} (\mathbf{P}_j - \text{mean}(\mathbf{P}_j))^2}, \quad (22)$$

where \mathbf{P}_j is j th column of \mathbf{P}_{ij} and $\text{mean}(\cdot)$ computes average. The range of D_{EFD} is $[0,1]$, the larger the value the more dissimilar are the two shapes. We obtain dissimilarity score by comparing the EFDs of a probe subject with EFDs of each of the gallery subjects for a gait period. In a similar manner to phase 1, this dissimilarity score is used to classify the subject.

The above shape analysis gives excellent classification results in the absence of shape variations between the gallery and probe sequences of the same subject. The classification performance decreases significantly if the shape of the gallery and probe subjects differ due to different activities (e.g., slow or fast walk vs walking with ball) undertaken by the same subject for the CMU MoBo data set, and carrying conditions (briefcase vs no briefcase) and shadows under feet for the USF HumanID data set. To reduce the effects of shape variations on the classification rate, a part-based shape analysis using EFDs is invoked when carrying condition is detected in a sequence.

The shape of an upright silhouette above the wrist is not affected by shape variations when the subject's hand carries a briefcase or a small bag. According to anthropometry, the position of the wrist as a fraction of body height is estimated to be $0.485H$ [41] measured from the bottom of the bounding rectangle. Thus, an analysis of the part of silhouettes enclosed in the region $(1 - 0.485H)$, i.e., $0.515H$ of the bounding rectangle measured from the top using EFDs remove the shape variations due to carrying conditions (briefcase vs no briefcase) among the gallery and probe sequences. The leg region of a silhouette enclosed between aH and bH (where $a = 0.750$ and $b = 0.9375$ [28]) removes the effect of shape distortion due to the presence of briefcase and shadows under feet. Thus, STM-SPP detects the presence of briefcase or a small bag automatically by examining the difference in the number of contour points enclosed in the region between $0.515H$ and $0.750H$. A substantial increase in the number

of contour points (e.g., for USF data set an increase of at least twenty contour points without applying polygon approximation and for the same phase of gait period between gallery and probe sequences) confirms the presence of briefcase for most of the cases. STM–SPP analyses the EFDs of the subjects carrying briefcase with shadows under feet in two parts, so as to avoid the variations in shape among gallery and probe sequences. We verified the effectiveness of this part-based shape analysis on all subjects carrying a briefcase for the USF data set.

If a subject carries a small item, e.g., ball, package, tiffin box, etc., with folded arms, it is unlikely that the shape of silhouettes above the shoulder and below the position of wrist, i.e., $0.515H$, from the top of the bounding rectangle will be affected. Experimental analysis for all the subjects holding a ball in CMU MoBo data set verifies the appropriateness of the assumption. We found that the segment of the silhouette enclosed between $0.225H$ from the top of the bounding rectangle, and the lower segment enclosed between $0.500H$ and the bottom of the bounding rectangle exclude the ball for all the twenty four subjects walking with ball in the CMU MoBo data set (Fig. 6(a)–(f)). Experimental analysis reveals that an increase in the number of contour points enclosed in the region between the anatomical position of wrist and the top of the bounding rectangle by at least 50 confirms the presence of a ball for all the subjects in the CMU MoBo data set without applying any morphological operation and polygon approximation technique. Thus, the effect of carrying small items on silhouette shape can be removed by segmenting it in two parts: (1) the upper segment spanning from the top of the bounding rectangle up to the shoulder and (2) the lower segment spanning from the anatomical position of wrist to the bottom of the bounding rectangle. For part-based EFD analysis, we obtain a dissimilarity score corresponding to each part and compute the average dissimilarity score.

3.4. Module 3: combining classifications

We use rank-summation based classifier combination rule [46] to combine the outputs of the two classifiers for improved reliability in human identification. We choose this combination rule due to its appropriateness in STM–SPP, as it enables to effectively combine the results of a small number of classifiers with a relatively large number of classes. It is also easier to implement than the score-based fusion strategy [47], as the latter requires transformation of the scores to a same scale in order to be comparable before being combined.

