A new approach for multi-view gait recognition on unconstrained paths ☆

D. López-Fernández *, F.J. Madrid-Cuevas [1], A. Carmona-Poyato [1], R. Muñoz-Salinas [1], R. Medina-Carnicer [1]

Department of Computing and Numerical Analysis, Maimónides Institute for Biomedical Research (IMIBIC), University of Córdoba, Córdoba, Spain

ARTICLE INFO

ABSTRACT

Direction changes cause difficulties for most of the gait recognition systems, due to appearance changes. We propose a new approach for multi-view gait recognition, which focuses on recognizing people walking on unconstrained (curved and straight) paths. To this effect, we present a new rotation invariant gait descriptor which is based on 3D angular analysis of the movement of the subject. Our method does not require the sequence to be split into gait cycles, and is able to provide a response before processing the whole sequence. A Support Vector Machine is used for classifying, and a sliding temporal window with majority vote policy is used to reinforce the classification results. The proposed approach has been experimentally validated on "AVA Multi-View Dataset" and "Kyushu University 4D Gait Database" and compared with related state-of-art work. Experimental results demonstrate the effectiveness of this approach in the problem of gait recognition on unconstrained paths.

1. Introduction

Research on human gait as a biometric feature for identification has received a lot of attention due to the apparent advantage that it can operate at a distance and can be applied discreetly without needing the active participation of the subject [1]. However, gait recognition performance is significantly affected by changes in various covariate conditions such as clothing [2], camera viewpoint [3,4], load carrying [5], and walking speed [6].

According to camera viewpoint, the previous work can be categorized into two approaches: view-dependent and view-independent approaches. View-dependent approaches assume that the viewpoint does not change while walking. In such methods, a change in the appearance, caused by a viewpoint change, will adversely affect to the recognition [7]. For example, when a subject walks along a curved trajectory, the observation angle between the walking direction of the subject and the camera optical axis is gradually changed during the gait cycle. Fig. 1 shows the influence of a curved path on the silhouette appearance. On the contrary, the view-independent approaches aim to recognize people under different viewing angles. However, some of them do not allow curved trajectories or direction changes during walking.

This paper presents a new approach to recognize people walking along curved trajectories on unconstrained paths. Some potential applications of this work are access control in special or restricted areas (e.g. military bases, governmental facilities) or smart video surveillance (e.g. bank offices). This work also can be used for staff identification on laboratories or medical isolation zones where subjects wear special clothes that do not allow them to show the face or use the fingerprint (e.g. protective clothing for viral diseases).

The rest of the paper is structured as follows. Section 2 presents the most relevant works related to ours, making a clear distinction between view-dependent and view-independent methods. Section 3 presents a new rotation invariant gait descriptor. Section 4 shows the details of our gait recognition method. An analysis of the performance is given in Section 5. Finally, we conclude this paper in Section 6.

2. Related work

2.1. View-dependent approaches

One of the earliest view-dependent approaches can be seen in [8], where it is used the width of the outer contour of the binarized silhouette from a side view, to build a descriptor which contains
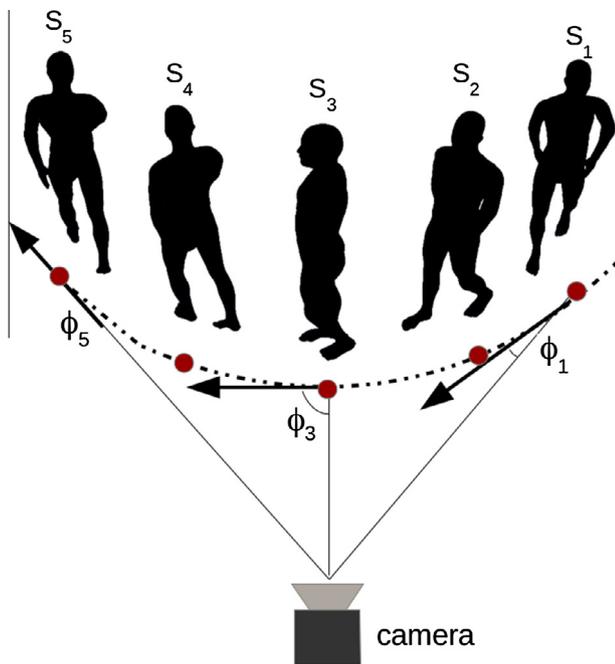

Fig. 1. In a curved path, the observation angle between the walking direction of the subject and optical axis of the camera is gradually changed, which affects the silhouette appearance.

both structural features and dynamic aspects of gait. Feature vectors derived from binary silhouettes have been also used to train Hidden Markov Models [9]. The contours of silhouettes have also been used [10,11].

In addition, in [12] it is presented a gait recognition method which analyses the shape of the silhouette using Procrustes Shape Analysis and Elliptic Fourier Descriptors. In [13] it is proposed a gait representation called Gait Energy Image (GEI), which is the average of all silhouette images for a single gait cycle.

Based on the idea of GEI, Depth Energy Image (DEI) was defined in [14], which is simply the average of the depth silhouettes taken along a gait cycle, over the front view. GEI is also extended in [15] to consider depth information from the side view, by means of a new feature called Depth Gradient Histogram Energy Image (DGHEI). In [16] a time-sliced averaged motion history image (TAMHI) alongside the histograms of oriented gradients (HOG) to generate gait signatures.

In [17] it is presented the Gait Energy Volume (GEV), a binary voxel-discretized volume which is spatially aligned and averaged over a gait cycle. The authors apply the GEV on partial reconstructions obtained with depth sensors from the front view of the individual. An extended work from GEV [17] that combines the frontal-view depth gait image and side-view 2D gait silhouette by means of a back-filling technique is presented in [18]. In [19], the depth and RGB frames from Kinect are register to obtain smooth silhouette shape along with depth information. A partial volume reconstruction of the frontal surface of each silhouette is done and the Pose Depth Volume (PDV) feature is derived from this volumetric model.

The performance of the above methods depends on the viewpoint. As was stated above, appearance changes due to viewing angle changes cause difficulties for most of the gait recognition methods, and this situation cannot be easily avoided in practical applications.

2.2. View-independent approaches

There are three major approach categories to sort out this problem [3], namely: (1) approaches that construct 3D gait information through multiple calibrated cameras; (2) approaches that extract gait features which are invariant to viewing angle changes; (3) approaches whose performance relies on learning mapping/projection relationship of gaits under various viewing angles.

Approaches of the first category are represented by [4,20–22]. In [21], a 3D approximation of a Visual Hull (VH) [23] is used to design a multi-modal and model-based gait recognition approach. Seely et al. [20] proposed an appearance-based approach which uses 3D volumetric data to synthesize silhouettes from a fixed viewpoint relative to the subject. The resulting silhouettes are then passed to a standard 2D gait analysis technique, such as the average silhouette.

Another approach that applies image-based rendering on a 3D VH model to reconstruct gait features under a required viewing angle is presented in [22]. This approach tries to classify the motion of a human in a view-independent way, but it has two drawbacks. On the one hand it considers only straight paths to estimate the position and orientation of a virtual camera. Tests were performed only on straight path motions. On the other hand, not all the 3D information available in the VH is used, because feature images are extracted from 2D images rendered only from a single view.

In [4], an observation angle at each frame of a gait sequence is estimated from the walking direction, by fitting a 2D polynomial curve to the foot points. Virtual images are synthesized from a 3D model, so that the observation angle of a synthesized image is the same that the observation angle for the real image of the subject, which is identified by using affine moment invariants extracted from images as gait features. The advantage of this method is that the setup assumes multiple cameras for training, but only one camera for testing. However, this approach requires to split the sequence into gait cycles and assumes that the gait phase of the first frame of a gait cycle of a subject is the same for each person in the database. Besides, shadows on the floor complicate the estimation of the foot points in silhouette images.

In the above four works, despite 3D models are used, the gait recognition scheme is based on silhouette analysis, what restricts a large amount of discriminant information because the recognition relies on single view silhouette analysis, instead of analyze the 3D information.

Approaches of the second category extract gait features which are invariant to viewing angle change. In [24], it is described a method to generate a canonical view of gait from any arbitrary view. This method can work with a single calibrated camera but the synthesis of a canonical view is only feasible from a limited number of initial views. The performance is significantly dropped when the angle between image plane and sagittal plane is large.

