
Entropy Volumes for Viewpoint Independent Gait Recognition

D. López-Fernández · F.J. Madrid-Cuevas · A. Carmona-Poyato ·
R. Muñoz-Salinas · R. Medina-Carnicer

Received: date / Accepted: date

Abstract Gait as biometrics has been widely used for human identification. However, direction changes cause difficulties for most of the gait recognition systems, due to appearance changes. This study presents an efficient multi-view gait recognition method that allows curved trajectories on completely unconstrained paths for indoor environments. Our method is based on volumetric reconstructions of humans, aligned along their way. A new gait descriptor, termed as Gait Entropy Volume (GEnV), is also proposed. GEnV focuses on capturing 3D dynamical information of walking humans through the concept of entropy. Our approach does not require the sequence to be split into gait cycles. A GEnV based signature is computed on the basis of the previous 3D gait volumes. Each signature is classified by a Support Vector Machine, and a majority voting policy is used to smooth and reinforce the classifications results. The proposed approach is experimentally validated on the “AVA Multi-View Gait Dataset (AVAMVG)” and on the “Kyushu University 4D Gait Database (KY4D)”. The results show that this new approach achieves promising results in the problem of gait recognition on unconstrained paths.

Keywords Gait Entropy Volume · Gait recognition · View-Independent · 3D reconstruction · Curved trajectories.

D. López Fernández · F.J. Madrid-Cuevas · A. Carmona-Poyato · R. Muñoz-Salinas · R. Medina-Carnicer
Department of Computing and Numerical Analysis, Edificio Einstein. Campus de Rabanales. Maimónides Institute for Biomedical Research (IMIBIC). University of Córdoba, 14071, Córdoba, Spain.
Tel.: +34-957212255
E-mail: D. López Fernández (i52lofed@uco.es)

1 Introduction

Biometrics is the science that deals with the identification of individuals from an anatomical and behavioural point of view. Some of the current biometric methods use face, voice, iris or fingerprint for human recognition due to its universality and uniqueness [4].

The gait is a human feature that contains information about the physical and psychological state of the person. What is especially interesting is that each individual describes a unique gait pattern, which means it can be used as a biometric indicator [9]. Gait as biometric feature for identification can be applied discreetly without needing the active participation of the individuals.

Previous studies on gait recognition have been classified into two categories: model-based approaches and appearance-based approaches. The model-based methods represent gait using the parameters of a body configuration model which is estimated over time, whereas appearance-based approaches characterize the human gait pattern by a compact representation, without having to develop any model for feature extraction and having practical application even with low quality images where the color and texture information is lost.

In addition, regarding viewing angle, the previous work can be categorized into two approaches: view-dependent and view-independent approaches. The view-dependent approaches assume that will not happen any appearance change during walking. In such methods, a change in the appearance, caused by a viewing angle change, will adversely affect to the performance [41]. 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 at all frames in one gait cycle. This is shown in Fig. 1.

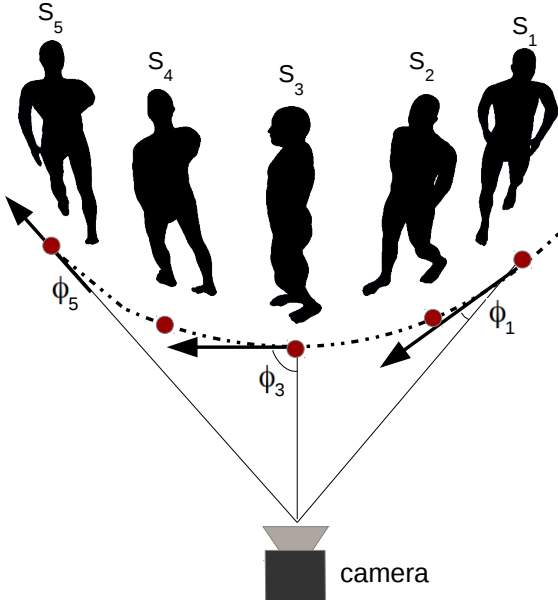


Fig. 1 In a curved path, the observation angle between the walking direction of the subject and optical axis of the camera is gradually changed, which affects the silhouette appearance.

While most of appearance-based approaches are view-dependent, the model-based approaches are generally invariant to rotational effects and slight variations in the viewpoint. However, they are characterised by complex searching and mapping processes, which increase the computational cost.

This paper presents an efficient view-independent and appearance-based method to recognize people walking along curved trajectories on completely unconstrained paths. We also propose a gait descriptor which focuses on capturing the maximum amount of dynamical gait information in a 3D sense.

Some potential applications of this work are access control in special or restricted areas (e.g. military bases and governmental facilities) or smart video surveillance where subjects do not know they are being monitored (e.g. bank offices). It also could 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).

This article is organized as follows. Section 2 describes works related to the topic of gait recognition. Section 3 explain the details of the proposed algorithm, gait descriptor and derived signatures. An analysis of

the proposed method and the performance is given in Section 4. Finally, we conclude this paper in Section 5.

2 Related work

2.1 View-dependent approaches

One of the earliest view-dependent approaches can be seen in [25], where the outer contour of the binarized silhouette from a lateral view is used to build a descriptor which contains both structural features and dynamic aspects of gait. The contours of silhouettes have also been used directly [40,18], and through their Fourier descriptors [33,29]. In [7], Hidden Markov Models have been trained from feature vectors derived from binary silhouettes.

In addition, in [8] it is presented a gait recognition method which combines spatio-temporal motion characteristics, statistical and physical parameters of a person for its classification. This is carried out by analysing the shape of the silhouette using Procrustes Shape Analysis and Elliptic Fourier Descriptors. In [15] it is proposed a gait representation called Gait Energy Image (GEI), which key idea is to compute the average of all silhouette images for a single gait cycle.

Based on the idea of GEI, Depth Energy Image (DEI) [35] consists in the average of frontal depth silhouettes for a gait cycle. In [17], a new feature called Depth Gradient Histogram Energy Image (DGHEI) is proposed to extend GEI by including depth information.

In [37] it is presented the Gait Energy Volume (GEV), which is an average voxel-discretized volume. The authors apply GEV on partial reconstructions obtained with depth sensors from the front view of the individual. The front view depth gait image and the side view 2D gait silhouette is combined by means of a back-filling technique in [38]. The front view depth image is also captured in [6].

A work closely related to our approach was presented in [1], in which the Gait Entropy Image (GEnI) is presented. GEnI encodes in a single image, the randomness of pixel values in the silhouette images over a complete gait cycle. More specifically, considering the intensity value of the silhouettes at a fixed pixel location as a discrete random variable, entropy measures the uncertainty associated with the random variable over a complete gait cycle. Dynamic body areas which undergo consistent relative motion during a gait cycle (e.g. leg, arms) lead to high gait entropy values, whereas those areas that remain static (e.g. torso) give rise to low values.

A human silhouette is extracted from the side view of the gait sequence. After applying size normalization and horizontal alignment to each silhouette image, gait cycles are segmented by estimating the gait frequency using a maximum entropy estimation technique. GEnI is defined as:

$$\text{GEnI}(x, y) = - \sum_{k=1}^K p_k(x, y) \log_2 p_k(x, y), \quad (1)$$

where x, y are the pixel coordinates and $p_k(x, y)$ is the probability of the pixel (x, y) for the label $k \in K$. The silhouettes are binary images, and therefore $K : \{0, 1\}$, so that $p_1(x, y) = \frac{1}{T} \sum_{t=1}^T I(x, y)$, and $p_0(x, y) = 1 - p_1(x, y)$, where T is the length of the gait cycle and I is the binary image.

In [2], the use of the GEnI descriptor is proposed to distinguish the dynamic and static areas of a GEI by measuring Shannon entropy at each pixel location. The authors use the GEnI to perform a feature selection, based on the relevance of gait features extracted from GEI, instead of using GEnI as gait descriptor directly as in [1].

These approaches [1, 2] have some drawbacks that are worth mentioning. Firstly, they are based on computing entropy over the side view of the gait sequence. However, some people tend to swing their arms from side to side while walking, and they often rotate their torso slightly. This fact lead us to think that some dynamic and structural information of the individual is lost when GEI or GEnI is only computed over the side view of the gait sequence, because by just using a single 2D image view, a large part of 3D gait information is discarded. The Fig. 2 shows the GEnI, computed over several gait cycles.

Secondly, these approaches implicitly assume that people walk along a straight line and their walking direction does not change during one gait cycle. However, in real-life situations people walk on curved trajectories in order to turn a corner or to avoid an obstacle. When the subject is walking along a curved path, the viewing angle change causes a decrease in the performance for most single-view based conventional methods, due to appearance changes.

2.2 View-independent approaches

Appearance changes due to viewing angle changes cause difficulties for most of the appearance-based gait recognition methods. This situation cannot be easily avoided in practical applications. There are three major approach categories to sort out this problem, 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 [27].

Approaches of the first category are represented by [36, 3, 21]. In [36] a polyhedral and surface-mapped 3D approximation of the Visual Hull [28] (VH) is used to design a multi-modal recognition approach, that combines both face and gait recognition. Although a polyhedral VH model is computed, the gait recognition scheme is based on silhouette analysis, which does not take advantage of all the available information because the recognition is based on single view silhouette analysis, instead of exploiting the 3D model.