Let $R^{(j)}(\theta)$ be the rank assigned to class θ by classifier $j \in J$, where J represents the set of classifiers consisting of two elements for STM–SPP. The sum of all ranks assigned to each class by all classifiers is

$$S(\theta) = \sum_j R^{(j)}(\theta). \quad (23)$$

The class with the lowest sum rank is chosen as the correct class for the probe sequence. Noting that Hu moments are linear combinations of normalised central moments that are invariant to changes in rotation, reflection and scale, to resolve cases where



Fig. 6. Illustration of part-based shape analysis: (a–c) upper segments and (d–f) lower segments of three subjects from CMU MoBo data set to exclude the ball.

two classes have the same sum rank, STM–SPP performs hierarchical contour matching [37] based on the Hu moments [23] between each image of the probe gait sequence and the corresponding images of the gallery sequences.

The translational invariant 2D $(p+q)$ th order central moments $(\mu_{p,q})$ of a contour $I(x,y)$ is

$$\mu_{p,q} = \sum_{i=0}^N I(x,y)(x-x_{avg})^p(y-y_{avg})^q, \quad (24)$$

where $x_{avg} = m_{10}/m_{00}$, $y_{avg} = m_{01}/m_{00}$ and N is the total number of pixels in the contour. To ensure objects of the same shape but dissimilar sizes give similar values, we use the normalised central moment

$$\eta_{p,q} = \frac{\mu_{p,q}}{m_{00}^{(p+q)/2+1}}. \quad (25)$$

The hierarchical contour matching technique involves the formation of contour trees [48] prior to the contour comparisons based on Hu moments. Since the resultant contour trees are susceptible to minor variations in the contours, all contours are approximated by a polygon using Douglas–Peucker approximation algorithm [49,37] having fewer vertices (as shown in Fig. 1(g)), for better comparison.

4. Experiments

We use the silhouettes from CMU MoBo data set [44] and USF HumanID gait challenge data set [13] to evaluate the performance of STM–SPP with variations in terms of walking speeds, carrying conditions, clothing, footwear and viewpoints. Figs. 7 and 8, respectively, show sample silhouettes of CMU MoBo data set and USF HumanID gait challenge data set. Experimental analyses on these two data sets enable us to make uniform comparisons with several related methods.

4.1. Experiments on CMU MoBo data set

CMU MoBo data set comprises sequences of 25 subjects performing four types of walk: slow walk; fast walk; walk holding a ball; and walk on an inclined plane of a treadmill. Each sequence is of approximately 11 s duration and is recorded at 30 frames/s

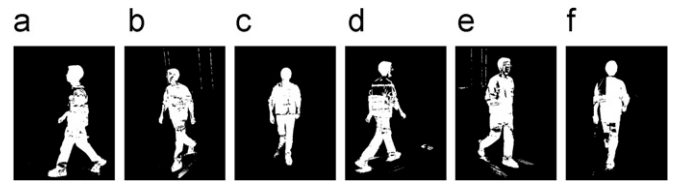


Fig. 7. Sample silhouettes of the same subject from CMU MoBo data set for six views.

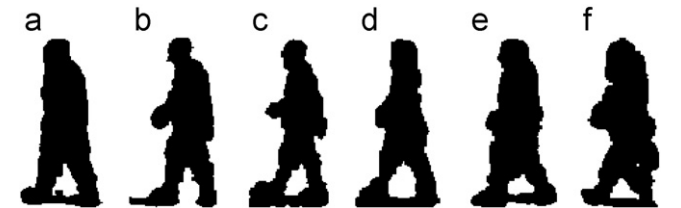


Fig. 8. Sample silhouettes of different subjects from USF Gait Data set: (a and b) walk on grass with different shoe types and viewpoints; (c and d) walk on concrete with different shoe types and viewpoints; and (e and f) walk on grass carrying a briefcase.

from six different views. We evaluate the performance of STM-SPP on the profile view of the silhouettes as illustrated in Fig. 7(a), using different gallery and probe sequences for slow walk, fast walk and walk with ball. We compare the performance of STM-SPP with the following silhouette based approaches: SSP [24], CMU1 [19] and VCR-C [20].

In order to compare STM-SPP with SSP, we also use holdout cross validation technique, in which gallery and probe sets correspond to different combination of walking speeds for each of twenty five subjects. Since STM-SPP is defined only on lateral view of the silhouettes, we consider lateral view of two sequences per subject (total 50 sequences) walking at slow pace (2.06 miles/h) and fast pace (2.82 miles/h), for performance comparison with SSP. Table 1 shows that the performance of SSP for profile view (obtained from [24]) degrades

Table 1

Top-rank identification rates of STM-SPP (in %) on MoBo data set for lateral view with rates of SSP from [24] enclosed in parentheses.

