**Multimedia Tools and Application manuscript No.**
(will be inserted by the editor)

# A Generic Codebook based Approach for Gait Recognition

**Muhammad Hassan Khan**[1,2] · **Muhammad Shahid Farid**[2] · **Marcin Grzegorzek**[3]

Received: date / Accepted: date

**Abstract** Gait refers to the walking style of a person and it has emerged as an important biometric feature for person identification. The gait recognition algorithms proposed in literature exploit various types of information from the gait video sequence, such as, the skeletal data, human body shape, and silhouettes; and use these features to recognize the individuals. This paper presents the proposal of using a generic codebook in gait recognition. The idea is built upon a novel gait representation which exploits the spatiotemporal motion characteristics of the individual for identification. In particular, we propose to use a set of sample gait sequences to construct a generic codebook and use it to build a gait signature for person identification. To this end, we chose synthetic gait sequences of CMU MoCap gait database due to its diversity in walking styles. A set of spatiotemporal features are extracted from these sequences to build a generic codebook. The motion descriptors of real gait sequences are encoded using this generic codebook and Fisher vector encoding; the classification is performed using support vector machine. An extensive evaluation of this novel proposal is carried out using five benchmark gait databases: TUM GAID, CASIA-C, NLPR, CMU MoBo, and CASIA-B. In all experiments, the generic codebook is used in feature encoding. The performance of the proposed algorithm is also compared with the state-of-the-art gait recognition techniques and the results show that the idea of using a generic codebook in gait recognition is practical and effective.

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

✉ M.H. Khan E-mail: hassan.khan@uni-siegen.de

[1] Research Group for Pattern Recognition, University of Siegen, Siegen, Germany.

[2] Punjab University College of Information Technology, University of the Punjab, Pakistan.

[3] Institute of Medical Informatics, University of Lübeck, Germany.

## 1 Introduction

Biometrics refers to measure the biological and behavioral characteristics to authenticate the identities of people, and it has received significant research efforts in the recent years due to its growing applications in authentication, access control and surveillance. Studies [34, 48] have shown that individuals can be identified using different distinguishing biological traits. These include: fingerprints, iris, DNA, facial features, earlobes structure, voice, and gait, which have been proven to be unique for each individual. Gait is a behavioral biometric that seeks to identify people using the way they walk. Person identification using gait has gained a wide interest in the community because of its advantages of unobtrusive and obtainable from a distance. This interest is strongly driven by the need of automated systems for person identification in security-sensitive and monitoring environments such as banks, military bases, airports and etc. In contrast to the conventional biometric features, gait does not require human interaction with the system which makes it the most suitable for surveillance systems. Moreover, gait biometrics can be collected at low resolution in a non-invasive and hidden manner. Although gait has some benefits over physiological biometrics, however it is challenging as many factors may affect it, such as, variation in clothing and footwear, walking speed and surface, injuries and other similar reasons. Gait may not be as powerful as other biometric modalities such as fingerprints to identify the individuals. Nevertheless, its ability to recognize human from distance and without any interaction with the system makes it irreplaceable in many applications such as visual surveillance.

Existing gait recognition methods can be classified into two broad categories: (1) model-based approaches and (2) model-free approaches. Model-based techniques build a gait signature using the human body structure and motion models [53]. The structural models [9, 16, 75] which may include stick figure, interlinked pendulum and ellipse fitting techniques are generally constructed based on the prior knowledge of human body shape. The motion models [12, 46, 60] exploit the motion information of the human body parts, such as, joint angle trajectories, rotation patterns of hip and thigh. Recent studies [9, 37, 51, 80] claim that such models are able to deal with the occlusion and the rotation problems to some extent. However, their performance highly depends on the localization of torso which is not easy to extract from the underlying model. These approaches are computationally expensive and are sensitive to the quality of video data, and therefore they are not considered suitable for real-world applications [80].

The model-free approaches do not use structural or human motion models; instead they usually operate on the sequence of extracted binary silhouettes of human from gait images. Such algorithms either construct a template image [5, 13, 29, 56, 57] or use the temporal information of human motion [8, 10, 25, 36, 38, 41] from the sequence of silhouettes and use them to recognize the individuals. In contrast to model-based gait recognition approaches, the model-free techniques generally perform better. Moreover, they are computationally efficient too. Although, numerous techniques have been proposed in literature claiming the excellent performance but they are sensitive to variations in silhouette shapes, walking surface, clothing and other similar reasons. Moreover, their performance also depend on the accuracy of human body silhouette segmentation [82], which is still a challenging problem in the literature.

In our recent work [40], a novel gait representation is proposed which is based on the spatial and temporal gait characteristics. In particular, it extracts dense trajectories from gait video sequence by tracking a set of points in the successive frames and formulates the local motion descriptors which are used to identify the individuals. This novel gait representation has a number of advantages over the conventional gait features. For example, in contrast to the existing gait recognition algorithms, the method proposed in [40] does not require the extraction of human body silhouette, contour, or other skeletal information. Moreover, it is a model-free approach and it does not involve any kind of human body segmentation or the gait-cycle computation. Therefore, its performance is particularly better than other competing approaches under low-resolution and low quality video sequences. In this paper, we present the concept of a generic codebook for gait recognition. The idea is applicable in the encoding of any feature that use a codebook to encode the local descriptors. However, due to excellent performance of our recent gait feature [40], we chose it to implement the concept of generic codebook. The advantages and major contributions of this paper are as follows:

- A novel proposal of a generic codebook is presented to encode the motion descriptors of the gait sequences. To the best of our knowledge it is the first time that a generic codebook approach is proposed for feature encoding in gait recognition algorithm. The idea of using the generic codebook for gait recognition can achieve a number of advantages over the conventional database-specific codebooks, including:
  - Usually a codebook is generated for specific gait recognition scenarios and it can be used to recognize the individuals in that particular environment only. Whereas the generic codebook can be effectively used to build a gait signature for any type of walk sequence.
  - A generic codebook generated from a large set of walking styles can serve as a universal codebook and can be used to encode any kind of individual's walk.
- To obtain a generic codebook, we chose the synthetic gait video sequences from CMU Motion Capture (mocap) database [1] due to its diverse walking styles. An extensive experimental evaluation is performed to assess the performance of the proposed technique on five benchmark gait databases, including both outdoor (TUM GAID, CASIA-C, NLPR) and indoor (CMU MoBo, CASIA-B) gait databases. The results are also compared with the state-of-the-art gait recognition techniques. Experimental evaluation demonstrates that using the generic codebook approach is both time and space efficient, and it achieves excellent recognition results.
- The influence of dimensionality reduction in the feature encoding in terms of recognition accuracy, computation time, and space requirements is also investigated and important findings are presented.

The rest of the paper is organized as follows: Sect. 2 reviews the related literature. Sect. 3 describes the proposed generic codebook for gait recognition. Sect. 4 outlines the proposed gait recognition algorithm based on the generic codebook. Experimental evaluation and comparisons with the state-of-the-art are presented in Sect. 5. In Sect. 6, we analyze the performance of the proposed method with database-specific codebook and also evaluate the impact of Principal Component Analysis (PCA) on the proposed gait features. The conclusions of this research are documented in Sect. 7.

## 2 Related Work

The existing gait recognition algorithms are generally categorized into model-based and model-free approaches.

### 2.1 Model-based gait recognition

The model-based gait recognition techniques construct the structural models or motion models from the human body shape and use them to identify the walkers. Lee et al. [46] proposed the modeling of human body structure using seven different ellipses representing the various body regions. They computed several statistical measurements on these regions over time such as: mean, standard deviation, location of its centroid, magnitude and phase of these moment based regions for gait and gender classification. The authors in [75] built a human body model using fourteen rigid parts connected to each other at joint locations and computed the joint angle trajectories at these locations to form a gait signature. Sivapalan et al. [60] used the ellipsoids fitting technique into four different components of lower limbs to build a 3D voxel model. The features derived from the ellipsoids are modeled using a Fourier representation to recognize the individual's gait. Bouchrika et al. [9] built a motion model using the elliptic Fourier descriptors to extract features from human joints and incorporated them to establish a model for person identification. Cunado et al. [16] proposed a gait feature by computing the angular motion of the hip and thigh using the Fourier series. The authors in [12] split the human body region into three parts and the variance of these parts over time are combined and used as gait signature. Model based approaches support view invariance but they are computationally expensive and sensitive to the video quality as well [9, 80].

