Universidad de Córdoba

Departamento de Informática y Análisis Numérico

TESIS DOCTORAL

CONTRIBUTIONS TO GAIT RECOGNITION USING MULTIPLE VIEWS

David López Fernández

Córdoba, marzo de 2016

TITULO: *CONTRIBUTIONS TO GAIT RECOGNITION USING MULTIPLE VIEWS*

AUTOR: *David López Fernández*

Universidad de Córdoba

Departamento de Informática y Análisis Numérico

TESIS DOCTORAL

CONTRIBUTIONS TO GAIT RECOGNITION USING MULTIPLE VIEWS

David López Fernández

Directores: Dr. Francisco José Madrid Cuevas
Dr. Ángel Carmona Poyato

Córdoba, marzo de 2016

TÍTULO DE LA TESIS: CONTRIBUTIONS TO GAIT RECOGNITION USING MULTIPLE VIEWS

DOCTORANDO/A: David López Fernández

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

El doctorando ha realizado su trabajo de una forma totalmente satisfactoria. Además de realizar todas las actividades académicas requeridas por el programa de Doctorado en el que está matriculado, ha realizado un completo estudio del arte sobre el tópico de la tesis.

El doctorando ha generado una completa base de datos para aplicar métodos multivista al reconocimiento del paso en secuencias de vídeo que actualmente está siendo muy solicitada en el campo de investigación a nivel internacional.

Además, como fruto de su trabajo de investigación, el doctorando ha propuesto varias formas de abordar el problema de la identificación de personas mediante el análisis de su forma de andar usando información fusionada a partir de múltiples vistas, avanzando en el campo de conocimiento al permitir que este reconocimiento se pueda realizar de manera independiente al punto de vista y sin estar restringido a trayectorias rectas.

Como resultado de su trabajo de investigación el doctorando ha publicado tres trabajos en revistas JCR de primer nivel en el campo. También ha participado en dos congresos internaciones con sendas comunicaciones.

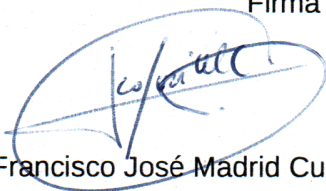A continuación se detalla la lista de contribuciones derivadas de esta tesis:

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. Machine Vision and Applications, 26(7):1079-1094, 2015. ISSN 1432-1769. doi: 10.1007/s00138-015-0707-9.

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, R. Muñoz-Salinas and R. Medina-Carnicer. Viewpoint-independent gait recognition through morphological descriptions of 3D human reconstructions. Image and Vision Computing, 48-49:1-13, 2016. ISSN 0262-8856. doi: 10.1016/j.imavis.2016.01.003.

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas and R. Medina-Carnicer. A new approach for multi-view gait recognition on unconstrained paths. Journal of Visual Communication and Image Representation, 38:396-406, 2016. ISSN 1047-3203. doi: 10.1016/j.jvcir.2016.03.020.

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, and Rafael Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In Activity Monitoring by Multiple Distributed Sensing, Lecture Notes in Computer Science, pages 26-39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3.

- D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Multi-view gait recognition on curved trajectories. In Proceedings of the 9th International Conference on Distributed Smart Cameras, ICDSC'15, pages 116-121, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3681-9. doi: 10.1145/2789116.2789122.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 28 de marzo de 2016

Firma del/de los director/es

Fdo.:Dr. Francisco José Madrid Cuevas   Fdo.: Dr. Ángel Carmona Poyato.

Contributions to Gait Recognition Using Multiple Views

Doctorando: David López Fernández.
Directores: Dr. Francisco José Madrid Cuevas.
                  Dr. Ángel Carmona Poyato.

Resumen en castellano:

Esta tesis se centra en la identificación de personas a través de la forma de caminar. El problema del reconocimiento del paso ha sido tratado mediante diferentes enfoques, en los dominios 2D y 3D, y usando una o varias vistas. Sin embargo, la dependencia con respecto al punto de vista, y por tanto de la trayectoria del sujeto al caminar sigue siendo aún un problema abierto.

Se propone hacer frente al problema de la dependencia con respecto a la trayectoria por medio de reconstrucciones 3D de sujetos caminando. El uso de reconstrucciones varias ventajas que cabe destacar. En primer lugar, permite explotar una mayor cantidad de información en contraste con los métodos que extraen los descriptores de la marcha a partir de imágenes, en el dominio 2D. En segundo lugar, las reconstrucciones 3D pueden ser alineadas a lo largo de la marcha como si el sujeto hubiera caminado en una cinta andadora, proporcionando así una forma de analizar el paso independientemente de la trayectoria seguida.

Este trabajo propone tres enfoques para resolver el problema de la dependencia a la vista:

1. Mediante la utilización de reconstrucciones volumétricas alineadas.

2. Mediante el uso de reconstrucciones volumétricas no alineadas.

3. Sin usar reconstrucciones.

Se proponen además tres tipos de descriptores. El primero se centra en describir el paso mediante análisis morfológico de los volúmenes 3D alineados. El segundo hace uso del concepto de entropía de la información para describir la dinámica del paso humano. El tercero persigue capturar la dinámica de una forma invariante a rotación, lo cual lo hace especialmente interesante para

ser aplicado tanto en trayectorias curvas como rectas, incluyendo cambios de dirección.

Estos enfoques han sido probados sobre dos bases de datos públicas. Ambas están específicamente diseñadas para tratar el problema de la dependencia con respecto al punto de vista, y por tanto de la dependencia con respecto a la trayectoria. Los resultados experimentales muestran que para el enfoque basado en reconstrucciones volumétricas alineadas, el descriptor del paso basado en entropía consigue los mejores resultados, en comparación con métodos estrechamente relacionados del Estado del Arte actual. No obstante, el descriptor invariante a rotación consigue una tasa de reconocimiento que supera a los métodos actuales sin requerir la etapa previa de alineamiento de las reconstrucciones 3D.

Contributions to Gait Recognition Using Multiple Views

Doctorando: David López Fernández.
Directores: Dr. Francisco José Madrid Cuevas.
  Dr. Ángel Carmona Poyato.

Abstract:

This thesis focuses on identifying people by the way they walk. The problem of gait recognition has been addressed using different approaches, both in the 2D and 3D domains, and using one or multiple views. However, the dependence on camera viewpoint, and therefore the dependence on the trajectory of motion still remains an open problem.

This work proposes to address the problem of dependence on the trajectory through the use of 3D reconstructions of walking subjects. The use of 3D models have several advantages that are worth mentioning. First, by the use of 3D reconstructions it is possible to exploit a greater amount of information in contrast to methods that extract descriptors from 2D images. Second, the 3D reconstructions can be aligned along the way as if the subject had walked on a treadmill, thus providing a way to recognize people regardless the path.

Three approaches are proposed in order to address the problem of the dependence with respect the trajectory:

1. Using aligned 3D reconstructions of walking humans.

2. By using volumetric unaligned volumetric reconstructions of walking humans.

3. Without using 3D reconstructions.

Three gait descriptors are also proposed. The first focuses on describing gait by means of morphological analysis of 3D aligned volumes. The second makes use of the concept of information entropy to describe the dynamics of human gait. The third aims to capture the dynamics of gait in a rotation invariant way, which makes it interesting for recognize people walking on both straight and curves path, and regardless direction changes.

These approaches have been tested on two public databases. Both databases are specifically designed to address the problem of dependence on the viewpoint, and therefore the dependence on the trajectory. Experimental results show that for the approach based on aligned volumetric reconstructions, the entropy-based gait descriptor achieved the best results compared to other closely related methods of the state-of-art. However, the rotation invariant gait descriptor achieves a recognition rate that overcomes the compared state-of-art methods, without requiring the prior step of alignment of the 3D gait reconstructions.

A mis padres.

Agradecimientos

La memoria de esta tesis es la culminación de un periodo fantástico de mi vida académica y personal. Sin lugar a dudas, el hecho de que esta tesis haya sido posible está estrechamente relacionado con un gran número de personas a las que no puedo más que ofrecerles mi más sincero agradecimiento.

He de agradecer profundamente a la persona que apostó por mí y me invitó a participar del apasionante mundo de la investigación, el doctor Francisco José Madrid, por sus valiosas enseñanzas, que no pueden ser debidamente ponderadas en estas líneas, y su amistad sincera. También al doctor Ángel Carmona, por su disponibilidad y supervisión científica.

A Dr. Rafael Medina, por la oportunidad de trabajar en el grupo de investigación Aplicaciones de la Visión Artificial de la Universidad de Córdoba, y por velar para que este trabajo pudiera llevarse a cabo en unas condiciones óptimas de financiación.

A los profesores Rafael Muñoz, Manuel Jesús Marín, Enrique Yeguas y Nicolás Luis Fernández, por todo lo aprendido de ellos, que no ha sido poco. A mis compañeros, que siempre mostraron su apoyo y su amistad, y a los que tanto aprecio. Nunca podría olvidar este periodo de trabajo junto a Sergio Garrido, Víctor Manuel Mondéjar, Manuel Ignacio López y Eusebio Jesús Aguilera. Con ellos he trabajado, me he divertido, he caído y me he vuelto a levantar.

A mis padres, que día a día me regalaron su presente para que yo pudiera alcanzar mi futuro. Si algo soy en esta vida es gracias a ellos. Papá y mamá, esta tesis es más vuestra que mía, y no puedo más que daros las gracias, una y otra vez. Gracias. También a mis hermanos, Antonio José y Miguel, y a mis hermanas Carmen María y Adriana, por su ayuda y su apoyo incondicional. Y por último a Mercedes, por compartir todos y cada uno de los buenos y malos momentos conmigo. Por estar a mi lado, por empujarme y dar cada paso conmigo, por animarme, y por iluminarme cuando algo se torna oscuro. Gracias a todos por vuestro amor, por vuestra compañía, y por ser mi familia.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Generally, biometric is a field of technology that uses automated methods for identifying or verifying a person based on anatomical or behavioural traits. The term biometrics is derived from a Greek word "Bio" means life and "metrics" means measure [1]. Thus, biometrics is the science and technology of measuring and analysing biological data. Biometrics is classified into two categories, anatomical and behavioural characteristics. Anatomical characteristics are related to the shape of the body. Examples include, but are not limited to fingerprints [2], iris [3], hand geometry [4], vein [5], DNA [6], Face [7] and Ear Recognition [8]. Behavioural characteristics are related to the pattern of behaviour of an individual and pay attention to the actions of a person. Examples include, but are not limited to voice [9], typing rhythm [10] and gait [11].

1.1   Gait recognition

The term "gait" refers to the walking pattern of a person. The walking pattern is cyclic in nature and may be composed of many gait cycles, where each gait cycle consists of at least two steps. What is especially interesting of gait is that each individual describes an unique gait pattern, which means it can be used as a biometric indicator [12]. Thus, gait as a biometric feature for identification has received in recent years, 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 [13].

The first gait recognition approach was developed by Niyogi and Adelson on a small gait database in 1994 [14]. Subsequently, the HumanID program, sponsored by Defense Advanced Research Projects Agency (DARPA) [15], assists greatly in advancing automatic gait recognition. Spurred by the HumanID program, many international famous universities and research institutes, such as the University of Southampton, the Massachusetts Institute of Technology (MIT), Carnegie Mellon University (CMU) and the Institute of Automation Chinese Academy of Sciences (CASIA), have made a lot of researches on gait recognition.

Consequently, many research papers have been published in recent years tackling the problem of human gait recognition using different sources of data like inertial sensors [16, 17], foot pressure [18], infrared images [19], depth images [20] or the traditional images [21, 22, 23]. For example, in [13, 11] we can find two surveys on this problem summarizing some of the most popular approaches. Some of them use explicit geometrical models of human bodies, whereas others use only image features.

However, the gait recognition performance of most methods is significantly affected by changes in various covariate conditions such as clothing [24], camera viewpoint [25, 26], load carrying [27], and walking speed [28]. According to camera viewpoint, the previous work can be categorized into two approaches: view-dependent and view-independent approaches. The 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 [29]. 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. 1.1 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.

Previous studies on gait recognition have been also classified into two categories: model-based approaches and model-free approaches [30]. The model-based methods represent gait using the parameters of a body configuration model which is estimated over time, whereas model-free 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

2
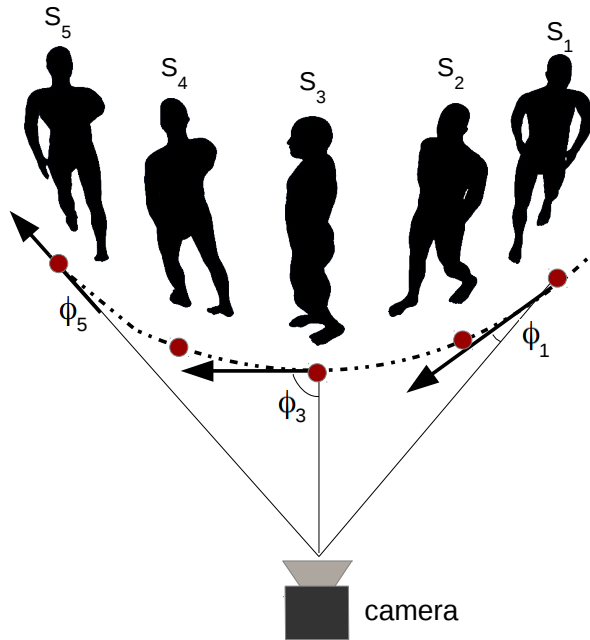
Figure 1.1: In a curved path, the observation angle between the walking direction of the subject and optical axis of the camera is gradually changed, which affects the silhouette appearance.

is lost. While most of model-free 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.

1.2 Applications

Some potential applications of gait recognition are access control [31, 13] in special or restricted areas (e.g. military bases and governmental facilities) and smart video surveillance where subjects do not know they are being monitored (e.g. bank offices) [32, 33]. 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).

(a) Gatwick auto-boarding biometric gate.

(b) Medical isolation zone.

(c) Vault door at the Winona National Bank.

(d) Military facilities.

Figure 1.2: Example of situations where gait recognition can be applied.

1.3 Aims of this work

The main goal of this dissertation is obtaining view-independent gait recognition algorithms which can recognize people independently of the trajectory of motion. In this way, we tackle such a challenge from the standpoint of 3D reconstructions.

The main hypothesis is that the use of 3D reconstructions can be a valid approach to address the challenge of gait recognition on unconstrained paths. The use of 3D reconstructions has several advantages that are worth mentioning. On the one side, 3D volumes of walking people can be aligned along their way, and therefore well-known 2D gait descriptors can be applied, independently of the trajectory of motion, to extract gait features from rendered projections of the aligned 3D-reconstructed volumes. On the other side, the use of volumetric information allows more information to be analysed in contrast to other methods of the literature which only compute

4

gait descriptors from silhouettes or 2D images, and therefore new 3D gait descriptors can be proposed.

1.4   Contributions

The research work carried out during the development of this dissertation has produced several articles that have been published in well-respected peer-reviewed journals and conferences.

- D. López-Fernández, F.J Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, R. Muñoz-Salinas, and R. Medina-Carnicer. Viewpoint-independent gait recognition through morphological descriptions of 3d human reconstructions. *Image and Vision Computing*, 48-49:1-13, 2016. ISSN 0262-8856. doi: 10.1016/j.imavis.2016.01.003. url: http://www.sciencedirect.com/science/article/pii/S0262885616300014

  This study presents a multi-view gait recognition method that allows curved trajectories on unconstrained paths in indoor environments. The recognition is based on volumetric analysis of the human gait, to exploit most of the 3D information enclosed in it. Two new morphological gait descriptors are presented. The first consists in computing three Cover by Rectangles (CR) descriptors [34] on the front, side, and top projections of the aligned 3D gait volumes. The second is the *Cover by Cubes*, which is defined as the union of all the cubes with the largest size that can fit into a volume belonging to the person. The proposed approach is experimentally validated on two gait databases. The results show that this new approach is able to identify people walking on curved paths.

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications*, 26(7):1079-1094, 2015. ISSN 1432-1769. doi: 10.1007/s00138-015-0707-9. url: http://link.springer.com/article/10.1007%2Fs00138-015-0707-9

  This work presents an efficient multi-view gait-recognition method that allows curved trajectories on completely unconstrained paths for indoor

environments. The 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. The proposed approach is experimentally validated on two gait databases, achieving promising results in the problem of gait recognition on unconstrained paths.

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. A new approach for multi-view gait recognition on unconstrained paths. *Journal of Visual Communication and Image Representation*, 38:396-406, 2016. ISSN 1047-3203. doi: 10.1016/j.jvcir.2016.03.020. url: `http://www.sciencedirect.com/science/article/pii/S1047320316300232`

  This paper proposes a new approach for multi-view gait recognition, which focuses on recognizing people walking on unconstrained (curved and straight) paths. A new rotation invariant gait descriptor is presented, which is based on 3D angular analysis of the movement of the subject. This approach is experimentally validated on two gait databases, 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.

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, and Rafael Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In *Activity Monitoring by Multiple Distributed Sensing, Lecture Notes in Computer Science*, pages 26-39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3. url: `http://link.springer.com/chapter/10.1007%2F978-3-319-13323-2_3`

  In this paper, we introduce the AVA Multi-View Dataset for Gait Recognition (AVAMVG). This paper provides a comprehensive description of the AVAMVG database, including the studio environment and camera setup, among other information. To validate the dataset, in this paper are extended some appearance-based 2D gait recognition methods to work with multi-view 3D data, obtaining very encouraging results.

- D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Multi-view gait recogni-

tion on curved trajectories. In *Proceedings of the 9th International Conference on Distributed Smart Cameras, ICDSC'15*, pages 116-121, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3681-9. doi: 10.1145/2789116.2789122. url: `http://dl.acm.org/citation.cfm?id=2789122`

This paper presents a method to recognize walking humans independently of the viewpoint and regardless direction changes on curved trajectories. This approach is based on 3D angular analysis of the movement of the walking humans. In this paper, it is also described a new rotation invariant gait descriptor which is computed directly from silhouettes obtained from multiple views, by using calibrating information instead of 3D reconstructions.

1.5   Structure of this document

This section explains how this document is organized. Chapter 2 contains an extensive literature review. This chapter is divided into three sections, view-dependent approaches, view-independent approaches and closely related works. A review on public databases is presented in Chapter 3. This chapter also describes a new dataset created in order to test multi-view gait recognition approaches. Chapter 4 addresses the topic of 3D reconstruction and gait alignment.

Chapter 5 studies the implementation of several morphological descriptors and their application on aligned 3D sequences of walking people. Chapter 6 presents an approach which focuses on capturing 3D dynamical information of gait by means of the concept of entropy information applied on volumetric reconstructions of humans. A rotation invariant gait descriptor which can be computed directly on 3D gait sequences without the previous step of alignment is presented in Chapter 7. Finally, Chapter 8 presents the main conclusions and future research lines derived from this dissertation.

Chapter 2

Background and Literature Review

The previous work can be categorized into two approaches: view-dependent and view-independent approaches. The third section of this chapter describes in detail some methods which are closely related with our proposed approaches. These methods are also used to compare and evaluate the performance of our proposals.

2.1 View-dependent approaches

One of the earliest model-free and view-dependent approaches can be seen in [35], where the width of the outer contour of the binarized silhouette from a side view is used to build a descriptor which contains both structural features and dynamic aspects of gait. Feature vectors derived from binary silhouettes have been also used to train Hidden Markov Models [36]. The contours of silhouettes have been used directly [37, 38], and through their Fourier descriptors [39, 40].

In addition, the authors of [41] present a gait recognition method which analyses the shape of the silhouette using Procrustes Shape Analysis and Elliptic Fourier Descriptors. The Gait Energy Image (GEI) descriptor is introduced by Han and Bhanu in [21], which is the average of all silhouette images for a single gait cycle. Silhouette images are also used in Lam *et*

al. [22] to generate the gait flow image (GFI). Liu *et al.* [42], to improve the gait recognition performance, propose the computation of Histogram of Oriented Gradients (HOG) [43] from popular gait descriptors as the GEI and the Chrono-Gait Image (CGI) [44].

In [45], the authors try to find the minimum number of gait cycles needed to carry out a successful recognition by using the GEI descriptor. Martín-Félez and Xiang [46] [47], using GEI as the basic gait descriptor, propose a new ranking model for gait recognition. This new formulation of the problem allows to leverage training data from different datasets, thus, improving the recognition performance. Lai *et al.* [48] proposed a novel discriminant subspace learning method (Sparse Bilinear Discriminant Analysis) that extends methods based on matrix-representation discriminant analysis to sparse cases, obtaining competitive results on gait recognition.

Based on the concept of GEI, Depth Energy Image (DEI) was defined in [49], which is simply the average of the depth silhouettes taken along a gait cycle, over the front view. GEI is also extended in [20] to consider depth information from the side view, by means of a new feature called Depth Gradient Histogram Energy Image (DGHEI). Depth information is also used in Chattopadhyay *et al.* [50] to address the problem of occlusion in frontal gait recognition. The Gait Entropy Image (GEnI) was presented in the work of Bashir *et al* [51]. GEnI encodes in a single image, the randomness of pixel values in the silhouette images over a complete gait cycle.

Analysis by morphological size distributions was proposed in [34]. In this work, video cameras are placed in hallways to capture longer sequences from the front view of walkers rather than the side view, which results in more gait cycles per gait sequence. Despite the high recognition rate, the main drawback of this model-free approach is the dependence with respect to the viewpoint. To obtain a gait representation directly from silhouettes, the authors proposed the use of a morphological descriptor, called Cover by Rectangles, which is defined as the union of all the largest rectangles that can fit inside a silhouette.

The Gait Energy Volume (GEV), a binary voxel-discretized volume which is spatially aligned and averaged over a gait cycle, is presented in [52]. The authors apply GEV on partial reconstructions obtained with depth sensors from the front view of the individual. An extended work from GEV that combines the frontal-view depth gait image and side-view 2D gait silhouette by means of a back-filling technique is presented in [53]. A spatio temporal

representation based on point clouds in a spherical coordinate space was proposed in [54], where frontal 3D point clouds of humans obtained with stereo cameras are used.

In [55], a 3D approximation of a Visual Hull (VH) [56] is used to design a multi-modal recognition approach. Although a VH model is computed, a gait recognition scheme based on silhouette analysis is applied, which restricts a large amount of discriminant information because the recognition is based on single view silhouette analysis, instead of analyse 3D information. Seely *et al.* [57] use 3D volumetric data to synthesise 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.

Ariyanto *et al.* [58] propose a model-fitting algorithm, correlation filters and dynamic programming to extract gait kinematics features. They use a structural model including articulated cylinders with 3D Degrees of Freedom (DoF) which are fitted to a visual hull shape to model the human lower legs. In [59], 3D data collected from a projector-camera system is used to fit 3D body models and reconstruct synthetic poses in a gait cycle.

2.2   View-independent approaches

Appearance changes due to viewing angle changes cause difficulties for most of the model-free 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 [25, 60].

Approaches of the first category are represented by Shakhnarovich *et al.* [55], Bodor *et al.* [61] and Iwashita *et al.* [26]. In the work of Shakhnarovich *et al.* [55], a polyhedral and surface-mapped 3D approximation of the visual hull [56] (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 3D information because the recognition is based on single view silhouette analysis, instead

of exploiting the 3D models.

Bodor *et al.* [61] applies image-based rendering on a 3D VH model to extract gait features under a required viewing angle. 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 the work of Iwashita *et al.* [26], 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 3D reconstructions, 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, as in the above two works, despite 3D volumes are used, descriptors are extracted from 2D images, so that, the amount of used information is restricted.

Approaches of the second category extract gait features which are invariant to viewing angle changes. A method to generate a canonical view of gait from any arbitrary view is described in [62]. The main disadvantage of this method is that the synthesis of a canonical view is only feasible from a limited number of initial views. The performance is significantly dropped down when the angle between image plane and sagittal plane is large.

In [63], a method based on homography to compute view-normalized trajectories of body parts obtained from monocular video sequences was proposed, but this method only works properly 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 [64], however this method seems not to be applicable when the person is walking nearly parallel to the optical axis. In [65] 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 [66]. 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 [67] is proposed the use of motion descriptors based on dense tracklets (i.e. short-term point trajectories). This method is able to recognize people in curved trajectories with promising results. In [68] were proposed and evaluated diverse strategies to improve tracklet-based gait recognition systems. Two sets of tracklet-based features were combined with audio features in [69], to evaluate how the fusion of audio and visual features can help in the challenging task of people identification based on their gait.

Zhao *et al.* [70] present a multi-camera approach for gait tracking and recognition. The video sequences are used as input, and then a human 3D model is set up. The lengths of key segments are extracted as static parameters, and the motion trajectories of lower limbs are used as dynamic features. A skeletal 3D model is also used in the work of Kastaniotis *et al.* [71], which presents a framework for pose-based gait recognition and identification, as well as gender recognition.

The approaches of the third category are based 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 [72], where LDA-subspaces are learned to extract discriminative information from gait features under each viewing angle.

A View Transformation Model (VTM) was introduced by [73] to transform gait features from different views into the same view. The method of Makihara *et al.* [73] creates a VTM based on frequency-domain gait features, obtained through Fourier Transformation. To improve the performance of this method, Kusakunniran *et al.* [74] 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 [25]. However, this method cannot deal 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 [25]: (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., 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.

2.3   Closely related work

As it was described in Section 2.1, Gait Entropy Image (GEnI) [51] 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. This representation is also less sensitive to changes in covariate conditions such as carrying and clothing.

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

$$\text{GEnI}(x, y) = - \sum_{k=1}^{K} p_k(x, y) \log_2 p_k(x, y), \qquad (2.1)$$

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 [75], 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. In this work 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 [51].

Figure 2.1: Several examples of GEnI, computed over a gait cycle. The gray level represents the entropy value in a pixel (x, y). As can be seen, legs and arms have high gait entropy value, whereas static areas as torso have low values of entropy.

These approaches [51, 75] have some drawbacks that are worth mentioning. First, these methods are view-dependent and therefore the individuals cannot walk freely in the scene. Second, 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. 2.1 shows the GEnI, computed over several gait cycles.

Another work closely related with our approaches is that of Barnich *et al.* [34]. In this work, the authors propose the analysis by morphological size distributions, by means of the Cover By Rectangles descriptor, proposed in [76]. Considering a silhouette S, the Cover by Rectangles, denoted by $\mathrm{C}(S)$ is defined as the union of all the largest rectangles that can fit inside of S. Figure 2.2 shows a graphical description of $\mathrm{C}(S)$. This union is unique and the cover $\mathrm{C}(S)$ has the following useful properties:

- The elements of the set overlap each other introducing redundancy (i.e. robustness).

- Each element (rectangle) of $\mathrm{C}(S)$ cover at least one pixel that belongs to no other rectangle.

- When displayed in the frame, the union of all rectangles reconstructs
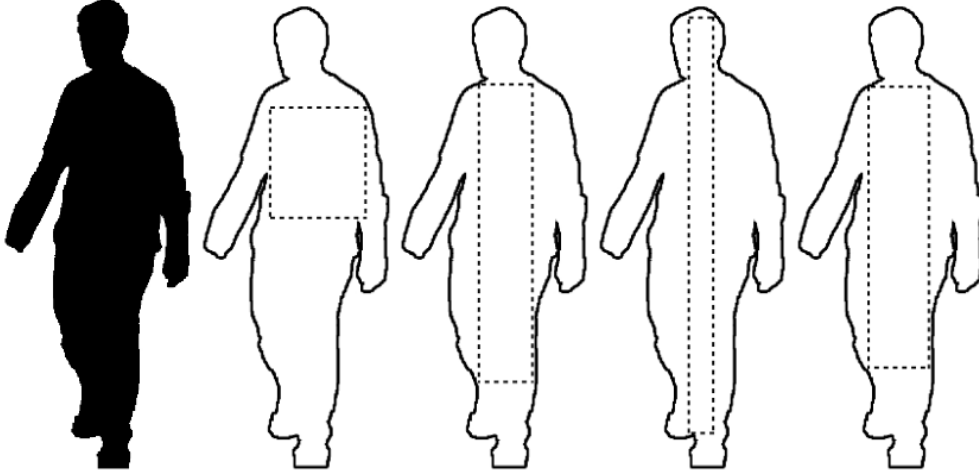
15

Figure 2.2: The Cover by Rectangles, C(S), is the union of all the largest rectangles that can be wedged inside of the silhouette.

S so that no information is ever lost.

The number of largest rectangles that will fit inside a binary silhouette can be very high (more than a thousand). It is thus impractical to use all the rectangles directly as a set of features. In order to find a more compact representation, the authors propose to operate in a size distribution density, as shown in Figure 2.3. As can be seen, the largest number of rectangles containing a given pixel is to be found inside the torso, and the tallest rectangles pass through both the legs and the head.

From a formal point of view, let α be the cardinality of C(S), *i.e.* $\alpha = \#C(S)$. The rectangles of $C(S)$ are indexed with a parameter d, so that $R_d(d = 1, ..., \alpha)$ are the rectangles of $C(S)$. The width and height of R_d are, respectively, denoted by w_d y h_d; they are upper-bounded by w^{max} and h^{max}: $\forall d, w_d \leq w^{max}$ and $h_d \leq h^{max}$. In order to build histograms, the widths and heights of the rectangles R_d are partitioned into M bins $B_W(i)$ and N bins $B_H(j)$

$$B^W(i) =]i\frac{w^{max}}{M}, (i+1)\frac{w^{max}}{M}],$$

$$B^H(j) =]j\frac{h^{max}}{N}, (j+1)\frac{h^{max}}{N}],$$

where $i = 0, ..., M-1$ y $j = 0, ..., N-1$.

Following the above notations, the histogram $hist^W(i)$ of the normal-
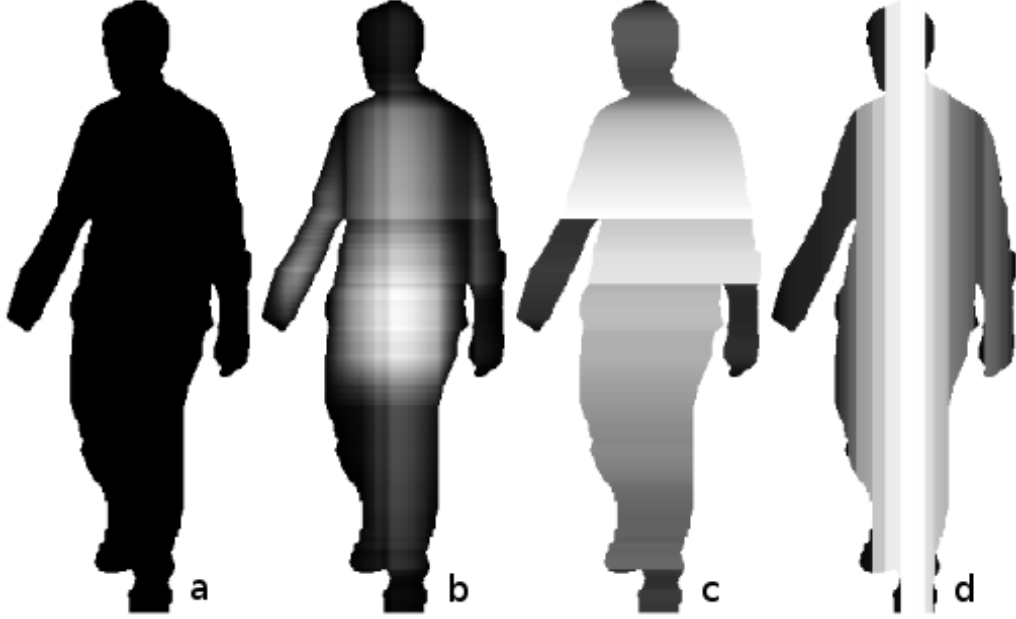
16

Figure 2.3: Size distributions based on the description provided by the cover $C(S)$. The image (a) represents a silhouette of a walking person, observed from a frontal viewpoint. A gray level of pixel (x, y) in images (b), (c) and (d) displays respectively the density of rectangles, the width of the widest rectangle and the height of the tallest rectangle where all these rectangles contain pixel (x, y).

ized widths is defined as

$$\text{hist}^W(i) = \tfrac{1}{\alpha} \#\{R_d | w_d \in B^W(i)\},$$

the histogram of the normalized heights similarly as

$$\text{hist}^H(j) = \tfrac{1}{\alpha} \#\{R_d | h_d \in B^H(i)\},$$

and the two-dimensional histogram $\text{hist}^{WxH}(i, j)$ as

$$\text{hist}^{WxH}(i, j) = \tfrac{1}{\alpha} \#\{R_d | w_d \in B^W(i), h_d \in B^H(j)\}.$$

These histograms are normalized respect to the largest rectangle of $C(S)$. In a continuous space, they would be scale invariant. Of the three histograms ($\text{hist}^W(i)$, $\text{hist}^H(j)$, $\text{hist}^{WxH}(i, j)$), the last one best describes S. However, its dimensionality is proportional to the product of the numbers of bins (MxN), which might be too high. In order to solve this tractabil-

ity issue, it is introduced the composite histogram $\text{hist}^{W+H}(k)$, with $k = 0, ..., M + N - 1$ defined as the strict concatenation of $\text{hist}t^W(i)$ and $\text{hist}^H(j)$ (marginal distributions). The authors introduce the following gait signature

$$\mathscr{G}_{\text{CR}}^{WxH}(i, j, t) =$$
$$\text{hist}^{WxH}(i, j, t - (L - 1)), ..., \text{hist}^{WxH}(i, j, t - 1), \text{hist}^{WxH}(i, j, t),$$

and a shortened version

$$\mathscr{G}_{\text{CR}}^{W+H}(k, t) = \text{hist}^{W+H}(k, t - (L - 1)), ..., \text{hist}^{W+H}(k, t - 1), \text{hist}^{W+H}(k, t),$$

which consists of n-uples of L consecutive histograms at time t. Video cameras are placed in hallways to capture longer sequences from the front view of walkers rather than the side view, which results in more gait cycles per gait sequence. Despite the high recognition rate, the main drawback of this model-free approach is the dependence with respect to the viewpoint, which restricts the movement of the individuals to just straight paths.

