Gait representation and recognition from temporal co-occurrence of flow fields

Hatem A. Rashwan, Miguel Ángel Garcia, Sylvie Chambon, Domenec Puig

Gait representation and recognition from temporal co-occurrence of flow fields

Hatem A. Rashwan[1] · Miguel Ángel García[3] · Sylvie Chambon[1] · Domenec Puig[2]

Abstract
This paper proposes a new gait representation that encodes the dynamics of a gait period through a 2D array of 17-bin histograms. Every histogram models the co-occurrence of optical flow states at every pixel of the normalized template that bounds the silhouette of a target subject. Five flow states (up, down, left, right, null) are considered. The first histogram bin counts the number of frames over the gait period in which the optical flow for the corresponding pixel is null. In turn, each of the remaining 16 bins represents a pair of flow states and counts the number of frames in which the optical flow vector has changed from one state to the other during the gait period. Experimental results show that this representation is significantly more discriminant than previous proposals that only consider the magnitude and instantaneous direction of optical flow, especially as the walking direction gets closer to the viewing direction, which is where state-of-the-art gait recognition methods yield the lowest performance. The dimensionality of that gait representation is reduced through principal component analysis. Finally, gait recognition is performed through supervised classification by means of support vector machines. Experimental results using the public CMU MoBo and AVAMVG datasets show that the proposed approach is advantageous over state-of-the-art gait representation methods.

Keywords Gait recognition · Action recognition · Temporal co-occurrence · Optical flow · PCA · Support vector machine

1 Introduction

Human gait is the repetitive pattern of motion of the limbs of a person while moving over a solid surface. Every person has a distinctive gait when walking or running and other specific gait features that characterize gender and age. Therefore, automatic gait recognition is of interest to a variety of applications, such as biometrics and video surveillance. That interest has driven the development of a wide variety of gait recognition algorithms that can be broadly classified into model-based and appearance-based approaches, which basically differ in the feature space utilized to represent the dynamics associated with gait prior to classification. While model-based approaches are only applicable to humans, appearance-based methods are generalizable to both humans and animals. Most of the approaches assume the application of a preprocessing stage in order to segment the target individual (foreground) from the background of the scene. Eventually, a shadow detection algorithm may also be required in the case of outdoor scenes. Furthermore, the extracted silhouettes are usually normalized to a predefined size in order to gain invariance to scale changes. The majority of gait recognition approaches have been devised from lateral views, that is, with a line of sight mostly perpendicular to the sagittal plane.

Model-based approaches assume some a priori knowledge about the geometry of the human body, which imposes several physical constraints on the motion of the lower limbs. For example, in [1], a 73 dimensional measurement vector

✉ Hatem A. Rashwan
hatem rashwan@ieee.org; hatem.mahmoud@enseeiht.fr

Miguel Ángel García
miguelangel.garcia@uam.es

Sylvie Chambon
sylvie.chambon@enseeiht.fr

Domenec Puig
domenec.puig@urv.cat

1 CNRS-IRIT, INP-ENSEEIHT, Université de Toulouse, Toulouse, France

2 Department of Electrical and Computer Engineering, Universitat Rovira i Virgil, Tarragona, Spain

3 Department of Electronic and Communications Technology, Universidad Autonoma de Madrid, Madrid, Spain

is formed by aggregating parameters such as the subject's speed, gait frequency, body proportions and coefficients of rotation sinusoidal models for the hip, knee and ankle. Mean and covariances are computed for the measurement vectors corresponding to every subject, and Bayesian classification is finally applied. In addition, in [2], least-squares linear regression is applied in order to extract the lines that represent the thighs and shins from the edges of the subject's silhouette. A genetic algorithm determines the harmonic coefficients that characterize the temporal variation of shin angles. Classification is performed through K-NN. In addition, [3] proposes a three-phase model that combines spatiotemporal shape and dynamic motion characteristics of silhouette contours to identify human's gait. For identifying a subject, the match scores obtained by analyzing the local and global gait characteristics obtained in the three phases are combined using a weighted sum. Alternatively, [4] proposes a set of dynamic features based on a human skeleton model for gait recognition invariant to the walking speed. In particular, time series of changing joint angles, angle phase differences, mass ratios of body parts and distances of body parts from the body center are computed over the entire gait sequence. Classification is performed by comparing the features extracted from the probe subjects with the tested ones using the Euclidean distance.

Alternatively, appearance-based methods only take into account the consecutive foreground regions corresponding to a same subject, with no assumptions about the latter other than the periodicity of its motion. The gait representation utilized by most appearance-based methods is based on one or several images (or 2D arrays) obtained from the sequence of foreground regions of the target subject during a gait period. Preliminary works include [5] where two images are generated from the subject's binary silhouettes: a binary Motion-Energy Image (MEI), which represents where motion has occurred in the last images, and a gray-scale Motion-History Image (MHI) in which intensity is proportional to the recency of motion. The means and covariances of the Hu's moments of those images are used as lower-dimensional features, while classification is performed through the nearest neighbor rule using the Mahalanobis distance between features. An extension of MHI is presented in [6]. In particular, the frames of half a gait period are partitioned into a constant number of disjoint subsets (temporal windows). The MHI is computed for every subset and a histogram of oriented gradients (HOG) obtained for each MHI. The gait period is then characterized by the set of HOGs. The Euclidean distance between sets of HOGs is used for recognition.

In turn, [7] proposed the Gait Energy Image (GEI) as the average of the normalized binary silhouettes corresponding to the same gait period. GEI has become a widely used gait representation thereafter. In that proposal, dimensionality reduction is done trough PCA and Multiple Discriminant Analysis (MDA), whereas classification is based on the minimum Euclidean distance between features. Similar gait representations based on the average of binary silhouettes were proposed in [8]. Where PCA is applied for dimensionality reduction and the Euclidean distance used to measure the similarity of the averaged silhouettes. A variation of GEI was proposed in [9], in which the lower subimage from the knees to the feet is removed. The result is the Head-Torso Image (HTI). The nearest neighbor rule is utilized for classification by using the sum of absolute differences as a distance. Following a different approach also based on GEI, [10] applies Gabor filters to the GEI representation in order to obtain three new images that encode the filters response over the different directions and scales. General Tensor Discriminant Analysis (GTDA) is applied in order to extract features and reduce the dimensionality of the problem. Finally, classification is done through Linear Discriminant Analysis (LDA). In that line, [11] applies the Radon transform to the GEI representation to compute the 2D projection image along varying angles between 0 and 180^{circ}. All the projections are appended to form a 1D Radon Transform vector, and PCA is applied to reduce the dimensionality of the final feature space. Classification is performed through the nearest neighbor rule by using the Euclidean distance.

Alternatively, in [12], GEI images are mapped to a Grassmannian manifold. Then, Grassmannian locality preserving discriminant analysis is used for improving recognition accuracy. Following a different approach, Kusakunniran et al. [13] proposed a gait recognition method based on calculating an adaptation of the local binary pattern (LBP) descriptor called a weighted binary pattern (WBP) given a sequence of aligned silhouettes during a gait cycle. The distance is then calculated to measure the similarity between the various gait signatures. In [14], two features based on multi-scale LBP (MLBP) and Gabor filter banks extracted locally from Regions of Interest (ROIs) are used to represent the dynamic areas in GEI. A spectra Regression Kernel Discriminant Analysis reduction algorithm is then used to reduce the dimensionality of these features. Alternatively, Wang et al. [15] propose the Chrono-Gait Image (CGI) as a gait representation also generated from the contours of the normalized binary silhouettes. In particular, every contour is mapped to an RGB color depending on its position within the gait period. All colored contours corresponding to a quarter of the gait period are added. The CGI image is obtained by accumulating the contour additions corresponding to different quarters of the gait period. PCA and LDA are used for dimensionality reduction, while classification is done by means of K-NN using the Euclidean distance. In addition, view-invariant gait recognition based on extracting parametric 3D gait models from three cameras and using partial similarity matching to improve recognition rates is

proposed in [16,17]. Sparse representation-based classification is then performed for gait classification.

All the aforementioned appearance-based methods encode the dynamic information of a gait period based on the visual appearance of the foreground region associated with the target subject. However, with the exception of [18], all those methods start with the binary silhouette of the subject, hence disregarding the internal appearance of the regions. In practice, only the shape of the region contour is thus taken into account. Notwithstanding, an alternative family of appearance-based gait recognition methods also take into account the internal appearance of the foreground regions by analyzing the optical flow fields computed within those regions, and encoding the variation of those fields over the gait period. The advantage is that those gait representations do not lose as much discriminative power as their counterparts as the angle between the walking and the viewing directions departs from 90^{circ}. In the extreme case where both directions are aligned (i.e., the line of sight is perpendicular to the coronal plane), most of the aforementioned methods have almost no shape information to discriminate among different gaits.

