Pattern Analysis and Applications (2019) 22:1629–1642
https://doi.org/10.1007/s10044-019-00793-4

INDUSTRIAL AND COMMERCIAL APPLICATION

Gait-based person re-identification under covariate factors

Emna Fendri[1] · Imen Chtourou[2] · Mohamed Hammami[1]

Received: 7 November 2017 / Accepted: 1 February 2019 / Published online: 11 February 2019

Abstract

Gait is recognized as an effective behavioral biometric trait. Gait pattern information can be captured and perceived from a distance thanks to its noninvasive and less intrusive nature. Therefore, gait could be well suited for person re-identification. However, semantic information like clothing and carrying bags has a remarkable influence on its accuracy. Unlike the existing solutions, this paper proposed a new method for gait-based person re-identification relying on dynamic selection of human parts. This method consists in computing a new person descriptor from relevant selected human parts. The selection of the most informative parts was achieved depending on the presence of semantic information. Our experiments were performed on the CASIA-B database revealing promising results and showing the effectiveness of the proposed method.

Keywords Gait · Dynamic selection · Re-identification · Semantic information

1 Introduction

In recent years, surveillance cameras have been extensively deployed almost everywhere. The analysis of data captured from these cameras serves a significant role in several applications. For example, in order to detect abnormal activities in surveillance videos or crowd behavior analysis, various kinds of activity modeling are proposed in the literature. In [36], they propose a generalized version of Laplacian regularized sparse coding for human activity recognition called *p*-Laplacian regularized sparse coding (pLSC). In addition, in this context, Liu et al. [35] present a new manifold regularized semi-supervised learning method called multiview Hessian regularized logistic regression (mHLR) to recognize actions effectively. In this paper, we will be interested in person re-identification which is defined as the process of establishing correspondence between frames of a person taken from different cameras. There are two classes of approaches for person re-identification: appearance-based and biometric-based approaches. The first category uses typical features to quantify the appearance. These features are of low level such as color and texture derived from clothing. The main disadvantage of the appearance features is that in the real world, people may wear similar clothes as dark clothes in winter or white apron like in the hospital. Thus, appearance-based features may not be informative enough about the identity in some situations. Several biometric features (e.g., face, gait, etc.) have therefore been developed and have shown promising results. However, the significant disadvantage met is that face features usually require proximal sensing, whereas gait is a biometric feature which is perceivable from a distance. Gait is also captured without a walker's attention, so walker rarely hides or disguises their gait consciously. In addition, it is less likely to be covered than other biometric features. Therefore, from a surveillance perspective, gait seems to be an appealing modality. However, several covariate factors may alter the performance of gait, including clothing, carrying bags, view angles, walking speed, and so on. Among these, clothing and carrying bags influence is generally unavoidable and has a significant effect on accuracy. Figure 1 shows the images of the same walking subject in different situations: carrying nothing (a) and carrying a bag (b). From these images, we can see that appearance is considerably different despite the fact that the

✉ Imen Chtourou
imene.chtourou@gmail.com

Emna Fendri
fendri.msf@gnet.tn

Mohamed Hammami
mohamed.hammami@fss.rnu.tn

1 MIRACL Laboratory, FSS, University of Sfax, Road Sokra km 4, BP 802, 3038 Sfax, Tunisia

2 MIRACL Laboratory, ENIS, University of Sfax, Road Sokra km 4, BP 1173, 3038 Sfax, Tunisia

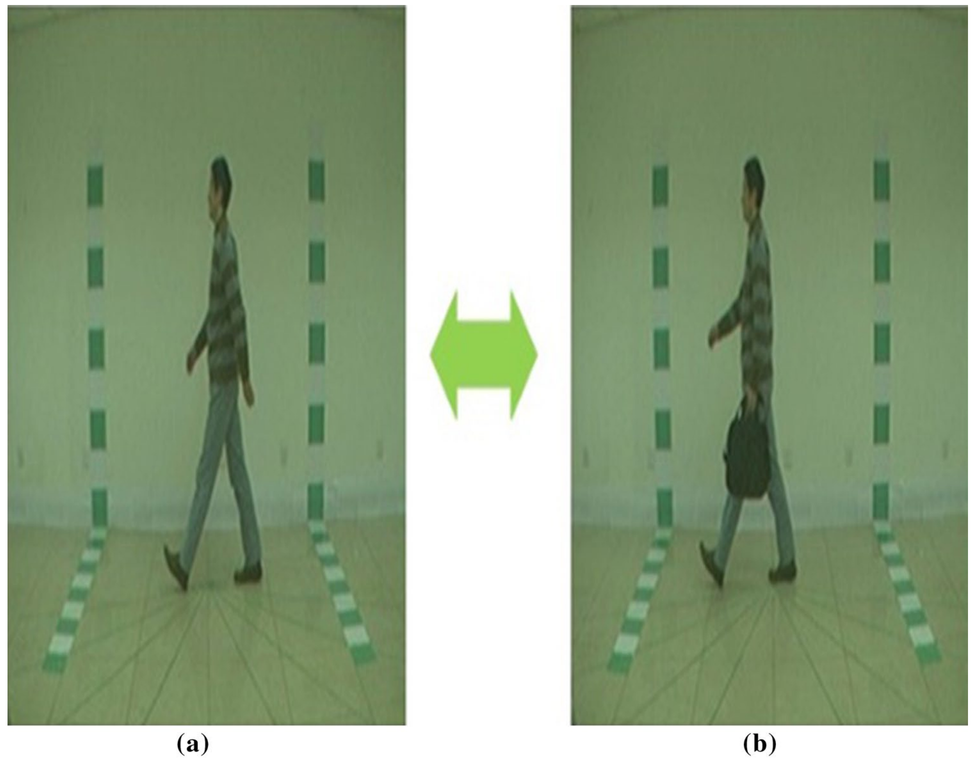


Fig. 1 Images of the same walking subject in different situations. The appearance of the subject carrying nothing (**a**), carrying a bag (**b**)

captured images are of the same subject. This difference causes the degradation of gait accuracy.

In this paper, a new method for a gait-based person re-identification was proposed addressing the challenge of carrying bags and clothing. We have used semantic attributes to model these challenges. This method is based on a dynamic selection of the most relevant parts for re-identification. Not only does this selection make the proposed method robust to the problems caused by the semantic attributes, it also contributes to speed up the process using some relevant parts instead of the whole-body image. This rapidity is crucial for a video surveillance application. The remainder of this paper was organized as follows. Section 2 provides a brief review of the related work, and Sect. 3 introduces the proposed gait-based person re-identification method. Our experimental results and analysis are presented in Sect. 4. Finally, Sect. 5 recapitulates our work and states our conclusions.

2 Related work

The human gait has always been an active research topic in recognition and re-identification [5, 53] for computer vision researchers. A large number of methods based on gait have been proposed. These methods are divided into two major categories: model-based and model-free methods. In the model-based methods [6, 12, 17, 23, 34, 42, 44, 47, 51, 54, 57], the human body is modeled by different geometrical shapes and the representation of gait is acquired using the model parameters. These methods suffer from their intensive computations and require high-quality images for extracting the model parameters from the human body parts. Therefore, they faced the difficulty of running in a real-time system. In the model-free methods, silhouette images are directly used to extract spatiotemporal motion information and different statistical features. These methods are further split up into two sub classes: frame-based and period-based methods. In the frame-based methods, the images are matched frame by frame [22, 48]. For a successful evaluation, a synchronization step has to precede the matching step. Sequences have to be aligned (or be in phase) at a preprocessing stage. However, in period-based methods, individual frames over a given period commonly known as a computed gait cycle (two stances) are integrated. One of the earlier feature representations in the model-free methods is gait energy image (GEI) introduced by Han and Bhanu [39]. GEI is computed by averaging silhouettes values, pixel by pixel, over an entire gait cycle period. Other methods have also been derived as variants of GEI like the gait entropy image (GenI) [21, 41], Chrono gait image (CGI) [37, 52], gait flow image (GFI) [3, 27], and frame difference energy image (FDEI) [5, 10]. Besides the multiple variants of GEI, an experimental study by Iwama et al. [19] shows that GEI, thanks to its simplicity, is the most stable and efficient type of features on their proposed dataset named OU-ISIR gait database.

