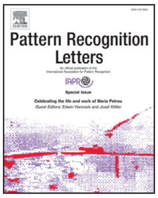
Pattern Recognition Letters 68 (2015) 327–335

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec



A framework for gait-based recognition using Kinect[☆]

Dimitris Kastaniotis[a,*], Ilias Theodorakopoulos[a], Christos Theoharatos[b], George Economou[a], Spiros Fotopoulos[a]

[a] Electronics Laboratory, Department of Physics, University of Patras, Patras 26500, Greece
[b] Computer Vision Systems, Irida Labs S.A., Platani, Patras 26504, Greece

ARTICLE INFO

Article history:
Available online 9 July 2015

Keywords
Pose-based gait recognition
Pose-based gender recognition
Human motion analysis
Sparse representation
Microsoft Kinect

ABSTRACT

Gait analysis has gained new impetus over the past few years. This is mostly due to the launch of low cost depth cameras accompanied with real time pose estimation algorithms. In this work we focus on the problem of human gait recognition. In particular, we propose a modification of a framework originally designed for the task of action recognition and apply it to gait recognition. The new scheme allows us to achieve complex representations of gait sequences and thus express efficiently the dynamic characteristics of human walking sequences. The representational power of the suggested model is evaluated on a publicly available dataset where we achieved up to 93.29% identification rate, 3.1% EER on the verification task and 99.11% gender recognition rate.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Human activity encompasses a plethora of biological cues. The human brain is able to perceive effortlessly complex biological motions within a small time interval [1]. Furthermore, many years of psychophysical research have shown that the human visual system is finely tuned to the social cues immanent in human movements [2]. For example, Johansson in [3] using a method called moving light display has demonstrated that the animation of a walking person elicits a compelling percept of a human figure in motion. In this method an otherwise invisible walking person was visualized using only 12 major joints, without any further information about the joints configuration.

It has been shown that gait comprises several unique characteristics that can be used in applications where discrimination among different persons is required. In particular, as early medical studies suggest, there are 24 different components to human gait [4] that establish its' uniqueness. Indeed, a number of studies have shown that identification of familiar people is feasible [5–7] while recognizing actions [8,9] gender, age [5,10] and intention is possible even for unfamiliar people.

Gait recognition is considered nowadays as one of the most active areas of biometrics [11]. Compared to other biometric traits, gait offers significant advantages. Primarily, gait is characterized by its' unobtrusiveness. That is, physical touch with the subject is not required, as for example in the case of fingerprint acquisition. Also, subjects' attention and cooperation is not needed in order to capture data. In particular, the utilization of gait analysis provides great flexibility to a biometric system, due to its' ability to operate from a distance. Additionally, while face recognition is one of the most commonly used methods for person identification, it suffers significant limitations. For example face can be easily changed due to intentional disguise etc. On the other hand, gait cannot be easily imitated, or intentionally faked. Furthermore, gait analysis, as opposed to face recognition, retains confidentiality [12,13] due to the fact that the type of incorporated data cannot be easily accessed by humans in order to retrieve the identity of the persons involved.

Over the past years, gait analysis has focused on videos from natural scenes mostly with respect to the view point. In this context, methods incorporated features stemming directly from raw video data. Very soon it was understood that in order to deal with tasks like human recognition it is sufficient to utilize information of the joints. These findings brought again in the foreground methods based on the analysis of a sequence of human poses estimated from video raw data. However, accurate pose estimation casts a major problem due to several intrinsic and extrinsic factors as human body non-rigidness and occlusions caused by human limbs or other objects. For these reasons and in order to track the joints of a human body specialized devices, using electromagnetic markers positioned in specific positions (joints) under a controlled environment were suggested. The high cost and complexity of these systems restricted the appliance of pose-based human motion analysis mostly into medical applications. Recently the launch of low cost depth sensors, like Microsoft's Kinect, made possible the design of practical applications based on human pose analysis like action recognition.

[☆] This paper has been recommended for acceptance by Paulo Lobato Correia

[*] Corresponding author. Tel.: +30 2610 99 6058; fax: +30 2610 99 7456.
E-mail address: dkastaniotis@upatras.gr, d.kastaniotis@gmail.com (D. Kastaniotis).

http://dx.doi.org/10.1016/j.patrec.2015.06.020
0167-8655/© 2015 Elsevier B.V. All rights reserved.

In this context and inspired by our prior work on human motion analysis [14,15], in this paper we propose a framework for pose-based gait recognition and identification as well as gender recognition. Our framework utilizes a multiple level pose data processing scheme, where pose data acquired using Microsofts' Kinect sensor are used in order to recognize a walking person. In the first level, the captured pose sequences are expressed as angular vectors (Euler angles) of eight selected limbs. Then, at the second stage, these trajectories (sequences of angular vectors) are mapped in the dissimilarity space resulting into a vector of dissimilarities. At the third level, dissimilarity vectors of pose-sequences are modeled via sparse representation. The derived sparse representation was used at the fourth stage in order to perform person as well as gender recognition, and found advantageous in terms of identification and verification accuracy compared to classical gait analysis features.

The rest of this paper is organized as follows: Section 2 provides an overview of Gait Motion Analysis. In Section 3 an outline of recently introduced methodologies is presented. The proposed approach is described in Section 4. Experimental results are reported and thoroughly discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Background

2.1. Human gait motion analysis

Over the past years analysis of human motion has become one of the most emerging topics in the computer vision community [16]. Gait is considered as a special category of human action. In the context of human motion analysis, gait recognition has been approached from two major directions- model free and model-based [17].

Model free approaches use a two-dimensional representation of a set of aligned images extracted from human body. They utilize the silhouette [18], the contour, or a one-dimensional signal representation of contour [19]. They are also able to work with low resolution data as they are considered insensitive to the quality of silhouettes extracted. Thus, are reckoned to be simple and computationally efficient methods compared to the model based ones. A negative characteristic of these appearance based techniques is their sensitivity to viewpoint and scale. With the emergence of RGB-D sensors, these methods have evolved by incorporating depth information [20–22].

On the other hand, model based approaches rely on a pre-defined model of the underlying kinematics in order to represent the data and capture gait dynamics. These methods are usually more complex and computational expensive than model free approaches, but can lead to more efficient representations. In this research direction, authors in [23] have shown experimentally, that pose estimation can lead to better representation of human motion compared to the appearance representation where low-level features are extracted directly from video data. Furthermore, they manifest that pose based representation is much more efficient even in the presence of noise. According to these findings, highly precise pose estimation is not essential for the task of human motion analysis. In general, model based approaches are scale and view invariant which makes them appropriate for practical applications. By definition, model based approaches use a framework that relies on a stick-figure model. They use static and/or dynamic features that include, stride and cadence, distances of human body parts, length of body parts, height and joint angles between sets of rigid parts or from motion capture data. For an extensive review of vision based gait recognition methods, we refer the reader to [19].