Recently, [19] proposed the discretization of the flow directions, encoding them into a single 5-bin histogram associated with every pixel. In particular, the first bin keeps the number of frames within the gait period where that pixel has no motion, whereas the other four bins keep the number of frames with motion up, down, left and right, respectively. A Motion Intensity Image (MII) is obtained from the values of the first bin for all pixels of the foreground region. Similarly, four Motion Direction Images (MDI) are obtained from the other bins. The MII and the first three MDIs are independently used for recognition. Component and Discriminant Analysis (CDA) is applied for dimensionality reduction. Classification is done using the Euclidean distance between feature vectors in the CDA subspace. A similar approach to MII [19] later proposed in [20] defines the Gait Flow Image (GFI) as a gait representation in which the value associated with each pixel is the number of frames within the gait period in which that pixel has no optical flow. GFIs can be recognized by direct matching or after dimension reduction using LDA followed by classification through the nearest neighbor rule.

Furthermore, [21] presented a gait recognition approach that combines low-level motion descriptors based on optical flow, which are extracted from dense local spatiotemporal features, with a multilevel gait encoding based on the Pyramidal Fisher Motion (PFM) gait descriptor. This method supports single and multiple viewpoints. The computed optical flow is represented by means of the Divergence-Curl-Shear (DCS) descriptor. Then, the DCS descriptors are used to build a person-level gait descriptor based on Fisher vectors. A Fisher vector is computed within each cell of the spatiotemporal grid of the subject's body. In order to build a pyramidal representation, different grid configurations are combined. Then, a final feature vector is computed as the concatenation of the cell-level Fisher vectors for all the levels of the pyramid. Finally, SVM is utilized for the classification stage to recognize the different gaits. In [22,23], based on optical flow estimated from consecutive frames, a discriminative biometric signature (DBS) is used for representing the gait cycle. The descriptor accounts for both spatial and temporal DBS features considering local and global cues of the motion process.

In addition, some works of action recognition are based on capturing the local motion information of a set of images [24,25]. Regarding to [24], the authors have proposed a descriptor based on combining histograms of oriented gradients (HOG) and histograms of optical flow (HOF) for human action recognition in a video. They utilized HOG to describe static appearance information, whereas HOF to capture the local motion information. In turn, [25] have introduced a descriptor based on motion boundary histograms (MBH) which relies on differential optical flow. For each interest point detected, they consider a volume of 32×32-pixel window along 15 consecutive frames, such that the window is centered at the trajectory of the corresponding point. That trajectory is estimated through optical flow and median filtered. Then, you divide the volume into $2 \times 2 \times 3$ cells and compute MBHx and MBHy for each cell. MBHx is an 8-bin histogram of the derivative of horizontal optical flow for the pixels within the cell. MBHy is the same for vertical. The dimension is 96 (i.e., $2 \times 2 \times 3 \times 8$) for both MBHx and MBHy. In addition, [26] proposed an extension for MBH by using 3D co-occurrence descriptors, which take into account the spatiotemporal context within a local 32×32-pixel window along a set of consecutive frames. They then apply the Bag-of-Features model for each vectorized 3D co-occurrence matrix, and leverage a multi-channel kernel SVM to combine channels for the descriptor.

This paper proposes a new gait representation that encodes the dynamics of a gait period through a 2D array of 17-bin histograms. Every histogram models the co-occurrence of optical flow states at every pixel of the normalized template that bounds the silhouette of the target subject. Similar to [19], four flow directions are considered (i.e., up, down, left, right) and the first histogram bin counts the number of frames over the gait period in which the optical flow for the corresponding pixel is null. However, each of the remaining 16 bins now represents a change in motion state (e.g., up/up, up/down, down/up, down/down, null/left, left/null, up/null, ...) and counts the number of frames in which the optical flow vector has changed from one state to the other during the gait period. Experimental results show that this representation is significantly more discriminant than previous proposals that only consider the magnitude and instantaneous direction of

optical flow [19,20], especially as the walking direction gets closer to the viewing direction, which is where state-of-the-art gait recognition methods yield the worst performance. The dimensionality of the proposed gait representation is reduced through PCA. Finally, gait recognition is performed through supervised classification by means of a set of Support Vector Machines (SVM).

The rest of this paper is organized as follows: Section 2 introduces the basic concepts of optical flow and describes the proposed gait representation, including the dimensionality reduction stage. The classification stage is presented in Sect. 3. Experimental results are shown and discussed in Sect. 4, including a comparison with state-of-the-art gait representation methods using the public CMU MoBo and AVAMVG datasets. Finally, conclusions and future lines are highlighted in Sect. 5.

2 Gait representation through temporal co-occurrence of flow fields

This section describes a new gait representation that encodes the dynamics of a gait cycle by statistically modeling the temporal co-occurrence of optical flow states during a complete gait period. Similar to previous approaches [7,19,20], a background subtraction scheme is previously applied to the input video sequence for extracting the foreground regions corresponding to the target subject. To gain invariance to scale, every foreground region is scaled such that its height is set to a predefined size h and its width adjusted accordingly in order to keep the original aspect ratio. The result is then centered into an $h \times w$ template. The gait period is obtained through maximum entropy estimation [27].

Given the set of n consecutive $h \times w$ templates belonging to a same gait period, $\{T_0, T_1, \ldots T_{n-1}\}$, an optical flow field is computed for every template. The flow field for the $i-th$ template is denoted as w_i and computed using two consecutive templates: T_i and T_j, where $j = (i+m)\%n$, $0 < m < n$ and $0 \leq i, j < n$, where % is the module operator. A state-of-the-art optical flow algorithm previously proposed by the authors in [28] is applied for estimating accurate dense flow fields. That algorithm is robust to noise and outliers in being based on a discontinuity-preserving filtering stage that applies stick tensor voting, Fig. 2(c).

Every optical flow field is a two-dimensional vector field w_i defined in the domain of the $h \times w$ templates. That field is constituted by both a vertical displacement field v_i and a horizontal displacement field u_i, $w_i = (u_i, v_i)$. Therefore, the flow vector corresponding to a given template pixel at coordinates (x, y) is $w_i(x, y) = (u_i(x, y), v_i(x, y))$.

Let $R_i(x, y)$ and $L_i(x, y)$ be two Boolean variables that indicate that the associated pixel has a positive (Right) or a negative (Left) horizontal displacement, respectively, which can be defined as:

$$L_i(x, y) = 1 - \mathcal{H}(u_i(x, y)), \tag{1}$$

$$R_i(x, y) = 1 - \mathcal{H}(-u_i(x, y)), \tag{2}$$

where $\mathcal{H}$ is the Heaviside function. In addition, let $H_i(x, y)$ be a Boolean variable that represents that the magnitude of the flow vector in the horizontal direction, $u_i(x, y)$, is null, that is, below a very small threshold ϵ (in this work, $\epsilon < 1.0$).

$$H_i(x, y) = \begin{cases} 1 & -\epsilon \leqslant u_i(x, y) \leqslant \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

If $H_i(x, y)$ is true, variables $R_i(x, y)$ and $L_i(x, y)$ are forced to the false state.

Similarly, let $D_i(x, y)$ and $U_i(x, y)$ be two mutually excluding Boolean variables that denote that the pixel's vertical displacement is positive (Down) or negative (Up), respectively. In addition, $V_i(x, y)$ is a Boolean variable that represents that the magnitude of the flow vector in the vertical direction, $v_i(x, y)$, is null. If $V_i(x, y)$ is true, $U_i(x, y)$ and $D_i(x, y)$ are set to false.

$$U_i(x, y) = 1 - \mathcal{H}(v_i(x, y)), \tag{4}$$

$$D_i(x, y) = 1 - \mathcal{H}(-v_i(x, y)), \tag{5}$$

$$V_i(x, y) = \begin{cases} 1 & -\epsilon \leqslant v_i(x, y) \leqslant \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The variation of the n flow fields estimated for the whole gait period is then statistically modeled by an $h \times w$ array of 17-bin histograms, with a histogram associated with every template pixel. The first bin (HV_i) counts the total number of templates over the gait period where the associated pixel has no apparent motion (i.e., $H_i(x, y) = V_i(x, y) = 1$). Therefore, that bin represents the magnitude of motion as:

$$HV_i(x, y) = H_i(x, y) \times V_i(x, y), \tag{7}$$

where $\times$ is the multiplication operator.