As it was described in Section 2.1, Seely *et al.* [57] use 3D volumetric data to synthesise silhouettes from a fixed viewpoint relative to the subject. In this work, the resulting silhouettes are then passed to a standard 2D gait analysis technique, such as the average silhouette. The sequences are collected from a multi-biometric tunnel, where the subjects just walk straight.

For the analysis, silhouettes are synthesized from a side-on, front-on and top-down orthogonal viewpoints, by taking the union of the voxels along X, Y and Z-axis, respectively, where the X-axis spans left to right, Y-axis spans front to back and Z-axis spans from the top to the bottom. The average silhouette is calculated in a similar manner to that of Han and Bhanu [21], where the centre of mass $C_i = (C_{i,x}, C_{i,y})$ is found for each frame i. The average silhouette is the found by summing the centre of mass aligned silhouettes

$$A(x, y) = \frac{1}{M}\sum_{i=0}^{M-1} J_i(x - C_{i,x}, y - C_{i,y}), \tag{2.2}$$

where A is the average silhouette, M is the number of frames in the gait cycle and J is the synthesized image of frame i. So that, this method requires the sequence to be split into gait cycles. The derived average silhouette is scale normalised so that it is 64 pixels high, whilst preserving the aspect ratio. That average silhouette is then treated as the feature vector. As well as the above works, this method do not allow the subjects to walk

freely in the scene because the view of the virtual silhouettes depends on the trajectory of the 3D volumes.

Chapter 3

Gait Datasets

Most current gait recognition methods require gait sequences captured from a single view, namely, from the side view or from the front view of a walking person [40, 77, 78, 79, 80, 75, 34, 21]. Hence, there are many existing databases which capture the gait sequences from a single view. However, new challenges in the topic of gait recognition, such as achieving the independence from the camera point of view, usually require multi-view datasets. In fact, articles related to multi-view and cross-view gait recognition have been increasingly published [26, 25, 81, 82, 66, 74, 61, 73, 67].

First, some of the existing multi-view datasets were recorded in controlled conditions, and in some cases, they made use of a treadmill [83, 84, 85]. An inherent problem associated with walking on a treadmill is that the human gait is not as natural as it should be, the gait speed is usually constant, and the subjects cannot turn right or left. They are not representative of human gait in a real world. Secondly, for other multi-view datasets, calibration information is not provided or cameras are not synchronized, e.g. CASIA Dataset B [86]. Some of the gait recognition methodologies require camera calibration to deal with 3D information.

In addition, there are not many multi-view datasets specifically designed for gait. Some of them are designed for action recognition, and therefore they do not contain gait sequences of enough length as to contain several gait cycles, because gait is a subset of them.

For this reason, we have created a new indoor dataset to test gait recognition algorithms, the "AVA Multi-View Dataset for Gait Recognition"

(AVAMVG) [87]. In Section 3.3 we briefly describe the camera setup, the database content, and the preprocessing steps carried out in order to further increase the applicability of the database. This dataset can be applied in workspaces where subjects cannot show the face or use the fingerprint, and even they have to wear special clothing, e.g. a laboratory. Furthermore, in this dataset people appear walking along both straight and curved paths, which makes this dataset suitable to test methods like [26], which are able to cope with curved paths. The cameras have been calibrated and therefore, methods based on 3D information can use this dataset to test, which is free only for research purposes.

Section 3.2 briefly describes the "Kyushu University 4D gait dataset", a recently and publicly available multi-view dataset which has also been used to test our gait recognition approaches, because it contains videos of people walking on both curved and straight trajectories.

3.1   Current datasets for gait recognition

The chronologic order of appearance of the different human action video datasets runs parallel to the challenges that the scientific community has been considering to face the problem of automatic and visual gait recognition. From this point of view, datasets can be divided into two groups: datasets for single-view gait recognition and datasets for multi-view gait recognition. Besides, the datasets can be divided in two subcategories: indoor and outdoor datasets.

Regarding single view datasets, indoor gait sequences are provided by the OU-ISIR Biometric Database [83]. OU-ISIR Treadmill Database is composed by four datasets, called A, B, C and D. Dataset A is composed of gait sequences of 34 subjects from side view with speed variation. The dataset B is composed of gait sequences of 68 subjects from side view with clothes variation up to 32 combinations. OU-ISIR gait dataset D contains 370 gait sequences of 185 subjects observed from the side view. The dataset D focuses on the gait fluctuations over a number of periods. The OU-ISIR gait dataset C focuses on view variation, but it is currently under preparation, and as far as we know, information about this has not been released yet. The OU-ISIR Large Population Dataset contains 4016 subjects, walking on straight paths. It is distributed in a form of silhouette sequences registered and size-normalized to 88 by 128 pixels size. The subjects are recorded by

just two cameras, and calibration information is not provided. Samples of still images (captured color images) are shown in Figure 3.1.

Figure 3.1: Samples from OU-ISIR Large Population Dataset.

In contrast with OU-ISIR, the first outdoor single-view dataset was from the Visual Computing Group of the UCSD (University of California, San Diego) [88]. The UCSD gait database includes six subjects with seven image sequences of each, from the side view. In addition to UCSD gait database, one of the most used outdoor datasets for single-view gait recognition is the USF HumanID database [15]. This database consists of 122 persons walking in elliptical paths in front of the camera.

Other outdoor walking sequences are provided in CASIA database, from the Center for Biometrics and Security Research of the Institute of Automation of the Chinese Academy of Sciences. CASIA Gait Database is composed by three datasets, one indoor and the other two outdoor. The indoor dataset can be also considered as a multi-view dataset, and therefore it will be discussed later. The outdoors datasets are named as Dataset A and Dataset C (infrared dataset), and they are described below.

Dataset A [89] includes 20 persons. Each person has 12 image sequences, 4 sequences for each of the three directions, i.e. parallel, 45 degrees and 90 degrees with respect to the image plane. The length of each sequence is not identical due to the variation of the walker's speed. Dataset C [90] was collected by an infrared (thermal) camera. It contains 153 subjects and takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The videos were all captured

at night.

More outdoor gait sequences are also found in the HID-UMD database [91], from University of Maryland. This database contains walking sequences of 25 people in 4 different poses. These are frontal view/walking-toward, frontal view/walking-away, frontal-parallel view/toward left, frontal-parallel view/toward right.

A database which contains both indoor and outdoor sequences is the Southampton Human ID gait database (SOTON Database) [84]. This database consists of a large population, which is intended to address whether gait is individual across a significant number of people in normal conditions, and a small population database, which is intended to investigate the robustness of biometric techniques to imagery of the same subject in various common conditions.

Currently, in real problems, more complex situations are managed. Thus, for example, outdoor scenarios may be appropriate to deal with real surveillance situations, where occlusions occur frequently. To address the challenge of occlusions, the TUM-IITKGP Gait Database is presented in [92]. On the other hand, some gait recognition methods solely focus on data extracted from a RGB video stream. The freely available TUM Gait from Audio, Image and Depth (TUM-GAID) database is presented in [93] to provide a means for multimodal gait recognition.

Other indoor and outdoor datasets have been specifically designed for action recognition. However, it is possible to extract a subset of gait sequences from them. Examples of this are Weizmann [94], KTH [95], Etiseo, Visor[96] and UIUC [97]. The Weizmann database contains the walking action among other 10 human actions, each action performed by nine people. KTH dataset contains six types of human actions performed several times by 25 people in four different scenarios. ETISEO and Visor were created to be applied in video surveillance algorithms. UIUC (from University of Illinois) consists of 532 high resolution sequences of 14 activities (including walking) performed by eight actors.

With the new gait recognition approaches that deal with 3D information, new gait datasets for multi-view recognition have emerged. One of the first multi-view published dataset was CASIA Dataset B[86]. Dataset B is a large multi-view gait database. There are 124 subjects, and the gait data was captured from 11 viewpoints. However, neither the camera position nor the camera orientation is provided, and the videos are not synchronized. Figure
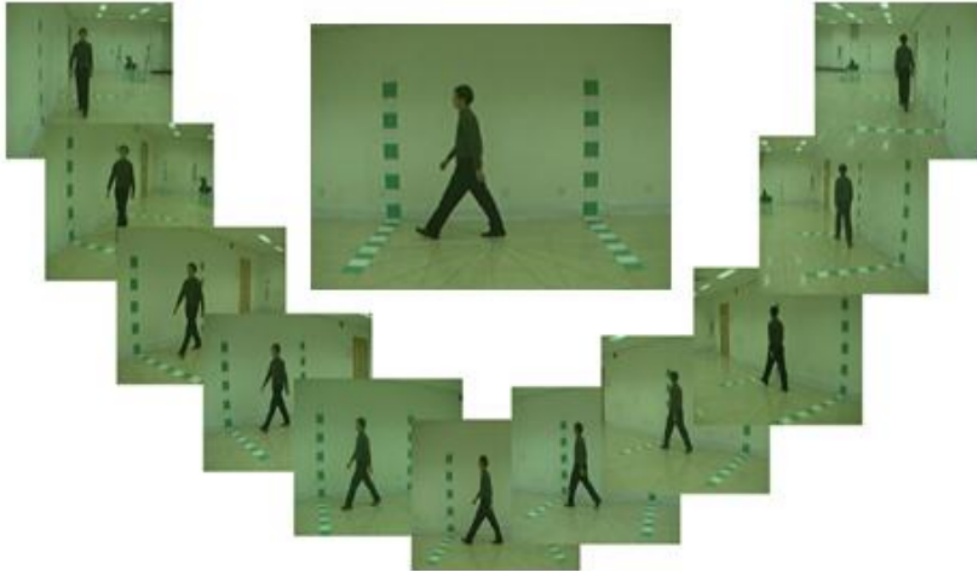
24

3.2 shows a sample from CASIA Database B.

Figure 3.2: Samples from CASIA Dataset B.

As can be seen in OU-ISIR and SOTON databases among others, treadmills are widely used, nonetheless, an inherent problem associated with walking on a treadmill is that the human gait is not as natural as in real situations. An example of using of the treadmill in a multiview dataset is presented in the CMU Motion of Body (MoBo) Database [85], which contains videos of 25 subjects walking on a treadmill from multiple views. Figure 3.3 shows a sample from MoBo Database. A summary of the current freely available gait datasets is shown in Table 3.2.

Other multiview datasets are specifically designed for action recognition rather than gait recognition. A summary of them can be seen in Table 3.1. The i3DPost Multi-View Dataset [98] was recorded using a convergent eight camera setup to produce high definition multi-view videos, where each video depicts one of eight people performing one of twelve different human actions. A subset for gait recognition can be obtained from this dataset. The actors enter the scene from different entry points, which seems to be suitable to test view-invariant gait recognition algorithms. However, the main drawback of this subset is the short length of the gait sequences, extracted from a bigger collection of actions.

The Faculty of Science, Engineering and Computing of Kingston University collected in 2010 a large body of human action video data named

Figure 3.3: Example of images from the CMU Motion of Body (MoBo) Database.

Table 3.1: Multi-view action datasets including walking. This table shows some of the most popular multiview action datasets, which contain walking sequences among other activities.

Database	Actions	Subjects	Source	Views	Path	Year
i3DPost [98]	13	8	Indoor	8	Straight	2009
MuHAVi [99]	17	14	Indoor	8	Straight	2010
IXMAS [100]	11	10	Indoor	5	Closed curve	2006

MuHAVi (Multicamera Human Action Video dataset) [99]. It provides a realistic challenge to objectively compare action recognition algorithms. There are 17 action classes (including walk and turn back) performed by 14 actors. A total of eight non-synchronized cameras are used. The main weakness of

MuHAVi is that the walking activity is carried out in an unique predefined trajectory. Due to this, this dataset is not very suitable to compare view-invariant gait recognition algorithms. Moreover, this dataset was specifically designed to test action recognition algorithms, and it does not contain gait sequences of enough length.

The INRIA[1] Xmas Motion Acquisition Sequences (IXMAS) database, reported in [100], contains five-view video and 3D body model sequences for eleven actions and ten persons. A subset for gait recognition challenges can be obtained from the INRIA IXMAS database. However, humans appear walking in very closed circle paths. Consequently, the dataset does not provide very realistic gait sequences.

---

[1] "Institut National de Recherche en Informatique et en Automatique", France

Table 3.2: Summary of existing gait datasets. The table below shows some features of the existing gait datasets. Some of the current databases are divided into other subsets, to deal with specific challenges, as clothes variation, carrying conditions, or multiple view recognition. The columns #Sub. and #Seq. show the number of subjects and sequences per subjects respectively. The sixth column indicates whether the database use a treadmill.

Database	Type of problem	#Sub.	#Seq.	Source	Tread.	Views	Path	Year
UCSD [88]	Shaded scenes	6	7	Outdoor	No	Side	Circular	1998
HID-UMD [91]	Undetermined	25	1	Outdoor	No	Front, side	Straight	2001
MoBo [85]	Multi-view recog.	25	4	Indoor	Yes	Six views	Straight	2001
SOTON [84] Large	Multiple purposes	$\tilde{1}00$	6	indoor, outdoor	Some	0, 45, 90	Straight	2002
SOTON [84] Small	Diff. walk. cond.	12	15	Indoor	No	0, 45, 90	Straight	
CASIA A [89]	Undetermined	20	12	Outdoor	No	0, 45, 90	Straight	2001
CASIA B [86]	Multi-view recog., diff. carrying cond.	124	10	Indoor	No	11 views	Straight	2005
CASIA C [90]	Diff. walk. cond.	153	10	Outdoor	No	Side	Straight	2005
USF HID [15]	Covariate cond.	122	Up to 5	Outdoor	No	Side	Elliptical	2005
TUM-IITKGP [92]	Occlusions	35	1	Indoor	No	Side	Straight	2011
TUM-GAID [93]	Audio+depth+RGB, diff. carrying cond.	305	1	Indoor	No	Side	Straight	2012
OU-ISIR [83] A	Speed variation	34	68	Indoor	Yes	Side	Straight	2012
OU-ISIR [83] B	Clothes variation	68	Up to 32	Indoor	Yes	Side	Straight	2012
OU-ISIR [83] D	Gait fluctuation	370	185	Indoor	Yes	Side	Straight	2012
OU-ISIR LP [101]	Covariate cond.	4016	1	Indoor	No	Side	Straight	
AVAMVG [87]	Multi-view recog., diff. trajectories	20	10	Indoor	No	Six views	Curved, straight	2013
KY4D [26]	Multi-view recog.	42	6	Indoor	No	16 views	Curved, straight	2014

3.2 Kyushu University 4D gait dataset

Kyushu University 4D Gait Database (KY4D)[2] [26] is composed of sequential 3D models and image sequences of 42 subjects walking along four straight and two curved trajectories, as indicated by black lines in Figure 3.4. The sequences were recorded by 16 cameras, at a resolution of 1032×776 pixels. As can be seen in Figure 3.4, KY4D gait sequences are captured by 16 cameras forming rings at two heights. The lower level comprises the cameras {7451527, 7172435, 7121059, 7451462, 7451476, 7340706, 7451471, 7230135}, whereas the upper level comprises the cameras {7451465, 7340709, 7340697, 7451466, 7451477, 7340692, 7451468, 7340700}.

KY4D provide sequential 3D models of subjects, as well as the intrinsic and extrinsic parameters of each camera. The radius of the curved trajectories varies between $1.5m$ and $3.0m$. The height of the lower camera set is approximately $1.2m$, and the distance from each camera to the center of the studio is about $3.5m$. Figure 3.5 shows a sample extracted from this dataset. Each frame corresponds to a camera view in a certain time t.
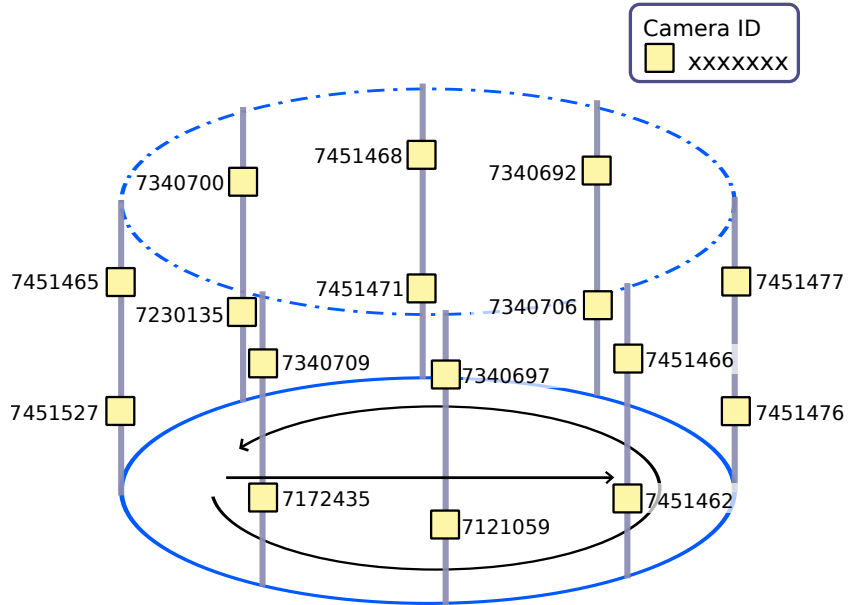

Figure 3.4: Experimental setup of KY4D.

---

[2]Publicly available at: http://robotics.ait.kyushu-u.ac.jp/research-e.php?content=db

Figure 3.5: Example of silhouettes from KY4D multiview dataset.

3.3 The AVA Multi-View Dataset for Gait Recognition

In this section we present the AVA Multi-View Dataset for Gait Recognition (AVAMVG)[3] [87]. This paper briefly describe the camera setup, the database content, and the preprocessing steps carried out in order to further increase the applicability of this database.

---

[3]Publicly available at: http://www.uco.es/grupos/ava/node/41

The AVA Multi-View Dataset for Gait Recognition

David López-Fernández$^{(\boxtimes)}$, Francisco José Madrid-Cuevas,
Ángel Carmona-Poyato, Manuel Jesús Marín-Jiménez,
and Rafael Muñoz-Salinas

Computing and Numerical Analysis Department, Campus de Rabanales,
Maimónides Institute for Biomedical Research (IMIBIC),
University of Córdoba, 14071 Córdoba, Spain
{i52lofed,fjmadrid,ma1capoa,mjmarin,rmsalinas}@uco.es

Abstract. In this paper, we introduce a new multi-view dataset for gait recognition. The dataset was recorded in an indoor scenario, using six convergent cameras setup to produce multi-view videos, where each video depicts a walking human. Each sequence contains at least 3 complete gait cycles. The dataset contains videos of 20 walking persons with a large variety of body size, who walk along straight and curved paths. The multi-view videos have been processed to produce foreground silhouettes. To validate our dataset, we have extended some appearance-based 2D gait recognition methods to work with 3D data, obtaining very encouraging results. The dataset, as well as camera calibration information, is freely available for research purposes.

1 Introduction

Research on human gait as a biometric for identification has received a lot of attention due to the apparent advantage that it can be applied discreetly on the observed individuals without needing their cooperation. Because of this, the automation of video surveillance is one of the most active topics in Computer Vision. Some of the interesting applications are, among others, access control, human-machine interface, crowd flux statistics, or detection of anomalous behaviours [1].

Most current gait recognition methods require gait sequences captured from a single view, namely, from the side view or from the front view of a walking person [2–9]. Hence, there are many existing databases which capture the gait sequences from a single view. However, new challenges in the topic of gait recognition, such as achieving the independence from the camera point of view, usually require multi-view datasets. In fact, articles related to multi-view and cross-view gait recognition have been increasingly published [10–16].

First, some of the existing multi-view current datasets were recorded in controlled conditions, and in some cases, they made use of a treadmill [17–19]. An inherent problem associated with walking on a treadmill is that the human gait

is not as natural as it should be, the gait speed is usually constant, and the subjects cannot turn right or left. They are not representative of human gait in a real world. Secondly, for other multi-view datasets, calibration information is not provided, e.g. [20]. Some of the gait recognition methodologies require camera calibration to deal with 3D information.

In addition, there are not many multi-view datasets specifically designed for gait. Some of them are designed for action recognition, and therefore they do not contain gait sequences of enough length as to contain several gait cycles, because gait is a subset of them.

For this reason, we have created a new indoor dataset to test gait recognition algorithms. This dataset can be applied in workspaces where subjects cannot show the face or use the fingerprint, and even they have to wear special clothing, e.g. a laboratory. Furthermore, in this dataset people appear walking along both straight and curved paths, which makes this dataset suitable to test methods like [10]. The cameras have been calibrated and the methods based on 3D information can use this dataset to test. The dataset is free only for research purposes.

This paper is organized as follows. Section 2 describes current datasets for gait recognition. Section 3 describes the AVA Multi-View Dataset for Gait Recognition (AVAMVG). Section 4 shows several application examples carried out to validate our database. Finally, we conclude this paper in Sect. 5.

2 Current Datasets for Gait Recognition

The chronologic order of appearance of the different human action video datasets runs parallel to the challenges that the scientific community has been considering to face the problem of automatic and visual gait recognition. From this point of view, datasets can be divided into two groups: datasets for single-view gait recognition and datasets for multi-view gait recognition. Besides, the datasets can be divided in two subcategories: indoor and outdoor datasets.

Regarding single view datasets, indoor gait sequences are provided by the OU-ISIR Biometric Database [17]. OU-ISIR database is composed by four treadmill datasets, called A, B, C and D. Dataset A is composed of gait sequences of 34 subjects from side view with speed variation. The dataset B is composed of gait sequences of 68 subjects from side view with clothes variation up to 32 combinations. OU-ISIR gait dataset D contains 370 gait sequences of 185 subjects observed from the lateral view. The dataset D focuses on the gait fluctuations over a number of periods. The OU-ISIR gait dataset C is currently under preparation, and as far as we know, information about this has not been released yet.

In contrast with OU-ISIR, the first available outdoor single-view database was from the Visual Computing Group of the UCSD (University of California, San Diego) [21]. The UCSD gait database includes six subjects with seven image sequences of each, from the side view. In addition to UCSD gait database, one of the most used outdoor datasets for single-view gait recognition is the USF HumanID database [22]. This database consists of 122 persons walking in elliptical paths in front of the camera.

Other outdoor walking sequences are provided in CASIA database, from the Center for Biometrics and Security Research of the Institute of Automation of the Chinese Academy of Sciences. CASIA Gait Database is composed by three datasets, one indoor and the other two outdoor. The indoor dataset can be also considered as a multi-view dataset, and therefore it will be discussed later. The outdoors datasets are named as Dataset A and Dataset C (infrared dataset), and they are described below.

Dataset A [23] includes 20 persons. Each person has 12 image sequences, 4 sequences for each of the three directions, i.e. parallel, 45 degrees and 90 degrees with respect to the image plane. The length of each sequence is not identical due to the variation of the walker's speed. Dataset C [24] was collected by an infrared (thermal) camera. It contains 153 subjects and takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The videos were all captured at night.

More outdoor gait sequences are also found in the HID-UMD database [25], from University of Maryland. This database contains walking sequences of 25 people in 4 different poses (frontal view/walking-toward, frontal view/walking-away, frontal-parallel view/toward left, frontal-parallel view/toward right).

A database containing both indoor and outdoor sequences is the Southampton Human ID gait database (SOTON Database) [18]. This database consists of a large population, which is intended to address whether gait is individual across a significant number of people in normal conditions, and a small population database, which is intended to investigate the robustness of biometric techniques to imagery of the same subject in various common conditions.

Currently, in real problems, more complex situations are managed. Thus, for example, outdoor scenarios may be appropriate to deal with real surveillance situations, where occlusions occur frequently. To address the challenge of occlusions, the TUM-IITKGP Gait Database is presented in [26].

Other indoor and outdoor datasets have been specifically designed for action recognition. However, it is possible to extract a subset of gait sequences from them. Examples of this are Weizmann [27], KTH [28], Etiseo, Visor [29] and UIUC [30]. The Weizmann database contains the walking action among other 10 human actions, each action performed by nine people. KTH dataset contains six types of human actions performed several times by 25 people in four different scenarios. ETISEO and Visor were created to be applied in video surveillance algorithms. UIUC (from University of Illinois) consists of 532 high resolution sequences of 14 activities (including walking) performed by eight actors.

With the new gait recognition approaches that deal with 3D information, new gait datasets for multi-view recognition have emerged. One of the first multi-view published dataset was CASIA Dataset B [20]. Dataset B is a large multi-view gait database. There are 124 subjects, and the gait data was captured from 11 viewpoints. Neither the camera position nor the camera orientation are provided.

As can be seen in OU-ISIR and SOTON databases among others, treadmills are widely used, nonetheless, an inherent problem associated with walking on a treadmill is that the human gait is not as natural as it should be. An example

of using of the treadmill in a multiview dataset is presented in the CMU Motion of Body (MoBo) Database [19], which contains videos of 25 subjects walking on a treadmill from multiple views. A summary of the current freely available gait datasets is shown in Table 2.

Table 1. Multi-view action datasets that include walking as an action. This table shows some of the most popular multiview action datasets, which contain walking sequences among other activities.

Database	Actions	Subjects	Source	Views	Path	Year
i3DPost [31]	13	8	Indoor	Eight views	Straight	2009
MuHAVi [32]	17	14	Indoor	Eight views	Straight	2010
IXMAS [33]	11	10	Indoor	Five views	Closed curve	2006

Other multiview datasets, specifically designed for action recognition rather than gait recognition, are described then. A summary of them can be seen in Table 1. The i3DPost Multi-View Dataset [31] was recorded using a convergent eight camera setup to produce high definition multi-view videos, where each video depicts one of eight people performing one of twelve different human actions. A subset for gait recognition can be obtained from this dataset. The actors enter the scene from different entry points, which seems to be suitable to test invariant view gait recognition algorithms. However, the main drawback of this subset is the short length of the gait sequences, extracted from a bigger collection of actions.

The Faculty of Science, Engineering and Computing of Kingston University collected in 2010 a large body of human action video data named MuHAVi (Multicamera Human Action Video dataset) [32]. It provides a realistic challenge to objectively compare action recognition algorithms. There are 17 action classes (including walk and turn back) performed by 14 actors. A total of eight non-synchronized cameras are used. The main weakness of MuHAVi is that the walking activity is carried out in an unique predefined trajectory. Due to this, this dataset is not very suitable to compare invariant-view gait recognition algorithms. This dataset was specifically designed to test action recognition algorithms, and it does not contain gait sequences of enough length.

The INRIA Xmas Motion Acquisition Sequences (IXMAS) database, reported in [33], contains five-view video and 3D body model sequences for eleven actions and ten persons. A subset for gait recognition challenges can be obtained from the INRIA IXMAS database. However, humans appear walking in very closed circle paths. Consequently, the dataset does not provide very realistic gait sequences.

3   AVA Multi-view Dataset for Gait Recognition

In this section we briefly describe the camera setup, the database content, and the preprocessing steps carried out in order to further increase the applicability of the database.

Table 2. Summary of existing gait datasets. The table below shows some features of the existing gait datasets. Some of the current databases are divided into other subsets, to deal with specific challenges, as clothes variation, carrying conditions, or multiple view recognition. The column sequences shows the number of available sequences per subject.

Database	Subset	Type of problem	Subjects	Sequences	Source	Treadmill	Views	Path	Year
UCSD [21]	N.A	Shaded scenes	6	7	Outdoor	No	Side	Circular	1998
HID-UMD [25]	N.A	Undetermined	25	1	Outdoor	No	Front, side	Straight	2001
MoBo [19]	N.A	Multi-view recognition	25	4	Indoor	Yes	Six views	Straight	2001
SOTON [18]	Large	Multiple purposes	~100	6	In-outdoor	Some seq.	0, 45, 90	Straight	2002
	Small	Diff. walk. cond.	12	15	Indoor	No	0, 45, 90	Straight	2002
	A [23]	Undetermined	20	12	Outdoor	No	0, 45, 90	Straight	2001
CASIA	B [20]	Multi-view recognition and diff. carrying cond.	124	10	Indoor	No	11 views	Straight	2005
	C [24]	Diff. walk. cond.	153	10	Outdoor	No	Side	Straight	2005
USF Human ID [22]	N.A	Covariate conditions	122	Up to 5	Outdoor	No	Side	Elliptical	2005
TUM-IITKGP [26]	N.A	Occlusions	35	1	Indoor	No	Side	Straight	2011
OU-ISIR [17]	A	Speed variation	34	68	Indoor	Yes	Side	Straight	2012
	B	Clothes variation	68	Up to 32	Indoor	Yes	Side	Straight	2012
	D	Gait fluctuation	370	185	Indoor	Yes	Side	Straight	2012
AVA	N.A	Multi-view recognition	20	10	Indoor	No	Six views	Curved and straight	2013

3.1 Studio Environment and Camera Setup

Six convergent IEEE-1394 FireFly MV FFMV-03M2C cameras are equipped in the studio where the dataset was recorded, spaced in a square of 5.8 m of side at a height of 2.3 m above the studio floor. The cameras provide 360° coverage of a capture volume of 5 m × 5 m × 2.2 m.

A natural ambient illumination is provided by four windows through which natural light enters into the scene. Video gait sequences were recorded at different times of day and the cameras were positioned above the capture volume and were directed downward.

Instead of using a screen backdrop of a specific color, as in [31], the background of the scene is the white wall of the studio. However, to facilitate foreground segmentation, the actors wear clothes of different color than the background scene.

Human gait is captured in 4:3 format with 640 × 480 pixels at 25 Hz. Synchronized videos from all six cameras were recorded uncompressed directly to disk with a dedicated PC capture box. All cameras were calibrated to extract their intrinsic (focal length, centre of projection, and distortion coefficients) and extrinsic (pose, orientation) parameters.

To get the intrinsics of each camera, we used a classical black-white chessboard based technique [34] (OpenCV), while for the extrinsics we used the Aruco library [35] 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. An example of extrinsics calibration based on Aruco library is shown in Fig. 1. Calibration of the studio multi-camera system can be done in less than 10 min using the above referenced techniques.
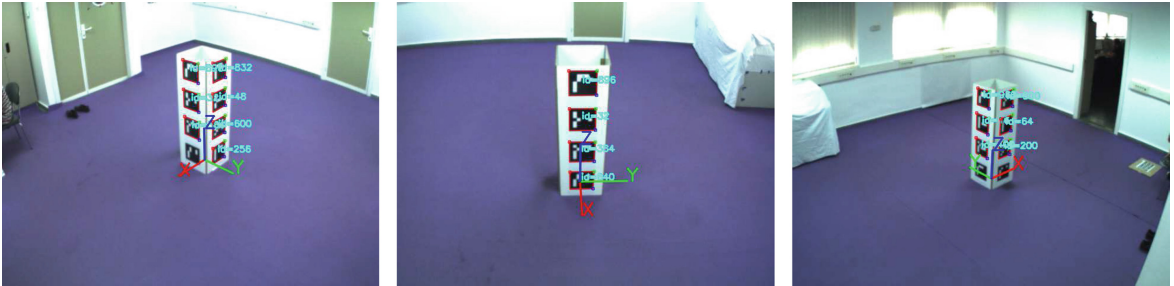

Fig. 1. 3D artifact with Aruco [35] board of markers, used for getting the pose and orientation of each camera.

3.2 Database Description

Using the camera setup described above, twenty humans (4 females and 16 males), participated in ten recording sessions each. Consequently, the database contains 200 multi-view videos or 1200 (6 × 200) single view videos. In the following paragraphs we describe the walking activity carried out by each actor of the database.
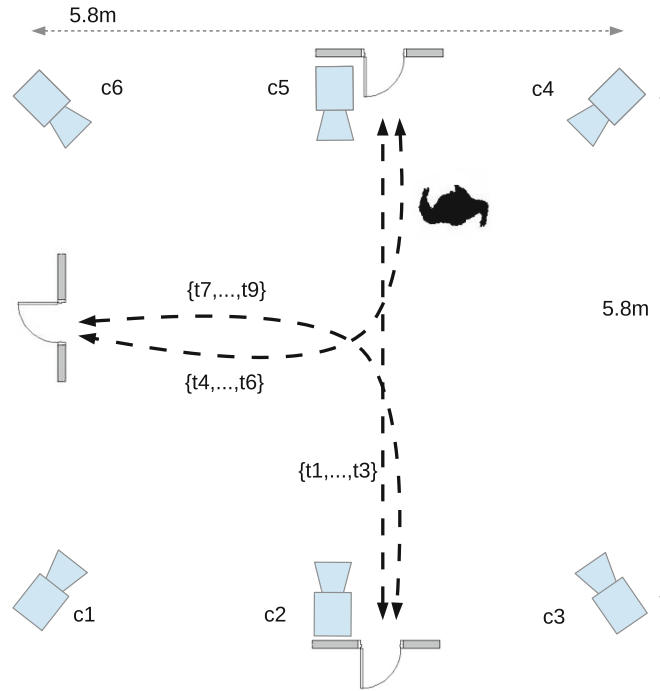
Fig. 2. Workspace setup for dataset recording, where $\{c1, ..., c6\}$ represent the set of cameras of the multiview dataset and $\{t1, ..., t9\}$ represent the different trajectories followed by each actor of the dataset.

Ten gait sequences were designed before the recording sessions. All actor depict three straight walking sequences ($\{t1, ..., t3\}$), and six curved gait sequences ($\{t4, ..., t9\}$), as if they had to turn a corner. The curved paths are composed by a first section in straight line, then a slight turn, and finally a final straight segment. These paths are graphically described in Fig. 2. In the last sequence actors describe a figure-eight path ($t10$).