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

In [21], an observation angle is estimated from the walking direction, by fitting a 2D polynomial curve to the foot points. Virtual images are synthesized from 3D models, 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 main advantage of this method is that the setup assumes multiple cameras for training, but only one camera for testing. It is able to recognize people walking on curved paths. However, as in the above two works, despite 3D models are used, features are extracted from 2D images, so that, the amount of available information is restricted. On the other hand, shadows on the floor complicate the estimation of the foot points in silhouette images.

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. The main disadvantage of this method is that the performance is significantly dropped when the angle between image plane and sagittal plane is large. Besides, the synthesis of a canonical view is only feasible from a limited number of initial views.



Fig. 2 Several examples of GENI, computed over a gait cycle. The gray level represent the entropy value in a pixel. As can be seen, legs and arms have high gait entropy value, whereas static areas as torso have low values of entropy.

In [22], a method based on homography to compute view-normalized trajectories of body parts obtained from monocular video sequences was proposed. However, 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 [23], but this method seems not to be applicable when the person is walking nearly parallel to the optical axis. In [16] 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.

A self-calibrating view-independent gait recognition based on model-based gait features is proposed in [13]. 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 robust from markerless motion; 2) it is not applicable for frontal view because the poses of the limbs become untraceable; and 3) this method assume that subjects walk along a straight line segment.

In [12] 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 [32] to transform gait features from different views into the same view. The method of Makihara *et al.* [32] creates a VTM based on frequency-domain gait features, obtained through Fourier Transformation. To improve the performance of this method, Kusakunniran *et al.* [26] 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 [27]. However, this method cannot cope with changes in the direction of motion.

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 [27]: (1) performance of gait recognition decreases as the viewing angle increases; (2) since the methods rely on supervised learning, it will be difficult for recognizing 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 single gait cycle (i.e., that people do not walk along curved trajectories). However, people often walk on curved trajectories in order to turn a corner or to avoid an obstacle.

3 Proposed method

This paper proposes a method to recognize people walking on unconstrained paths, even if they walk along curved trajectories or change direction. The gait descriptors are extracted from 3D aligned human reconstructions, so that a greater amount information is analysed in contrast to other related work, which compute the gait descriptors just from 2D images, discarding a significant part of the 3D gait information.

In gait recognition, the dynamic of the information is very useful, because it represents temporal transitions in human behaviour. We propose to use the dynamic of information about the relative motion on aligned 3D gait reconstructions by measuring entropy at each voxel location.

Some entropy-based algorithms [1,2] split the gait sequence into gait cycles, and to do this, the sequence has to be analysed from the beginning to the end in order to obtain the gait frequency. However, our gait-recognition algorithm provides the name for the person as soon as possible without splitting the gait sequence into gait cycles nor computing the whole gait sequence before providing a response, what makes our method less restrictive than many other techniques described in the literature.

Besides, some of these methods use single lateral cameras. However, by using a single lateral camera, the individual would leave the field of view very soon, so the length of the gait sequence is restricted. Since 3D gait volumes are centred with respect to a global reference system and aligned along their way, we can get more images from the volumes along a sequence.

The proposed recognition algorithm is shown in Fig. 3. The algorithm consists of five steps that are carried out at each time t . Entropy is computed on a sliding temporal window of size L . These steps are exposed in detail in this section:

1. Silhouette extraction of each camera's view by a background subtraction technique [19].
2. 3D reconstruction from silhouettes captured from several viewpoints, by Shape from Silhouette algorithm [10].
3. Person detection and gait alignment.
4. Gait signature update.
5. Classification of gait signature by a machine learning algorithm whose output is the identity of a person known by the multi-class classification system.

The aim of the first three steps of the algorithm is to generate a 3D volume with occupancy information of the person at time t . In addition, the last two steps of the algorithm perform the signature update and the gait classification.

3.1 3D Reconstruction, detection and alignment

Our method starts by computing a 3D reconstruction of the individual. To do this, we need the silhouettes from multiple calibrated cameras. Calibrating a multi-camera setup is a simple process that can be done in a few minutes.

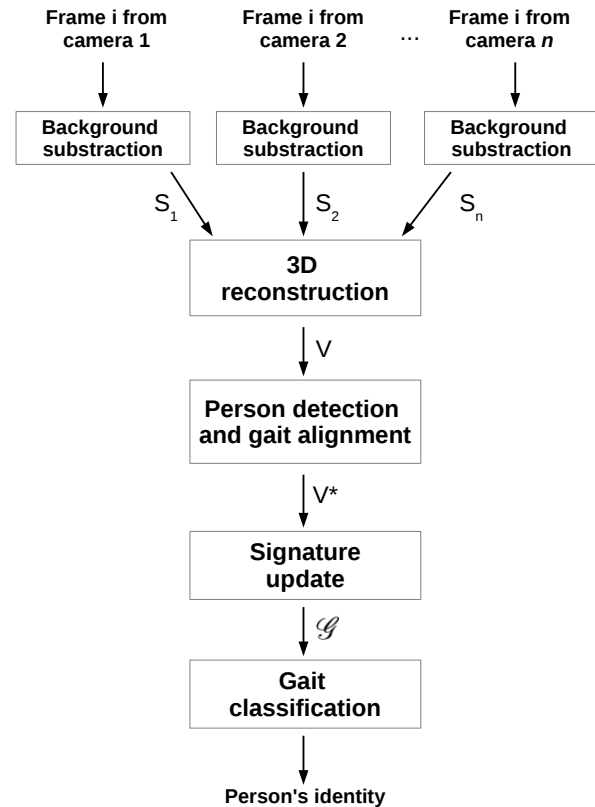


Fig. 3 Steps of our gait-recognition algorithm at time t . S_1 , S_2 and S_n represent human silhouettes extracted from the input frames, whereas V , V^* , and \mathcal{G} represent a reconstructed 3D scene, an aligned 3D human volume, and a gait signature, respectively.

After the 3D reconstruction, the gait volumes are aligned and centred with respect to a global reference system, so that the generation of the descriptors can be done as if the person had walked on a treadmill in a certain direction. In the following we explain these steps.

3.1.1 Silhouette extraction

As we have a static background, we use Horprasert's algorithm [19] to obtain silhouettes of the walking humans. Horprasert's algorithm is able to deal with local and global perturbations, such as illumination changes, shadows and lightening in controlled environments. This algorithm is able to detect moving objects on color images, in a scene that may also contain shadows. After this, we filter the noise of binary images through morphological operations such as opening and closing.

3.1.2 3D reconstruction

Since our method generates the gait descriptors from 3D occupation volumes, a 3D reconstruction procedure, such as the Shape from Silhouette (SfS) standard algorithm is required.

We assume a three-dimensional work area that is divided into cubes of the same size called voxels. Let us also 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 [10], 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 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. This procedure requires calibration parameters, such as the camera matrix, distortion coefficients (intrinsic parameters), pose and orientation (extrinsic parameters) of each camera.

At 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 gait sequence.

3.1.3 Volume detection and alignment

Since we have a V_t reconstructed volume of a person in an instant t along the way, it is required a mechanism of detection and alignment to achieve the independence which refers to the point of view. So that the individual can walk freely in the scene without the orientation and direction of its motion can affect to the subsequent generation of gait descriptors.

For this purpose, a 3D reconstruction of the scene is carried out at time t , and then the volume belonging to a person has to be detected. Although there is only one individual in the scene, we assume that reconstructed shadows as well as noise can coexist, due to poor segmentation results.

By obtaining the ground marginal distribution of occupied voxels (ground projection of the volume), we detect the volume belonging to a person as that which has a greater volume than a certain threshold θ , and its volume has fully entered into the workspace.

When the volume of a person has been detected, the centroid p of the ground region corresponding to the detected volume is calculated and used to estimate the trajectory, which is determined by the displacement vector, defined as:

$$\mathbf{v}_t = p_t - p_{t-1}, \quad (2)$$

where t is the current time, p_t is the current position of the centroid, and p_{t-1} is the last known position. The angle of the displacement vector is calculated using the expression:

$$\alpha_t = \arctan \frac{v_{t_y}}{v_{t_x}}. \quad (3)$$

Ground projections of the individual can be seen in Fig. 5, where the principal axis and displacement vector are represented at several moments of the gait sequence.

The angle of the displacement vector is used to construct a rotation matrix, which is applied to align the gait volume by changing the coordinate system, rotating it about the vertical axis. Then, to reduce the workspace where the descriptor will be computed, the aligned gait volume is translated into a bounding-box of average human's size, so that the workspace where the descriptor will be computed is reduced.

Although we assume a constant walking speed, the individual could vary moderately the walking speed at certain times of the sequence. It could happen, for example, when the individual is depicting a closed curved path.