In turn, each of the remaining 16 bins models a co-occurrence of optical flow states between a pair of templates separated by m frames, T_i and T_j. The 16 bins are noted as: HL_i, HR_i, VU_i, VD_i, LH_i, RH_i, UV_i, DV_i, LL_i, LR_i, RL_i, RR_i, UU_i, UD_i, DU_i, DD_i. For example, $RH_i(x, y)$ counts the number of templates where both $H_i(x, y)$ and $R_j(x, y)$ are true, indicating that the horizontal displacement is null at template i and the horizontal component of the flow vector then becomes positive at template j, that is, m templates later:

$$HR_i(x, y) = H_i(x, y) \times R_j(x, y). \tag{8}$$

The same pixel would also contribute to, for instance, bin VU_i provided $U_j(x, y)$ and $V_i(x, y)$ were true, meaning that

the vertical displacement at template i is null and m templates later becomes negative.

$$VU_i(x, y) = V_i(x, y) \times U_j(x, y). \tag{9}$$

Notice that both $R_j(x, y)$ and $U_j(x, y)$ can be true simultaneously, indicating that the angle of the flow vector is in the first quadrant. In that case, $L_j(x, y)$, $D_j(x, y)$, $H_j(x, y)$ and $V_j(x, y)$ will be false. Every 17-bin histogram is normalized through the L_1 norm, which yields slightly better results than the L_2 norm and other normalization policies that have been evaluated, see Fig. 1. For a gait cycle, the $h \times w$ array of 17 bins is computed between every pair of frames. Finally, an accumulation $h \times w$ histogram of 17-bins is computed. For instance, the bin related to $LL(x, y)$ can be defined as:

$$LL(x, y) = \sum_{i=0}^{n-1} LL_i(x, y), \tag{10}$$

The complete equations are shown in the Annex.

Similar to [19], the $h \times w$ array of 17-bins histograms can also be interpreted as a set of 17 gray-level images of $h \times w$ pixels each, such that every image keeps the values of one of the bins. Thus, the first bin yields the Motion Intensity Image (MII) [19], in which the lower the value the higher the percentage of motion during the gait period. That image resembles the popular Gait Energy Image (GEI) [7]. However, since the GEI is computed using binary silhouettes instead of optical flow fields, the measurement of motion intensity is indirect and far less accurate than the one utilized for generating the MII. The remaining 16 bins yield corresponding images referred to in this work as Motion Co-occurrence Images (MCI).

Figure 2 shows two consecutive frames from one of the video sequences belonging to the CMU dataset (a, b), the computed optical flow field for those frames (c), the MII (d) and the 16 MCIs (e-p) obtained for the gait period, which comprises 34 image frames ($n = 34$) in this example. The minimum separation between pairs of frames has been set to $m = 1$, which as shown in Sect. 4 yields optimal or near-optimal results for both slow and fast walking speeds. Regarding the MII, the lowest values representing high motion are observed at the legs of the subject as expected. As for the MCIs, the first two rows show the motion co-occurrences for the horizontal (e–h) and vertical (i–l) components. In turn, the last two rows show the co-occurrences between null motion and appreciable motion states (m–p), and vice-versa (q–t). Notice that all co-occurrence images are different and provide distinctive and complementary information that, as a whole, conveys the dynamics of the analyzed gait.

Once the motion patterns of the gait cycle are represented through the proposed $h \times w$ array of 17-bin histograms, it is necessary to reduce the dimensionality of that array in order to make it more compact and hence manageable. Similarly [7,11,15,29], this is done through PCA, which applies an orthogonal transformation in order to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. When the number of principal components is less than the number of original variables, the dimensionality of the problem is reduced while retaining as much of the variance in the dataset as possible, which is useful for further discrimination tasks. The principal components are the right singular vectors of the Singular Value Decomposition (SVD) of the centered data matrix. The dimensionality is reduced by only considering the first r largest singular values and their corresponding right singular vectors.

In particular, the MII and each of the MCIs are arranged as a single hw column vector. The original $h \times w$ array of 17-bin histograms is then represented as an $hw \times 17$ matrix X. Therefore, the number of variables in this problem is 17 and the number of observations hw. Using SVD, a 17×17 matrix of right singular vectors W is obtained, along with a 17×17 diagonal matrix of singular values Σ. The principal components of X are the columns of W, which are sorted in descending order of their corresponding singular values. Matrix W is then truncated along its columns to retain the first r singular vectors, yielding W_r, which is a $17 \times r$ array. In this work, r has experimentally been set to six as shown in Sect. 4. Matrix W_r is then mapped to a single $17r$ column vector, which is the final gait representation (gait vector).

3 Gait recognition

Gait recognition in this work is cast as a multiclass supervised classification problem based on support vector machines (SVMs). In particular, a binary SVM is trained for every considered gait class. A one-versus-all training approach is applied. Thus, during the offline training stage, every SVM is trained with the gait vectors corresponding to the class associated with that SVM as positive examples, and the gait vectors corresponding to the remaining classes as negative examples. In turn, during the online classification stage, an input gait vector is classified into the class corresponding to the SVM with the largest output function, thus following a winner-takes-all strategy. The experimental results conducted in this work have yielded the best classification results by using nonlinear SVMs with a kernel based on a Gaussian radial basis function (RBF) ($\gamma = 0.2$) and soft margin parameter ($C = 1$). In addition, the mapping kernel RBF is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \tag{11}$$

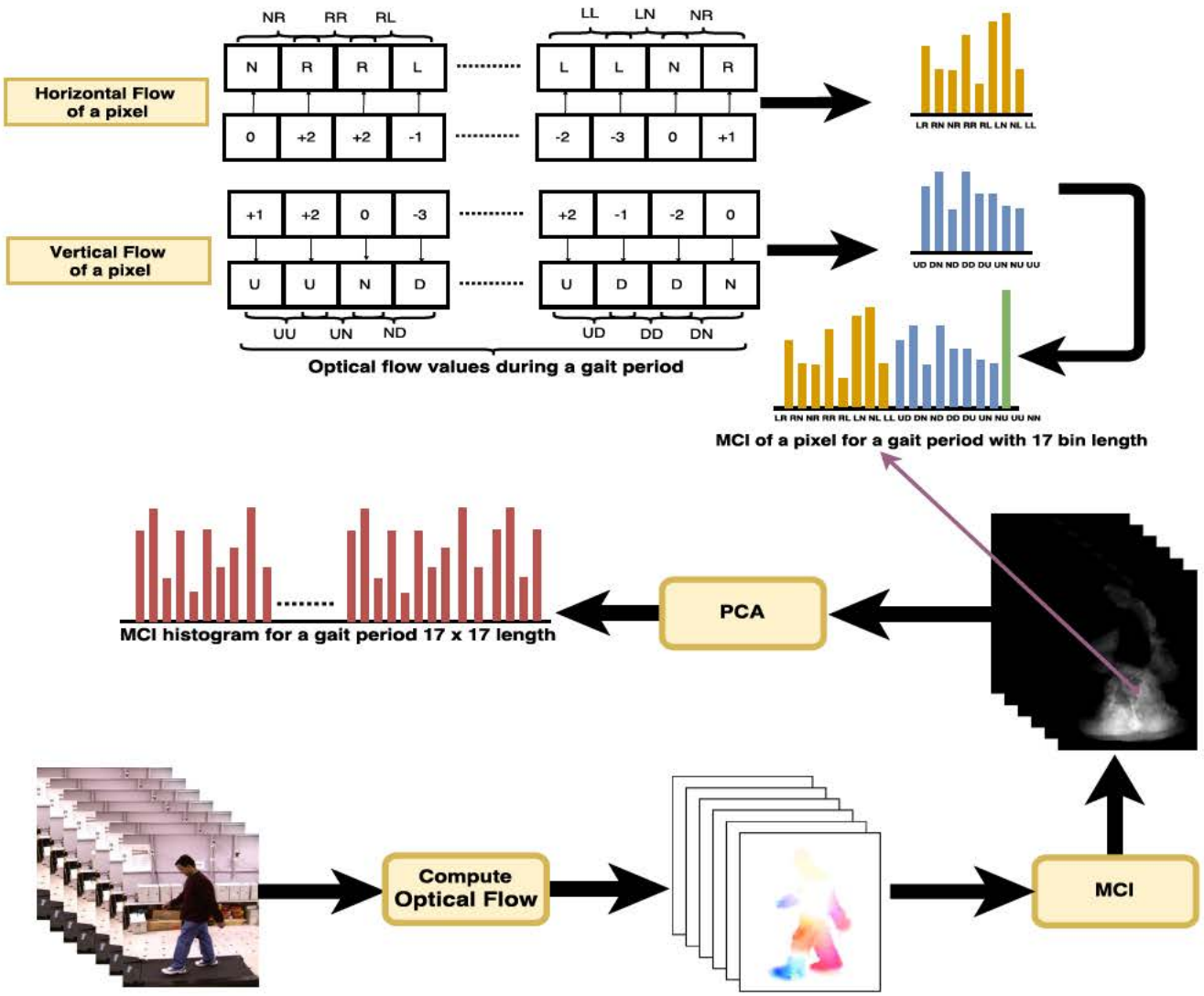

Fig. 1 Example of computation of MCI for a sequence of images during a gait period

where $\gamma = 1/2\sigma^2$, $\|x_i - x_j\|^2$ is the squared Euclidean distance between the two feature vectors x_i and x_j, and σ is a free parameter of the standard deviation.