3.3   Multi-view Video Preprocessing

The raw video sequences were preprocessed to further increase the applicability of the database. To obtain the silhouettes of actors, we have used the Horprasert's algorithm [36]. This algorithm is able to detect moving objects in a static background scene that contains shadows on color images, and it is also able to deal with local and global perturbations such as illumination changes, casted shadows and lightening.

In Fig. 3, several walking subjects of the AVA Multi-View Dataset for Gait Recognition are shown.

4   Database Application Examples

In this section, we carry out several experiments to validate our database. First, we use a Shape from Silhouette algorithm [37] to get 3D reconstructed human

Fig. 3. Example of our multiview dataset. People walking in different directions, from multiple points of view.

Fig. 4. 3D reconstructed gait sequences. Example of reconstructed gait sequences, sampled at 2 Hz, where each point represents the center of a squared voxel.

volumes along the gait sequence. The whole gait sequences can be reconstructed, as shown in Fig. 4. Then, these gait volumes are aligned and centred respect to a global reference system. After this, we can get rendered projections of these volumes to test 2D-based gait recognition algorithms. By this way, we can test view-dependent gait recognition algorithms on any kind of path, either curved or straight.

4.1 Gait Recognition Based on Rendered Gait Entropy Images

One of the most cited silhouette-based representations is the Gait Energy Image (GEI) [9], which represents gait using a single grey scale image obtained by averaging the silhouettes extracted over a complete gait cycle. In addition to GEI, a gait representation called Gait Entropy Image (GEnI) is proposed in [7]. GEnI encodes in a single image the randomness of pixel values in the silhouette images over a complete gait cycle. Thus it is a compact representation which is an ideal starting point for feature selection. In fact, GEnI was proposed to measure the relevance of gait features extracted from the GEI.

As we have aligned gait volumes, we can use rendered side projections of the aligned volumes to compute the GEI. In addition, the GEnI can be computed by calculating Shannon entropy for each pixel of the silhouette images, rendered from side projections of the aligned volumes. In this way, the methods proposed in [7,9] can be tested in a view invariant way. Figure 5 shows the GEI and GEnI descriptors computed over rendered images of the aligned sequence.
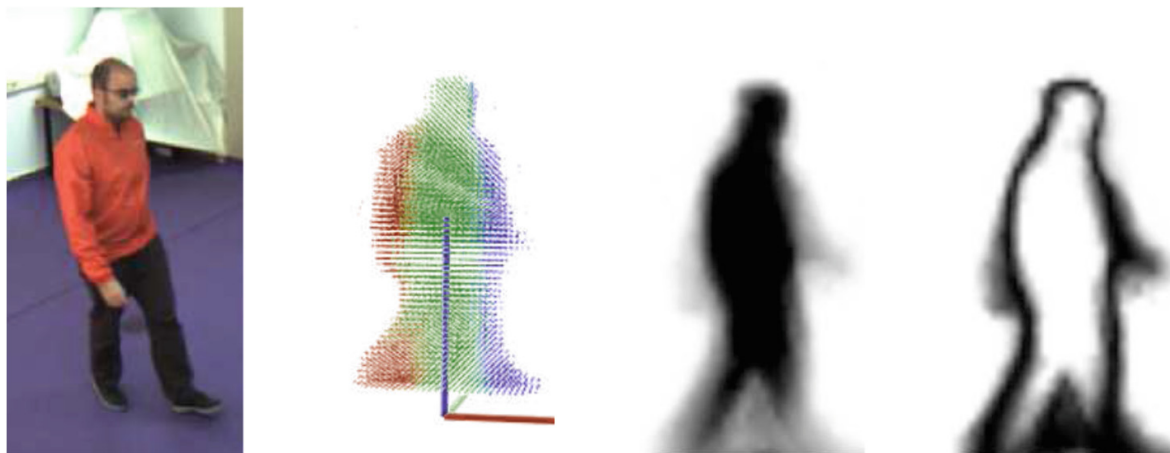

Fig. 5. GEI and GEnI. The leftmost image shows a walking subject. The reconstructed volumes are aligned along the gait sequence, as can be seen in the second image. The two last images show the GEI and GEnI computed over rendered images of the aligned sequence, respectively.

We designed a hold-out experiment where the gallery set is composed by the 1st, 2nd, 4th, 5th, 7th and 8th sequences and probe set is formed by 3rd, 6th, and 9th sequences of the AVA Multi-View dataset. The recognition rate

obtained with the application of the gait descriptors proposed in [7,9] are shown in Table 3.

Table 3. Results of the algorithm proposed in [7] **on the AVA Multi-View dataset**, based on silhouettes. We report the recognition rate in %, comparing GEI with the GEnI by direct template matching, using the AVA Multiview Dataset.

Database (probe)	GEI	GEnI
AVA Multiview Dataset (AVA B)	94.6	98.1

4.2 Front View Gait Recognition by Intra- and Inter-frame Rectangle Size Distribution

In [8], video cameras are placed in hallways to capture longer sequences from the front view of walkers rather than the side view, which results in more gait cycles per gait sequence. To obtain a gait representation, a morphological descriptor, called Cover by Rectangles (CR), is defined as the union of all the largest rectangles that can fit inside a silhouette. Despite of the high recognition rate, a drawback of this approach is the dependence with respect to the angle of the camera.

According to the authors, Cover by Rectangles has the following useful properties: (1) the elements of the set overlap each other, introducing redundancy (i.e. robustness), (2) each rectangle covers at least one pixel that belongs to no other rectangle, and (3) the union of all rectangles reconstructs the silhouette so
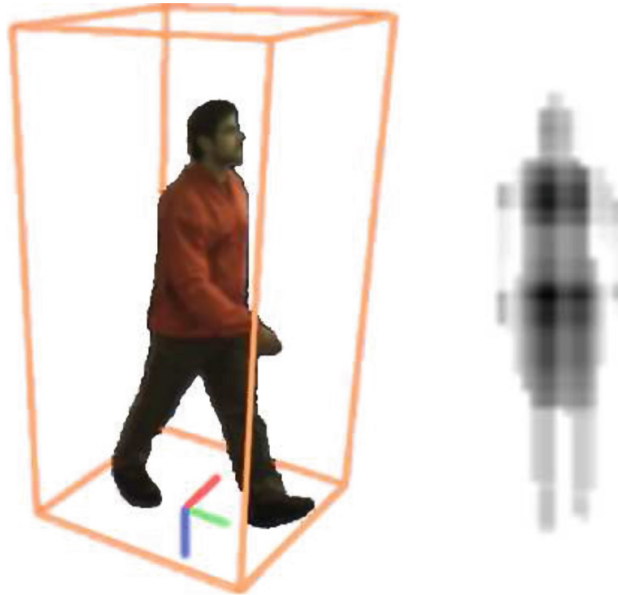

Fig. 6. Cover by Rectangles descriptor. Bounding box of a walking human (left), Cover by Rectangles descriptor (right). A gray level on pixel displays the density of rectangles that contains that pixel.
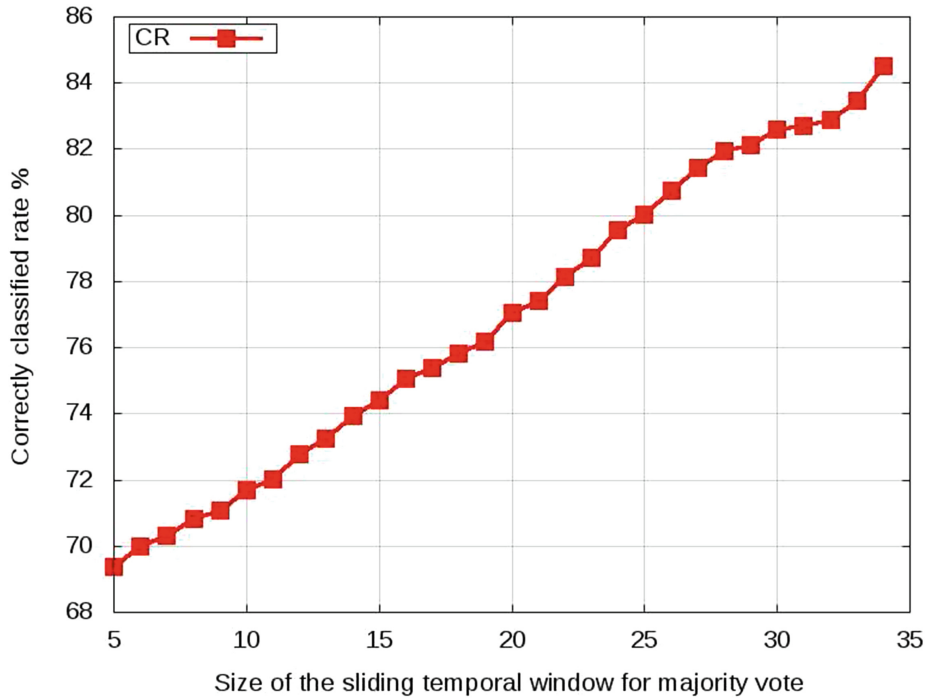
Fig. 7. Recognition rate obtained with the application of the appearance based algorithm proposed in [8]**.** Since we have aligned the reconstructed volumes along the gait sequence, we can use a frontal projection of them. We show the effect on the classification rate of using a sliding temporal window for voting.

that no information is ever lost. A representation of this descriptor can be seen in Fig. 6.

As we have aligned gait volumes, we can use front-rendered projections of the aligned volumes to compute the CR, and therefore the method proposed in [8] can be tested in a view invariant way.

To test the algorithm with the AVA Multi-view Dataset, we use a leave-one-out cross-validation. Each fold is composed by a tuple formed by a set of 20 sequences (one sequence per actor) for testing, and by the remaining eight sequences of each actor for training, i.e. 160 sequences for training and 20 sequences for test. We use SVM with Radial Basis Functions, since we obtained better results than with others classifiers. To make the choice of SVM parameters independent of the sequence test data, we cross-validate the SVM parameters on the training set. For this experiment, the features vector size was set to $L = 20$, and the histogram size with which the highest classification rate is achieved is $M = N = 25$ (see [8]). With the CR descriptor applied on the frontal volume projection, we obtain a maximum accuracy of 84.52 %, as can be seen in Fig. 7.

5   Conclusions

In this paper, we present a new multi-view database containing gait sequences of 20 actors that depict ten different trajectories each. The database has been

specifically designed to test multi-view and 3D based gait recognition algorithms. The dataset contains videos of 20 walking persons (men and women) with a large variety of body size, who walk along straight and curved paths. The cameras have been calibrated and both calibration information and binary silhouettes are also provided.

To validate our database, we have carried out some experiments. We began with the 3D reconstruction of volumes of walking people. Then, we aligned and centred them respect to a global reference system. After this, since we have reconstructed and aligned gait sequences, we used rendered projections of these volumes to test some appearance-based algorithms that work with silhouettes to identify an individual by his manner of walking.

This dataset can be applied in workspaces where subjects cannot show the face or use the fingerprint, and even they have to wear special clothing, e.g. a laboratory. The dataset is free only for research purposes[1].

References

1. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst. Man Cybern. **34**, 334–352 (2004)
2. Lee, C.P., Tan, A.W.C., Tan, S.C.: Gait recognition via optimally interpolated deformable contours. Pattern Recogn. Lett. **34**, 663–669 (2013)
3. Das Choudhury, S., Tjahjadi, T.: Gait recognition based on shape and motion analysis of silhouette contours. Comput. Vis. Image Underst. **117**, 1770–1785 (2013)
4. Zeng, W., Wang, C.: Human gait recognition via deterministic learning. Neural Netw. **35**, 92–102 (2012)
5. Roy, A., Sural, S., Mukherjee, J.: Gait recognition using pose kinematics and pose energy image. Sig. Process. **92**, 780–792 (2012)
6. Huang, X., Boulgouris, N.: Gait recognition with shifted energy image and structural feature extraction. IEEE Trans. Image Process. **21**, 2256–2268 (2012)
7. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recogn. Lett. **31**, 2052–2060 (2010)
8. Barnich, O., Van Droogenbroeck, M.: Frontal-view gait recognition by intra- and inter-frame rectangle size distribution. Pattern Recogn. Lett. **30**, 893–901 (2009)
9. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 316–322 (2006)
10. Iwashita, Y., Ogawara, K., Kurazume, R.: Identification of people walking along curved trajectories. Pattern Recogn. Lett. **48**(0), 60–69 (2014)
11. Kusakunniran, W., Wu, Q., Zhang, J., Li, H.: Gait recognition under various viewing angles based on correlated motion regression. IEEE Trans. Circuits Syst. Video Technol. **22**, 966–980 (2012)

---

[1] Full database access information: http://www.uco.es/grupos/ava/node/41.

12. Krzeszowski, T., Kwolek, B., Michalczuk, A., Świtoński, A., Josiński, H.: View independent human gait recognition using markerless 3D human motion capture. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 491–500. Springer, Heidelberg (2012)

13. Lu, J., Tan, Y.P.: Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. Pattern Recogn. Lett. **31**, 382–393 (2010)

14. Goffredo, M., Bouchrika, I., Carter, J., Nixon, M.: Self-calibrating view-invariant gait biometrics. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **40**, 997–1008 (2010)

15. 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)

16. Bodor, R., Drenner, A., Fehr, D., Masoud, O., Papanikolopoulos, N.: View-independent human motion classification using image-based reconstruction. Image Vis. Comput. **27**, 1194–1206 (2009)

17. Makihara, Y., Mannami, H., Tsuji, A., Hossain, M., Sugiura, K., Mori, A., Yagi, Y.: The ou-isir gait database comprising the treadmill dataset. IPSJ Trans. Comput. Vis. Appl. **4**, 53–62 (2012)

18. Shutler, J., Grant, M., Nixon, M.S., Carter, J.N.: On a large sequence-based human gait database. In: Proceedings of RASC, pp. 66–72. Springer (2002)

19. Gross, R., Shi, J.: The cmu motion of body (mobo) database. Technical report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA (2001)

20. Yu, S., Tan, D., Tan, T.: A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 4, pp. 441–444 (2006)

21. Nixon, M.S., Tan, T.N., Chellappa, R.: Human Identification Based on Gait, vol. 4. Springer, New York (2006)

22. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The humanid gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 162–177 (2005)

23. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1505–1518 (2003)

24. Tan, D., Huang, K., Yu, S., Tan, T.: Efficient night gait recognition based on template matching. In: 18th International Conference on Pattern Recognition, ICPR 2006, vol. 3, pp. 1000–1003 (2006)

25. Chalidabhongse, T., Kruger, V., Chellappa, R.: The umd database for human identification at a distance. Technical report, University of Maryland (2001)

26. Hofmann, M., Sural, S., Rigoll, G.: Gait recognition in the presence of occlusion: a new dataset and baseline algorithm. In: Proceedings of 19th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG), Plzen, Czech Republic, 31 January 2011–03 February 2011

27. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: Tenth IEEE International Conference on Computer Vision, ICCV 2005, vol. 2, pp. 1395–1402 (2005)

28. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local svm approach. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol. 3, pp. 32–36 (2004)

29. Vezzani, R., Cucchiara, R.: Video surveillance online repository (visor): an integrated framework. Multimedia Tools Appl. **50**, 359–380 (2010)

30. Tran, D., Sorokin, A.: Human activity recognition with metric learning. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 548–561. Springer, Heidelberg (2008)
31. Gkalelis, N., Kim, H., Hilton, A., Nikolaidis, N., Pitas, I.: The i3dpost multi-view and 3d human action/interaction database. In: Proceedings of the 2009 Conference for Visual Media Production, CVMP '09, pp. 159–168. IEEE Computer Society, Washington, DC (2009)
32. Singh, S., Velastin, S., Ragheb, H.: Muhavi: a multicamera human action video dataset for the evaluation of action recognition methods. In: 2010 Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 48–55 (2010)
33. Weinland, D., Ronfard, R., Boyer, E.: Free viewpoint action recognition using motion history volumes. Comput. Vis. Image Underst. **104**, 249–257 (2006)
34. Bradski, G., Kaehler, A.: Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly, Cambridge (2008)
35. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F., Marín-Jiménez, M.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recogn. **47**, 2280–2292 (2014)
36. Horprasert, T., Harwood, D., Davis, L.S.: A statistical approach for real-time robust background subtraction and shadow detection. In: Proceedings of IEEE ICCV, pp. 1–19 (1999)
37. Díaz-Más, L., Muñoz-Salinas, R., Madrid-Cuevas, F., Medina-Carnicer, R.: Shape from silhouette using dempster shafer theory. Pattern Recogn. **43**, 2119–2131 (2010)

Chapter 4

3D Reconstruction and Gait Alignment

This dissertation proposes some model-free approaches to recognize walking humans independently of the viewpoint and regardless direction changes. These approaches focus on capturing 3D morphological, structural and dynamical information of the gait from volumetric reconstructions of walking humans.

This chapter start by describing the computational process involved in the generation of a 3D human reconstruction, from silhouettes extracted from several viewpoints. This procedure requires calibration parameters, such as the camera matrix, distortion coefficients (intrinsic parameters), pose and orientation (extrinsic parameters) of each camera. This point is covered in detail in Section 4.1.

After the 3D reconstruction, the individual is detected and the 3D volumes corresponding to a gait sequence are aligned and centered with respect to a global reference system, in order to achieve the independence with respect to the trajectory of motion. This point is described in Section 4.2.

4.1 3D reconstruction

Shape-from-Silhouette (SfS) is a well-known approach to reconstruct the 3D structure of objects using a set of silhouettes obtained from different views.

Baumgart [102] was the first to introduce concepts about the 3D geometric modelling, but it was not until 1991 that Laurentini [56] defined the concept of the Visual Hull (VH) as the closest 3D solid equivalent to the real object that explains the silhouettes extracted. The VH is the geometric intersection of all visual cones explaining the projection of a silhouette in its corresponding camera image. Consequently, as the number of views increases, so does the precision of the reconstructed object [103].
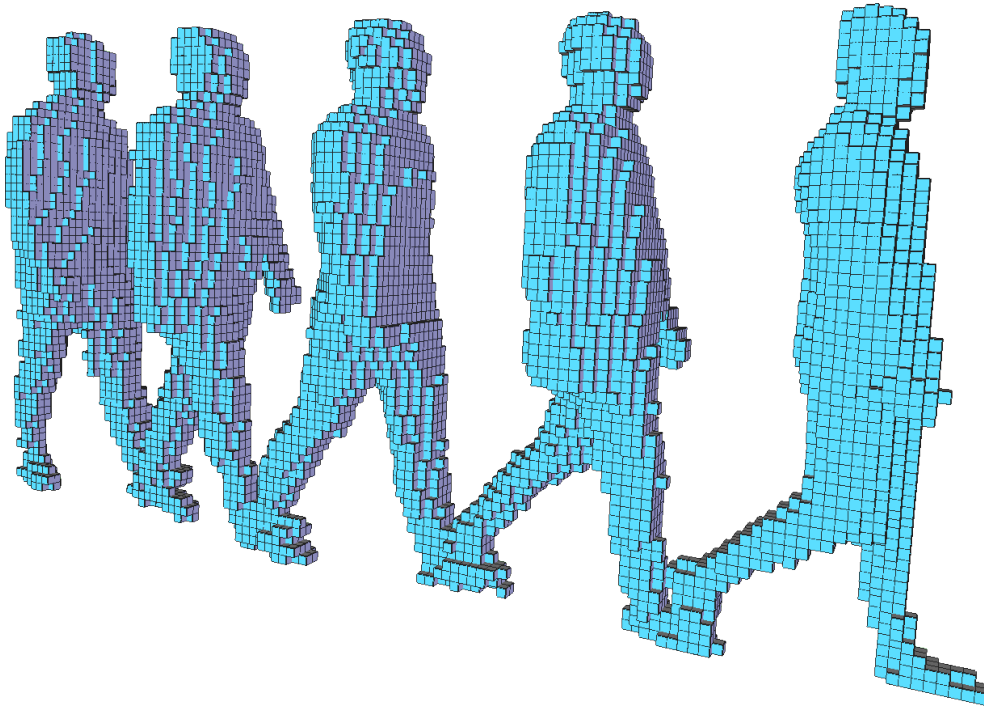

Figure 4.1: Example of reconstructed gait sequence, sampled at 2Hz, where each cube represents a voxel.

Since this work propose methods that generate gait descriptors from 3D occupation volumes or VH, it is important to first explain the basic concepts of a standard SfS method.

It is assumed a three-dimensional work area that is divided into cubes of the same volume called voxels. The voxel set is denoted by $\mathbf{V} = \{v^i = \{x, y, z\} | i = 1, \ldots, n\}$, where n represents the total number of voxels and $v^i = \{x, y, z\}$ the center of the i-th voxel.

Let us also assume that there is a set of cameras placed at known locations (extracted using calibration) and that we have a background subtraction method that obtains the silhouettes of the foreground objects. We

denote these foreground images as $\mathbf{F} = \{\mathcal{F}_c | c = 1, \dots, C\}$, where C is the number of cameras. A pixel $(x, y) \in \mathcal{F}_c$ is *true* if it is classified as belonging to the foreground and *false* otherwise.

SfS methods examine voxel projections in the foreground images in order to determine whether they belong to the shape of objects. This is achieved by means of a projection test. 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 a border and a decision must be made.

Projection tests play an essential role in SfS algorithms. The most simple one consist in projecting only the center of the voxel. More complex approaches consist in testing either all the pixels or a subset of pixels within the polygon formed by the voxel projection. Either way, the result is a boolean decision indicating whether the voxel is occupied or not.

An overview of the SfS algorithm is shown in Algorithm 1. At the first, all voxels are assumed to be occupied. Then, voxels projections are examined in all the images using the projection test. If the projection test indicates that the voxel does not belong to the silhouette of any foreground image, the it is considered as not occupied independently of its projection on the rest of images. The result of the algorithm is the set of VHs of the objects in the scene.

Algorithm 1 Classical SfS algorithm.

Require:
    Foreground images. $\mathbf{F} : \{\mathcal{F}_c\}$
    Projection Test Function. $PT(v^i, \mathcal{F}_c)$
 1: **for all** $v^i \in \mathbf{V}$ **do**
 2:     $v^i \leftarrow occupied$
 3:     **for all** $\mathcal{F}_c \in \mathbf{F}$ **do**
 4:         **if** $PT(v^i, \mathcal{F}_c)$ is false **then**
 5:             $v^i \leftarrow \neg ocuppied$
 6:             examine next voxel
 7:         **end if**
 8:     **end for**
 9: **end for**
10: **return** VH

We assume that we may have inconsistent silhouettes, due to poor segmentation results. It could cause false positive and false negative errors in the reconstruction stage. To face this inconvenient, the SfS algorithm proposed by [104] could be used. Fig. 4.1 shows a 3D reconstructed gait sequence.

To minimize the computational time, SfS could take advantage of the power of Graphics Processing Units (GPU), as it was proved in [105] and [106].

4.2   Volume detection and alignment

Given a reconstructed volume V_t of a person at each instant t along the way, it is required a mechanism of detection and gait alignment to achieve the independence which refers to the trajectory. This process will allow the individual to walk freely in the scene, without adversely affect to the subsequent generation of gait descriptors.

We assume that although there is only one individual in the scene, reconstructed shadows as well as noise can coexist, due to poor segmentation. 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 ϕ, and its volume has fully entered into the workspace. The value of ϕ is tuned up accordingly to the average corporal volume for humans and the resolution of the 3D reconstructions.

We consider that a voxel size of $0.27 \times 10^{-4} m^3$ is enough to get detailed 3D human reconstructions. The average corporal volume for humans is $66.4L = 0.6640 \times 10^{-1} m^3$ measured by the water displacement method in 521 people aged $17 - 51$ years [107]. Using a voxel size of $0.27 \times 10^{-4} m^3$, the number of voxels belonging to a person in the 3D volume should be about 2459. With a value of $1 \times 10^3 < \phi < 2459$ the system should be able to detect an adult human.

When the volume belonging to a person has been detected, the centroid p of the ground projection is calculated. Then, the volume is moved into a bounding-box of average adult human's size, so that the workspace where the descriptor will be computed is bounded. This process is illus-
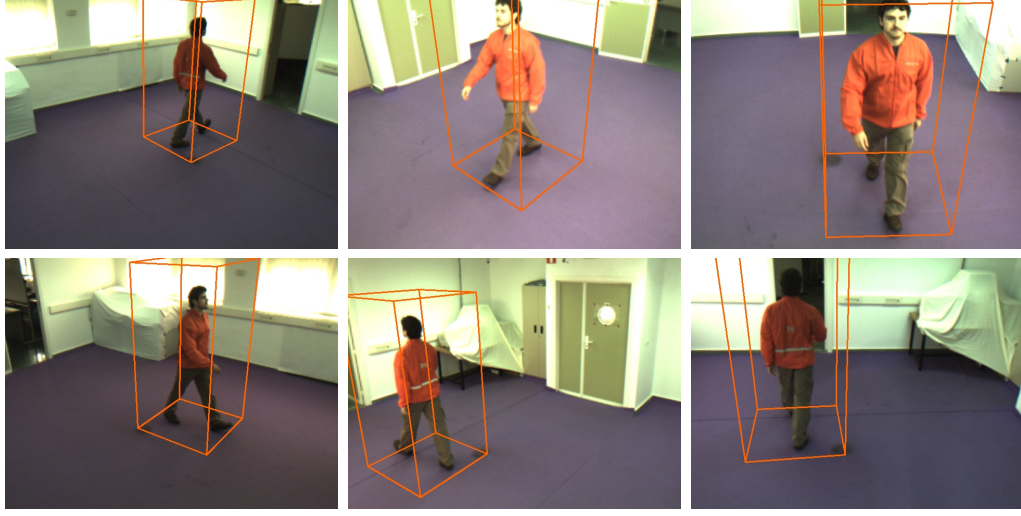
Figure 4.2: Bounding-box with size of an average adult human, where the descriptor will be computed on the 3D reconstructed data at time t.

trated in Fig. 4.2. The estimation of the direction of path is determined by the displacement vector, defined as

$$\vec{v}_t = p_t - p_{t-1}, \tag{4.1}$$

where t is the current time, p_t is the centroid's current position, and p_{t-1} is the last known position of the centroid.

The angle of the displacement vector is calculated using the expression

$$\alpha_t = \arctan \frac{v_{t_y}}{v_{t_x}}. \tag{4.2}$$

An example of extraction of the displacement vector angle can be seen in Fig. 4.3, where top projections of the individual can be seen in several moments of the gait, and the principal axis (perpendicular to displacement vector) is represented. The reconstructed volume is rotated about the body vertical axis using the angle α_t.

Although we assume a constant walking speed, an individual could vary moderately the walking speed in a certain moment of the gait. It could happen, for example, when the individual is describing a curved closed path.

If the walking speed is very low at time t, $|\vec{v}_t|$ will be too small, which could result in a noisy estimation of the angle α_t. To attenuate this noise

Figure 4.3: The displacement vector (red line) of the individual is computed at each time. The principal axis (blue line) is perpendicular to the displacement vector.

in the α_t estimation and smooth the path, we propose a weighted average of the displacement vector angle as follows

$$\bar{\alpha}_t = \alpha_t \cdot \beta + \bar{\alpha}_{t-1} \cdot (1 - \beta), \qquad (4.3)$$

where

$$\beta = \frac{||\vec{v}_t||}{\max_{i=0..t}\{||\vec{v}_i||\}}. \qquad (4.4)$$

The aim of β is to reduce this noise in the estimation of the alignment angle, giving more or less weight to the current estimation depending on the walking speed. For example, if the walking speed is decreasing and it becomes too small, the magnitude of the displacement vector may not be large enough, causing oscillations in the estimation of the angle. In this case, it would be right to give less importance to the current angle estimation. However, if the walking speed is increasing, it would be more appropriate to give more weight to the current angle estimation. A method to decrease over time the denominator in Eq. (4.4) should be applied if the gait sequence were too large. Thus, the whole gait sequence can be centered and aligned along the path as it is illustrated in Fig. 4.4. In the following, the aligned volume is denoted by V^*.
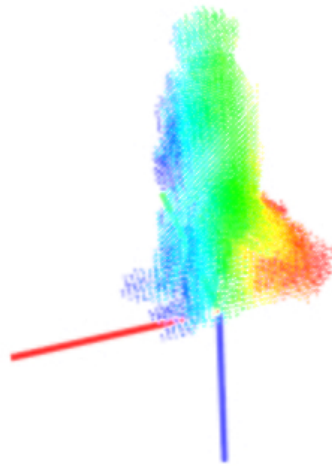
Figure 4.4: Union of all aligned 3D volumes over the gait sequence. Different colors represent the distance to the center on the X axis.

Chapter 5

Morphological Descriptions of 3D Human Reconstructions

In this chapter we present a model-free approach to recognize walking humans independently of the viewpoint and regardless direction changes. This new approach focuses on capturing 3D morphological and structural information from the gait through volumetric reconstructions of the walking human [108]. The use of volumetric information allows more information to be analysed in contrast to other related works, which only compute gait descriptors from silhouettes, discarding an important part of the dynamical and structural information of the gait. This method extends the input domain for the morphological gait descriptors presented in [34], from 2D silhouettes to 3D reconstructions of the individuals, aligned along the way.

Our method relies on morphological analysis of series of 3D occupation volumes which are generated from multi-view video sequences. For each time of the gait sequence, a 3D occupation volume is generated by combining information from multiple silhouette images, from several points of view. Then, this gait volume is aligned and centered with respect to a global reference system. Next, our gait descriptor is computed on each 3D gait volume in order to provide information about their 3D appearance.

Two new gait descriptors are presented. The first is based on computing Cover by Rectangles C(S) descriptors (see Section 2.3) on the front, side, and top projections of the aligned volume V^*. The second is the *Cover by Cubes*, which is defined as the union of all the cubes with the largest size that can fit into a volume belonging to a person.

Main publications associated to this chapter:

- D. López-Fernández, F.J Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, R. Muñoz-Salinas, and R. Medina-Carnicer. Viewpoint-independent gait recognition through morphological descriptions of 3d human reconstructions. *Image and Vision Computing*, 48-49:1-13, 2016. ISSN 0262-8856. doi: 10.1016/j.imavis.2016.01.003. url: `http://www.sciencedirect.com/science/article/pii/S0262885616300014`

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, and R. Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In *Activity Monitoring by Multiple Distributed Sensing, Lecture Notes in Computer Science*, pages 26-39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3. url: `http://link.springer.com/chapter/10.1007%2F978-3-319-13323-2_3`

Viewpoint-independent gait recognition through morphological descriptions of 3D human reconstructions ☆

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

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

ARTICLE INFO

ABSTRACT

Many studies have confirmed gait as a robust biometric feature for identification of individuals. 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 unconstrained paths in indoor environments. The recognition is based on volumetric analysis of the human gait, to exploit most of the 3D information enclosed in it. Appearance-based gait descriptors are extracted from 3D gait volumes and temporal patterns of them are classified using a Support Vector Machine with a sliding temporal window for majority voting. The proposed approach is experimentally validated on the "AVA Multi-View Dataset (AVAMVG)" and on the "Kyushu University 4D Gait Database (KY4D)". The results show that this new approach is able to identify people walking on curved paths.

© 2016 Elsevier B.V. All rights reserved.

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 be applied discreetly on the observed individual without needing the active participation of the individual.

Previous studies on gait recognition have been classified into two categories: model-based and model-free approaches. The model-based methods extract gait features by fitting a model to input images, whereas model-free approaches characterize the human gait pattern by a compact representation, without having to develop any articulated 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 the appearance will not change during walking. In such methods, a change in the appearance, caused by a view change, will adversely affect performance [1]. Fig. 1 shows the influence of a curved path on the silhouette appearance. As one of the advantages of gait as biometric is that it does not need the cooperation of the individual, the trajectory of motion cannot be restricted to just straight paths.

On the other hand, the use of volumetric information allows more information to be analyzed in contrast to methods which only compute gait descriptors from silhouettes or 2D images. This paper presents an efficient view-independent method to recognize people walking along unconstrained (curved and straight) trajectories. This approach focuses on capturing 3D morphological and structural information from volumetric reconstructions of walking humans, which are previously aligned along the way. The main contribution is that our method allows direction changes, achieving a good recognition rate on unconstrained paths.

Some potential applications of this work are access control in special or restricted areas (e.g. military bases, governmental facilities and laboratories) or smart video surveillance (e.g. bank offices) [2].

This article is organized as follows. Section 2 describes works related to the topic of gait recognition. Section 3 explains the details
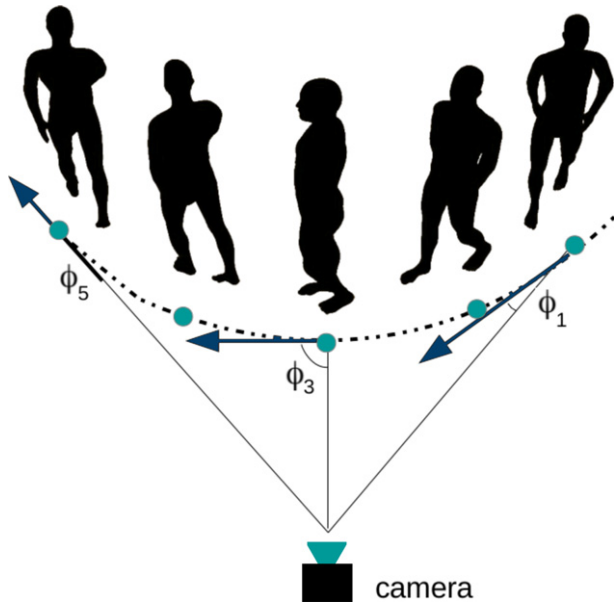
Fig. 1. In a curved path, the observation angle between the walking direction of the subject and optical axis of the camera is gradually changed, which affects the silhouette appearance.

of the proposed algorithm and describes three new descriptors which obtain information from 3D occupancy volumes. 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

The previous work can be categorized into two approaches: view-dependent and view-independent approaches. In the following we describe works related to both categories.