2.2. Biological inspired human motion analysis

Since human brain achieves state of the art performance in biological motion perception and inference of attributes contained in gait motion, biological inspired frameworks have been developed, trying to imitate brains' functionality. In [24] authors proposed a system that imitates the functionality of Visual cortex ventral and dorsal pathways. These two paths are specialized for the analysis of form and Optic-Flow information respectively. Beauchamp et al. [25] showed that point-light displays of human actions activate the ventral temporal cortex, although this activation is less strong than for whole body displays. Particularly, the form pathway analyses biological movements by recognizing 'snapshots' sequences of body shapes. On the other hand the motion pathway recognizes biological movements by analyzing optic-flow patterns. Their common characteristic is that they comprise hierarchies of neural feature detectors of increasing complexity.

In this context, recently Castrodad and Sapiro proposed in [26] a framework where human actions are modeled using a two-level sparse representation scheme. In the first level, action-specific dictionaries are constructed using sparse modeling techniques, whereas a test action is sparsely represented using the superset of these individual dictionaries. The test action is considered as a mixture of the training actions, and the proportion of the mixture is computed based on the origins of the selected dictionary atoms during sparse representation. At the second level, the mixtures are modeled again using sparse representation techniques, and the recognition of the test action is based on the sparse representation of the corresponding mixture of the test action. Using this technique, increasingly complex features are learned at higher levels of processing. Indeed architectures with increasingly complex features representations have shown excellent performance compared to shallow ones. For example authors in [15], in a similar manner with [26], incorporated the SRC (Sparse Representation based Classification) classifier in order to perform skeleton based human action recognition. The presented results have shown excellent performance compared with state of the art techniques. Targeting the problem of model-free gait recognition in [27], in an analogous approach with [26], Jiwen et al. proposed a method for view invariant gait based identity and gender recognition. In particular in their recognition pipeline they incorporated human silhouettes which they first assigned into several clusters. Then for every cluster they computed a cluster-based average gait image (C-AGI) and represented it using sparse representation. In order to cope with the problem of view invariance they incorporated a distance metric learning technique which minimizes the intra- class and maximizes the interclass distance. Classification is then performed using the SRC [28].

2.3. Applications

Gait analysis is utilized in many application areas spanning from biometrics to human computer interaction and clinical diagnosis [29]. Furthermore, it can be used to infer human attributes like age [30], weight and mental state [31]. Such attributes – also called soft biometrics – can be used in gathering population statistics that are valuable to marketing and personalized advertisement applications. Medical applications are mostly focused in neurological diseases [32,33], surgery prognosis and Parkinson diagnosis [34]. In another application, in [35], gait was employed to assess the fall risk of elderly people using stride-to-stride gait variability through information extracted from a Microsoft Kinect sensor.

3. Related work

With the launch of RGB- D, many attempts have been made to enhance the traditional model free methods with the information provided from a depth sensor. In [20], an extended version of Gait Energy Image (GEI) to 3D is introduced. In particular, they proposed binary voxel volumes, which is analogous to two dimensional silhouettes, and are spatially aligned and averaged over a gait cycle. They focused

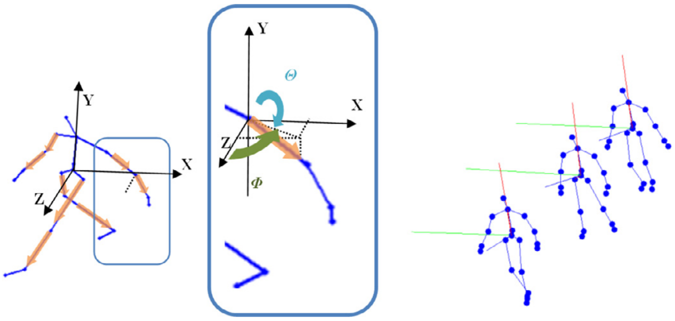


Fig. 1. Left: Skeleton model, and the selected skeletal primitives for pose representation. In the zoomed region the computation of the Euler angles $\{\varphi,\theta\}$ for one skeletal primitive according to the human attached coordinate system is presented. Right: The axes of the coordinate system are color coded with red $(y'y)$, green$(x'x)$ and blue $(z'z)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

on the lower body, using only features extracted from legs as they claim that appearance changes due to carrying goods or due to hand movements. In [21] GEI is modified in order to make use of information provided by the Kinect Sensor, improving gait recognition. Their system utilizes the depth information so as to achieve a better silhouette extraction that leads to an improvement of the original GEI, the Depth Gradient Histogram Energy Image (DGHEI). In [22], subjects walking in different directions were used to explore whether depth information can be beneficial for gait classification purposes. They extract 2D and 3D features based on shape descriptors using only depth information.

To the best of authors' knowledge, only three papers have been presented, utilizing pose data acquired using Kinect, for gait analysis and gait biometrics. In order to take advantage of the efficient, fast and reasonably accurate pose estimation that Kinect provides, in [36] a number of body features along with step length and speed were combined. In [37] authors incorporated the mean, standard deviation and maximum value of three angles for each one of the left and right legs, a total of 18 features that were whitened before running K-means. More recently [14] targeting the task of real time gender recognition, incorporated the skeleton data captured from Kinect sensor in order to identify the gender of a walking person by aggregating features into histograms and then classifying samples using an SVM classifier. The framework proposed in this work enhances the initial findings of [14] that "gait encompasses information about gender" and extends these to the field of identity recognition.

4. Proposed framework

4.1. Overview

Assume that a human body pose is described via a set of 3D coordinates, corresponding to the position of the joints of a skeletal model, captured in each frame from an RGB-D sensor. Then, the goal of a pose-based gait recognition application is to process a sequence of poses corresponding to a walking person, in order to identify this person. Microsoft Kinect, is a sensor which is able to provide this type of data in real-time and at a frame rate of 30 fps. The employed SDK implements a state-of-the-art algorithm [38], producing 3D coordinates for a 20-joint skeletal model, illustrated in Fig. 1. In order to deal with the inaccuracies of pose estimation and at the same time deliver significant discriminative power to the system, instead of static characteristics, we utilize here a scheme that exploits dynamic characteristics, related to the unique walking style of each person.

Based on these facts and further motivated by the biological evidence, presented in the previous sections and recently published work [26,15], that indicates the effectiveness of multiple-level representation architectures in discriminating between different human motions, in this work a framework for pose-based gait recognition is proposed. The pipeline of this framework incorporates a hierarchy of representations in order to perform both gait as well as gender recognition.

In order to recognize a walking person from a given gait sequence, the 3D coordinates of the joints (at every frame) are initially processed and a set of features is extracted resulting into an appropriate representation. Then a dissimilarity value between a walking sequence and each sequence belonging to set of labeled training sequences is calculated, using classical dissimilarity measures. Thus, the new walking sequence is now represented as a vector of dissimilarity values to a set of labeled sequences. Next, a sparse representation of the dissimilarity-space vector is computed, utilizing the corresponding dissimilarity representations of the labeled training data as a codebook. At the final stage, recognition is performed via application of the minimum reconstruction error criterion to the final sparse representation of the walking sequence. An overview of the proposed framework is illustrated in Fig. 2. In the following sections we provide a detailed description of the implemented pipeline.