4 Experimental results

In all experiments, we assume that the subjects are previously extracted from original images. In particular, our dataset [30] already contains the bounding box and silhouette of every subject. In turn, for the dataset in [31], we used pedestrian detection based on the HOG descriptor and background subtraction based on Codebook models to extract the bounding boxes and silhouettes, respectively. Both algorithms are implemented in OpenCV.

4.1 Experiments on the CMU MoBo dataset

The proposed gait representation and recognition method have been evaluated using the public video sequences of the CMU Motion of Body (MoBo) dataset [30]. That database contains video sequences corresponding to four different gaits (slow walk, fast walk, incline walk, walking with a ball) performed on a treadmill by 25 subjects. Every individual was simultaneously recorded by six high-resolution color cameras evenly distributed around the treadmill as shown in Fig. 3. Thus, every sequence corresponds to one subject executing a particular gait and captured from a specific viewpoint. Therefore, a total of 600 video sequences have been processed, each containing 340 image frames approximately. A gait vector has been computed for every sequence as described in Sect. 2.

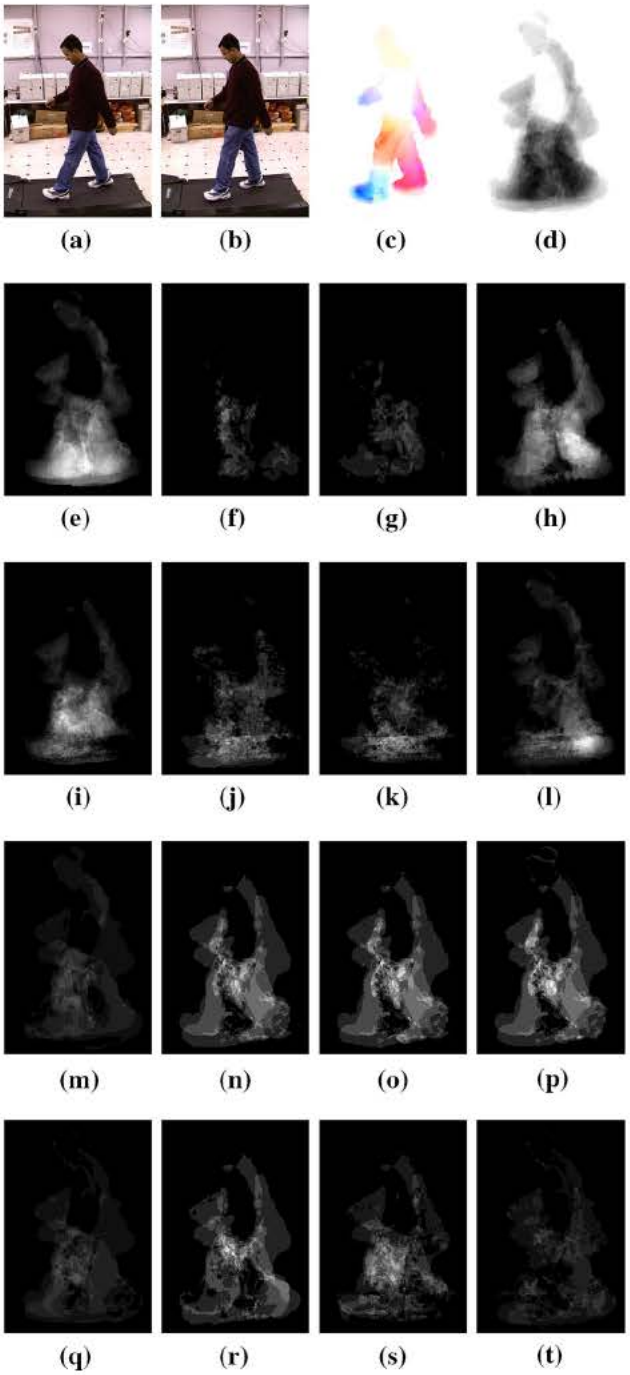

Fig. 2 **a**, **b** Two consecutive frames within sequence 87 from the CMU MoBo dataset, **c** corresponding flow field, **d** MII, **e** RR, **f** RL, **g** LR, **h** LL, **i** UU, **j** UD, **k** DU, **l** DD, **m** HR, **n** HL, **o** VU, **p** VD, **q** RH, **r** LH, **s** UV and **t** DV

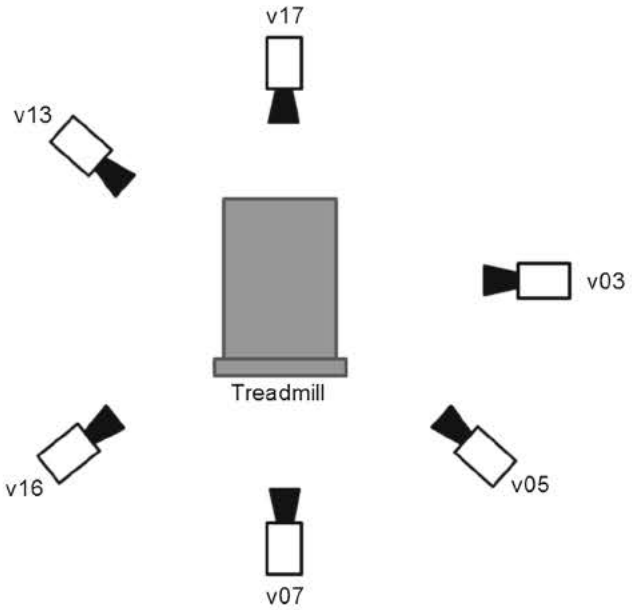

Fig. 3 Six camera viewpoints considered in the CMU MoBo dataset

As indicated in Sect. 3, one SVM per gait has been trained with positive examples of gait vectors corresponding to the six viewpoints of the associated gait for the 25 subjects. The remaining training gait vectors associated with the alternative gaits have been used as negative examples. The training set was obtained through holdout cross-correlation by randomly picking half of the available video sequences (i.e., 300 training gait vectors). The remaining vectors constituted the test set. The holdout cross-correlation process was repeated 30 times, thus obtaining 30 different pairs of training and test sets. The final classification rates reported in this work have been obtained by averaging the rates corresponding to every test set.

The first experiment evaluates the effect of the number of principal components, r, on the classification rates for the different tested gaits with disregard of the camera viewpoint. Table 1 shows the average classification rates for the tested gaits by considering the first r singular vectors, with r ranging between 2 and 10. Notice that classification rates do not grow significantly beyond 6 components. Therefore, $r = 6$ is a convenient trade-off between classification accuracy and computational cost. Thus, the classification rate is:

$$CR(\%) = R_c/R_t, \tag{12}$$

where R_c and R_t are the number of subjects correctly identified by the SVM prediction and the total number of subjects in the dataset, respectively.