### 2.2 Model-free gait recognition

The model-free gait recognition approaches usually operate on the sequence of extracted human silhouettes. Such approaches either construct a template image, compute the temporal information of human motion, or use the statistical measurements from silhouette shape for individual recognition. The classical gait recognition approach using template image is proposed in [52]. The human body silhouettes are extracted using background modeling and averaged over time in a gait-cycle to obatin a representation known as gait energy image (GEI). A number of extensions in GEI are also proposed, *e.g.,* [5, 28, 67, 73, 83]. Whytock et al. [78] proposed a skeleton based descriptor known as Skeleton Variance Image (SVIM). They extracted skeleton information from the silhouette images and combined it with the motion information to form a gait representation. Zeng et al. [82] used the statistical measurements, such as, height to width ratio, silhouette area, width of the contour and centroid of the contour, computed from person's silhouette to estimate the individual's gait. In [26], height and width of normalized and scaled human silhouette over time are used for gait recognition. Tan et al. [64] developed a normalized pseudo-height and width (NPHW) histogram using silhouette images for gait recognition. The authors in [28] proposed a histogram binning to capture the edges and depth gradient of depth silhouette to recognize the gait. The authors in [47, 50] analyzed the principal components' coefficients obtained from the silhouette images and wavelet descriptors using Independent Component Analysis (ICA) to get the more independent gait features.

Shape analysis of human's silhouette is also exploited in many gait recognition techniques. Wang et al. [76] used the Procrustes Mean Shape (PMS) computed from the sequence of silhouettes as gait signature. PMS represents the both motion and body shape into a unified descriptor and similarity is measured using Procrustes Distance. Benabdelkader et al. [7] used the concept of Self Similarity Plot (SSP) to encode the projection of gait dynamics. The SSP used to construct a gait descriptor comprises a matrix of cross-correlation between each pair of silhouette in the sequences. Wang et al. [77] plot the 2D silhouettes into 1D normalized distance signal using contour unwrapping. The variations in the shape of 1D silhouette over time are used to approximate the gait pattern. Dadashi et al. [17] improved [77] and analyzed the 1D signals through wavelet packet transform and classified them using transductive support vector machine.

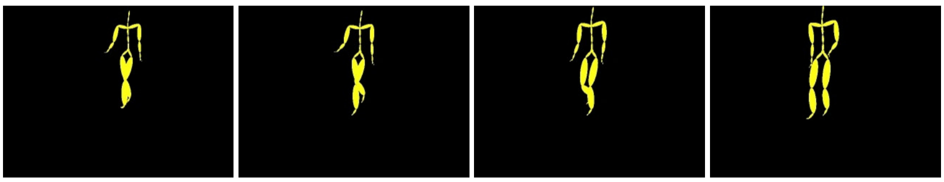
Motion information from gait video sequences is also exploited for person identification. In [11], spatiotemporal cuboids of optical flow are used to obtain a high level gait representation. Kusakunniran et al. [41] extracts the space-time interest points from the video sequences in spatiotemporal domain and proposed a Histogram of Space-Time Interest Points Descriptors (HSD) for gait recognition. The method proposed in [15] used spatiotemporal motion characteristics, and statistical and physical parameters of silhouette's contour for recognition. Hu et al. [30] used local binary patterns computed from optical flow to form the representation of gait and temporal relationship are learned incrementally for detection.

The silhouette based model-free approaches are computationally efficient and they have demonstrated convincing recognition results on various benchmark gait databases. However, they are generally very sensitive to variations in the silhouette shapes and thus are highly dependent on the precise silhouette segmentation. An inaccurate segmentation may lower the recognition accuracy [82]. In contrast, the proposed approach is model-free and does not involve any kind of human body segmentation or other gait related characteristics such as gait cycle estimation.

## 3 Proposed Generic Codebook for Gait Recognition

In this section, we present the idea of generic codebook and describe the algorithm to generate it using the Gaussian mixture model (GMM). This codebook is then used with the algorithm proposed in [40] for gait recognition. We name the proposed technique 'Gait Recognition using Generic Codebook' (GRGC).

Usually the codebooks are application specific and are generated using a subset of the dataset, we use either for training or for validation. The idea of 'generic codebook' advocates the concept of using a single codebook built on one dataset, and to encode the sequences of other datasets. Thus, selection of dataset used to generate the generic codebook is important as it directly deals with feature encoding and may effects the recognition performance of algorithm. For this purpose, a dataset with a large variety of walking styles is momentous for good performance of gait recognition algorithm. A codebook obtained from a dataset with limited walking styles would loose important gait cues in encoding, resulting in poor performance. To this end, we explored various existing gait datasets and found that a subset of CMU Motion Capture (mocap) database [1] is the most appropriate for generic codebook



(a) 65th frame (b) 70th frame (c) 75th frame (d) 80th frame

**Fig. 1** Sample synthetic images from mocap dataset used to obtain the generic codebook.

**Table 1** Details of CMU Motion Capture (mocap) gait database.

| Number Video Sequences | 81 |
|---|---|
| Normal walk | 73 |
| Slow walk | 5 |
| Brisk walk | 1 |
| Stride walk | 2 |
| Frame Rate | 30 f/s |
| Resolution | 320×240 |

generation. We chose the synthetic video sequences of walk from mocap dataset to generate the generic codebook. These sequences cover a large range of walk types, *e.g.,* normal, slow, fast, exaggerated stride, brisk walk, wander, etc. A total of 81 video sequences are used to generate the generic codebook. Fig. 1 shows few sample synthetic images of a normal walk sequence from the selected dataset. Further details of this dataset are presented in Tab. 1.

The gait recognition algorithm [40] that we are aiming to use with the proposed generic codebook, exploits spatiotemporal features to capture the distinctive motion characteristics of human gait. To build the generic codebook, the spatiotemporal motion features are extracted from each mocap sequence using optical flow field and their motion information is encoded in local descriptors. It is important to mention here that we do not use the motion descriptors of gallery or probe sets from any other gait database.

## 3.1 Feature Extraction

Numerous feature computation techniques have been proposed in recent years and have been successfully exploited in various computer vision problems; SIFT (Scale Invariant Feature Transform) [49], SURF (Speeded-Up Robust Feature) [6], HOG (Histogram of Oriented Gradient) [18], HOF (Histogram of Optical Flow) [44], MBH (Motion Boundary Histogram) [19], 2DMED(Two-Dimensional Maximum Embedding Difference) [71], locality preserving-based algorithms [70, 72], and trajectory [74] are a few to mention. Recently, dense trajectories have shown excellent results in action recognition [54, 74]. Our motivation to use dense trajectories for gait recognition is because they encode the local motion patterns and they can be easily extracted from video sequences.

To obtain trajectories, a set of dense points is selected from each frame and tracked in successive frames using the displacement information from a dense optical flow field. Let $P_t = (x_t, y_t)$ be a point in frame $t$ which is tracked in frame $t+1$ by median filtering in a dense optical flow field. The set of points in subsequent frames are concatenated to form a trajectory: $P_t, P_{t+1}, P_{t+2}, \ldots, P_{t+L}$. The displacement vector $\Delta P_t$ is defined as the distance between the corresponding points $P_t$ and $P_{t+1}$, given as:

$$\Delta P_t = P_{t+1} - P_t = [x_{t+1} - x_t, y_{t+1} - y_t]$$

The sequence of displacement vectors represents the shape of the trajectory which describes the local motion pattern. Given a trajectory of length $L$, a sequence $S$ of displacement vectors is formed.

$$S = [\Delta P_t, \Delta P_{t+1}, \ldots, \Delta P_{t+L-1}], \tag{1}$$

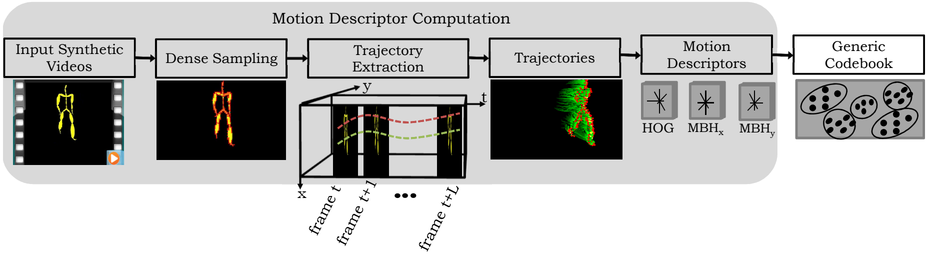



**Fig. 2** Proposed generic codebook generation for gait recognition.

The sequence vector $S$ is normalized to obtain the trajectory $\mathscr{T}$:

$$\mathscr{T} = \frac{1}{\sum_{j=t}^{t+L-1} \left\| \Delta P_j \right\|} \cdot [\Delta P_t, \Delta P_{t+1}, \ldots, \Delta P_{t+L-1}] \tag{2}$$

Wang et al. [74] proposed the computation of HOG and HOF along the dense trajectories. The spatial derivatives are also computed along the horizontal and the vertical components of the optical flow and their orientation information is quantized into histograms. These descriptors (*i.e.,* histograms) encode the relative motion information between pixels along the respective axis, known as $\mathrm{MBH}_x$ and $\mathrm{MBH}_y$ respectively. In [40] different descriptors including, HOG, HOF, MBH, and their various combinations are evaluated for their performance. The empirical evaluations performed on TUM GAID database showed that HOG and MBH together performs the best. Therefore, we chose HOG and MBH ($\mathrm{MBH}_x$, $\mathrm{MBH}_y$) descriptors to construct our gait signature.