In [25], a method based on homography to compute view-normalized trajectories of body parts obtained from monocular video sequences was proposed. But this method efficiently works only for a limited range of views. Planar homography has also been used to reduce the dependency between the motion direction and the camera optical axis [26], however this method seems not to be applicable when the person is walking nearly parallel to the optical axis. In [27] view-invariant features are extracted from GEI. Only parts of gait sequences that overlap between views are selected for gait matching, but this approach cannot cope with large view angle changes under which gait sequences of different views can have little overlap. Neither it can be applied to recognize people walking on curved trajectories.

A self-calibrating view-independent gait recognition based on model-based gait features is proposed in [28]. The poses of the lower limbs are estimated based on markerless motion estimation. Then, they are reconstructed in the sagittal plane using viewpoint rectification. This method has two main drawbacks that are worth mentioning: (1) the estimation of the poses of the limbs is not
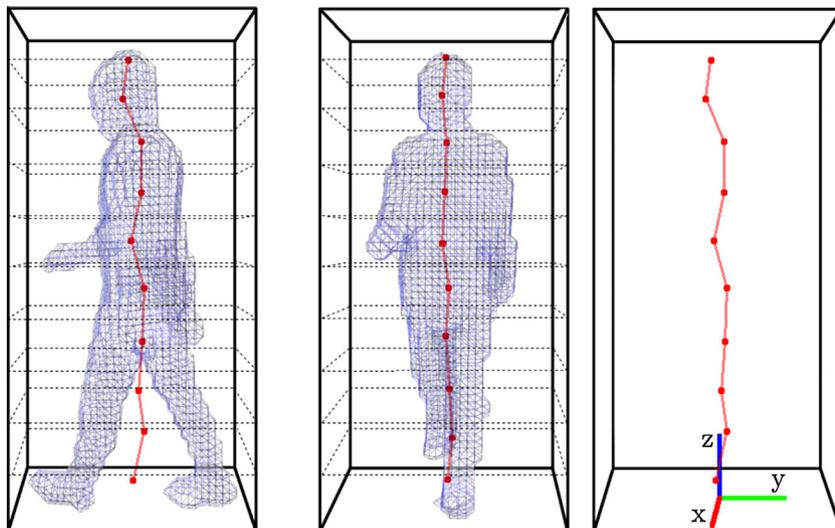
Fig. 2. The reconstructed model is divided into 3D stacked areas of the same size called slices (regions within dotted lines). Centroids are computed on each slice (red points). The gait feature is composed by a set of inner angles between the line joining each pair of consecutive centroids (red line) and the *z*-axis in R3. Best viewed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

robust from markerless motion; (2) it is not applicable for frontal view because the poses of the limbs become untraceable; and (3) it is assumed that subjects walk along a straight line segment.

In [29] is proposed the use of motion descriptors based on densely sampled short-term trajectories. This method is able to recognize people in curved trajectories with promising results.

The approaches of the third category rely on learning mapping/projection relationship of gaits under various viewing angles. The trained relationship may normalize gait features from different viewing angles into shared feature spaces. An example from this category can be read in [30], where LDA-subspaces are learned to extract discriminative information from gait features under each viewing angle.

A View Transformation Model (VTM) was introduced by [31] to transform gait features from different views into the same view. The method of Makihara et al. [31] creates a VTM based on frequency-domain gait features, obtained through Fourier Transformation. To improve the performance of this method, Kusakunniran et al. [32] created a VTM based on GEI optimized by linear discriminant analysis. A sparse-regression-based VTM for gait recognition under various views is also proposed in [3]. However, this method cannot deal with changes in the direction of motion. Neither it can be applied to recognize people walking on curved trajectories.

Although methods of the third category have better ability to cope with large view angle changes compared to other works, some common challenges are the following [3]: (1) performance of gait recognition decreases as the viewing angle increases; (2) since the methods rely on supervised learning, it is difficult to recognize gait under untrained/unknown viewing angles, (3) these methods implicitly assume that people walk along straight paths and that their walking direction does not change during a gait cycle (i.e., that people do not walk along curved trajectories).

Most of the view independent methods restrict the view angle change to a few angles, and they do not take into account curved trajectories. However, people sometimes walk on curved trajectories so as to turn a corner or to avoid an obstacle.

3. Proposed descriptor

This work presents a method to recognize humans walking on unconstrained paths, even with curved or straight trajectories,

and regardless direction changes. Thus, we propose a new gait descriptor that is able to cope with rotation changes, while preserving enough discriminatory information from the gait. In contrast to other related works, which discard a significant part of 3D information by computing the gait descriptors just from 2D images, our descriptor focuses on capturing 3D dynamical information of gait.

Let us assume that a workspace can be divided into N cubes of the same size (called voxels). This workspace contains information about the occupation, and can be denoted by:

$$V = \{v^i | i = (i_x, i_y, i_z)\} \mid i \in \mathbb{N}^3 \tag{1}$$

where $0 \leqslant i_x < N_x$, $0 \leqslant i_y < N_y$, $0 \leqslant i_z < N_z$, $i = (i_x, i_y, i_z)$ represents the voxel in Cartesian coordinates and $v^i \in \{1, 0\}$ depending on whether the voxel is occupied or empty. We assume a function $f : \mathbb{N}^3 \mapsto \mathbb{R}^3$ to map from voxel coordinates to scene coordinates. For the sake of simplicity, we also assume that the reference system of the monitored area is placed at the floor plane, in the center of the volume. Therefore, *x*- and *y*-axis are on that plane, whereas *z*-axis extends up.

Then, the workspace is divided into $H \in \mathbb{N}^+$ horizontal slices, as shown in Fig. 2. Let us also define a slice $S(h)$, $0 \leqslant h < H$ as a subset of voxels:

$$S(h) = \left\{ v^i \mid v^i \in V \wedge \left\lfloor h\frac{N_z}{H} \right\rfloor \leqslant i_z < \left\lfloor (h+1)\frac{N_z}{H} \right\rfloor \right\} \tag{2}$$

where N_z is the number of voxels of the discretized area with respect to the *z*-axis. The centroid $C_h = (\bar{x}, \bar{y}, \bar{z})$ of each slice $S(h)$ can be denoted by:

$$C_h = \frac{1}{|S(h)|} \sum_{v^i \in S(h)} v^i f(i) \tag{3}$$

where $|S(h)| = N_x \times N_y \times \frac{N_z}{H}$ represents the number of voxels of the slice $S(h)$. Next, we define the acute angle β_h between the normal vector to the floor plane ($\vec{Z} = (0, 0, 1)$) and the vector joining each pair of consecutive centroids as:

$$\alpha_h = \arccos\left(\frac{\vec{Z} \cdot \overrightarrow{C_h C_{h+1}}}{\| \overrightarrow{C_h} \| \| \overrightarrow{C_{h+1}} \|} \right), \quad 0 \leqslant h < H-2, \tag{4}$$