In the second experiment, the proposed gait representation based on temporal co-occurrence of flow fields, referred to as MCI, has been compared to three alternative gait representations: the widely used GEI representation [7], and the two gait representations based on flow fields, respectively, proposed in [19], hereafter referred to as GFF, and [20], referred to as GFI. In order to compare those representations with the one proposed in this work solely in terms of discrimination capabilities, both the dimensionality reduction stage (PCA-based) and gait recognition stage (SVM-based) proposed in

Table 1 Classification rates with the proposed technique for a varying number r of principal components between 2 and 10

Gait	2	3	4	5	6	8	10
Slow	0.88	0.88	0.92	0.95	0.97	0.97	0.97
Fast	0.86	0.89	0.92	0.95	0.97	0.97	0.97
Incline	0.87	0.89	0.93	0.95	0.96	0.97	0.97
Ball	0.87	0.89	0.92	0.95	0.96	0.97	0.96
Average	0.87	0.89	0.92	0.95	0.96	0.97	0.97

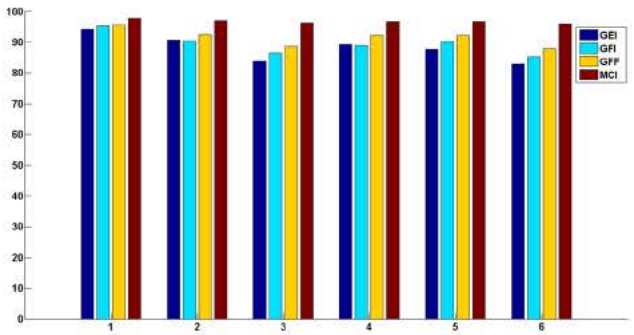

Fig. 4 Average classification rates for a slow gait from the CMU MoBo dataset yielded by the four gait representation methods: GEI, GFI, GFF, MCI (proposed method), distinguishing among the six camera viewpoints: 1 (v03), 2 (v05), 3 (v07), 4 (v13), 5 (v16) and 6 (v17), see Fig. 3

this work have also been applied to those three alternative representations instead of the corresponding stages proposed in their original proposals. In that way, the difference of classification rates among them will only be attributable to the descriptive power of each gait representation.

The average classification rates corresponding to the four evaluated gaits for each of the four tested gait representations are shown in Figs. 4, 5, 6 and 7, distinguishing among each of the six camera viewpoints. Notice that the proposed representation (MCI) is significantly superior to the others for recognizing all gaits from the viewpoints aligned with the walking direction (i.e., v07 and v17). This is the hardest scenario, as shown by the drop in performance close to ten percent that undergo the other methods. In contrast, the proposed representation has a more stable behavior, with classification rates above 95 percent in all cases. These results show that the temporal co-occurrence of flow fields is able to better capture the dynamics of the gait patterns independently of the speed of the subject and the camera viewpoint, especially when the silhouette of the target does not correspond to a lateral view, as well as for fast walking speeds.

In addition, Tables 2, 3, 4 and 5 show the confusion matrices among gaits (including all their viewpoints) corresponding to the proposed gait representation (MCI), GFF, GFI and GEI, respectively. The proposed method yields the

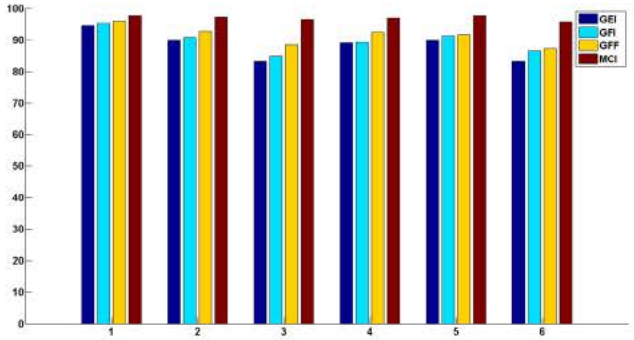

Fig. 5 Average classification rates for a fast gait from the CMU MoBo dataset yielded by the four gait representation methods: GEI, GFI, GFF, MCI (proposed method), distinguishing among the six camera viewpoints: 1 (v03), 2 (v05), 3 (v07), 4 (v13), 5 (v16) and 6 (v17), see Fig. 3

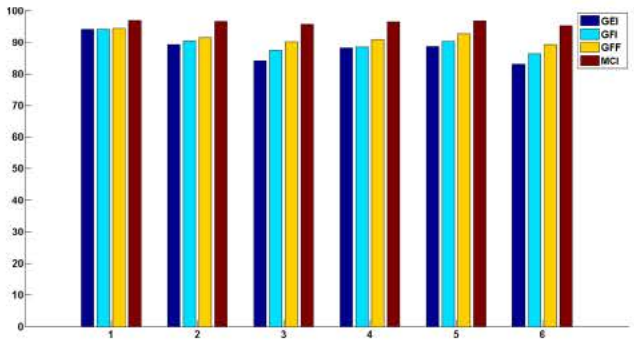

Fig. 6 Average classification rates for an incline gait from the CMU MoBo dataset yielded by the four gait representation methods: GEI, GFI, GFF, MCI (proposed method), distinguishing among the six camera viewpoints: 1 (v03), 2 (v05), 3 (v07), 4 (v13), 5 (v16) and 6 (v17), see Fig. 3

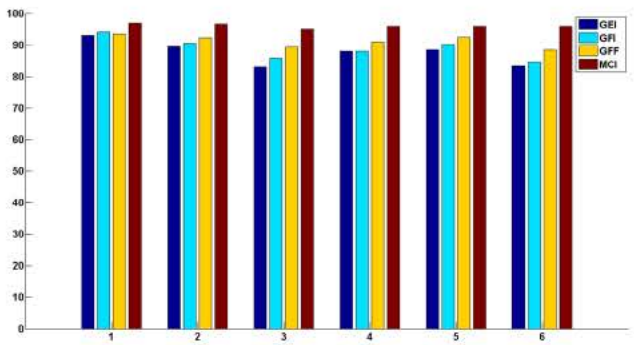

Fig. 7 Average classification rates for carrying ball gait from the CMU MoBo dataset yielded by the four gait representation methods: GEI, GFI, GFF, MCI (proposed method), distinguishing among the six camera viewpoints: 1 (v03), 2 (v05), 3 (v07), 4 (v13), 5 (v16) and 6 (v17), see Fig. 3

best performance and the lowest error rates among the alternative gait representation methods.

Another experiment aims at evaluating the effect of the time span parameter m introduced in Sect. 2, which represents the number of frames between templates whose

Table 2 Confusion matrix for the four tested gaits with the proposed method (MCI+PCA+SVM)

	Slow	Fast	incline	Ball
Slow	96.0	2.0	1.5	0.5
Fast	2.5	97.0	0.5	0.0
Incline	2.9	2.0	95.0	0.1
Ball	1.8	2.0	0.2	96.0

Table 3 Confusion matrix for the four tested gaits with GFF+PCA+SVM

	Slow	Fast	Incline	Ball
Slow	92.4	2.8	3.3	1.5
Fast	3.7	91.1	2.2	3.0
Incline	3.5	1.8	92.0	2.7
Ball	2.5	3.5	3.6	90.4

Table 4 Confusion matrix for the four tested gaits with GFI+PCA+SVM

	Slow	Fast	Incline	Ball
Slow	90.7	4.5	2.5	2.3
Fast	5.0	90.2	2.1	2.7
Incline	4.5	3.4	88.9	3.2
Ball	4.2	2.8	4.0	89.0

Table 5 Confusion matrix for the four tested gaits with GEI+PCA+SVM

	Slow	Fast	Incline	Ball
Slow	89.2	4.1	2.2	4.5
Fast	4.8	89.0	3.2	3.0
Incline	5.5	5.8	87.5	1.2
Ball	6.0	3.2	2.7	88.1

flow fields are considered for the computation of the co-occurrence of optical flow states. Figure 8 shows the average classification rates obtained for the four tested gaits corresponding to one of the two most demanding viewpoints (v07), with different values of m ranging between 1 and 15. Notice that the performance deteriorates as m increases. Therefore, m has been set to one in this work, since it is the value that yields optimal performance for the fast gait and near-optimal performance for the other tested gaits.

Finally, Table 6 shows the CPU times corresponding to the test stage (i.e., computation of the gait representation and classification) for all the tested gait representations by using the same dimensionality reduction stage based on PCA and the classification stage based on SVMs. All methods have been implemented in MATLAB and executed on an Intel Dual Core at 3.2 GHz. The execution time of the proposed technique is in the same range as the other methods based on optical flow fields.

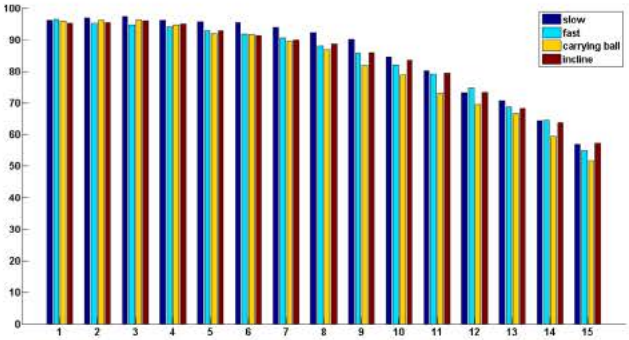

Fig. 8 Average classification rates with the proposed method (MCI+PCA+SVM) for the 4 tested gaits with viewpoint v07 and values of the time span parameter m between 1 and 15

Table 6 CPU times in seconds of the gait representation and classification stages for the four tested gait representations

Method	Representation	Classification	Total
GEI	7.063	0.425	7.488
GFI	34.452	0.846	35.284
GFF	26.376	1.154	27.530
MCI	32.244	2.251	34.495

4.2 Experiments on the AVAMVG dataset

In order to further assess the robustness of the proposed method, it has also been evaluated using the AVA Multi-View Database for Gait Recognition (AVAMVG) [31]. This database contains 1200 video sequences of 20 subjects (16 males and 4 females). There are 60 sequences associated with every subject, which correspond to 10 different paths captured from 6 viewpoints, as shown in Fig. 9. Three paths correspond to straight segments ($tr01, \ldots, tr03$), six paths to curved segments ($tr04, \ldots, tr09$), as shown in Fig. 9a, and the last path is an $tr08$-shaped segment ($tr10$), as shown in Fig. 9b.