Probe	Gallery		
Walk type	Slow walk	Fast walk	Walk with ball
Slow walk	100(100)	94(54)	93
Fast walk	91(32)	100(100)	84
Walk with ball	82	82	100

Table 2

Top-rank identification rates (in %) of CMU1 (from [19]) and STM-SPP on CMU MoBo data set for profile views only.

Gallery	Probe	CMU1	STM-SPP
Profile, slow	Profile, fast	76	96
Profile, slow	Profile, ball	92	93

Table 3

Top-rank identification rates (in %) of STM-SPP on MoBo data set for across-activities with rates of VCR-C from [20] enclosed in parentheses.

Activity	Slow walk	Fast walk	Walk with ball
Slow walk	100(100)	95(80)	93(48)
Fast walk	96(84)	100(100)	84(48)
Walk with ball	82(68)	82(48)	100(92)

Table 4

Top-rank identification rates (in %) for different methods on full version of USF HumanID gait challenge data set using the gallery set (G, A, R, NB, M/N) of 122 subjects. The rates for GEI from [1] are enclosed in parentheses. Keys for covariates: 1, view; 2, shoe; 3, surface; 4, briefcase; 5, time; 6, clothes.

Exp.	Probe [no. of subjects]	Covariates	Identification rate (P_i)%				Verification rate (P_V)%	
			Baseline (GEI)		STM-SPP		Baseline	STM-SPP
			Rank 1	Rank 5	Rank 1	Rank 5	$P_F = 1(10)\%$	$P_F = 1(10)\%$
A	G, A, L, NB, M/N [122]	1	73(90)	88(94)	92	96	82(94)	88(100)
B	G, B, R, NB, M/N [54]	2	78(91)	93(94)	95	98	87(94)	94(100)
C	G, B, L, NB, M/N [54]	2, 1	48(81)	78(93)	84	95	65(94)	86(98)
D	C, A, R, NB, M/N [121]	3	32(56)	66(78)	72	80	44(80)	80(94)
E	C, B, R, NB, M/N [60]	3, 2	22(64)	55(81)	68	84	35(76)	74(84)
F	C, A, L, NB, M/N [121]	3, 1	17(25)	42(56)	29	59	20(60)	50(82)
G	C, B, L, NB, M/N [60]	3, 2, 1	17(36)	38(53)	40	61	28(55)	52(76)
H	G, A, R, BF, M/N [120]	4	61(64)	85(90)	69	92	72(91)	83(95)
I	G, B, R, BF, M/N [60]	4, 2	57(60)	78(83)	60	84	67(85)	76(93)
J	G, A, L, BF, M/N [120]	4, 1	36(60)	62(82)	64	85	48(76)	65(92)
K	G, A/B, R, NB, N [33]	5, 2, 6	3(6)	12(27)	20	30	6(24)	21(58)
L	C, A/B, R, NB, N [33]	3, 5, 2, 6	3(15)	15(21)	18	27	6(24)	19(52)

significantly when the probe and gallery sequences differ in walking speed, whereas STM-SPP which analyses shape at 10 specific phases of gait period and uses DTW for gait period estimation overcomes the effects of variations in walking speed, and outperforms SSP for the profile view.

Table 2 shows that STM-SPP outperforms CMU1 for profile view. A comparison with VCR-C is presented in Table 3 for subjects walking parallel to the image plane. The shape based approach of VCR-C which uses stance correlation, shows encouraging results for subjects performing same activities with varying speed. Although speed variations are also accounted for in STM-SPP, like VCR-C the performance of STM-SPP degrades when shape of the subject's silhouettes change due to different activities (e.g., slow walk vs walk with ball). The better performance of STM-SPP over VCR-C is attributed to the validation based on physical parameters (i.e., gait period and step length) which are independent of body shape, and part-based shape analysis using EFDs.

4.2. Experiments on USF HumanID gait challenge data set

We evaluate STM-SPP on both the earlier, smaller version (452 sequences from 74 subjects, data acquired in May only) and the full version (1870 sequences from 122 subjects, data acquired in May and November) of USF HumanID gait challenge data set available at <http://www.GaitChallenge.org>. This data set comprises sequences of subjects walking on elliptical paths and provides up to thirty two possible 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 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)) [13]. The 33 subjects that are common in the May and November data sets account for time covariate. The number of subjects in the probe set is enclosed in square bracket in both Tables 4 and 5. There are no common sequences between the gallery set and any of the probe sets, and not all subjects participated in all experiments [1,13].