If the walking speed of the individual is very slow in an instant t , $|\mathbf{v}_t|$ will be too small, which could result in a noisy estimation of the angle α_t . To attenuate this noise, and thus smooth the path, we propose a weighted average of the displacement vector angle as follows:

$$\bar{\alpha}_t = \alpha_t \cdot \beta + \bar{\alpha}_{t-1} \cdot (1 - \beta), \quad (4)$$

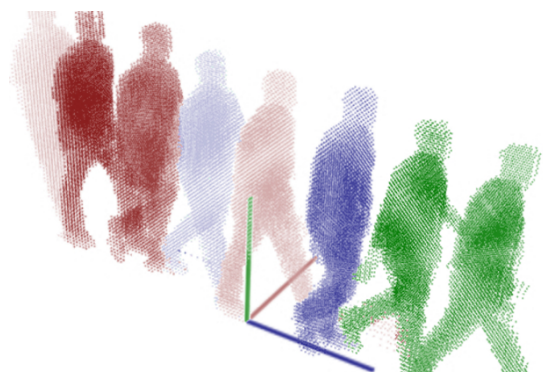


Fig. 4 Example of reconstructed gait sequence, sampled at 2Hz, where each point represents the center of a squared voxel. The time instant is represented by different colors.

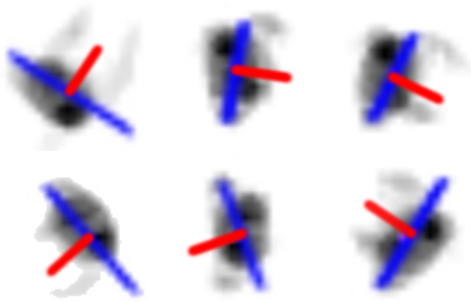


Fig. 5 Displacement vector (red line) of the individual is computed at each time. The principal axis (blue line) is perpendicular to the displacement vector.

where

$$\beta = \frac{\|\mathbf{v}_t\|}{\max_{i=0..t} \{\|\mathbf{v}_i\|\}}. \quad (5)$$

The person may be in some of these states of motion:

- Constant speed. If the speed of the individual is constant (and it is the maximum speed known) then $\beta = 1$.
- Acceleration: if the individual is increasing the walking speed, then $\beta \simeq 1$, so we give more importance to the current path angle. This case is similar to the constant speed.
- Deceleration: if the individual is slowing its motion, the modulus of the displacement vector may not be large enough, causing oscillations in the angle. In this case it would be ideal to give less importance to the current path angle. Therefore, it must be that $\beta \simeq 0$.

A method to decrease over time the denominator in Eq. 5 should be applied if the gait sequence were too large. The whole reconstructed gait sequence can be centered and aligned with respect to the same coordinate system. It is illustrated in Fig. 6.

3.2 Gait Identification

The algorithm step that handles up the gait identification consists of two basic steps, described below.

3.2.1 Descriptor generation

The first one is the generation of the gait descriptor. Our gait descriptor is based on computing the entropy associated to the voxel values of the 3D gait sequence.

Instead of using the Shannon's logarithmic definition of entropy, considered in [1] and [2], we use a definition of entropy based on the exponential behaviour

of information-gain, proposed and justified by *Pal et al.* in [34].

To provide the name of the person at time t , without computing the entire gait sequence, the probability of voxel occupation has to be computed over a sliding temporal window of size L . Given an aligned gait volume V_t^* , let us denote the probability of voxel occupation (p_1) at time t as:

$$p_1(x, y, z) = \frac{1}{L} \sum_{i=t-L}^t V_i^*(x, y, z), \quad (6)$$

and p_0 as:

$$p_0(x, y, z) = 1 - p_1(x, y, z) \quad (7)$$

where L refers to the number of previous volumes on which the exponential entropy will be computed.

So that, the uncertainty associated with a voxel value over the L previous volumes can be computed, at time t , as:

$$\text{GENV}(x, y, z) = m \left(\sum_{k \in \{0,1\}} p_k(x, y, z) e^{(1-p_k(x, y, z))} - 1 \right), \quad (8)$$

where x , y and z are the voxel coordinates, and m is a normalizing constant defined as $m = 1/(e^{1-1/2} - 1)$.

GENV gives an insight into the information content of the gait sequence as the intensity value at voxel location (x, y, z) , which is proportional to its entropy value.

Several candidate signatures can be proposed here. The first approach suggest the use of the whole GENV descriptor as feature vector:

$$\mathcal{G}^{\text{GENV}} = \text{GENV}(x, y, z). \quad (9)$$

However, its dimensionality is proportional to the size of the voxelset, which might be too high (thousands of features). In addition, we can use marginal distributions of the entropy volume to reduce dimensionality without loss of information. GENV and marginal distributions of it are shown in Fig. 6.

According to this, we propose the following candidate signatures:

$$\begin{aligned} - \mathcal{G}_F^{\text{GENV}}(z, y) &= \frac{1}{N_x} \sum_{x=0}^{N_x} \text{GENV}(x, y, z), \\ - \mathcal{G}_S^{\text{GENV}}(z, x) &= \frac{1}{N_y} \sum_{y=0}^{N_y} \text{GENV}(x, y, z), \\ - \mathcal{G}_T^{\text{GENV}}(x, y) &= \frac{1}{N_z} \sum_{z=0}^{N_z} \text{GENV}(x, y, z). \end{aligned}$$

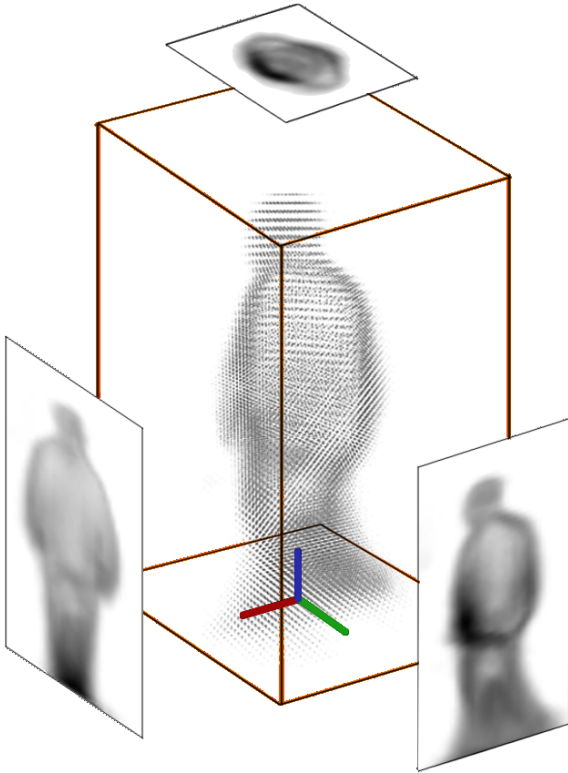


Fig. 6 In the center of the image, a 3D representation of GEnV computed over L 3D-reconstructed and aligned volumes of an individual. Voxels are represented as points. Intensity on gray level represent the entropy value corresponding to that voxel. Marginal distributions of GEnV are also shown.

The definition of the above signatures leads to think that some of them might provide more information than others. Hence, combinations of them can be used, in order to obtain a more discriminative combined signature. Therefore, the combined signature is defined as follows:

$$\mathcal{G}_{F \oplus S \oplus T}^{\text{GEnV}} = \mathcal{G}_F^{\text{GEnV}}, \mathcal{G}_S^{\text{GEnV}}, \mathcal{G}_T^{\text{GEnV}}. \quad (10)$$

where \oplus represents concatenation. So let us denote the set of possible combinations as:

$$\text{view} : \{S, F, T, S \oplus F, S \oplus T, F \oplus T, S \oplus F \oplus T\} \quad (11)$$

Similarly, since our algorithm carries out the alignment of 3D-reconstructed gait volumes, we can also compute GEnI of Bashir et al. [1] on binarized marginal distributions (silhouettes) of V_t^* . So, let us denote the following signatures, which can also be combined:

- $\mathcal{G}_F^{\text{GEnI}}$, GEnI computed on binarized marginal distributions of V^* along the X-axis,
- $\mathcal{G}_S^{\text{GEnI}}$, GEnI computed on binarized marginal distributions of V^* along the Y-axis,

- $\mathcal{G}_T^{\text{GEnI}}$, GEnI computed on binarized marginal distributions of V^* along the Z-axis.

Finally, the total set of proposed signatures is represented as:

$$\mathcal{G}_{\text{view}}^{\text{desc}} \quad (12)$$

where $\text{desc} : \{\text{GEnV}, \text{GEnI}\}$.

3.2.2 Gait classification

The gait signature obtained at time t is the feature vector used for recognition. Each of these feature vectors are assigned to a class label that corresponds to one of the individuals in the database. This idea is well known as multi-class classification system.

We use a Support Vector Machine (SVM) for training and classification. SVM is a partial case of kernel-based methods [5]. It maps feature vectors into a higher-dimensional space using a kernel function and builds an optimal linear discriminating function in this space or an optimal hyper-plane that fits into the training data.

Our recognition algorithm provides the identity of the person as soon as possible, and it does not require the sequence to be split into gait cycles. This makes our method less restrictive compared to other techniques of the literature. At each new 3D volume, a class label is produced, based on the L previous ones.

A majority vote policy over a sliding temporal window of size W is used, in order to reinforce and smooth the results over time, so that the use of this window causes a delay of $L + W$ frames in obtaining the identity of the subject. In the subsequent volumes the system gives a response at the rate of a label per new volume. The majority voting system over a sliding temporal window is shown in Fig. 7.