Input video sequences must first be processed to segment the walking subjects. Firstly, the RGB images are converted into HSI color space. In this work, background subtraction based on mixtures of Gaussian is applied to the Hue and Saturation channels, which are more robust to illumination changes. The subjects' silhouettes are then processed through morphological operations in order to improve their quality.

The performance of MCI for the six viewpoints of the AVAMVG dataset is compared with the three standard gait representation methods: GEI, GFI and GFF. Figure 10 shows that the proposed approach is robust to variations of the camera viewpoints, and can correctly detect more than 95% of

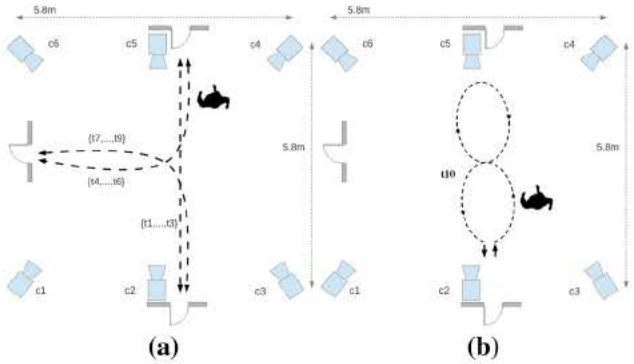

Fig. 9 AVAMVG setup and different paths: **a** $tr01$ to $tr09$ and **b** $tr10$

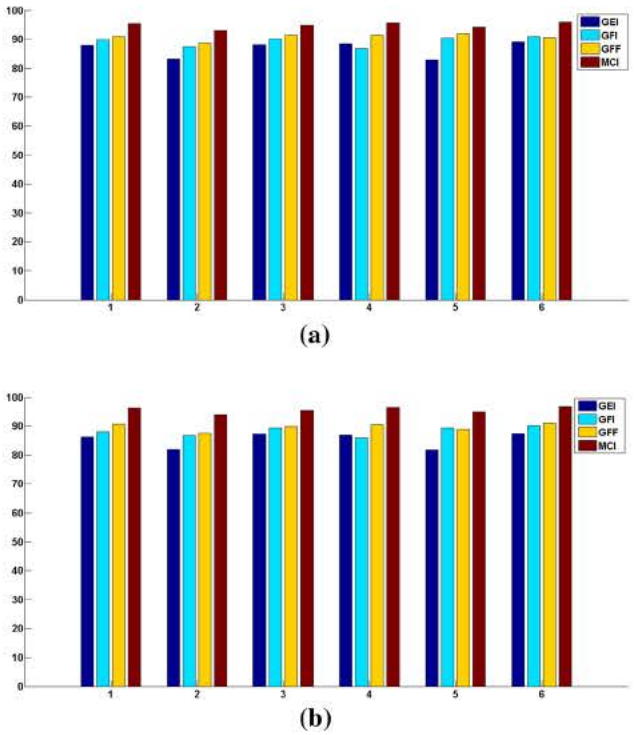

Fig. 10 Average classification rates for **a** path $tr01$ and **b** path $tr04$ from the AVAMVG dataset yielded by the four gait representation methods: GEI, GFI, GFF, MCI (proposed method), distinguishing among the six camera viewpoints: c1 , c2 , c3 , c4 , c5 and c6

the subjects' gait. Actually, the alternative methods yield less accuracy, especially when the camera is aligned with the walking direction, which is consistent with the results previously reported on the CMU MoBo dataset.

In addition, the four gait representation methods, GEI, GFI, GFF, MCI (proposed method), have been compared in order to recognize the path segments of the subjects in the AVAMVG dataset. In particular, the six viewpoints of the subjects have been used to train four SVMs, such that each SVM is associated with one path segment ($tr01$, $tr04$, $tr07$ and $tr10$). As shown in Fig. 11, the results show the robustness of the proposed gait recognition method with respect to

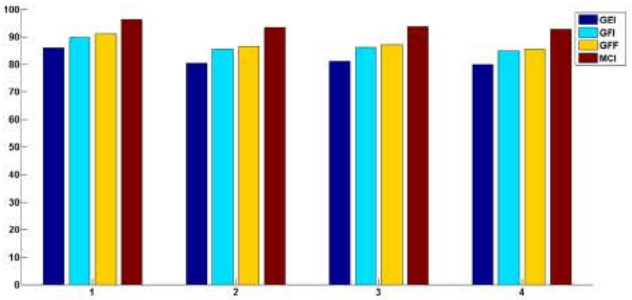

Fig. 11 Average classification rates for four paths ($tr01$, $tr04$, $tr07$ and $tr10$) from the AVAMVG dataset yielded by the four gait representation methods: GEI, GFI, GFF, MCI (proposed method)

Table 7 Classification rates with the proposed method and the five variations of the approach proposed in [21] (PFM)

Method	tr04	tr07	tr10
MCI+PC	89.6	90.7	91.4
PFM [21]	75.0	91.3	81.0
PFM+PCAL100 [21]	72.5	91.5	77.2
PFM+PCAL100+PCAH256 [21]	73.8	90.0	81.1
PFM+PCAL50+PCAH256+pyr [21]	75.0	90.0	87.3
PFM+PCAL100+PCAH256+pyr [21]	71.3	92.5	82.3

the three other methods with a variation of walking paths. MCI successfully detects the correct segment for the four tested paths in the AVAMVG dataset (more than 94%), whereas the other methods do not provide satisfactory recognition rates, especially for the curved paths ($tr04$, $tr07$ and $tr10$).

In this experiment, the proposed approach has been compared with the method presented in [21] (PFM) with the same multiview dataset (AVAMVG). In [21], the PFM approach was evaluated by training different variations of a PFM model with the three straight paths ($tr01$, $tr02$, $tr03$) and then by testing on the other curved paths ($tr04$, $tr07$, $tr10$). As mentioned in [21], they used five different variations of PFM: (a) PFM denotes a single-level PFM obtained by concatenating the descriptors extracted from both the top and bottom half of the subject's body with Fisher vectors obtained using a Gaussian Mixture Model (GMM) of size 150. (b) PFM+PCAL100 corresponds to applying PCA to reduce the dimensionality of the low-level motion descriptors to 100 before building the PFM. (c) PFM+PCAL100+PCAH256 corresponds to PFM with a reduction in the dimensionality of the low-level descriptors to 100 and of the final PFM descriptor to 256. (d) PFM+PCAL50+PCAH256+pyr corresponds to a two-level pyramidal configuration where the first level has no spatial partitions and the second level is obtained by dividing the bounding box into two parts along the vertical axis. In addition, PCA is applied in order to reduce the dimensionality of the low-level descriptors to 50 and of the final

Table 8 Recall (R), precision (P) and specificity (S) values with HOF, MBH, 3D-CoHOF,3D-CoMBH and MCI descriptors with slow, fast, incline and carrying ball gaits

Methods	MCI			MBH			HOF			3D-CoHOF			3D-CoMBH		
	R	P	S	R	P	S	R	P	S	R	P	S	R	P	S
Slow	96.6	96.4	96.4	97.1	96.9	96.8	95.2	95.1	95.3	93.9	93.7	93.3	94.4	94.1	94.2
Fast	96.9	96.8	96.8	97.4	97.2	97.2	95.7	95.5	95.6	92.4	92.1	92.5	94.5	94.2	94.3
Incline	94.3	94.9	94.8	95.1	94.6	94.6	93.1	93.2	92.9	91.5	91.2	90.5	93.6	93.3	93.1
Ball	96.5	96.0	96.0	96.2	95.9	95.9	95.0	94.9	94.7	92.0	91.8	91.4	94.1	94.0	93.8

PFM vector to 256. (e) PFM + PCAL100 + PCAH256 + pyr is similar to PFM + PCAL50 + PCAH256 + pyr, but reducing the dimensionality of the low-level descriptors to 100 instead of 50.