Table 4 shows the results on the full version of the data set in terms of identification rate (P_i) at ranks 1 and 5, and verification rate (P_V) at the false alarm rates of 1% and 10%, to enable a comparison with that of the baseline method [13] and GEI [1]. We report the verification rates for baseline method obtained using z-normalised similarity scores. The identification rates achieved by GEI for the twelve challenge experiments after combining the real and synthetic gait features are enclosed in parentheses.

Table 5

Top-rank identification rates (in %) for silhouette-based algorithms on earlier, smaller USF HumanID gait challenge data set (data acquired in May only) using the gallery set (G, A, R) of 71 subjects. The rates for baseline, CASIA, CMU2, CMU1 and GEI are from [13,14,25,25,1], respectively. Keys for covariates: 1, view; 2, shoe; 3, surface.

Exp.	Probe [no. of subjects]	Covariate	Baseline	CASIA	CMU2	CMU1	GEI	STM-SPP
A	G, A, L [71]	1	87	70.42	85	87	100	100
B	G, B, R [41]	2	81	58.54	81	81	90	94
C	G, B, L [41]	2, 1	54	51.22	60	66	85	89
D	C, A, R [70]	3	39	34.33	23	21	47	73
E	C, B, R [44]	3, 2	33	21.43	17	19	57	69
F	C, A, L [70]	3, 1	29	27.27	25	27	32	40
G	C, B, L [44]	3, 2, 1	26	14.29	21	23	31	36