### 3.2 Generic Codebook Generation

To build the generic codebook, we randomly selected one million features from each motion descriptor of mocap sequences. The GMM describes the distribution over feature space [35] and can be expressed as:

$$p(X;\theta) = \sum_{i=1}^{K} w_i \mathscr{N}(X;\mu_i,\Sigma_i) \tag{3}$$

where $X$ is local motion descriptors (in our experiment we use HOG and MBH), $\theta = \{w_i, \mu_i, \Sigma_i \,|\, i = 1, 2, ...., K\}$ is the set of model parameters, $i$ is the number of mixture component, $w_i$ is the weight of $i$th component, and $\mathscr{N}(X;\mu_i,\Sigma_i)$ represents the $D$-dimensional Gaussian distribution with mean vector $\mu_i$ and covariance matrix $\Sigma_i$. For a given feature set $X = \{x_1, \ldots, x_t\}$, the optimal parameters of GMM are learned through maximum likelihood estimation [21]. The soft assignment of descriptor $x_t$ to cluster $i$, also known as posterior probability is defined as,

$$q_t(i) = \frac{w_i \mathscr{N}(x_t;\mu_i,\Sigma_i)}{\sum_{j=1}^{K} w_j \mathscr{N}(x_t;\mu_j,\Sigma_j)} \tag{4}$$

We assume that each model represents a specific motion pattern shared by the descriptors in the codebook. Unlike the k-means clustering, which performs hard assignment, the Expectation maximization (EM) algorithm [2] of GMM performs soft assignments of feature descriptor to each mixture component and it provides not only the mean information of code words, but also the shape of their distribution. Therefore, the local descriptors will be assigned to multiple components in a weighted manner using the posterior component probability given by the descriptor. Fig. 2 demonstrates the process of generating a generic codebook.

## 4 Generic Codebook based Gait Recognition

The proposed generic codebook is used with the gait recognition algorithm presented in [40]. The algorithm extracts the spatiotemporal features HOG and MBH from the real gait video sequences, as described in Sect. 3.1. These descriptors are encoded using Fisher vector encoding [58] and generic codebook computed in Sect. 3.2. The Fisher vector representation contains the information of local descriptors by its deviation from the generative model (*i.e.,* GMM) and this deviation is computed using the gradient of the log-likelihood with respect to the model parameters. Specifically, it comprises the average first and second order differences of descriptors from the centers of GMM.

### 4.1 Feature encoding

For a given feature set $X = \{x_t, t = 1, ...., T\}$ from local descriptors, it is modeled into a vector using the probability density function $p(X;\theta)$ (3). $X$ can be mapped into a vector by computing the gradient vector of its log-likelihood function with respect to the current model parameters $\theta$ [62]:

$$F_X = \frac{1}{T} \nabla_\theta logp(X;\theta), \tag{5}$$

where $F_X$ is representing the FV and $\nabla_\theta$ is the gradient of the log-likelihood function which describes the contribution of parameters in the model generation. Let $x_t$ is a $D$-dimensional local descriptor, $q_t(i)$ is the soft assignments of descriptor $x_t$ to $i$th Gaussian component (4). Assuming that the covariance matrices $\sum_i$ are diagonal and can be represented as $\sigma_i$. The gradient vectors with respect to mean $\mu_i$ and covariance $\sigma_i$ are defined as [55]:

$$u_i = \frac{1}{T\sqrt{w_i}} \sum_{t=1}^{T} q_t(i) \frac{x_t - \mu_i}{\sigma_i} \tag{6}$$

$$v_i = \frac{1}{T\sqrt{2w_i}} \sum_{t=1}^{T} q_t(i) \left[ \frac{(x_t - \mu_i)^2}{\sigma_i^2} - 1 \right], \tag{7}$$

where $u_i$ and $v_i$ are $D$-dimensional gradient vectors and also known as the first order and second order differences of local descriptor to Gaussian components, respectively. The Fisher encoding for the set of local descriptors $X$ is computed by concatenating the all $u$ and $v$ for all $K$ components. That is,

$$f = [u_1^\top, v_1^\top, u_2^\top, v_2^\top, ....., u_K^\top, v_K^\top]^\top \tag{8}$$

Since, the final gradient vector $f$ consists of $u_i$ and $v_i$ vectors for $i = 1, 2, ...., K$ components, and each vector is $D$-dimensional, therefore, the total size of encoded vector is $2KD$. We encode the HOG, $MBH_x$ and $MBH_y$ descriptors using the above formulation and fused them in a representation level fusion, as described in our earlier work [40]. A block diagram of the proposed gait recognition algorithm is shown in Fig. 3.

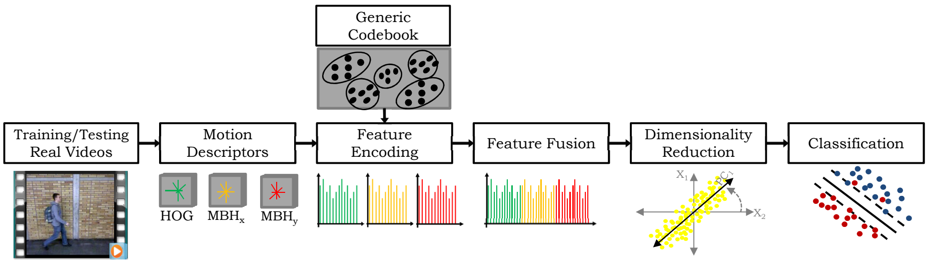



**Fig. 3** Proposed gait recognition algorithm using generic codebook.

## 4.2 Classification

The similarity between two samples $X$ and $Y$ can be measured using the Fisher kernel (FK) [58], which can be defined as a dot-product between the feature vectors of $X$ and $Y$. That is,

$$FK(X,Y) = f_X' \cdot f_Y, \tag{9}$$

where $f_X$ and $f_Y$ are representing 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 linear support vector machine (SVM) to solve this problem. The SVM has emerged as an efficient tool for large sparse dataset with huge number of instances and features. In the implementation of the proposed algorithm, LIBLINEAR SVM library[1] [23] is used to classify the encoded gait features.

Support vector machine is considered a powerful tool for solving classification problems in many computer vision applications [33, 39]. It first maps the training samples in high dimensional space and then extracts a hyper-plane between the different classes of objects using the principle of maximizing the margin. Because of this principle, the generalization error of SVM is theoretically independent from the feature dimension [33]. In a given set of labeled instances $(x_i, y_i), x_i \varepsilon \mathbb{R}$ and $y_i = \{-1, +1\}$, the following optimization problem is solved,

$$\min_{w} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} max(0, 1 - y_i w^T x_i)^2, \tag{10}$$

where $w$ is known as the weight vector, $C > 0$ is the penalty parameter and $max(0, 1 - y_i w^T x_i)^2$ is the loss function. The objective is to maximizing the margin, *i.e.,* minimizing the regularization term $\|w\|^2$ augmented with a term $C\sum_{i=1}^{N} max(0, 1 - y_i w^T x_i)^2$ to penalize the mis-classification and margin errors. The soft margin parameter $C$ plays an important role in maximizing the margin and minimizing the loss function. In particular, we performed 10-fold cross validation to validate the model with the selection of $C$ before training the actual model on the full training database.

[1] https://www.csie.ntu.edu.tw/~cjlin/liblinear/

**Fig. 4** Two sample images from NLPR gait database in a lateral view.

## 5 Experimental Evaluation and Results

The performance of the proposed method is evaluated on five benchmark gait databases: NLPR [77], CMU MoBo [27], TUM GAID [28], CASIA-B [81], and CASIA-C [67]. We also compare the performance of the proposed method with the state-of-the-art gait recognition algorithms. The comparison is performed in terms of recognition accuracy and execution time complexity. In all experiments, the generic codebook is used in the proposed algorithm to compute the recognition results. To build a generic codebook, we computed the local motion descriptors from the synthetic video sequences in mocap dataset and one million features are randomly selected from each local descriptor. The codebook's clustering size ($K$) is empirically chosen and is fixed to 256 in this study. The local descriptors of real video sequences are encoded using the above formulation, as described in Sect. 4. The size of encoded vector is $2KD$, where $D$ is the 96 dimensional long local descriptor.

The proposed method is implemented using Matlab software and experiments are carried out on Intel core i5 2.6GHz computer with 8GB RAM.