Before training a SVM model, 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 [20].

4 Experiments and discussion

In this section we describe the datasets we have used, and then we present the experiments conducted to evaluate the proposed gait recognition method and signatures.

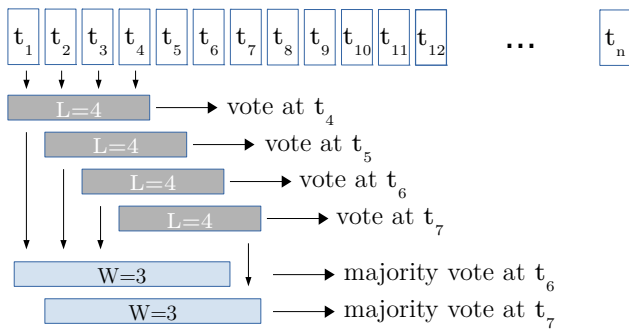


Fig. 7 Majority vote policy over a sliding temporal window. In the example, the size of the signature is set to $L=4$, and the size of the voting window is set to $W=3$. This means that the probability of occupation p_1 (see Eq. 6) is calculated over the previous $L = 4$ volumes.

4.1 Datasets description

In order to be used by our algorithm, the dataset must contain 2D gait images captured by multiple synchronized cameras, which have to be calibrated.

Two synchronized multi-view datasets have been used to perform our experiments, the AVA Multi-View Dataset for Gait Recognition (AVAMVG) ¹ [31] and the Kyushu University 4D Gait Database (KY4D) ² [21].

In AVAMVG, 20 subjects perform 9 walking trajectories in an indoor environment. Each trajectory is recorded by 6 synchronized IEEE-1394 FireFly MV FFMV-03M2C color cameras placed around a room that is crossed by the subjects during the performance, according to the distribution shown in diagram of Fig. 8. The actors enter into the scene from different entry points, which makes this dataset suitable to test view-independent gait recognition methods. Trajectories $\{t_1, \dots, t_3\}$ are straight while $\{t_4, \dots, t_9\}$ are curved.

The video sequences of AVAMVG have a resolution of 640×480 pixels, and they were recorded at a rate of 25 frames per second. An example of this dataset is shown in Fig. 9.

With respect to KY4D Gait Database, 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. The setup is shown in Fig. 10. Although KY4D Gait Database provides sequential 3D models of subjects, we have reconstructed them with the same SfS method and resolution parameters used for getting the AVAMVG models.

¹ Publicly available at: <http://www.uco.es/investiga/grupos/ava/node/41>

² Publicly available at: <http://robotics.ait.kyushuu.ac.jp/researche.php?content=db>

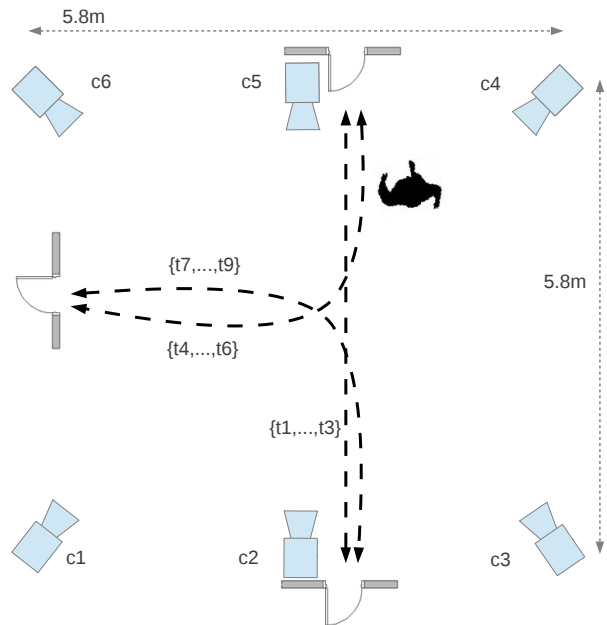


Fig. 8 Workspace setup of AVAMVG dataset, where $\{c_1, \dots, c_6\}$ represent the set of cameras of the multi-view setup and $\{t_1, \dots, t_9\}$ represent the different trajectories performed by each actor.

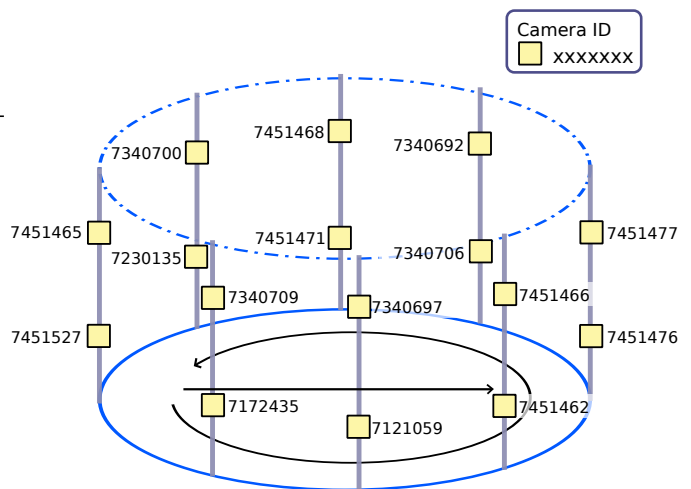


Fig. 10 Experimental setup of KY4D.

As far as we know, there are others well-known multi-camera databases, such as the CMU Motion of Body (MoBo) Database [14] and CASIA Dataset B [42]. However, since these databases do not include information on camera parameters, 3D models of walking people cannot be obtained. Moreover, in the case of CASIA, videos are not synchronized [39]. Therefore, we could not use these databases in the experiments of the present study.



Fig. 9 Example of AVAMVG dataset. People walking in different directions, from multiple points of view.

4.2 Experimental setup

We describe below the different experiments performed to test our gait recognition method, 3D gait descriptors and derived signatures.

- **Experiment A: Baseline.** We adopt the approach described in [1] as baseline, which will be used to benchmark the performance of our proposed method. As we have aligned the 3D volumes, we can compute entropy on binarized marginal distributions of the reconstructed volumes (silhouettes) and test GEnI independently of the trajectory, even with curved paths. For this experiment, we use all trajectories (linear and curved) of the AVAMVG and KY4D databases. As it was described in [1], matching based on minimal distances between GEnI descriptors is carried out. Before matching, a PCA+LDA feature reduction process is performed. We use a leave-one-out cross-validation strategy for both databases.
- **Experiment B: Gait Entropy Volume (GEnV).** The aim of this experiment is to evaluate the performance obtained by using the $\mathcal{G}^{\text{GEnV}}$ signature. Since its dimensionality is very large (proportional to the 3D reconstruction resolution), we also aim to evaluate the impact of the dimensionality reduction (PCA) and the effect of improving the class separability by preserving as much of the class discriminatory information as possible (LDA), on final recognition performance. We use SVM and a leave-one-out cross-validation strategy with all trajectories (linear and curved) for both AVAMVG and KY4D databases.
- **Experiment C: Single signatures.** In this experiment we try to determine the most discriminative GEnV-based single signature. We also aim to compare the obtained performance by using single signatures based on the GEnV, with the obtained performance by using single signatures based on GEnI. As in the experiment B, we also apply PCA+LDA in order to achieve the best data representation and class separability simultaneously. We use a SVM and leave-one-out cross-validation strategy with all trajectories (linear and curved) of both AVAMVG and KY4D databases.
- **Experiment D: Combined signatures.** The aim of this experiment is to find out the most discriminative combined GEnV signature. As in the Experiment C, we also aim to compare the obtained performance by using combined GEnV signatures with the performance obtained by using combined GEnI signatures. We apply PCA+LDA, SVM, and leave-one-out cross-validation strategy with all trajectories (linear and curved) for both databases.
- **Experiment E: Majority vote policy over a sliding temporal window.** This experiment focuses on testing the effect of the majority vote policy over a sliding temporal window on the final performance. We select the most discriminative signatures that were found out in the experiment D and we test them on both databases.
- **Experiment F: Number of cameras.** This experiment focuses on determining the number of cameras which are required in the reconstruction step to achieve a good performance.
- **Experiment G: Training and testing with different camera setup.** This experiment aims to test the effect of training and testing with different subsets of cameras.
- **Experiment H: Training on straight paths and testing on curved paths.** In this experiment, we

Table 1 Results of GEnI baseline approach (Experiment A). We report the quantitative results obtained with the approach described in [1]. From the fourth to the fifth columns are indicated the dimensionality reduction techniques applied. The sixth column indicates the best PCA energy found for each descriptor. Each row corresponds to a database.

Database	Signature	Without Dim.Red.	PCA	PCA+LDA	ϵ
AVAMVG	GEnI [1] (baseline)	79.23%	74.06%	90.27%	0.95
KY4D	GEnI [1] (baseline)	87.97%	84.82%	87.67%	0.95

Table 2 Results obtained by using GEnV as gait signature (Experiment B). From the third to the fourth columns are indicated the dimensionality reduction techniques applied. The fifth column indicates the best PCA energy found for each descriptor. Each row corresponds to a database. The size of the sliding temporal window for majority voting is set to $W = 1$.