4.2. Feature extraction and representation

In order to efficiently represent the dynamic characteristics of gait we select a number of limbs that are found to be mostly informative

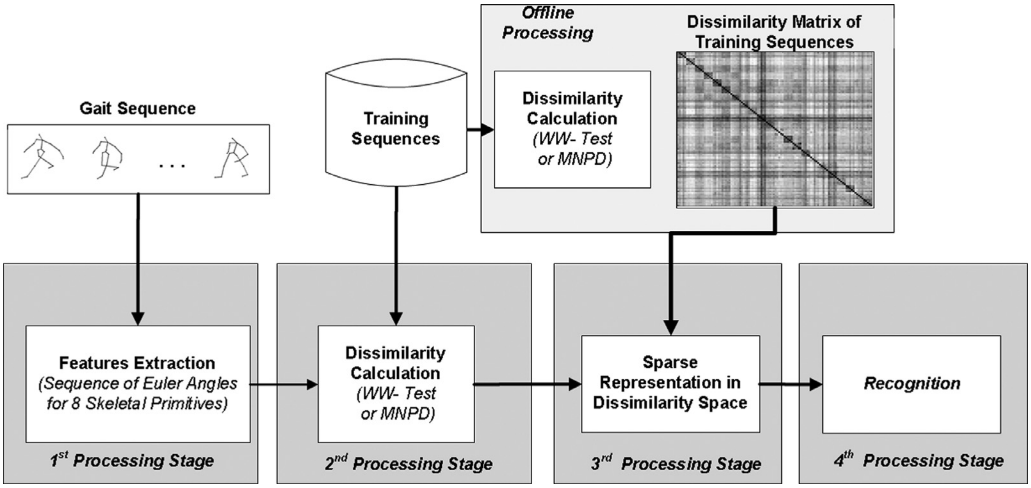

Offline Processing
Dissimilarity Matrix of Training Sequences
Gait Sequence
Training Sequences
Dissimilarity Calculation (WW- Test or MNPD)
Features Extraction (Sequence of Euler Angles for 8 Skeletal Primitives)
1st Processing Stage
Dissimilarity Calculation (WW- Test or MNPD)
2nd Processing Stage
Sparse Representation in Dissimilarity Space
3rd Processing Stage
Recognition
4th Processing Stage

Fig. 2. Overview of the proposed framework.

[15,14]. These are the upper and lower arm and upper and lower leg in both sides. Then we encode the direction of every limb using two Euler angles as shown in Fig. 1. The computation of Euler angles with respect to the sensor coordinate system does not yield a view invariant representation. In order to cope with this limitation we compute a coordinate system that is attached to the human torso as shown in Fig. 1. The computation of the coordinate system attached to the torso is performed as follows. First, the line connecting the center of hips with the center of shoulders is considered as the first axis. Then the perpendicular line starting at the center of the hips to the left thigh is used as the second axis. The third axis is simply computed as the cross product of the first two (Fig. 1).

Techniques based on sparse representations are known for their exceptional discriminative power. In order to allow such a representation, at the second level of the proposed learning scheme, first a pose sequence has to be transformed from an arbitrary sized set of 16 dimensional vectors in the Euler Angles space, into a fixed-length vector representation. While several techniques have been proposed, here we selected the one based on a representation in the dissimilarity space [39]. According to this approach every gait sequence can be represented as a vector of dissimilarities with respect to some prototype samples. This kind of representation is computationally efficient and enhances the discriminability between different walking styles.

Furthermore, in the dissimilarity space one can make use of standard classifiers as shown in [40,41]. In particular authors in [40] demonstrate that in dissimilarity space the generalization performance achieved by quadratic and linear classifiers is improved compared to the k-NN classifier. It is worth to notice that this holds especially true for the case of small representative data which is usually the case in biometric and forensics applications where the training data may be difficult to acquire. Recently authors in [41] evaluated a linear SVM in a publicly available dissimilarity dataset. The reported classification performance encourages the use of classifiers following this representation scheme. Additionally, in [42] authors highlighted the advantages regarding kernel generalization of schemes based on sparse representation indicating the overall efficiency introduced by proximity representations.

In order to achieve such a representation in the gait sequences case, a proximity measure between multidimensional spaces is required. Traditionally this can be done under one of the following four main perspectives: (a) Estimation of a non-linear wrapping between trajectories, by incorporating dynamic programming [43], (b) working in the graph domain and using graph similarity measures [44], (c) using statistical tests between multivariate distributions [45] and (d) via computation of appropriate distances [46].

In this work we have evaluated two different approaches to the dissimilarity estimation. Firstly, the multidimensional Wald–Wolfowitz (WW) test [45], a statistical test checking the hypothesis of whether two sets of samples derived from the same multidimensional distribution, was used as a dissimilarity measure. The resultant W-index is a negative value indicating the strength of this hypothesis. When the W-index is non-negative, the hypothesis is considered valid. Thus, dissimilarity values can naturally derive from the WW test as the absolute values of W-index in case of negative test values, otherwise as zeros in case of valid hypothesis. Additionally, and due to its low computational overhead in comparison to WW test, the Mutual Nearest Point Distance (MNPD) [47] is also used here as an alternative measure. It is known that in order to compare trajectories, MNPD performs quite well in many applications and as we present in the experimental results this is also the case of gait sequences.

4.3. Dissimilarity representation

Given a training set of gait sequences with n samples $\mathbf{X} = \{g_1, g_2, \dots, g_n\}$ where g_i corresponds to the angular representation of a gait sequence as described in Section 4.2. Also assume that we have a subset of our training set (namely the prototypes) $P = \{p_1, p_2, \dots, p_k\}$ such that $P \subseteq X$. Using these prototypes, dissimilarity representation is a function that allows us to map our training samples (gait sequences) in the dissimilarity space by using the following formula $F(\bullet, P) : X \rightarrow \mathbb{R}^k$.

A common approach followed by the procedure of prototype selection which is also adopted in this work is to use every training sample as a prototype and thus $P := \mathbf{X}$ and $k = n$. Therefore, given a gait sequence gas well as a set of prototypes P, we can map sequence gas follows:

$$\mathbf{y} = F(g, P) = [d(g, p_1), d(g, p_2), \dots, d(g, p_n)]^T \tag{1}$$

Assuming that we have a set of test sequences g_t that we want to encode, then in the distance matrix between the train and test objects, every test object is represented by the corresponding column of the dissimilarity matrix $D = [\mathbf{y}_1^T, \dots, \mathbf{y}_M^T]$, where $\mathbf{y}_i$ is a column vector. In the following section we introduce the sparse representation of these gait sequences depicted as dissimilarity vectors.

4.4. Sparse representation of gait sequences

Given a matrix C whose columns are a set of codebook elements namely atoms and a column vector $\mathbf{y}$ corresponding to the dissimilarity vector of a gait sequence as presented in formula (1), then the goal of sparse coding is to find a coefficient vector $\mathbf{x}$ such that $\mathbf{y} = C\mathbf{x}$ and $\|\mathbf{x}\|_0$ is minimized. This can be expressed as the following optimization problem:

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = C\mathbf{x} \tag{2}$$

where $\|\cdot\|_0$ denotes the l^0 norm which corresponds to the sum of a vector's nonzero entries. The problem in (2) though is NP-hard combinatorial and in general is difficult even to approximate the solution [48]. In [49] authors showed that one can approximate the solution by replacing the l^0 norm with the l^1 in case where the solution of (2) is sparse enough. In this case, the result can be obtained in polynomial time using linear programming methods ([50]). Hence the problem defined in (2) is now reformulated as follows:

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = C\mathbf{x}, \tag{3}$$

where $\|\cdot\|_1$denotes the ℓ^1 norm of a vector. Another important issue here is the computation of the matrix $\mathbf{C}$- namely the dictionary matrix. According to the procedure of sparse modeling this matrix has to be estimated from the training data. Several techniques have been proposed towards the solution of this problem [51–53]. Commonly a recursive scheme is utilized where given a random initialized matrix $\mathbf{C}$, the algorithm iteratively updates the sparse coefficients and then given these coefficients, applies an update to the dictionary. The convergence is achieved when a cost function is minimized. At the time of the convergence the achieved solution results in a dictionary appropriate for the sparse representation of the feature vectors.