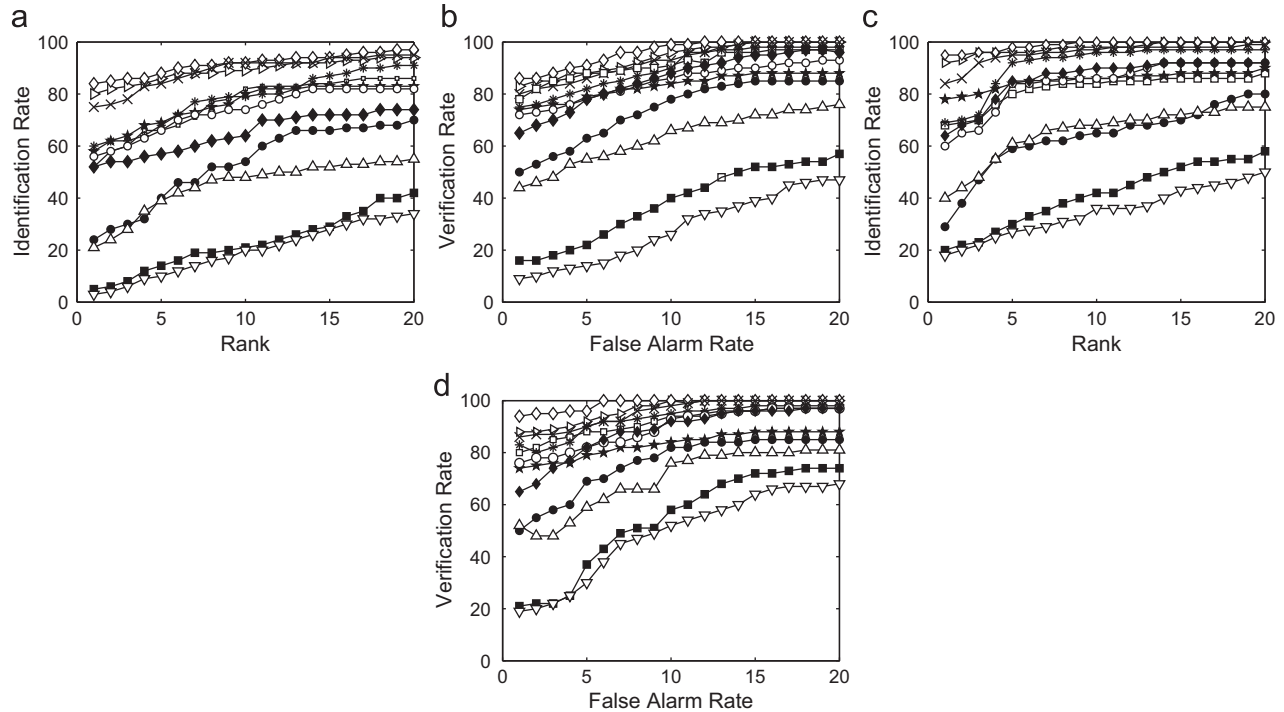


Fig. 9. Performance on 12 challenge experiments of USF data set. Identification mode (CMC): (a) PSA and (c) combined. Verification mode (ROC): (b) PSA and (d) combined. Keys: '▷', Exp. A (probe: G, A, L, NB, M/N); '◊', Exp. B (probe: G, B, R, NB, M/N); '×', Exp. C (probe: G, B, L, NB, M/N); '□', Exp. D (probe: C, A, R, NB, M/N); '★', Exp. E (probe: C, B, R, NB, M/N); '•', Exp. F (probe: C, A, L, NB, M/N); '△', Exp. G (probe: C, B, L, NB, M/N); '✱', Exp. H (probe: G, A, R, BF, M/N); '○', Exp. I (probe: G, B, R, BF, M/N); '◆', Exp. J (probe: G, A, L, BF, M/N); '■', Exp. K (probe: G, A/B, R, NB, N); '▽', Exp. L (probe: C, A/B, R, NB, N).

All methods perform satisfactorily for experiments A–J but poorly for experiments K and L, with STM-SPP achieving the best performances in all experiments, followed by GEI. The better performance of STM-SPP in experiment G than in experiment F in spite of an additional covariate, namely shoe type, is attributed to the fewer subjects participating in experiment G and thus the smaller likelihood of including subjects from the class which is difficult to identify across all the experiments for STM-SPP [13]. Also, the validation based on physical parameters contributes to making STM-SPP robust against across-day gait variations, e.g., the same subject wearing different shoes.

Table 5 shows the results on the smaller, earlier version of the data set (No-Briefcase data) to enable a comparison with Baseline, CASIA [14], CMU2 [25], CMU1 and GEI. Note that we present the identification rates at rank 1 of CMU2 obtained by weighted correlation similarity measure, as its performance is better than the identification rates at rank 1 obtained by median weighted distance similarity measure presented in [25]. We compare the performance of STM-SPP with the identification rates of GEI obtained by fusing real and synthetic gait templates, as it shows higher performance than the identification rates obtained

separately by using real and synthetic gait templates. Table 5 shows that STM-SPP outperforms the other methods for all experiments.

The performance of STM-SPP for the 12 challenge experiments of the data set is measured by two different modes of experimental analysis, namely identification mode and verification mode, using Cumulative Match Characteristic (CMC) and Receiver Operating Characteristic (ROC) curves, respectively, following [50].

Identification refers to an attempt to determine the identity of an unknown subject by comparing a subject's probe sequence to all the gallery sequences in the database. In contrast to open-set identification, a closed-set identification always guarantees the existence of the subject in the database. In this paper, we analyse the closed set identification performance of STM-SPP on the profile view by taking out one subject as the probe sequence and train it on all the subjects of the data set including the probe sequence. The identification result is represented by a CMC curve, i.e., the probability of correct matches vs ranks. According to this curve, the probability of correct identification at rank r implies that the probability of correct match is among the top r similarity scores, and the performance at rank 1 represents the correct

classification rate (CCR), i.e., the identification rate. Fig. 9(a) and (c) shows that the identification rates at rank 1 range from 3% to 84% for PSA, which are increased to a range of 18–92% using the classifier combination rule.

Verification refers to an attempt to confirm a subject's claimed identity by a one-to-one comparison of the probe sequence to one or more gallery sequences corresponding to the subject of the claimed identity in terms of false alarm rate (the probability that the method incorrectly matches the probe sequence to a non-matching gallery sequence) and verification rate (the probability that the method succeeds to correctly detect a match between the probe and gallery sequences). An ROC curve is a graphical representation of the relationship between false alarm rate and verification rate of the classifier as its discrimination threshold is varied. Fig. 9(b) and (d) shows that the verification rates using PSA range from 9% to 86% at a false alarm rate of 1%, and increase to 19–94% at a false alarm rate of 1% by combining the classifiers, for all the twelve challenge experiments. It is evident from Table 4 and Fig. 9, that STM-SPP is least affected by variation in shoe types, followed by about 30 degrees change in viewpoint. However, time (i.e., when the data set was generated) has the most impact on the performance of STM-SPP, as it implicitly means the same subjects wearing different clothes and shoes.