Table 7 shows that the proposed method (MCI + PCA) achieves correct classification rates above 89% with $tr04$ and 91% with $tr10$, yielding significantly better results than the five variations of [21]. In turn, the proposed scheme yields a recognition rate of more than 90% for path $tr07$. Although PFM + PCAL100 + PCAH256 + pyr yields the best correct classification rates for $tr07$, this algorithm has a much larger computational complexity than the proposed technique due to both the use of GMM models of 150 Gaussians for computing the Fisher vector features and the application of PCA to both the low-level descriptors and the final PFM descriptor.

4.3 Comparison with action recognition methods

The proposed descriptor has also been compared with four descriptors, namely HOF[24], MBH [25], 3D-CoHOF [26] and 3D-CoMBH [26], used for human action recognition. As proposed in [25], the pixel window size is 32×32. In turn, the path length has been set equal to the gait cycle (i.e., 33 frames for the CMU MoBo dataset). The $32 \times 32 \times 33$ pixel volume has been split into $2 \times 2 \times 3$ cells.

For HOF, orientations are quantized into 8 bins, with full orientation and magnitudes being used for weighting. An additional zero bin has been added (i.e., 9 bins), as proposed in [24]. The final descriptor size is 108 for HOF (i.e., $2 \times 2 \times 3 \times 9$). For MBH, both MBHx and MBHy feature vectors have 96 dimensions. 3D-CoHOF, 3D-CoMBHx and 3D-CoMBHy have been defined as described in [26], with each corresponding 3D co-occurrence descriptor along with three offsets: $offset_x$, $offset_y$ and $offset_t$, and where each offset is a vector of length $4 \times 4 \times 2 \times 2 \times 3$. SVM is then used for gait recognition.

In this experiment, the recall, precision and specificity values with HOF, MBH, MCI, 3D-CoHOF and 3D-CoMBH have been computed for the four gaits (i.e., slow, fast, incline and carrying ball) of the CMU MoBo dataset. As shown in Table 8, the MBH descriptor yields the best results among the three descriptors for both slow and fast gaits. In turn, the proposed descriptor MCI yields the best results for the two other gaits (i.e., incline and carrying ball). Meanwhile, the standard HOF and MBH descriptors yield better accuracy than the extended descriptors 3D-CoHOF and 3D-CoMBH, although their accuracy is less than the one of the proposed MCI descriptors.

Table 9 CPU times in seconds of the gait representation and classification stages for HOF, MBH, 3D-CoHOF,3D-CoMBH and MCI

Method	descriptor	Classification
HOF	22.2	2.85
MBH	38.7	2.92
3D-CoHOF	32.8	3.51
3D-CoMBH	45.3	4.60
MCI	16.9	2.25

Table 9 shows the CPU times corresponding to the test stage (i.e., computation of the gait representation and classification) for the five descriptors: HOF, 3D-CoHOF, MBH, 3D-CoMBH and MCI. They are implemented in C++. The same classification stage based on SVMs has been applied implemented in MATLAB. The proposed descriptor is significantly faster than the four other descriptors, which must compute the trajectories of the subjects over the gait cycle. The execution time of the classification stage based on the five descriptors is in the same range. Therefore, although MBH yields significantly better classification rates than MCI in some cases, the latter is twice faster.

5 Conclusion

This paper proposes a new gait representation that encodes the dynamics of a gait period through a 2D array of 17-bin histograms. Every histogram models the co-occurrence of optical flow states at every pixel of the normalized template that bounds the silhouette of a target subject. Five flow states (up, down, left, right, null) are considered. The first histogram bin counts the number of frames over the gait period in which the optical flow for the corresponding pixel

is null. In turn, each of the remaining 16 bins represents a pair of flow states and counts the number of frames in which the optical flow vector has changed from one state to the other during the gait period. The dimensionality of the proposed gait representation is reduced through principal component analysis. Finally, gait recognition is performed through supervised classification by means of support vector machines. Experimental results using the CMU MoBo and AVAMVG datasets show that the temporal co-occurrence of flow fields is able to better capture the dynamics of the gait patterns independently of the speed of the subject, the path shape and the camera viewpoint, especially when the walking direction is aligned with the camera viewpoint and for fast walking speeds, which is where state-of-the-art gait recognition methods yield the lowest performance. Future work aims at applying the proposed gait recognition technique to a robust surveillance system for both indoor and outdoor scenarios.

Appendix

Each of the 17-bin histograms models a co-occurrence of optical flow states between a pair of templates separated by m frames, T_i and T_j, where $j = (i+m)\%n$, $0 < m < n$ and $0 \leq i,\ j < n$, with n being the number of frames in a gait period. Those bins are noted as: HV_i, HR_i, HL_i, LR_i, LL_i, LH_i, RR_i, RL_i, RH_i, VU_i, VD_i, UU_i, UD_i, UV_i, DU_i, DD_i, DV_i.

$$HV_i(x,y) = H_i(x,y) \times V_j(x,y), \tag{13}$$
$$HR_i(x,y) = H_i(x,y) \times R_j(x,y), \tag{14}$$
$$HL_i(x,y) = H_i(x,y) \times L_j(x,y), \tag{15}$$
$$LR_i(x,y) = L_i(x,y) \times R_j(x,y), \tag{16}$$
$$LL_i(x,y) = L_i(x,y) \times L_j(x,y), \tag{17}$$
$$LH_i(x,y) = L_i(x,y) \times H_j(x,y), \tag{18}$$
$$RR_i(x,y) = R_i(x,y) \times R_j(x,y), \tag{19}$$
$$RL_i(x,y) = R_i(x,y) \times L_j(x,y), \tag{20}$$
$$RH_i(x,y) = R_i(x,y) \times H_j(x,y), \tag{21}$$
$$VU_i(x,y) = V_i(x,y) \times U_j(x,y), \tag{22}$$
$$VD_i(x,y) = V_i(x,y) \times D_j(x,y), \tag{23}$$
$$UU_i(x,y) = U_i(x,y) \times U_j(x,y), \tag{24}$$
$$UD_i(x,y) = U_i(x,y) \times D_i(x,y), \tag{25}$$
$$UV_i(x,y) = U_i(x,y) \times V_j(x,y), \tag{26}$$
$$DU_i(x,y) = D_i(x,y) \times U_j(x,y), \tag{27}$$
$$DD_i(x,y) = D_i(x,y) \times D_j(x,y), \tag{28}$$
$$DV_i(x,y) = D_i(x,y) \times V_j(x,y). \tag{29}$$

Finally, an accumulation $h \times w$ histogram of 17 bins is computed. The 17 bins can be defined as:

$$HV(x,y) = \sum_{i=0}^{n-1} HV_i(x,y), \tag{30}$$
$$HR(x,y) = \sum_{i=0}^{n-1} HR_i(x,y), \tag{31}$$
$$HL(x,y) = \sum_{i=0}^{n-1} HL_i(x,y), \tag{32}$$
$$LR(x,y) = \sum_{i=0}^{n-1} LR_i(x,y), \tag{33}$$
$$LL(x,y) = \sum_{i=0}^{n-1} LL_i(x,y), \tag{34}$$
$$LH(x,y) = \sum_{i=0}^{n-1} LH_i(x,y), \tag{35}$$
$$RR(x,y) = \sum_{i=0}^{n-1} RR_i(x,y), \tag{36}$$
$$RL(x,y) = \sum_{i=0}^{n-1} RL_i(x,y), \tag{37}$$
$$RH(x,y) = \sum_{i=0}^{n-1} RH_i(x,y), \tag{38}$$
$$VU(x,y) = \sum_{i=0}^{n-1} VU_i(x,y), \tag{39}$$
$$VD(x,y) = \sum_{i=0}^{n-1} VD_i(x,y), \tag{40}$$
$$UU(x,y) = \sum_{i=0}^{n-1} UU_i(x,y), \tag{41}$$
$$UD(x,y) = \sum_{i=0}^{n-1} UD_i(x,y), \tag{42}$$
$$UV(x,y) = \sum_{i=0}^{n-1} UV_i(x,y), \tag{43}$$
$$DU(x,y) = \sum_{i=0}^{n-1} DU_i(x,y), \tag{44}$$
$$DD(x,y) = \sum_{i=0}^{n-1} DD_i(x,y), \tag{45}$$
$$DV(x,y) = \sum_{i=0}^{n-1} DV_i(x,y). \tag{46}$$