$$\beta_h = \min\{\alpha_h, 180 - \alpha_h\}, \tag{5}$$

Frame i from camera 1

Silhouette i from camera 1

Frame i from camera 2

Silhouette i from camera 2

3D Reconstruction and person detection

Descriptor generation and signature update

Gait classification

Person's identity

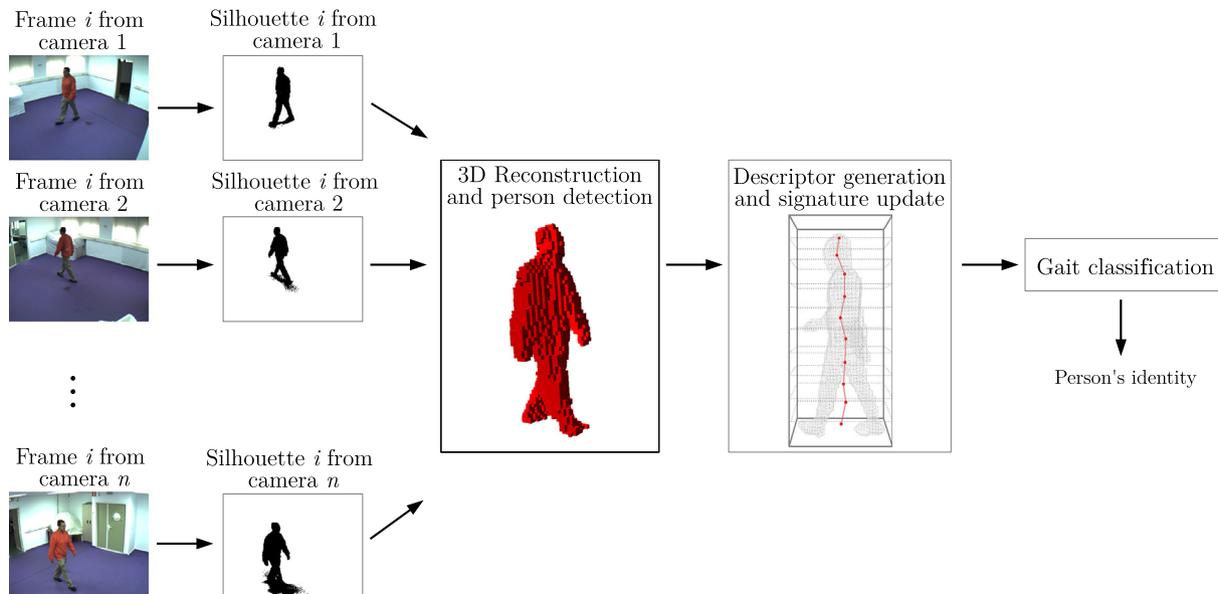Frame i from camera n

Silhouette i from camera n

Fig. 3. Pipeline of our approach.

where $\overrightarrow{C_h C_{h+1}}$ is the vector connecting the C_h and C_{h+1} centroids. Thus, for each instant t, our descriptor is a tuple of angular measurements that we can define as:

$$D_{H,t} = (\beta_{(0,t)}, \beta_{(1,t)}, \ldots, \beta_{(H-2,t)}). \tag{6}$$

In addition, to preserve the height of the subject as feature, if the slice is empty (e.g. partitions above the head), then the centroid corresponds to the center of the slice (i.e., the slice is considered fully occupied before computing its centroid).

The angular measurements are calculated on vectors in $\mathbf{R}^3$. We can say that our descriptor is rotation invariant and, therefore, the features extracted do not depend on the walking direction. Furthermore, even though is possible that two different subjects may have similar structure, differences on the dynamic of movement should help to differentiate them.

4. Proposed framework

Series of 3D occupation volumes are generated from multi-view video sequences at a rate of a 3D volume per time. Once a person has entered into the scene, our rotation invariant gait descriptor is computed on each volume. Because of the invariant properties of our gait descriptor, the direction of walking has no adverse effect on the recognition. The gait signature is updated at time on the basis of the previous gait descriptors.

The proposed algorithm consists of five steps which predict the identity of a walking human at time t. Following are described these steps in detail.

1. Silhouette extraction of each camera's view by a background subtraction technique [33].
2. 3D reconstruction from silhouettes captured from several viewpoints, by a Shape from Silhouette algorithm (SfS) [34].
3. Person detection.
4. Coarse-to-fine descriptor generation and gait signature update.
5. Classification of gait signature by a machine learning algorithm.

The aim of the first three stages of the algorithm is to generate a 3D volume with occupancy information of the lópe person at time t. On the other hand, the last two stages of the algorithms perform the feature extraction, signature generation and gait classification. The pipeline of our approach is shown in Fig. 3.

4.1. Feature extraction on reconstructed gait volumes

As previously indicated, we compute a 3D reconstruction for each frame of a gait sequence. In order to do this, we need to obtain silhouettes from multiple calibrated cameras. Then, when the individual has been detected, we extract features from the gait volume and use them to update the gait signature.

4.1.1. Silhouette extraction

The first step of our pipeline consists in obtaining the silhouettes of the walking subject. For this, we use a statistical approach for real-time robust background subtraction presented by Horprasert et al. in [33]. This approach is able to cope with local and global perturbations, such as illumination changes, casted shadows and highlights in controlled environments on static backgrounds.

Several silhouettes obtained by this algorithm are shown in Fig. 6. As it can be seen, despite the use of an advanced background subtraction technique, the silhouette is not perfectly defined. We should note that the performance of the recognition method also rely on the consistency of the silhouettes images, and therefore, of the 3D reconstructions.

After the background subtraction, we carry out a filtering through morphological operations as opening and closing. We do not do any other post-process operation.

4.1.2. 3D reconstruction

Since our method computes the gait descriptor from a 3D occupation volume, it requires a 3D reconstruction procedure, such as the Shape from Silhouette (SfS) standard algorithm. We assume a three-dimensional work area that is divided into cubes of the same volume called voxels. Let us assume that there is a set of cameras placed at known locations and that we have the silhouettes of the foreground objects, obtained by a background subtraction method. As described in more detail in [34], SfS method examine voxel projections in the foreground images in order to determine whether they belong to the shape of objects or not. Each voxel is projected in all the foreground images and if its projection lays completely into a silhouette in all the foreground images, then it is considered
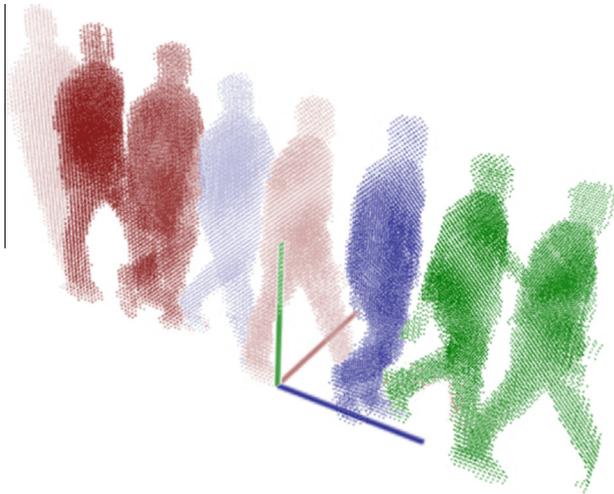
Fig. 4. Example of reconstructed segment of a gait sequence, sampled at 2 Hz, where each point represents a squared voxel. The time instant is represented by different colors. Best viewed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

occupied. However, if the voxel projects in a background region in any of the images, it is considered unoccupied. Finally, if the voxel projects partially in a foreground region, it is considered to belong to an edge and a decision must be made. We base this decision on the area of the projected voxel that lays into the silhouette. In the end, the result is a Boolean decision $(0, 1)$ indicating whether the region of the space represented by the voxel is empty or occupied. Fig. 4 shows the 3D reconstruction of a fragment belonging to a gait sequence.

In order to get a 3D reconstruction through SfS, calibration information for a multi-camera setup is also required. A classical black-white chessboard based technique [35] (OpenCV) can be used to get the intrinsics of each camera. For the extrinsics, we recommend Aruco library [36] whose detection of boards (several markers arranged in a grid) have two main advantages. First, since there is more than one marker, it is less likely to lose them all at the same time. Second, the more markers detected, the more points available for computing the camera extrinsics. Calibrating a multi-camera setup is a simple task that can be done in a few minutes using the above referenced techniques. To minimize the computational time, SfS could take advantage of the power of Graphics Processing Units (GPU), as it was proved in [37,38].

4.1.3. Person detection

It is assumed that although there is only one person in the scene, reconstructed shadows as well as noise can coexist, due to a poor segmentation. Because of this, it is required to detect whether the subject has fully entered into the scene, and track it. To detect the person, we use a threshold η, which refers to the number of occupied voxels corresponding to the size of a person. The volume belonging to a person is that which has a number of occupied voxels greater than η. This threshold is experimentally fixed in Section 5.

In addition to this, we consider that the subject has fully entered into the scene when the contour of the ground marginal distribution of occupied voxels P_z is separated by at least one voxel from the scene boundaries. So, let us define the ground marginal distribution of occupied voxels as the integral over the z-axis:

$$P_z(x, y) = \frac{1}{N_z} \sum_{i_z=0}^{N_z-1} v^{(x,y,i_z)}. \tag{7}$$

4.2. Gait identification

We next describe the steps employed by our system to extract the gait features, generate the gait signature and provide the name of the person.