4.3. Computational complexity analysis

The time for recognising a subject depends on the size of the data sets as well as on their characteristics, i.e., the number of cases in which similar dissimilarity scores are obtained by PSA and the number of subjects carrying items in different sequences. The processing time (measured using the computer system clock) for comparing all the 10 Rf-ROI's one at a time with the Rt-ROIs obtained from the silhouette images of a subject's gait period based on Hu moments and determining the minimum S_{score} in each case for extracting the 10 specific phases of the gait period is 5 s using OpenCV 2.1 in Microsoft Visual Studio 2008 Express Edition environment on an Intel (R) Core (TM) i7 processor working at 2.93 GHz with 4 GB RAM and 500 GB HDD running Windows 7 operating system. The combined processing time to obtain the dissimilarity scores between a gallery subject and a probe subject using both PSA and EFDs along with estimation of the subject's physical parameters is 45 s/gait period.

The baseline method has a very high computational complexity as it performs repeated intersequence spatio-temporal correlation between gallery and probe silhouette sequences to obtain the similarity measure [4,14]. Instead of processing an entire gait sequence, the real-time method in [3] identifies a previously known subject by analysing its silhouettes over a gait period spanning up to 25 frames to reduce computational complexity. STM-SPP further reduces it by analysing the shape of contours instead of silhouettes at the double support phase and 10 specific phases of the gait period by using PSA and EFDs, respectively. Furthermore, it converts the contour at the double support phase into 1D shape signal based on 28 landmarks before applying PSA. Since the configuration matrices formed by the shape signals have one column only, the space and time complexity of PSA is linear, i.e., $O(k)$, where $k=28$ for STM-SPP. Despite the effectiveness of DTW, it has a quadratic time and space complexity ($O(MN)$) (where M and N denote the lengths of the two time-varying sequences being compared) which limits its usefulness to small sequences. However, the number of frames of a gait period does not usually exceed 35 and this ensures the suitability of using DTW in STM-SPP. The 10 specific phases of a gait period are obtained by Region-Of-Interest (ROI) of contour matching based on Hu moments. The use of ROI helps to speed up execution time, as it enables processing of subregion of an image.

The NNC used for validating similar dissimilarity scores requires storage of entire gallery database, thus requiring much memory space and increased execution time. However, its use is limited only to resolving similar dissimilarity scores obtained by PSA, and therefore it does not increase the overall computational complexity of STM-SPP significantly. Since the computational complexity of EFDs is quadratic, i.e., $O(b_c m)$ (where b_c denotes the number of elliptic Fourier coefficients considered and m denotes the number of points of the polygon-approximated contour), it increases the processing time of STM-SPP. To address this, instead of analysing all frames of a gait period, we analyse the contours by EFDs only at its 10 specific phases (i.e., keeping $a_r=10$) while capturing most of the significant gait characteristics. The contours are approximated with reduced number of points, i.e., $m=128$, which reduces processing time.

5. Conclusion

The proposed two-phase gait recognition method, STM-SPP, analyses the shape of silhouette contours of a human subject in a video sequence. In the first phase, STM-SPP performs subject classification based on a dissimilarity score by comparing distances of landmarks (anatomical, mathematical and pseudo) from centre-of-mass of the contour in the double support phase of a gait period using PSA. This classification performance is substantially enhanced by validating similar dissimilarity scores based on spatio-temporal gait characteristics and physical parameters of human body in the presence of limited variations in view, clothing and footwear. In the second phase, STM-SPP characterises the silhouette contours by EFDs at the 10 specific phases of a gait period to obtain gait signatures and uses a dissimilarity score to classify the subjects. A part-based shape analysis using EFDs is applied to reduce the impact of shape variation between gallery and probe silhouette contours on the classification rate when across-day variations due to carrying conditions are detected. The outputs of the two classifiers are combined effectively by rank-summation based classifier combination rule, where a tie in ranking is resolved by contour matching based on Hu moments.

STM-SPP has several desirable advantages, which make it suitable for real-world applications. The shape analysis at the 10 specific phases of a gait period and gait period detection by the application of DTW aids STM-SPP to deal with varying walking speeds of the same subject under different circumstances. STM-SPP is also robust to subjects carrying small items and limited across-day gait variations, but not significant change of styles, e.g., pants vs skirts or long coats, massive leg injury, variations of camera viewpoints, etc. It is also robust against missing or distorted frames to some extent mainly due to partial occlusions and segmentation imperfections. It is insensitive to colour and texture of the subject's clothing, as it analyses the shape of the contours. Since its feature space is simplified, it does not require any dimensionality reduction technique like PCA and multiple discriminant analysis as in [1,14]. The attractiveness of the STM-SPP is the ease of implementation with low computational complexity. Experimental analyses on two publicly available data sets show that STM-SPP significantly outperforms several related silhouette-based gait recognition methods. However, it suffers from the following limitations which require further attention for its advancement:

- Walking direction: STM-SPP is designed to identify human subjects only for lateral views of the gallery and probe sequences. Although, lateral views of the walking subjects capture most of the significant gait characteristics, it is not

always possible to capture image frames from the side of a subject, especially in hallways [3]. Hence, future developments are required to enable STM–SPP to address unconstrained human movements especially in cluttered scenes.

