

# Spatiotemporal Features of Human Motion for Gait Recognition

Muhammad Hassan Khan · Muhammad Shahid Farid · Marcin Grzegorzek

Received: date / Accepted: date

**Abstract** Gait is a novel biometric feature that offers human identification at a distance and without physical interaction with the imaging device. Moreover, it performs well even in low-resolution which makes it ideal for use in numerous human identification applications, *e.g.*, visual surveillance, monitoring and access control systems. Most existing gait based human identification solutions extract human body silhouettes, contours, or shapes from the images and construct gait features. Therefore, the performance of such algorithms highly depends upon the accuracy of human body segmentation, which is still a challenging problem in literature. In this paper, we propose a new gait recognition algorithm which uses the spatial and temporal motion characteristics of human gait for individual identification without needing the silhouette extraction. The proposed algorithm extracts a set of spatiotemporal descriptors from gait videos. The extracted descriptors are encoded using the Fisher vector encoding and Gaussian mixture model based codebook. The encoded features are classified using a simple linear support vector machine to recognize the individuals. The proposed gait recognition method is evaluated on five widely used gait databases, including indoor (CMU-MoBo, CASIA-B) and outdoor (NLPR, CASIA-C, TUM GAID) gait databases. The results reveal that our method showed excellent performance on all five databases and outperformed the state-of-the-art gait recognition approaches.

**Keywords** Gait recognition, Spatiotemporal features, Fisher vector encoding, Feature evaluation

---

M.H. Khan (✉) · M. Grzegorzek  
Research Group for Pattern Recognition, University of Siegen, Germany.  
E-mail: hassan.khan@uni-siegen.de

M.S. Farid  
College of Information Technology, University of the Punjab, Pakistan.

## 1 Introduction

Biometrics has emerged as an important research area in the recent years due to its vast applications in various fields such as surveillance, authentication and etc. A number of human physiological characteristics *e.g.*, face, iris, fingerprints, deoxyribonucleic acid (DNA) have been shown to be unique and therefore they can be used for individual identification [28, 42]. Lately, the gait biometric has received attention of research community for human identification. It refers to the way a person walks. Gait is different from other biometric modalities in which the person interaction is needed with the system for identification such as fingers must be placed on the scanners to get their print, face must be in a specific position in front of a camera to be captured, DNA requires sample of blood or skin to be used for analysis; the gait on the other hand does not require any interaction with the imaging device. It works in a hidden manner and the individuals are generally unaware that they are being captured by a camera installed at some secret place. Moreover, the performance of gait based human identifiers is less susceptible to noise as compared to other biometrics and that is why they perform pretty well even in low quality and low resolution videos. Gait may not be as precise and powerful biometric as some other modalities are, but the aforementioned characteristics make the gait an ideal choice for visual surveillance applications.

The gait recognition algorithms are generally categorized into two types: model-based algorithms and model-free approaches. In the former category, the gait recognition algorithms use the human motion, body structures, and joint positions to track and recognize the individuals [57]. The structural models are usually generated based on the prior knowledge of the shape of human body [7, 23]. The motion models [3, 10] exploit the motion parameters of human body-parts like joint angle trajectories, rotation patterns of hip and

thigh. The gait recognition algorithm proposed in [3] computes features from joints' locations using elliptic Fourier descriptors which are used to recognize the gait. The authors in [10] modeled a gait feature by exploiting the angular motion of the hip and thigh using the Fourier series. Wang et al. [53] modeled human body using fourteen rigid parts connected to each other at joint locations and used these locations to obtain joint angle trajectories which are used as gait descriptor. The model-based approaches are somewhat robust to occlusion but they suffer with number of limitations such as, the performance of these methods depends upon the torso localization and also requires high quality gait videos [3, 57]. Moreover, these methods are computationally expensive too.

The model-free gait recognition approaches usually extract human silhouettes from gait video and derive different information from these silhouettes images for gait identification. Numerous techniques construct a template image from the segmented human body silhouettes and use it for person identification *e.g.*, [1, 8, 17, 24, 30, 37, 41, 51]. The technique proposed in [30] computes silhouette images and averages the results over a gait cycle to estimate the person identity, known as gait energy image (GEI). Several improvements over GEI have also proposed *e.g.*, [1, 24, 41, 46, 51, 60]. The approaches [14, 36, 59] extract different gait related parameters from the silhouette images and use them for person identification. Goffredo *et al.* [14] computed the height and width features from the normalized and scaled silhouette region over time and used them as gait dynamics. In [59], radial basis function network and deterministic learning is used to classify the height and width ratio and contour centroid over time to approximate the individual's gait.

Shape analysis on the sequence of silhouettes and projection of silhouettes in different directions have also explored in gait identification [21, 44, 45, 49, 54]. The 2D silhouette are mapped into 1D normalized distance signal in the algorithm described in [55]. The changes in shape of 1D silhouette over time are used to approximate the gait patterns. The methods in [44, 45] used the normalized segmented silhouette region and projected them in different directions. They consider a vector of projective values in each projection as gait feature. Motion information has also been extensively used for gait recognition *e.g.*, [6, 9, 20]. Castro et al. [6] computed the spatiotemporal cuboids of optical flow from the video sequences and fed to convolutional neural network to obtain a high level gait representation. The authors in [9] exploit the spatiotemporal motion characteristics and statistical and physical parameters of silhouettes for recognition.