Database	Signature	PCA	PCA+LDA	ϵ
AVAMVG	$\mathcal{G}^{\text{GEnV}}$	77.63%	70.10%	0.95
KY4D	$\mathcal{G}^{\text{GEnV}}$	78.46%	94.52%	0.95

Table 3 Recognition results obtained by using a single signature based on GEnV compared to the results obtained by using a single signature based on GEnI, on the AVAMVG gait database (Experiment C). From the third to the fourth columns are indicated the dimensionality reduction techniques applied. The fifth column indicates the best PCA energy found for each descriptor. The size of the sliding temporal window for majority voting is set to $W = 1$.

Signature	Without Dim. Red.	PCA	PCA+LDA	ϵ
$\mathcal{G}_S^{\text{GEnV}}$	97.19%	97.03%	96.84%	0.95
$\mathcal{G}_S^{\text{GEnI}}$	95.34%	94.81%	92.63%	0.85
$\mathcal{G}_F^{\text{GEnV}}$	94.27%	94.15%	91.94%	0.95
$\mathcal{G}_F^{\text{GEnI}}$	90.78%	89.47%	87.06%	0.85
$\mathcal{G}_T^{\text{GEnV}}$	70.73%	65.86%	58.41%	0.90
$\mathcal{G}_T^{\text{GEnI}}$	60.50%	57.27%	50.52%	0.85

use linear trajectories for training and curved trajectories for testing. We compare the best signatures found in the previous experiments with some related methods that are able to recognize people on curved trajectories.

Before conducting the experiments, we need to determine the value of several basic parameters of our method. Thus, considering the 3D reconstruction step, the first parameter to be determined is the voxel size. We consider that a voxel size of $0.27 \times 10^{-4} m^3$ is enough to get a detailed 3D human reconstruction.

The average corporal volume for humans is $66.4l$ measured by the water displacement method in 521 people aged between 17 and 51 years [11]. Using a voxel size of $2.7cl$, the number of voxels belonging to a person in the 3D volume should be about 2459. Thus, for $\theta > 1 \times 10^3$ (See Section 3.1.3) the system should be able to detect both children and adults.

The number of L volumes where entropy is computed is set to $L = 20$, because with a rate of 25 volumes per seconds, this value roughly matches with the average length of a gait cycle.

Lastly, regarding the dimensionality reduction, we tested a range from $\epsilon = 0.75$ to $\epsilon = 0.99$ for the percentage of variance (energy contained in the components signal) that PCA should retain.

4.3 Results

We present the results of the experiments that were described in the previous section.

We have adopted the approach described in [1] as baseline. Table 1 shows its performance for each database. As we have 3D aligned gait volumes, the GEnI can be computed over binarized marginal distributions of the aligned volumes, i.e silhouettes. We perform matching between GEnI features and we use a leave-one-out cross-validation strategy. In the case of the AVAMVG dataset, each fold is composed by a tuple formed by a set of 20 gait sequences (one sequence per actor) for testing, and by the remaining eight trajectories of each actor for training, i.e. 8×20 sequences for training and 20 sequences for test. For the KY4D gait dataset, each fold is composed by 42 sequences for testing (one se-

Table 4 Recognition results obtained by using a single signature based on GEnV compared to the results obtained by using a single signature based on GEnI, using the KY4D gait database (Experiment C). From the third to the fourth columns are indicated the dimensionality reduction techniques applied. The fifth column indicates the best PCA energy found for each descriptor. The size of the sliding temporal window for majority voting is set to $W = 1$.

Signature	Without Dim. Red.	PCA	PCA+LDA	ϵ
$\mathcal{G}_S^{\text{GEnV}}$	90.92%	92.74%	92.50%	0.95
$\mathcal{G}_S^{\text{GEnI}}$	88.90%	89.53%	87.88%	0.90
$\mathcal{G}_F^{\text{GEnV}}$	84.42%	84.75%	84.40%	0.99
$\mathcal{G}_F^{\text{GEnI}}$	77.25%	78.34%	76.06%	0.90
$\mathcal{G}_T^{\text{GEnV}}$	78.42%	78.68%	76.35%	0.95
$\mathcal{G}_T^{\text{GEnI}}$	76.43%	74.55%	73.60%	0.85

Table 5 Comparative results obtained with combined GEnV signatures and combined GEnI signatures using the AVAMVG dataset (Experiment D). From the second to the third columns are indicated the dimensionality reduction techniques applied. The fourth column indicates the best PCA energy found for each descriptor. The size of the sliding temporal window for majority voting is set to $W = 1$.

Signature	PCA	PCA+LDA	ϵ
$\mathcal{G}_{S\oplus F}^{\text{GEnV}}$	97.94%	97.95%	0.95
$\mathcal{G}_{S\oplus F}^{\text{GEnI}}$	96.82%	95.44%	0.85
$\mathcal{G}_{S\oplus T}^{\text{GEnV}}$	96.51%	95.99%	0.95
$\mathcal{G}_{S\oplus T}^{\text{GEnI}}$	96.03%	94.63%	0.80
$\mathcal{G}_{F\oplus T}^{\text{GEnV}}$	92.74%	92.82%	0.99
$\mathcal{G}_{F\oplus T}^{\text{GEnI}}$	90.04%	90.27%	0.90
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GEnV}}$	97.20%	97.16%	0.95
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GEnI}}$	97.52%	97.29%	0.90

quence per actor) and by the remaining five sequences of each actor (i.e. 42×5 sequences) for training. To make the choice of SVM parameters independent of the sequence test data, we cross-validate the SVM parameters on the training set. As can be seen, the obtained accuracy is 90.27% for the AVAMVG dataset and 87.67% for the KY4D dataset. It corresponds to Experiment A (see Sec. 4.2).

We show in Table 2 the performance of $\mathcal{G}^{\text{GEnV}}$ as gait signature. Due to the high dimensionality, which is proportional to the 3D gait volume resolution (about 74×10^3 features in a volume with base of $1m^2$ and height of $2m$, with voxel size of $2.7cl$), we have applied dimensionality reduction techniques. In this experiment, the number of features in AVAMVG after applying dimensionality reduction is 101 with PCA, and $C - 1 = 19$ with PCA+LDA where C is the number of classes in our multi-class system. On the other hand, for the KY4D database, the number of features is 670 after PCA, and 41 after PCA+LDA.

Table 6 Comparative results obtained with combined GEnV signatures and combined GEnI signatures using the KY4D dataset (Experiment D). From the second to the third columns are indicated the dimensionality reduction techniques applied. The fourth column indicates the best PCA energy found for each descriptor. The size of the sliding temporal window for majority voting is set to $W = 1$.

Signature	PCA	PCA+LDA	ϵ
$\mathcal{G}_{S\oplus F}^{\text{GEnV}}$	94.23%	94.04%	0.95
$\mathcal{G}_{S\oplus F}^{\text{GEnI}}$	92.40%	91.18%	0.90
$\mathcal{G}_{S\oplus T}^{\text{GEnV}}$	94.34%	93.76%	0.95
$\mathcal{G}_{S\oplus T}^{\text{GEnI}}$	93.91%	92.18%	0.90
$\mathcal{G}_{F\oplus T}^{\text{GEnV}}$	90.62%	89.47%	0.99
$\mathcal{G}_{F\oplus T}^{\text{GEnI}}$	87.58%	86.12%	0.95
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GEnV}}$	95.00%	95.13%	0.95
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GEnI}}$	94.28%	93.17%	0.90

We use SVM with Radial Basis Functions (RBF) and a leave-one-out cross-validation strategy. To make the choice of SVM parameters independent of the sequence test data, we cross-validate the SVM parameters on the training set. In this experiment, for the sake of simplicity, we disabled the sliding temporal window for majority voting ($W = 1$). As can be seen, by using this signature on the AVAMVG gait dataset, the accuracy on subject identification is rather far from baseline. However, with the KY4D gait database, the accuracy is higher than the baseline. It corresponds to Experiment B (see Sec. 4.2).

In Table 3 and Table 4 we compare the obtained performance by using a single signature (not combined) based on GEnV, with the obtained performance by using a single GEnI signature, for the AVAMVG dataset and the KY4D database respectively. It corresponds to experiment C (see Sec. 4.2). This experiment was carried out without applying the sliding temporal window of majority vote ($W = 1$). In order to achieve the best data representation and class separability simulta-

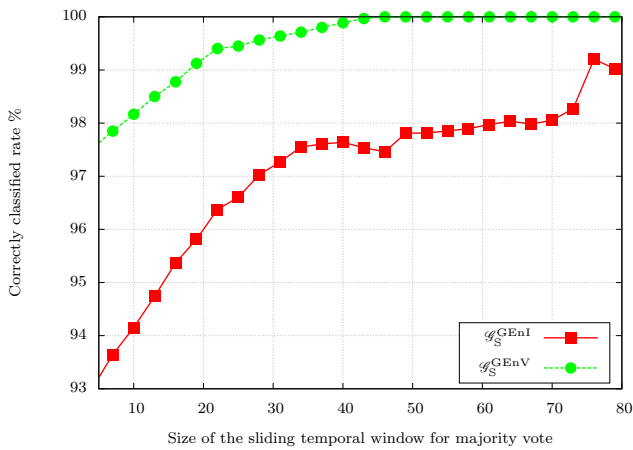


Fig. 11 Performance of $\mathcal{G}_S^{\text{GENV}}$ and $\mathcal{G}_S^{\text{GENI}}$ on AVAMVG for different lengths for the majority voting window.