GEI is treated differently: there are several works that use pixels of GEI as descriptors others use descriptors on the GEI like local binary pattern (LBP), fuzzy local binary pattern (FLBP), etc.

Works that use pixels of GEI as descriptors are: the work of sivapalan et al. [50] which use GEI as a 1D vector after the concatenation of these columns and use the entire GEI image. In addition, the method proposed by [40] explored the whole image of GEI using RankSVM for gait recognition. The drawback of these methods is the size of the feature vector which requires an important memory space. Therefore, there exist several works which used descriptors on GEI image: the work presented in [24] focused on the relevance of LBP (local binary pattern) in extracting texture characteristics in the whole GEI image and the region delimited by the legs. The proposed representation technique is capable of capturing variations in gait due to change in cloth, carrying a bag, and different instances of normal walking conditions more effectively, but a lot of useful information which may be used to enhance accuracy is discarded. Lee et al. [29] adopt the transient binary pattern operator (TBP: transit binary patterns) after dividing the GEI image into equal regions. This method presents a combination of spatiotemporal approach and texture descriptors to extract the temporal patterns in gait cycles. However, parameters like number of bits considered for the binary patterns and the number of pixels in each cell region for the TBP operator need to be calculated. The method in [7] applied the fuzzy local binary pattern (FLBP) on each region of the image after dividing the GEI image into five and seven non-overlapping regions. The advantage of this method is that it investigates the effect of partitioning the image over the use of the whole image. An area for improvement is still to be made in order to improve results. The method in [32] is based on golden ratio. A two-dimensional Gabor filter (2D Gabor) is adopted to extract features from GEI. It uses four different clothing models to identify unaltered area of the test GEI that is used for recognition. The method have find out the part of clothing and discard it, but this experiment is only available on clothing condition. The work achieved by [33] is based on the Haralick features extracted from gait energy image (GEI). These features are extracted locally by dividing vertically or horizontally the GEI into two or three equal regions of interest, respectively. The drawback of these works is that they discard a lot of relevant information. The method presented by Iwashita et al. [20] uses affine moment invariants on the five horizontal parts of the GEI image. This method can deal with any unknown covariate condition changes, but it assumes that gallery set is normal (i.e., images carrying nothing) which made it suitable for cooperative setting. Choudhury et al. [11] use entropy to compute the descriptor from the limb region of the leg of the gait energy image (GEI). The method in [15] also used only the lower part of the human body for recognition. Experimental results show the performance of these methods, but useful information would be excluded. The method presented in [45] used the top and bottom parts of gait as selected features to reduce the influence of covariate factors. In addition, Li et al. [31] use the head and feet of the gait to construct a structural gait energy image (SGEI) for recognition. The method can cope with the clothing and carrying variations pretty well, but relevant information may be discarded. This information can be used to enhance accuracy. Liu et al. [38] have explored Gabor feature on gait representation in order to enhance the appearance re-identification accuracy. Alotaibi et al. [1] propose a feature selection method based on the gait energy image. They describe an augmentation technique to overcome some of the problems associated with the intra-class gait variations, as well as if the amount of the training data is relatively small. They used dictionary learning with sparse coding and LDA to seek the best discriminative data representation before feeding it to the NC classifier. In [25], gait motions were encoded based on a set of spatiotemporal interest points from a raw gait video. These interest points were detected by using Harris corner detector from the regions with significant movements of human body in local video volumes.

Most of the previous works select a few parts which are generally assumed to be affected by clothing and carrying bags. However, some of the chosen parts are not really all affected by these covariate factors. Besides, these methods fail to notice some uninfluenced parts.

The main contribution of this work was to select, depending on the existence of semantic attributes (clothing and carrying bags), the relevant parts. In order to locate the covariate factors, we have used semantic information. We proved that a dynamic selection of parts can enhance the re-identification accuracy.

3 Proposed method

Our proposed method for a gait-based person re-identification has the ability to re-identify a person independent of his status (wearing coat, carrying bags, etc.). In this paper, we have focused on controlled environments where individuals are seen from a side view. Figure 2 shows the flowchart of the proposed method. Given a sequence of gait images of a person P, our dynamic gait-based person re-identification method consists of two main steps: (1) signature extraction and (2) signature matching. The first step extracts discriminative features from a gait image sequence. The second step matches the probe person P with a set of gallery people $G = \left\{G_1, G_2, \ldots, G_N\right\}$.

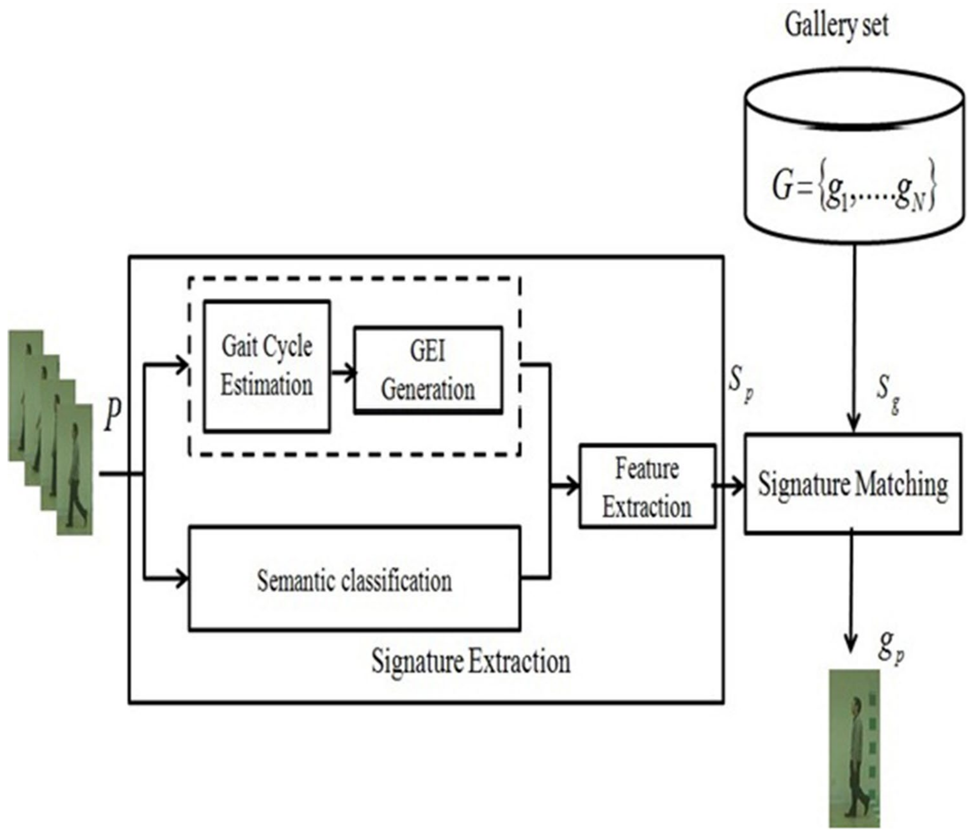


Fig. 2 Flowchart of the proposed method

3.1 Signature extraction

Given gait image sequences of a person, this step aims to extract the discriminative and informative features from the different body parts for both the probe person S_p and the gallery set $S_G = \left\{ S_{g_1}, S_{g_2}, \ldots, S_{g_N} \right\}$. First, we start by the estimation of the gait cycle and then a gait energy image (GEI) is generated. Based on the semantic classification, we extract features from each relevant part that is unaffected by semantic attributes.