### 5.1 Performance on NLPR Gait Database

The NLPR gait database comprises the walk sequences of 20 subjects. The database was captured in an outdoor environment and each subject has four sequences of walk in three different viewing angles. The video sequences were captured at 30 frames/second (f/s) and the original video resolution is $352 \times 240$. In all experiment, we used the sequences recorded in the lateral view. Fig. 4 shows two sample lateral view images from the database. We assign three sequences of each subject to gallery set and the forth one to probe set. Tab. 2 shows the performance of the proposed method and the results achieved by the other competing techniques. The results show that the proposed algorithm outperforms all existing methods and achieves 100% average recognition accuracy.

**Table 2** Comparison of average recognition accuracy (%) on NLPR gait database. Best results are marked in bold.

| Method | Accuracy |
|---|---|
| Wavelet+ICA [50] | 82.5 |
| Partial silhouette [59] | 85.0 |
| PSC [43] | 97.5 |
| NN [45] | 87.5 |
| 2D polar-plane [14] | 92.5 |
| Gait+Face+Dynamic [79] | 90.0 |
| Gait+Face+Distance [24] | 90.0 |
| PSA [76] | 88.8 |
| Curves+NN [61] | 89.3 |
| STC+PCA [77] | 82.5 |
| Proposed | **100.0** |

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

**Fig. 5** Sample images from CMU MoBo gait database. (a) Slow walking, (b) fast walking, (c) slow walking with ball and (d) walking at certain slope (*i.e.,* incline).

## 5.2 Performance on CMU MoBo Gait Database

The CMU Motion of Body (MoBo) gait database contains the gait sequences of 25 subjects. The database was recorded in an indoor environment while the subjects were walking on a treadmill. Each subject in the database has four different walk patterns: slow walk ($S$), fast walk ($F$), slow walk with a ball in hands ($B$), and incline walk ($I$) at slope of $15°$. The video sequences were recorded at 30 f/s, using 6 different viewing angles. The size of the videos is $640 \times 480$ in 24-bit color resolution. Fig. 5 shows few sample images from the database to demonstrate the different walking scenarios.

The video sequences in lateral view are used to evaluate the performance of proposed method under different conditions: walk surface, walk speed and carrying conditions. To increase the number of instances for classification, all the video sequences are divided into three sub-sequences. We conduct two different types of experiments:

1. within the same condition: where the same type of walking scenarios are used in probe and gallery set,
2. across the condition: where different type of walking scenarios are used in probe and gallery set.

Tab. 3 summarizes the description of sixteen experiments on CMU MoBo gait database. The performance comparison of proposed method with state-of-the-art techniques is outlined in

**Table 3** List of sixteen experiments on CMU MoBo gait database.

(a) Same condition experiments.

| Exp. | Gallery Set | Probe Set |
|---|---|---|
| A | Slow walk | Slow walk |
| B | Fast walk | Fast walk |
| C | Walk with ball | Walk with ball |
| D | Incline walk | Incline walk |

(b) Across condition experiments.

| Exp. | Gallery Set | Probe Set |
|---|---|---|
| E | Slow walk | Fast walk |
| F | Slow walk | Incline walk |
| G | Slow walk | Walk with ball |
| H | Fast walk | Slow walk |
| I | Fast walk | Incline walk |
| J | Fast walk | Walk with ball |
| K | Walk with ball | Slow walk |
| L | Walk with ball | Fast walk |
| M | Walk with ball | Incline walk |
| N | Incline walk | Slow walk |
| O | Incline walk | Fast walk |
| P | Incline walk | Walk with ball |

**Table 4** Comparison of recognition results (%) on CMU MoBo gait database. Each column corresponds to a different experiment and average (Avg) is computed as the mean score of all the experiments. Best results are marked in bold.

| Methods | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* | *J* | *K* | *L* | *M* | *N* | *O* | *P* | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SC [69] | 100 | 100 | 92 | 92 | 80 | 48 | 48 | 84 | 28 | 48 | 68 | 48 | 28 | 32 | 44 | 0 | 60 |
| Partial silhouette [59] | 88 | 88 | - | - | 32 | - | - | 28 | - | - | - | - | - | - | - | - | 59 |
| ICA [47] | 100 | 100 | 100 | 100 | - | - | 79.2 | 64 | - | - | - | - | - | - | - | - | 90.5 |
| SSP [7] | 100 | 100 | - | - | 54 | - | - | 32 | - | - | - | - | - | - | - | - | 71.5 |
| Eigen features [31] | 95.8 | 95.8 | 95.4 | - | - | - | - | 75 | - | - | - | - | - | - | - | - | 90.5 |
| HMM [32] | 72 | 68 | 91 | - | 56 | - | - | 59 | - | - | - | - | - | - | - | - | 69.2 |
| Shape kinematics [68] | 100 | 100 | 92 | 92 | 80 | 48 | 48 | 84 | 28 | 48 | 68 | 48 | 12 | 32 | 44 | 0 | 59 |
| STM-SPP [15] | 100 | 100 | 100 | - | 94 | - | 93 | 91 | - | 84 | 82 | 82 | - | - | - | - | 91.8 |
| 3D ellipsoid [60] | 100 | 100 | 100 | - | 78.6 | - | 70.5 | - | - | - | - | 61 | - | - | - | - | 85.1 |
| WBP [42] | 100 | 100 | 100 | 98.7 | 92 | - | 72.67 | 92 | - | 60.7 | 74.7 | 63.3 | - | - | - | - | 85.4 |
| Uniprojective [65] | 100 | 100 | 96 | - | 72 | - | - | 60 | - | - | - | - | - | - | - | - | 85.6 |
| NDDP [66] | 100 | 100 | 96 | - | 88 | - | - | 80 | - | - | - | - | - | - | - | - | 92.8 |
| HSD [41] | 100 | 100 | **100** | - | 92 | - | - | - | - | - | 88 | 84 | - | - | - | - | 94 |
| SDL [82] | 100 | 100 | 98.7 | - | 96 | - | 86.7 | 92 | - | 88 | 86.7 | 88 | - | - | - | - | 92.9 |
| Proposed | **100** | **100** | 96 | **100** | **100** | **96** | **92** | **100** | **100** | **96** | **92** | **92** | **88** | **100** | **100** | **88** | **96.3** |

Tab. 4. The statistics show that the proposed method achieves excellent results in all the experiments, except experiment *C* where HSD [41] performs better than our method. It can be noted from the results that the performance of the proposed technique is consistently better than other competing algorithms on almost all experiments. In particular, in the difficult set of experiments *E* to *P*, the compared techniques performs rather poor and our method consistently achieves more than 90% recognition accuracy in most of the experiments. The proposed method obtained the highest average recognition rate 96.3% outperforming all the compared methods.

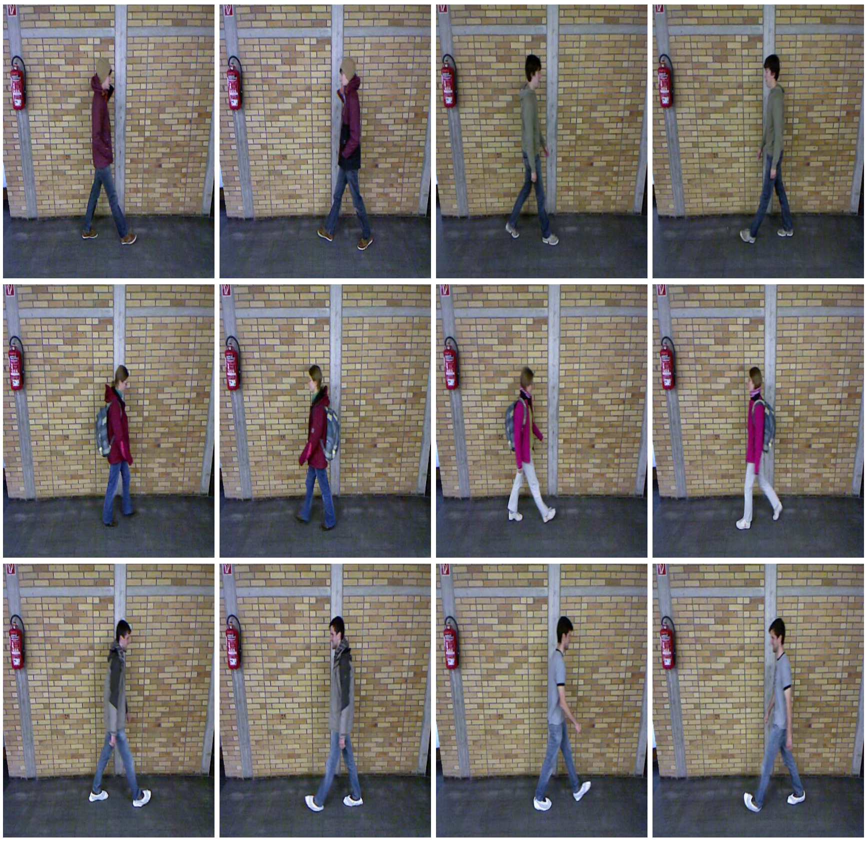


**Fig. 6** Sample images from TUM GAID database. Top row: Normal walking, middle row: walking with backpack, and bottom row: walking with coating shoes. The first two images of each row are representing the walking sequences (from left-to-right and right-to-left), recorded in the first session and the rest two are recorded in second session for the same subject.