- **Clothing invariance:** STM–SPP is not robust against significant clothing variations between gallery and probe sequences, such as pants vs long skirts, shorts vs down jackets, trendy coats, etc. Also, the extraction of landmarks in the limb region is impossible in the case of a subject wearing long skirt or long coat as the clothing keeps the subject's limbs covered. Therefore, future work will involve improvements of STM–SPP using part-based clothing categorisation to achieve substantial clothing invariance.
- **Dynamic gait characteristics:** STM–SPP analyses sequences of deforming shape of contours over a gait period, but does not incorporate the dynamic motion characteristics of gait which also play an important role in human identification. Since the more appropriate gait signatures are utilised the better is the performance of any gait recognition algorithm, we will consider oscillatory trajectories of joints in future work, giving arm-swing a consideration.

Acknowledgments

The authors would like to thank Warwick Postgraduate Research Scholarship for providing the funds for this research.

References

- [1] J. Han, B. Bhanu, Individual recognition using gait energy image, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2) (2006) 316–322.
- [2] L. Wang, H. Ning, T. Tan, Fusion of static and dynamic body biometrics for gait recognition, *IEEE Transactions on Circuits and Systems for Video Technology* 14 (2) (2004) 149–158.
- [3] O. Barnich, M.V. Droogbroeck, Frontal-view gait recognition by intra- and inter-frame rectangle size distribution, *Pattern Recognition Letters* 30 (9) (2009) 893–901.
- [4] M.S. Nixon, T.N. Tan, R. Chellappa, *Human Identification Based on Gait*, Springer Science+Business Media, New York, USA, 2006.
- [5] F. Tafazzoli, R. Safabakhsh, Model-based human gait recognition using leg and arm movements, *Engineering Applications of Artificial Intelligence* 23 (2010) 1237–1246.
- [6] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviours, *IEEE Transactions on Systems, Man, and Cybernetics, Part C* 34 (3) (2004) 334–352.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] L. Havasi, Z. Szlavik, T. Sziranyi, Higher order symmetry for non-linear classification of human walk detection, *Pattern Recognition Letters* 27 (2006) 822–829.
- [11] Y. Ran, Q. Zheng, R. Chellappa, T.M. Stat, Applications of a simple characterization of human gait in surveillance, *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 40 (4) (2010) 1009–1019.
- [12] N.V. Boulgouris, K.N. Plataniotis, D. Hatzinakos, Gait recognition using linear time normalization, *Pattern Recognition* 39 (2006) 969–979.
- [13] 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.
- [14] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (12) (2003) 1505–1518.
- [15] J.D. Shutter, M.S. Nixon, C.J. Harris, Statistical gait description via temporal moments, in: *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, 2000, pp. 291–295.
- [16] J.D. Shutter, M.S. Nixon, Zernike velocity moments for sequence-based description of moving features, *Image and Vision Computing* 24 (2006) 343–356.
- [17] L. Wang, T. Tan, W. Hu, H. Ning, Automatic gait recognition based on statistical shape analysis, *IEEE Transactions on Image Processing* 12 (9) (2003) 1–13.
- [18] C. BenAbdelkader, R. Culter, L. Davis, Stride and cadence as a biometric in automatic person identification and verification, in: *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 2002.
- [19] R. Collins, R. Gross, J. Shi, Silhouette-based human identification from body shape and gait, in: *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 2002.
- [20] A. Veraraghavan, A.K. Roy-Chowdhury, R. Chellappa, Matching shape sequences in video with applications in human movement analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (12) (2005) 1896–1909.
- [21] F.P. Kuhl, C.R. Giardina, Elliptic Fourier features of a closed contour, *Computer Graphics and Image Processing* 18 (1982) 236–258.
- [22] I.L. Dryden, K.V. Mardia, *Statistical Shape Analysis*, second ed., John Wiley & Sons Ltd, England, 1998.
- [23] M.K. Hu, Pattern recognition by moment invariants, *Proceedings of the Institute of Radio Engineers* 49 (9) (1961) 1428.
- [24] C. BenAbdelkader, R.G. Cutler, L.S. Davis, Gait recognition using image self-similarity, *EURASIP Journal on Advances in Signal Processing* (4) (2004) 572–585.