3.1.1 Gait cycle estimation

Although the walking style differs from one person to another, the process of walking is the same for all human beings. A gait cycle is a periodic repetition of swing and stance phase, respectively [18]. It begins with the heel strike of one leg and finishes when the same foot touches the ground. It is defined by the set of silhouettes between two consecutive mid-stances of the same type. In this work, we have based on the variation of bounding box's width presented by [29]. Indeed, the curve representing the bounding box's width reaches a local minimum in "Midstance" posture and reaches a local maximum in "Double Support" posture. A half-period is then defined by the postures between two successive local minimums or two successive local maxima, a period being associated with two successive half-periods (cf. Fig. 3).

3.1.2 GEI generation

After the gait period is estimated, the GEI can be computed. Given N binary gait silhouette image frames $S_t(x, y)$, the gray-level image is defined as:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^{N} S_t(x, y) \tag{1}$$

where t is the frame number in the sequence (point in time) and (x, y) are the pixels coordinates. Figure 4a shows a sequence of gait silhouette along one gait cycle and (b) shows one example of gait energy image (GEI). GEI has two main regions: static and dynamic areas. The high intensity pixels correspond to the static parts of the body (top part of the body). This part contains the body shape information. Low-intensity pixels correspond to the dynamic parts of the body (lower part of the body). After the GEI generation, the image is divided into different-sized non-overlapping parts. We have based on an anatomical study [13, 30] to divide the GEI image into seven parts as shown in Fig. 5: head, chest

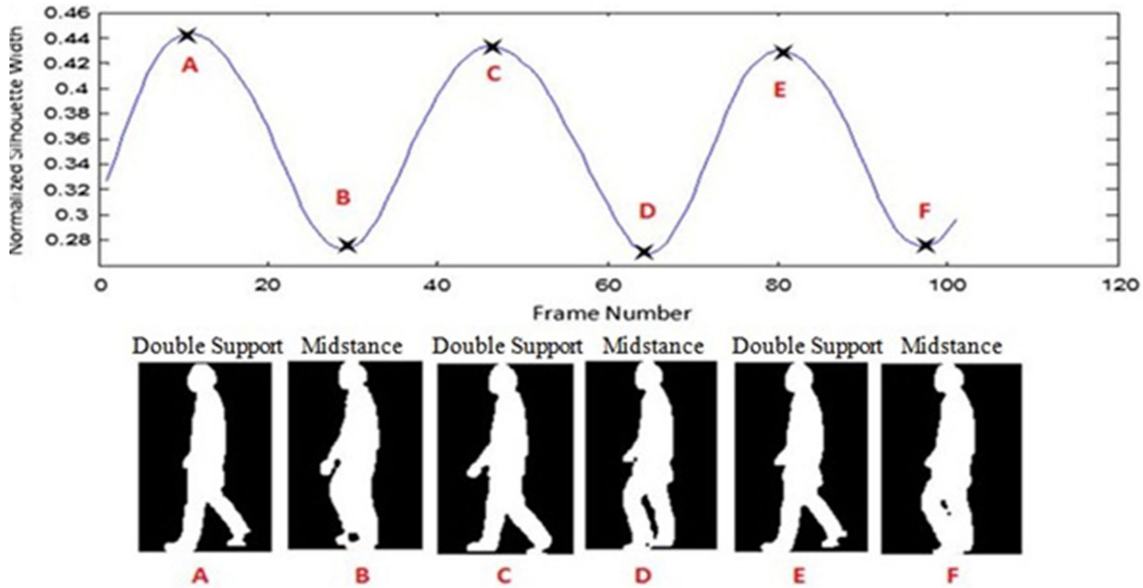


Fig. 3 Top row shows the time series signal of silhouette width. Bottom row shows the corresponding gait image at local maximum (A, C, and E) and local minimum (B, D, and F) with corresponding posture

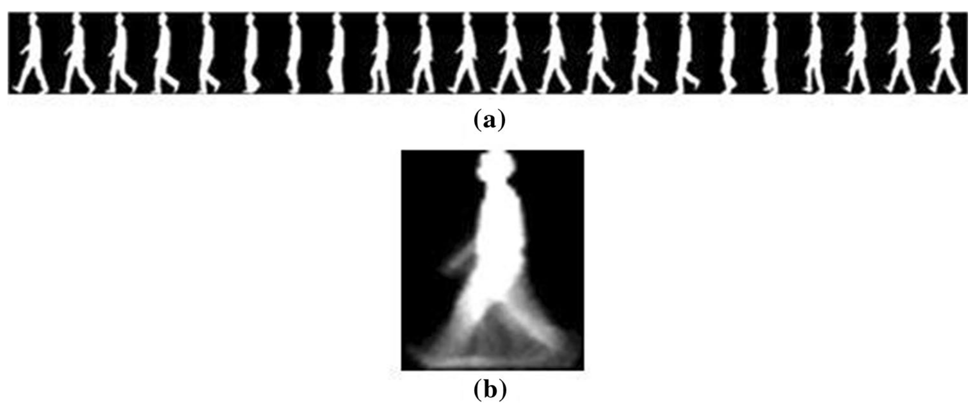


Fig. 4 **a** Sequence of gait silhouette, **b** gait energy image (GEI)

1, chest 2, knee 1, knee 2, foot 1, and foot 2. These parts are helpful for semantic attribute classification.

3.1.3 Semantic classification

This is the first attempt to adopt semantic attributes in a gait-based person re-identification. Based on the assumption that not all features contribute equally to the re-identification process, we have used semantic attributes in order to discard parts that may alter accuracy and negatively impact performance. The semantic classification process is achieved by a first step which is the detection of semantic attributes; then, relying on this detection, a dynamic selection of the relevant parts is performed. The semantic attributes are a form that concerns the use of simple terms that people use to describe each other linguistically. Figure 6 presents images of the same person, the first image is carrying a Single Shoulder bag (a) and the second is carrying nothing (b). The corresponding gait energy image is shown in both cases single shoulder bag (a) and carrying nothing (b). We can see from the differences in the back-body part that it is ineffective to compare the similarity between these two corresponding body parts. The GEI image is widely affected by the person's appearance. Therefore, using semantic attributes information such as backpack and single shoulder bag to improve the gait-based person re-identification might be a promising solution.

In our proposed method, we are interested in semantic attributes that can affect the representation of GEI image. We have used four different semantic attributes: carried objects (backpack, single shoulder bag, and handbag) and clothes (coat). This is a pioneer work which takes into account the presence of semantic attributes in gait-based person re-identification task. Inspired by [2, 16, 28, 43], the general

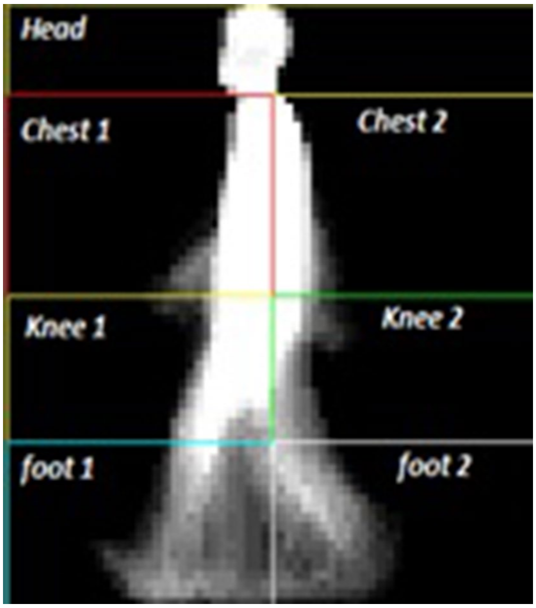


Fig. 5 GEI divided into seven parts (carrying nothing subject)

process of the detection of semantic attributes requires two stages: an off-line stage and online one as shown in Fig. 7. The off-line stage involves three major substeps. First, we start with a data preparation of the training database related to semantic attribute. Second, a predictive model for each class of attribute is constructed. Third, a validation step is required. In the data preparation step, the goal is to construct a two-dimensional table from the training database. Each table row represents a bounding box's image, and each column represents a feature. In the last column, the semantic attribute class is saved. Low-level color and texture features have shown their robustness in describing pedestrian images [28]. Therefore, we have used this collection of color features (i.e., color histograms in RGB, HSV, and YCbCr color spaces) and texture features (i.e., Gabor and Schmid filters) to model each semantic attribute. Table 1 shows the repartition of train and test sets for the five semantic attributes from the CASIA-B database.