The model-free gait recognition techniques achieve better results compared to their counterpart model-based approaches, and they are computationally efficient too. However, the bottleneck in these techniques is the precise silhouette

segmentation, as a poor segmentation can adversely affect the recognition performance [59]. This paper presents a novel spatiotemporal features based gait representation using dense trajectories to model the human gait. The motivation in using the dense trajectories is because they contain the local motion patterns of walk and they can be easily computed from the gait video sequences using optical flow field. The extracted motion information is encoded using Fisher vector and Gaussian mixture model based codebook. These encoded features are classified using a simple linear support vector machine (SVM).

The proposed approach, unlike existing techniques, does not require gait cycle estimation or human body extraction and therefore it performs better in low-resolution videos and poor quality videos. The experimental evaluation performed on CASIA-C gait database (Sect. 3.5) which consists of low resolution video sequences, reveals the efficacy of the proposed method in unfavorable and challenging conditions, as the video sequences are recorded at night using a thermal imaging camera. The proposed gait features comprises the persons static appearance and the relative motion information which are capable to capture even a small variance in the gait. High recognition accuracy obtained with a simple linear SVM also confirm the discriminative nature of the proposed features for gait recognition. Preliminary results of this research were presented in [19]; however, in this paper a number of improvements are proposed which are summarized in the following.

- In this study, we investigated the influence of dimensionality reduction on feature encoding. High dimensional features inevitably increase the computational complexity and the storage requirements. It is inferred from a large experimentation that applying principal component analysis (PCA) after feature encoding not only improves the recognition rate, it also reduces the classification time and the storage requirements of the proposed algorithm.
- It is also concluded that dimensionality reduction prior to feature encoding results in loss of few important cues for gait recognition, which may lower the recognition accuracy. Indeed, this information is important which significantly improves the efficiency of our method.
- In this research, an extensive experimental evaluation is performed. The proposed technique is evaluated on five large benchmark gait databases and the recognition results are compared with state-of-the-art techniques.

## 2 Proposed Gait Recognition Method

The proposed gait recognition technique consists of four steps. First, the dense trajectories are extracted for a set of sample points in a video sequence using optical flow field, and for each tracked point the appearance and motion in-

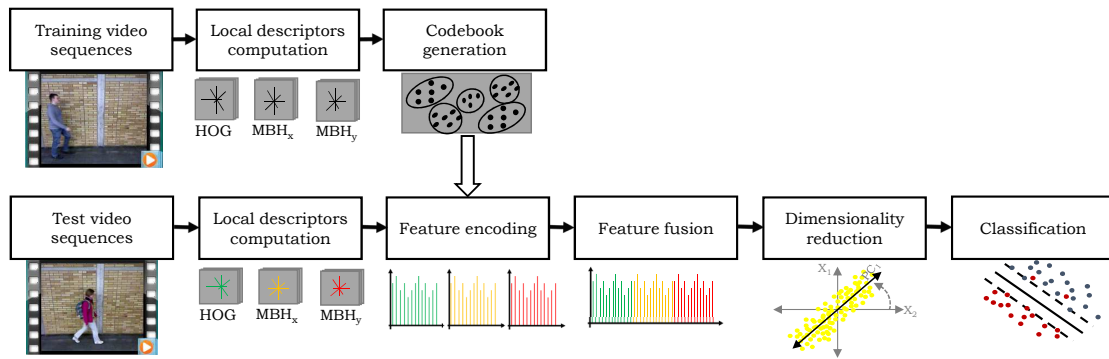


Fig. 1 Schematic diagram of the proposed gait-based human recognition algorithm.

formation of a walker is computed within a spatiotemporal patch and saved in local descriptors. Second, we randomly selected one million local descriptors to build a codebook. The codebook is constructed using Gaussian mixture models (GMM), and the local descriptors are encoded using this codebook and Fisher vector encoding scheme in the third step. Finally, we used a linear SVM to classify the encoded features. The performance of the proposed algorithm is evaluated on five widely used benchmark gait databases and the results are compared with the existing techniques. The recognition results confirm the superior performance of the proposed method. Fig. 1 shows the block diagram of the proposed method. Each step of the proposed method is presented in the following sections.

## 2.1 Motion Descriptor Computation

Recently, dense trajectories have proven to be effective in action recognition [33,52]. Since they encode the movement patterns of a person’s walk and can be easily computed directly from the gait video sequences, we used them to compute the gait features for person identification. In order to compute trajectories, we select a set of dense points from a frame and they are being tracked in the subsequent frames. The tracking is performed using displacement information in a dense optical flow field. In particular, each selected point  $P_t = (x_t, y_t)$  in frame  $t$  is tracked in the subsequent frame  $t + 1$  and so forth. The concatenation of these tracked points form a trajectory. Assuming that  $S$  represents a sequence of displacement vector and can be computed as,

$$S = (\Delta P_t, \Delta P_{t+1}, \dots, \Delta P_{t+L-1}), \quad (1)$$

where  $L$  is the trajectory length and  $\Delta P_t = (P_{t+1} - P_t)$ . The resultant vector  $S$  is normalized:

$$S' = \frac{(\Delta P_t, \dots, \Delta P_{t+L-1})}{\sum_{j=t}^{t+L-1} \|\Delta P_j\|} \quad (2)$$

The Eq. 2 describes the shape of the trajectory. The authors in [52] proposed the computation of local descriptors

around the tracked points to capture the appearance and motion information for action recognition. The local descriptors are computed within space-time volume and their information is saved in two histograms namely, histogram of oriented gradient (HOG) and histogram of optical flow (HOF). Furthermore, two additional descriptors are computed by taking the spatial derivatives along horizontal ( $x$ -axis) and vertical ( $y$ -axis) components of the optical flow and their information is saved in two histograms namely,  $MBH_x$  (*i.e.* Motion Boundary Histogram) and  $MBH_y$  respectively. They describe the relative motion between the pixels. We quantized the orientation information of each descriptor into 8-bin histogram. A ninth bin is used only for the HOF descriptor. The descriptors are normalized with  $L_2$ -norm. In order to assess the efficacy of these descriptors for gait recognition, we evaluated their several combinations in [19] and concluded that the combination of MBH and HOG outperforms the others feature combinations. The main reason for the superior performance of this combination is that HOG captures the person’s static appearance and MBH encodes the relative motion information between the pixels, therefore when used collectively they have a greater impact in recognizing the individuals using their appearance and motion information. Moreover, they are capable to capture small variances in the gait patterns.