4.2.1. Descriptor generation and gait signature update

The first step of our classification system is the generation of the gait descriptor $D_{(H,t)}$ at time t. The gait descriptor can be computed on a detected gait volume as was described in Section 3. Then, the gait signature can be built as a time series of gait descriptors obtained from the 3D reconstructed gait sequence.

In order to combine different description levels, we propose a coarse-to-fine refinement. We define the number of levels as:

$$0 < l \leqslant \lfloor \log_2 H \rfloor, \tag{8}$$

so that the first level descriptor contains features extracted from a volume divided into 2 slices, the second level descriptor contains features extracted from a volume divided into $H = 2^2$ slices, and so on until we have divided the volume into $H = 2^l$ slices. We can now concatenate the level descriptors to represent our coarse-to-fine descriptor as:

$$\mathscr{D}_{(l,t)} = (D_{(2,t)}, D_{(2^2,t)}, \ldots, D_{(2^l,t)}). \tag{9}$$

The gait signature is a temporal pattern of gait, a sample that feeds a classifier producing a class label corresponding to the identity of a particular person. Our signature is updated at every moment of the walking, and it allows to take place a synchronous classifying process. Thus, we define the gait signature $\mathscr{G}$ on a sliding temporal window of size L. Let us denote $\mathscr{G}$ as:

$$\mathscr{G}_{(l,t)} = (\mathscr{D}_{(l,t-L+1)}, \ldots, \mathscr{D}_{(l,t-1)}, \mathscr{D}_{(l,t)}), \tag{10}$$

which consist of a concatenation of L consecutive descriptors. In other words, our gait signature is updated at each instant of the gait by concatenating successive gait descriptors into a sliding temporal window of size L.

Our gait signature preserves the temporal consistency and has several advantages that are worth mentioning. First, the gait phase of the first frame of a gait sequence of a subject does not have to be the same for each person in the database. Second, it does not require the sequence to be split into gait cycles, and therefore it is not necessary to estimate the gait period. This makes our method less restrictive compared to other techniques from the literature such as [3,4,39] among others.

4.2.2. Classification

The gait signature $\mathscr{G}_{(l,t)}$ is in fact the feature vector used for classification. Each feature vector is assigned to a class label that corresponds to one of the person in the database. This idea is well known as multi-class classification system.

We adopt the subspace Component and Discriminant Analysis, based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), which seeks to project the original features to a subspace of lower dimensionality so that the best data representation and class separability can be achieved simultaneously [40]. Then we use a Support Vector Machine (SVM) [41] for training and classification.

The gait signature is based on the L previous volumes, and a possibly different class label can be produced for each new gait signature at each time. In order to smooth and reinforce the results over time, we use a majority vote policy over a sliding temporal window of size W. Our recognition algorithm provides the identity of the person as soon as possible. However, as the gait signature information is computed on L previous volumes, the use of this window causes a delay of $L + (W - 1)$ frames in obtaining the
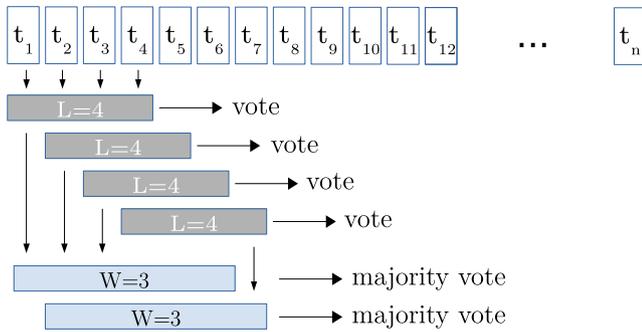
Fig. 5. Majority vote policy over a sliding temporal window. In the example, the size of the signature is set to $L = 4$, and the size of the voting window is set to $W = 3$.

identity. The majority voting system over a sliding temporal window is represented in Fig. 5.

5. Experiments and discussion

In order to validate our approach, we carry out diverse experiments on two publicly available datasets: the "AVA Multi-View Dataset for Gait Recognition" [42] and the "Kyushu University 4D Gait Database" [4]. In this section we try to answer, among others, the following questions:

- Is our descriptor a valid approach to recognize walking humans independently of the viewpoint? Is our proposal effective on curved trajectories?
- What level of refinement for our coarse-to-fine gait descriptor is required to achieve the best recognition rate?
- What is the effect of using PCA and PCA + LDA-based dimensionality reduction on the recognition performance?
- What is the influence of the sliding temporal window for majority voting policy on the recognition rate?
- How many cameras are needed to achieve good performance?
- Can the proposed model generalize well on unrestricted walking trajectories compared to other related works?

5.1. Datasets description

The first dataset where we perform our experiments is the "AVA Multi-View Dataset for Gait Recognition" (AVAMVG)[2] [42]. In AVAMVG, 20 subjects perform 9 walking trajectories in an indoor environment. Each trajectory is recorded by 6 color cameras placed around a room that is crossed by the subjects during the performance, according to the scheme of Fig. 6.

The video sequences of AVAMVG have a resolution of 640 × 480 pixels, and were recorded at a rate of 25 frames per second. For each actor, 9 gait sequences are captured in several trajectories as described in the figure by $\{t1, \ldots, t9\}$. Of these trajectories, 3 are straight ($\{t1, \ldots, t3\}$) and 6 are curved ($\{t4, \ldots, t9\}$). An example of this dataset is shown in Fig. 7, in which several subjects walk along different paths, from multiple viewpoints.

"Kyushu University 4D Gait Database" (KY4D)[3] [4], it is composed of sequential 3D models and image sequences of 42 subjects walking along four straight and two curved trajectories. The sequences were recorded by 16 cameras, at a resolution of 1032 × 776 pixels. Although the KY4D Gait Database also provide sequential 3D models of subjects, we have reconstructed them with
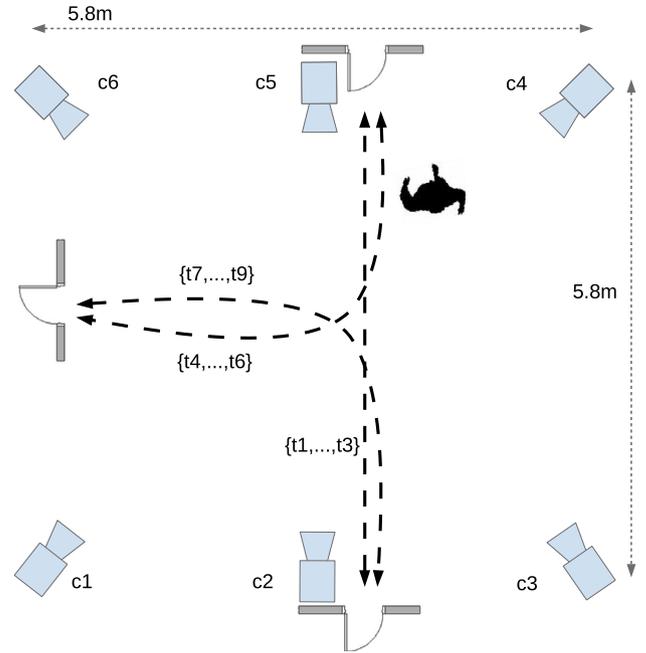
Fig. 6. Workspace setup used by AVAMVG dataset, where $\{c1, \ldots, c6\}$ represent the set of cameras of the multiview dataset and $\{t1, \ldots, t9\}$ represent the different trajectories followed by each actor of the dataset.

the same SfS method and resolution parameters used for the AVAMVG models. The intrinsics and extrinsics camera parameters are available for both databases. The camera setup of KY4D is shown in Fig. 8.

The aim of our approach is to recognize people walking on unconstrained paths, therefore we need databases containing video sequences of people walking on various types of trajectories, including curved paths. There are other publicly available gait databases [43], such as the "CASIA Dataset B" [44], the "CMU Motion of Body (MoBo)" [45], which are for changes on camera viewpoint and that include 2D gait images captured by multiple cameras. However, since these databases do not include people walking on curved trajectories, our approach cannot be tested on them.