## 5.3 Performance on TUM GAID database

The TUM Gait Audio Image and Depth (TUM GAID) database is one of the largest gait database. It comprises gait video sequences of 305 subjects and in total $3,370$ video sequences. The database was recorded at 30 f/s using Microsoft Kinect in two different sessions at outdoor environment of Technical University of Munich, Germany. The first session of recording was held in January 2012, which are the coldest days of the year in the region and temperature is around -5°C. Therefore, the subjects in these videos were wearing heavy jackets and winter boots. The gait sequences of 176 subjects were recorded in this session. The second session of recording was held in April 2012 and the temperature was around +15°C in the region. Therefore, the subjects were wearing significantly different clothes and shoes from the first session. The gait sequences of 161 subjects were recorded in this session. There is a subset of 32 subjects who participated in both recording sessions, thus the database comprises recording of 305 subjects. The variation in clothing, shoes, lighting and other captured properties make this database extremely challenging in the field of gait

**Table 5** Walk variations and abbreviations in TUM GAID database.

| Winter Session | | Summer Session | |
|---|---|---|---|
| Walk Type | Symbol | Walk Type | Symbol |
| Normal walking | N | Normal walking after time | TN |
| Walking with backpack | B | Walking with backpack after time | TB |
| Walking with coating shoes | S | Walking with coating shoes after time | TS |

**Table 6** Comparison of recognition results (%) on TUM GAID gait database. The symbols $N$, $B$, $S$, $TN$, $TB$, and $TS$ represents the experiments and average (Avg) is computed as sum of the weighted mean scores. Best results are marked in bold.

| Method | $N$ | $B$ | $S$ | $TN$ | $TB$ | $TS$ | Avg |
|---|---|---|---|---|---|---|---|
| GEI [28] | 99.4 | 27.1 | 56.2 | 44.0 | 6.0 | 9.0 | 56.0 |
| GEV [28] | 94.2 | 13.9 | 87.7 | 41.0 | 0.0 | 31.0 | 61.4 |
| SEIM [78] | 99.0 | 18.4 | 96.1 | 15.6 | 3.1 | 28.1 | 66.0 |
| GVI [78] | 99.0 | 47.7 | 94.5 | 62.5 | 15.6 | 62.5 | 77.3 |
| SVIM [78] | 98.4 | 64.2 | 91.6 | 65.6 | 31.3 | 50.0 | 81.4 |
| DGHEI [28] | 99.0 | 40.3 | 96.1 | 50.0 | 0.0 | 44.0 | 87.3 |
| CNN-SVM [11] | 99.7 | 97.1 | 97.1 | 59.4 | 50.0 | **62.5** | 94.2 |
| CNN-NN128 [11] | 99.7 | 98.1 | 95.8 | 62.5 | 56.3 | 59.4 | 94.2 |
| Proposed | **99.7** | **99.0** | **98.4** | **68.8** | **56.3** | 53.1 | **95.3** |

recognition. Fig. 6 display few sample images from the TUM GAID database to show the substantial variation in the appearance of participants.

The database contains the ten video sequences for each subject with three different variations listed in Tab. 5. In all experiments, we used the same division of gallery and probe sets as outlined in [28]. In this division, first four recordings of normal walk are assigned to gallery set and the remaining two sequence of normal walk, walk with backpack and walk with coating shoes are assigned to probe set, separately, in three different experiments namely: $N$, $B$ and $S$. For the rest of experiments, the gallery set was same but the gait sequences of normal walk after time, walk with backpack after time and walk with coating shoes after time are used in probe set separately, namely: $TN$, $TB$ and $TS$, respectively. The performance comparison of the proposed method with the state-of-the-art techniques is outlined in Tab. 6. The results reveal that our method outperforms all the compared methods in all experiments, except in $TS$ where the CNN-SVM [11] performs better than the proposed method. The average recognition accuracy of the proposed algorithm turns to the highest 95.3% amongst the competing methods.

### 5.4 Performance on CASIA-B gait database

The CASIA-B is a large database containing the gait sequences of 124 subjects. The video sequences are recorded at 25 f/s using 11 different viewing angles in a controlled laboratory environment. The walk sequences are recorded in three different variations in walking style: normal walking (*nm*), walk with bag (*bg*) and walking in a coat (*cl*). Ten video sequences are recorded for each subject, including: six sequences of *nm* and two sequences of each *bg* and *cl*. Fig. 7 displays few sample images from the database in lateral view demonstrating the variations in the walking styles in this database.

(a) (b) (c)

**Fig. 7** Sample images from CASIA-B gait database. (a) Normal walking, (b) walking in coat and (c) walk with bag in a lateral view.

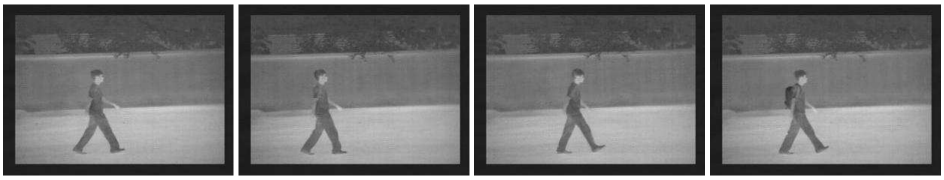
**Table 7** Comparison of recognition results (%) on CASIA-B gait database in lateral view. Each column of *nm*, *bg* and *cl* corresponds to a different experiment and the average (Avg) is computed as the mean score of all the experiments. Best results are marked in bold.

| Experiment | *nm* | *bg* | *cl* | Avg |
|---|---|---|---|---|
| TM [3] | 97.6 | 52.0 | 32.7 | 60.8 |
| GEI [73] | 91.6 | 31.7 | 24.0 | 49.1 |
| CGI [73] | 88.0 | 43.7 | 43.0 | 58.2 |
| iHMM [30] | 94.0 | 45.2 | 42.9 | 60.7 |
| AEI+2DLPP [83] | 98.4 | 91.9 | 72.2 | 87.5 |
| Baseline method [81] | 97.6 | 52.0 | 32.2 | 60.8 |
| GEnI [4] | 98.3 | 80.1 | 33.5 | 70.7 |
| RF+FSS+CDA [22] | 100.0 | 50.0 | 33.1 | 61.0 |
| HSD [41] | 94.5 | 62.9 | 58.1 | 71.8 |
| SDL [82] | 98.4 | 93.5 | **90.3** | 94.1 |
| Proposed Method | **100.0** | **98.4** | 86.7 | **95.0** |

In our experimental evaluations, we used the videos recorded in the lateral view. The first four sequences of *nm* are assigned to the gallery set and the remaining sequences of normal walk, walk with bag and walk in coat for all 124 subjects are assigned to probe set in three different experiments separately, namely: *nm*, *bg* and $cl$. The results achieved by our method and the state-of-the-art techniques are outlined in Tab. 7. The SDL algorithm [82] performs better then our method in *cl* experiment. In *nm* and *bg* experiments, the proposed method outperforms all the compared methods, and it achieves the best average recognition accuracy 95.0%.

## 5.5 CASIA-C gait database

The CASIA-C is another large gait database comprising the gait sequences of 153 subjects. The video sequences are recoded at 25 f/s under different conditions at night using a low-resolution thermal camera. The database contains walk sequences with four variations: normal walking ($fn$), slow walking ($fs$), fast walking ($fq$), and walking with backpack ($fb$). Fig. 8 shows the sample walking styles in CASIA-C database. Each subject in the database has ten sequences of gait, including, four sequences of normal walking and two sequences for each of the rest walking styles. The proposed method is evaluated on CASIA-C gait database to demonstrate its robustness in terms of carrying objects, walking speed and illumination conditions during the walk.



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

**Fig. 8** Sample images from CASIA-C gait database. (a) Normal walking, (b) slow walking, (c) fast walking and (d) walk with backpack.