2.1. View-dependent approaches

One of the earliest model-free and view-dependent approaches can be seen in [3], where the width of the outer contour of the binarized silhouette from a side view is used to build a descriptor which contains both structural features and dynamic aspects of gait. Feature vectors derived from binary silhouettes have been also used to train Hidden Markov Models [4]. The contours of silhouettes have been used directly [5,6], and through their Fourier descriptors [7,8].

In addition, the authors of Ref. [9] present a gait recognition method which analyzes the shape of the silhouette using Procrustes Shape Analysis and Elliptic Fourier Descriptors. The Gait Energy Image (GEI) descriptor is introduced in [10], which is the average of all silhouette images for a single gait cycle. Silhouette images are also used by Lam et al. [11] to generate the gait flow image (GFI).

Based on the idea of GEI, Depth Energy Image (DEI) was defined in [12], which is simply the average of the depth silhouettes taken along a gait cycle, over the front view. GEI is also extended in [13] to consider depth information from the side view, by means of a new feature called Depth Gradient Histogram Energy Image (DGHEI). Depth information is also used by Chattopadhyay et al. [14] to address the problem of occlusion in frontal gait recognition.

The Gait Energy Volume (GEV), a binary voxel-discretized volume which is spatially aligned and averaged over a gait cycle, is presented in [15]. The authors apply GEV on partial reconstructions

obtained with depth sensors from the front view of the individual. An extended work from GEV [15] that combines the frontal-view depth gait image and side-view 2D gait silhouette by means of a back-filling technique is presented in [16]. A spatio-temporal representation based on point clouds in a spherical coordinate space was proposed in [17], where frontal 3D point clouds of humans obtained with stereo cameras are used.

A work closely related to our proposed approach in terms of analysis by morphological size distributions was proposed in [18]. In this work, video cameras are placed in hallways to capture longer sequences from the front view of walkers rather than the side view, which results in more gait cycles per gait sequence. Despite the high recognition rate, the main drawback of this model-free approach is the dependence with respect to the viewpoint. To obtain a gait representation directly from silhouettes, the authors proposed the use of a morphological descriptor, called Cover by Rectangles, which is defined as the union of all the largest rectangles that can fit inside a silhouette.

In [19], a 3D approximation of a Visual Hull (VH) [20] is used to design a multi-modal recognition approach. Although a VH model is computed, a gait recognition scheme based on silhouette analysis is applied, which restricts a large amount of discriminant information because the recognition is based on single view silhouette analysis, instead of analyze 3D information. Seely et al. [21] use 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. The sequences are collected from a multi-biometric tunnel, where the subjects just walk straight.

Ariyanto and Nixon [22] propose a model-fitting algorithm, correlation filters and dynamic programming to extract gait kinematics features. They use a structural model including articulated cylinders with 3D Degrees of Freedom (DoF) which are fitted to a visual hull shape to model the human lower legs. In [23], 3D data collected from a projector–camera system is used to fit 3D body models and reconstruct synthetic poses in a gait cycle.

2.2. View-independent approaches

Appearance changes due to viewing angle changes cause difficulties for most of the model-free 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 [24].

Approaches of the first category are represented in [25,26]. Bodor et al. [25] apply image-based rendering on a 3D VH model to reconstruct gait features under a required viewing angle. 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 [26], 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 3D reconstructions, 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, as in the above two works, despite 3D volumes are used, descriptors are extracted from 2D images, so that, the amount of used 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. A method to generate a canonical view of gait from any arbitrary view is described in [27]. The main disadvantage of this method is that the synthesis of a canonical view is only feasible from a limited number of initial views. The performance is significantly dropped down when the angle between image plane and sagittal plane is large.

In [28], a method based on homography to compute view-normalized trajectories of body parts obtained from monocular video sequences was proposed. But this method only works properly 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 [29], however this method seems not to be applicable when the person is walking nearly parallel to the optical axis. In [30] 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 [31]. 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.

Zhao et al. [32] present a multi-camera approach for gait tracking and recognition. The video sequences are used as input, and then a human 3D model is set up. The lengths of key segments are extracted as static parameters, and the motion trajectories of lower limbs are used as dynamic features. A skeletal 3D model is also used in the work of Kastaniotis et al. [33], which presents a framework for pose-based gait recognition and identification, as well as gender recognition.

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 [34], where LDA-subspaces are learned to extract discriminative information from gait features under each viewing angle.

A View Transformation Model (VTM) was introduced in [35] to transform gait features from different views into the same view. The method of Makihara t al. [35] creates a VTM based on frequency-domain gait features, obtained through Fourier Transformation. To improve the performance of this method, Kusakunniran t al. [36] 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 [24]. However, this method cannot deal 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 [24]: (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

We propose a model-free approach to recognize walking humans independently of the viewpoint and regardless direction changes. Our approach focuses on capturing 3D morphological and structural information from the gait through volumetric reconstructions of the walking humans.

The use of volumetric information allows more information to be analyzed in contrast to other related works, which only compute gait descriptors from silhouettes, discarding an important part of the dynamical and structural information of the gait. So that, our method extends the input domain for the morphological gait descriptors used in [18], from 2D silhouettes to 3D reconstructions of the individuals, aligned along the way.

Our approach relies on morphological analysis of series of 3D occupation volumes which are generated from the multi-view video sequences. For each time of the gait sequence, a 3D occupation volume is obtained by combining information from multiple silhouette images, from several points of view. Then, this gait volume is aligned and centered with respect to a global reference system. Next, our gait descriptor is computed from each 3D gait volume in order to provide information about their 3D appearance.

A gait signature is built by aggregating descriptors. 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. The proposed recognition algorithm is shown in Fig. 2, where the identity of a walking human is predicted at each time t. The algorithm consists of five steps which are exposed in detail in this section:

1. Silhouette extraction of each camera's view by a background subtraction technique [37].
2. 3D reconstruction from silhouettes captured from several viewpoints, by a Shape from Silhouette algorithm (SfS) [38].
3. Person detection and gait alignment.
4. Gait descriptor generation, which is used to update the gait signature.
5. Classification of gait signature by a machine learning algorithm.

The first three steps of the algorithm generate a 3D volume with occupancy information of the person at time t, whereas the last three steps perform the feature extraction, signature generation and gait classification.

3.1. 3D reconstruction, detection and alignment

We start by computing a 3D reconstruction of the individual from silhouettes extracted from several viewpoints. This procedure requires calibration parameters, such as the camera matrix, distortion coefficients (intrinsic parameters), pose and orientation (extrinsic parameters) of each camera.

After the 3D reconstruction, the individual is detected and the 3D volumes corresponding to a gait sequence are aligned and centered with respect to a global reference system, so that the generation of the descriptors can be made as if the person had walked on a treadmill in a certain direction.

3.1.1. 3D reconstruction

Since our method generates gait descriptors from 3D occupation volumes or VH, a 3D reconstruction procedure, such a Shape from Silhouette algorithm [38] is required. Fig. 3 shows a 3D reconstructed gait sequence.
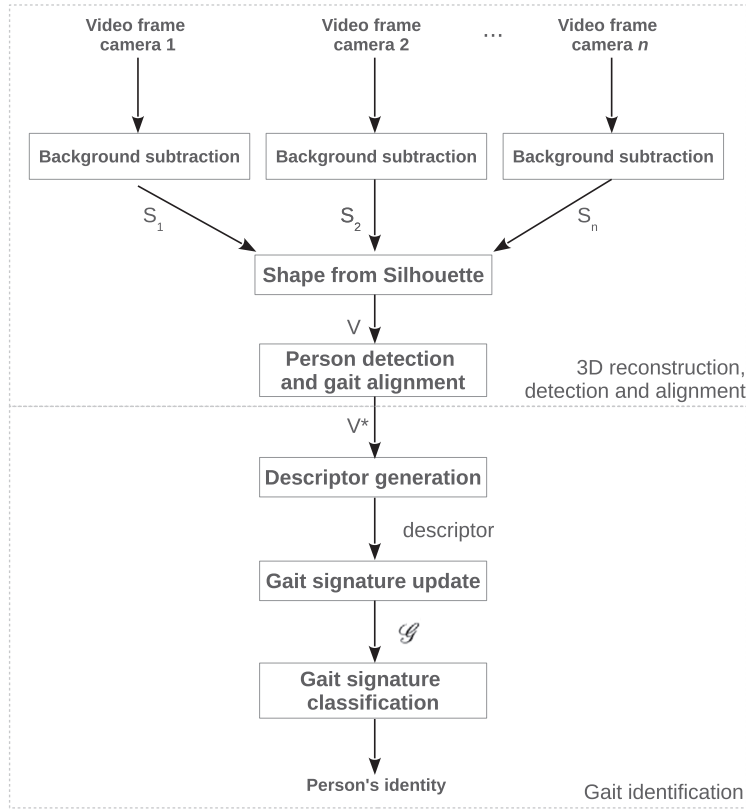
Fig. 2. Steps of our gait-recognition algorithm at time *t*.

3.1.2. Volume detection and alignment

Given a reconstructed volume V_t of a person at each instant t along the way, it required a mechanism of detection and alignment to achieve the independence which refers to the viewpoint. This process will allow the individual to walk freely in the scene, without adversely affecting the subsequent generation of gait descriptors.

We assume that although there is only one individual in the scene, reconstructed shadows as well as noise can coexist, due to poor segmentation. By obtaining the ground marginal distribution

Fig. 3. Example of reconstructed gait sequence, sampled at 2Hz, where each cube represents a voxel.

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 ϕ, and its volume has fully entered into the workspace. The value of ϕ is tuned up accordingly to the average corporal volume for humans and the resolution of the 3D reconstructions. This is described in Section 4.2.

When the volume belonging to a person has been detected, the centroid p of the ground projection is calculated. Then, the volume is moved into a bounding-box of average adult human's size, so that the workspace where the descriptor will be computed is bounded. This process is illustrated in Fig. 4. The estimation of the direction of path is determined by the displacement vector, defined as:

$$\vec{v}_t = p_t - p_{t-1}, \tag{1}$$

where t is the current time, p_t is the centroid's current position, and p_{t-1} is the last known position of the centroid.

The angle of the displacement vector is calculated using the expression:

$$\alpha_t = \arctan \frac{v_{t_y}}{v_{t_x}}. \tag{2}$$

An example of extraction of the displacement vector angle can be seen in Fig. 5, where top projections of the individual can be seen in several moments of the gait, and the principal axis (perpendicular to displacement vector) is represented. The reconstructed volume is rotated about the body vertical axis.

Although we assume a constant walking speed, an individual could vary moderately the walking speed in a certain moment of the
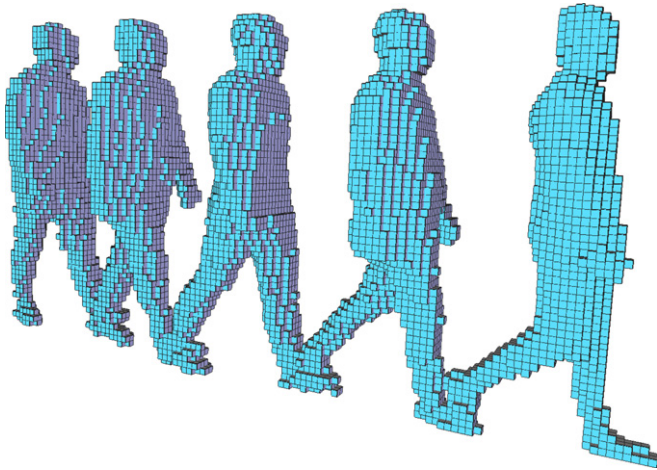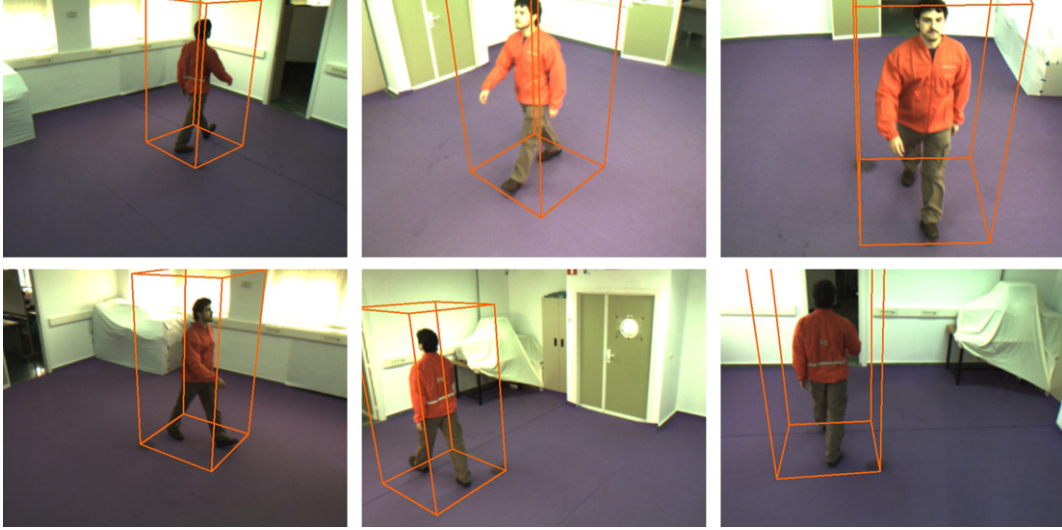
Fig. 4. Bounding-box with size of an average adult human, where the descriptor will be computed on the 3D reconstructed data at time *t*.

gait. It could happen, for example, when the individual is describing a curved closed path.

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

$$\bar{\alpha}_t = \alpha_t \cdot \beta + \bar{\alpha}_{t-1} \cdot (1 - \beta), \tag{3}$$

where

$$\beta = \frac{||\vec{v}_t||}{\max_{i=0\ldots t}\{||\vec{v}_i||\}}. \tag{4}$$

The aim of β is to reduce this noise in the estimation of the alignment angle, giving more or less weight to the current estimation depending on the walking speed. For example, if the walking speed is decreasing and it becomes too small, the magnitude of the displacement vector may not be large enough, causing oscillations in the estimation of the angle. In this case, it would be right to give less importance to the current angle estimation. However, if the walking speed is increasing, it would be more appropriate to give more weight to the current angle estimation. A method to decrease over time the denominator in Eq. (4) should be applied if the gait sequence were too large. Thus, the whole gait sequence can be centered and aligned along the path as it is illustrated in Fig. 6.

3.2. Gait identification

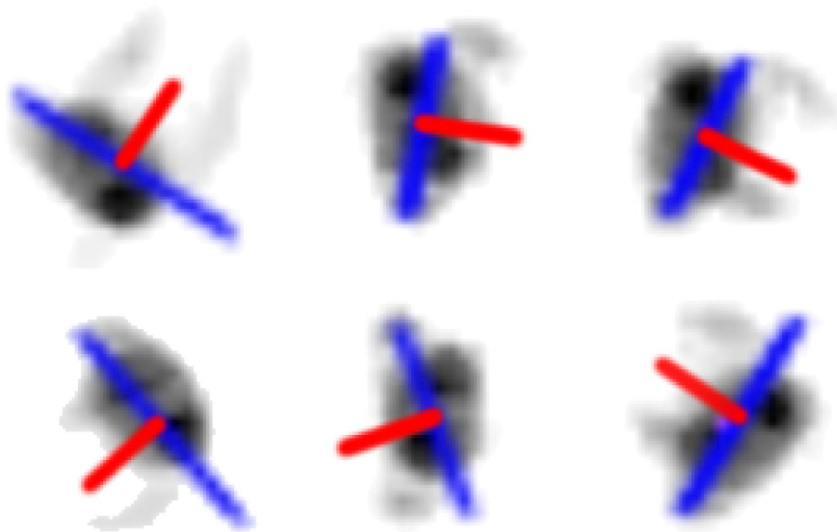The algorithm steps that handle up the gait identification are described below.

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.
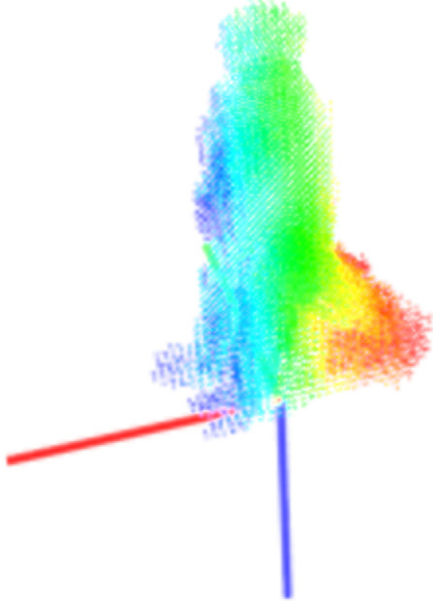
Fig. 6. Union of all aligned 3D volumes over the gait sequence. Different colors represent the distance to the center on the X axis.

3.2.1. Descriptor generation

The first step is the generation of the gait descriptor on the V_t volume. We propose the three following candidate descriptors.

- Cover by Rectangles from frontal volume projection:
  The Cover by Rectangles descriptor, denoted here as CR(S), was proposed by Barnich and Van Droogenbroeck [18]. It is defined as the union of all the largest rectangles that can fit inside of a silhouette S. In Barnich's method, video cameras are placed in hallways to capture longer sequences from the front view of walkers rather than the side view, which results in more gait cycles per gait sequence. The main drawback of this method is the dependence on the viewpoint.
  Each silhouette is then converted to an intra-frame histogram, which compacts the width and height distributions of the set of all the rectangles that can be wedged inside the silhouette. In order to build the histogram, the widths and heights of the rectangles are discretized into M and N bins respectively.
  As the occupation volumes have been aligned to achieve the view independence, a virtual camera can be placed in front of the volumes to obtain projections on which the CR(S) can be computed, as can be seen in Fig. 7 (b).
- Cover by Cubes:
  We propose a new gait descriptor defined as the union of all the cubes with the largest size that can fit into a volume belonging to the person. It is called *Cover by Cubes*.
  Considering V as the 3D volume of a person in a moment of the gait, CC(V) is the union of all cubes of maximum size that can fit in it. The new descriptor deals with three-dimensional domain spatial information, and like *Cover by Rectangles*, it has the following useful properties:

  - The elements of the set overlap each other, introducing redundancy (i.e. robustness).
  - Each element (cube) of CC(V) covers at least one voxel that belongs to no other cube.

- The union of all cubes reconstructs the volume V so that no information is ever lost.

Let $\alpha = \#$CC(V) be the cardinality of the set CC(V). The cubes of CC(V) are indexed with a parameter e, so that $\mathcal{C}_e(e = 1, \ldots, \alpha)$ are the cubes of CC(V). The width, height and depth of $\mathcal{C}_e$ are, respectively, denoted by w_e, h_e and d_e; and they are upper-bounded by w^{max}, h^{max} and d^{max}, so $\forall e$, $w_e \leq w^{max}$, $h_e \leq h^{max}$ and $d_e \leq d^{max}$.

In order to build histograms, the widths, heights and depths of the cubes R_e are discretized into M bins $B_W(i)$, N bins $B_H(j)$ and D bins $B^D(k)$

$$B^W(i) = \left[i\frac{w^{max}}{M}, (i+1)\frac{w^{max}}{M}\right), \tag{5}$$

$$B^H(j) = \left[j\frac{h^{max}}{N}, (j+1)\frac{h^{max}}{N}\right), \tag{6}$$

$$B^D(k) = \left[k\frac{d^{max}}{D}, (k+1)\frac{d^{max}}{D}\right) \tag{7}$$

where $i = 0, \ldots, M-1; j = 0, \ldots, N-1$ and $k = 0, \ldots, D-1$.
Three histograms are defined, $H^W(i)$, $H^H(j)$ and $H^D(k)$:

$$H^W(i) = \frac{1}{\alpha}\#\left\{\mathcal{C}_e|w_e \in B^W(i)\right\}, \tag{8}$$

$$H^H(j) = \frac{1}{\alpha}\#\left\{\mathcal{C}_e|h_e \in B^H(j)\right\}, \tag{9}$$

$$H^D(k) = \frac{1}{\alpha}\#\left\{\mathcal{C}_e|d_e \in B^D(k)\right\}, \tag{10}$$

and the three-dimensional histogram $H^{W \times H \times D}$ as:

$$H^{W \times H \times D}(i,j,k) = \frac{1}{\alpha}\#\left\{\mathcal{C}_e|w_e \in B^W(i), h_e \in B^H(j), d_e \in B^D(k)\right\}. \tag{11}$$

All these histograms are normalized taking into account the number of cubes of maximum size of the volume.

From the four histograms, $H^W(i)$, $H^H(j)$, $H^D(k)$ and $H^{W \times H \times D}(i,j,k)$, the latter is the one that better describes V. However, its dimensionality is proportional to the product of the numbers of bins ($M \times N \times D$), which might be too high, e.g., for embedded systems. To deal with such situation, a composite histogram is proposed, $H^{W+H+D}(l)$, with $l = 0, \ldots, M+N+D-1$ defined as the concatenation of $H^W(i)$, $H^H(j)$ and $H^D(k)$ (marginal distributions).

An example of Cover by Cubes histograms is shown in Fig. 8. In this figure, the joint distribution and the marginal distributions of the histograms can be seen, corresponding to Eqs. (11), (8), (9), and (10).

- Cover by Rectangles from top, side, and frontal volume projections:
  The availability of 3D gait volumes leads us to that think we can use several projections of the gait volumes, instead of just using the frontal projection of them, in order to exploit the 3D information of gait.
  So following the idea about the use of virtual cameras, we also propose a new descriptor based on computing the CR descriptor on the front, side, and top projections of the volume V, as it is shown in Fig. 7. Its concatenation can be denoted as CRP(V).
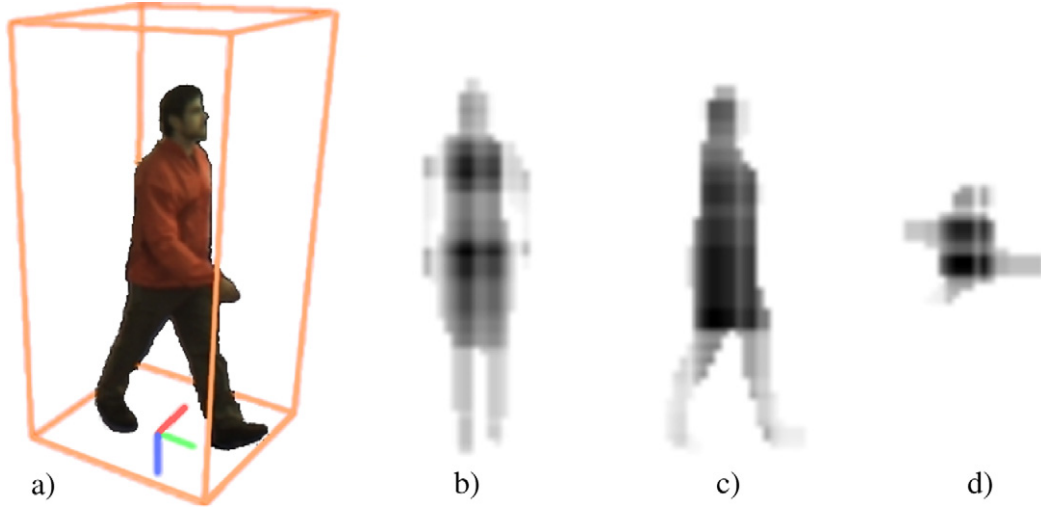
Fig. 7. Cover by Rectangles descriptors computed on the front (b), side (c), and top (d) view projections of a 3D volume reconstruction (a). A gray level of pixel in images (b) and (d) displays the density of rectangles that contain that pixel.

$$\text{CRP}(V) = \text{CR}(side), \text{CR}(front), \text{CR}(top), \tag{12}$$

where *side*, *front* and *top* are lateral, frontal and top rendered projections of *V*, respectively.

3.2.2. Gait signature update

The action immediately prior to the classification by a machine learning algorithm is the generation of the sample or vector of features. A sample is generated at every moment of the walking, which enables synchronous classification. This is known as $\mathscr{G}$ or gait signature, and represents a temporal pattern of movement of the person.

For the CR descriptor, obtained on the front projection of the volume, the signature $\mathscr{G}$ can be obtained by combining a number *L* of successive histograms into a single spatio-temporal (inter-frame) gait signature, as it was described in [18]. The signature can be made by the combination of the marginal or joint distribution of the histograms and it needs to be updated in every time *t*.

However, with the new proposed gait descriptors (CC and CRP) which get information from 3D occupancy volumes instead of getting

it from silhouettes, it is necessary to reformulate the procedure to construct the signature $\mathscr{G}$. The way in which this signature is built depends on the descriptor.

With regard to CC descriptor, the gait *signature* $\mathscr{G}$ relies on temporal series of descriptors obtained from the 3D volumes of a person's gait sequence. So if *t* refers to the current time, and H(i,j,k,t) is the Cover by Cubes descriptor obtained from the volume V_t, we have two possible signatures as follows:

$$\mathscr{G}^{W \times H \times D}(i,j,k,t) = H^{W \times H \times D}(i,j,k,t-(L-1)),\ldots,H^{W \times H \times D}(i,j,k,t) \tag{13}$$

which consist of *n*-uples of *L* consecutive histograms, and a shortened version as:

$$\mathscr{G}^{W+H+D}(o,t) = H^{W+H+D}(o,t-(L-1)),\ldots,H^{W+H+D}(o,t) \tag{14}$$

where $o = 0,\ldots,M+N+D-1$.

Similarly, for CRP descriptors, the gait signature $\mathscr{G}$ can be also composed by aggregating, on a sliding window, *L* CRP descriptors (joint or marginal distribution of the histograms computed on rendered side, top and front projections of *V*).
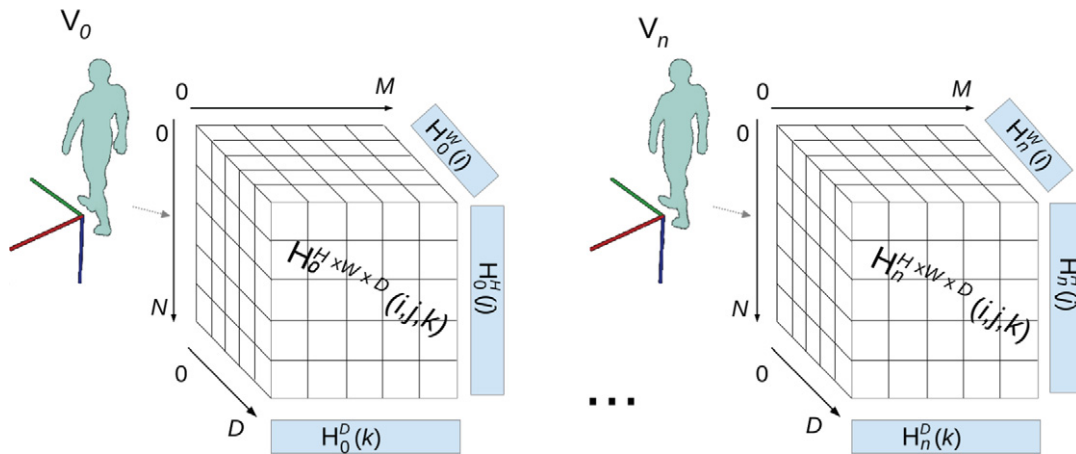

Fig. 8. Cover by Cubes histograms, where *n* refers to the length of the gait sequence. $M \times N \times D$ is the number of bins of the histogram $H^{W \times H \times D}$, and the blue rectangles represent marginal distributions.

3.2.3. Classification

The gait signature obtained at time t is the feature vector used for recognition. Each of these feature vectors is 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 adopt a Support Vector Machine (SVM) [39] for training and classification.

Our recognition algorithm provides the identity of the person as soon as possible, without having to split the gait sequence into gait cycles. This makes our method less restrictive compared to other techniques from the literature. A possibly different class label can be produced for each new gait signature, on the basis of L previous volumes.

To smooth and reinforce the results over time, a majority vote policy over a sliding temporal window of size W is used. As the gait signature information is computed on L previous volumes, the use of this window causes a delay of $L + W$ frames in obtaining the identity. Fig. 9 shows an example of majority voting system over a sliding temporal window, with $L = 5$ and $W = 3$.

4. Experiments and discussion

In this section we start by describing the used datasets, and then we present the experimental results.

4.1. Datasets description

To perform a 3D reconstruction by the SfS algorithm, the dataset must be multiview and calibration information have to be provided. Two datasets have been used to carry out the experiments, the "AVA Multi-View Dataset for Gait Recognition (AVAMVG)"[2] [40] and the "Kyushu University 4D Gait Database (KY4D)"[3] [26].

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 distribution shown in the diagram of Fig. 10.

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. 11, in which several subjects are walking along different paths, from multiple viewpoints. Calibration parameters for the cameras of AVAMVG have been obtained with Aruco library [41].

With respect to KY4D Gait Database, it is composed of sequential 3D reconstructions 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 reconstructions of subjects, we have reconstructed them with the same SfS method and resolution parameters used for reconstructing the 3D AVAMVG models.

As far as we know, there are other well-known multi-camera databases, as CMU Motion of Body (MoBo) Database [42] and CASIA Dataset B [43]. However, since these databases do not include information on camera parameters, 3D reconstructions of walking people cannot be obtained. Therefore, we did not use these databases in the experiments of the present study.
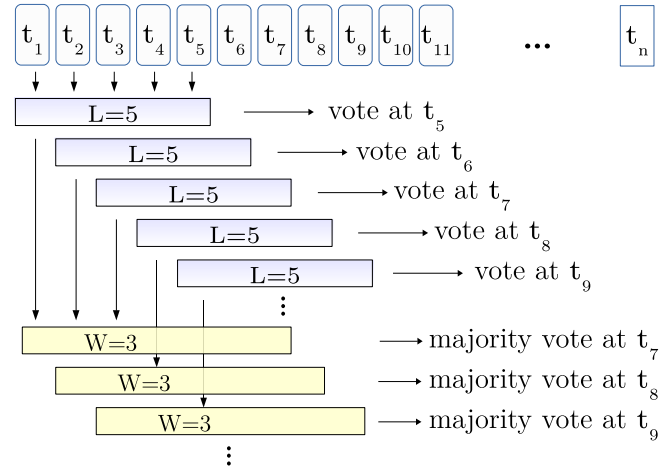
Fig. 9. Majority vote policy over a sliding temporal window. In this example, the size of the signature is set to $L = 5$, and the size of the voting window is set to $W = 3$.

4.2. Experimental results

In this section, we present the results of multiple experiments run on both gait datasets. 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 consider that a voxel size of $0.27 \times 10^{-4} \mathrm{m}^3$ is enough to get detailed 3D human reconstructions.

The average corporal volume for humans is $66.4L = 0.6640 \times 10^{-1} \mathrm{m}^3$ measured by the water displacement method in 521 people aged $17 - 51$ years [44]. Using a voxel size of $0.27 \times 10^{-4} \mathrm{m}^3$, the number of voxels belonging to a person in the 3D volume should be
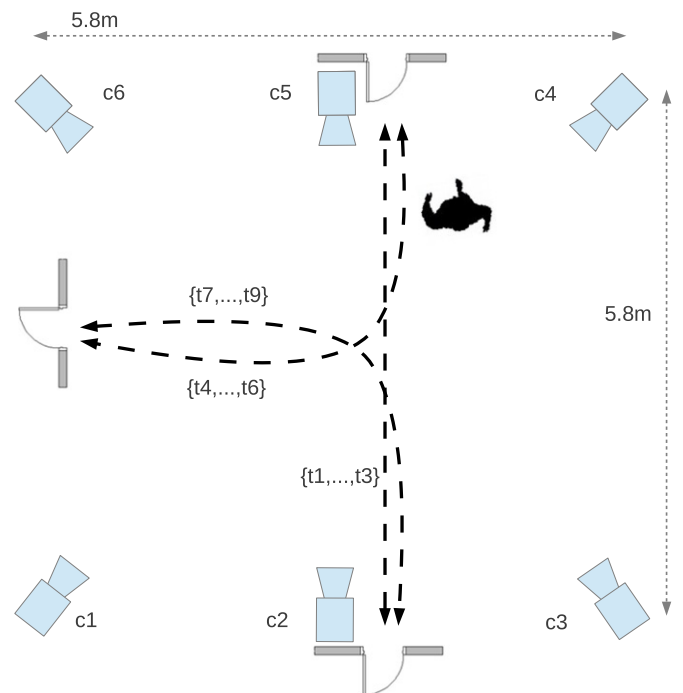

Fig. 10. Workspace setup used by AVAMVG Dataset, where $\{c1, \ldots, c6\}$ represent the set of cameras of the multiview dataset and $\{t1, \ldots, t9\}$ represent the different trajectories followed by each actor of the dataset.

Fig. 11. Example of AVAMVG multiview dataset. People walking in different directions, from multiple points of view.

about 2459. With a value of $1 \times 10^3 < \phi < 2459$ (see Section 3.1.2) the system should be able to detect an adult human.

Regarding the volume alignment, Table 1 shows the mean error in the estimation angle for each trajectory of the KY4D dataset. We also report the 95% confidence interval on the mean, assuming that the data are normally distributed.

As it was proved in [18], the number of silhouettes (volumes in our case) aggregated in a single gait signature can be set in $L = 20$, because we have a rate of 25 volumes per second, and $L = 20$ responds to a signature of about 1 s which matches the length of a gait cycle.