In order to reduce the complexity imposed by the computation of the dictionary we adopted the method proposed by Wright [28]. According to that scheme, the dictionary can be constructed directly using the training data. The hypothesis made in this approach is based on the assumption that a training set consisting of n objects, obtained from k classes is given. Furthermore, for simplicity assume that all classes of the training set have the same number of objects e.g. m objects. We denote as $C_i = [\mathbf{v}_1^i, \mathbf{v}_2^i, \dots, \mathbf{v}_m^i] \in \mathbb{R}^{n \times m}$ the matrix that holds the m feature vectors of the ith class from the training set, as column vectors. Then, a vector $y \epsilon \mathbb{R}^n$ obtained from a class i which is represented in the training set, can be reconstructed as a linear combination of the corresponding training vectors with negligible reconstruction error i.e.

$$\mathbf{y} = \sum_{j=1}^{m} \mathbf{v}_j^i x_j^i = C_i \mathbf{x}_i, \tag{4}$$

where $x_j^i \in \mathbb{R}$ are the sparse coefficients that form the coefficient vector $\mathbf{x}_i = [x_1^i, x_2^i, \ldots, x_m^i]^T$. In our framework, following the dissimilarity representation at the first level, the matrix that contains all vectors of the training set is simply the dissimilarity matrix computed on the training objects and can be expressed as $C = [C_1, C_2, \ldots, C_k] \in \mathbb{R}^{n\times n}$. Thus, the linear representation of the test vector $\mathbf{y}$ can be equivalently expressed as:

$$\mathbf{y} = C\mathbf{x}, \tag{5}$$

In Eq. (4) $\mathbf{x} = [0, \ldots, 0, x_1^i, \ldots, x_m^i, 0, \ldots, 0]^T \in \mathbb{R}^m$ is the coefficient vector. It is important to notice that the elements of $\mathbf{x}$ that are nonzero are only associated with the ith category. Therefore, the elements of vector $\mathbf{x}$ contain information regarding the identity of a given test object. Thus given a test sample (object), we are able to identify its category via its sparsest solution using equations (5). What we expect (ideally) is $\mathbf{x}$ to have non-zero entries only with objects of its actual category.

According to the sparsity requirements of Eq. (2), matrix $\mathbf{C}$, has to be overcomplete. Thus the cardinality of the training set has to be much larger than the vectors dimensionality. Commonly the problem is approached as follows. Given the dimensionality of vectors we select as many prototypes as needed to achieve over completeness. In our case though, the cardinality of prototypes equals the dimensionality of vectors. This problem can be simply solved by reducing the dimensionality of dissimilarity vectors via Principal Component Analysis (PCA). This is reasonable as the vector space formed by the dissimilarity vectors is endowed with the Euclidean metric and the inner product. This procedure also reduces significantly the complexity of the linear solvers. One has to notice here that this procedure consists of an atypical form of prototype selection due to the fact that every principal direction is also a set of weights on the training objects dissimilarities.

After the introduction of PCA, the optimization problem is reformulated as follows:

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{W}^T C\mathbf{x} = \hat{\mathbf{y}} \tag{6}$$

where $\mathbf{W} \in \mathbb{R}^{n\times r}$ with $r \ll n$ is a projection matrix and $\hat{\mathbf{y}} = \mathbf{W}^T\mathbf{y}$. Authors in [28] have shown that in case where matrix $\mathbf{W}$ is a random orthoprojection matrix which does not cause significant deterioration in the classification stage. Following the criteria presented in Section 4.5, $\hat{\mathbf{x}}$ can be used for identification of a test sample.

4.5. Evaluation criteria

The ideal hypothesis that the coefficient vector $\hat{\mathbf{x}}$ should not have zero entries in positions other than those associated with a single class does not always hold. To cope with this problem each vector $\hat{\mathbf{y}}$ is being classified based on the ability of the coefficients associated with each class to reconstruct $\hat{\mathbf{y}}$ in terms of smallest residual. Therefore, the classification operation can be formulated as follows:

$$\text{identity}(\hat{\mathbf{y}}) = \arg\min_i \left\|\hat{\mathbf{y}} - \mathbf{W}^T C\delta_i(\hat{\mathbf{x}})\right\|_2 \tag{7}$$

where $\|\cdot\|_2$ denotes the $\boldsymbol{l}^2$ norm and $\delta_i(\hat{\mathbf{x}}) \in \mathbb{R}^n$ is a new vector whose only nonzero entries correspond to the indices of $\hat{\mathbf{x}}$ that are associated with class i.

Recent advances in the insights of the classification SRC algorithm [54] show that in order to classify an object correctly the projection of its feature vector on a sub-space spanned by the training samples of this class has to be as close as possible to the original vector. Concurrently, the opposite should hold for the projection of the feature vector to the subspace spanned by all the other training samples – it to be as far as possible from the original vector. This "double-checking" is fundamental for achieving good classification performance under the presence of noise or small sample sets. Another important feature of all recognition systems is to determine whether or not a test sample is presented in the training set. An appropriate index is proposed in [28].

As mentioned before, the coefficient vector $\hat{\mathbf{x}}$, derived from the sparse representation of a sample that belongs to a class available in the training set, contains enough information to classify this sample, based on the expectation that the nonzero coefficients would be concentrated mostly on locations associated with one class. Pursuing further that reasoning the nonzero coefficients for an invalid sample is expected to be spread widely among multiple classes. To quantify this property the *Sparsity Concentration Index* (SCI) can be defined as follows:

$$\text{SCI}(\hat{\mathbf{x}}) = \frac{(k \cdot \max\left\|\delta(\hat{\mathbf{x}})\right\|_1 / \left\|\hat{\mathbf{x}}\right\|_1) - 1}{k-1} \in [0{,}1] \tag{8}$$

For a coefficient vector $\hat{\mathbf{x}}$ if $\text{SCI}(\hat{\mathbf{x}}) = 1$, the test sample is represented using only training samples from a single person, and if $\text{SCI}(\hat{\mathbf{x}}) = 0$, the sparse coefficients are spread evenly over all classes. Using SCI and a single threshold $\tau \in (0, 1)$ a sample can be accepted as valid if $\text{SCI}(\hat{\mathbf{x}}) \geq \tau$, otherwise rejected as invalid.