**Table 8** Comparison of recognition results (%) on CASIA-C gait database. Each column of $fn$, $fs$, $fq$ and $fb$ corresponds to a different experiment. Average (Avg) is computed as the mean score of all the experiments. Best results are marked in bold.

| Experiment | $fn$ | $fs$ | $fq$ | $fb$ | Avg |
|---|---|---|---|---|---|
| AEI+2DLPP [83] | 88.9 | 89.2 | 90.2 | 79.7 | 87.0 |
| WBP [42] | 99.0 | 86.4 | 89.6 | 80.7 | 88.9 |
| NDDP [66] | 97.0 | 83.0 | 83.0 | 17.0 | 70.0 |
| OP [63] | 98.0 | 80.0 | 80.0 | 16.0 | 68.5 |
| HSD [41] | 97.0 | 86.0 | 89.0 | 65.0 | 84.2 |
| Wavelet packet [17] | 93.0 | 83.0 | 85.0 | 21.0 | 70.5 |
| Pseudo shape [64] | 98.4 | 91.3 | 93.7 | 24.7 | 77.03 |
| Gait curves [20] | 91.0 | 65.4 | 69.9 | 25.5 | 62.9 |
| HTI [67] | 94.0 | 85.0 | 88.0 | 51.0 | 79.5 |
| NDDP [66] | 97.0 | 83.0 | 83.0 | 17.0 | 70 |
| Uniprojective [65] | 97.0 | 84.0 | 88.0 | 37.0 | 76.5 |
| SDL [82] | 95.4 | 91.2 | 92.5 | 81.7 | 90.2 |
| Proposed | **100.0** | **99.0** | **100.0** | **99.7** | **99.6** |

For the first experiment denoted as $fn$, we assigned the first three sequences of normal walk to the gallery set and the remaining forth sequence is assigned to probe set. In the next three experiments, namely: $fs$, $fq$ and $fb$ the gallery set was the same and the gait sequences of slow walking, fast walking and walking with backpack for all 153 subjects are separately used in probe set. Tab. 8 summarizes the recognition results achieved by the proposed and the other competing methods. The results show that our method outperforms the existing methods in all experiments. Overall, the proposed method obtained the highest average recognition rate 99.6%.

The results presented in Tables 2, 4 and 6– 8 confirm the effectiveness of the proposed gait recognition algorithm. All the five benchmark gait databases used in performance evaluation contain a diverse variety of gait sequences. The proposed algorithms showed very convincing results outperforming the state-of-the-art in most experiments. High recognition accuracy also confirm that the proposed generic codebook is capable to effectively encode the gait features of various walking styles. In particular, the average recognition performance of the proposed algorithm is the highest on all five gait databases.

## 6 Performance Analysis & Discussion

In this section, we analyze the performance of the proposed method with generic codebook and database-specific codebook to quantitatively assess the benefits of using the generic

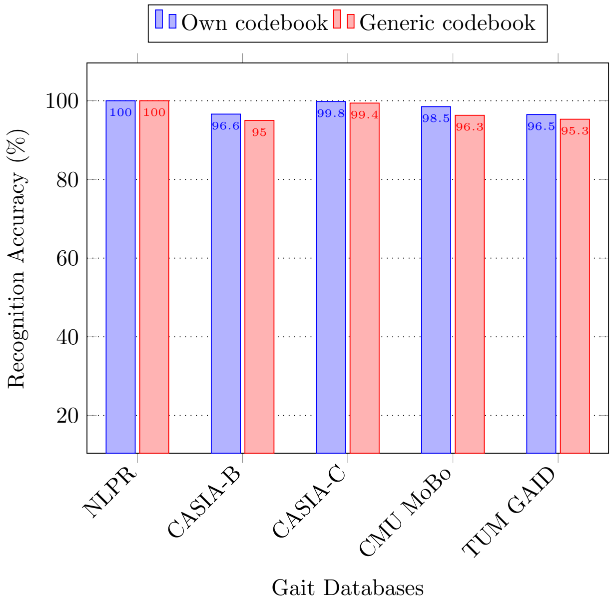



**Fig. 9** Performance comparison of the proposed algorithm on all five gait databases with database-specific codebook and the generic codebook.

codebook. We also investigate the impact of employing the Principal Component Analysis (PCA) to the encoded features in our proposed method, in terms of recognition accuracy, classification time, and space requirement.

## 6.1 Performance analysis with generic codebook vs. database-specific codebook

From the previous section, we recall that the proposed generic codebook based gait recognition technique demonstrated promising recognition results on all the five benchmark gait databases, outperforming the compared methods in most of the experiments. However, analysis of the performance of proposed method using the gait database-specific codebooks instead of generic codebook would be an interesting investigation. To this end, we computed a separate codebook for each gait database and repeated all the experiments reported in Sect. 5 to measure the recognition accuracy. The average recognition results achieved by the proposed algorithm on each database using its own codebook and using the generic codebook are illustrated in Fig. 9.

The results show that the proposed algorithm with database-specific codebook performs marginally better than the generic codebook; the average difference of recognition accuracy is around 1%, which reveals that the idea of generic codebook is effective for gait recognition. The number of advantages we achieve with this small loss in accuracy are manifold, such as, in contrast to database-specific codebook where it is computed for each database, the generic codebook is generated once for all databases, thus it is computationally efficient and it also requires less storage. In case of database-specific codebook, change in the gait database would require the regeneration of the codebook, this however, would not be required in the generic codebook. Moreover, the motion descriptors of a new gait database

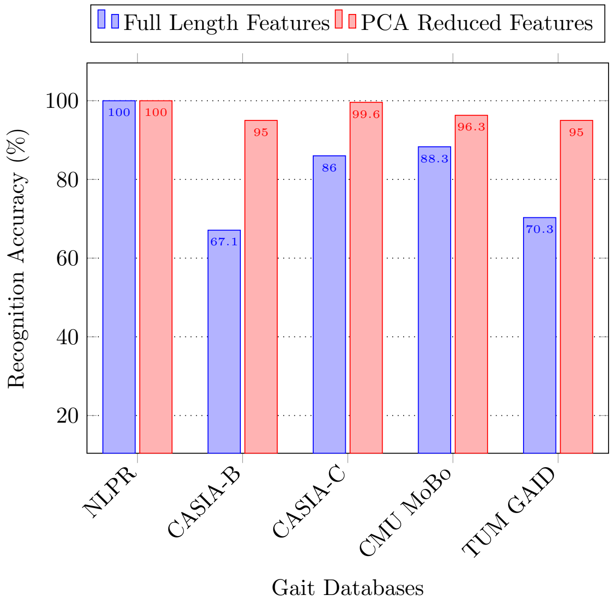



**Fig. 10** Impact on the recognition accuracy of the proposed generic codebook based algorithm with and without applying PCA on encoded features. The graph shows the average recognition accuracy on each database.

can be directly encoded using the generic codebook, without the need of recomputing the codebook.

### 6.2 Impact of using PCA on the performance of the proposed algorithm

In the next set of experiments, we analyze the influence of Principal Component Analysis (PCA) for dimensionality reduction on our encoded features. We compute recognition accuracy, classification time, and the space complexity to objectively assess the impact of using PCA reduced feature and full-length feature with the proposed algorithm. The higher dimension features inevitably increase the computational complexity and it requires more storage, making the real-time implementation difficult and costly. The PCA is exploited to map the data into a lower dimensional space by preserving as much data variance as possible. This leads to easier model building and provides more discriminative features to the gait recognition system. Let $l$ is the length of the encoded feature ($l = 2KD$, where $K$ is the number of clusters and $D$ is the dimension of local descriptors, having values 256 and 96 respectively [40].) and $n$ be the total number of features in both the gallery and the probe sets. Since, $l \gg n$, the dimension of each feature is reduced to $n-1$. The recognition accuracy achieved by the proposed algorithm with PCA reduced features and full-length features on all the five gait databases is presented in Fig. 10. The results show that with reduced dimension features, the classification accuracy of the proposed algorithm is significantly higher (up to 27%) than the full-length features.

We also analyze the impact of applying PCA on our encoded features for the classification time and the space requirement. For each database, we computed the size (in Megabytes) of the features before and after applying PCA to estimate the percentage mem-

**Table 9** Analysis of classification time and space requirement of the proposed algorithm on full length and PCA reduced length encoded features.

| Database | Size (MB) | | | Time (Sec.) | | |
|---|---|---|---|---|---|---|
| | Full | Reduced | % Saving | Full | Reduced | % Speedup |
| NLPR | 71.90 | 0.05 | 99.93 | 5.44 | 0.02 | 99.58 |
| CASIA-B | 917.70 | 8.60 | 99.06 | 172.88 | 13.51 | 92.18 |
| CASIA-C | 2252.80 | 20.80 | 99.08 | 2663.94 | 38.08 | 98.57 |
| CMU | 164.90 | 0.19 | 99.88 | 30.81 | 0.14 | 99.55 |
| TUM | 1638.40 | 24.80 | 98.49 | 441.48 | 39.90 | 90.96 |
| Avg. | - | - | 99.29 | - | - | 96.17 |

ory saving achieved due to PCA. The execution time for the classification of proposed algorithm with full-length and reduced-length encoded features is also computed and the speedup gain is calculated. The results are presented in Tab. 9. The results show that applying PCA on encoded features has significantly reduced the classification time and the space requirements. The average percentage saving of memory (space) is more than 99% and the average speedup achieved with PCA is more than 96%. One can note from the results presented in Fig. 10 and Tab. 9, that the PCA based feature reduction not only improves the recognition accuracy, it also significantly reduces the classification time and the storage requirement.