## 2.2 Codebook Generation

Usually, a set of local descriptors are used to construct a fixed length vector to represent an image or video. We encoded HOG and MBH descriptors using Fisher vector (FV) encoding and Gaussian mixture model (GMM) based codebook. The FV is based on the principle of Fisher kernel [38] which incorporates the advantages of both discriminative and generic approaches using a kernel from generative model of the data. The GMM defines a collection of several Gaussian distributions over the feature space [33] and can be expressed as:

$$p(X | \theta) = \sum_{i=1}^K w_i \mathcal{N}(x | \mu_i, \Sigma_i) \quad (3)$$

where  $K$  represents the number of components of GMM,  $w_i$  is the weight of  $i$ th component with the constraint that  $\sum_{i=1}^K w_i = 1$ ,  $\mu_i$  and  $\Sigma_i$  are respectively representing the mean vector and covariance matrix of the  $i$ th component. Additionally,  $\theta = \{w_i, \mu_i, \Sigma_i, i = 1, 2, \dots, K\}$  is the list of parameters for all  $K$  components of GMM and  $\mathcal{N}(x | \mu_i, \Sigma_i)$  represents the Gaussian distribution with  $D$ -dimensions:

$$\mathcal{N}(x | \mu_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} e^{-\frac{1}{2}(x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i)} \quad (4)$$

For a given feature set  $X = \{x_1, \dots, x_T\}$ , we used maximum likelihood estimation algorithm [31] to estimate the optimal parameters of GMM. It performs soft assignment of a data point to each component. The assignment score which is also known as posterior probability demonstrates the association of data point to the respective component in GMM and can be defined as,

$$q_t(i) = \frac{w_i \mathcal{N}(x_t | \mu_i, \Sigma_i)}{\sum_{j=1}^K w_j \mathcal{N}(x_t | \mu_j, \Sigma_j)} \quad (5)$$

We consider that each component of GMM describes a particular appearance and motion pattern contributed by the local descriptors. Since GMM uses an iterative Expectation maximization (EM) algorithm [11, 25] which implements soft assignment of a descriptor to all components, the descriptor would be assigned to multiple components using the posterior estimates of the descriptor.

### 2.3 Feature Encoding

Feature encoding is the process of transforming the local descriptors of an image or video into a fixed length vector. Lately, FV encoding became quite popular due to its superior performance in many image processing applications [34, 38]. We chose FV to encode our local descriptors. The encoding of FV starts by learning a GMM (Sect. 2.2). Let  $X = \{x_t | t = 1, \dots, T\}$  be a set of local descriptors which is transformed into a vector using (3).

$$F_X = \frac{1}{T} \nabla_{\theta} \log p(X | \theta), \quad (6)$$

where  $F_X$  represents the resultant Fisher vector and  $\nabla_{\theta}$  is the gradient of the log-likelihood. Let  $x_t$  be a  $D$ -dimensional local descriptor, and its assignment to the  $i$ th Gaussian component (5) is represented as  $q_t(i)$ . Assuming that the covariance matrices  $\Sigma_i$  are diagonal and can be represented as  $\sigma_i$ , the gradient vector can be represented as in [35]:

$$u_i = \frac{1}{T \sqrt{w_i}} \sum_{t=1}^T q_t(i) \frac{x_t - \mu_i}{\sigma_i} \quad (7)$$

where mean  $\mu_i$  represents the mean of the  $i$ th component.

$$v_i = \frac{1}{T \sqrt{2w_i}} \sum_{t=1}^T q_t(i) \left[ \frac{(x_t - \mu_i)^2}{\sigma_i^2} - 1 \right], \quad (8)$$

where  $u_i$  and  $v_i$  are  $D$ -dimensional gradient vectors with respect to  $\mu_i$  and  $\sigma_i$ . The Fisher encoding for the set of local descriptors  $X$  is obtained as a concatenation of  $u$  and  $v$ .

$$f = [u_1^\top, v_1^\top, u_2^\top, v_2^\top, \dots, u_K^\top, v_K^\top]^\top \quad (9)$$

The final gradient vector  $f$  consists of  $u_i$  and  $v_i$  vectors for  $i = 1, 2, \dots, K$ . The size of  $f$  is  $2KD$ . The MBH<sub>x</sub>, MBH<sub>y</sub> and HOG descriptors are encoded as described above and they are fused using the representation level fusion [33]. Each of the feature vectors representing a set of local descriptors are computed separately and a global representation is obtained by concatenating them.