The decision to build the gait signatures with joint or marginal distributions of histograms depends on the amount of training data and memory available for the classification process. If all the other parameters are kept unchanged, the use of joint distributions should lead to get better results. However, the dimensionality of the corresponding feature space using joint distributions is $M \times N \times L$ for CR [18], $M \times N \times 3 \times L$ for CRP, or $M \times N \times D \times L$ for CC, which is too expensive to compute. Alternatively, the dimensionality of the feature space using marginal distributions of histograms is $(M+N) \times L$ for CR, $(M+N) \times 3 \times L$ for CRP and $(M+N+D) \times L$ for CC. For example, for a value of $M = N = D = 25$ and $L = 20$, the feature space of the CC descriptor has a dimensionality of 312500 using joint distributions, compared to 1500 features using marginal distributions.

In several previous experiments, we noted that if we use joint distributions and SVM, the accuracy is lower than if we use marginal distributions. Maybe it can occur because the statistical significance of the joint distribution of histograms is much lower than the statistical significance of the marginal distributions of them. Moreover, may be impracticable to compute the joint distribution of histograms such as CC or CRP because of the high dimensionality. Therefore, we focus our experimentation on the use of marginal distributions of the histograms.

In order to determine M, N and D (the number of bins), we tested values ranging from 5 to 25, with step 5 on both AVAMVG and KY4D

gait databases. It was observed that large values of M, N or D generally lead to better performance. However, the performance saturates with 20 bins and above, depending on the descriptor.
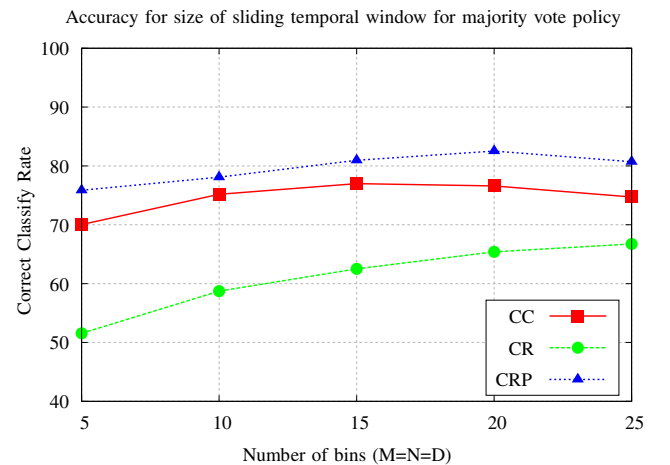
Fig. 12. Performance of each descriptor for different histogram sizes on AVAMVG gait database. The sliding temporal window for majority voting policy is disabled ($W = 1$).
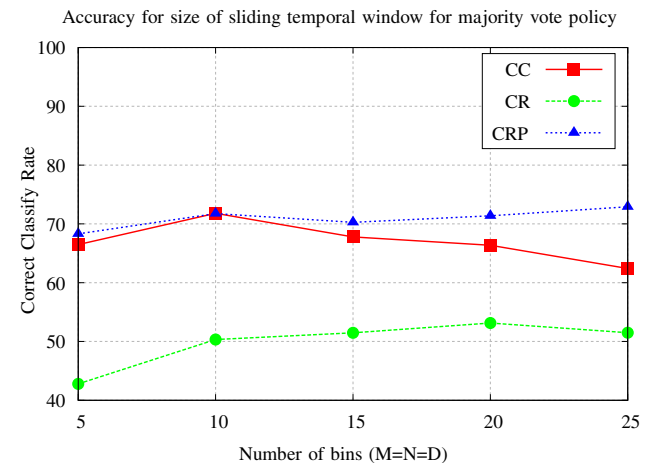
Fig. 13. Performance of each descriptor for different histogram sizes on KY4D gait database. The sliding temporal window for majority voting policy is disabled ($W = 1$).

Table 1
Mean error in the angle estimation (degrees) of the volume alignment step, for each trajectory of KY4D dataset. The 95% confidence interval on the mean is also shown.

Straight paths				Curved paths	
Tr. 1	Tr. 2	Tr. 3	Tr. 4	Tr. 5	Tr. 6
0.73 ± 0.04	0.76 ± 0.06	0.74 ± 0.05	0.76 ± 0.04	2.16 ± 0.34	2.15 ± 0.30

Fig. 14. Performance of each descriptor on AVAMVG database for different lengths for the majority voting window.
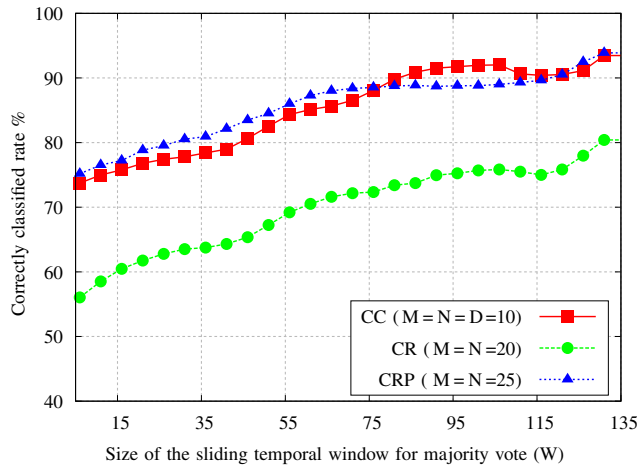
Fig. 15. Performance of each descriptor on KY4D database for different lengths for the majority voting window.

Depending on the size of the training dataset and the resolution of the 3D reconstructed volumes, the statistical significance of all the bins of the histograms needs to be taken into account. The first series of experiments consisted in determining the appropriate number of bins for each descriptor. For the sake of simplicity, we restricted to the case where $M=N$ or $M=N=D$, and disabled the majority vote on the previous W frames (or equivalently set W to 1).

We use a leave-one-out cross-validation strategy. Regarding experiments on the AVAMVG dataset, each fold is composed by a tuple formed by a set of 20 sequences (one sequence per actor) for

Table 3

Correct classification rate on KY4D. These results correspond to a cross-validation experiment. Each column corresponds to a test trajectory, using the remaining trajectories as training set. Each row corresponds to a different method. Each entry contains the percentage of correct recognition for each tuple trajectory-setup.

Method	Tr. 1	Tr. 2	Tr. 3	Tr. 4	Tr. 5	Tr. 6	Mean
CRP($M = N = 25$, $W = 135$)	92.6%	100%	100%	97.5%	84.9%	87.8%	93.8%
CC($M = N = D = 10$, $W = 135$)	97.5%	97.5%	95.1%	97.5%	82.9%	90%	93.4%
Seely et al. [21]	95.1%	100%	97.5%	100%	68.2%	72.5%	88.8%
Ariyanto and Nixon [22]	41.4%	41.4%	43.9%	53.6%	19.5%	17.5%	36.2%

testing, and by the remaining eight sequences 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 sequence per actor) and by the remaining five sequences of each actor (i.e. 42×5 sequences) for training.

We use a C-SVC SVM, which allows imperfect separation of classes with penalty multiplier C for outliers. Several SVM kernels were tested, and finally we selected Radial Basis Function 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. We report the improvement with respect to different kernels in Table 4.

Fig. 12, shows the performance of each descriptor with $L = 20$ and different histogram sizes on the AVAMVG dataset. We use marginal distributions of histogram and we get the best results with $M = N = 25$ for the CR descriptor, $M = N = 20$ for the CRP descriptor, and $M = N = D = 15$ for the CC descriptor.
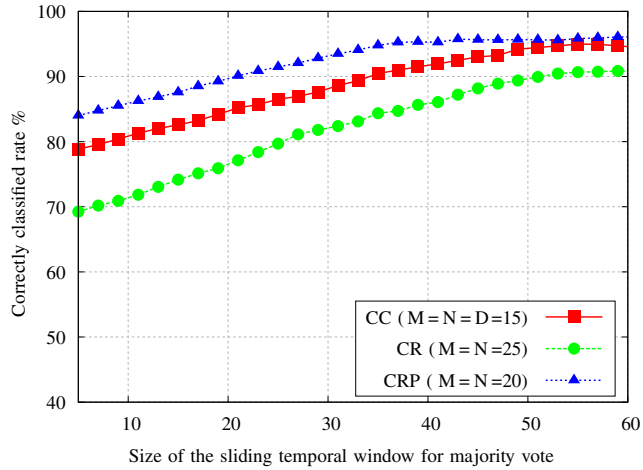
On the other hand, Fig. 13 shows the performance of each descriptor with different histogram sizes, applied on the KY4D gait dataset. We use marginal distributions, and in this case, we get the best results with $M = N = 20$ for the CR descriptor, $M = N = 25$ for the CRP descriptor, and $M = N = D = 10$ for the CC descriptor. In this experiment, for the sake of simplicity, we disabled the sliding temporal window for majority vote.

The second series of experiments that were carried out consisted in determining the optimum size of the sliding temporal window for majority voting. In Figs. 14 and 15 we show how the accuracy increases with respect to the size of the sliding temporal window for majority voting on both datasets, using the histogram sizes selected in the previous experiment.

The sliding temporal window of majority voting stage improves the performance of the method with any of the three proposed descriptors for both datasets. Nevertheless, the size of the sliding temporal window for voting is limited by the number of available gait signatures for each sequence.

For the AVAMVG, with CR descriptors applied on frontal volume projections, we obtain a maximum accuracy of 90.8%. Nevertheless, using the CC descriptor computed on the entire volume, the accuracy is about 94.5% and finally, with the CRP descriptor, the system was able to correctly identify up to 96.1% of subjects. On the other hand,

Table 2

Correct classification rate on AVAMVG. These results correspond to a cross-validation experiment. Each column corresponds to a test trajectory, using the remaining trajectories as training set. Each row corresponds to a different method. Each entry contains the percentage of correct recognition for each tuple trajectory-setup.

Method	Tr. 1	Tr. 2	Tr. 3	Tr. 4	Tr. 5	Tr. 6	Tr. 7	Tr. 8	Tr. 9	Mean
CRP($M = N = 20, W = 60$)	100%	88%	100%	99.3%	99.2%	97.7%	96.2%	84.8%	100%	96.1%
CC($M = N = D = 15, W = 60$)	100%	96%	75.5%	98.6%	87.8%	99.1%	99.5%	94%	100%	94.5%
Seely et al. [21]	90%	80%	94.7%	90%	60%	100%	80%	84.2%	90%	85.4%
Ariyanto and Nixon [22]	55%	45%	52.6%	45%	26.3%	35%	35%	31.5%	40%	40.6%

Table 4
Correct classification rate on KY4D with respect to different kernels on KY4D dataset.

Descriptor	Linear	Polynomial	Sigmoid	RBF
$CR(M = N = 20, W = 135)$	78.3%	78.4%	78.0%	80.4%
$CC(M = N = D = 10, W = 135)$	89.3%	91.8%	92.2%	93.4%
$CRP(M = N = 25, W = 135)$	90.2%	91.5%	90.1%	93.8%

Table 5
Comparative of recognition results on AVAMVG dataset [40] . Each row corresponds to a different method. The second row indicates the training trajectory. The third and fourth columns indicate the tested trajectory. The fifth column shows the average value of the recognition rate for the curved paths.

Method	Training path	Curve t4	Curve t7	Mean
$CRP(M = N = 20, W = 60)$	Straight {t1,t2,t3}	45%	81.5%	63.2%
Seely et al. [21]	Straight {t1,t2,t3}	55%	70%	62.5%
Iwashita et al. [26]	Straight {t1,t2,t3}	35.1%	37.7%	36.4%
Ariyanto and Nixon [22]	Straight {t1,t2,t3}	30%	30%	30%

Table 6
Comparative of recognition results on KY4D gait dataset [26] . Each row corresponds to a different method. The second column indicates the training trajectory. The third and fourth columns indicate the tested trajectory. The fifth column shows the average value of the recognition rate for the curved paths.

Method	Training trajectories	Curve 1	Curve 2	Mean
$CRP(M = N = 25, W = 135)$	Straight {1,2,3,4}	60.4%	81.7%	71%
Iwashita et al. [26]	Straight {1,2,3,4}	61.9%	71.4%	66.6%
Seely et al. [21]	Straight {1,2,3,4}	19.5%	35%	27.2%
Ariyanto and Nixon [22]	Straight {1,2,3,4}	12.1%	15%	13.5%

on the KY4D dataset we obtain a maximum accuracy of 80.4% for CR, 93.4% for CC and 93.8% for CRP descriptors.

Tables 2 and 3 show detailed results for the leave-one-out experiment on AVAMVG and KY4D datasets respectively. We compare the accuracy of our approach for CRP and CC signatures with the accuracy of state-of-art methods such as [21,22] on AVAMVG and KY4D datasets. For the comparison with Ariyanto and Nixon we used the best kinematics features proved in [22], whereas for the case of Seely et al. [21] we have used the side-on, front-on, top-down average silhouettes. Since these methods are not designed to cope with curved trajectories, we have aligned the gait volumes along the path (see Section 3.1.2). The resolution of the reconstructed volumes was the same for all cases.

The third series of experiments that we performed consisted in testing our method, which allows completely free trajectories, with the method presented in [26], for identification of people walking along curved trajectories. Comparative of recognition results on both datasets are shown in Tables 5 and 6. In this experiment, we used the straight trajectories for training, and the curved trajectories for testing.

As we can observe in Tables 4 and 5, the performance of compared methods is dropped down when the training set does not contain curved trajectories. Moreover, we have noticed a decrease of performance of the method presented in [26] when it is trained with the straight paths and tested with the curves of AVAMVG. We think it may be due to the low number of cameras of AVAMVG and therefore to the quality of the 3D reconstructions. Besides that, in the AVAMVG dataset, depending on the viewpoint and performed trajectory, people appear at diverse scales, even showing partially occluded body parts.

We think that when the subject walks on a curved path, the gait pattern is consequently modified, as we can see in Fig. 16. For this reason, we consider that it is not entirely correct to train the classifier of our model-free approach with straight paths only.

5. Conclusions

This paper has proposed a method to recognize walking humans independently of the viewpoint and regardless direction changes. The method focuses on capturing 3D morphological and structural information from volumetric reconstructions of the gait. The main contribution is that the method achieves a good recognition rate on completely unconstrained paths, allowing direction changes, in contrast to others view-independent approaches where the view change is restricted to a few angles. In our method, the individual can walk freely in the scene without adversely affect to the recognition.

For this purpose, it was designed a mechanism of person detection and gait alignment based on 3D reconstructions. In order to extract information from the 3D volumes, three gait morphological descriptors are proposed. The first one is the Cover by Rectangles (CR) [18] applied on rendered front projections of the gait volume. The second is composed by an aggregation of three Cover by Rectangles descriptors computed on the top, side, and frontal projections of the gait volume (CRP). Lastly, the third new proposed descriptor is called Cover by Cubes (CC) and it is defined as the union of all the cubes with the largest size that can fit into a gait volume of a person.
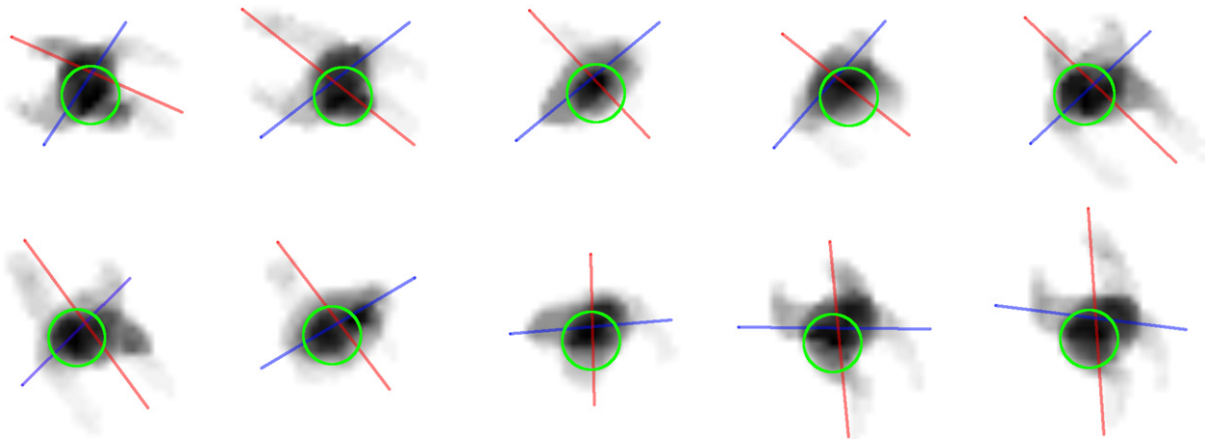

Fig. 16. Example of a curved gait cycle. We show several ground marginal distributions of occupied voxels. The velocity vector is represented by a red line, the blue line represents the torso main axis, and the position of the head is represented by a green circle. We can note that in a curved trajectory, the person rotates his/her torso and leans towards the walking direction.

The experimental results show that the CRP descriptor is the most reliable for using with our gait recognition method, providing good results in both AVAMVG and KY4D gait databases. The experimental results also show what is the optimal size for the histograms of each descriptor on each dataset. Finally, by using a majority vote policy on a sliding temporal window, the system is able to correctly identify up to 96% of the subjects of the AVAMVG gait database and nearly 94% of subjects of the KY4D dataset.

Acknowledgments

References

[1] S. Yu, D. Tan, T. Tan, Modelling the effect of view angle variation on appearance-based gait recognition, in: P.J. Narayanan, S.K. Nayar, H.-Y. Shum (Eds.), Computer Vision - ACCV 2006, Lecture Notes in Computer Science 3851, Springer Berlin Heidelberg 2006, pp. 807–816. ISBN 978-3-540-31219-2. http://dx.doi.org/10.1007/11612032_81.

[2] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 34 (3) (2004) 334–352. ISSN 1094-6977. ISSN. http://dx.doi.org/10.1109/TSMCC.2004.829274.

[3] A. Kale, N. Cuntoor, B. Yegnanarayana, A.N. Rajagopalan, R. Chellappa, Gait analysis for human identification, Audio- and Video-based Biometric Person Authentication 2688, Springer Berlin Heidelberg 2003, pp. 706–714. ISBN 978-3-540-40302-9.

[4] M.-H. Cheng, Meng-F. Ho, C.-L. Huang, Gait analysis for human identification through manifold learning and {HMM}, Pattern Recogn. 41 (8) (2008) 2541–2553. ISSN 0031-3203. ISSN. http://dx.doi.org/10.1016/j.patcog.2007.11.021.

[5] 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. ISSN 0162-8828. ISSN. http://dx.doi.org/10.1109/TPAMI.2003.1251144.

[6] S. Hong, H. Lee, E. Kim, Automatic gait recognition using width vector mean, Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on, 2009. pp. 647–650. http://dx.doi.org/10.1109/ICIEA.2009.5138285.

[7] S.D. Mowbray, M.S. Nixon, Automatic gait recognition via Fourier descriptors of deformable objects, in: J. Kittler, M.S. Nixon (Eds.), Audio- and Video-based Biometric Person Authentication, Lecture Notes in Computer Science> 2688, Springer Berlin Heidelberg 2003, pp. 566–573. ISBN 978-3-540-40302-9. http://dx.doi.org/10.1007/3-540-44887-X_67.

[8] C.P. Lee, A.W.C. Tan, S.C. Tan, Gait recognition via optimally interpolated deformable contours, Pattern Recogn. Lett. 34 (6) (2013) 663–669. ISSN 0167-8655. ISSN. http://dx.doi.org/10.1016/j.patrec.2013.01.013.

[9] S.D. Choudhury, T. Tjahjadi, Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors, Pattern Recogn. 45 (9) (2012) 3414–3426. ISSN 0031-3203. ISSN best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011). http://dx.doi.org/10.1016/j.patcog.2012.02.032.

[10] J. Han, B. Bhanu, Individual recognition using gait energy image, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2) (2006) 316–322. ISSN 0162-8828. ISSN. http://dx.doi.org/10.1109/TPAMI.2006.38.

[11] T.H.W. Lam, K.H. Cheung, J.N.K. Liu, Gait flow image: a silhouette-based gait representation for human identification, Pattern Recogn. 44 (4) (2011) 973–987. ISSN 0031-3203. ISSN. http://dx.doi.org/10.1016/j.patcog.2010.10.011.

[12] C. Rougier, E. Auvinet, J. Meunier, M. Mignotte, J.A de Guise, Depth energy image for gait symmetry quantification, Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, 2011. pp. 5136–5139. ISSN 1557-170X. ISSN. http://dx.doi.org/10.1109/IEMBS.2011.6091272.

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

[14] P. Chattopadhyay, S. Sural, J. Mukherjee, Frontal gait recognition from occluded scenes, Pattern Recogn. Lett. 63 (2015) 9–15. ISSN 0167-8655. ISSN. http://dx.doi.org/10.1016/j.patrec.2015.06.004.

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

[16] 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, Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on, 2012. pp. 1–8. http://dx.doi.org/10.1109/DICTA.2012.6411694.

[17] J. Ryu, S. Kamata, Front view gait recognition using Spherical Space Model with Human Point Clouds, Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011. pp. 3209–3212. ISSN 1522-4880. ISSN. http://dx.doi.org/10.1109/ICIP.2011.6116351.

[18] O. Barnich, Ma. Van Droogenbroeck, Frontal-view gait recognition by intra- and inter-frame rectangle size distribution, Pattern Recogn. Lett. 30 (10) (2009) 893–901. ISSN 0167-8655. ISSN. http://dx.doi.org/10.1016/j.patrec.2009.03.014.

[19] G. Shakhnarovich, L. Lee, T. Darrell, Integrated face and gait recognition from multiple views, Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 1, 2001. pp. 439–446. ISSN 1063-6919. ISSN. http://dx.doi.org/10.1109/CVPR.2001.990508.

[20] A. Laurentini, The visual hull concept for silhouette-based image understanding, IEEE Trans. Pattern Anal. Mach. Intell. 16 (2) (1994) 150–162. ISSN 0162-8828. ISSN. http://dx.doi.org/10.1109/34.273735.

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

[22] G. Ariyanto, M.S. Nixon, Model-based 3D gait biometrics, Biometrics (IJCB), 2011 International Joint Conference on, 2011. pp. 1–7. http://dx.doi.org/10.1109/IJCB.2011.6117582.

[23] K. Yamauchi, B. Bhanu, H. Saito, Recognition of walking humans in 3D: initial results, Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, 2009. pp. 45–52. ISSN 2160-7508. ISSN. http://dx.doi.org/10.1109/CVPRW.2009.5204296.

[24] W. Kusakunniran, Q. Wu, J. Zhang, H. Li, Gait recognition under various viewing angles based on correlated motion regression, IEEE Trans. Circuits Syst. Video Technol. 22 (6) (2012) 966–980. ISSN 1051-8215. ISSN. http://dx.doi.org/10.1109/TCSVT.2012.2186744.

[25] 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. ISSN 0262-8856. ISSN. http://dx.doi.org/10.1016/j.imavis.2008.11.008.

[26] Y. Iwashita, K. Ogawara, R. Kurazume, Identification of people walking along curved trajectories, Pattern Recogn. Lett. 48 (2014) 60–69. ISSN 0167-8655. ISSN celebrating the life and work of Maria Petrou. http://dx.doi.org/10.1016/j.patrec.2014.04.004.

[27] A. Kale, A. Sundaresan, A.N. Rajagopalan, N.P. Cuntoor, A.K. Roy-Chowdhury, V. Kruger, R. Chellappa, Identification of humans using gait, IEEE Trans. Image Process. 13 (9) (2004) 1163–1173. ISSN 1057-7149. ISSN. http://dx.doi.org/10.1109/TIP.2004.832865.

[28] 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. ISSN 0031-3203. ISSN. http://dx.doi.org/10.1016/j.patcog.2009.05.006.

[29] S. Jeong, T.-h. Kim, J. Cho, Gait recognition using description of shape synthesized by planar homography, J. Supercomput. 65 (1) (2013) 122–135. ISSN 0920-8542. ISSN. http://dx.doi.org/10.1007/s11227-013-0897-8.

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

[31] M. Goffredo, I. Bouchrika, J.N. Carter, M.S. Nixon, Self-calibrating view-invariant gait biometrics, IEEE Trans. Syst. Man Cybern. B Cybern. 40 (4) (2010) 997–1008. ISSN 1083-4419. ISSN. http://dx.doi.org/10.1109/TSMCB.2009.2031091.

[32] G. Zhao, G. Liu, H. Li, M. Pietikainen, 3D gait recognition using multiple cameras, Automatic face and gesture recognition, 2006. FGR 2006. 7th International Conference on, 2006. pp. 529–534. http://dx.doi.org/10.1109/FGR.2006.2.

[33] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, S. Fotopoulos, A framework for gait-based recognition using Kinect, Pattern Recogn. Lett. 68, Part 2 (2015) 327–335. ISSN 0167-8655. ISSN special Issue on "Soft Biometrics". http://dx.doi.org/10.1016/j.patrec.2015.06.020.

[34] N. Liu, Y.-P. Tan, View invariant gait recognition, Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, 2010. pp. 1410–1413. ISSN 1520-6149. ISSN. http://dx.doi.org/10.1109/ICASSP.2010.5495466.

[35] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Gait recognition using a view transformation model in the frequency domain, in: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision — ECCV 2006, Lecture Notes in Computer Science 3953, Springer Berlin Heidelberg 2006, pp. 151–163. ISBN 978-3-540-33836-9. http://dx.doi.org/10.1007/11744078_12.

[36] W. Kusakunniran, Q. Wu, H. Li, J. Zhang, Multiple views gait recognition using View Transformation Model based on optimized Gait Energy Image, Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009. pp. 1058–1064. http://dx.doi.org/10.1109/ICCVW.2009.5457587.

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

[38] L. Díaz-Más, R. Muñoz-Salinas, F.J. Madrid-Cuevas, R. Medina-Carnicer, Shape from silhouette using Dempster–Shafer theory, Pattern Recogn. 43 (6) (2010) 2119–2131. ISSN 0031-3203. ISSN. http://dx.doi.org/10.1016/j.patcog.2010.01.001.

[39] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Disc. 2 (2) (1998) 121–167. ISSN 1384-5810. ISSN. http://dx.doi.org/10.1023/A:1009715923555.

[40] D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jimínez, R. Muñoz-Salinas, The AVA multi-view dataset for gait recognition, Activity Monitoring by Multiple Distributed Sensing Springer International Publishing 2014, pp. 26–39. ISBN 978-3-319-13322-5. http://dx.doi.org/10.1007/978-3-319-13323-2_3.

[41] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, M.J. Marín-Jiménez, Automatic generation and detection of highly reliable fiducial markers under occlusion, Pattern Recogn. 47 (6) (2014) 2280–2292. ISSN 0031-3203. ISSN. http://dx.doi.org/10.1016/j.patcog.2014.01.005.

[42] R. Gross, J. Shi, The CMU Motion of Body (MoBo) Database, Robotics Institute, Pittsburgh, PA, 2001.

[43] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, Image Processing (ICIP), 2011 18th IEEE International Conference on, 2011. pp. 2073–2076. ISSN 1522-4880. ISSN. http://dx.doi.org/10.1109/ICIP.2011.6115889.

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

Chapter 6

Entropy Volumes

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

This chapter proposes a new method to recognize people walking on unconstrained paths, even if they walk along curved trajectories or change direction [109]. 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.

Some entropy-based algorithms [51, 75] 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 is synchronous and provides the name of 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 leaves the field of view very soon, so the length of the gait sequence is restricted. Since our 3D gait volumes are centered with respect to a global reference system and aligned along their way, we can get more rendered images from the volumes along a sequence.

Main publications associated to this chapter:

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications*, 26(7):1079-1094, 2015. ISSN 1432-1769. doi: 10.1007/s00138-015-0707-9. url: `http://link.springer.com/article/10.1007%2Fs00138-015-0707-9`

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, and R. Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In *Activity Monitoring by Multiple Distributed Sensing, Lecture Notes in Computer Science*, pages 26-39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3. url: `http://link.springer.com/chapter/10.1007%2F978-3-319-13323-2_3`

CrossMark

ORIGINAL PAPER

Entropy volumes for viewpoint-independent gait recognition

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

Abstract Gait as biometrics has been widely used for
human identification. However, direction changes cause dif-
ficulties 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
i52lofed@uco.es

[1] Department of Computing and Numerical Analysis,
Maimónides Institute for Biomedical Research (IMIBIC),
University of Córdoba, Edificio Einstein, Campus de
Rabanales, 14071 Córdoba, Spain

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 an
unique gait pattern, which means it can be used as a biometric
indicator [10]. 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 clas-
sified 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 colour and texture infor-
mation 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 sub-
ject 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
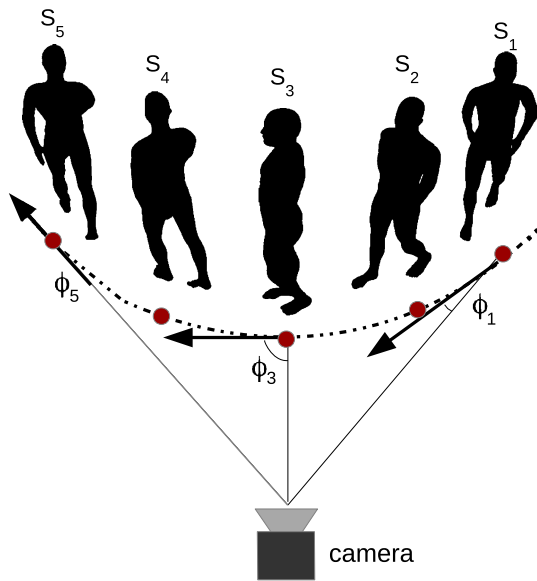cycle. This is shown in Fig. 1.

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 characterized 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 Sect. 4. Finally, we conclude this paper in Sect. 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 [18,40], and through their Fourier descriptors [29,33]. In [8], hidden Markov models have been trained from feature vectors derived from binary silhouettes.

In addition, in [9] 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 [35] consists in the average of frontal depth silhouettes for a gait cycle. In [17], a new feature called depth gradient histogram energy image 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 [7].

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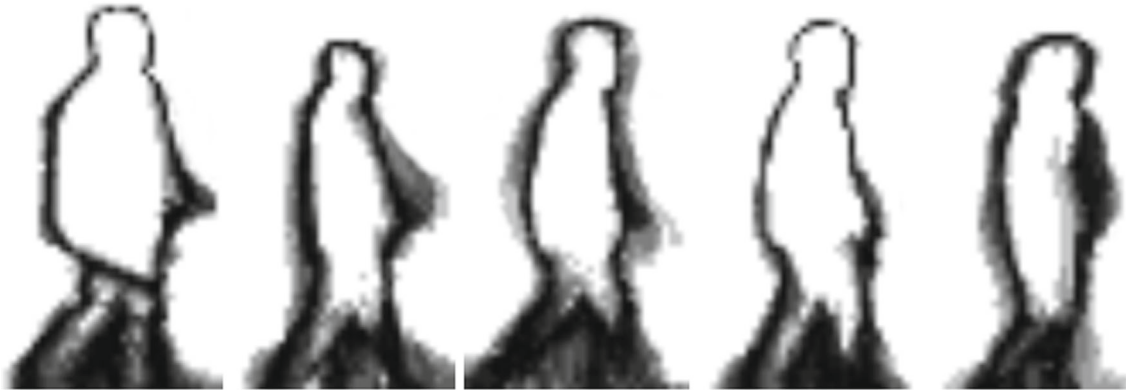
Fig. 2 Several examples of GEnI, computed over a gait cycle. The *gray* level represents the entropy value in a pixel. As can be seen, legs and arms have high gait entropy value, whereas static areas as torso have low values of entropy

In [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. Figure 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 [3,21,36]. 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. [24] 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.

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 assumes that subjects walk along a straight line segment.

In [6] 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 (LDA). 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 leaves 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 (SfS) algorithm [11].
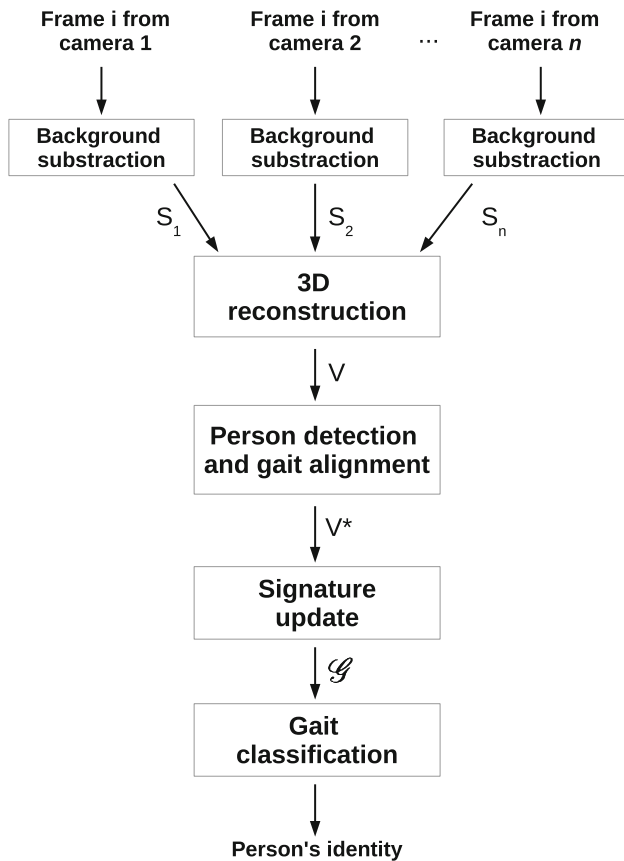
Frame i from camera 1   **Frame i from camera 2**   ···   **Frame i from camera _n_**

↓   ↓   ↓

| Background substraction | Background substraction | Background substraction |

S_1 ↘ S_2 ↓ S_n ↙

3D reconstruction

↓ V

Person detection and gait alignment

↓ V*

Signature update

↓ $\mathcal{G}$

Gait classification

↓

Person's identity

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

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.