The proposed generic codebook, source code of the proposed gait recognition approach, and a sample database with computed features are made available online at[2] for the use of the biometric research community and to make the results reproducible, reported in this paper.

## 7 Conclusion

In this paper, a new generic codebook based gait recognition approach is presented. The idea of a generic codebook is proposed for gait biometric which presents a number of advantages over the conventional database-specific codebooks. For instance, rather than computing a database-specific codebook for each database, the generic codebook is generated once for all databases; it is computationally efficient and requires less storage. Moreover, the motion descriptors of a new gait database can be directly encoded using the generic codebook, without the need of recomputing the codebook. The generic codebook is built using the motion descriptors of a large synthetic gait sequences with a variety of gait styles. A gait recognition algorithm based on spatiotemporal motion characteristics and the generic codebook to encode the gait features is proposed. In contrast to existing gait based person identification techniques, the proposed model-free approach does not require any kind of human body segmentation or gait cycle estimation. The recognition results on five large benchmark gait databases confirm the effectiveness of proposed method. Though the gait sequences recorded in lateral view contains the most significant gait characteristics of a person, however in future we plan to extend the proposed method to recognize the unconstrained movements.

[2] `http://www.di.unito.it/~farid/Research/GRGC.html`

## References


1. CMU motion capture database. `http://mocap.cs.cmu.edu/`
2. Anzai, Y.: Pattern Recognition and Machine Learning. Elsevier (2012)
3. Bashir, K., Xiang, T., Gong, S.: Feature selection for gait recognition without subject cooperation. In: BMVC, pp. 1–10 (2008)
4. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: IET ICDP, pp. 1–6 (2009)
5. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognit. Lett. **31**(13), 2052–2060 (2010)
6. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: ECCV 2006, pp. 404–417 (2006)
7. BenAbdelkader, C., Cutler, R.G., Davis, L.S.: Gait recognition using image self-similarity. EURASIP J. Adv. Signal Process. **2004**(4), 1–14 (2004)
8. Bouchrika, I., Carter, J.N., Nixon, M.S.: Towards automated visual surveillance using gait for identity recognition and tracking across multiple non-intersecting cameras. Multimed. Tools Appl. **75**(2), 1201–1221 (2016)
9. Bouchrika, I., Nixon, M.: Model-based feature extraction for gait analysis and recognition. In: ICCV, pp. 150–160. Springer (2007)
10. Castro, F., Marín-Jiménez, M., Mata, N., Muñoz-Salinas, R.: Fisher motion descriptor for multi-view gait recognition. Int. J. Pattern Recognit. Artif. Intell. **31**(01), 1756002 (2017). DOI 10.1142/S021800141756002X. URL `http://www.worldscientific.com/doi/abs/10.1142/S021800141756002X`
11. Castro, F.M., Marín-Jiménez, M.J., Guil, N., de la Blanca, N.P.: Automatic learning of gait signatures for people identification. In: International Work-Conference on Artificial Neural Networks, pp. 257–270. Springer (2017)
12. 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)
13. Chen, C., et al.: Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognit. Lett. **30**(11), 977–984 (2009)
14. 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)
15. Choudhury, S.D., Tjahjadi, T.: Silhouette-based gait recognition using procrustes shape analysis and elliptic fourier descriptors. Pattern Recognit. **45**(9), 3414–3426 (2012)
16. 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)
17. Dadashi, F., et al.: Gait recognition using wavelet packet silhouette representation and transductive support vector machines. In: 2nd Int. Congress on Image and Signal Process., pp. 1–5 (2009)
18. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE CVPR, vol. 1, pp. 886–893 vol. 1 (2005). DOI 10.1109/CVPR.2005.177
19. Dalal, N., Triggs, B., Schmid, C.: Human detection using oriented histograms of flow and appearance. In: ECCV, pp. 428–441 (2006)
20. DeCann, B., Ross, A.: Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment. In: SPIE Defense, Security, and Sensing, pp. 76670Q–76670Q. International Society for Optics and Photonics (2010)
21. 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. **1977**, 1–38 (1977)
22. Dupuis, Y., Savatier, X., Vasseur, P.: Feature subset selection applied to model-free gait recognition. Image Vis. Comput. **31**(8), 580–591 (2013)
23. Fan, R.E., et al.: Liblinear: A library for large linear classification. J. Mach. Learn. Res **9**(Aug), 1871–1874 (2008)
24. Geng, X., Wang, L., Li, M., Wu, Q., Smith-Miles, K.: Distance-driven fusion of gait and face for human identification in video. In: Image and Vision Computing Conference. Image and Vision Computing New Zealand (2007)
25. Goffredo, M., Bouchrika, I., Carter, J.N., Nixon, M.S.: Performance analysis for automated gait extraction and recognition in multi-camera surveillance. Multimed. Tools Appl. **50**(1), 75–94 (2010)
26. Goffredo, M., Carter, J.N., Nixon, M.S.: Front-view gait recognition. In: IEEE Int. Conf. Biometrics: Theory, Appl. and Systems (BTAS), pp. 1–6. IEEE (2008)
27. Gross, R., Shi, J.: The CMU motion of body (mobo) database (2001). Gait Video Sequences
28. Hofmann, M., Bachmann, S., Rigoll, G.: 2.5D gait biometrics using the depth gradient histogram energy image. In: IEEE BATS Conf., pp. 399–403 (2012)

29. Hu, M., Wang, Y., Zhang, Z.: Cross-view gait recognition with short probe sequences: From view transformation model to view-independent stance-independent identity vector. Int. J. Pattern Recognit. Artif. Intell. **27**(06), 1350017 (2013). DOI 10.1142/S0218001413500171. URL `http://www.worldscientific.com/doi/abs/10.1142/S0218001413500171`
30. Hu, M., Wang, Y., Zhang, Z., Zhang, D., Little, J.J.: Incremental learning for video-based gait recognition with lbp flow. IEEE Trans. Cybern. **43**(1), 77–89 (2013)
31. Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A., Chellappa, R.: Gait analysis for human identification. In: Int. Conf. on Audio-and Video-Based Biometric Person Authentication, pp. 706–714. Springer (2003)
32. Kale, A., Sundaresan, A., Rajagopalan, A., Cuntoor, N.P., Roy-Chowdhury, A.K., Kruger, V., Chellappa, R.: Identification of humans using gait. IEEE Trans. Image Process. **13**(9), 1163–1173 (2004)
33. Khan, M., 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)
34. Khan, M.H.: Human Activity Analysis in Visual Surveillance and Healthcare, vol. 45. Logos Verlag Berlin GmbH (2018)
35. Khan, M.H., Farid, M.S., Grzegorzek, M.: Person identification using spatiotemporal motion characteristics. In: Proc. Int. Conf. Image Process. (ICIP), pp. 166–170. IEEE (2017)
36. Khan, M.H., Farid, M.S., Grzegorzek, M.: Using a generic model for codebook-based gait recognition algorithms. In: Int. Workshop Biometrics Forensics (IWBF), pp. 1–7. IEEE (2018)
37. Khan, M.H., Farid, M.S., Grzegorzek, M.: Spatiotemporal features of human motion for gait recognition. Signal Image Video Process. **13**(2), 369–377 (2019)
38. Khan, M.H., Farid, M.S., Zahoor, M., Grzegorzek, M.: Cross-view gait recognition using non-linear view transformations of spatiotemporal features. In: Proc. Int. Conf. Image Process. (ICIP), pp. 773–777. IEEE (2018)
39. Khan, M.H., Helsper, J., Farid, M.S., Grzegorzek, M.: A computer vision-based system for monitoring vojta therapy. J. Med. Informat. **113**, 85–95 (2018)
40. Khan, M.H., Li, F., Farid, M.S., Grzegorzek, M.: Gait recognition using motion trajectory analysis. In: M. Kurzynski, M. Wozniak, R. Burduk (eds.) Proc. Int. Conf. Comput. Recognit. Systems (CORES), *Advances in Intelligent Systems and Computing*, vol. 578, chap. 8, pp. 73–82. Springer, Cham (2017)
41. Kusakunniran, W.: Attribute-based learning for gait recognition using spatio-temporal interest points. Image Vis. Comput. **32**(12), 1117–1126 (2014)
42. Kusakunniran, W., Wu, Q., Li, H., Zhang, J.: Automatic gait recognition using weighted binary pattern on video. In: IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS), pp. 49–54. IEEE (2009)
43. Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Pairwise shape configuration-based psa for gait recognition under small viewing angle change. In: IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS), pp. 17–22. IEEE (2011)
44. Laptev, I., Marszalek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: IEEE CVPR, pp. 1–8 (2008). DOI 10.1109/CVPR.2008.4587756
45. Lee, H., Hong, S., Kim, E.: An efficient gait recognition based on a selective neural network ensemble. Int. J. Imaging Syst. and Technol. **18**(4), 237–241 (2008)
46. Lee, L., Grimson, W.E.L.: Gait analysis for recognition and classification. In: Proc. Int. Conf. Automatic Face and Gesture Recognit., pp. 155–162. IEEE (2002)
47. Liang, J., Chen, Y., Hu, H., Zhao, H.: Appearance-based gait recognition using independent component analysis. In: Int. Conf. on Natural Computation, pp. 371–380. Springer (2006)
48. Loula, F., Prasad, S., Harber, K., Shiffrar, M.: Recognizing people from their movement. J. Exp. Psychol.-Hum. Percept. **31**(1), 210 (2005)
49. Lowe, D.G.: Object recognition from local scale-invariant features. In: EEE ICCV, vol. 2, pp. 1150–1157 vol.2 (1999). DOI 10.1109/ICCV.1999.790410
50. Lu, J., Zhang, E., Jing, C.: Gait recognition using wavelet descriptors and independent component analysis. In: Int. Symp. Neural Networks, pp. 232–237. Springer (2006)
51. Lun, R., Zhao, W.: A survey of applications and human motion recognition with microsoft kinect. Int. J. Pattern Recognit. Artif. Intell. **29**(05), 1555008 (2015). DOI 10.1142/S0218001415550083. URL `http://www.worldscientific.com/doi/abs/10.1142/S0218001415550083`
52. Man, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
53. Nixon, M., et al.: Model-based gait recognition. In: Enclycopedia of Biometrics, pp. 633–639. Springer (2009). URL `https://eprints.soton.ac.uk/268238/`
54. 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)
55. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: ECCV, pp. 143–156. Springer (2010)

