Gait-based Gender Recognition using Pose Information for Real Time Applications

Dimitrios Kastaniotis, Ilias Theodorakopoulos, George Economou and Spiros Fotopoulos
Electronics Laboratory, Physics Dept.
University of Patras
Patras, Greece
{dkastaniotis, iltheodorako}@upatras.gr, {economou, spiros}@physics.upatras.gr

Abstract— **Biological cues inherent in human motion play an important role in the context of social communication. While recognizing the gender of other people is important for humans, security, advertisement and population statistics systems could also benefit from such kind of information. In this work for first time we propose a method suitable for real time gait based gender recognition relying on poses estimated from depth images. We provide evidence that pose based representation estimated by depth images could greatly benefit the problem of gait analysis. Given a gait sequence, in every frame the dynamics of gait motion are encoded using an angular representation. In particular several skeletal primitives are expressed as two Euler angles that cast votes into aggregated histograms. These histograms are then normalized, concatenated and projected onto a PCA basis in order to form the final sequence descriptor. We evaluated our method on a newly created dataset -UPCVgait - captured with Microsoft Kinect, consisting of 5 gait sequences performed by 30 subjects. An RBF kernel SVM used for classification in a leave one person out scheme on gait sequences of arbitrary length as well as on variable number of frames confirms the efficiency of our method.**

Keywords— real time gender recognition; depth imaging; gait sequence; histogram encoding; svm classification

I. INTRODUCTION

Human brain is able to effortlessly process biological motion. Gait constitutes a human motion containing several unique characteristics that can be proved useful in inferring social cues inherent in human movement. Indeed, research on this field has evinced that given a gait sequence, a person is able to recognize familiar people as well as the gender, age, weight, intention and mental state even for unfamiliar ones [1]. Biometrics is the scientific field that deals with methods aiming to identify a person or to infer several attributes like gender, age, mental state etc., by utilizing biological information. In contrast to other modalities like iris, face, voice etc, gait possesses several unique characteristics. The most important is the unobtrusiveness in which physical contact as well as cooperation with the subject is not required. Moreover, gait is hard to imitate, able to work at a distance while also retains confidentiality as opposed to face.

Human visual system is finely tuned for perceiving the social cues immanent in human movement. Despite humans ability to efficiently process complex biological motions in short time, they are not able to process large amount of information concurrently. Additional to this natural restriction, human performance in processing visual information typically degrades after some time.

Several applications ranging from biometrics and security to population statistics and even medical assessment could benefit from the information conveyed in gait sequences. Usually such applications require the analysis of large amounts of visual information that is prohibitive for human beings. In this context, a great effort is being made in order to imitate human brain ability in biological perception, attempting to infer high level information and achieve meaningful representations of face, iris, palm-print, gait, and voice. Despite the significant performance achieved by imitation of biological perception, as a drawback this often leads to high cost, complex, computational and memory demanding systems that are not appropriate to operate in real time. In this context the need for a reliable and low cost system that would efficiently approximate human brain performance has emerged.

In this work, we propose a low cost system for pose based gender recognition that relies only on depth images captured by a low cost RGB- D Sensor [2]. The proposed method is robust to view point changes, occlusion, differences in clothing and carrying objects, while can operate in real time utilizing only a small amount of system resources. Furthermore our method can operate in absence of light while can work even with a few number of frames, which is essential for real time applications (where tracking for long time periods still remain a challenging task). Also we present a newly dataset for posed based gait gender recognition captured using the Kinect sensor that we make publicly available. Our dataset is totally balanced and is suggested as a benchmark for pose based, gait gender recognition. Experimental results demonstrate the contribution of pose estimation in the problem of gait analysis as well as the efficiency of our proposed method as opposed to other recent relying on silhouettes extracted from depth images.

The rest of this paper is structured as follows. In the next section we present the background of gait analysis and in Section 3 the related work. Next in section 4 we introduce our method and in section 5 the dataset used for evaluation. In section 6 we provide the experimental results, and in section 7 we make a thorough discussion about the proposed method and its advantages over others.

978-1-4673-5807-1/13/$31.00 ©2013 IEEE DSP2013

II. BACKGROUND

Traditionally, two major approaches to gait analysis have been proposed, namely model-free and model-based methods.