5.2. Experimental results

This section explains the experimentation carried out to test our proposal. First of all, we need to determine the value of several parameters of our method. Thus, considering the 3D reconstruction stage, the first relevant parameter is the voxel size. We tested several voxel sizes, i.e. 0.015 m (3.3×10^{-6} m³), 0.03 m (2.7×10^{-5} m³), 0.06 m (21.6×10^{-5} m³), 0.09 m (72.9×10^{-5} m³) and 0.12 m (172.8×10^{-5} m³) of voxel side. Table 1 shows the influence of the voxel size on the recognition rate. The best results for both databases were found with a voxel side of 0.03 m (2.7×10^{-5} m³).

The average corporal volume for humans is $66.4L = 6.64 \times 10^{-2}$ m³ measured by the water displacement method in 521 people aged 17–51 years [46]. Using a voxel size of 2.7×10^{-5} m³, the number of voxels belonging to a person in a 3D volume should be about 2459. Thus, with a value of $\eta > 1 \times 10^3$ (see Section 4.1.3) the system should be able to efficiently detect when a person is in the scene.

With regards to the number of refinement levels (see Section 4.2.1), $l = 6$ is the maximum allowed with the above described voxel size and scene resolution (note that 2^l must be less than or

Fig. 7. Example of AVAMVG multiview dataset. People walking in different directions, from multiple points of view. Below the color images are shown their respective silhouettes, which have been obtained by using the background subtraction algorithm of Horprasert et al. [33]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
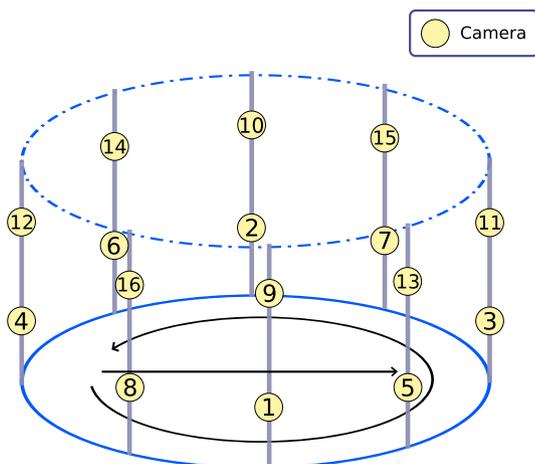

Fig. 8. Experimental setup of KY4D. Each camera is represented by a circle and a number which shows the order in which it was selected to evaluate the performance changes with respect to the number of cameras (see Section 5.2 for further details).

equal to N_z). The length of the signature is set to $L = 20$ and $L = 30$ for KY4D and AVAMVG respectively, because these values roughly match with the average length of a gait cycle in these databases.

Table 1
Correct classification rate [%] for both AVAMVG and KY4D datasets with different voxel sizes and values for the parameter l. Best results are marked in bold. The signature length is set to $L = 20$ for KY4D and $L = 30$ for AVAMVG. The size of the sliding temporal window for majority voting is set to $W = 1$ (see Section 4.2 for further details).

Voxel side (m)	l	AVAMVG [42] PCA + LDA	KY4D [4] PCA + LDA
0.015	5	75.73	68.21
0.015	6	74.49	68.69
0.03	**5**	**91.22**	**89.52**
0.03	**6**	**91.55**	**88.72**
0.06	4	83.52	72.59
0.06	5	89.92	74.63
0.09	3	64.81	44.34
0.09	4	85.80	56.88
0.12	3	51.23	29.33
0.12	4	79.60	42.59

We use a k-fold cross-validation strategy, where k corresponds to the number of trajectories. On the one hand, the AVAMVG dataset consists of 20 subjects performing 9 trajectories each, therefore each fold is composed by a tuple formed by a set of 20 sequences (one trajectory or sequence per actor) for testing, and by the remaining eight trajectories of each actor for training, i.e. 20×8 sequences for training and 20 sequences for test. It corresponds to a 9-fold cross-validation. On the other hand, since the KY4D dataset consists of 42 subjects and 6 trajectories, each fold is

Table 2
Correct classification rate [%] for both AVAMVG and KY4D datasets and several values for the parameter *l*. We use a *k*-fold cross-validation strategy, where *k* corresponds to the number of trajectories. The size of the sliding temporal window for majority voting is set to $W = 1$. The signature length is set to $L = 20$ for KY4D and $L = 30$ for AVAMVG. The voxel side is set to 0.03 m. Best results are marked in bold. (See main text for further details.)

l	AVAMVG [42]		KY4D [4]	
	PCA	PCA + LDA	PCA	PCA + LDA
1	10.98	N.A	12.13	N.A
2	53.22	45.50	57.25	N.A
3	69.38	63.22	74.84	73.16
4	84.74	83.83	85.03	85.38
5	92.24	91.22	**87.40**	**89.52**
6	**92.13**	**91.55**	86.59	88.72

Table 3
Number of features [AVG] for both AVAMVG and KY4D datasets and several values for the parameter *l*.

l	Without Dim. Red.	AVAMVG [42]		KY4D [4]	
		PCA	PCA + LDA	PCA	PCA + LDA
1	20	8.11	N.A	9.66	N.A
2	80	30.11	20	28.66	N.A
3	220	88.55	20	62.50	42
4	520	222.22	20	159.16	42
5	1140	550.77	20	394.00	42
6	2400	1277.22	20	911.50	42

composed by 42 sequences (one sequence per actor) for testing and by the remaining five sequences of each actor (i.e. 42×5 sequences) for training. It corresponds to a 6-fold cross-validation.

We use a C-SVC SVM, which allows imperfect separation of classes with penalty multiplier for outliers. We use Radial Basis Function as SVM kernels, since we obtained better results than with linear, polynomial, or sigmoid kernels. We set the same weight to all classes. To make the choice of SVM parameters independent of the sequence test data, we cross-validate the SVM parameters on the training set. Note that curved paths are sometimes longer than straight paths. In addition, some subjects walk faster than others and therefore cause a greater number of votes on the confusion matrix. To cope with this issue, we normalize by class the results of each trajectory.

In order to achieve the best data representation and class separability simultaneously, we apply PCA + LDA to the training and test data (see Section 4.2.2). Here we tested several SVM kernels, and finally we selected a C-SVC SVM with Radial Basis Function since we obtained better results than with linear, polynomial, or sigmoid kernels. With regard to PCA, we only retain 95% of the variance.

Table 2 shows the recognition rate for several values of the parameter *l* on AVAMVG and KY4D databases, with a voxel size 2.7×10^{-5} m³. It also shows the effect of the dimensionality reduction on the recognition rate. In this experiment, for the sake of simplicity, we disabled the sliding temporal window for majority voting ($W = 1$). As can be seen, the best results are obtained with high coarse-to-fine refinement level for the spatial division of the human body region. These values correspond to $H = 64$ for AVAMVG and $H = 32$ for KY4D. The average on number of features can be seen in Table 3. As can be observed, the number of features is considerably lower with PCA + LDA than with PCA. Therefore, if the system can be trained off-line, LDA allows SVM to handle feature spaces of lower dimensionality, and the identity of the individual could be given in less time.

We next conducted experiments in which we applied the sliding temporal window for majority voting policy. We use a *k*-fold cross-validation strategy where *k* is the number of trajectories, similar to the first experiment. As can be seen in Figs. 9 and 10, the use of a majority voting policy over a sliding temporal window significantly improves the performance of our method, which is close to achieving the perfect recognition. By using this approach, the results are smoothed and reinforced over time. However, the size of the window is limited by the number of available gait signatures in each sequence.