Once the data preparation step is defined, our task is to perform machine learning using different classifiers in order to prepare predictive model for each class of semantic attribute. There are several algorithms of supervised learning in the literature. Each has its advantages and disadvantages. We used three supervised algorithms from different families, like the support vector machines (SVM) [8], tree bagger-based decision tree (FT) [59], and the neural networks (NN) [56]. The canonical correlation analysis network is useful for multiview image classification. Yang et al. [55] propose a canonical correlation analysis networks which allow the extraction of two different view features from an image and construct a final representation of this image. However, in this paper, we will not address the situation in which sample images are represented by two view features. Further, trained classes are tested for the classification accuracy and their corresponding results are presented in Sect. 4. We have opted for correct classification accuracy (CCA) which denotes the ratio of correctly classified images with the total number of images for the validation step. For the online stage, given a new pedestrian's bounding box, we start by extracting features to represent the global information of each image. Labels are generated depending on the semantic information of the image. These feature vectors design the input of the pre-learned selected predictive model.

3.1.4 Feature extraction

After the detection of the semantic information, a selection substep of parts is needed. This step consists of a dynamic selection of parts from the divided GEI image. Based on the fact that we have treated four different types of semantic attributes which are: single shoulder bag, handbag, backpack, and clothing (coat), we know that to remove the influence of the semantic information, parts which are affected should be discarded and only the uninfluenced parts can be used for our task. For the case of outerwear, relevant parts

Fig. 6 A GEI image showing a person carrying single shoulder bag (**a**), carrying nothing (**b**)

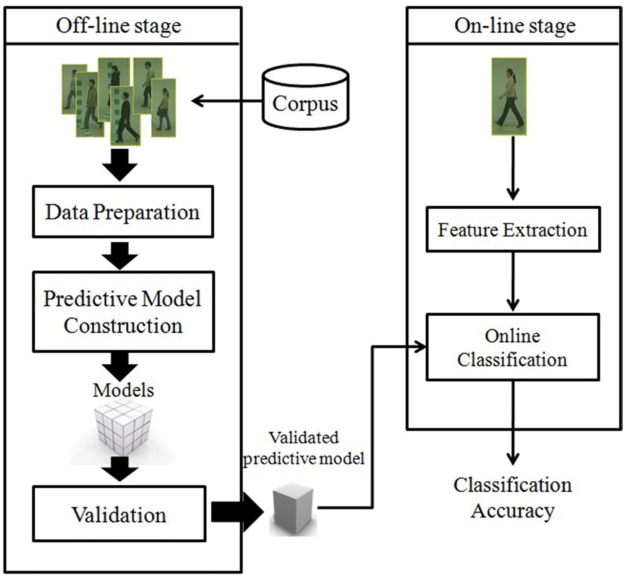


Fig. 7 General process of the detection of semantic attributes

Table 1 Train and test sets repartition of the five semantic attributes from CASIA-B database

Attribute	#Train	#Test
Carrying nothing	1488	496
Outerwear	744	248
Single shoulder bag	522	174
Handbag	132	44
Back pack	90	30

are head, foot 1, and foot 2. Parts that are in gray will be discarded (cf. Fig. 8). For the case of handbag, relevant parts are head, torse 2, torse 3, foot 1, and foot 2 (cf. Fig. 9). For the case of backpack, relevant parts are head, torse 2, high 1, foot 1, and foot 2 (cf. Fig. 10). For the case of single shoulder bag, relevant parts are head, torse 2, high 1, foot 1, and foot 2 (cf. Fig. 11).

Consequently, in order to determine the set of relevant parts V, we have divided the GEI image into seven parts $\{G_{H1}(x,y),\dots,G_{H7}(x,y)\}$. c is the coefficient related to each part. The set of relevant parts V is determined as follows:

$$V = \{c_1 \cdot G_{H1}(x,y),\dots,c_i \cdot G_{Hi}(x,y),\dots,c_7 \cdot G_{H7}(x,y)\}$$
$$\begin{cases} c_i = 0 & \text{if semantic attribute exist in part } c_i \\ \text{otherwise} & c_i = 1 \end{cases} \tag{2}$$

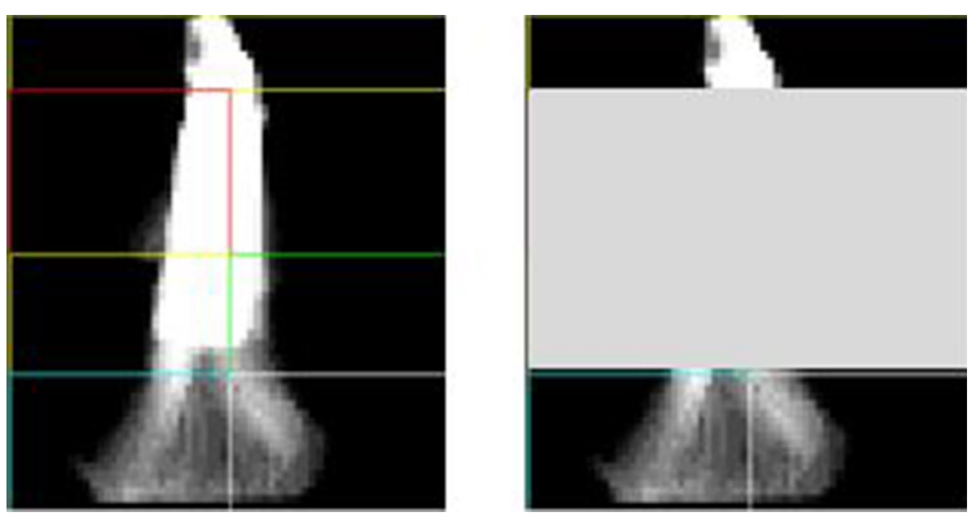
where i is the index of part $G_{H_i}(x,y), i\epsilon[1,\dots,7]$.

The most important task is the extraction of the salient and suitable feature to successfully capture the gait characteristics. After an extensive study, we have adopted P-LBP (partial LBP) for this task. It is a modified local binary pattern (LBP) introduced by [26]. The P-LBP can be applied on GEI to extract many meaningful features. Like the original LBP, it takes the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number (0 or 1). This is defined as follows:

$$\text{P-LBP}(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \tag{3}$$

where i_c corresponds to the value of the center pixel (x_c, y_c), i_n to the value of the eight surrounding pixels, and function $s(x)$ is defined as:

Fig. 8 Relevant parts used for outerwear: gray parts will be discarded



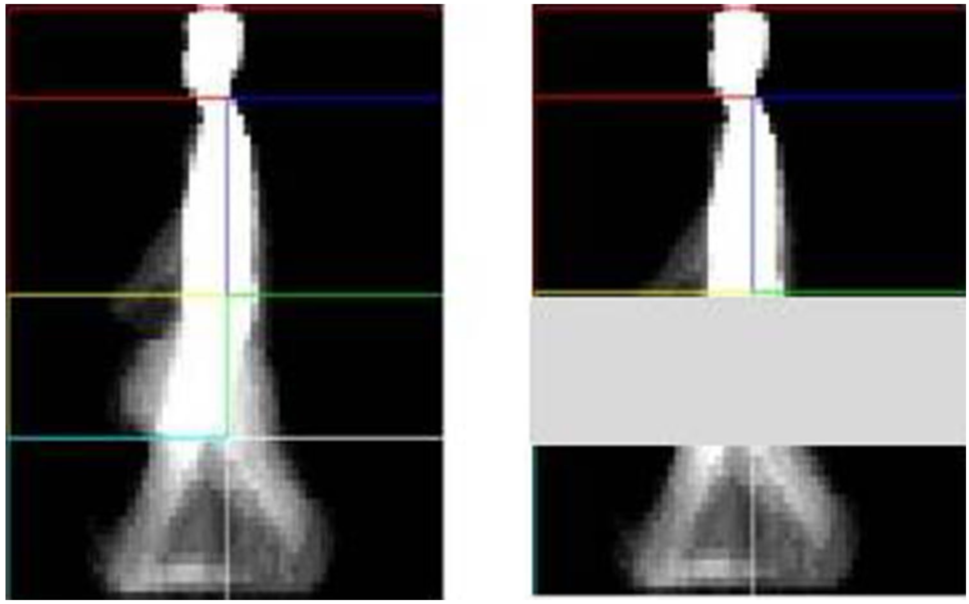


Fig. 9 Relevant parts used for handbag: gray parts will be discarded

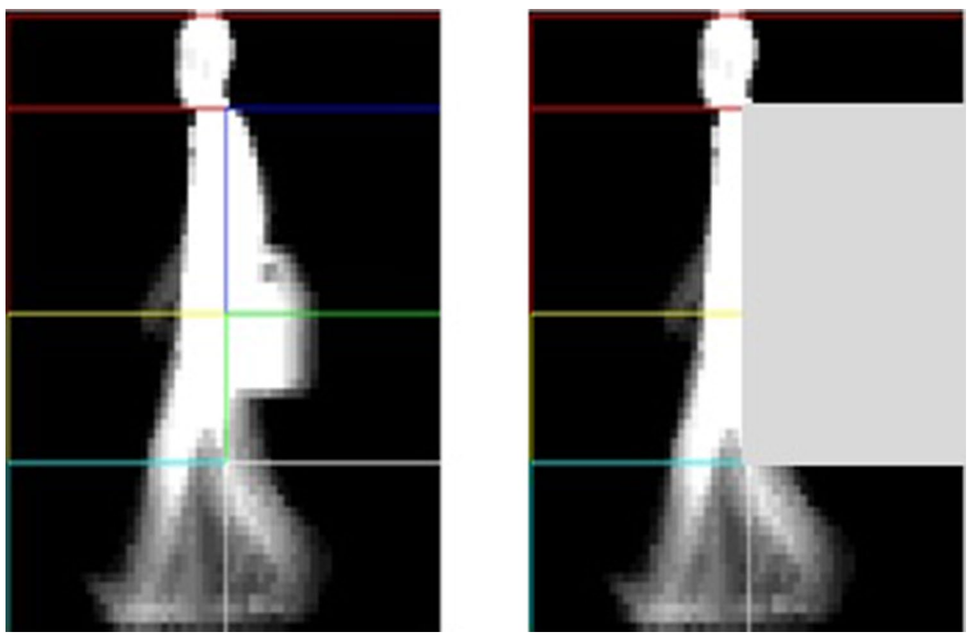


Fig. 10 Relevant parts used for single shoulder bag: gray parts will be discarded

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{4}$$

The original LBP uses a histogram to collect local statistics of a binary pattern in each of the image parts where LBP is carried on. Nevertheless, the histogram failed to capture the detailed local texture information. Therefore, P-LBP immediately recorded the bit sequences above to maintain more local information. The binary sequence would be further used as gait signature for each relevant part. Below is the pseudo-algorithm of the signature extraction step of the relevant parts.

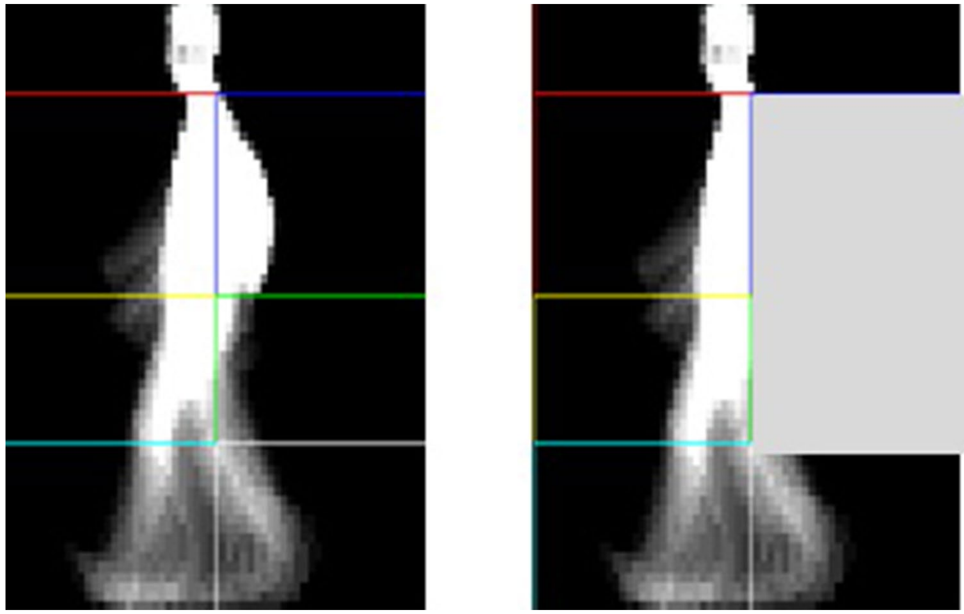


Fig. 11 Relevant parts used for back pack: gray parts will be discarded

---

Algorithm 1: The pseudo algorithm of the signature extraction step of the relevant parts

Input : N Silhouette images extracted over one gait cycle: $S_t(x, y)$ $t = 1, 2, ...N$
Output: Relevant Features set F

Generate GEI using Equation.1: G(x,y)
Divide GEI into 7 non-overlapping parts : $G_{H1}(x, y), G_{H2}(x, y), G_{H3}(x, y), G_{H4}(x, y), G_{H5}(x, y), G_{H6}(x, y)$ and $G_{H7}(x, y)$
Determine relevant parts according to semantic attributes classification
For each relevant part, compute the P-LBP features $F_{(Hi)}$ $i < 7$
Generate relevant feature set : $F_{(H)}$

3.2 Signature matching

A matching step is necessary in the end of our proposed method. In order to calculate the matching scores, this step aims to measure the distance between the probe person signature S_{p} and each gallery image signature $S_{gi}, i \in 1, \ldots, N$. Common parts none affected by semantic attribute for the gallery and probe image are used. After that, the gallery set is arranged according to similarity in order to generate a ranked list. Lastly, the identity of the most similar person g_{p} (i.e., top rank person) is assigned to the probe image, as reformulated in Eqs. (5) and (6). It is assumed that the identity of the probe image belongs to the gallery set. In our work, the Euclidean distance was used as a metric

$$\mathrm{id}(p) = \mathrm{id}(g_{\mathrm{p}}) \tag{5}$$

$$g_{\mathrm{p}} = \underset{i \in 1, \ldots, N}{\arg\min} [\mathrm{dist}(S_{gi}, S_{\mathrm{p}})] \tag{6}$$

where

$$\mathrm{dist} = \frac{\sum_{i=1}^{\mathrm{nbpart}} \mathrm{dist}_i}{\mathrm{nbpart}} \tag{7}$$

$\mathrm{nbpart} = \text{number of relevant parts}$.

4 Experimental evaluation

In this section, we detailed the different experimental setups in order to validate the performance of our method. In our experiments, we used the public CASIA-B database that covers several challenges, which makes it suitable for a re-identification scenario. Different experimentations have been presented to evaluate our proposed method. The first experiment is realized in order to choose the most suitable descriptor for our proposed method. The experiment was carried on GEI images carrying nothing (i.e., normal). The second experiment revealed the results from semantic attributes classification. Third experiment shows the contribution of our proposed method based on dynamic selection of parts. This experiment was carried out on a mixture of normal, carrying bag, and wearing coat GEI images. Finally, in the fourth experiment we compared our method with some state-of-the-art methods. Before presenting these different experiments, we briefly described the used database, evaluation metric, and experimental protocol.