### 2.4 Classification

The similarity between two samples  $X$  and  $Y$  can be measured using the Fisher kernel (FK) [38]. It is defined as a dot-product between the feature vectors of  $X$  and  $Y$ :  $FK(X, Y) = f_X^\top f_Y$ . Here  $f_X$  and  $f_Y$  represent the Fisher vectors for samples  $X$  and  $Y$ , respectively. A non-linear kernel machine using FK as a kernel is similar to a linear kernel machine using  $f_X$  as feature vector. The main advantage of using such an explicit vector formulation is that we can exploit any simple linear classifier which can learn very efficiently. We used a simple linear SVM to solve this problem. In the implementation of the proposed algorithm, LIBLINEAR SVM library [13] is used to classify the encoded gait features. Before the actual model is trained on full training database, a 10-fold cross validation is performed to validate the model by selecting the optimal value of its parameters.

## 3 Experiments and Results

The performance of the proposed algorithm is evaluated on five large gait databases: TUM GAID [17], CMU MoBo [15], NLPR [55], CASIA-B [58] and CASIA-C [46] databases. Each database contains gait sequences captured in different environment with several variations in walking style, clothing, walking speed and etc. For example, in TUM GIAD database, three different walking styles are captured: normal walk, walk with coating shoes and walk with backpack. Moreover, the database is recorded in two different seasons, thus the clothing and the environment are significantly different. A summary of the characteristics of gait databases used in the experimental evaluation is presented in Tab. 1. For each database, a codebook is separately computed using a training set, a subset of the database, and is used to encode

**Table 1** Summary of gait databases used in performance evaluation. Size represents the number of gait video sequences in the database used in experimental evaluation.

Database	Environment	Size	Walk Scenarios
TUM GAID	Outdoor	3370	Left-to-right and right-to-left walk with three variations: normal walk, walk with coating shoes, walk with backpack
CMU MoBo	Indoor	300	Slow walk, fast walk, slow walk with ball in hands, slow walk at certain slope
NLPR	Outdoor	80	Left-to-right walk, right-to-left walk
CASIA-B	Indoor	1240	Normal walk, walk with coat, walk with bag
CASIA-C	Outdoor	1530	Normal walk, fast walk, slow walk, walk with backpack

the local descriptors of that particular database. In all experiments, one million randomly selected local descriptors from the training set are used to build a codebook with GMM. The number of components  $K$  in GMM are empirically computed and set to  $2^8$  in all experiments and the length of each local descriptor is 96. The same training and testing sets are used in evaluating the performance of the proposed and the compared gait recognition methods.

We also assess the influence of dimensionality reduction on the classification accuracy of our computed features. It is observed that dimensionality reduction significantly improves the recognition accuracy of the proposed method and reduces the computational time and the storage requirements. This improvement is due to the impact of PCA on GMM, which decorrelates the different dimensions during the projection, and since we are using a diagonal covariance matrix in Gaussian distribution, it is considered more advantageous [43]. Let  $d$  be the dimension of our feature vector and  $n$  be the total number of instances in the training and the testing sets. Since  $d \gg n$ , therefore we have applied PCA on our feature vectors and reduced their dimensions to  $n - 1$  to alleviate the curse of the dimensionality problem. It is empirically concluded that average recognition accuracy of the proposed algorithm on five gait databases is significantly increased (up to 27%), when the dimension of the final feature vector is reduced. Furthermore, using PCA the average percentage saving of memory on all the five databases is more than 50% and the average speedup achieved in classification time is more than 96%.

### 3.1 Evaluation on TUM GAID database

TUM GAID [17] is one of the biggest gait database, comprising 3,370 gait sequences of 305 subjects. The gait sequences were recorded in two different seasons. Each subject in the database has ten walk sequences which include six sequences of normal walk ( $N$ ), two sequences of walk with coating shoes ( $S$ ) and two sequences of walk with backpack

**Table 2** Performance comparison of proposed algorithm on TUM GAID database. The recognition results (%) of each experiment  $N$ ,  $S$ ,  $B$ ,  $TN$ ,  $TS$  and  $TB$  are described in the respective columns. Avg represents the average performance of the respective algorithm in all experiments. Best results are marked in bold.

Algorithm	$N$	$S$	$B$	$TN$	$TS$	$TB$	Avg
GEI [17]	99.4	56.2	27.1	44.0	9.0	6.0	56.0
GEV [17]	94.2	87.7	13.9	41.0	31.0	0.0	61.4
SEIM [56]	99.0	96.1	18.4	15.6	28.1	3.1	66.0
GVI [56]	99.0	94.5	47.7	62.5	62.5	15.6	77.3
SVIM [56]	98.4	91.6	64.2	65.6	50.0	31.3	81.4
DGHEI [17]	99.0	96.1	40.3	50.0	44.0	0.0	87.3
CNN-SVM [6]	99.7	97.1	97.1	59.4	<b>62.5</b>	50.0	94.2
CNN-NN128 [6]	99.7	95.8	98.1	62.5	59.4	56.3	94.2
H2M [4]	99.4	98.1	100.0	71.9	43.8	63.4	95.5
PFM [5]	99.7	99.0	99.0	<b>78.1</b>	54.9	62.0	96.0
Proposed	<b>99.7</b>	<b>99.7</b>	<b>100.0</b>	68.8	53.1	<b>71.9</b>	<b>96.5</b>

( $B$ ). A subset of 32 subjects in the database who took part in both recordings have ten more walk sequences (*i.e.* in total 20 sequences), and they are described as normal walk after time ( $TN$ ), walk with coating shoes after time ( $TS$ ) and walk with backpack after time ( $TB$ ). The time variation, where the shoes, clothing, illumination and other recording properties are significantly changed, make this database extremely challenging. A similar division of gait sequences into the training (*i.e.* gallery) and the testing (*i.e.* probe) sets is used, defined in [17]. Specifically, the first four sequences of  $N$  for each person (*i.e.*  $N_1 - N_4$ ) are used in the training set and  $N_5 - N_6$ ,  $S_1 - S_2$  and  $B_1 - B_2$  are used in the testing sets separately, with three different experiments, namely  $N$ ,  $S$  and  $B$ . Moreover, three more experiments are conducted namely  $TN$ ,  $TS$  and  $TB$  where the sequences of  $N_7 - N_8$ ,  $S_3 - S_4$  and  $B_3 - B_4$  are used in the testing set separately, while the training set is similar as in the previous set of experiments.