The results of Tables 4 and 5 show detailed results of the *k*-fold cross-validation experiment, which have been obtained by testing on each trajectory and training on the remaining $k - 1$ trajectories. It can be observed that our approach achieves good recognition
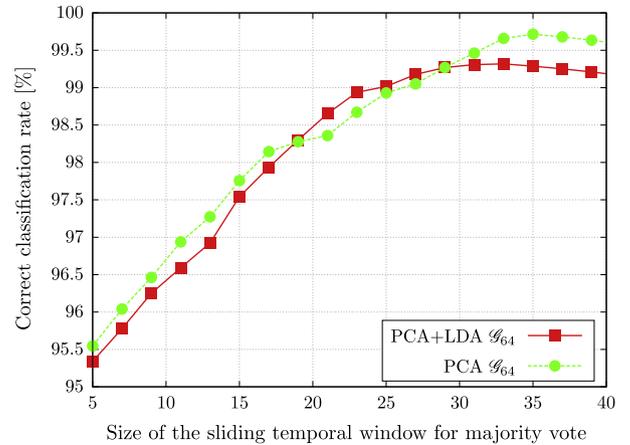

Fig. 9. Performance of our descriptor on the AVAMVG database for different lengths of the majority voting window.
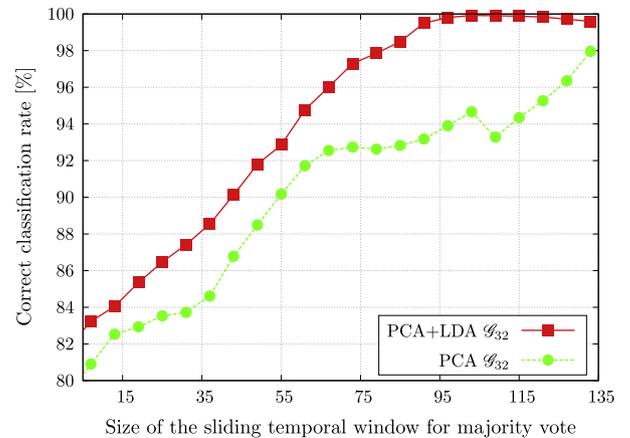

Fig. 10. Performance of our descriptor on the KY4D database for different lengths of the majority voting window.

rates for both dataset, even with curved paths. In this experiment, we have selected the optimal number of coarse-to-fine subdivisions of the human body region that we found in the first experiment for each database. Moreover, we have added the case where the use of the sliding temporal window for majority voting achieved the best results.

Our method does not require accurate models for feature extraction. In order to determine the number of cameras that should be employed and its effect on the performance, we have designed a *k*-fold cross validation experiment where *k* refers to the number of distinct trajectories. As in the others experiments, to make the choice of SVM parameters independent of the sequence test data, we cross-validate the SVM parameters on the

Table 4
Correct classification rate on AVAMVG [%]. Each column corresponds to a test trajectory, using the remaining trajectories as training set. Each row corresponds to a different configuration of the gait descriptor. Each entry contains the percentage of correct recognition for each tuple trajectory-setup. Best results are marked in bold.

Experiment	Straight paths			Curved paths						AVG
	t1	t2	t3	t4	t5	t6	t7	t8	t9	
$\mathscr{G}_{64}$-PCA-W = 1	97.53	90.68	98.02	93.38	79.98	92.11	92.30	91.76	93.45	92.13
$\mathscr{G}_{64}$-PCA-W = 35	**100**	**100**	**100**	99.58	**98.73**	**100**	99.74	**100**	99.34	**99.71**
$\mathscr{G}_{64}$-PCA + LDA-W = 1	94.84	88.47	97.43	93.57	83.85	91.56	92.29	91.56	90.45	91.55
$\mathscr{G}_{64}$-PCA + LDA-W = 32	99.47	98.52	**100**	**99.64**	98.55	99.55	**99.77**	98.93	**99.75**	99.35

Table 5
Correct classification rate on KY4D [%]. Each column corresponds to a test trajectory, using the remaining trajectories as training set. Each row corresponds to a different configuration of the gait descriptor. Each entry contains the percentage of correct recognition for each tuple trajectory-setup. Best results are marked in bold.

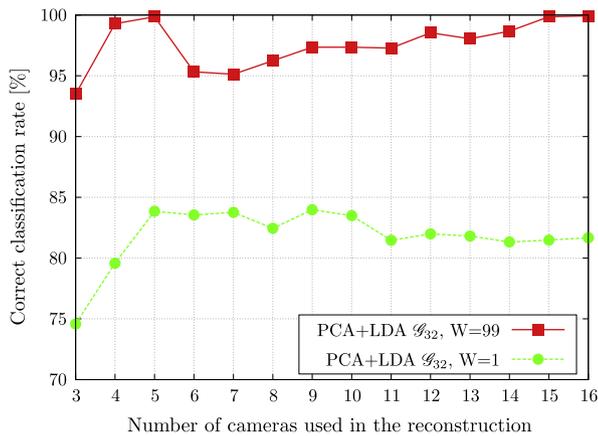| Experiment	Straight paths				Curved paths		AVG
	t1	t2	t3	t4	t5	t6	
$\mathscr{G}_{32}$-PCA-W = 1	93.12	97.55	96.44	96.16	54.25	86.86	87.39
$\mathscr{G}_{32}$-PCA-W = 130	**97.56**	**100**	**100**	**100**	90.24	**100**	97.96
$\mathscr{G}_{32}$-PCA + LDA-W = 1	94.98	98.62	99.10	97.22	58.09	89.09	89.51
$\mathscr{G}_{32}$-PCA + LDA-W = 99	**97.56**	**100**	**100**	**100**	**100**	**100**	**99.59**

Fig. 11. Performance of our descriptor on KY4D database for an increasing number of cameras.

Table 6
Correct classification rate [%] on AVAMVG gait dataset. Each row corresponds to a different method. The second column indicates the training trajectory. The third and fourth columns indicate the tested trajectory. For the method of Iwashita et al., we set $K = 5$ and $M = 40$ (see Section 4 of [4]). For the method of Castro et al., we selected PFM + PCAL100 + PCAH256 + pyr and $K = 150$ (see Section 3, Table II of [29]). For the method of Seely et al. [20] we have used the side-on, front-on, top-down average silhouettes (see Section 5 of [20]).

Method	Training trajectories	t4	t7	AVG
$\mathscr{G}_{64}, W = 57$, PCA + LDA	straight {t1,t2,t3}	90.69	96.57	93.63
$\mathscr{G}_{64}, W = 30$, PCA + LDA	straight {t1,t2,t3}	89.85	94.26	92.05
Castro et al. [29]	straight {t1,t2,t3}	85.00	95.00	90.00
Seely et al. [20]	straight {t1,t2,t3}	55.00	70.00	62.50
Iwashita et al. [4]	straight {t1,t2,t3}	35.14	37.71	36.42

training set. We selected the signature configuration that achieved the best performance in the previous experiments and then we tested it with a set of KY4D models which have been reconstructed using a number of cameras in the range 3–16. Fig. 9 shows the order in which the cameras were selected. For a two-camera reconstruction, we selected cameras 1 and 2. For a four-camera reconstruction, we selected cameras 1, 2, 3 and 4, and so on. This arrangement was motivated by the results exposed in the work

Table 7
Correct classification rate [%] on KY4D gait dataset. Each row corresponds to a different method. The second column indicates the training trajectory. The third and fourth columns indicate the tested trajectory. The results of the method of Iwashita et al. are taken directly from [4]. The results of [29] has been obtained by combining all the viewpoints of KY4D dataset by majority voting, PFM + PCAL150 + PCAH256 + pyr and $K = 200$ (see Section 3 of [29]). For the method of Seely et al. [20] we have used the side-on, front-on, top-down average silhouettes (see Section 5 of [20]).

Method	Training trajectories	Curve 1	Curve 2	AVG
$\mathscr{G}_{64}, W = 130$, PCA + LDA	straight {t1,t2,t3,t4}	68.29	77.50	72.89
$\mathscr{G}_{64}, W = 20$, PCA + LDA	straight {t1,t2,t3,t4}	63.16	73.53	68.34
Iwashita et al. [4]	straight {t1,t2,t3,t4}	61.90	71.40	66.65
Castro et al. [29]	straight {t1,t2,t3,t4}	58.50	61.00	59.75
Seely et al. [20]	straight {t1,t2,t3,t4}	19.51	35.00	27.25

of Takahashi et al. [47]. As can be seen in Fig. 11, with just 4 calibrated cameras, our method is able to correctly classify nearly 99% of individuals, independently of the path, even with curved trajectories.

5.3. Comparison with related work

We have compared our method with the recently published approaches of Iwashita et al. [4] and Castro et al. [29] because these methods are able to recognize people walking on curved trajectories, and they are therefore closely related with our aim. We have also compared with Seely et al. [20] because this method is an appearance-based approach which uses 3D reconstructed models. Since this method is not designed to cope with curved trajectories, we have aligned the gait volumes along the path.