4.1 Gait database

To evaluate the performance of a typical re-identification system, it is necessary to have two sets of person images. The first set, called gallery, gathers subject signatures known by the system. The second set, called probe, consists of subjects to be re-identified by the system. The probe and the gallery sets can be generated by selecting samples containing different challenges. The re-identification system compares the extracted signature for a probe person and matches it with one from the gallery set. As mentioned above, in our work we used the CASIA-B gait database [58]. It contains

124 subjects (individuals), each of which consists of 10 series with three situations; these three situations are "carrying bags," which means subjects appear with a bag, "wearing coat," which means subjects appear with a coat, and "carrying nothing" means normal subjects. Each image was scaled to 64×64 pixels. The gait silhouettes used were in 90 degree (side view) viewing angle as this view provides more gait information than the silhouettes taken from other view angles. Figure 12a–e shows an example of person's images from the CASIA-B database with various situations.

4.2 Evaluation metric and experimental protocol

To evaluate the proposed method, the accuracy at major top ranks and the cumulative matching characteristic (CMC) curves were reported. The CMC curve represents the probability of finding the correct match in the top r matches. In other words, a rank-r recognition rate shows the percentage of the probes that were correctly recognized from the top r matches in the gallery. In our experiments, we followed the common experimental protocols in person re-identification presented in [16]. Fifty-one image pairs were randomly selected as probe and gallery sets. The experiments were performed 10 times.

4.3 Results

4.3.1 First serie of experiments: descriptor choice

This section aimed to attest the proposed gait feature based on different descriptors such as partial local binary pattern (P-LBP), histogram of oriented gradients (HOG), and local binary pattern (LBP). In this experiment, we used 204 GEI images from CASIA-B database. Gallery and probe set contain subjects carrying nothing (i.e., normal). Table 2 shows comparative results between LBP, HOG, and P-LBP. It is evident that P-LBP outperforms the other used descriptor like HOG and LBP over the entire range of ranks. P-LBP outperforms HOG and LBP showing an advantage in terms of rank-1 of $\approx$ 23.92% and $\approx$ 47.05%, respectively. This confirms that the partial local binary pattern (P-LBP) is the most convenient descriptor for gray-level image (GEI: gait energy image) for gait-based person re-identification task.

Table 2 Experimental results on the CASIA-B database (%)

	Rank 1	Rank 3	Rank 5
LBP	39.6078	59.2157	68.2353
HOG	62.7451	83.3333	89.6078
P-LBP	86.666	94.509	96.078

4.3.2 Second serie of experiments: semantic attributes classification

Our aim in this section was to reveal the results obtained from semantic attributes classification. We used all the 124 person images from CASIA-B database, where 10 sequences per subject were used. Overall, we gathered 1240 person images. Our database contains about 744 carrying nothing (normal), 248 carrying bags (bag), and 248 wearing coats (coat) person's images. The annotation process was carried out according to the situation of person's images and the image number for each category of bags (i.e., single shoulder bag, handbag, and backpack). We have used five classes which are: single shoulder bag, handbag, backpack, coat,

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

Fig. 12 Example of person's images in different situations (**a–e**): carrying nothing, wearing coat, carrying single shoulder bag, carrying backpack, carrying handbag [58] and normal. Consequently, the five classes correspond to five semantic attributes. Each one can be annotated by two possibilities: 1 indicating the existence of the attribute and 0 showing the absence of the attribute. We divided the data in such a way that each attribute had an equal number of positive and negative samples. Classification of these five semantic attribute was performed using three popular supervised classifiers, namely support vector machines (SVM) [8], neural network (NN) [56], and tree bagger derived from random forest (FT) [59]. For SVM classifier, we used LIBSVM [49]. In the neural network (NN) structure, the neurons are associated with layers. The first and last layers are called input and output layers, respectively, as they represent inputs and outputs of the overall network. The remaining layers are known as the hidden layers. We have adopted 10 hidden neurons and backpropagation algorithm [46] for training. For tree bagger, MATLAB's TreeBagger implementation of bagged decision trees was used with 100 trees. Table 3 shows that the neural network (NN) gives better accuracy than SVM [49] and tree bagger (FT) [59] for 4 semantic attributes (i.e., wearing coat, single shoulder bag, backpack, and handbag) higher than 89%. For the semantic attribute carrying nothing (i.e., normal), the support vector machine (SVM) shows a better performance than the neural network (NN). This confirms that neural network (NN) is more precise and efficient for detecting semantic attributes that alter the human shape.

4.3.3 Third serie of experiments: dynamic selection of human parts

In real life, people may have similar or different appearances in the probe and gallery sets. The situation of each sample in the probe and the gallery sets was randomly chosen. Probe and gallery sets may contain a mixture of subject's images carrying bags (bag), wearing coats (coat) and carrying nothing (normal). In this serie of experiments, we performed three evaluations. First, we present results of using all the seven parts like shown in Fig. 5. Second, we displayed the advantage of using dynamic selection of relevant parts. Third, inspired by the assumption that the dynamic body parts [4, 14], head shape and the neck [45] contain discriminative information, we have explored invariant parts which are composed of three parts (i.e., head, foot 1, and foot 2) as shown in Fig. 5. The underlying idea is based also on the fact that all our chosen invariant parts are the common parts unaffected by the four semantic attributes. These four attributes represent the three types of the situation carrying bags (bag) which are: single shoulder bag, backpack, and handbag and the situation of wearing coats (coat).

Our proposed method based on dynamic selection mitigates the clothing and carrying variation problems very well. The results from Fig. 13 and Table 4 clearly show that our method yields the best matching rate and a much superior performance with 52.3529 at rank 1, 69.8039 at rank 3, and 78.2353% at rank 5. The advantage of our method is particularly evident when appearance of people changes (gallery and probe sets are different). Overall, our method can handle different appearances, particularly when people change their clothes, carry different types of bags, or even carry nothing (Fig. 13).

Apart from person re-identification accuracy, the computational complexity is essential to perform real-time video surveillance applications. Our proposed method clearly shows an important reduction in the computational cost when the selection of parts is achieved prior to applying the P-LBP. Table 5 presents a comparison of the complexity cost where gallery and probe sets contain a mixture of subject's image carrying bags (bag), wearing coats (coat), and carrying nothing (normal). We denoted this set as (bag–coat–normal). The process of selecting the relevant parts and then applying the P-LBP denoted "Times with selection" requires approximately half the times of the process of applying the P-LBP and then selecting the parts that are altered by the semantic attributes which is denoted as "Times without selection." This gain

Table 4 Performance comparison of P-LBP with dynamic selection (%)

	Rank 1	Rank 3	Rank 5
P-LBP 7 parts	49.411	59.215	63.725
P-LBP dynamic selection	52.352	69.803	78.235
P-LBP invariant parts	47.843	68.235	74.509

Table 3 Classification accuracy for each attribute (%)

	Carrying bag			Wearing coat	Carrying nothing
	Single shoulder bag	Backpack	Handbag		
Neural network (NN)	89.4	94.4	92.3	89.9	86.2
Support vector machine (SVM)	79.885	73.333	84.09	80.645	89.314
Tree bagger (FT)	78.2	76.7	65.9	73.4	72.0