Recognition rate measure is used to assess the performance of gait recognition algorithms. The recognition results achieved by our algorithm and their comparison with the state-of-the-art techniques [4–6, 17, 56] are outlined in Tab. 2. The statistics show that in the first set of experiments  $N$ ,  $S$  and  $B$ , the proposed algorithm outperforms all competing methods. It achieved the recognition accuracy of 99.7%, 99.7% and 100%, respectively. In the next three experiments,  $TN$ ,  $TS$  and  $TB$ , our algorithm performs the best on  $TB$  experiment achieving 71.9% recognition accuracy. However, on  $TN$  and  $TS$  experiments, respectively, PFM [5] and CNN-SVM [6] perform better than our algorithm. Overall, our algorithm achieves the best average recognition accuracy of 96.5%.

### 3.2 Evaluation on CMU MoBo database

The CMU MoBo database [15] contains the gait videos of 25 subjects while they are walking on a treadmill. The database

**Table 3** Performance comparison of proposed algorithm on CMU MoBo gait database. The composition of training and testing sequences in each experiment is denoted as *Training – Testing*. The recognition results (%) of each experiment are described in the respective columns and average column demonstrates the average performance of the respective algorithm in all experiments. The first four columns describe the recognition accuracies when similar walking styles are used in the training and the testing sets while the other twelve columns demonstrate the accuracies with different walking styles. The best results are marked in bold.

Algorithm	<i>S–S</i>	<i>F–F</i>	<i>B–B</i>	<i>I–I</i>	<i>S–F</i>	<i>S–I</i>	<i>S–B</i>	<i>F–S</i>	<i>F–I</i>	<i>F–B</i>	<i>B–S</i>	<i>B–F</i>	<i>B–I</i>	<i>I–S</i>	<i>I–F</i>	<i>I–B</i>	Average
Partial silhouette [39]	88.0	88.0	-	-	32.0	-	-	28.0	-	-	-	-	-	-	-	-	59.0
Shape kinematics [49]	100.0	100.0	92.0	92.0	80.0	48.0	48.0	84.0	28.0	48.0	68.0	48.0	12.0	32.0	44.0	20.0	59.0
PFM [5]	-	-	-	-	92.0	100.0	<b>100.0</b>	92.0	96.0	83.3	48.0	48.0	44.0	100.0	96.0	87.5	82.2
GEI [41]	100.0	100.0	100.0	-	95.0	-	80.0	-	-	-	-	70.0	-	-	-	-	90.8
STM-SPP [9]	100.0	100.0	100.0	-	94.0	-	93.0	91.0	-	84.0	82.0	82.0	-	-	-	-	91.8
WBP [21]	100.0	100.0	100.0	98.67	92.0	-	72.67	92.0	-	60.67	74.67	63.33	-	-	-	-	85.4
Uniprojective [44]	100.0	100.0	96.0	-	72.0	-	-	60.0	-	-	-	-	-	-	-	-	85.6
NDDP [45]	100.0	100.0	96.0	-	88.0	-	-	80.0	-	-	-	-	-	-	-	-	92.8
HSD [20]	100.0	100.0	100.0	-	92.0	-	-	-	-	-	88.0	84.0	-	-	-	-	94.0
SDL [59]	100.0	100.0	98.7	-	96.0	-	86.7	92.0	-	88.0	86.7	88.0	-	-	-	-	92.9
Proposed	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	96.0	<b>100.0</b>	<b>100.0</b>	<b>96.0</b>	<b>96.0</b>	<b>96.0</b>	<b>96.0</b>	<b>100.0</b>	<b>100.0</b>	<b>96.0</b>	<b>98.5</b>

**Table 4** Performance comparison of proposed algorithm on NLPR gait database. The best recognition result (%) is marked in bold.

Algorithm	Accuracy	Algorithm	Accuracy
ICA [29]	82.5	Partial silhouette [39]	85.0
WBP [21]	100.0	HSD [20]	100.0
NN [22]	87.5	2D polar-plane [8]	92.5
3D deform. [14]	100	PSA [54]	88.8
STC+PCA [55]	82.5	GEI [32]	95.0
Proposed	<b>100.0</b>		

comprises four variations of walk: slow walk (*S*), walk with ball in hands (*B*), fast walk (*F*) and walk at certain slope (*I*). The sequences are recorded from six different viewing-angles. The gait sequences recorded in lateral view are used to assess the performance of the proposed algorithm in dealing with the variation in walking surface (*i.e.* incline), speed and carrying condition. In order to increase the number of gait sequences, similar to [59] we split the sequences into three equal parts. Two type of experiments are conducted, (1) similar walking styles are used in the training and the testing sets, (2) different walking styles are used in the training and the testing sets. The recognition accuracies obtained by the proposed algorithm and their comparison with competing methods are listed in Tab. 3. The proposed algorithm achieved excellent results, outperformed all competing methods in fifteen experiments and achieved the highest average recognition accuracy of 98.5%.