In this work, we chose to incorporate a verification criterion which utilizes information from both the first and second levels of processing, providing a boost in verification performance, as indicated by the experimental results. To this purpose we combine the above rule based on SCI index, along with the classical verification criterion of thresholding the minimum dissimilarity of a test sample to the training data (*MinDiss*), information derived from the first level of our scheme:

$$\text{Verification}(\hat{\mathbf{x}}) = \begin{cases} 1 & \text{if } \left(\text{SCI}(\hat{\mathbf{x}}) > \tau_1\right) \wedge \left(\min_i (y_i) < \tau_2\right) \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

where y_i is the element of the vector of dissimilarities computed at the first level, τ_1, τ_2 are threshold values and $\wedge$ denotes the logical conjunction operator.

4.6. Hierarchical data processing

The proposed architecture can be seen as a four stage data processing pipeline where at every stage a function is applied to the data resulting into a hierarchical representation. This deep architecture has proved to be advantageous compared to shallow approaches [26]. In particular, at the first stage skeletal data are used in order to extract feature vectors by computing the two Euler angles of eight selected limbs. This results into multidimensional trajectories as we mentioned in the previous section. At the second stage, these trajectories are mapped in the dissimilarity space resulting into dissimilarity vectors. These vectors are used then at the third stage where using an unsupervised learning technique, a vector of sparse coefficients is learned for every dissimilarity vector. Next these sparse vectors are finally classified at the fourth stage in the context of gait and gender recognition.

The selected representation has several advantages. First it can be used in a variety of tasks regarding human motion (action, gait, gender). This can be achieved by simply changing the classification rules at the final (fourth) stage. Second, the proposed architecture is characterized by low complexity. This is due to the fact that an unsupervised learning technique is incorporated only once in the third stage. Third, the resulted pipeline exhibits all the advantages of a deep architecture, allowing complex mappings of feature vectors that is usually difficult to achieve with shallow architectures.

The experimental results presented in the following sections confirm our initial belief that such a representation is competent for difficult biometric tasks.

5. Experimental results

In order to evaluate our method we used a publicly available dataset originally presented in [14]. As opposed to the authors in [14] here we evaluate our method on both human gait and gender recognition. Next we provide a short description of the dataset.

5.1. Dataset

The incorporated dataset [14] consists of pose sequences involving people walking in a straight line direction and was captured using the Microsoft Kinect sensor. The sensor was placed at 1.70 m above the ground, at the left of the walking path, with sensor's principal direction at an angle of $\sim 30^\circ$ relative to the walking line. Sequences from 30 persons – 15 women and 15 men between the ages of 23 and 55 – were captured during our experimental setup. Each person was asked to walk in a straight direction, without any visual aid drawn on the floor of the corridor indicating a straight path, at their own normal walking speed. There were captured 5 sequences in three separate sessions during the same day. Each captured sequence consists of 55 to 120 frames, depending mostly on the walking speed of the observed person. The pose estimation was performed using the provided SDK, at a frame rate of $\sim$30 fps.

5.2. Recognition task

In order to evaluate the recognition performance of the proposed framework, we used an experimental protocol where 3 randomly selected sequences were used as the training sequences for each person, and the remaining 2 were used as tests. Once the recognition rate is obtained the procedure is repeated again using a new randomization of the training and test sets. The final recognition rate is calculated as the average from 20 iterations of the test procedure.

In order to compare the performance of our framework, with recently proposed approaches to the pose-based gait recognition problem using Kinect, we have evaluated the methods proposed by authors in [36,37] using the dataset proposed in [14] and the same experimental protocol. More specifically, Preis et al. used the height, the length of legs, torso, both lower legs, both thighs, both upper arms and forearms, the step length, and the speed as features forming a 14-dimensional feature vector for each sequence. Ball et al. used the mean, standard deviation and maximum value of three angles for each of the left and right legs of the extracted skeleton. The angle of the upper leg relative to the vertical, the angle of the lower leg relative to the upper leg, and the angle of the foot relative to the horizontal form the 18-dimensional feature vector representing each sequence. During evaluation of both these methods the classification was performed using linear SVM.

Additionally, in order to support the choice of a multiple level architecture for the proposed framework, we also present results regarding recognition using only first and second level information i.e. pairwise dissimilarities between each test sample and the training samples, with the incorporation of a standard k-NN classifier with $k = 3$ neighbors. Finally, to demonstrate the contribution of the sparse representation, we have also evaluated the classification scheme proposed by authors in [41]. In this scheme linear SVM classifier is incorporated in order to classify data represented into dissimilarity space, and is considered state-of-the-art method for classifying dissimilarity data. Thus, for comparison purposes dissimilarity data from the first level were directed as input to the above scheme so as to measure the recognition performance. The protocol followed was the same for all experiments.

Table 1 summarizes the results for all the above experiments on the recognition task. The proposed framework was evaluated using both WW-Test [45] and MNPD [47] as dissimilarity functions. The projection matrix **W** of Eq. (5) was computed using PCA on the dissimilarity vectors of the respective training set, and consists of the 20 principal orientations. The results clearly show that the proposed framework using WW-test as first level dissimilarity function performs better compared to all the other evaluated methods, achieving 93.29% recognition rate. Classical gait analysis features incorporated in [36,37] performed very poorly on the dataset proposed in [14], advocating our approach of utilizing dynamic characteristics of pose sequences in order to perform gait recognition using commercial RGB-D sensors. On our framework, working with dissimilarity vectors, k-NN was able to achieve 80.54% recognition rate which is respectable, but significantly lower than the performance achieved with the incorporation of sparse representation. The benefits regarding the addition of another level of representation (sparse representation of dissimilarity vectors) before the final classification are also evident in the performance assessment of the classification scheme of [41], where 86.93% recognition rate was achieved. The latter, also supports the incorporation of sparse representation as the most appropriate method to model the second (higher) level relations of the data.

Table 1
Identification rates comparison (%) between the proposed scheme (SRC) for the two evaluated dissimilarity functions.

Method/results	WW	MNPD
Proposed scheme	93.29	77.27
Duin et al. (SVM in dissimilarity space)	86.93	57.84
First-level identification (k-NN)	80.59	53.18
Ball et Al. (SVM)	12.05	
Preis et al. (using SVM)	43.18	

5.2.1. Evaluating the effect of dimensionality

In order to study the effect of dimensionality induced by the projection matrix **W**, and the size of training set on the performance of the proposed framework, additional experimentation has been performed.

A number of pose sequences, between 1 and 4, were randomly selected for each person, in order to form the training set, and the remaining sequences were used as test sequences. Each experimental setup was repeated 20 times using different (random) permutations of the dataset as training and testing sets and the results were averaged. The whole procedure is repeated, projecting each time in different number of dimensions. The results are illustrated in Fig. 3 both for WW-test and MNPD.

It is evident that the dimensionality of the space where the dissimilarity vectors are projected plays an essential role to the attained performance. Actually, the optimum dimensionality is approximately the same as the number of involved persons. It is a reasonable conclusion, since classification into dissimilarity space is equivalent to classifying neighboring patterns, the number of which is expected to be equal to the number of classes. It is essential though, the dissimilarity vectors of each class, occupying their own subspaces in the initial dissimilarity space, to be able to span across their own principal direction into the reduced space, enabling the direction-sensitive SRC to easily discriminate them. Another notable fact is that even if only one sequence per person is available in the training set, the proposed method achieves a respectable 84.5% recognition rate using WW-test. This behavior constitutes a remarkable advantage for the proposed classification framework, compared to e.g. SVM where the training procedure is difficult when a very small amount of positive samples are available. This ability is partially due to the incorporation of dissimilarity representation of data, combined with the mechanism [54] of SRC and is essential in applications where training data are difficult to acquire.