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 colour 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 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 [11], 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. Figure 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.
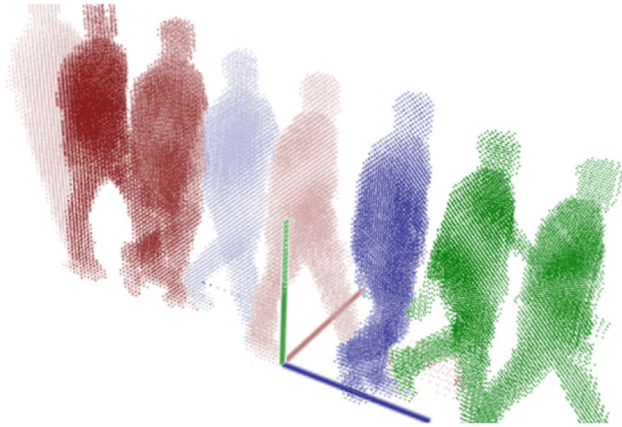
Fig. 4 Example of reconstructed gait sequence, sampled at 2 Hz, where each *point* represents the *centre* of a *squared* voxel. The time instant is represented by different *colours* (colour figure online)

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:

$$\mathbf{v}_t = p_t - p_{t-1}, \tag{2}$$

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:

$$\alpha_t = \arctan \frac{v_{t_y}}{v_{t_x}}. \tag{3}$$

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
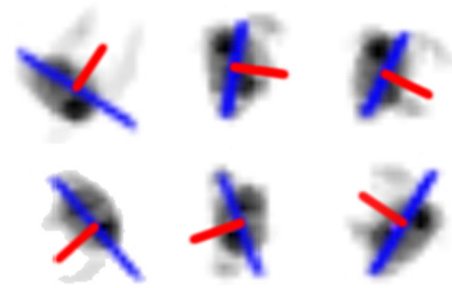

Fig. 5 Displacement vector (*red line*) of the individual is computed at each time. The principal axis (*blue line*) is *perpendicular* to the displacement vector (colour figure online)

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:

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

where

$$\beta = \frac{||\mathbf{v}_t||}{\max_{i=0\dots t} \{||\mathbf{v}_i||\}}. \tag{5}$$

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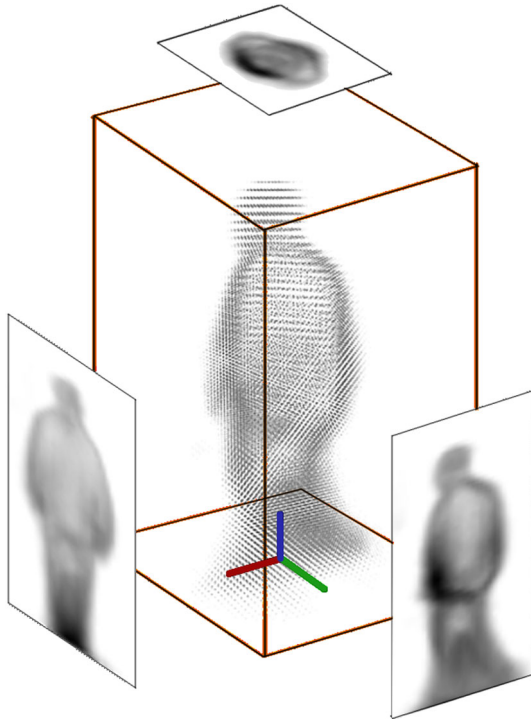
Fig. 6 In the *centre of the image*, a 3D representation of GEnV computed over L 3D-reconstructed and aligned volumes of an individual. Voxels are represented as *points*. Intensity on *gray* level represent the entropy value corresponding to that voxel. Marginal distributions of GEnV are also shown

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 [1,2], we use a definition of entropy based on the exponential behaviour of information-gain, proposed and justified by Pal and Pal [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:

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

and p_0 as:

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

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:

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

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

Gait entropy volume (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:

$$\mathscr{G}^{\text{GEnV}} = \text{GEnV}(x, y, z). \qquad (9)$$

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:

- $\mathscr{G}_F^{\text{GEnV}}(z, y) = \frac{1}{N_x} \sum_{x=0}^{N_x} \text{GEnV}(x, y, z),$
- $\mathscr{G}_S^{\text{GEnV}}(z, x) = \frac{1}{N_y} \sum_{y=0}^{N_y} \text{GEnV}(x, y, z),$
- $\mathscr{G}_T^{\text{GEnV}}(x, y) = \frac{1}{N_z} \sum_{z=0}^{N_z} \text{GEnV}(x, y, z).$

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:

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

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

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

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:

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

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

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

$$\mathscr{G}_{\mathrm{view}}^{\mathrm{desc}} \qquad\qquad (12)$$

where desc : {GEnV, 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.
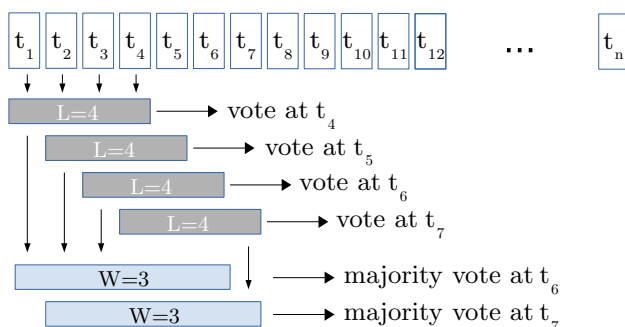

Fig. 7 Majority vote policy over a sliding temporal window. In the example, the size of the signature is set to $L = 4$, and the size of the voting window is set to $W = 3$. This means that the probability of occupation p_1 (see Eq. 6) is calculated over the previous $L = 4$ volumes

Before training a SVM model, we adopt the subspace component and discriminant analysis, based on principal component analysis (PCA) and 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.

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 colour 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 $\{t1, \ldots, t3\}$ are straight while $\{t4, \ldots, t9\}$ 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.

As far as we know, there are others well-known multi-camera databases, such as the CMU motion of body database [14] and CASIA Dataset B [42]. However, since these data-

---

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

[2] Publicly available at: http://robotics.ait.kyushu-u.ac.jp/research-e.php?content=db.
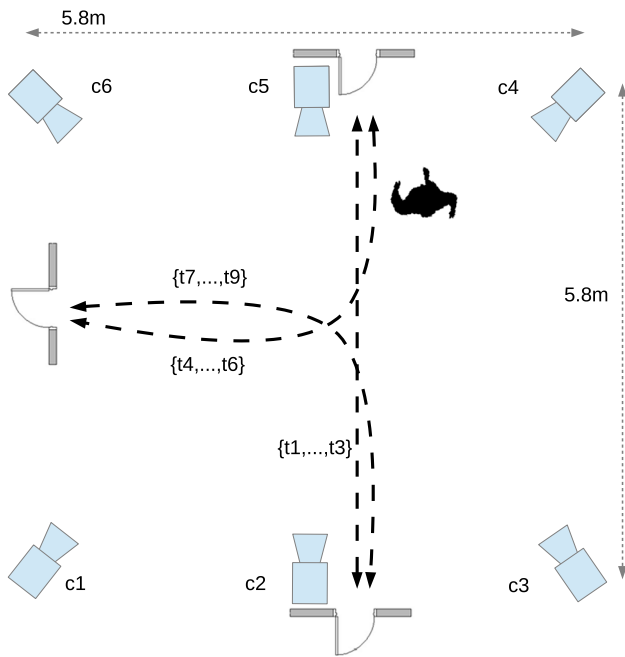
Fig. 8 Workspace setup of AVAMVG dataset, where $\{c1, \ldots, c6\}$ represent the set of cameras of the multi-view setup and $\{t1, \ldots, t9\}$ represent the different trajectories performed by each actor

bases 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.

4.2 Experimental setup

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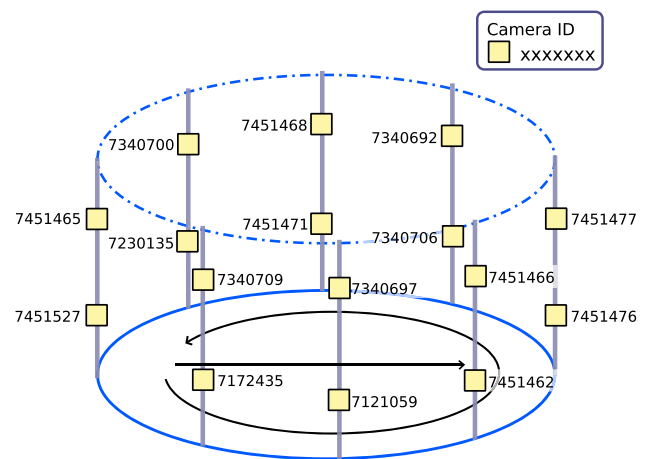

Fig. 10 Experimental setup of KY4D

- *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: GEnV.* The aim of this experiment is to evaluate the performance obtained by using the $\mathscr{G}^{\mathrm{GEnV}}$ signature. Since it 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

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

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 aim 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 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 0.27×10^{-4} m^3 is enough to get a detailed 3D human reconstruction.

The average corporal volume for humans is 66.4 l measured by the water displacement method in 521 people aged between 17 and 51 years [12]. Using a voxel size of 2.7 cl,

the number of voxels belonging to a person in the 3D volume should be about 2459. Thus, for $\theta > 1 \times 10^3$ (see Sect. 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 sequence 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 Sect. 4.2).

We show in Table 2 the performance of $\mathscr{G}^{GEnV}$ 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 $2m$,

Table 1 Results of GEnI baseline approach (Experiment A)

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

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

Table 2 Results obtained by using GEnV as gait signature (Experiment B)

Database	Signature	PCA (%)	PCA+LDA (%)	ϵ
AVAMVG	$\mathscr{G}^{\mathrm{GEnV}}$	77.63	70.10	0.95
KY4D	$\mathscr{G}^{\mathrm{GEnV}}$	78.46	94.52	0.95

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 $W = 1$

with voxel size of 2.7*cl*), 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.

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 Sect. 4.2).

In Tables 3 and 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 Sect. 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 simultaneously, 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 $\mathscr{G}_S^{\mathrm{GEnV}}$ 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 Sect. 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. Figures 11, 12,

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)

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

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 $W = 1$

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)

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

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 $W = 1$

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
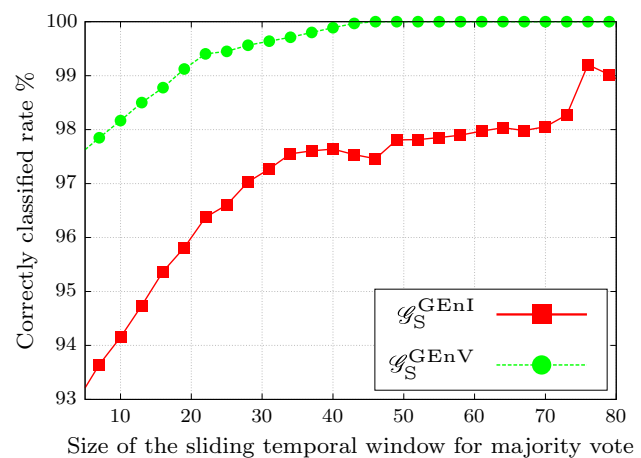

Fig. 11 Performance of $\mathscr{G}_S^{\mathrm{GEnV}}$ and $\mathscr{G}_S^{\mathrm{GEnI}}$ on AVAMVG for different lengths for the majority voting window

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 Sect. 4.2). In Tables 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.

As can be seen, both $\mathscr{G}^{\text{GEnV}}_{S \oplus F}$ and $\mathscr{G}^{\text{GEnV}}_{S \oplus F \oplus T}$ 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,
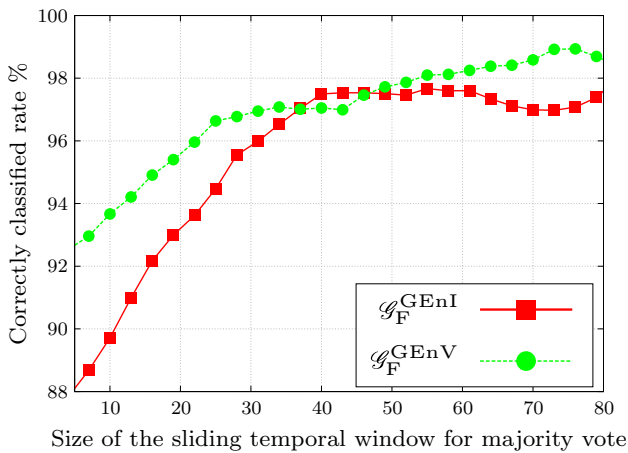

Fig. 12 Performance of $\mathscr{G}^{\text{GEnV}}_F$ and $\mathscr{G}^{\text{GEnI}}_F$ on AVAMVG for different lengths for the majority voting window
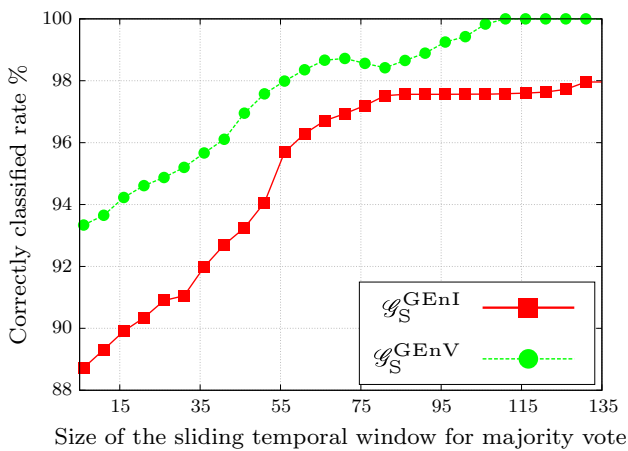
Fig. 13 Performance of $\mathscr{G}^{\text{GEnV}}_S$ and $\mathscr{G}^{\text{GEnI}}_S$ on KY4D for different lengths for the majority voting window

Fig. 14 Performance of $\mathscr{G}^{\text{GEnV}}_F$ and $\mathscr{G}^{\text{GEnI}}_F$ on KY4D for different lengths for the majority voting window

Table 5 Comparative results obtained with combined GEnV signatures and combined GEnI signatures using the AVAMVG dataset (Experiment D)

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

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 $W = 1$

Table 6 Comparative results obtained with combined GEnV signatures and combined GEnI signatures using the KY4D dataset (Experiment D)

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

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 $W = 1$
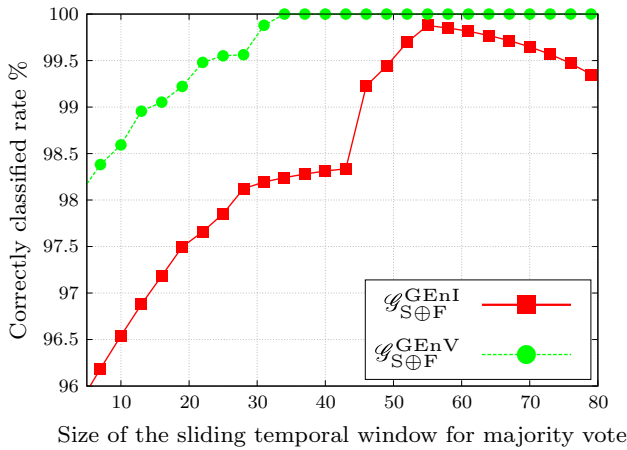
Fig. 15 Performance of $\mathscr{G}_{S\oplus F}^{\text{GEnV}}$ and $\mathscr{G}_{S\oplus F}^{\text{GEnI}}$ on AVAMVG for different lengths for the majority voting window
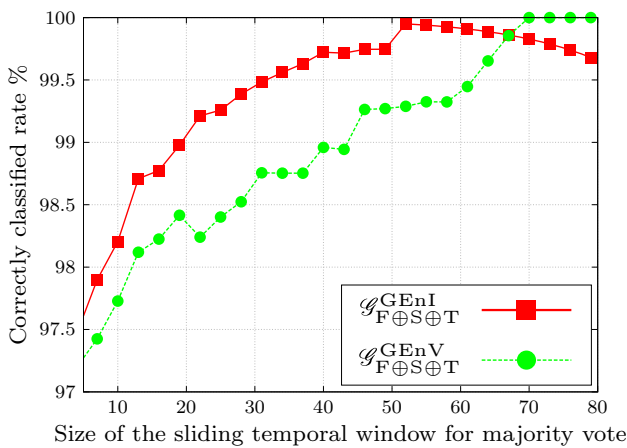

Fig. 16 Performance of $\mathscr{G}_{S\oplus F\oplus T}^{\text{GEnV}}$ and $\mathscr{G}_{S\oplus F\oplus T}^{\text{GEnI}}$ on AVAMVG for different lengths for the majority voting window

Fig. 17 Performance of $\mathscr{G}_{S\oplus F}^{\text{GEnV}}$ and $\mathscr{G}_{S\oplus F}^{\text{GEnI}}$ on KY for different lengths for the majority voting window

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

we cannot 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 $\mathscr{G}_{S\oplus F\oplus T}^{\text{GEnV}}$ is greater than the computational cost of $\mathscr{G}_{S\oplus F}^{\text{GEnV}}$. Therefore, for a system in operation, $\mathscr{G}_{S\oplus F}^{\text{GEnV}}$ 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. Figures 15, 16, 17 and 18 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.

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 parame-
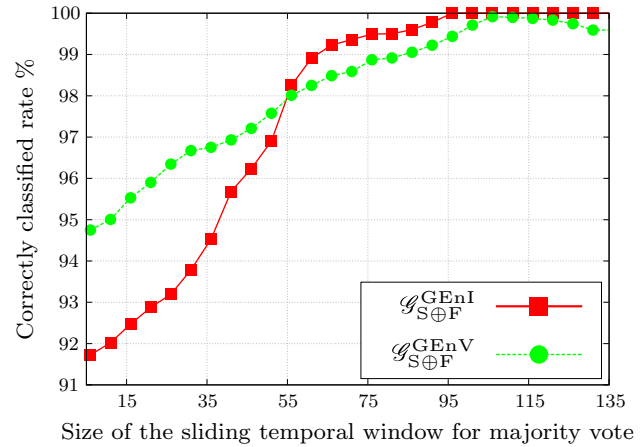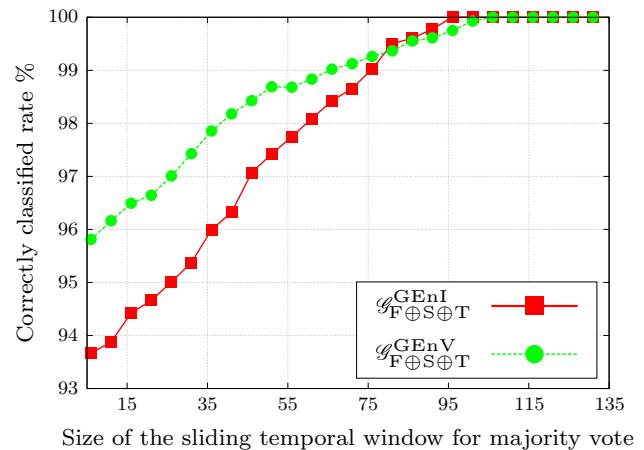
ters independent of the sequence test data, we cross-validate the SVM parameters on the training set. We selected the signature configuration that achieved the best performance in the previous experiments ($\mathscr{G}_{S\oplus F}^{\text{GEnV}}$ and $\mathscr{G}_{S\oplus F\oplus T}^{\text{GEnV}}$) and then we tested them with a set of KY4D models which have been reconstructed using a number of cameras in the range 2–16. As can be seen in Fig. 19, with just three 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 $\mathscr{G}_{S\oplus F}^{\text{GEnV}}$ and 94.26 % with $\mathscr{G}_{S\oplus F\oplus T}^{\text{GEnV}}$.

The results shown in Tables 7 and 8 correspond to Experiment H (see Sect. 4.2). For the corresponding tests on
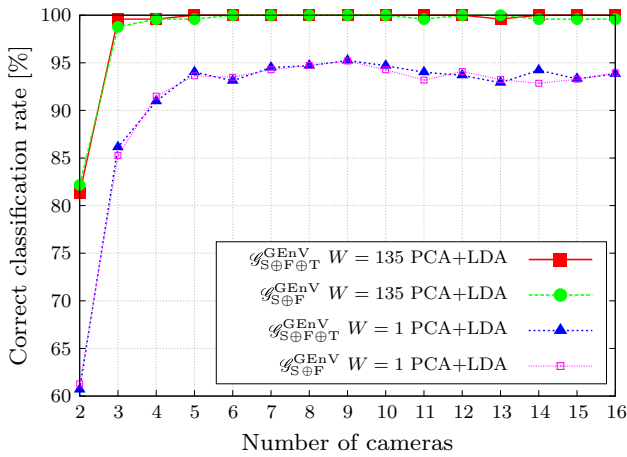
Fig. 19 Performance of $\mathscr{G}_{S\oplus F\oplus T}^{GEnV}$ and $\mathscr{G}_{S\oplus F}^{GEnV}$ on KY4D database for an increasing number of cameras

Table 7 Comparative of recognition results on AVAMVG dataset (Experiment F)

Method	Test curve $t4$ (%)	Test curve $t7$ (%)	AVG (%)
Castro et al. [6]	95	95	95
Iwashita et al. [21]	35.14	37.71	36.42
$\mathscr{G}_{S\oplus F}^{GEnV}$ (PCA+LDA)	97.61	98.41	98.01
$\mathscr{G}_{S\oplus F\oplus T}^{GEnV}$ (PCA+LDA)	97.62	98.84	98.23
$\mathscr{G}_{S\oplus F\oplus T}^{GEnI}$ (PCA+LDA)	96.20	98.42	97.31

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 $W = 80$

Table 8 Comparative of recognition results on KY4D dataset (Experiment F)

Method	Test curve 1 (%)	Test curve 2 (%)	AVG (%)
Iwashita et al. [21]	61.90	71.40	66.65
$\mathscr{G}_{S\oplus F}^{GEnV}$ (PCA+LDA)	44.92	42.08	43.50
$\mathscr{G}_{S\oplus F\oplus T}^{GEnV}$ (PCA+LDA)	57.83	79.28	68.55
$\mathscr{G}_{S\oplus F\oplus T}^{GEnI}$ (PCA+LDA)	55.26	87.31	71.28

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 $W = 135$

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 Sect. 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 [6,21]. 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. [6] only can be tested on the AVAMVG dataset, because it need colour 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 $\mathscr{G}_{S\oplus F\oplus T}^{GEnI}$ and $\mathscr{G}_{S\oplus F\oplus T}^{GEnV}$ 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 Experiments D and E. For this reason, on unconstrained paths, we suggest the use of $\mathscr{G}_{S\oplus F}^{GEnV}$ or $\mathscr{G}_{S\oplus F\oplus T}^{GEnV}$ 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 other 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 AVAMVG [31] and on the KY4D [21]. The experimental results show that GEnV-based signatures such as $\mathscr{G}_{S\oplus F}^{GEnV}$ and $\mathscr{G}_{S\oplus F\oplus T}^{GEnV}$ are the most reli-

able 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.

References

1. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: 3rd International Conference on Crime Detection and Prevention (ICDP'09), pp. 1–6 (2009)
2. Bashir, K., Xiang, T., Gong, S.: Gait recognition without subject cooperation. Pattern Recognit. 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 Vis. Comput. **27**(8), 1194–1206 (2009)
4. Bolle, R., Pankanti, S.: Biometrics, Personal Identification in Networked Society. Kluwer Academic Publishers, Norwell (1998)
5. Burges, C.J.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov. **2**, 121–167 (1998)
6. Castro, F.M., Marín-Jiménez, M.J., Medina-Carnicer, R.: Pyramidal fisher motion for multiview gait recognition. In: Proceedings of the International Conference on Pattern Recognition (2014)
7. Chattopadhyay, P., Roy, A., Sural, S., Mukhopadhyay, J.: Pose depth volume extraction from RGB-D streams for frontal gait recognition. J. Vis. Commun. Image Represent. **25**(1), 53–63 (2014)
8. Cheng, M.H., Ho, M.F., Huang, C.L.: Gait analysis for human identification through manifold learning and hmm. Pattern Recognit. **41**(8), 2541–2553 (2008)
9. Choudhury, S.D., Tjahjadi, T.: Silhouette-based gait recognition using procrustes shape analysis and elliptic Fourier descriptors. Pattern Recognit. **45**(9), 3414–3426 (2012)
10. Cutting, J.E., Kozlowski, L.T.: Recognizing friends by their walk: gait perception without familiarity cues. Bull. Psychon. Soc. **9**, 353–356 (1977)
11. Díaz-Más, L., Muñoz Salinas, R., Madrid-Cuevas, F.J., Medina-Carnicer, R.: Shape from silhouette using Dempster–Shafer theory. Pattern Recognit. **43**(6), 2119–2131 (2010)
12. Drinkwater, D., Ross, W.: Anthropometric fractionation of body mass. In: Kinanthropometry, vol. 2, pp. 178–189. University Park Press, Baltimore (1980)
13. Goffredo, M., Bouchrika, I., Carter, J., Nixon, M.: Self-calibrating view-invariant gait biometrics. IEEE Trans. Syst. Man Cybern. Part B Cybern. **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 (2001)
15. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **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 (ICIP'05), 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: IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 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. Artif. Intell. Eng. **13**(4), 359–366 (1999)
21. Iwashita, Y., Ogawara, K., Kurazume, R.: Identification of people walking along curved trajectories. Pattern Recognit. Lett. **48**, 60–69 (2014)
22. Jean, F., Albu, A.B., Bergevin, R.: Towards view-invariant gait modeling: computing view-normalized body part trajectories. Pattern Recognit. **42**(11), 2936–2949 (2009)
23. Jeong, S., Kim, T.H., Cho, J.: Gait recognition using description of shape synthesized by planar homography. J. Supercomput. **65**(1), 122–135 (2013)
24. Kale, A., Chowdhury, A., Chellappa, R.: Towards a view invariant gait recognition algorithm. In: Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance, 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 (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: 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 Trans. Circuits Syst. Video Technol. **22**(6), 966–980 (2012)
28. Laurentini, A.: The visual hull concept for silhouette-based image understanding. IEEE Trans. Pattern Anal. Mach. Intell. **16**(2), 150–162 (1994)
29. Lee, C.P., Tan, A.W., Tan, S.C.: Gait recognition via optimally interpolated deformable contours. Pattern Recognit. Lett. **34**(6), 663–669 (2013)
30. Liu, N., Tan, Y.P.: View invariant gait recognition. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP'10), 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'14, *Lecture Notes in Computer Science*, vol. 8703. Springer, Berlin (2014)
32. Makihara, Y., Sagawa, R., Mukaigawa, Y., Echigo, T., Yagi, Y.: Gait recognition using a view transformation model in the frequency domain. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) Computer Vision—ECCV 2006. Lecture Notes in Computer Science, vol. 3953, pp. 151–163. Springer, Berlin (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, Berlin (2003)
34. Pal, N., Pal, S.: Entropy: a new definition and its applications. IEEE Trans. Syst. Man Cybern. **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: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01), vol. 1, pp. I-439–I-446 (2001)
37. Sivapalan, S., Chen, D., Denman, S., Sridharan, S., Fookes, C.: Gait energy volumes and frontal gait recognition using depth images. In: Jain, A.K., Ross, A., Prabhakar, S., Kim, J. (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: International Conference on Digital Image Computing Techniques and Applications (DICTA'12), 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 Trans. Pattern Anal. Mach. Intell. **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: Narayanan, P., Nayar, S., Shum, H.Y. (eds.) Computer Vision—ACCV 2006. Lecture Notes in Computer Science, vol. 3851, pp. 807–816. Springer, Berlin (2006)
42. Zheng, S., Zhang, J., Huang, K., He, R., Tan, T.: Robust view transformation model for gait recognition. In: 18th IEEE International Conference on Image Processing (ICIP'11), pp. 2073–2076 (2011)

A. Carmona-Poyato received his title of Agronomic Engineering and Ph.D. degree from the University of Cordoba (Spain), in 1986 and 1989, respectively. Since 1990 he has been working with the Department of Computing and Numerical Analysis of the University of Cordoba, currently he is an assistant professor. His research is focused on image processing and 2D object recognition.
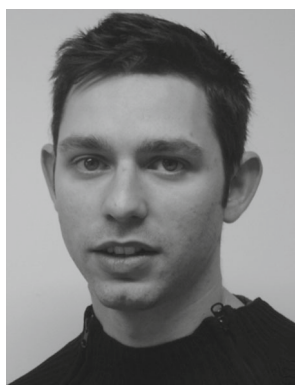
R. Muñoz-Salinas was born in Córdoba, Spain, in 1979. He received both his M.S. degree in Computer Science in 2005 and his Ph.D. in Computer Science in 2006, both from the University of Granada, Spain. He is currently Assistant Professor at the University of Cordoba. His research interests include Autonomous Robots, Soft Computing, Stereo Vision and Tracking.

D. López-Fernández received the Bachelor degree in Computer Science from Córdoba University (Spain), in 2011. Currently he is a Ph.D. student and researcher at University of Córdoba. His research is focused mainly on 3D reconstructions, gait analysis and recognition, and evaluation of computer vision algorithms.

R. Medina-Carnicer received the Bachelor degree in Mathematics from University of Sevilla (Spain). He received the Ph.D. in Computer Science from the Polytechnic University of Madrid (Spain) in 1992. Since 1993 he has been a lecturer of Computer Vision in Cordoba University (Spain). His research is focused on edge detection, evaluation of computer vision algorithms and pattern recognition.

F. J. Madrid-Cuevas received the Bachelor degree in Computer Science from Málaga University (Spain) and the Ph.D. degree from Polytechnic University of Madrid (Spain), in 1995 and 2003, respectively. Since 1996 he has been working with the Department of Computing and Numerical Analysis of Cordoba University, currently as assistant professor. His research is focused on image processing, 2D object recognition and gait recognition using multi-view computer vision algorithms.

Chapter 7

Rotation Invariant Descriptor based on 3D Angular Analysis

This chapter presents two methods to recognize walking humans on unconstrained paths, even with curved or straight trajectories, and regardless direction changes. Both methods are based on 3D angular analysis of the movement of the walking humans. The first [110] takes advantage of 3D reconstructions to extract a rotation invariant gait descriptor, whereas the second [111] is able to extract the rotation invariant gait features directly from silhouettes obtained from multiple views, by using calibration information instead of 3D reconstructions.

A new rotation invariant gait descriptor is presented in the following sections. The methods based on this descriptor are able to cope with rotation changes, while preserving enough discriminatory information from the gait. They focuses on capturing 3D dynamical information of gait, in contrast to other related works which discard a significant part of 3D information.

Main publications associated to this chapter:

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. A new approach for multiview gait recognition on unconstrained paths. *Journal of Visual Communication and Image Representation*, 38:396-406, 2016. ISSN 1047-3203. doi: 10.1016/j.jvcir.2016.03.020. url: `http://www.sciencedirect.com/science/article/pii/S1047320316300232`

- D. López-Fernández, F. J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz Salinas, and R. Medina-Carnicer. Multi-view gait recognition on curved trajectories. In *Proceedings of the 9th International Conference on Distributed Smart Cameras, ICDSC'15*, pages 116-121, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3681-9. doi: 10.1145/2789116.2789122. url: `http://dl.acm.org/citation.cfm?id=2789122`

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, and R. Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In *Activity Monitoring by Multiple Distributed Sensing, Lecture Notes in Computer Science*, pages 26-39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3. url: `http://link.springer.com/chapter/10.1007%2F978-3-319-13323-2_3`

7.1 Angular analysis on 3D human reconstructions

The following paper presents a detailed description and experimental results of a new method that uses 3D human reconstructions to extract gait features which are invariant to rotation changes.

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.

© 2016 Elsevier Inc. All rights reserved.

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
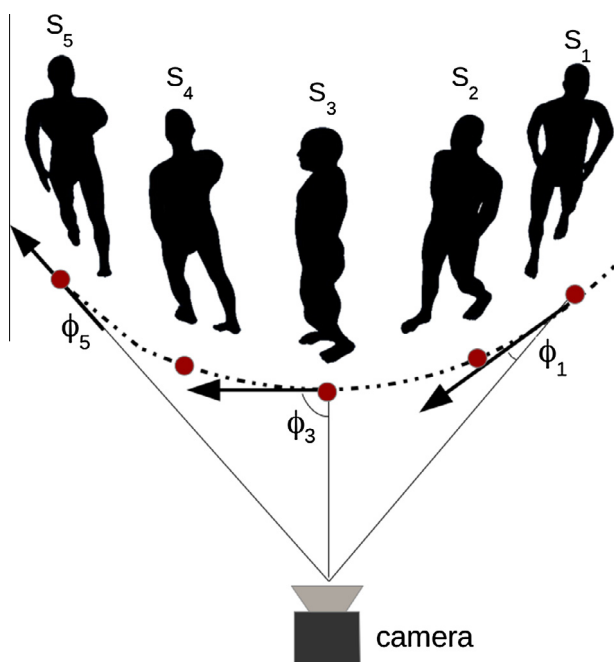

Fig. 1. In a curved path, the observation angle between the walking direction of the subject and optical axis of the camera is gradually changed, which affects the silhouette appearance.