References

1. Bazin, A.I., Nixon, M.S.: Gait verification using probabilistic methods, In: Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05) - Volume 1 - Volume 01, series WACV-MOTION '05, pp. 60–65. IEEE Computer Society, Washington, DC, 2005. https://doi.org/10.1109/ACVMOT.2005.55
2. He, W., Li, P.: Gait recognition using the temporal information of leg angles, In: 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 2010, vol. 5, pp. 78–83 (2010)
3. Choudhury, S.D., Tjahjadi, T.: Gait recognition based on shape and motion analysis of silhouette contours. Comput. Vis. Image Underst. **117**(12), 1770–1785 (2013)
4. Kovac, J., Peer, P.: Human skeleton model based dynamic features for walking speed invariant gait recognition. Math. Prob. Eng. **2014**, 1 (2014)
5. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 257–267 (2001)
6. Lee, C.P., Tan, A.W., Tan, S.C.: Time-sliced averaged motion history image for gait recognition. J. Vis. Commun. Image Represent. **25**(5), 822–826 (2014)
7. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Trans. Pattern Anal. Mach. Intell. **28**(2), 316–322 (2006)
8. Hosseini, N.K., Nordin, M.J.: Human gait recognition: A silhouette based approach. J. Autom. Control Eng. **1**(2), 103–105 (2013)
9. Tan, D., Huang, K., Yu, S., Tan, T.: Efficient night gait recognition based on template matching. In: Proceedings of the 18th International Conference on Pattern Recognition - Volume 03, series ICPR '06, pp. 1000–1003. IEEE Computer Society, Washington, DC (2006). https://doi.org/10.1109/ICPR.2006.478
10. Tao, D., Li, X., Wu, X., Maybank, S.J.: General tensor discriminant analysis and gabor features for gait recognition. IEEE Trans. Pattern Anal. Mach. Intell. **29**(10), 1700–1715 (2007)
11. Hayder Ali, C.A.E.G.M., Dargham, J.: C.A.E.G.M., Dargham, Jamal: Gait recognition using gait energy image. Int. J. Signal Process. Image Proc. Pattern Recognit. **4**, 3.141–3.152 (2011)
12. Tee, C., Goh, M., Teoh, A.: Gait recognition using sparse grassmannian locality preserving discriminant analysis, In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 2989–2993 (2013)
13. Kusakunniran, W., Wu, Q., Li, H., Zhang, J.: Automatic gait recognition using weighted binary pattern on video. In: Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, 2009. AVSS '09, pp. 49–54 (2009)
14. Lishani, A.O., Boubchir, L., Khalifa, E., Bouridane, A.: Human gait recognition using GEI-based local multi-scale feature descriptors. Multimed. Tools Appl. **77**, 1–16 (2018)
15. Wang, C., Zhang, J., Wang, L., Pu, J., Yuan, X.: Human identification using temporal information preserving gait template. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2164–2176 (2012)
16. Tang, J., Luo, J., Tjahjadi, T., Guo, F.: Robust arbitrary-view gait recognition based on 3d partial similarity matching. IEEE Trans. Image Process. **26**(1), 7–22 (2017)
17. Jia, N., Li, C.-T., Sanchez, V., Liew, A.W.-C.: Fast and robust framework for view-invariant gait recognition. In: 5th International Workshop on Biometrics and Forensics (IWBF), 2017, pp. 1–6. IEEE (2017)
18. BenAbdelkader, C., Cutler, R., Nanda, H., Davis, L.S.: Eigengait: Motion-based recognition of people using image self-similarity. In: Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, series AVBPA '01, pp. 284–294. Springer, London (2001). http://dl.acm.org/citation.cfm?id=646073.677457
19. Bashir, K., Xiang, T., Gong, S.: Gait representation using flow fields. In: Proceedings of the British Machine Vision Conference, pp. 113.1–113.11. BMVA Press (2009)
20. Lam, T.H.W., Cheung, K.H., Liu, J.N.K.: Gait flow image: a silhouette-based gait representation for human identification. Pattern Recognit. **44**(4), 973–987 (2011)
21. Castro, F.M., Marín-Jimenez, M.J., Medina-Carnicer, R.: Pyramidal fisher motion for multiview gait recognition. In: Proceedings of the 2014 22Nd International Conference on Pattern Recognition, series ICPR '14, pp. 1692–1697. IEEE Computer Society, Washington, DC (2014). https://doi.org/10.1109/ICPR.2014.298
22. Mahfouf, Z., Bouchrika, I., Merouani, H.F., Harrati, N.: Gait biometrics via optical flow motion features for people identification. In: 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), 2016, pp. 312–321. IEEE (2016)
23. Mahfouf, Z., Merouani, H.F., Bouchrika, I., Harrati, N.: Investigating the use of motion-based features from optical flow for gait recognition. Neurocomputing **283**, 140–149 (2018)
24. Laptev, I., Marszaek, M., Schmid, C., Rozenfeld, B.: Learning realistic human actions from movies. In: CVPR (2008)
25. Wang, H., Kläser, A., Schmid, C., Liu, C.-L.: Dense trajectories and motion boundary descriptors for action recognition. Int. J. Comput. Vis. **103**(1), 60–79 (2013)
26. Peng, X., Qiao, Y., Peng, Q.: Motion boundary based sampling and 3d co-occurrence descriptors for action recognition. Image Vis. Comput. **32**(9), 616–628 (2014)
27. 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)
28. Rashwan, H.A., García, M.A., Puig, D.: Variational optical flow estimation based on stick tensor voting. IEEE Trans. Image Process. **22**(7), 2589–2599 (2013)
29. Lee, H., Hong, S., Kim, E.: An efficient gait recognition with backpack removal. EURASIP J. Adv. Signal Process **2009**, 4.61–4.67 (2009). https://doi.org/10.1155/2009/384384
30. Gross, R., Shi, J.: The CMU motion of body (mobo) database. Robotics Institute, Pittsburgh, PA, Technical Report CMU-RI-TR-01-18 (2001)
31. Lopez-Fernandez, A.C.P.M.M.-J.D., Madrid-Cuevas, F.J., Muoz-Salinas, R.: The AVA multi-view dataset for gait recognition (AVAMVG) In: International Workshop on Activity Monitoring by Multiple Distributed Sensing (AMMDS) (2014)

Hatem A. Rashwan received the B.S. and M.S. degrees in electrical engineering from South Valley University (Egypt) in 2002, 2007. He received the PhD degree in Computer Vision from Rovira i Virgili University in 2014. Between 2004 and 2009, he joined the Electrical Engineering Department, South Valley University as an Assistant Lecturer. From Jan. 2010 until Oct. 2014, he joined IRCV Group, Department of Computer Science and Mathematics at Rovira i Virgili University (Spain) as a PhD student and research assistant. From Nov. 2014 until 2017, he is a PostDoc in VORTEX

group, IRITCNRS, INP-Toulouse, University of Toulouse (France). From 2018 until now, he is a Beatriu de Pinós researcher in URV. His research interests include image processing, computer vision, machine learning and pattern recognition.

Miguel Ángel García received the B.S., M.S., and Ph.D. degrees in computer science from the Polytechnic University of Catalonia (Barcelona, Spain) in 1989, 1991, and 1996, respectively. He joined the Department of Software at the Polytechnic University of Catalonia in 1996 as an Assistant Professor. From 1997 to 2006, he was with the Department of Computer Science and Mathematics at Rovira i Virgili University (Tarragona, Spain), where he was the Head of Intelligent Robotics and Computer Vision group. In 2006, he joined the Department of Informatics Engineering at Autonomous University of Madrid (Spain), where he is currently Associate Professor. His research interests include mobile robotics, image processing, and 3-D modeling.

Sylvie Chambon received the Ph.D. degree in computer science from the University of Toulouse, Toulouse, France, working on "Colour stereoscopic matching with occlusions." From 2006 to 2007, she was a Postdoctoral Researcher with Télécom Paris, working on in multimodal registration of medical images. From 2008 to 2011, she was a Permanent Researcher with IFSTTAR (the French institute of science and technology for transport, development and networks), working on segmentation of thin structures and, in particular, road cracks. Since September 2011, she has been an Assistant Professor in the Institut de Recherche en Informatique de Toulouse (IRIT), INP-ENSEEIHT, Universit of Toulouse. Her research interest includes matching, feature detection and tracking, and segmentation of thin structures and urban scenes.

Domenec Puig received the M.S. and Ph.D. degrees in computer science from the Polytechnic University of Catalonia (Barcelona, Spain) in 1992 and 2004, respectively. In 1992, he joined the Department of Computer Science and Mathematics at Rovira i Virgili University (Tarragona, Spain), where he is currently Professor. Since July 2006, he is the Head of the Intelligent Robotics and Computer Vision group at the same university. His research interests include image processing, texture analysis, perceptual models for image analysis, scene analysis, and mobile robotics.