56. Rokanujjaman, M., Islam, M.S., Hossain, M.A., Islam, M.R., Makihara, Y., Yagi, Y.: Effective part-based gait identification using frequency-domain gait entropy features. Multimed. Tools Appl. **74**(9), 3099–3120 (2015)
57. Samangooei, S., Nixon, M.S.: Performing content-based retrieval of humans using gait biometrics. Multimed. Tools Appl. **49**(1), 195–212 (2010)
58. Sánchez, J., et al.: Image classification with the fisher vector: Theory and practice. Int. J. Comput. Vis. **105**(3), 222–245 (2013)
59. Shaikh, S.H., Saeed, K., Chaki, N.: Gait recognition using partial silhouette-based approach. In: Int. Conf. Signal Process. and Integrated Netw. (SPIN), pp. 101–106. IEEE (2014)
60. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: 3d ellipsoid fitting for multi-view gait recognition. In: IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS), pp. 355–360. IEEE (2011)
61. Su, H., Huang, F.: Gait recognition using principal curves and neural networks. In: Int. Symp. Neural Networks, pp. 238–243. Springer (2006)
62. Sun, C., Nevatia, R.: Large-scale web video event classification by use of fisher vectors. In: IEEE Int. Workshop Appl. Comput. Vis. (WACV), pp. 15–22. IEEE (2013)
63. Tan, D., Huang, K., Yu, S., Tan, T.: Orthogonal diagonal projections for gait recognition. In: Proc. Int. Conf. Image Process. (ICIP), vol. 1, pp. I–337. IEEE (2007)
64. Tan, D., Huang, K., Yu, S., Tan, T.: Recognizing night walkers based on one pseudoshape representation of gait. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 1–8. IEEE (2007)
65. Tan, D., Huang, K., Yu, S., Tan, T.: Uniprojective features for gait recognition. In: Int. Conf. Biometrics (ICB), pp. 673–682. Springer (2007)
66. Tan, D., Yu, S., Huang, K., Tan, T.: Walker recognition without gait cycle estimation. In: Int. Conf. on Biometrics, pp. 222–231 (2007)
67. Tan, D., et al.: Efficient night gait recognition based on template matching. In: Proc. Int. Conf. Pattern Recognit. (ICPR), vol. 3, pp. 1000–1003 (2006)
68. Veeraraghavan, A., Chowdhury, A.R., Chellappa, R.: Role of shape and kinematics in human movement analysis. In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 1, pp. I–730. IEEE (2004)
69. Veeraraghavan, A., Roy-Chowdhury, A.K., Chellappa, R.: Matching shape sequences in video with applications in human movement analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**(12), 1896–1909 (2005)
70. Wan, M., Lai, Z., Yang, G., Yang, Z., Zhang, F., Zheng, H.: Local graph embedding based on maximum margin criterion via fuzzy set. Fuzzy Sets Syst. **318**, 120–131 (2017)
71. Wan, M., Li, M., Yang, G., Gai, S., Jin, Z.: Feature extraction using two-dimensional maximum embedding difference. Information Science. **274**, 55–69 (2014)
72. Wan, M., Yang, G., Gai, S., Yang, Z.: Two-dimensional discriminant locality preserving projections (2ddlpp) and its application to feature extraction via fuzzy set. Multimed. Tools Appl. **76**(1), 355–371 (2017)
73. Wang, C., et al.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
74. Wang, H., Schmid, C.: Action recognition with improved trajectories. In: IEEE ICCV, pp. 3551–3558 (2013)
75. 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)
76. 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)
77. 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)
78. Whytock, T., Belyaev, A., Robertson, N.: Dynamic distance-based shape features for gait recognition. J. Math. Imaging Vis. **50**(3), 314–326 (2014)
79. Wu, Q., Wang, L., Geng, X., Li, M., He, X.: Dynamic biometrics fusion at feature level for video based human recognition. In: Proc. of Image and Vis. Computing New Zealand, pp. 152–157. Citeseer (2007)
80. Yang, Y., Tu, D., Li, G.: Gait recognition using flow histogram energy image. In: Proc. Int. Conf. Pattern Recognit. (ICPR), pp. 444–449 (2014)
81. 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)
82. Zeng, W., Wang, C., Yang, F.: Silhouette-based gait recognition via deterministic learning. Pattern Recognit. **47**(11), 3568–3584 (2014)

83. Zhang, E., Zhao, Y., Xiong, W.: Active energy image plus 2dlpp for gait recognition. Signal Processing **90**(7), 2295–2302 (2010)

**Muhammad Hassan Khan** obtained the B.S. and M.Phil. degrees in Computer Science from BZ University and University of the Punjab, Pakistan respectively, and the Ph.D. in computer science from University of Siegen, Germany. He is currently an Assistant Professor with the College of Information Technology, University of the Punjab, Lahore, Pakistan. His research interests include Pattern Recognition and Machine Learning techniques particularly in visual surveillance and vision based biometric recognition, and human activity recognition in the challenging healthcare or therapeutic interventions systems based on algorithmically monitored health condition to stimulate the physical and mental well-being.

**Muhammad Shahid Farid** received the B.S., M.Sc., and M.Phil. degrees in computer science from the University of the Punjab, Lahore, Pakistan, in 2004, 2006, and 2009, respectively. He received his Ph.D. from the Dipartimento di Informatica, Università degli Studi di Torino, Torino, Italy in 2015. He worked as Postdoc researcher for a short term in 2017 at the Department of Computer Science and Engineering, Qatar University, Doha, Qatar. Currently, Dr. Farid is an Assistant Professor with College of Information Technology, University of the Punjab, Lahore, Pakistan. He worked on 3D television technology particularly on efficient representation & coding of multiview videos, novel view synthesis techniques and quality assessment of 3D videos. His research interests also include information fusion, biometrics, image segmentation and medical image analysis. He has been a member of Technical Program Committee for several conferences, including, IEEE FIT, ICACS, and

ICOSST, and served as reviewer in numerous journal.

**Marcin Grzegorzek** is Professor of Medical Informatics with a research focus on Medical Data Science at the University of Lübeck, Germany. He was leading the Research Group for Pattern Recognition at the University of Siegen, Germany from October 2010 to September 2018. He studied Computer Science at the Silesian University of Technology, did his PhD at the Pattern Recognition Lab at the University of Erlangen-Nuremberg, worked scientifically as Postdoc in the Multimedia and Vision Research Group at the Queen Mary University of London and at the Institute for Web Science and Technologies at the University of Koblenz-Landau, did his habilitation at the AGH University of Science and Technology in Kraków. He published more than 100 papers in pattern recognition, image processing, machine learning, and multimedia analysis.