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
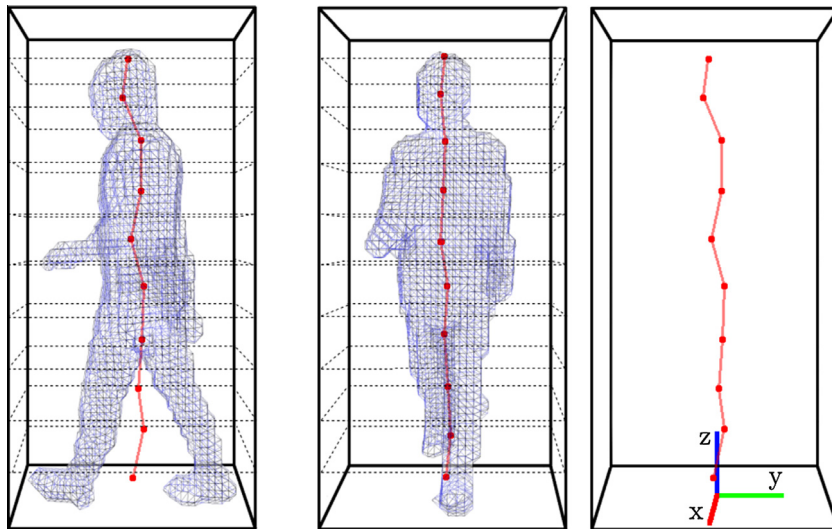
Fig. 2. The reconstructed model is divided into 3D stacked areas of the same size called slices (regions within dotted lines). Centroids are computed on each slice (red points). The gait feature is composed by a set of inner angles between the line joining each pair of consecutive centroids (red line) 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:

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

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:

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

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:

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

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:

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

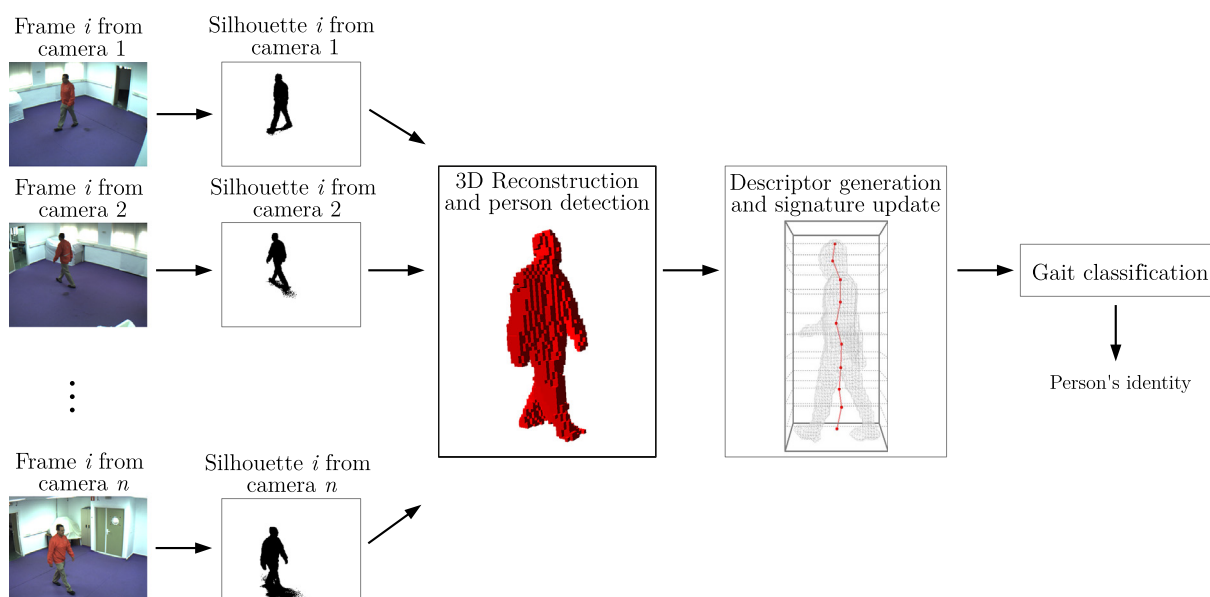$$\beta_h = \min\{\alpha_h, 180 - \alpha_h\}, \qquad (5)$$

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:

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

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 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
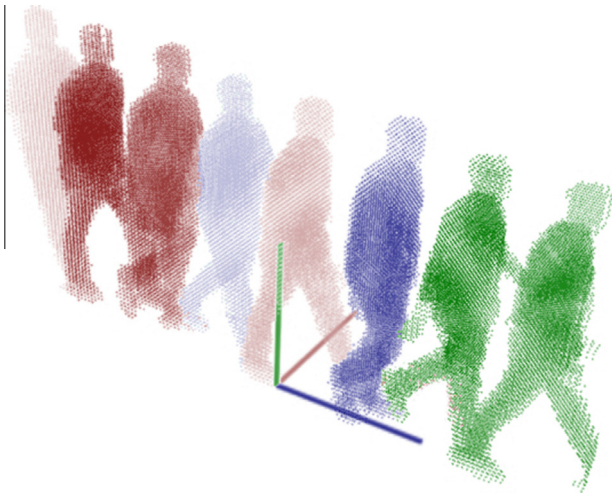
Fig. 4. Example of reconstructed segment of a gait sequence, sampled at 2 Hz, where each point represents a squared voxel. The time instant is represented by different colors. Best viewed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the web 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
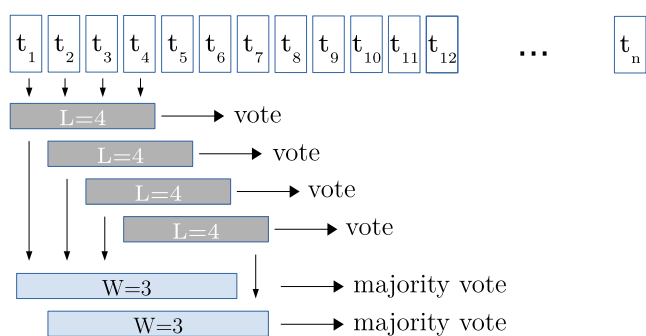
Fig. 5. Majority vote policy over a sliding temporal window. In the example, the size of the signature is set to $L = 4$, and the size of the voting window is set to $W = 3$.

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
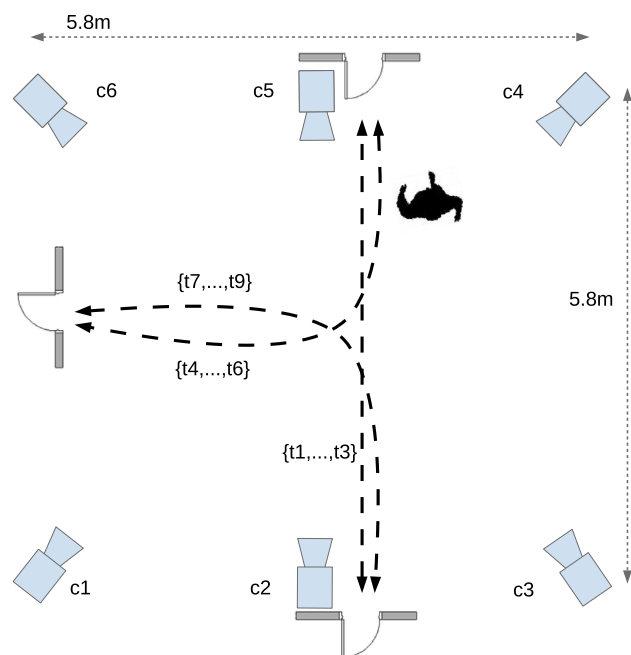
Fig. 6. Workspace setup used by AVAMVG dataset, where $\{c1, \ldots, c6\}$ represent the set of cameras of the multiview dataset and $\{t1, \ldots, t9\}$ represent the different trajectories followed by each actor of the dataset.

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.)
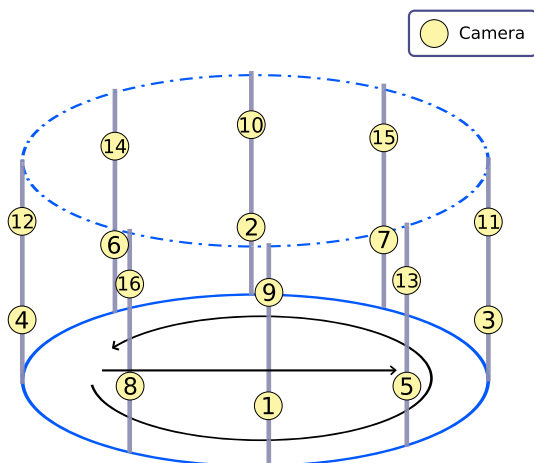

Fig. 8. Experimental setup of KY4D. Each camera is represented by a circle and a number which shows the order in which it was selected to evaluate the performance changes with respect to the number of cameras (see Section 5.2 for further details).

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

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.

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
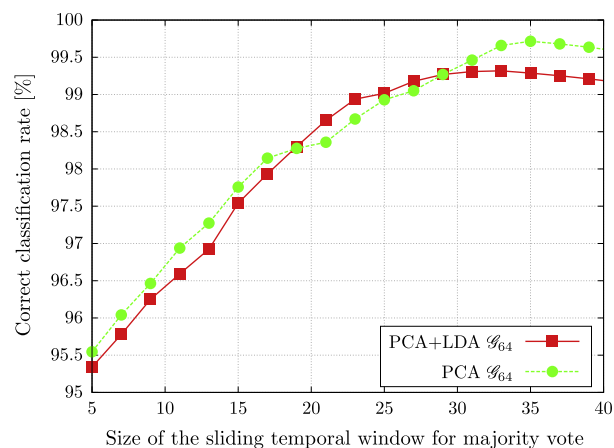

Fig. 9. Performance of our descriptor on the AVAMVG database for different lengths of the majority voting window.
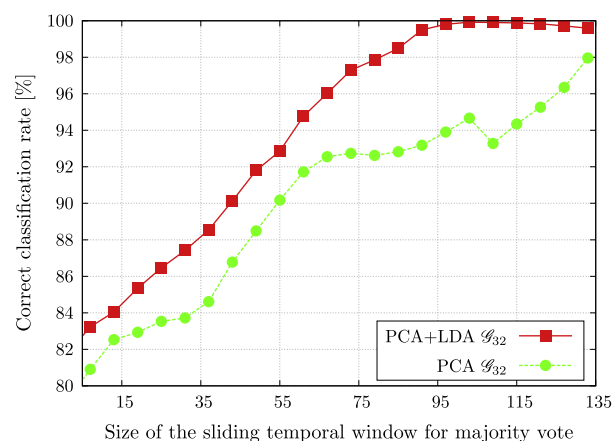

Fig. 10. Performance of our descriptor on the KY4D database for different lengths of the majority voting window.

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.

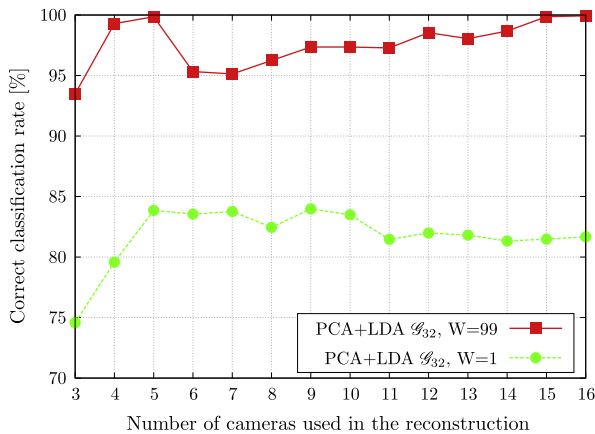| Experiment	Straight paths				Curved paths		AVG
	t1	t2	t3	t4	t5	t6	
$\mathscr{G}_{32}$-PCA-W = 1	93.12	97.55	96.44	96.16	54.25	86.86	87.39
$\mathscr{G}_{32}$-PCA-W = 130	**97.56**	**100**	**100**	**100**	90.24	**100**	97.96
$\mathscr{G}_{32}$-PCA + LDA-W = 1	94.98	98.62	99.10	97.22	58.09	89.09	89.51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

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 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

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
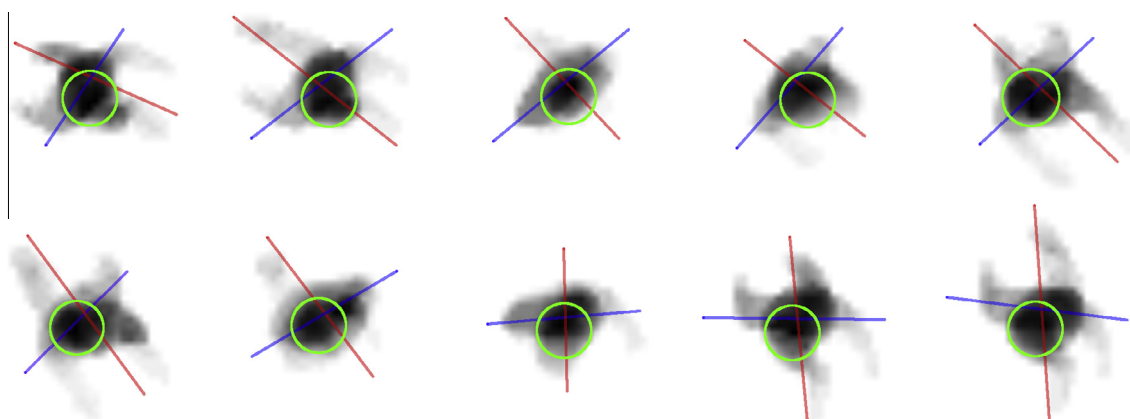
Fig. 12. Example of a curved gait cycle. We show several ground marginal distributions of occupied voxels (see Section 4.1.3). The velocity vector is represented by a red line, the blue line represents the torso main axis, and the position of the head is represented by a green circle. We can note that 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

[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.

Recognition, ICPR 2014, Stockholm, Sweden, August 24–28, 2014, pp. 1692–1697, http://dx.doi.org/10.1109/ICPR.2014.298.

7.2 Angular analysis without 3D human reconstructions

This section presents a method to recognize walking humans independently of the viewpoint and regardless direction changes on curved trajectories. In contrast to the above work, this approach aims to extract 3D dynamical information of gait without using 3D reconstructions.

Multi-view gait recognition on curved trajectories

D. López-Fernández
Department of Computing and
Numerical Analysis.
Maimónides Institute for
Biomedical Research
(IMIBIC).
University of Córdoba, Spain.
i52lofed@uco.es

F.J. Madrid-Cuevas
Department of Computing and
Numerical Analysis.
Maimónides Institute for
Biomedical Research
(IMIBIC).
University of Córdoba, Spain.
fjmadrid@uco.es

A. Carmona-Poyato
Department of Computing and
Numerical Analysis.
Maimónides Institute for
Biomedical Research
(IMIBIC).
University of Córdoba, Spain.
ma1capoa@uco.es

R. Muñoz-Salinas
Department of Computing and
Numerical Analysis.
Maimónides Institute for
Biomedical Research
(IMIBIC).
University of Córdoba, Spain
rmsalinas@uco.es

R. Medina-Carnicer
Department of Computing and
Numerical Analysis.
Maimónides Institute for
Biomedical Research
(IMIBIC).
University of Córdoba, Spain
rmedina@uco.es

ABSTRACT

Appearance changes due to viewing angle changes cause difficulties for most of the gait recognition methods. In this paper, we propose a new approach for multi-view recognition, which allows to recognize people walking on curved paths. The recognition is based on 3D angular analysis of the movement of the walking human. A coarse-to-fine gait signature represents local variations on the angular measurements along time. A Support Vector Machine is used for classifying, and a sliding temporal window for majority vote policy is used to smooth and reinforce the classification results. The proposed approach has been experimentally validated on the publicly available "Kyushu University 4D Gait Database". The results show that this new approach achieves promising results in the problem of gait recognition on curved paths.

CCS Concepts

•**Security and privacy** → **Biometrics;** •**Computing methodologies** → **Tracking;** *Activity recognition and understanding; Motion capture;* •**Applied computing** → Surveillance mechanisms;

Keywords

Gait recognition; 3D descriptor; independent-view; curved paths.

1. INTRODUCTION

Researches on human gait as a biometric feature for identification have received a lot of attention due to the advantage that it can operate from a distance and can be applied discreetly without needing the active participation of the observed individual [7]. However, gait recognition performance is significantly affected by changes in various covariate conditions such as clothing [6], camera viewpoint [12, 9], load carrying [16], and walking speed [18].

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 will not change while walking [13, 2, 5, 17]. In such methods, a change in the appearance, caused by a viewpoint change, adversely affects to the recognition [19]. 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 during a gait cycle. Fig. 1 shows this problem and 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 for multi-view gait recognition which allows to identify people walking along both curved and straight paths. Some potential applications of this work is smart video surveillance (e.g. bank offices, government facilities, or underground stations) and access control or monitoring in special or restricted areas (e.g. military bases or medical isolation zones where subjects wear special clothing that does not allow to show the face or use the fingerprint).

The rest of the paper is structured as follows. After presenting in Section 2 the related work, we describe our proposed framework for gait recognition in Section 3. Section 4 is devoted to the experimental results. And, finally, we conclude this paper in Section 5.
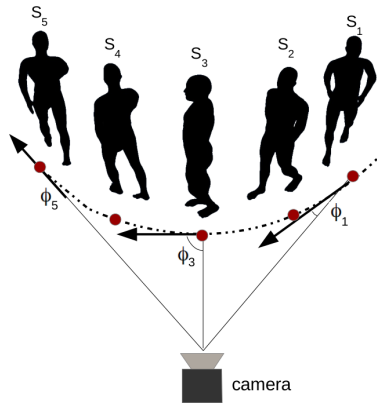
Figure 1: In a curved path, the observation angle between the walking direction of the subject and optical axis of the camera is gradually changed, which affects the silhouette appearance.

2. RELATED WORK

Appearance changes due to viewing angle changes cause difficulties for most of the 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 [12].

Approaches of the first category are represented by [1, 9]. The approach of Bodor *et al.* [1] tries to classify the motion of a human in a view-independent way, but it also 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 [9], 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, as in the above two works, despite 3D models are used, descriptors are computed from silhouettes and they based on 2D information, so that 3D information is discarded.

Approaches of the second category extract gait features which are invariant to viewing angle change. In [10], 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 mo-

tion direction and the camera optical axis [11], however this method seems not to be applicable when the person is walking nearly parallel to the optical axis. In [4] 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 [3]. 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.

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 [14], where LDA-subspaces are learned to extract discriminative information from gait features under each viewing angle.

A View Transformation Model (VTM) was introduced by [15] to transform gait features from different views into the same view. The method of Makihara *et al.* [15] creates a VTM based on frequency-domain gait features, obtained through Fourier Transformation. A sparse-regression-based VTM for gait recognition under various views is also proposed in [12]. However, this method cannot deal with changes in the direction of motion and cannot 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 [12]: (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).

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 FRAMEWORK

This work presents a method to recognize walking humans independently of the viewpoint and regardless direction changes on curved trajectories. Our approach aims to extract 3D dynamical information of gait. The body human region is vertically divided into 3D stacked areas of the same size called slices and then we compute the centroid of each slice. The gait feature is composed by a set of acute angles between the line joining each pair of consecutive centroids and the z-axis (z-axis extends up) in $\mathbb{R}^3$.

The proposed algorithm consists of four steps that predict the identity of a walking human a time t. Following are
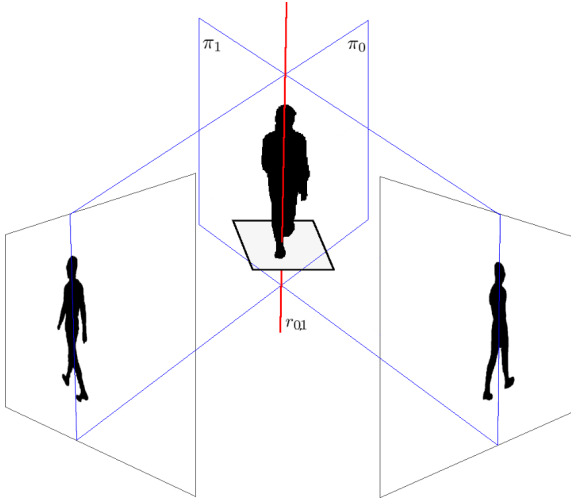
Figure 2: The principal axis of the silhouettes is back-projected to get a plane. Then, the location of the individual in the scene is determined by the intersection between the line of intersection of the two planes and the floor plane.
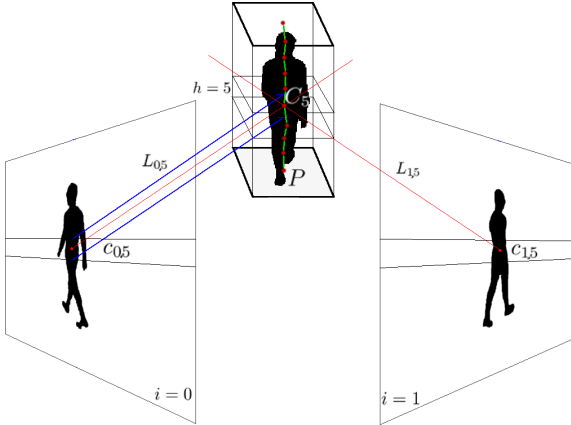

Figure 3: The centroid C_h is obtained by finding the point closest to the set of rays $\{L_{i,h} \mid 0 \leqslant i < N\}$. See main text for further details.

described these steps in detail.

3.1 Tracking

We assume a set of N calibrated cameras. Since cameras have been calibrated, the internal and external camera parameters are known. We also assume the floor to be flat and its position in 3D space to be known.

The first step of our algorithm is to determine the location of the individual in the scene. For that, we start by obtaining the principal axis of the silhouette for each camera view i by Principal Component Analysis. Next, for each view, we back-project this line in order to get the plane $\pi_i \in \{\pi_0, \pi_1, ..., \pi_{N-1}\}$, as Fig. 2 shows.

It is assumed a function $f \colon \mathbb{R}^3 \mapsto \mathbb{R}^3$ to map from camera local coordinates to scene world coordinates. Then we map each plane π_i from local camera coordinates to scene world coordinates. Let us denote $r_{i,j}$ as the intersection line between the planes π_i and π_j, where $0 \leqslant i < N$ and $0 \leqslant j < N$.

We denote F as a set of candidate foot points, obtained by intersecting the lines $r_{i,j}$ with the floor plane, without repetition, so that the cardinality of the set is $|F| = \binom{N}{2}$. Finally, the location of the individual is denoted by:

$$P = \frac{1}{|F|} \sum_{i=0}^{|F|} F_i. \tag{1}$$

3.2 Descriptor generation

Given a foot point $P : (P_x, P_y, 0)$, the 3D scene is vertically divided into $H \in \mathbb{N}^+$ parts, called slices. Let us denote $p^{i,h}$ and $p^{i,h+1}$ as the projections on the image view i of the 3D points $(P_x, P_y, h\frac{Z}{H})$ and $(P_x, P_y, (h+1)\frac{Z}{H})$ respectively, where Z is the total height of the 3D scene. We compute the 2D centroid $c_{i,h} = \{\bar{x}, \bar{y}\}$ on the bounding box enclosing the pixels $(0, p_y^{i,h+1})$ and $(w, p_y^{i,h})$, where w is the width of the image.

Then, using the 2D centroid $c_{i,h}$ and the calibration data for the view i, we can backproject the ray $L_{i,h}$ passing through the image view point $c_{i,h}$. In order to obtain an approximation of the 3D centroid C_h of the slice h in the scene, we find the point closest to the set of rays $\{L_{i,h} \mid 0 \leqslant i < N\}$. We propose to solve it by minimizing the sum of squared distances.

Next we define the angle between the normal vector to the floor plane ($\vec{Z} = (0, 0, 1)$) and the vector joining each pair of consecutive centroids as:

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

$$\beta_h = \min\{\alpha_h, 180 - \alpha_h\}, \tag{3}$$

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:

$$D_{H,t} = (\beta_{(0,t)}, \beta_{(1,t)}, ..., \beta_{(H-2,t)}). \tag{4}$$

If the slice h is empty (e.g. slices above the head) or $|\{L_{i,h}\}| < 2$, C_h cannot be estimated. In such cases, to preserve the height of the subject as feature, we set $\beta_{h,t} = 0$. Fig. 3 shows the descriptor generation process.

3.3 Signature update

The first step of our classification system is the generation of the gait descriptor $D_{(H,t)}$ at time t. 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:

$$0 < l \leqslant \lfloor \log_2 H \rfloor, \tag{5}$$

so that the first level descriptor contains features extracted from the scene divided into 2 slices, the second level descriptor contains features extracted from the scene divided into $H = 2^2$ slices, and so on until we have divided the scene into

$H = 2^l$ slices. We can now concatenate the level descriptors to represent our coarse-to-fine descriptor as:

$$\mathscr{D}_{(l,t)} = (D_{(2,t)}, D_{(2^2,t)}, ..., D_{(2^l,t)}). \qquad (6)$$

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:

$$\mathscr{G}_{(l,t)} = (\mathscr{D}_{(l,t-L+1)}, ..., \mathscr{D}_{(l,t-1)}, \mathscr{D}_{(l,t)}), \qquad (7)$$

which consists 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 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 [12, 9] among others.

3.4 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.

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 [8]. Then we use a Support Vector Machine (SVM) 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 identity.

4. OVERVIEW OF THE EXPERIMENTS

In order to validate our approach, we carry out diverse experiments on the publicly available "Kyushu University 4D Gait Database" [9]. With these experiments we try to answer, among others, the following questions:

- Is our descriptor a valid approach to recognize walking humans independently of the viewpoint, even with curved trajectories?

- What level of refinement for our coarse-to-fine gait descriptor is required to achieve the best recognition rate?

- 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?

4.1 Dataset description

"Kyushu University 4D Gait Database" (KY4D) [1] [9], 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 studio setup is shown in Figure 4. Despite 3D models are available, we do not use them, because this work relies on camera calibration for getting 3D information.
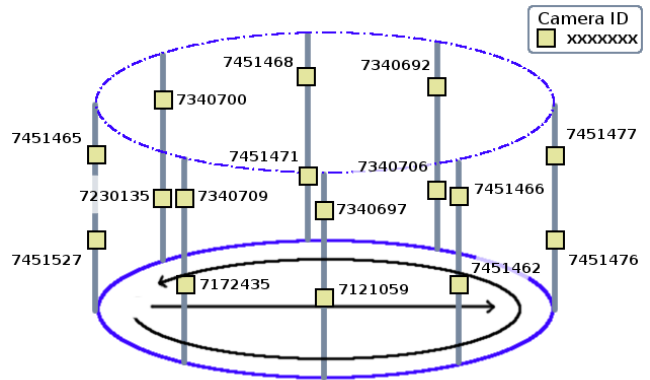

Figure 4: Experimental setup of KY4D.

As can be seen in Figure 4, KY4D gait sequences are captured by 16 cameras forming rings at two heights. The lower level comprises the cameras {7451527, 7172435, 7121059, 7451462, 7451476, 7340706, 7451471, 7230135}, whereas the upper level comprises the cameras {7451465, 7340709, 7340697, 7451466, 7451477, 7340692, 7451468, 7340700}.

4.2 Experimental results

Next we need to determine the value of several parameters of our method. According to the length of the gait signature, we set to $L = 20$ the number of frames where our descriptor is computed, because this value roughly matches with the average length of a gait cycle. Regarding the number of levels (see Section 3.3), we tested l from 1 (2 slices) to 6 (2^6 slices).

We use a leave-one-out cross-validation strategy. Thus, each fold is 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. We use a C-SVC SVM with Radial Basis Function (RBF) kernel, since we obtained better results than with linear, polynomial, or sigmoid kernel. So, we have to adjust the parameter C and the gamma value for the RBF kernel. To make the choice of these 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. Moreover, some subjects walk faster than others and therefore cause a greater number of votes. To cope with this issue, we normalize by class the results of each trajectory.

---

[1]Publicly available at: http://robotics.ait.kyushu-u.ac.jp/research-e.php?content=db

| | Straight paths | | | | Curved paths | | |
Experiment	t1	t2	t3	t4	t5	t6	AVG
upper-$\mathscr{G}_{32}$-PCA-W=1	44.63	51.80	46.78	48.13	28.74	58.45	46.42
upper-$\mathscr{G}_{32}$-PCA-W=120	70.73	78.04	82.92	87.80	54.34	85.00	76.47
upper-$\mathscr{G}_{32}$-PCA+LDA-W=1	49.85	57.07	51.12	52.68	29.70	56.05	49,41
upper-$\mathscr{G}_{32}$-PCA+LDA-W=120	80.48	80.48	90.24	82.92	52.17	72.50	76,46
lower-$\mathscr{G}_{32}$-PCA-W=1	84.86	87.73	88.97	89.49	52.24	78.02	80.21
lower-$\mathscr{G}_{32}$-PCA-W=120	95.12	100	100	100	**97.82**	**100**	**98.82**
lower-$\mathscr{G}_{32}$-PCA+LDA-W=1	88.24	89.82	89.10	90.30	52.42	78.52	81.40
lower-$\mathscr{G}_{32}$-PCA+LDA-W=120	**97.56**	**100**	**100**	**100**	91.30	97.50	97.72

Table 2: 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.

l	PCA	PCA+LDA
1	5.99	N.A
2	31.54	32.92
3	56.98	56.38
4	74.75	74.68
5	80.21	**81.40**
6	79.56	80.69

Table 1: Correct classification rate [%] on the lower set of cameras for several values of the parameter l. The size of the sliding temporal window for majority voting is set to $W = 1$. Best result is marked in bold.

In order to achieve the best data representation and class separability simultaneously, we apply PCA+LDA to the training and test data (see Section 3.4). With regard to PCA, we only retain 95% of the variance. In the classification step, 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.

The recognition rate on the lower set of cameras for several values of the parameter l is shown in Table 1. 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 ($W = 1$). We obtained the best results with $l = 5$ and PCA+LDA. Besides the recognition rate, the number of features is considerably lower with PCA+LDA than with PCA.

We next conducted experiments in which we applied the sliding temporal window for majority voting policy. As can be seen in Figure 5, the use of a majority voting policy over a sliding temporal window significantly improves the performance of our method. However, the performance obtained with the lower set of cameras is greater than with the upper set (see Section 4.1). This may be caused by the tilt of the upper set of cameras. Further analysis of this issue is left for a future study. The size of the window is limited by the number of available gait signatures for each sequence.

The results of Table 2 shows detailed results for the leave-one-out experiment, with $l = 5$. We show the effect of applying the sliding temporal window for majority voting compared with $W = 1$ (disabled window). As can be seen, we have obtained better results with the lower set of cameras than with the upper set. This could be due to the tilt of the cameras. This issue is left for a future study.

In order to determine the number of cameras that should
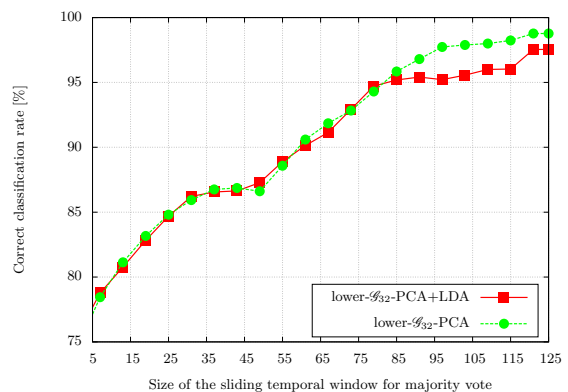

Figure 5: Performance of our descriptor on the lower level of cameras of KY4D database for different lengths of the majority voting window.
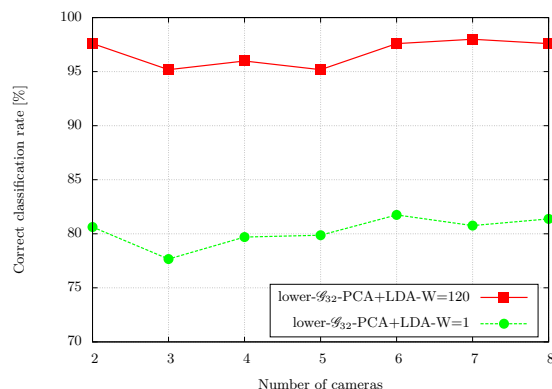

Figure 6: Performance of our descriptor for an increasing number of cameras.

be employed and its effect on the performance, we have designed a leave-one-out cross validation experiment. We selected the signature configuration that achieved the best performance in the previous experiments and then we tested it with a variable number of cameras of the lower set in the range 2 to 8. As can be seen in Figure 6, with just 2 calibrated cameras, our method is able to correctly classify up to 95% of individuals, independently of the path, even with curved trajectories. This is because at least two rays are needed ($|\{L_{i,h}\}| < 2$) to obtain the intersection C_h.

5. CONCLUSIONS

This paper has proposed a new approach to recognize walking humans independently of the viewpoint and regardless direction changes on curved trajectories. Our approach allow people to walk freely in the scene, in contrast to others view-independent approaches which restrict the view change to a few angles.

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.

This approach does not require the sequence to be split into gait cycles, and the results are smoothed and reinforced over time by using a sliding temporal window for majority voting policy. Experimental results show that our method is able to reach a correct classification rate up to 95%.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos. View-independent human motion classification using image-based reconstruction. *Image and Vision Computing*, 27(8):1194–1206, 2009.

[2] S. D. Choudhury and T. Tjahjadi. Gait recognition based on shape and motion analysis of silhouette contours. *Computer Vision and Image Understanding*, 117(12):1770 – 1785, 2013.

[3] M. Goffredo, I. Bouchrika, J. Carter, and M. Nixon. Self-calibrating view-invariant gait biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(4):997–1008, Aug 2010.

[4] J. Han, B. Bhanu, and A. Roy-Chowdhury. A study on view-insensitive gait recognition. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages III–297–300, Sept 2005.

[5] M. Hofmann, S. Bachmann, and 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)*, pages 399–403, Sept 2012.

[6] M. A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281 – 2291, 2010.

[7] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(3):334–352, Aug 2004.

[8] P. Huang, C. Harris, and M. Nixon. Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering*, 13(4):359 – 366, 1999.

[9] Y. Iwashita, K. Ogawara, and R. Kurazume. Identification of people walking along curved trajectories. *Pattern Recognition Letters*, 48(0):60 – 69, 2014. Celebrating the life and work of Maria Petrou.