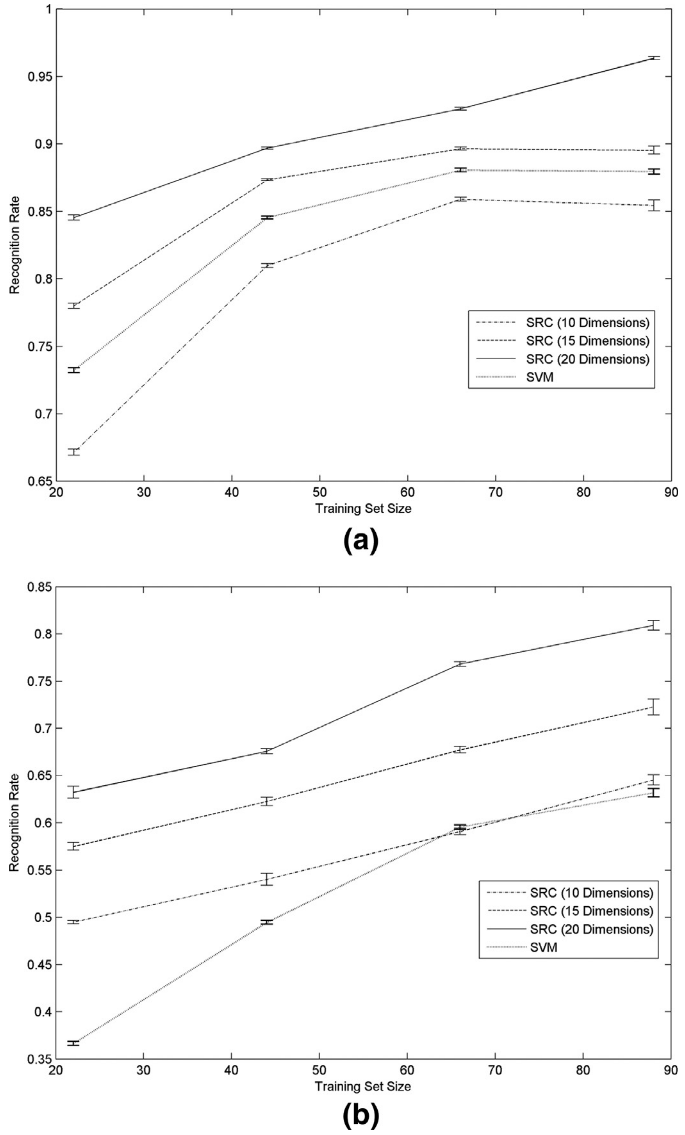


Fig. 3. Effect of projection dimensionality and training set's size on the performance of the proposed scheme, compared to SVM for: (a) WW test and (b) MNPD.

5.2.2. *Contribution of upper and lower body*

Aiming to further examine the contribution of upper and lower body parts to the overall recognition performance, we applied our framework to the input data partially. More specifically, the input to the first level was either 8-dimensional vectors with the Euler angles corresponding to the 4 skeletal primitives of the arms, or 8-dimensional vectors with the corresponding features from the legs. Results given in Fig. 4 illustrate the recognition rates for the upper and lower body features, using the experimental procedure described above. Results for full body features are also given for comparison reasons. The results indicate that the necessary information to discriminate persons from their walking exists both in the upper and the lower components of their estimated poses. Furthermore, the selected multi-level architecture of the proposed framework proves to be capable to extract this information, in a way that provides sufficient recognition performance. Partial incorporation of body features though, exhibit instability when the training set is limited to a single training sample per person, and the performance deteriorates faster, compared to the incorporation of all features. Therefore, given the fact that in real-world conditions the hands may be busy with activities irrelevant to walking, the incorporation of full body features seems the most reasonable choice.

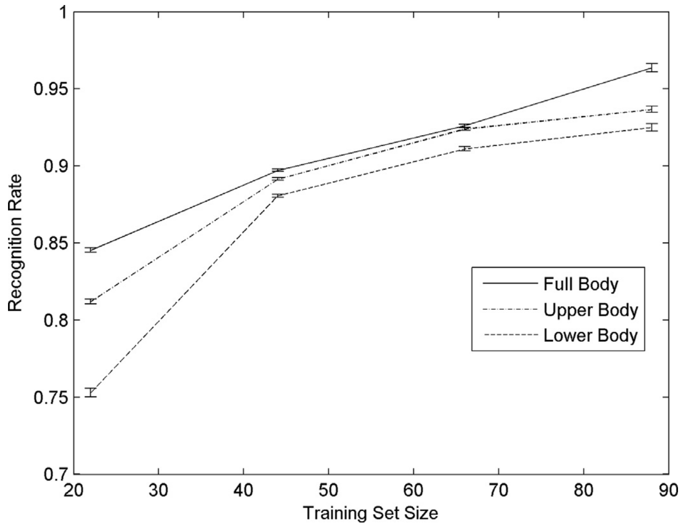


Fig. 4. Identification performance comparison using only upper, lower or full body components.

5.3. *Verification task*

The evaluation of the proposed framework on the verification task was performed by randomly selecting three sequences from each person in order to form the training set, and the remaining two for the test set. Then, 15 persons were randomly selected and the corresponding data were removed from the training set. Thus, 50% of the test samples are outliers, and the system is expected to be able to detect them. Three criteria were evaluated: the Sparsity Concentration Index (*SCI*), the minimum dissimilarity (*MinDiss*) and the combination of those, as described in Section 4.5. In order to quantify the performance of our method, ROC curves were constructed by sweeping the value of the corresponding thresholds, and measuring the False Acceptance Rate (FAR) and the Verification Rate (VR) for each threshold value. The combined criterion (*SCI&MinDiss*) requires two threshold values, one for each criterion. In order to simplify the evaluation procedure, the *MinDiss* criterion was calculated using dissimilarity values that had been normalized with respect to the maximum dissimilarity of the corresponding training set. Given this, the two threshold values were set to be equal. The experimental procedure was performed 100 times, using different randomization of training, testing and outliers sets, and the results were averaged. Fig. 5 illustrates the obtained ROC curves corresponding to WW-test (fig. 5a) and MNPD (Fig 5b) dissimilarity functions.

Results indicate that there is a clear advantage regarding the incorporation of information from both levels of processing during the verification procedure. The equal error rate (EER) for the *SCI&MinDiss* criterion is 3.1% and 5.1% for WW-test and MNPD respectively. The *SCI* criterion performed sufficiently, achieving EER of 5.2% and 7.1% for WW-test and MNPD respectively, and the corresponding rates for *MinDiss* are 10.2% and 7.5%.