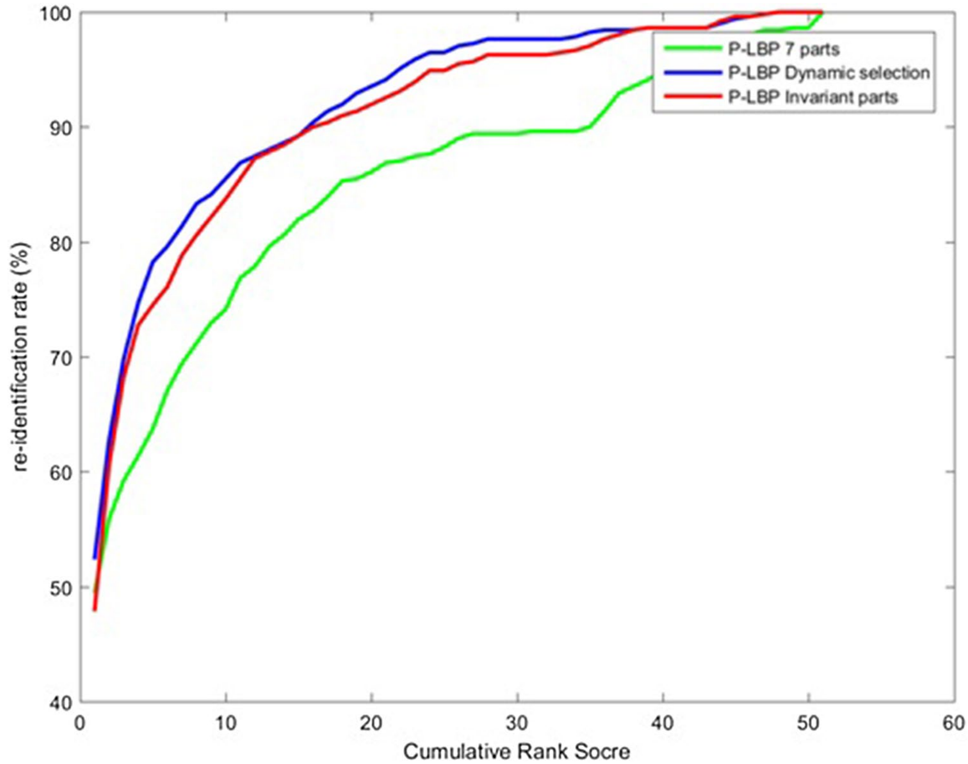


Fig. 13 Re-ID results: the CMC curves obtained

Table 5 Comparison of computational cost on CASIA-B database

	Times without selection	Times with selection
Bag–coat–normal (s)	811	465

Table 6 Simple setting: comparison of matching rate on CASIA-B database (%)

	NM	BG	CL	AVG
STIPs + BoW [25]	94.5	60.9	58.5	71.3
Reducing-based selection [1]	98.4	86.7	94.8	93.3
Proposed method	96.7	78.2	78.2	84.6

in execution time is essential for a real-time application context. Besides reducing the computation cost, removing the irrelevant parts has another potential benefit related to the storage requirements. As experimentally shown, our proposed method based on a dynamic selection provides a good agreement between re-identification accuracy and computational complexity.

4.3.4 Fourth serie of experiments: comparing gait-based person re-identification methods

In this section, we aimed to evaluate the performance of our proposed method using two different settings: simple setting and challenging setting.

4.3.4.1 Simple setting For the protocol of simple setting, the first four sequences of the normal condition, namely "nm-01," "nm-02," "nm-03," and "nm-04," are used as the gallery (training) set. The two remaining sequences of set NM ("nm-05" and "nm-06"), the two sequences of set BG ("bg-01" and "bg-02"), and the two sequences of set CL ("cl-01" and "cl-02") are used as a probe set for testing. We carried out our experiments under 90 view angle. We have compared our proposed method with recent works of gait recognition: STIPs + BoW [25] and reducing-based method [1]. Table 6 shows the obtained results. It is obvious that our proposed method with an average (AVG) of 84.6% outperforms STIPs + BoW− based method [25] with an average (AVG) of 71.3% for the normal (NM), bag (BG), and coat (CL). The reducing-based method [1] outperforms our proposed method for the normal (NM), bag (BG), and coat (CL) with an average (AVG) of 93.3 %. However, this method depends on several parameters for feature selection. In addition, it is time-consuming as authors in this paper use the augmentation of gait representations. This made it unpractical for real-time application.

Table 7 Challenging setting: comparison of matching rate on CASIA-B database (%)

	Rank 1	Rank 5	Rank 10	Rank 15	Rank 20
Method 1 [38] (score-level fusion)	12.745	30.784	43.137	54.117	63.725
Method 1 [38] (feature-level fusion)	25.490	45.098	62.745	66.666	78.431
Method 2 [40]	19.607	58.627	69.607	76.274	82.745
Proposed method	52.352	78.235	85.490	89.215	93.529

4.3.4.2 Challenging setting In this section, we achieved using a challenging setting comparisons with some related state-of-the-art methods as follows:

Method 1 [38] has explored gait features for enriching the appearance re-identification representations. The idea is based on the assumption that appearance features are not discriminative in some cases, e.g., dark clothes in winter, color may be imprecise by different camera intrinsic parameters.

Method 2 [40] has considered GEI-RSVM which uses gait energy image (GEI) feature [39] and the ranking SVM [9] model.

Adopting the same experimental protocol, we re-implemented both of these works. For the method 1 [38], we have reported results with score-level fusion and feature-level fusion. Experiments were conducted where gallery and probe contain a mixture of subject's images carrying nothing (normal), carrying bags (bag), and wearing coats (coat). Table 7 shows that the proposed method performs much better than method 1 [38] and method 2 [40]. We can see that the matching rates at rank 1, 5, and 10 are 52,352, 78,235, and 85,490% for our proposed method, while those of method 1 [38] are only 12.745, 30.784, and 43.137% and those of method 2 [40] are 19.607, 58.627, and 69.607% for rank 1, rank 5, and rank 10. The effectiveness of our method can be explained by the fact that the idea of dynamic selection may discard parts that can affect accuracy. This fact enhances the matching rate even in a challenging setting.

5 Conclusion

Due to the increasing demand on visual surveillance systems, gait as a behavioral feature from a distance has gained more interests. Many allied studies have also demonstrated that gait can be used as a useful biometric feature for a person re-identification. This paper presented a novel gait-based person re-identification method relying on a dynamic selection of human parts. In this paper, we dealt with the covariate factors (wearing coat and carrying bags). To represent these factors, we used semantic attributes. We proposed a dynamic selection of relevant parts that may contain more information. Parts that are affected by semantic attributes were eliminated. The CASIA-B database was used to evaluate the proposed method. The experimental results show that the proposed method can achieve an impressive performance. It is also suitable for different appearances, particularly when people change their clothes, carry different types of bags, or carry nothing. One of the future perspectives is to explore our proposed method on different view angles.

References

1. Alotaibi M, Mahmood A (2017) Reducing covariate factors of gait recognition using feature selection and dictionary-based sparse coding. Signal Image Video Process 11(6):1131–1138
2. An L, Chen X, Kafai M, Yang S, Bhanu B (2013) Improving person re-identification by soft biometrics based reranking. In: 2013 7th international conference on distributed smart cameras (ICDSC). IEEE, pp 1–6
3. Arora P, Srivastava S et al (2016) Human gait recognition using gait flow image and extension neural network. In: Proceedings of the 2nd international conference on computer and communication technologies. Springer, Berlin, pp 1–10
4. Bashir K, Xiang T, Gong S (2010) Gait recognition without subject cooperation. Pattern Recognit Lett 31(13):2052–2060
5. Bedagkar-Gala A, Shah SK (2014) Gait-assisted person re-identification in wide area surveillance. In: Asian conference on computer vision. Springer, Berlin, pp 633–649
6. Benouis M, Senouci M, Tlemsani R, Mostefai L (2016) Gait recognition based on model-based methods and deep belief networks. Int J Biomet 8(3–4):237–253
7. Binsaadoon AG, El-Alfy ESM (2016) Gait-based recognition for human identification using fuzzy local binary patterns. In: ICAART (2), pp 314–321
8. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proceedings of the 5th annual workshop on computational learning theory. ACM, pp 144–152
9. Chapelle O, Keerthi SS (2010) Efficient algorithms for ranking with svms. Inf Retr 13(3):201–215
10. Chen C, Liang J, Zhao H, Hu H, Tian J (2009) Frame difference energy image for gait recognition with incomplete silhouettes. Pattern Recognit Lett 30(11):977–984
11. Choudhury SD, Tjahjadi T (2015) Robust view-invariant multiscale gait recognition. Pattern Recognit 48(3):798–811
12. Cunado D, Nixon MS, Carter JN (2003) Automatic extraction and description of human gait models for recognition purposes. Comput Vis Image Underst 90(1):1–41
13. Dempster WT, Gaughran GR (1967) Properties of body segments based on size and weight. Dev Dyn 120(1):33–54
14. Dupuis Y, Savatier X, Vasseur P (2013) Feature subset selection applied to model-free gait recognition. Image Vis Comput 31(8):580–591
15. Gabriel-Sanz S, Vera-Rodriguez R, Tome P, Fierrez J (2013) Assessment of gait recognition based on the lower part of the human body. In: 2013 international workshop on biometrics and forensics (IWBF). IEEE, pp 1–4