[10] F. Jean, A. B. Albu, and R. Bergevin. Towards view-invariant gait modeling: Computing view-normalized body part trajectories. *Pattern Recognition*, 42(11):2936 – 2949, 2009.

[11] S. Jeong, T.-h. Kim, and J. Cho. Gait recognition using description of shape synthesized by planar homography. *The Journal of Supercomputing*, 65(1):122–135, 2013.

[12] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(6):966–980, June 2012.

[13] C. P. Lee, A. W. Tan, and S. C. Tan. Gait recognition via optimally interpolated deformable contours. *Pattern Recognition Letters*, 34(6):663 – 669, 2013.

[14] N. Liu and Y.-P. Tan. View invariant gait recognition. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1410–1413, March 2010.

[15] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In *Computer Vision - ECCV 2006*, number 3953 in Lecture Notes in Computer Science, pages 151–163. Springer Berlin Heidelberg, Jan. 2006.

[16] S. Singh and K. Biswas. Biometric gait recognition with carrying and clothing variants. In *Pattern Recognition and Machine Intelligence*, volume 5909 of *Lecture Notes in Computer Science*, pages 446–451. Springer Berlin Heidelberg, 2009.

[17] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes. Gait energy volumes and frontal gait recognition using depth images. In *International Joint Conference on Biometrics (IJCB)*, pages 1–6, Washington DC, USA, Oct 2011. IEEE.

[18] A. Tsuji, Y. Makihara, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 717–722, June 2010.

[19] S. Yu, D. Tan, and T. Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In *Computer Vision - ACCV 2006*, volume 3851 of *Lecture Notes in Computer Science*, pages 807–816. Springer Berlin Heidelberg, 2006.

Chapter 8

Conclusions

This chapter lists the main conclusions of this dissertation in Section 8.1. Finally, Section 8.2 enumerates some open research lines that will be continued in the near future.

8.1 Conclusions and main contributions

Along this dissertation, the main goal we have pursued has been achieving new 3D gait recognition methods able to identify people independently of the trajectory of motion. As it was stated in Section 1.3, we have tackled such a main goal from the standpoint of 3D reconstructions.

It has been demonstrated that the use of 3D reconstructions in gait recognition have several advantages. Firstly, the use of a methodology based on 3D reconstructions and gait alignment is a way to adapt, by means of projections, well-known 2D gait recognition descriptors (such as the GEnI [75] and the cover $CR(S)$ [34]) in order to recognize people independently of the trajectory of motion, even with curved trajectories and direction changes. Secondly, a greater amount of dynamical information can be extracted and analysed from the 3D domain, leading to a better accuracy 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.

Chapters 4 to 6 have covered the topic of 3D reconstructions and gait alignment, and have presented two contributions to extract a gait signature

from aligned 3D volumes. The first contribution [108] is described in Chapter 5 and proposes two new morphological descriptors which are listed below:

- The aggregation of three CR descriptors computed on the top, side, and frontal projections of the gait volume (CRP).

- The Cover by Cubes (CC), which is defined as the union of all the cubes with the largest size that can fit into a gait volume of a person.

As can be seen in [108] and Chapter 5, the experimental results show that the $CRP(M = N = 20)$ signature provides good results on both AVAMVG and KY4D gait databases. By using a sliding temporal window for majority voting, the system is able to correctly identify up to 96% of the subjects of the AVAMVG gait database and nearly 94% of subjects of the KY4D dataset.

The second contribution [109] is described in Chapter 6 and proposes the Gait Entropy Volume (GEnV), which focuses on capturing 3D dynamic information of a walking human through the concept of entropy, applied on aligned volumetric reconstructions. Several signatures based on marginal distributions of the GEnV, and combinations of them, have also been proposed in order to get a better recognition rate.

The experimental results show that $\mathscr{G}_{F \oplus S \oplus T}^{GEnV}$ and $\mathscr{G}_{F \oplus S}^{GEnV}$ signatures are the most reliable for using with our gait recognition method on unconstrained path, providing a recognition rate of 100% and 99.59% with both signatures on AVAMVG and KY4D, respectively.

The third contribution [110] is described in Chapter 7. In Section 7.1 is presented a new rotation invariant gait descriptor which can be applied on 3D reconstructed volumes without the need for a previous alignment step. What is specially interesting of this descriptor is that it can be applied to characterize the gait of an individual independently of the trajectory of motion, and regardless direction changes. With this approach, the recognition rate is 99.35% and 99.59% on AVAMVG and KY4D, respectively.

The use of a sliding temporal window for majority voting policy has improved the classification results of the described approaches. It was proposed to smooth and reinforce the results over time. We have also exhaustively compared our approaches with several close-related state of the art works such as [34, 75, 26, 57, 58, 67], using the both AVAMVG and KY4D publicly available databases.

However, the use of 3D reconstructions had also two main disadvantages. The first is the need for a multi-view camera setup. The second is the need for calibration information. In Section 7.2, we propose a method [111] whose aim is to extract 3D dynamical information of gait without using 3D reconstructed volumes, but its performance seems to be seriously affected by the camera tilting.

8.2 Future research

This dissertation has addressed the problem of the independence from the trajectory in gait recognition, from the standpoint of 3D reconstructions. However, the use of 3D reconstructions have also two main disadvantages. The first is the need for a multi-view camera setup. The second is the need for calibration information.

We consider this thesis as a milestone in a long road towards unrestricted automatic video-surveillance based on gait. We have detected the following future research lines:

- The development of unconstrained gait recognition methods, i.e. without camera calibration on single-view datasets. E.g. the View Transformation Models (VTM) do not require camera calibration. The performance of these methods relies on learning mapping/projection relationship of gaits under various viewing angles. However, in VTM it is difficult to recognize gait under untrained viewing angles, and these methods assume people walk along straight paths. They could be improved to cope with direction changes.

- The design of methods that do not require neither camera calibration nor multi-view sequences would allow to use more populated databases, such as OU-ISIR.

- In Section 7.2, we propose a method whose aim is to extract 3D dynamical information of gait without using 3D reconstructed volumes, but its performance seems to be seriously affected by the camera tilting. The correction of the camera tilting before computing the gait descriptor could be a significant improvement of this method.

- Occlusion handling is a major problem in gait recognition. Typically,

during occlusion, only portions of the body are visible. The individuals can be occluded by fixed or moving objects, but they can also be self-occluded. The self-occlusion of a human body is a difficult problem which occurs mainly in 2D-based gait recognition methods. We consider the use of multiple cameras as a possible way to address different types of occlusions.

- The design of methods to address other covariate conditions, such as clothing and load or bag carrying.

Bibliography

[1] J.P. Singh and S. Jain. Person identification based on gait using dynamic body parameters. In *Trendz in Information Sciences Computing (TISC), 2010*, pages 248–252, Dec 2010. doi: 10.1109/TISC.2010. 5714649.

[2] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti. Toward unconstrained fingerprint recognition: A fully touchless 3-D system based on two views on the move. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(2):202–219, Feb 2016. ISSN 2168-2216. doi: 10.1109/TSMC.2015.2423252.

[3] C. Tan and A. Kumar. Accurate iris recognition at a distance using stabilized iris encoding and zernike moments phase features. *IEEE Transactions on Image Processing*, 23(9):3962–3974, Sept 2014. ISSN 1057-7149. doi: 10.1109/TIP.2014.2337714.

[4] W. Kang and Q. Wu. Pose-invariant hand shape recognition based on finger geometry. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 44(11):1510–1521, Nov 2014. ISSN 2168-2216. doi: 10.1109/TSMC.2014.2330551.

[5] L. Mirmohamadsadeghi and A. Drygajlo. Palm vein recognition with local texture patterns. *Biometrics, IET*, 3(4):198–206, 2014. ISSN 2047-4938. doi: 10.1049/iet-bmt.2013.0041.

[6] L. Roewer. DNA fingerprinting in forensics: past, present, future. *Investigative Genetics*, 4(1):1–10, 2013. ISSN 2041-2223. doi: 10.1186/ 2041-2223-4-22.

[7] Y. Xu, X. Fang, X. Li, J. Yang, J. You, H. Liu, and S. Teng. Data uncertainty in face recognition. *IEEE Transactions on Cybernetics*, 44 (10):1950–1961, Oct 2014. ISSN 2168-2267. doi: 10.1109/TCYB.2014. 2300175.

[8] A. Pflug and C. Busch. Ear biometrics: a survey of detection, feature extraction and recognition methods. *Biometrics, IET*, 1(2):114–129, June 2012. ISSN 2047-4938. doi: 10.1049/iet-bmt.2011.0003.

[9] S. Schall, S. Kiebel, B. Maess, and K. von Kriegstein. Voice identity recognition: functional division of the right STS and its behavioral relevance. *Journal of Cognitive Neuroscience*, 27(2):280–291, Feb 2014. ISSN 0898-929X. doi: 10.1162/jocn_a_00707.

[10] A.A. Ahmed and I. Traore. Biometric recognition based on free-text keystroke dynamics. *IEEE Transactions on Cybernetics*, 44(4):458–472, April 2014. ISSN 2168-2267. doi: 10.1109/TCYB.2013.2257745.

[11] L. Liu, W. Jia, and Y. Zhu. Survey of gait recognition. In D. Huang, K. Jo, H. Lee, H. Kang, and V. Bevilacqua, editors, *Emerging Intelligent Computing Technology and Applications. With Aspects of Artificial Intelligence*, volume 5755 of *Lecture Notes in Computer Science*, pages 652–659. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-04019-1. doi: 10.1007/978-3-642-04020-7_70.

[12] J.E. Cutting and L.T. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, 9:353–356, 1977.

[13] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 34(3):334–352, Aug 2004. ISSN 1094-6977. doi: 10.1109/TSMCC.2004.829274.

[14] S.A. Niyogi and E.H. Adelson. Analyzing and recognizing walking figures in XYT. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pages 469–474, Jun 1994. doi: 10.1109/CVPR.1994.323868.

[15] S. Sarkar, P.J. Phillips, Z. Liu, I. Robledo-Vega, P. Grother, and K.W. Bowyer. The humanID gait challenge problem: Data sets, performance, and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:162–177, 2005.

[16] T.T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi. Similar gait action recognition using an inertial sensor. *Pattern Recognition*, 48(4):1289–1301, 2015. ISSN 0031-3203. doi: 10.1016/j.patcog.2014.10.012.

[17] T.T. Ngo, Y. Makihara, H. Nagahara, Y. Mukaigawa, and Y. Yagi. The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognition*, 47(1):228–237, 2014. ISSN 0031-3203. doi: 10.1016/j.patcog.2013.06.028.

[18] S. Zheng, K. Huang, T. Tan, and D. Tao. A cascade fusion scheme for gait and cumulative foot pressure image recognition. *Pattern Recognition*, 45(10):3603–3610, 2012. ISSN 0031-3203. doi: 10.1016/j.patcog.2012.03.008.

[19] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin. Infrared gait recognition based on wavelet transform and support vector machine. *Pattern Recognition*, 43(8):2904–2910, 2010. ISSN 0031-3203. doi: 10.1016/j.patcog.2010.03.011.

[20] M. Hofmann, S. Bachmann, and 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)*, pages 399–403, Sept 2012. doi: 10.1109/BTAS.2012.6374606.

[21] J. Han and B. Bhanu. Individual recognition using gait energy image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28 (2):316–322, Feb 2006. ISSN 0162-8828. doi: 10.1109/TPAMI.2006.38.

[22] T.H.W. Lam, K.H. Cheung, and J.N.K. Liu. Gait flow image: A silhouette-based gait representation for human identification. *Pattern Recognition*, 44(4):973–987, 2011. ISSN 0031-3203. doi: 10.1016/j.patcog.2010.10.011.

[23] Zhuowen Lv, Xianglei Xing, Kejun Wang, and Donghai Guan. Class energy image analysis for video sensor-based gait recognition: A review. *Sensors*, 15(1):932, 2015. ISSN 1424-8220. doi: 10.3390/s150100932.

[24] M.A. Hossain, Y. Makihara, J. Wang, and Y. Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, 43(6):2281–2291, 2010. ISSN 0031-3203. doi: 10.1016/j.patcog.2009.12.020.

[25] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li. Gait recognition under various viewing angles based on correlated motion regression. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(6):966–980, June 2012. ISSN 1051-8215. doi: 10.1109/TCSVT.2012.2186744.

[26] Y. Iwashita, K. Ogawara, and R. Kurazume. Identification of people walking along curved trajectories. *Pattern Recognition Letters*, 48: 60–69, 2014. ISSN 0167-8655. doi: 10.1016/j.patrec.2014.04.004. Celebrating the life and work of Maria Petrou.

[27] S. Singh and K.K. Biswas. Biometric gait recognition with carrying and clothing variants. In *Pattern Recognition and Machine Intelligence*, volume 5909 of *Lecture Notes in Computer Science*, pages 446–451. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-11163-1. doi: 10. 1007/978-3-642-11164-8_72.

[28] A. Tsuji, Y. Makihara, and Y. Yagi. Silhouette transformation based on walking speed for gait identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 717–722, June 2010. doi: 10.1109/CVPR.2010.5540144.

[29] Shiqi Yu, Daoliang Tan, and Tieniu Tan. Modelling the effect of view angle variation on appearance-based gait recognition. In P.J. Narayanan, S.K. Nayar, and H. Shum, editors, *Computer Vision - ACCV 2006*, volume 3851 of *Lecture Notes in Computer Science*, pages 807–816. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-31219-2. doi: 10.1007/11612032_81.

[30] S. Shirke, S.S. Pawar, and K. Shah. Literature review: Model free human gait recognition. In *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*, pages 891–895, April 2014. doi: 10.1109/CSNT.2014.252.

[31] A. Derbel, D. Vivet, and B. Emile. Access control based on gait analysis and face recognition. *Electronics Letters*, 51(10):751–752, 2015. ISSN 0013-5194. doi: 10.1049/el.2015.0767.

[32] I. Bouchrika and M.S. Nixon. People detection and recognition using gait for automated visual surveillance. In *Conference on Crime and Security. The Institution of Engineering and Technology*, pages 576–581, June 2006.

[33] R. Chellappa, A. Roy-Chowdhury, and S. Zhou. *Recognition of Humans and Their Activities Using Video*. Morgan and Claypool, 2006. ISBN 9781598290073. doi: 10.2200/S00002ED1V01Y200508IVM001.

[34] O. Barnich and M. Van Droogenbroeck. Frontal-view gait recognition by intra- and inter-frame rectangle size distribution. *Pattern Recogn.*

Lett., 30(10):893–901, July 2009. ISSN 0167-8655. doi: 10.1016/j. patrec.2009.03.014.

[35] A. Kale, N. Cuntoor, B. Yegnanarayana, A.N. Rajagopalan, and R. Chellappa. Gait analysis for human identification. In *Audio- and Video-Based Biometric Person Authentication*, volume 2688, pages 706–714. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-40302-9.

[36] M. Cheng, M. Ho, and C. Huang. Gait analysis for human identification through manifold learning and HMM. *Pattern Recognition*, 41(8):2541–2553, 2008. ISSN 0031-3203. doi: 10.1016/j.patcog.2007.11.021.

[37] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, Dec 2003. ISSN 0162-8828. doi: 10.1109/TPAMI.2003.1251144.

[38] S. Hong, H. Lee, and E. Kim. Automatic gait recognition using width vector mean. In *4th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 647–650, May 2009. doi: 10.1109/ICIEA. 2009.5138285.

[39] S. Mowbray and M.S. Nixon. Automatic gait recognition via Fourier descriptors of deformable objects. In J. Kittler and M.S. Nixon, editors, *Audio- and Video-Based Biometric Person Authentication*, volume 2688 of *Lecture Notes in Computer Science*, pages 566–573. Springer Berlin Heidelberg, 2003. ISBN 978-3-540-40302-9. doi: 10.1007/3-540-44887-X_67.

[40] C.P. Lee, A.W.C. Tan, and S.C. Tan. Gait recognition via optimally interpolated deformable contours. *Pattern Recognition Letters*, 34(6): 663–669, 2013. ISSN 0167-8655. doi: 10.1016/j.patrec.2013.01.013.

[41] S.D. Choudhury and T. Tjahjadi. Silhouette-based gait recognition using Procrustes shape analysis and elliptic Fourier descriptors. *Pattern Recognition*, 45(9):3414–3426, 2012. ISSN 0031-3203. doi: 10.1016/ j.patcog.2012.02.032. Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

[42] Y. Liu, J. Zhang, C. Wang, and L. Wang. Multiple HOG templates for gait recognition. In *21st International Conference on Pattern Recognition (ICPR)*, pages 2930–2933, Nov 2012.

[43] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893 vol. 1, June 2005. doi: 10.1109/CVPR.2005.177.

[44] Chen Wang, Junping Zhang, Jian Pu, Xiaoru Yuan, and Liang Wang. Chrono-gait image: A novel temporal template for gait recognition. In *Proceedings of the 11th European Conference on Computer Vision: Part I*, ECCV'10, pages 257–270, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-15548-0, 978-3-642-15548-2.

[45] R. Martín-Félez, J. Ortells, and R.A. Mollineda. Exploring the effects of video length on gait recognition. In *21st International Conference on Pattern Recognition (ICPR)*, pages 3411–3414, Nov 2012.

[46] R. Martín-Félez and T. Xiang. *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part I*, chapter Gait Recognition by Ranking, pages 328–341. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-33718-5. doi: 10.1007/978-3-642-33718-5_24.

[47] R. Martín-Félez and T. Xiang. Uncooperative gait recognition by learning to rank. *Pattern Recognition*, 47(12):379–3806, 2014. ISSN 0031-3203. doi: 10.1016/j.patcog.2014.06.010.

[48] Z. Lai, Y. Xu, Z. Jin, and D. Zhang. Human gait recognition via sparse discriminant projection learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(10):1651–1662, Oct 2014. ISSN 1051-8215. doi: 10.1109/TCSVT.2014.2305495.

[49] C. Rougier, E. Auvinet, J. Meunier, M. Mignotte, and J.A. de Guise. Depth energy image for gait symmetry quantification. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5136–5139, Aug 2011. doi: 10.1109/IEMBS.2011.6091272.

[50] P. Chattopadhyay, S. Sural, and J. Mukherjee. Frontal gait recognition from occluded scenes. *Pattern Recognition Letters*, 63:9–15, 2015. ISSN 0167-8655. doi: 10.1016/j.patrec.2015.06.004.

[51] K. Bashir, T. Xiang, and S. Gong. Gait recognition using gait entropy image. In *3rd International Conference on Crime Detection and Prevention (ICDP)*, pages 1–6, 2009.

[52] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes. Gait energy volumes and frontal gait recognition using depth images. In *International Joint Conference on Biometrics (IJCB)*, pages 1–6, Oct 2011. doi: 10.1109/IJCB.2011.6117504.

[53] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and 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)*, pages 1–8, Dec 2012. doi: 10.1109/DICTA.2012.6411694.

[54] J. Ryu and S. Kamata. Front view gait recognition using Spherical Space Model with Human Point Clouds. In *18th IEEE International Conference on Image Processing (ICIP)*, pages 3209–3212, Sept 2011. doi: 10.1109/ICIP.2011.6116351.

[55] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 439–446, 2001. doi: 10.1109/CVPR.2001.990508.

[56] A. Laurentini. The visual hull: a new tool for contour-based image under- standing. In *7th Scandinavian Conference on Image Processing*, pages 993–1002, 1991.

[57] R.D. Seely, S. Samangooei, M. Lee, J.N. Carter, and M.S. 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)*, pages 1–6, Sept 2008. doi: 10.1109/BTAS.2008.4699353.

[58] G. Ariyanto and M.S. Nixon. Model-based 3D gait biometrics. In *International Joint Conference on Biometrics (IJCB)*, pages 1–7, Oct 2011. doi: 10.1109/IJCB.2011.6117582.

[59] K. Yamauchi, B. Bhanu, and H. Saito. Recognition of walking humans in 3D: Initial results. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops.*, pages 45–52, June 2009. doi: 10.1109/CVPRW.2009.5204296.

[60] T. Connie, K.O.M. Goh, and A. Beng Jin Teoh. A review for gait recognition across view. In *Information and Communication Technology (ICoICT), 2015 3rd International Conference on*, pages 574–577, May 2015. doi: 10.1109/ICoICT.2015.7231488.

[61] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos. View-independent human motion classification using image-based reconstruction. *Image and Vision Computing*, 27(8):1194–1206, 2009. ISSN 0262-8856. doi: 10.1016/j.imavis.2008.11.008.

[62] A. Kale, A.K.R. Chowdhury, and R. Chellappa. Towards a view invariant gait recognition algorithm. In *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*, pages 143–150, July 2003. doi: 10.1109/AVSS.2003.1217914.

[63] F. Jean, A.B. Albu, and R. Bergevin. Towards view-invariant gait modeling: Computing view-normalized body part trajectories. *Pattern Recognition*, 42(11):2936–2949, 2009. ISSN 0031-3203. doi: 10.1016/j. patcog.2009.05.006.

[64] S. Jeong, T. Kim, and J. Cho. Gait recognition using description of shape synthesized by planar homography. *The Journal of Supercomputing*, 65(1):122–135, 2013. ISSN 0920-8542. doi: 10.1007/ s11227-013-0897-8.

[65] J. Han, B. Bhanu, and A.K. Roy-Chowdhury. A study on view-insensitive gait recognition. In *IEEE International Conference on Image Processing (ICIP)*, volume 3, pages III–297–300, Sept 2005. doi: 10.1109/ICIP.2005.1530387.

[66] M. Goffredo, I. Bouchrika, J.N. Carter, and M.S. Nixon. Self-calibrating view-invariant gait biometrics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(4):997–1008, Aug 2010. ISSN 1083-4419. doi: 10.1109/TSMCB.2009.2031091.

[67] F.M. Castro, M.J. Marín-Jiménez, and R. Medina-Carnicer. Pyramidal fisher motion for multiview gait recognition. In *22nd International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, August 24-28, 2014*, pages 1692–1697, 2014. doi: 10.1109/ICPR.2014.298.

[68] M.J. Marín-Jiménez, F.M. Castro, A. Carmona-Poyato, and N. Guil. On how to improve tracklet-based gait recognition systems. *Pattern Recognition Letters*, 68, Part 1:103–110, 2015. ISSN 0167-8655. doi: 10.1016/j.patrec.2015.08.025.

[69] F.M. Castro, M.J. Marín-Jiménez, and N. Guil. *Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015 Proceedings, Part I*, chapter Empirical Study of Audio-Visual Features Fusion for Gait Recognition,

pages 727–739. Springer International Publishing, Cham, 2015. ISBN 978-3-319-23192-1. doi: 10.1007/978-3-319-23192-1_61.

[70] G. Zhao, G. Liu, H. Li, and M. Pietikainen. 3D gait recognition using multiple cameras. In *7th International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 529–534, April 2006. doi: 10.1109/FGR.2006.2.

[71] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos. A framework for gait-based recognition using kinect. *Pattern Recognition Letters*, 68, Part 2:327–335, 2015. ISSN 0167-8655. doi: 10.1016/j.patrec.2015.06.020. Special Issue on Soft Biometrics.

[72] N. Liu and Y. Tan. View invariant gait recognition. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1410–1413, March 2010. doi: 10.1109/ICASSP.2010.5495466.

[73] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. Gait recognition using a view transformation model in the frequency domain. In A. Leonardis, H. Bischof, and A. Pinz, editors, *Computer Vision - ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 151–163. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-33836-9. doi: 10.1007/11744078_12.

[74] W. Kusakunniran, Q. Wu, H. Li, and 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)*, pages 1058–1064, Sept 2009. doi: 10.1109/ICCVW.2009.5457587.

[75] K. Bashir, T. Xiang, and S. Gong. Gait recognition without subject cooperation. *Pattern Recogn. Lett.*, 31(13):2052–2060, October 2010. ISSN 0167-8655.

[76] O. Barnich, S. Jodogne, and M. Van Droogenbroeck. Robust analysis of silhouettes by morphological size distributions. In J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, volume 4179 of *Lecture Notes in Computer Science*, pages 734–745. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-44630-9. doi: 10.1007/11864349_67.

[77] S. D.Choudhury and T. Tjahjadi. Gait recognition based on shape and motion analysis of silhouette contours. *Computer Vision and Image*

Understanding, 117(12):1770–1785, December 2013. ISSN 1077-3142. doi: 10.1016/j.cviu.2013.08.003.

[78] W. Zeng and C. Wang. Human gait recognition via deterministic learning. *Neural Networks*, 35:92–102, 2012. ISSN 0893-6080. doi: 10.1016/j.neunet.2012.07.012.

[79] A. Roy, S. Sural, and J. Mukherjee. Gait recognition using pose kinematics and pose energy image. *Signal Processing*, 92(3):780–792, March 2012. ISSN 0165-1684. doi: 10.1016/j.sigpro.2011.09.022.

[80] X. Huang and N.V. Boulgouris. Gait recognition with shifted energy image and structural feature extraction. *IEEE Transactions on Image Processing*, 21(4):2256–2268, April 2012. ISSN 1057-7149. doi: 10. 1109/TIP.2011.2180914.

[81] T. Krzeszowski, B. Kwolek, A. Michalczuk, A. Świtoński, and H. Josiński. View independent human gait recognition using markerless 3D human motion capture. In *International Conference on Computer Vision and Graphics*, volume 7594 of *Lecture Notes in Computer Science*, pages 491–500. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33563-1. doi: 10.1007/978-3-642-33564-8_59.

[82] J. Lu and Y. Tan. Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. *Pattern Recognition Letters*, 31(5): 382–393, 2010. ISSN 0167-8655. doi: 10.1016/j.patrec.2009.11.006.

[83] Y. Makihara, H. Mannami, A. Tsuji, M.A. Hossain, K. Sugiura, A. Mori, and Y. Yagi. The OU-ISIR Gait Database Comprising the Treadmill Dataset. *IPSJ Trans. on Computer Vision and Applications*, 4:53–62, Apr. 2012.

[84] J. Shutler, M. Grant, M. S. Nixon, and J.N. Carter. On a large sequence-based human gait database. In *Proc RASC*, pages 66–72. Springer Verlag, 2002.

[85] R. Gross and J. Shi. The CMU Motion of Body (MoBo) Database. Technical Report CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.

[86] S. Yu, D. Tan, and T. Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *18th International Conference on Pattern Recognition (ICPR)*, volume 4, pages 441–444, 2006.

[87] D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, and Rafael Muñoz-Salinas. The AVA Multi-View Dataset for Gait Recognition. In *Activity Monitoring by Multiple Distributed Sensing*, Lecture Notes in Computer Science, pages 26–39. Springer International Publishing, 2014. ISBN 978-3-319-13322-5. doi: 10.1007/978-3-319-13323-2_3.

[88] M.S Nixon, T.N Tan, and R. Chellappa. *Human identification based on gait*, volume 4 of *4*. Springer, 2006. ISBN 978-0-387-29488-9. doi: 10.1007/978-0-387-29488-9.

[89] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1505–1518, 2003. ISSN 0162-8828.

[90] D. Tan, K. Huang, S. Yu, and T. Tan. Efficient night gait recognition based on template matching. In *18th International Conference on Pattern Recognition (ICPR)*, volume 3, pages 1000–1003, 2006.

[91] T. Chalidabhongse, V. Kruger, and R. Chellappa. The UMD database for human identification at a distance. Technical report, University of Maryland, 2001.

[92] M. Hofmann, S. Sural, and G. Rigoll. Gait recognition in the presence of occlusion: A new dataset and baseline algorithm. In *Proc. 19th Intern. Conf. on Computer Graphics, Visualization and Computer Vision (WSCG), Plzen, Czech Republic*, 2011.

[93] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll. The TUM Gait from Audio, Image and Depth (GAID) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195–206, 2014. ISSN 1047-3203. doi: 10.1016/j.jvcir.2013.02.006. Visual Understanding and Applications with RGB-D Cameras.

[94] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri. Actions as space-time shapes. In *10th IEEE International Conference on Computer Vision (ICCV)*, volume 2, pages 1395–1402 Vol. 2, 2005.

[95] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local SVM approach. In *17th International Conference on Pattern Recognition (ICPR)*, volume 3, pages 32–36 Vol.3, 2004.

[96] R. Vezzani and R. Cucchiara. Video Surveillance Online Repository (ViSOR): an integrated framework. *Multimedia Tools Appl.*, 50(2): 359–380, November 2010. ISSN 1380-7501.

[97] D. Tran and A. Sorokin. Human activity recognition with metric learning. In *Computer Vision - ECCV 2008*, volume 5302 of *Lecture Notes in Computer Science*, pages 548–561. Springer Berlin Heidelberg, 2008. ISBN 978-3-540-88681-5.

[98] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas. The I3DPost Multi-View and 3D Human Action/Interaction Database. In *Proceedings of the 2009 Conference for Visual Media Production*, CVMP '09, pages 159–168, Washington, DC, USA, 2009. IEEE Computer Society. ISBN 978-0-7695-3893-8. doi: 10.1109/CVMP.2009.19.

[99] S. Singh, S.A. Velastin, and H. Ragheb. MuHAVi: A Multicamera Human Action Video Dataset for the Evaluation of action recognition methods. In *Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 48–55, 2010.

[100] D. Weinland, R. Ronfard, and E. Boyer. Free viewpoint action recognition using motion history volumes. *Comput. Vis. Image Underst.*, 104 (2):249–257, November 2006. ISSN 1077-3142.

[101] H. Iwama, M. Okumura, Y. Makihara, and Y. Yagi. The OU-ISIR gait database comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. on Information Forensics and Security*, 7, Issue 5:1511–1521, Oct. 2012.

[102] B.G. Baumgart. *Geometric Modeling for Computer Vision*. PhD thesis, Stanford, CA, USA, 1974. AAI7506806.

[103] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, Feb 1994. ISSN 0162-8828. doi: 10.1109/34.273735.

[104] L. Díaz-Más, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and R. Medina-Carnicer. Shape from silhouette using Dempster-Shafer theory. *Pattern Recognition*, 43(6):2119–2131, 2010. ISSN 0031-3203. doi: 10.1016/j.patcog.2010.01.001.

[105] G. Haro. Shape from silhouette consensus. *Pattern Recognition*, 45(9): 3231–3244, 2012. ISSN 0031-3203. doi: 10.1016/j.patcog.2012.02.029.

Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011).

[106] S. Yous, H. Laga, M. Kidode, and 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, pages 71–77, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-912-8. doi: 10.1145/1321261.1321274.

[107] D.T. Drinkwater and W.D. Ross. Anthropometric fractionation of body mass. In *Kinanthropometry*, volume 2, pages 178–189. University Park Press, Baltimore, 1980.

[108] D. López-Fernández, F.J Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, R. Muñoz-Salinas, and R. Medina-Carnicer. Viewpoint-independent gait recognition through morphological descriptions of 3d human reconstructions. *Image and Vision Computing*, 48–49:1 – 13, 2016. ISSN 0262-8856. doi: 10.1016/j.imavis.2016.01.003.

[109] D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications*, 26(7):1079–1094, 2015. ISSN 1432-1769. doi: 10.1007/s00138-015-0707-9.

[110] D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. A new approach for multi-view gait recognition on unconstrained paths. *Journal of Visual Communication and Image Representation*, 38:396 – 406, 2016. ISSN 1047-3203. doi: 10.1016/j.jvcir.2016.03.020.

[111] D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Multi-view gait recognition on curved trajectories. In *Proceedings of the 9th International Conference on Distributed Smart Cameras*, ICDSC '15, pages 116–121, New York, NY, USA, 2015. ACM. ISBN 978-1-4503-3681-9. doi: 10.1145/2789116.2789122.

Appendix A

Impact report

- D. López-Fernández, F.J Madrid-Cuevas, A. Carmona-Poyato, M.J. Marín-Jiménez, R. Muñoz-Salinas, and R. Medina-Carnicer. Viewpoint-independent gait recognition through morphological descriptions of 3d human reconstructions. Image and Vision Computing, 48-49:1-13, 2016. ISSN 0262-8856. doi: 10.1016/j.imavis.2016.01.003.

Journal: Image and Vision Computing
Impact factor: JCR 2014: 1.587

Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING
Position of journal: 22 of 104
Quartile in category: Q1

Category: COMPUTER SCIENCE, THEORY & METHODS
Position of journal: 22 of 102
Quartile in category: Q1

Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE
Position of journal: 52 of 123
Quartile in category: Q2

Category: ENGINEERING, ELECTRICAL & ELECTRONIC
Position of journal: 96 of 249
Quartile in category: Q2

Category: OPTICS
Position of journal: 41 of 87

Quartile in category: Q2

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. Entropy volumes for viewpoint-independent gait recognition. *Machine Vision and Applications*, 26(7):1079-1094, 2015. ISSN 1432-1769. doi: 10.1007/s00138-015-0707-9.

Journal: Machine Vision and Applications
Impact factor: JCR 2014: 1.351

Category: COMPUTER SCIENCE, CYBERNETICS
Position of journal: 11 of 24
Quartile in category: Q2

Category: ENGINEERING, ELECTRICAL & ELECTRONIC
Position of journal: 113 of 249
Quartile in category: Q2

Category: COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE
Position of journal: 64 of 123
Quartile in category: Q3

- D. López-Fernández, F.J. Madrid-Cuevas, A. Carmona-Poyato, R. Muñoz-Salinas, and R. Medina-Carnicer. A new approach for multi-view gait recognition on unconstrained paths. Journal of Visual Communication and Image Representation, 38:396-406, 2016. ISSN 1047-3203. doi: 10.1016/j.jvcir.2016.03.020.

Journal: Journal of Visual Communication and Image Representation
Impact factor: JCR 2014: 1.218

Category: COMPUTER SCIENCE, INFORMATION SYSTEMS
Position of journal: 57 of 139
Quartile in category: Q2

Category: COMPUTER SCIENCE, SOFTWARE ENGINEERING
Position of journal: 39 of 104
Quartile in category: Q2