We show the results of these experiments in Tables 6 and 7. In the case of the AVAMVG dataset, we trained with linear trajectories {t1,t2,t3} (all in the same set), and tested on curved trajectories t4 and t7 (see corresponding columns). For the KY4D dataset, we trained on linear trajectories {t1,t2,t3,t4} (all in the same set) and tested on curved trajectories t5 and t6. The percentage of relative difference on the average results between our proposal and the proposals of Iwashita et al. [4] and Castro et al. [29] is 8.56% and 18.02% respectively for KY4D, and 61.10% and 3.87% respectively for AVAMVG.

We have noticed a low performance of the method of Iwashita et al. when it is trained with straight paths and tested with curves of the AVAMVG dataset. In the AVAMVG dataset, depending on the
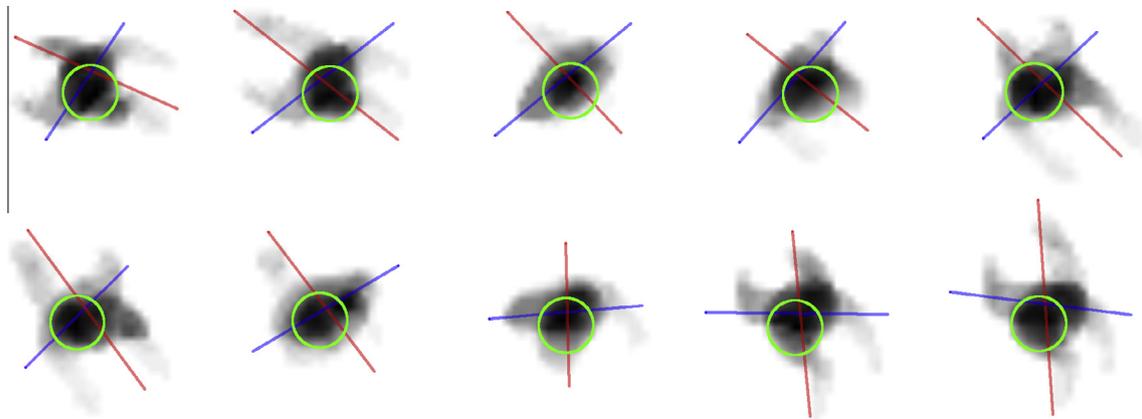
Fig. 12. Example of a curved gait cycle. We show several ground marginal distributions of occupied voxels (see Section 4.1.3). The velocity vector is represented by a red line, the blue line represents the torso main axis, and the position of the head is represented by a green circle. We can note that in a curved trajectory, the person rotates his/her torso and leans towards the walking direction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

viewpoint and trajectory, people appear at diverse scales, even showing partially occluded body parts. The method presented in [4] is based on high accuracy adaptive virtual image synthesis. In that method, affine moment invariants are used to describe the shape properties of the synthesized silhouettes. However, as can be seen in results of Table 6, it seems to decrease performance when silhouettes are rendered from inaccurate and inconsistent models (e.g., those that are reconstructed from poor segmentation results (see Fig. 7)). The results of Table 6 demonstrate that our method is robust against inaccurate and inconsistent models.

On the other hand, we know that on curved trajectories some persons tend to lean towards to the turning direction. Furthermore, some of them tend to rotate the torso and move the head towards the walking direction. It is shown in Fig. 12, where the torso main axis is drawn by a blue line, the velocity vector is drawn by a red line, and the head is indicated by a circle. The first two images of the top row and the last two images of the bottom row clearly show the leaning of the individual when it is depicting a curved trajectory. This could explain the low recognition rate obtained when the method is trained just with straight paths and it is tested with curved trajectories. As can be seen in Tables 6 and 7, the recognition rates fall well below to the results of Tables 2, 4 and 5, when the system is trained with both curved and straight trajectories. For these stated reasons, in order to identify people walking on curved trajectories, we suggest training the system with both straight and curved trajectories.

6. Conclusions

This paper has proposed a new gait recognition approach to identify people independently of the path, and regardless direction changes. In contrast to other view-independent approaches which restrict the view change to a few angles and cannot cope with curved trajectories, our approach allow people to walk freely in the scene without adversely affecting to the recognition, even with curved trajectories.

A new rotation invariant gait descriptor has been proposed to cope with rotation changes on curved trajectories, while preserving enough discriminatory information from the gait. Our descriptor focuses on capturing 3D dynamical information of gait, unlike other related works which discard a significant part of 3D information by computing the gait descriptors just from 2D images.

This approach does not require the sequence to be split into gait cycles, because the gait signature is built on a sliding temporal window. In addition, another sliding temporal window for majority vote policy is used to smooth and reinforce the results over time. The experiments have been conducted on two datasets, and they have shown that our approach is able to reach a correct classification rate close to 100%.

Despite of using 3D models, we have proved that our descriptor does not require high-accurate reconstructions, and it efficiently works with only four calibrated cameras.

References

[1] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Trans. Syst., Man, Cybernet., Part C: Appl. Rev. 34 (3) (2004) 334–352, http://dx.doi.org/10.1109/TSMCC.2004.829274.

[2] M.A. Hossain, Y. Makihara, J. Wang, Y. Yagi, Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control, Pattern Recogn. 43 (6) (2010) 2281–2291, http://dx.doi.org/10.1016/j.patcog.2009.12.020.

[3] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Gait recognition under various viewing angles based on correlated motion regression, IEEE Trans. Circ. Syst. Video Technol. 22 (6) (2012) 966–980, http://dx.doi.org/10.1109/TCSVT.2012.2186744.

[4] Y. Iwashita, K. Ogawara, R. Kurazume, Identification of people walking along curved trajectories, Pattern Recogn. Lett. 48 (0) (2014) 60–69, http://dx.doi.org/10.1016/j.patrec.2014.04.004. Celebrating the life and work of Maria Petrou.

[5] S. Singh, K. Biswas, Biometric gait recognition with carrying and clothing variants, in: Pattern Recognition and Machine Intelligence, Lecture Notes in Computer Science, vol. 5909, Springer, Berlin Heidelberg, 2009, pp. 446–451, http://dx.doi.org/10.1007/978-3-642-11164-8_72.

[6] A. Tsuji, Y. Makihara, Y. Yagi, Silhouette transformation based on walking speed for gait identification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 717–722, http://dx.doi.org/10.1109/CVPR.2010.5540144.

[7] S. Yu, D. Tan, T. Tan, Modelling the effect of view angle variation on appearance-based gait recognition, in: Computer Vision – ACCV 2006, Lecture Notes in Computer Science, vol. 3851, Springer, Berlin, Heidelberg, 2006, pp. 807–816, http://dx.doi.org/10.1007/11612032_81.

[8] A. Kale, N. Cuntoor, B. Yegnanarayana, A. Rajagopalan, R. Chellappa, Gait analysis for human identification, Audio- and Video-Based Biometric Person Authentication, vol. 2688, Springer, Berlin Heidelberg, 2003, pp. 706–714.

[9] M.-H. Cheng, M.-F. Ho, C.-L. Huang, Gait analysis for human identification through manifold learning and HMM, Pattern Recogn. 41 (8) (2008) 2541–2553, http://dx.doi.org/10.1016/j.patcog.2007.11.021.

[10] L. Wang, T. Tan, H. Ning, W. Hu, Silhouette analysis-based gait recognition for human identification, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1505–1518, http://dx.doi.org/10.1109/TPAMI.2003.1251144.

[11] C.P. Lee, A.W. Tan, S.C. Tan, Gait recognition via optimally interpolated deformable contours, Pattern Recogn. Lett. 34 (6) (2013) 663–669, http://dx.doi.org/10.1016/j.patrec.2013.01.013.

[12] S. Das Choudhury, T. Tjahjadi, Silhouette-based gait recognition using procrustes shape analysis and elliptic fourier descriptors, Pattern Recogn. 45 (9) (2012) 3414–3426, http://dx.doi.org/10.1016/j.patcog.2012.02.032.

[13] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (2006) 316–322, http://dx.doi.org/10.1109/TPAMI.2006.38.