- [25] D. Tolliver, R. Collins, Gait shape estimation for identification, in: *Proceedings of the International Conference on Audio- and Video-based Biometric Person Authentication*, 2003.
- [26] 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.
- [27] K. Bashir, T. Xiang, S. Gong, Gait recognition without subject cooperation, *Pattern Recognition Letters* 31 (2010) 2052–2060.
- [28] C. Wang, J. Zhang, J. Pu, X. Yuan, L. Wang, Chrono-Gait Image: a novel temporal template for gait recognition, in: *European Conference on Computer Vision, Part 1, LNCS 6311*, 2010, pp. 257–270.
- [29] 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.
- [30] 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.
- [31] D. Tau, 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.
- [32] Z. Liu, S. Sarkar, Improved gait recognition by gait dynamics normalization, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (6) (2006) 863–876.
- [33] H. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, MPCA: multilinear principal component analysis of tensor objects, *IEEE Transactions on Neural Networks* 19 (1) (2008) 18–39.
- [34] Q. Meng, B. Li, H. Holstein, Recognition of human periodic movements from unstructured information using a motion-based frequency domain approach, *Image and Vision Computing* 24 (2006) 795–809.
- [35] S.D. Mowbray, M.S. Nixon, Automatic gait recognition via Fourier descriptors of deformable objects, in: *Proceedings of the Audio Visual Biometric Person Authentication*, Guildford (UK), 2003, pp. 556–573.
- [36] K. Kim, T.H. Chalidabhongse, D. Harwood, L. Davis, Real-time foreground-background segmentation using codebook model, *Real-Time Imaging* 11 (2005) 172–185.
- [37] G. Bradski, A. Kaehler, *Learning OpenCV Computer Vision with the OpenCV Library*, O'Reilly Media, Sebastopol, 2008.
- [38] L. Jianzhuang, L. Wenqing, T. Yupeng, Automatic thresholding of gray-level pictures using two-dimension Otsu method, in: *International Conference on Circuits and Systems, China*, 1991.
- [39] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1) (1979) 62–66.
- [40] M. Nixon, A.S. Aguado, *Feature Extraction and Image Processing*, second ed., Elsevier, London, 2006.
- [41] D.A. Winter, *Biomechanics and Motor Control of Human Movement*, third ed., John Wiley & Sons, New Jersey, 2004.
- [42] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, second ed., Addison-Wesley, USA, 1992.
- [43] M. Müller, *Information Retrieval for Music and Motion*, Springer, New York, 2007.
- [44] R. Gross, J. Shi, The CMU Motion of Body (MoBo) Database, Technical report CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, 2001.
- [45] J.M. Le Minor, M. Schmittbuhl, Importance of elliptic Fourier methods for morphology of complex outlines: application to the distal human femur, *Technical report, Surgical and Radiologic Anatomy* 21 (6) (1999) 387–391.
- [46] O. Melnik, Y. Vardi, C. Zhang, Mixed group ranks: preference and confidence in classifier combination, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26 (8) (2004) 973–981.
- [47] B. Achermann, H. Bunke, Combination of Classifiers on the Decision Level for Face Recognition, University of Bern, IAM-96-002, 1996.
- [48] H. Carr, J. Snoeyink, M. van de Panne, Progressive topological simplification using contour trees and local spatial measures, in: *Proceedings of the 15th Western Computer Graphics Symposium*, Big White, British Columbia, 2004.

- [49] D. Douglas, T. Peucker, Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *Cartographica: The International Journal for Geographic Information and Geovisualization* 10 (2) (1973) 112–122.
- [50] J. Phillips, H. Moon, S. Rizvi, P. Rause, The FERET evaluation methodology for face recognition algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (10) (2000) 1090–1104.

Sruti Das Choudhury received B.Tech. in information technology from West Bengal University of Technology, India and M.Tech. in computer science and application from Calcutta University, India, in 2005 and 2009, respectively. She is currently pursuing Ph.D. in engineering at the University of Warwick, U.K. Her research interest is gait recognition.

Tardi Tjahjadi received B.Sc. in mechanical engineering from University College London in 1980, and M.Sc. in management sciences in 1981 and Ph.D. in total technology in 1984 from UMIST, U.K. He has been an associate professor at Warwick University since 2000. His research interests include image processing and computer vision.