5.4. *Gender recognition*

Another task of significant interest for a variety of applications is the gait-based gender recognition. In order to demonstrate the effectiveness of the proposed framework upon this task, we have implemented an appropriate experimental procedure, slightly modified compared to the above. More specifically, a number of pose sequences were selected for each person in order to form the test set,

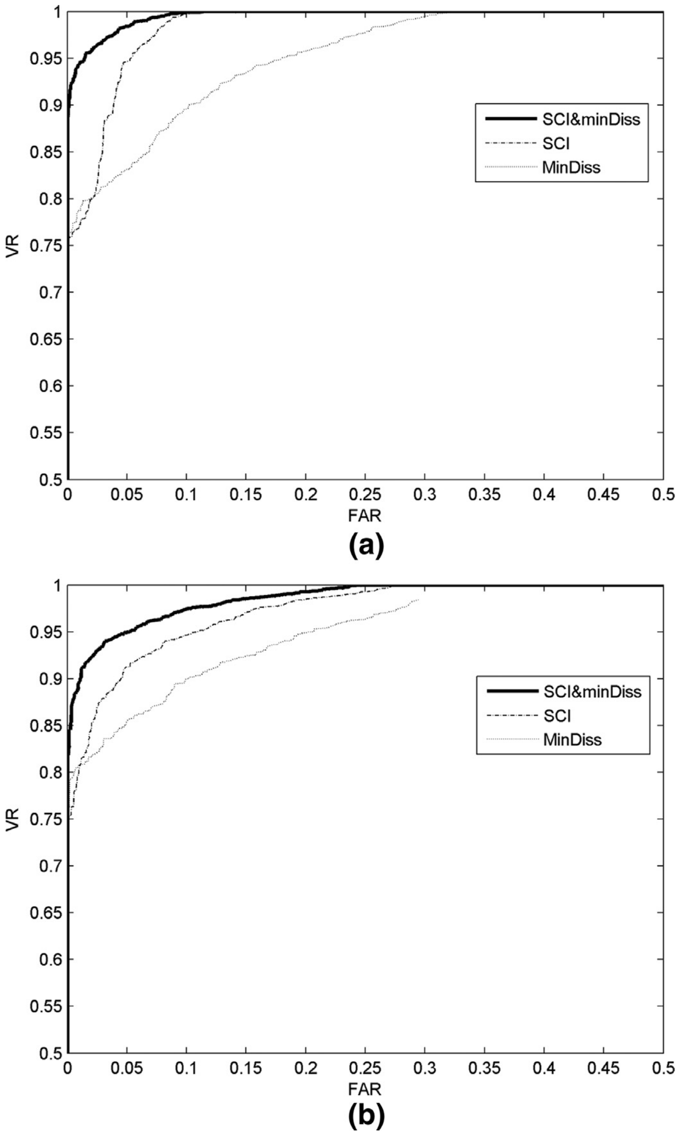


Fig. 5. ROC curves on the verification task (50% of true outliers) for (a) WW test and (b) MNPD.

Table 2
Recognition rates on the gender recognition task for variable training set size.

Training set size	20%	60%	100%
WW	94.45	97.64	99.11
MNPD	94.25	96.30	97.45

and the rest of the sequences- excluding samples from the same person- consisted the training set. Thus, the gender classification is performed using information only from other peoples' motion characteristics.

The results given in Table 2 show that the proposed framework is able to perform gender classification, achieving a recognition rate of ~94% even with a single sequence per person in the training set, for both the dissimilarity functions. The highest rates were 99.1% and 98.0% for WW-test and MNPD respectively.

6. Conclusions

In this work we presented a method for gait recognition based on a hierarchical representation of gait trajectories. For evaluation purposes, a publicly available dataset [14] consisting of walking sequences from 30 individuals, acquired using Microsoft Kinect was used. A thorough evaluation was performed on recognition and verification tasks, as well as in gender recognition. The proposed scheme was able to obtain high-level performance, achieving up to 96% identification rate, 99% gender recognition rate, and 3.1% EER on verification task. We showed how different components of our framework contribute to the overall performance, while comparing to the performance achieved by other recently proposed methods on a public available dataset. The suggested processing scheme proves to be essential to the obtained performance. Furthermore, we showed that the selected sparse representation approach in combination with the hierarchical representation exhibits significant advantages such as improved recognition performance, improved verification results and robustness to the small training data sample size.

Regarding future research directions, we are interested into extending the application of the proposed method into similar domains focusing on the detection of walking abnormalities, falling prediction, gait-based mental state assessment etc. Furthermore in the context of biological inspired human motion analysis we are interested to add more layers incorporating learning algorithms (both supervised and unsupervised) in order to study the representational power of these approaches.

References

[1] J. Lange, M. Lappe, A model of biological motion perception from configural form cues, J. Neurosci. 26 (11) (2006) 2894–2906.
[2] J. Shi, X. Weng, S. He, Y. Jiang, Biological motion cues trigger reflexive attentional orienting, Cognition 117 (3) (2010) 348–354.
[3] G. Johansson, Visual perception of biological motion and a model for its analysis, Percept. Psychophys. 14 (2) (1973) 201–211.
[4] 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.
[5] L. Kozlowski, J. Cutting, Recognizing the sex of a walker from a dynamic point-light display, Percept. Psychophys. 21 (6) (1977) 575–580.
[6] F. Loula, S. Prasad, K. Harber, M. Shiffrar, Recognizing people from their movement, J. Exp. Psychol. Hum. Percept. Perform. 31 (1) (2005) 210–220.
[7] N. Troje, C. Westhoff, M. Lavrov, Person identification from biological motion: effects of structural and kinematic cues, Percept. Psychophys. 67 (4) (2005) 667–675.
[8] W.H. Dittrich, Action categories and the perception of biological motion, Perception 22 (1) (1993) 15–22.
[9] J.F. Norman, S.M. Payton, J.R. Long, L.M. Hawkes, Aging and the perception of biological motion, Psychol. Aging 19 (1) (2004) 219–225.
[10] N.F. Troje, Decomposing biological motion: a framework for analysis and synthesis of human gait patterns, J. Vis. 2 (5) (2002) 371–387.
[11] X. Qinghan, Technology review: biometrics-technology, application, challenge, and computational intelligence solutions, IEEE Comput. Intell. Mag. 2 (2) (2007) 5–25.
[12] N.V. Boulgouris, D. Hatzinakos, K.N. Plataniotis, Gait recognition: a challenging signal processing technology for biometric identification, IEEE Signal Process. Mag. 22 (6) (2005) 78–90.
[13] M.S. Nixon, J.N. Carter, Automatic recognition by gait, Proc. IEEE 94 (11) (2006) 2013–2024.
[14] D. Kastaniotis, I. Theodorakopoulos, G. Economou, S. Fotopoulos, Gait-based gender recognition using pose information for real time applications, in: Proceedings of the 18th International Conference on Digital Signal Processing, DSP, 1–3 July 2013, 2013, pp. 1–6.
[15] I. Theodorakopoulos, D. Kastaniotis, G. Economou, S. Fotopoulos, Pose-based human action recognition via sparse representation in dissimilarity space, J. Vis. Commun. Image Represent. 25 (1) (2014) 12–23.
[16] R. Poppe, Vision-based human motion analysis: an overview, Comput. Vis. Image Underst. 108 (1–2) (2007) 4–18.
[17] M.S. Nixon, J.N. Carter, Advances in Automatic Gait Recognition, Paper presented at the IEEE Face and Gesture Analysis 2004, FG04, 2004.
[18] A. Roy, S. Sural, J. Mukherjee, Gait recognition using pose kinematics and pose energy image, Signal Process. 92 (3) (2012) 780–792.
[19] W. Jin, M. She, S. Nahavandi, A.A. Kouzani, Review of vision-based gait recognition methods for human identification, in: Proceedings of International Conference on Digital Image Computing: Techniques and Applications, DICTA, 1–3 December 2010, 2010, pp. 320–327.
[20] S. Sivapalan, D. Chen, S. Denman, S. Sridharan, C.B. Fookes, Gait energy volumes and frontal gait recognition using depth images, in: Proceedings of Paper presented at the International Joint Conference on Biometrics, Washington DC, USA, 2011.