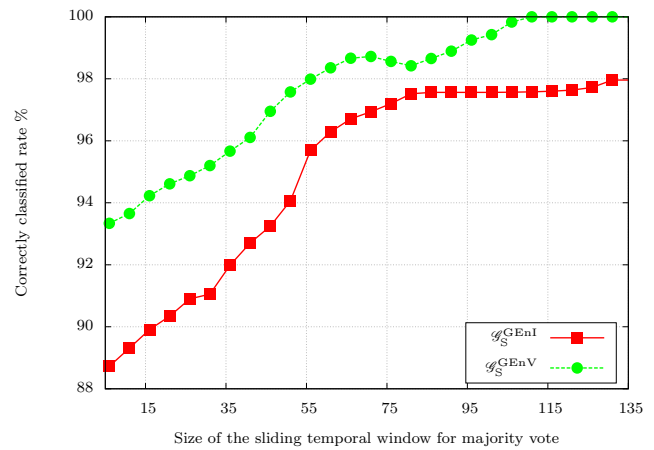


Fig. 13 Performance of $\mathcal{G}_S^{\text{GENV}}$ and $\mathcal{G}_S^{\text{GENI}}$ on KY4D for different lengths for the majority voting window.

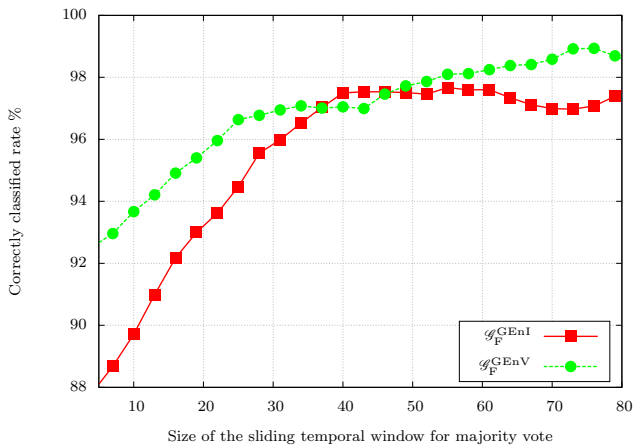


Fig. 12 Performance of $\mathcal{G}_F^{\text{GENV}}$ and $\mathcal{G}_F^{\text{GENI}}$ on AVAMVG for different lengths for the majority voting window.

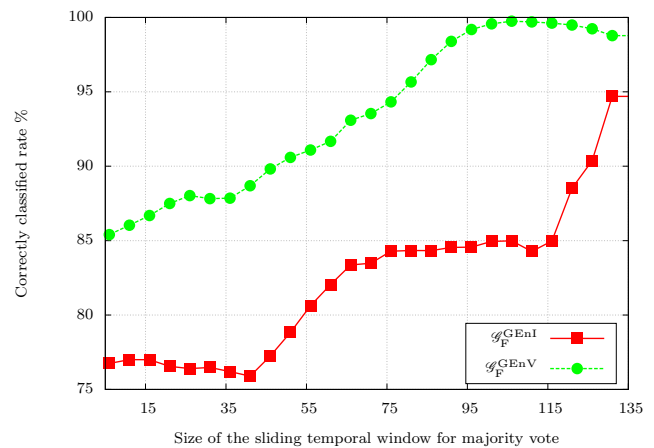


Fig. 14 Performance of $\mathcal{G}_F^{\text{GENV}}$ and $\mathcal{G}_F^{\text{GENI}}$ on KY4D for different lengths for the majority voting window.

neously, we apply PCA+LDA to the training and test data. A SVM-RBF with leave-one-out cross-validation strategy is used for training and classification. As we may observe, the $\mathcal{G}_S^{\text{GENV}}$ signature improves results fairly well compared to baseline on both gait databases.

The obtained accuracy by using LDA is very similar than the obtained accuracy by just using PCA. However, the number of features with LDA is significantly reduced compared with PCA. If the system can be trained off-line, LDA allows SVM to handle a feature space of low dimensionality, and the identity of the individual can be given in less time.

Experiment E (see Sec. 4.2) focuses on testing the effect, on the final performance, of the majority voting policy over a sliding temporal window. The size of the sliding temporal window for majority voting is limited by the number of available gait signatures for each sequence. Fig. 11, 12, 13 and 14 show how the accuracy for selected signatures (using PCA+LDA) increases with

respect to the size of the sliding temporal window for majority voting. As can be seen, the use of the sliding temporal window for majority voting increases the performance significantly, achieving the perfect identification in some cases by using single signatures based on GENV.

We have hypothesized that the combination of signatures will lead to a better performance, because the concept of 3D gait volume enables us to get a larger amount of dynamical and structural information of gait.

In order to find out the most discriminative combined signature on both gait databases, we have carried out the Experiment D (see Sec. 4.2). In Table 5 and 6 we report the performance obtained by using combined GENV signatures, compared with the performance obtained by using combined GENI signatures. We use SVM-RBF with a leave-one-out cross-validation strategy.

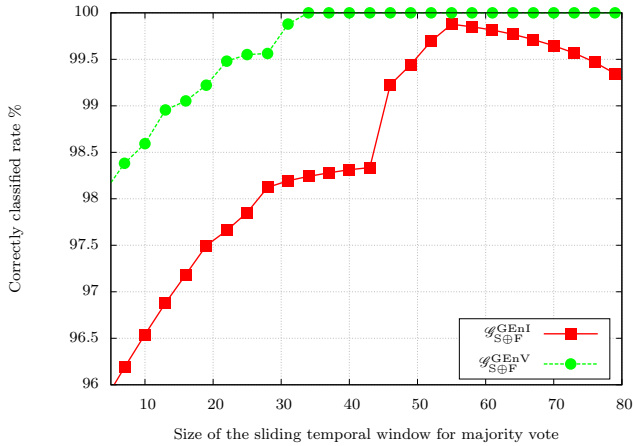


Fig. 15 Performance of $\mathcal{G}_{S\oplus F}^{GEnV}$ and $\mathcal{G}_{S\oplus F}^{GEnI}$ on AVAMVG for different lengths for the majority voting window.

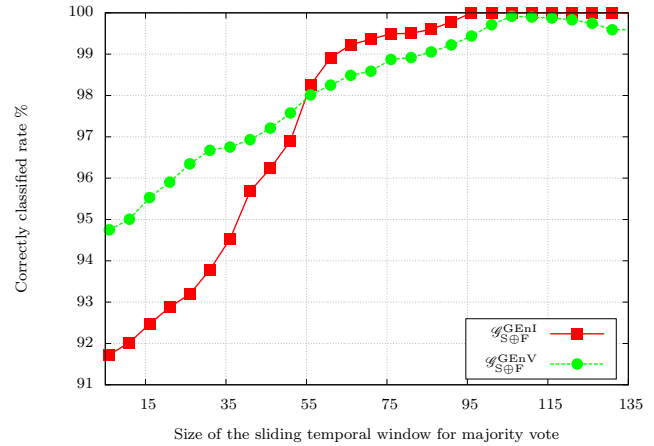


Fig. 18 Performance of $\mathcal{G}_{S\oplus F}^{GEnV}$ and $\mathcal{G}_{S\oplus F}^{GEnI}$ on KY for different lengths for the majority voting window.

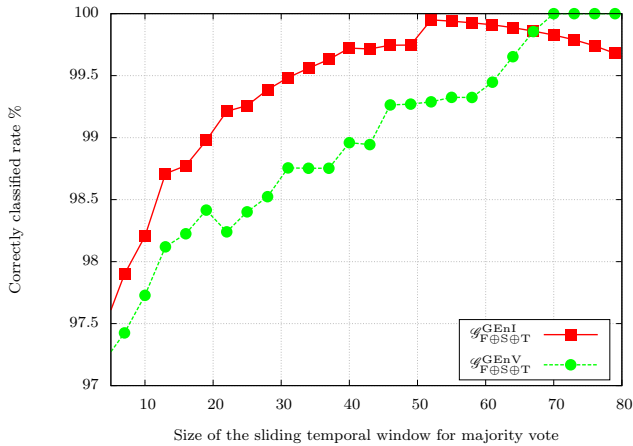


Fig. 16 Performance of $\mathcal{G}_{S\oplus F\oplus T}^{GEnV}$ and $\mathcal{G}_{S\oplus F\oplus T}^{GEnI}$ on AVAMVG for different lengths for the majority voting window.