### 3.3 Evaluation on NLPR gait database

The NLPR gait database [55] comprises the walk sequences of 20 subjects. Each subject has four walk sequences captured from three different viewing-angles:  $0^\circ$ ,  $45^\circ$  and  $90^\circ$ . Each subject in the database walks twice, from right-to-left and from left-to-right. The performance of the proposed algorithm is evaluated on the gait sequences recorded at  $0^\circ$

**Table 5** Performance comparison of proposed algorithm on CASIA-B gait database. The recognition results (%) of each experiment *nm*, *cl* and *bg* are described in the respective columns. Avg represents the average performance of the respective algorithm in all three experiments. The best results are marked in bold.

Algorithm	<i>nm</i>	<i>cl</i>	<i>bg</i>	Avg
GEI [51]	91.6	24.0	31.7	49.1
CGI [51]	88.0	43.0	43.7	58.2
AEI+2DLPP [60]	98.4	72.2	91.9	87.5
Baseline method [58]	97.6	32.2	52.0	60.8
GEnI [2]	98.3	33.5	80.1	70.7
RF+FSS+CDA [12]	100.0	33.1	50.0	61.0
HSD [20]	94.5	58.1	62.9	71.8
DCS+H2M [4]	100.0	72.6	99.2	90.6
PFM [5]	100.0	85.5	100.0	95.2
SDL [59]	98.4	<b>90.3</b>	93.5	94.1
Proposed	<b>100.0</b>	89.9	<b>100.0</b>	<b>96.6</b>

viewing-angle. The first three recordings of each subject are used in the training set and the rest of the fourth recording is used in the testing set. The recognition result obtained by the proposed algorithm on NLPR database and its comparison with other competing techniques is reported in Tab. 4. The proposed algorithm along with [14, 20, 21] achieves the 100% recognition rate on NLPR gait database.

### 3.4 Evaluation on CASIA-B gait database

CASIA-B [58] is another large gait database comprising the gait sequences of 124 subjects. The sequences are captured from 11 different viewing-angles. Each subject in the database has ten sequences of gait with three variations: six sequences of normal walk (*nm*), two of walk while wearing a coat (*cl*) and two of walk with bag (*bg*). The sequences recorded in the lateral view are used in the evaluation, and three different experiments namely *nm*, *cl* and *bg* are conducted. In all experiments, the first four normal walk sequences from all 124 subjects are placed in the training set, and the rest

of two *nm*, *cl* and *bg* sequences are used in the testing set, separately. The recognition accuracies obtained by the proposed algorithm and their comparison with competing methods are presented in Tab. 5. The results reveal that the proposed algorithm outperforms the existing methods in *nm* and *bg*, while SDL [59] performs slightly better than our method in *cl*. Our method achieves the highest average recognition accuracy of 96.6%.

### 3.5 Evaluation on CASIA-C database

The CASIA-C database [46] comprises the gait videos of 153 subjects with four walking scenarios: normal walk (*fn*), fast walk (*fq*), slow walk (*fs*) and walk with backpack (*fb*). The database contains low resolution videos which are recorded at night using a thermal imaging camera. Each subject in the database has four recordings of *fn* and two recordings of each *fq*, *fs* and *fb*. This database aims at evaluating the performance of gait recognition methods under variations in the walking speed, carrying condition, and illumination changes. Four different experiments namely *fn*, *fq*, *fs* and *fb* are conducted. In all experiments, the first three walk sequences of *fn* from all 153 subjects are placed in the training set. In the first experiment the forth remaining sequence of *fn* is placed in the testing set while the rest of *fq*, *fs* and *fb* sequences are placed in the testing set, separately, for the next three experiments. The recognition results obtained by the proposed algorithm and the competing methods are reported in Tab. 6. The results shows that our method outperforms the existing methods in most experiments and obtains the highest average recognition accuracy of 99.8%.

One can conclude from the results presented in the previous sections that the proposed gait recognition algorithm performed consistently better than many state-of-the-art gait recognition methods and that it is strong in resisting the changes in walk due to variations in clothing, shoes, backpacks, bags, jackets, and illumination. The source code of the proposed gait recognition algorithm is publicly released to reproduce the results reported in this paper. The source code along with a sample database and computed features are made available online at <sup>1</sup>.

## 4 Conclusions

A new gait representation based on spatiotemporal characteristics of human motion is presented in this paper. Unlike most existing methods which require the segmented silhouette of human body region to compute features for gait recognition; the proposed method does not require such information, and it directly operates on the video sequences.

**Table 6** Performance comparison of proposed algorithm on CASIA-C gait database. The recognition results (%) of each experiment *fn*, *fq*, *fs* and *fb* are described in the respective columns. Avg represents the average performance of the respective algorithm in all four experiments. The best results are marked in bold.

Experiment	<i>fn</i>	<i>fq</i>	<i>fs</i>	<i>fb</i>	Avg
AEI+2DLPP [60]	88.9	90.2	89.2	79.7	87.0
WBP [21]	99.0	89.6	86.4	80.7	88.9
NDDP [45]	97.0	83.0	83.0	17.0	70.0
HSD [20]	97.0	89.0	86.0	65.0	84.2
HTI [46]	94.0	88.0	85.0	51.0	79.5
Uniprojective [44]	97.0	88.0	84.0	37.0	76.5
GEI [47]	90.0	80.0	-	70.0	80.0
RSM [16]	100.0	99.6	<b>99.7</b>	96.2	98.9
SDL [59]	95.4	92.5	91.2	81.7	90.2
PFM [5]	100.0	100.0	98.7	99.3	99.5
Proposed	<b>100.0</b>	<b>100.0</b>	99.4	<b>100.0</b>	<b>99.8</b>

The proposed method tracks a set of points in successive frames to compute the dense trajectories which are used to specify a spatiotemporal patch to compute the local descriptors in order to capture the person's static appearance and motion information. The local descriptors are encoded using Fisher vector encoding and classification is performed using linear SVM. The experimental evaluation on five large well-known benchmark gait databases showed that the proposed algorithm is highly accurate. In future, we plan to extend the proposed method for view-invariant gait recognition and also explore the probabilistic modeling techniques to be able to track the individuals even in occlusions.