[21] M. Hofmann, S. Bachmann, G. Rigoll, 2.5D gait biometrics using the depth gradient histogram energy image, in: Proceedings of the IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems, BTAS, 23–27 September 2012, 2012, pp. 399–403.

[22] R. Borràs, À. Lapedriza, L. Igual, Depth information in human gait analysis: an experimental study on gender recognition, in: A Campilho, M Kamel (Eds.), Image Analysis and Recognition, 7325, Springer, 2012, pp. 98–105. Lecture Notes in Computer Science.

[23] A. Yao, J. Gall, G. Fanelli, L.V. Gool, Does human action recognition benefit from pose estimation?, in: Jesse Hoey, Stephen McKenna, Emanuele Trucco (Eds.) Proceedings of the British Machine Vision Conference, September 2011, BMVA Press, 2011, pp. 67.1–67.11.

[24] H. Jhuang, T. Serre, L. Wolf, T.A. Poggio, Biologically inspired system for action recognition, in: Proceedings of the IEEE 11th International Conference on Computer Vision, ICCV, 14–21 October 2007, 2007, pp. 1–8.

[25] M.S. Beauchamp, K.E. Lee, J.V. Haxby, A. Martin, FMRI responses to video and point-light displays of moving humans and manipulable objects, J. Cogn. Neurosci. 15 (7) (2003) 991–1001.

[26] A. Castrodad, G. Sapiro, Sparse modeling of human actions from motion imagery, Int. J. Comput. Vis. 100 (1) (2012) 1–15.

[27] L. Jiwen, W. Gang, P. Moulin, human identity and gender recognition from gait sequences with arbitrary walking directions, IEEE Trans. Inf. Forensics Secur. 9 (1) (2014) 51–61.

[28] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, M. Yi, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.

[29] J. Behrens, C. Pfüller, S. Mansow-Model, K. Otte, F. Paul, A.U. Brandt, Using perceptive computing in multiple sclerosis – the Short Maximum Speed Walk test, J. NeuroEng. Rehabil. 11 (2014) 89.

[30] L. Jiwen, T. Yap-Peng, Gait-based human age estimation, IEEE Trans. Inf. Forensics Secur. 5 (4) (2010) 761–770, doi:10.1109/tifs.2010.2069560.

[31] L. Sigal, D. Fleet, N. Troje, M. Livne, Human attributes from 3D pose tracking, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), Proceedings of European Conference on Computer Vision Computer Vision, ECCV, vol. 6313, Berlin, Heidelberg, Springer, 2010, pp. 243–257.

[32] D. Hodgins, The importance of measuring human gait, Med. Device Technol. 19 (5) (2008) 44–47.

[33] B.E. Maki, Gait changes in older adults: predictors of falls or indicators of fear, J. Am. Geriatr. Soc. 45 (3) (1997) 313–320.

[34] J. Barth, J. Klucken, P. Kugler, T. Kammerer, R. Steidl, J. Winkler, J. Hornegger, B. Eskofier, Biometric and mobile gait analysis for early diagnosis and therapy monitoring in Parkinson's disease, in: Proceedings of Conference of the IEEE Engineering in Medicine and Biology Society, 2011, 6090226.

[35] E.E Stone, M. Skubic, Passive in-home measurement of stride-to-stride gait variability comparing vision and Kinect sensing, in: Proceedings of IEEE Annual International Conference on Engineering in Medicine and Biology Society, EMBC, August 30–September 3 2011, 2011, pp. 6491–6494.

[36] J. Preis, M. Kessel, M. Werner, Gait Recognition with Kinect, in: C. Linnhoff-Popien (Ed.), Proceedings of the 1st International Workshop on Kinect in Pervasive Computing, 2012, Newcastle.

[37] A. Ball, D. Rye, F. Ramos, M. Velonaki, Unsupervised clustering of people from 'skeleton' data, in: Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction, HRI, 5–8 March 2012, 2012, pp. 225–226.

[38] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake, Real-time human pose recognition in parts from single depth images, in: Proceedings of the Paper presented at IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[39] E. Pękalska, R. Duin, The Dissimilarity Representation for Pattern Recognition: Foundations and Applications, World Scientific Publishing Company, 2005.

[40] E. Pękalska, R.P.W. Duin, P. Paclík, Prototype selection for dissimilarity-based classifiers, Pattern Recognit. 39 (2) (2006) 189–208.

[41] R.W. Duin, E. Pękalska, M. Loog, Non-euclidean dissimilarities: causes, embedding and informativeness, in: M Pelillo (Ed.), Similarity-Based Pattern Analysis and Recognition. Advances in Computer Vision and Pattern Recognition, Springer, 2013, pp. 13–44.

[42] S. Gao, I.-H. Tsang, L.-T. Chia, Kernel sparse representation for image classification and face recognition, in: K Daniilidis, P Maragos, N Paragios (Eds.), Proceedings of European Conference on Computer Vision Computer Vision, ECCV, vol. 6314, Berlin,Heidelberg, Springer, 2010, pp. 1–14.

[43] R. Belman, Dynamic Programming, Princenton University Press, 1957.

[44] H. Bunke, On a relation between graph edit distance and maximum common subgraph, Pattern Recognit. Lett. 18 (8) (1997) 689–694.

[45] J.H. Friedman, L.C. Rafsky, Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests, Ann. Stat. 4 (1979) 697–717.

[46] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the Hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. 15 (9) (1993) 850–863, doi:10.1109/34.232073.

[47] S.-C. Fang, H.-L. Chan, Human identification by quantifying similarity and dissimilarity in electrocardiogram phase space, Pattern Recognit. 42 (9) (2009) 1824–1831.

[48] E. Amaldi, V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, Theor. Comput. Sci. 209 (1–2) (1998) 237–260.

[49] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inf. Theor. 47 (7) (2006) 2845–2862.

[50] D.L. Donoho, For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution, Commun Pure Appl. Math. 59 (6) (2006) 797–829.

[51] K. Labusch, E. Barth, T. Martinetz, Sparse coding neural gas: learning of over complete data representations, Neurocomputing 72 (7–9) (2009) 1547–1555.

[52] M. Elad, M. Aharon, Image denoising via sparse and redundant representations over learned dictionaries, IEEE Trans. Image Process. 15 (12) (2006) 3736–3745.

[53] P.O. Hoyer, Non-negative matrix factorization with sparseness constraints, J. Mach. Learn. Res. 5 (2004) 1457–1469.

[54] D. Zhang, Y. Meng, F. Xiangchu, Sparse representation or collaborative representation: which helps face recognition? in: Proceedings of IEEE International Conference on Computer Vision, CV, 6–13 November 2011, 2011, pp. 471–478.