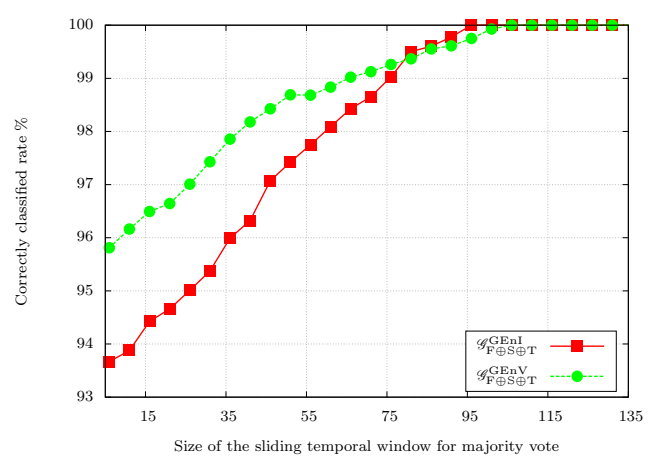


Fig. 19 Performance of $\mathcal{G}_{S\oplus F\oplus T}^{GEnV}$ and $\mathcal{G}_{S\oplus F\oplus T}^{GEnI}$ on KY for different lengths for the majority voting window.

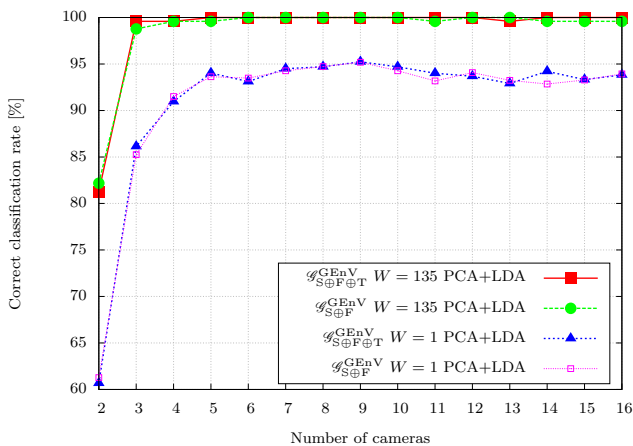


Fig. 17 Performance of $\mathcal{G}_{S\oplus F\oplus T}^{GEnV}$ and $\mathcal{G}_{S\oplus F}^{GEnV}$ on KY4D database for an increasing number of cameras.

As can be seen, both $\mathcal{G}_{S\oplus F}^{GEnV}$ and $\mathcal{G}_{S\oplus F\oplus T}^{GEnV}$ signatures provide good results on AVAMVG and KY4D gait

databases. However the accuracy's confidence intervals at 95% obtained for both measures are overlapped. For this reason, we can not conclude that one is better than the other from the point of view of accuracy. On the other hand, from the point of view of computational cost, the cost of $\mathcal{G}_{S\oplus F\oplus T}^{GEnV}$ is greater than the computational cost of $\mathcal{G}_{S\oplus F}^{GEnV}$. Therefore, for a system in operation, $\mathcal{G}_{S\oplus F}^{GEnV}$ could be the most appropriate signature.

As part of experiment E, we have tested the effect, on the final performance, of the majority voting policy over a sliding temporal window and combined signatures. Figs. 15, 16, 18 and 19 show the accuracy for the selected signatures (using PCA+LDA) with respect to the size of the sliding temporal window for majority voting. As we may observe, the use of the sliding temporal window for majority voting increases the performance significantly. The accuracy reaches 100% for all the signatures based on GEnV on both dataset.

Table 7 Comparative of recognition results on AVAMVG dataset (Experiment F). Training on trajectories $t1+t2+t3$ (straight paths). Each row corresponds to a different method. The second and third columns indicate the tested trajectory. The size of the sliding temporal window for majority voting is set to $W = 80$.

Method	Test Curve t4	Test Curve t7	AVG
Castro <i>et al.</i> [12]	95%	95%	95%
Iwashita <i>et al.</i> [21]	35.14%	37.71%	36,42%
$\mathcal{G}_{S\oplus F}^{\text{GenV}}$ (PCA+LDA)	97.61%	98.41%	98,01%
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GenV}}$ (PCA+LDA)	97.62%	98.84%	98,23%
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GenI}}$ (PCA+LDA)	96.20%	98,42%	97,31%

Table 8 Comparative of recognition results on KY4D dataset (Experiment F). Training on trajectories $1+2+3+4$ (straight paths). Each row corresponds to a different method. The second and third columns indicate the tested trajectory. The size of the sliding temporal window for majority voting is set to $W = 135$.

Method	Test Curve 1	Test Curve 2	AVG
Iwashita <i>et al.</i> [21]	61.90%	71.40%	66,65%
$\mathcal{G}_{S\oplus F}^{\text{GenV}}$ (PCA+LDA)	44.92%	42.08%	43,50%
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GenV}}$ (PCA+LDA)	57.83%	79.28%	68,55%
$\mathcal{G}_{S\oplus F\oplus T}^{\text{GenI}}$ (PCA+LDA)	55.26%	87.31%	71,28%

In order to determine the number of cameras that should be employed and its effect on the performance, we have designed a leave-one-out cross validation experiment. As in the other experiments, to make the choice of SVM parameters independent of the sequence test data, we cross-validate the SVM parameters on the training set. We selected the signature configuration that achieved the best performance in the previous experiments ($\mathcal{G}_{S\oplus F}^{\text{GenV}}$ and $\mathcal{G}_{S\oplus F\oplus T}^{\text{GenV}}$) and then we tested them with a set of KY4D models which have been reconstructed using a number of cameras in the range 2 to 16. As can be seen in Fig. 17, with just 3 calibrated cameras, our method is able to correctly classify up to 95% of individuals, independently of the path, even with curved trajectories.

To perform experiment G, we have reconstructed all the gait sequences of KY4D from two subsets of cameras. The subset A is composed by cameras {07451471, 07121059, 07451527, 07451476}, whereas the subset B is composed by cameras {07340706, 07172435, 07230135, 07451462}. Then, we designed a leave-one-out cross-validation experiment, but using the subset A for training and subset B for testing. We obtained 90.57% of recognition rate with $\mathcal{G}_{S\oplus F}^{\text{GenV}}$ and 94.26% with $\mathcal{G}_{S\oplus F\oplus T}^{\text{GenV}}$.

The results shown in Tables 7 and 8 correspond to experiment H (see Sec. 4.2). For the corresponding tests on AVAMVG, we have trained on linear trajectories {t1, t2, t3} (all in the same set), and tested on curved trajectories {t4, t7} (see corresponding columns). For the

KY4D dataset, we have trained on linear trajectories {1,2,3,4} (all in the same set) and tested on curved trajectories {Curve 1 and Curve 2} (see Sec. 4.1 for more details). We have compared our best signatures with other related methods that are able to recognize people on curved trajectories, such as [12] and [21]. The size of the sliding temporal window for majority voting of our method is set to $W = 80$ for AVAMVG and $W = 135$ for KY4D. The method of Castro *et al.* [12] only can be tested on the AVAMVG dataset, because it need color or grayscale images, and the KY4D dataset only provides binary silhouettes and 3D models.

We have notice a low performance of the method presented in [21] when it is trained with straight paths and tested with curves of the AVAMVG dataset. We think it may be due to the low number of cameras of AVAMVG (6 cameras), in contrast with the number of cameras of the KY4D gait dataset (16 cameras). On the one hand, it seems reasonable to think that fewer cameras leads us to obtain 3D reconstructions of lower precision. Besides the number of cameras, the quality of silhouettes is also a factor that must be considered. On the other hand, in the AVAMVG dataset, depending on the viewpoint and performed trajectory, people appear at diverse scales, even showing partially occluded body parts, which adversely affects to the performance of [21].

If the classifier is trained with just linear trajectories, both $\mathcal{G}_{S\oplus F\oplus T}^{\text{GenI}}$ and $\mathcal{G}_{S\oplus F\oplus T}^{\text{GenV}}$ signatures provide

good results. However, training with both linear and curved trajectories leads to get better performance, as we can see in results of Experiments D and E, and in that case, GEnV based signatures provide the best performance. We think when a subject is walking along a curved path, the gait pattern is consequently modified, and signatures based on GEnV are able to better capture dynamical 3D information than signatures based on GEnI, as was probed in Experiment D and E. For this reason, on unconstrained paths, we suggest the use of $\mathcal{G}_{S\oplus F}^{\text{GEnV}}$ or $\mathcal{G}_{S\oplus F\oplus T}^{\text{GEnV}}$ signatures.

5 Conclusions

This paper has proposed a method to recognize walking humans independently of the viewpoint, even with curved trajectories. Our method achieves a good recognition rate on unconstrained paths, in contrast to others view-independent approaches which restrict the view change to a few angles on straight trajectories.

A new gait descriptor, called GEnV has also been proposed. GEnV focuses on capturing 3D dynamic information of walking humans through the concept of entropy, applied on volumetric reconstructions. The use of volumetric reconstructions allows more information to be analysed in contrast to other related works, which only compute the gait descriptors from 2D images, discarding a significant part of 3D dynamical information of the gait. Several signatures based on GEnV have also been proposed in order to get better recognition rate.