## References

- Alotaibi, M., Mahmood, A.: Reducing covariate factors of gait recognition using feature selection and dictionary-based sparse coding. *Signal Image Video Process.* **11**(6), 1131–1138 (2017)
- Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: *IET ICDP*, pp. 1–6 (2009)
- Bouchrika, I., Nixon, M.: Model-based feature extraction for gait analysis and recognition. In: *IEEE ICCV*, pp. 150–160 (2007)
- Castro, F., Marín-Jiménez, M., Guil, N.: Multimodal features fusion for gait, gender and shoes recognition. *Mach. Vis. Appl.* pp. 1–16 (2016)
- Castro, F., et al.: Fisher motion descriptor for multiview gait recognition. *Int. J. Pattern Recognit. Artif. Intell.* **31**(01), 1756,002 (2017)
- Castro, F.M., Marín-Jiménez, M.J., Guil, N., Pérez de la Blanca, N.: Automatic learning of gait signatures for people identification. In: *Advances in Computational Intelligence*, pp. 257–270, 2017
- Chai, Y., et al.: A novel human gait recognition method by segmenting and extracting the region variance feature. In: *Proc. Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, pp. 425–428 (2006)
- Chen, S., Gao, Y.: An invariant appearance model for gait recognition. In: *Proc. IEEE Int. Conf. Multimed. and Expo (ICME)*, pp. 1375–1378. *IEEE* (2007)
- Choudhury, S.D., Tjahjadi, T.: Silhouette-based gait recognition using procrustes shape analysis and elliptic fourier descriptors. *Pattern Recognit.* **45**(9), 3414–3426 (2012)

<sup>1</sup> <http://www.pr.informatik.uni-siegen.de/en/gait-recognition-based-spatiotemporal-features-human-motion>