[14] C. Rougier, E. Auvinet, J. Meunier, M. Mignotte, J. de Guise, Depth energy image for gait symmetry quantification, in: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011, pp. 5136–5139, http://dx.doi.org/10.1109/IEMBS.2011.6091272.

[15] M. Hofmann, S. Bachmann, G. Rigoll, 2.5d gait biometrics using the depth gradient histogram energy image, in: IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2012, pp. 399–403, http://dx.doi.org/10.1109/BTAS.2012.6374606.

[16] C.P. Lee, A.W. Tan, S.C. Tan, Time-sliced averaged motion history image for gait recognition, J. Vis. Commun. Image Represent. 25 (5) (2014) 822–826, http://dx.doi.org/10.1016/j.jvcir.2014.01.012.

[17] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, C. Fookes, Gait energy volumes and frontal gait recognition using depth images, in: International Joint Conference on Biometrics (IJCB), IEEE, Washington DC, USA, 2011, pp. 1–6, http://dx.doi.org/10.1109/IJCB.2011.6117504.

[18] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, C. Fookes, The backfilled GEI – a cross-capture modality gait feature for frontal and side-view gait recognition, in: International Conference on Digital Image Computing Techniques and Applications (DICTA), 2012, pp. 1–8, http://dx.doi.org/10.1109/DICTA.2012.6411694.

[19] P. Chattopadhyay, A. Roy, S. Sural, J. Mukhopadhyay, Pose depth volume extraction from RGB-D streams for frontal gait recognition, J. Vis. Commun. Image Represent. 25 (1) (2014) 53–63.

[20] R. Seely, S. Samangooei, M. Lee, J. Carter, M. Nixon, The University of Southampton Multi-Biometric Tunnel and introducing a novel 3D gait dataset, in: 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems. BTAS 2008, 2008, pp. 1–6, http://dx.doi.org/10.1109/BTAS.2008.4699353.

[21] G. Shakhnarovich, L. Lee, T. Darrell, Integrated face and gait recognition from multiple views, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 2001, pp. I-439–I-446, http://dx.doi.org/10.1109/CVPR.2001.990508.

[22] R. Bodor, A. Drenner, D. Fehr, O. Masoud, N. Papanikolopoulos, View-independent human motion classification using image-based reconstruction, Image Vis. Comput. 27 (8) (2009) 1194–1206, http://dx.doi.org/10.1016/j.imavis.2008.11.008.

[23] A. Laurentini, The visual hull concept for silhouette-based image understanding, IEEE Trans. Pattern Anal. Mach. Intell. 16 (2) (1994) 150–162, http://dx.doi.org/10.1109/34.273735.

[24] A. Kale, A. Roy-Chowdhury, R. Chellappa, Towards a view invariant gait recognition algorithm, in: IEEE Conference on Advanced Video and Signal Based Surveillance, 2003, pp. 143–150, http://dx.doi.org/10.1109/AVSS.2003.1217914.

[25] F. Jean, A.B. Albu, R. Bergevin, Towards view-invariant gait modeling: computing view-normalized body part trajectories, Pattern Recogn. 42 (11) (2009) 2936–2949, http://dx.doi.org/10.1016/j.patcog.2009.05.006.

[26] S. Jeong, T.-h. Kim, J. Cho, Gait recognition using description of shape synthesized by planar homography, J. Supercomput. 65 (1) (2013) 122–135, http://dx.doi.org/10.1007/s11227-013-0897-8.

[27] J. Han, B. Bhanu, A. Roy-Chowdhury, A study on view-insensitive gait recognition, in: IEEE International Conference on Image Processing (ICIP), vol. 3, 2005, pp. III-297–300, http://dx.doi.org/10.1109/ICIP.2005.1530387.

[28] M. Goffredo, I. Bouchrika, J. Carter, M. Nixon, Self-calibrating view-invariant gait biometrics, IEEE Trans. Syst., Man, Cybernet., Part B: Cybernet. 40 (4) (2010) 997–1008, http://dx.doi.org/10.1109/TSMCB.2009.2031091.

[29] F.M. Castro, M.J. Marín-Jiménez, R.M. Carnicer, Pyramidal fisher motion for multiview gait recognition, in: 22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24–28, 2014, pp. 1692–1697, http://dx.doi.org/10.1109/ICPR.2014.298.

[30] N. Liu, Y.-P. Tan, View invariant gait recognition, in: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), 2010, pp. 1410–1413, http://dx.doi.org/10.1109/ICASSP.2010.5495466.

[31] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Gait recognition using a view transformation model in the frequency domain, in: Computer Vision – ECCV 2006, Lecture Notes in Computer Science, vol. 3953, Springer, Berlin Heidelberg, 2006, pp. 151–163, http://dx.doi.org/10.1007/11744078_12.

[32] W. Kusakunniran, Q. Wu, H. Li, J. Zhang, Multiple views gait recognition using view transformation model based on optimized gait energy image, in: IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009, pp. 1058–1064, http://dx.doi.org/10.1109/ICCVW.2009.5457587.

[33] T. Horprasert, D. Harwood, L.S. Davis, A statistical approach for real-time robust background subtraction and shadow detection, in: Proc. IEEE ICCV, 1999, pp. 1–19.

[34] L. Díaz-Más, R. Muñoz-Salinas, F. Madrid-Cuevas, R. Medina-Carnicer, Shape from silhouette using Dempster–Shafer theory, Pattern Recogn. 43 (6) (2010) 2119–2131, http://dx.doi.org/10.1016/j.patcog.2010.01.001.

[35] G. Bradski, A. Kaehler, Learning OpenCV: Computer Vision with the OpenCV Library, O'Reilly, Cambridge, MA, 2008.

[36] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, M. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recogn. 47 (6) (2014) 2280–2292, http://dx.doi.org/10.1016/j.patcog.2014.01.005.

[37] G. Haro, Shape from silhouette consensus, Pattern Recogn. 45 (9) (2012) 3231–3244, http://dx.doi.org/10.1016/j.patcog.2012.02.029. Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

[38] S. Yous, H. Laga, M. Kidode, K. Chihara, GPU-based shape from silhouettes, in: Proceedings of the 5th International Conference on Computer Graphics and Interactive Techniques in Australia and Southeast Asia, GRAPHITE '07, ACM, New York, NY, USA, 2007, pp. 71–77, http://dx.doi.org/10.1145/1321261.1321274.

[39] K. Bashir, T. Xiang, S. Gong, Cross-view gait recognition using correlation strength, in: Proceedings of the British Machine Vision Conference, BMVA Press, 2010, pp. 109.1–109.11, http://dx.doi.org/10.5244/C.24.109.

[40] P. Huang, C. Harris, M. Nixon, Recognising humans by gait via parametric canonical space, Artif. Intell. Eng. 13 (4) (1999) 359–366, http://dx.doi.org/10.1016/S0954-1810(99)00008-4.

[41] C.J. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Disc. 2 (2) (1998) 121–167, http://dx.doi.org/10.1023/A:1009715923555.

[42] D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, R. Muñoz-Salinas, The AVA multi-view dataset for gait recognition, in: Activity Monitoring by Multiple Distributed Sensing, Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 26–39, http://dx.doi.org/10.1007/978-3-319-13323-2_3.

[43] J.M. Chaquet, E.J. Carmona, A. Fernández-Caballero, A survey of video datasets for human action and activity recognition, Comput. Vis. Image Underst. 117 (6) (2013) 633–659, http://dx.doi.org/10.1016/j.cviu.2013.01.013.

[44] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, in: 18th IEEE International Conference on Image Processing (ICIP), 2011, pp. 2073–2076, http://dx.doi.org/10.1109/ICIP.2011.6115889.

[45] R. Gross, J. Shi, The CMU Motion of Body (MOBO) Database, Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.

[46] D. Drinkwater, W. Ross, Anthropometric fractionation of body mass, Kinanthropometry, vol. 2, University Park Press, Baltimore, 1980, pp. 178–189.

[47] T. Takahashi, O. Matsugano, I. Ide, Y. Mekada, H. Murase, Planning of multiple camera arrangement for object recognition in parametric eigenspace, in: 18th International Conference on Pattern Recognition, 2006. ICPR 2006, vol. 1, 2006, pp. 603–606, http://dx.doi.org/10.1109/ICPR.2006.937.