We have tested the classification performance for each proposed gait signature on the AVA Multi-View Gait Dataset (AVAMVG) [31] and on the Kyushu University 4D Gait Database (KY4D) [21]. The experimental results show that GEnV based signatures such as $\mathcal{G}_{S\oplus F}^{\text{GEnV}}$ and $\mathcal{G}_{S\oplus F\oplus T}^{\text{GEnV}}$ are the most reliable signatures for using with our gait recognition method on unconstrained path, providing good results in both gait databases. Finally, by using a majority voting policy on a sliding temporal window, the system is able to reach a perfect identification of individuals for both datasets.

Acknowledgements This work has been developed with the support of the Research Projects called TIN2012-32952 and BROCA both financed by Science and Technology Ministry of Spain and FEDER.

Thanks to Kurazume and Iwashita Laboratory, Graduate School of Information Science and Electrical Engineering, Kyushu University, for providing the KY4D gait database.

References

1. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: *Crime Detection and Prevention (ICDP 2009)*, 3rd International Conference on, pp. 1–6 (2009)
2. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. *Pattern Recogn. Lett.* **31**(13), 2052–2060 (2010)
3. Bodor, R., Drenner, A., Fehr, D., Masoud, O., Papanikolopoulos, N.: View-independent human motion classification using image-based reconstruction. *Image and Vision Computing* **27**(8), 1194–1206 (2009)
4. Bolle, R., Pankanti, S.: *Biometrics, Personal Identification in Networked Society*. Kluwer Academic Publishers, Norwell, MA, USA (1998)
5. Burges, C.J.: A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* **2**, 121–167 (1998)
6. Chattopadhyay, P., Roy, A., Sural, S., Mukhopadhyay, J.: Pose depth volume extraction from rgb-d streams for frontal gait recognition. *Journal of Visual Communication and Image Representation* **25**(1), 53 – 63 (2014)
7. Cheng, M.H., Ho, M.F., Huang, C.L.: Gait analysis for human identification through manifold learning and hmm. *Pattern Recognition* **41**(8), 2541 – 2553 (2008)
8. Choudhury, S.D., Tjahjedi, T.: Silhouette-based gait recognition using procrustes shape analysis and elliptic fourier descriptors. *Pattern Recognition* **45**(9), 3414 – 3426 (2012)
9. Cutting, J.E., Kozlowski, L.T.: Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society* **9**, 353–356 (1977)
10. Díaz-Más, L., Muñoz Salinas, R., Madrid-Cuevas, F.J., Medina-Carnicer, R.: Shape from silhouette using dempster-shafer theory. *Pattern Recognition* **43**(6), 2119 – 2131 (2010)
11. Drinkwater, D., Ross, W.: *Anthropometric fractionation of body mass*. In: *Kinanthropometry*, vol. 2, pp. 178–189. University Park Press, Baltimore (1980)
12. F.M.Castro, M.J., Medina-Carnicer, R.: Pyramidal fisher motion for multiview gait recognition. In: *Proceedings of the International Conference on Pattern Recognition (2014)*
13. Goffredo, M., Bouchrika, I., Carter, J., Nixon, M.: Self-calibrating view-invariant gait biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* **40**(4), 997–1008 (2010)
14. Gross, R., Shi, J.: The cmu motion of body (mobo) database. *Tech. Rep. CMU-RI-TR-01-18*, Robotics Institute, Pittsburgh, PA (2001)
15. Han, J., Bhanu, B.: Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(2), 316–322 (2006)
16. Han, J., Bhanu, B., Roy-Chowdhury, A.: A study on view-insensitive gait recognition. In: *IEEE International Conference on Image Processing, 2005. ICIP 2005*, vol. 3, pp. III–297–300 (2005)
17. Hofmann, M., Bachmann, S., Rigoll, G.: 2.5d gait biometrics using the depth gradient histogram energy image. In: *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pp. 399–403 (2012)
18. Hong, S., Lee, H., Kim, E.: Automatic gait recognition using width vector mean. In: *4th IEEE Conference on Industrial Electronics and Applications*, pp. 647–650 (2009)
19. Horprasert, T., Harwood, D., Davis, L.S.: A statistical approach for real-time robust background subtraction and shadow detection. In: *Proc. IEEE ICCV*, pp. 1–19 (1999)

20. [Huang, P., Harris, C., Nixon, M.: Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering* **13**\(4\), 359 – 366 \(1999\)](#)
21. [Iwashita, Y., Ogawara, K., Kurazume, R.: Identification of people walking along curved trajectories. *Pattern Recognition Letters* **48**\(0\), 60 – 69 \(2014\)](#)
22. [Jean, F., Albu, A.B., Bergevin, R.: Towards view-invariant gait modeling: Computing view-normalized body part trajectories. *Pattern Recognition* **42**\(11\), 2936–2949 \(2009\)](#)
23. [Jeong, S., Kim, T.h., Cho, J.: Gait recognition using description of shape synthesized by planar homography. *The Journal of Supercomputing* **65**\(1\), 122–135 \(2013\)](#)
24. [Kale, A., Chowdhury, A., Chellappa, R.: Towards a view invariant gait recognition algorithm. In: *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*, pp. 143–150 \(2003\)](#)
25. [Kale, A., Cuntoor, N., Yegnanarayana, B., Rajagopalan, A., Chellappa, R.: Gait analysis for human identification. In: *Audio- and Video-Based Biometric Person Authentication*, vol. 2688, pp. 706–714. Springer Berlin Heidelberg \(2003\)](#)
26. [Kusakunniran, W., Wu, Q., Li, H., Zhang, J.: Multiple views gait recognition using view transformation model based on optimized gait energy image. In: *2009 IEEE 12th International Conference on Computer Vision Workshops \(ICCV Workshops\)*, pp. 1058–1064 \(2009\)](#)
27. [Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Gait recognition under various viewing angles based on correlated motion regression. *IEEE Transactions on Circuits and Systems for Video Technology* **22**\(6\), 966–980 \(2012\)](#)
28. [Laurentini, A.: The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**\(2\), 150–162 \(1994\)](#)
29. [Lee, C.P., Tan, A.W., Tan, S.C.: Gait recognition via optimally interpolated deformable contours. *Pattern Recognition Letters* **34**\(6\), 663 – 669 \(2013\)](#)
30. [Liu, N., Tan, Y.P.: View invariant gait recognition. In: *2010 IEEE International Conference on Acoustics Speech and Signal Processing \(ICASSP\)*, pp. 1410–1413 \(2010\)](#)
31. [López-Fernández, D., Madrid-Cuevas, F.J., Carmona-Poyato, A., Marín-Jiménez, M.J., Muñoz Salinas, R.: The AVA Multi-View Dataset for Gait Recognition \(AVAMVG\). In: *2nd Workshop Activity Monitoring by Multiple Distributed Sensing, AMMDS. ICPR 2014, Lecture Notes in Computer Science*, vol. 8703. Springer Berlin Heidelberg \(2014\)](#)
32. [Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: A. Leonardis, H. Bischof, A. Pinz \(eds.\) *Computer Vision - ECCV 2006*, no. 3953 in *Lecture Notes in Computer Science*, pp. 151–163. Springer Berlin Heidelberg \(2006\)](#)
33. [Mowbray, S.D., Nixon, M.S.: Automatic gait recognition via fourier descriptors of deformable objects. In: *Audio Visual Biometric Person Authentication*, pp. 566–573. Springer \(2003\)](#)
34. [Pal, N., Pal, S.: Entropy: a new definition and its applications. *Systems, Man and Cybernetics, IEEE Transactions on* **21**\(5\), 1260–1270 \(1991\)](#)
35. [Rougier, C., Auvinet, E., Meunier, J., Mignotte, M., de Guise, J.: Depth energy image for gait symmetry quantification. In: *IEEE International Conference on Engineering in Medicine and Biology Society*, pp. 5136 –5139 \(2011\)](#)
36. [Shakhnarovich, G., Lee, L., Darrell, T.: Integrated face and gait recognition from multiple views. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I-439–I-446 vol.1 \(2001\)](#)
37. [Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Gait energy volumes and frontal gait recognition using depth images. In: A.K. Jain, A. Ross, S. Prabhakar, J. Kim \(eds.\) *International Joint Conference on Biometrics, IJCB*, pp. 1–6. IEEE \(2011\)](#)
38. [Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: The backfilled GEI - a cross-capture modality gait feature for frontal and side-view gait recognition. In: *2012 International Conference on Digital Image Computing Techniques and Applications \(DICTA\)*, pp. 1–8 \(2012\)](#)
39. [Wang, L., Geng, X.: *Behavioral Biometrics For Human Identification: Intelligent Applications*, 1 edn. Medical Information Science Reference \(2009\)](#)
40. [Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**\(12\), 1505 – 1518 \(2003\)](#)
41. [Yu, S., Tan, D., Tan, T.: Modelling the effect of view angle variation on appearance-based gait recognition. In: P. Narayanan, S. Nayar, H.Y. Shum \(eds.\) *Computer Vision - ACCV 2006, Lecture Notes in Computer Science*, vol. 3851, pp. 807–816. Springer Berlin Heidelberg \(2006\)](#)
42. [Zheng, S., Zhang, J., Huang, K., He, R., Tan, T.: Robust view transformation model for gait recognition. In: *Image Processing \(ICIP\), 2011 18th IEEE International Conference on*, pp. 2073–2076 \(2011\)](#)