16. Gray D, Tao H (2008) Viewpoint invariant pedestrian recognition with an ensemble of localized features. Comput Vis ECCV 2008:262–275
17. Gu J, Ding X, Wang S, Wu Y (2010) Action and gait recognition from recovered 3-D human joints. IEEE Trans Syst Man Cybern Part B (Cybern) 40(4):1021–1033
18. Hosseini NK, Nordin MJ (2013) Human gait recognition: a silhouette based approach. J Autom Control Eng 1(2):259–267
19. Iwama H, Okumura M, Makihara Y, Yagi Y (2012) The ou-isir gait database comprising the large population dataset and performance evaluation of gait recognition. IEEE Trans Inf Forensics Secur 7(5):1511–1521
20. Iwashita Y, Uchino K, Kurazume R (2013) Gait-based person identification robust to changes in appearance. Sensors 13(6):7884–7901
21. Khalid B, Tao X, Shaogang G (2009) Gait recognition using gait entropy image. In: 3rd international conference on crime detection and prevention (ICDP 2009)
22. Khedher MI (2014) Ré-identification de personnes à partir des séquences vidéo. PhD thesis, Institut National des Télécommunications
23. Kovač J, Peer P (2014) Human skeleton model based dynamic features for walking speed invariant gait recognition. Math Probl Eng 2014:15
24. Kumar HM, Nagendraswamy H (2014) LBP for gait recognition: a symbolic approach based on GEI plus RBL of GEI. In: 2014 international conference on electronics and communication systems (ICECS). IEEE, pp 1–5
25. Kusakunniran W (2014) Attribute-based learning for gait recognition using spatio-temporal interest points. Image Vis Comput 32(12):1117–1126
26. Kusakunniran W, Wu Q, Li H, Zhang J (2009) Automatic gait recognition using weighted binary pattern on video. In: 6th IEEE international conference on advanced video and signal based surveillance, 2009. AVSS'09. IEEE, pp 49–54
27. Lam TH, Cheung KH, Liu JN (2011) Gait flow image: a silhouette-based gait representation for human identification. Pattern Recognit 44(4):973–987
28. Layne R, Hospedales TM, Gong S (2014) Attributes-based re-identification. In: Person re-identification. Springer, Berlin, pp 93–117
29. Lee CP, Tan AW, Tan SC (2015) Gait recognition with transient binary patterns. J Vis Commun Image Represent 33:69–77
30. Li N, Xu Y, Yang XK (2010) Part-based human gait identification under clothing and carrying condition variations. In: 2010 international conference on machine learning and cybernetics (ICMLC), vol 1. IEEE, pp 268–273
31. Li X, Chen Y (2013) Gait recognition based on structural gait energy image. J Comput Inf Syst 9(1):121–126
32. Liang Y, Li CT, Guan Y, Hu Y (2016) Gait recognition based on the golden ratio. EURASIP J Image Video Process 2016(1):22
33. Lishani AO, Boubchir L, Khalifa E, Bouridane A (2017) Human gait recognition based on Haralick features. Signal Image Video Process 11:1–8
34. Liu D, Ye M, Li X, Zhang F, Lin, L.: Memory-based gait recognition. In: BMVC (2016)
35. Liu W, Liu H, Tao D, Wang Y, Lu K (2015) Multiview hessian regularized logistic regression for action recognition. Signal Process 110:101–107
36. Liu W, Zha ZJ, Wang Y, Lu K, Tao D (2016) *p*-laplacian regularized sparse coding for human activity recognition. IEEE Trans Ind Electron 63(8):5120–5129
37. Liu Y, Zhang J, Wang C, Wang L (2012) Multiple hog templates for gait recognition. In: 2012 21st international conference on pattern recognition (ICPR). IEEE, pp 2930–2933
38. Liu Z, Zhang Z, Wu Q, Wang Y (2015) Enhancing person re-identification by integrating gait biometric. Neurocomputing 168:1144–1156
39. Man J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322
40. Martín-Félez R, Xiang T (2014) Uncooperative gait recognition by learning to rank. Pattern Recognit 47(12):3793–3806
41. Nandy A, Pathak A, Chakraborty P (2017) A study on gait entropy image analysis for clothing invariant human identification. Multimed Tools Appl 76(7):9133–9167
42. Nixon M et al (2009) Model-based gait recognition. Encyclopedia of biometrics. Springer, Heidelberg, pp 633–639
43. Prosser BJ, Zheng WS, Gong S, Xiang T, Mary Q (2010) Person re-identification by support vector ranking. In: BMVC, vol 2, p 6 (2010)
44. Rafi M, Khammari H, Wahidabanu R, Taj Y (2013) A model based approach for gait recognition system. Int J Soft Comput Eng (IJSCE) 3:2231–2307
45. Rida I, Almaadeed S, Bouridane A (2016) Gait recognition based on modified phase-only correlation. Signal Image Video Process 10(3):463–470
46. Rumelhart DE, Hinton GE, Williams RJ et al (1998) Learning representations by back-propagating errors. Cognit Model 5(3):1
47. Saadoon A, Nordin MJ (2015) An automatic human gait recognition system based on joint angle estimation on silhouette images. J Theor Appl Inf Technol 81(2):277
48. Sarkar S, Phillips PJ, Liu Z, Vega IR, Grother P, Bowyer KW (2005) The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans Pattern Anal Mach Intell 27(2):162–177
49. Schölkopf B, Smola AJ (2002) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge
50. Sivapalan S, Chen D, Denman S, Sridharan S, Fookes C (2013) Histogram of weighted local directions for gait recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 125–130
51. Tafazzoli F, Safabakhsh R (2010) Model-based human gait recognition using leg and arm movements. Eng Appl Artif Intell 23(8):1237–1246
52. Wang C, Zhang J, Pu J, Yuan X, Wang L (2010) Chrono-gait image: a novel temporal template for gait recognition. Comput Vis ECCV 2010:257–270
53. Wei L, Tian Y, Wang Y, Huang T (2015) Swiss-system based cascade ranking for gait-based person re-identification. In: AAAI, pp 1882–1888
54. Yamauchi K, Bhanu B, Saito H (2009) Recognition of walking humans in 3D: initial results. In: 2009 CVPR Workshops 2009. IEEE computer society conference on computer vision and pattern recognition workshops. IEEE, pp 45–52
55. Yang X, Liu W, Tao D, Cheng J (2017) Canonical correlation analysis networks for two-view image recognition. Inf Sci 385:338–352
56. Yegnanarayana B (2009) Artificial neural networks. PHI Learning Pvt. Ltd, New Delhi
57. Zeng W, Wang C, Li Y (2014) Model-based human gait recognition via deterministic learning. Cognit Comput 6(2):218–229
58. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th international conference on pattern recognition, ICPR 2006, vol 4, pp 441–444
59. Zighed DA, Rakotomalala R (2000) Graphes d'induction: apprentissage et data mining. Hermes Paris, Paris