10. Cunado, D., Nixon, M.S., Carter, J.N.: Automatic extraction and description of human gait models for recognition purposes. *Comput. Vis. Image Underst.* **90**(1), 1–41 (2003)
11. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. Ser. B-Stat. Methodol.* pp. 1–38 (1977)
12. Dupuis, Y., Savatier, X., Vasseur, P.: Feature subset selection applied to model-free gait recognition. *Image Vis. Comput.* **31**(8), 580–591 (2013)
13. Fan, R.E., et al.: Liblinear: A library for large linear classification. *J. Mach. Learn. Res* **9**(Aug), 1871–1874 (2008)
14. Goffredo, M., Carter, J.N., Nixon, M.S.: Front-view gait recognition. In: 2nd IEEE Int. Conf. on Biometrics: Theory, Applications and Systems, pp. 1–6. IEEE (2008)
15. Gross, R., Shi, J.: The cmu motion of body (mobo) database. Carnegie Mellon University (2001)
16. Guan, Y., Li, C.T.: A robust speed-invariant gait recognition system for walker and runner identification. In: IEEE Int. Conf. on Biometrics (ICB), pp. 1–8 (2013)
17. Hofmann, M., Bachmann, S., Rigoll, G.: 2.5D gait biometrics using the depth gradient histogram energy image. In: Proc. IEEE BATS Conf., pp. 399–403 (2012)
18. Khan, M.H., et al.: Automatic recognition of movement patterns in the vojta-therapy using RGB-D data. In: Proc. Int. Conf. Image Process. (ICIP), pp. 1235–1239 (2016)
19. Khan, M.H., Li, F., Farid, M.S., Grzegorzec, M.: Gait recognition using motion trajectory analysis. In: Proc. 10th Int. Conf. on Computer Recognition Systems (CORES) 2017, pp. 73–82. Springer
20. Kusakunniran, W.: Attribute-based learning for gait recognition using spatio-temporal interest points. *Image Vis. Comput.* **32**(12), 1117–1126 (2014)
21. Kusakunniran, W., et al.: Automatic gait recognition using weighted binary pattern on video. In: Proc. 6th IEEE AVSS, pp. 49–54. (2009)
22. Lee, H., Hong, S., Kim, E.: An efficient gait recognition based on a selective neural network ensemble. *Int. J. Imaging Syst. Technol.* **18**(4), 237–241 (2008)
23. Lee, L., Grimson, W.E.L.: Gait analysis for recognition and classification. In: 5th IEEE Int. Conf. on Automatic Face and Gesture Recognit., pp. 155–162. IEEE (2002)
24. Lishani, A.O., Bouchir, L., Khalifa, E., Bouridane, A.: Human gait recognition based on haralick features. *Signal Image Video Process.* **11**(6), 1123–1130 (2017)
25. Little, R.J.A., Rubin, D.B.: *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics. Wiley, New York (1987)
26. Liu, J., et al.: Skeleton-based action recognition using spatio-temporal lstm network with trust gates. *IEEE Trans. Pattern Anal. Mach. Intell.* pp. 1–1 (2017).
27. Liu, J., et al.: Skeleton-based human action recognition with global context-aware attention lstm networks. *IEEE Trans. Image Process.* **27**(4), 1586–1599 (2018)
28. Loula, F., et al.: Recognizing people from their movement. *J. Exp. Psychol.-Hum. Percept.* **31**(1), 210 (2005)
29. Lu, J., Zhang, E., Jing, C.: Gait recognition using wavelet descriptors and independent component analysis. In: Int. Symposium on Neural Networks, pp. 232–237. Springer (2006)
30. Man, J., et al.: Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(2), 316–322 (2006)
31. McLachlan, G., Peel, D.: *Finite mixture models*. John Wiley & Sons (2004)
32. Nizami, I.F., et al.: Multi-view gait recognition fusion methodology. In: 3rd IEEE Conf. on Industrial Electronics and Applications, pp. 2101–2105. IEEE (2008)
33. Peng, X., Wang, L., Wang, X., Qiao, Y.: Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice. *Comput. Vis. Image Underst.* **150**, 109–125 (2016)
34. Peng, X., Zou, C., Qiao, Y., Peng, Q.: Action recognition with stacked fisher vectors. In: ECCV, pp. 581–595 (2014)
35. Perronnin, F., et al.: Improving the fisher kernel for large-scale image classification. In: ECCV, pp. 143–156 (2010)
36. Raheja, J.L., Chaudhary, A., Nandhini, K., Maiti, S.: Pre-consultation help necessity detection based on gait recognition. *Signal Image Video Process.* **9**(6), 1357–1363 (2015)
37. Rida, I., Almaadeed, S., Bouridane, A.: Gait recognition based on modified phase-only correlation. *Signal Image Video Process.* **10**(3), 463–470 (2016)
38. Sánchez, J., et al.: Image classification with the fisher vector: Theory and practice. *Int. J. Comput. Vis.* **105**(3), 222–245 (2013)
39. Shaikh, S.H., Saeed, K., Chaki, N.: Gait recognition using partial silhouette-based approach. In: Signal Processing and Integrated Networks (SPIN), 2014 Int. Conf. on, pp. 101–106 (2014)
40. Simonyan, K., Zisserman, A.: Two-stream convolutional networks for action recognition in videos. In: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (eds.) *Advances in Neural Information Processing Systems 27*, pp. 568–576 (2014)
41. Sivapalan, S., et al.: Gait energy volumes and frontal gait recognition using depth images. In: Proc. Int. Joint Conf. Biometrics (IJCB), pp. 1–6 (2011)
42. Stevenage, S.V., Nixon, M.S., Vince, K.: Visual analysis of gait as a cue to identity. *Appl. Cogn. Psychol.* **13**(6), 513–526 (1999)
43. Sun, C., Nevatia, R.: Large-scale web video event classification by use of fisher vectors. In: Proc. IEEE Workshop on Applications of Computer Vision (WACV), pp. 15–22. IEEE (2013)
44. Tan, D., Huang, K., Yu, S., Tan, T.: Uniprojective features for gait recognition. In: Proc. Int. Conf. Biom., pp. 673–682 (2007)
45. Tan, D., et al.: Walker recognition without gait cycle estimation. In: Proc. Int. Conf. Biom., pp. 222–231 (2007)
46. Tan, D., et al.: Efficient night gait recognition based on template matching. In: Proc. ICPR, vol. 3, pp. 1000–1003 (2006)
47. Vaidya, S., Shah, K.: Real time video surveillance system. *Int. J. Comput. Appl.* **86**(14) (2014)
48. Varol, G., Laptev, I., Schmid, C.: Long-term temporal convolutions for action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(6), 1510–1517 (2018)
49. Veeraraghavan, A., Chowdhury, A.R., Chellappa, R.: Role of shape and kinematics in human movement analysis. In: Proc. IEEE CVPR, vol. 1, pp. 1–730 (2004)
50. Veeraraghavan, A., et al.: Matching shape sequences in video with applications in human movement analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(12), 1896–1909 (2005)
51. Wang, C., et al.: Human identification using temporal information preserving gait template. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2164–2176 (2012)
52. Wang H., et al.: Dense Trajectories and Motion Boundary Descriptors for Action Recognition. *Int. J. Comput. Vis.* **103**(1), 60–79 (2013)
53. Wang, L., Ning, H., Tan, T., Hu, W.: Fusion of static and dynamic body biometrics for gait recognition. *IEEE Trans. Circuits Syst. Video Technol.* **14**(2), 149–158 (2004)
54. Wang, L., Tan, T., Hu, W., Ning, H.: Automatic gait recognition based on statistical shape analysis. *IEEE Trans. Image Process.* **12**(9), 1120–1131 (2003)
55. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(12), 1505–1518 (2003)
56. Whytock, T., et al.: Dynamic distance-based shape features for gait recognition. *J. Math. Imaging Vis.* **50**(3), 314–326 (2014)
57. Yang, Y., Tu, D., Li, G.: Gait recognition using flow histogram energy image. In: Proc. Int. Conf. Pattern Recognit. (ICPR), pp. 444–449 (2014)
58. Yu, S., et al.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: Proc. Int. Conf. Pattern Recognit. (ICPR), vol. 4, pp. 441–444 (2006)
59. Zeng, W., et al.: Silhouette-based gait recognition via deterministic learning. *Pattern Recognit.* **47**(11), 3568–3584 (2014)
60. Zhang, E., Zhao, Y., Xiong, W.: Active energy image plus 2dlpp for gait recognition. *Signal Process.* **90**(7), 2295–2302 (2010)