Model- free which are also called appearance based methods rely on low level processing and are able to work with low resolution images [3]. They usually employ the silhouette or the contour of a segmented human body which they use in order to encode human motion characteristics in the form of a template. While they are very computationally efficient, they also have some significant drawbacks mostly due to changes in view point, lighting and clothing. These methods are poor in preserving the dynamics of human motion as the final template is usually formed by averaging a number of frames of a gait cycle and are inadequate to capture static features. Also, the need of gait cycle detection introduces more errors while also increases the overall complexity. Model based approaches are based on a pre-defined model of the underlying kinematics in order to represent the data and capture gait dynamics. While they are more computational expensive than model free methods, they can lead to more informative representations while also being robust to view point, scale, clothing variances and the carrying of objects. They can be used to obtain several static and dynamic features by tracking the trajectories of several body parts. Primarily dynamic features consisted by stride, cadence and velocity, while other more recent methods rely on joint angle trajectories [4]. For a survey on gait based recognition approaches we refer the reader to [5].

Traditionally the problem of 3D pose estimation which aimed to eliminate the need for view-based models and constrained domains was based on synchronized multiple cameras. Given an initial pose and several estimated body segments, body segments in next frames are inferred and the motion of every segment is computed using a gradient descent scheme. In order to deal with rapid movements stochastic tracking techniques as for example particle filters were proposed. The large and impractical number of particles needed to sample the high- dimensional state space was reduced by an annealed particle filter or by a Particle filter with an adaptive particle numbers whose aim is to decompose the search space and decrease the computational complexity [6]. While these methods achieve remarkably good results, they are still expensive, computational demanding, often require a pose initialization, and mostly they are not able to recover from failures in tracking. For example recently, using synchronized binocular video and a 3D pose tracker based on a modified Annealed Particle Filter, authors were able to process one frame every 2 minutes using the Matlab programming language. As they report the tracker performed well except in cases where legs were close, while they also noticed some rare cases where leg identities were switched. Furthermore they collected MoCap data that used as regularization in terms of domain adaptation in order to overpass the problem of overfiting that appears in small amounts of noisy video pose data [1].

Classically, depth imaging devices were very expensive thereby unsuitable for everyday applications. With the launch of low cost devices like Microsoft Kinect, many generalizations of classical methods from the field of gait analysis have reformed in order to incorporate depth information. The impact of this progress was first emerged in the field of traditional appearance based methods [7- 10] as well as in model based methods [4], [11], [12]. Early methods exploited depth information mostly by means of providing a more accurate silhouette extraction thus augmenting older descriptors based solely on color. Although this replacement of color with depth information led to significant performance improvements, in fact the strength provided by depth imaging was not fully exploited, namely; the problem of view point and scale still remains [8].

However, the intrinsic lighting and color invariance in Depth imaging have proven to be much more beneficial to the pose estimation problem rather than this of accurate silhouette extraction. Recently an algorithm able to fast and reliable compute a human pose from Depth data was proposed by Shotton et.al. [13]. This algorithm formed the core algorithm of Microsoft Kinect SDK pose estimation algorithm, works at a pixel level and is robust into recovering from failure as it can accurate provide a robust pose estimation in every frame. It operates at 30fps on a personal computer and at 200fps on a consumer device by taking advantage of the GPU. From a perspective of gait analysis and medical assessment applications where accuracy cannot compensate with cost, several studies [14] based on the pose estimation from depth images proposed by [13] prove the reliability of Kinect sensor and the accompanying pose estimation algorithm as oppose to invasive Motion Capture Systems (MoCap Systems). Furthermore, in the context of pose estimation, Yao et al. in [15] showed that pose based approaches could be beneficial for the task of action recognition. In this context [11] proposed a pose based representation method in order to recognize actions captured with a depth sensor. Joint positions are encoded into spatial histograms and thus the dynamics of the human motion are analyzed. In a comparison of their algorithm with that proposed by [16] that is based on histograms of silhouette points, they showed that a pose representation can achieve significant better results in the task of action recognition.

In another approach, authors in [12] targeting the task of human identification, tried to encode the style of several individuals by fusing 13 static and dynamic features. They used the height, the length of legs, torso, both lower legs, both thighs, both upper arms, forearms, the steplength, and the speed. Later we provide experimental results on our dataset using the features proposed by [12] and we show that static features could also be discriminative for the task of gender recognition.

III. RELATED WORK

In the context of real time gender recognition using RGB-D sensors, authors in [9] proposed a new publicly available dataset 53 persons (36 male and 17 female) consisted by depth gait sequences performed and captured by multiple view angles. The rationale of their work is to transfer the problem of gait analysis from depth frames to a problem of describing shape distributions [17]. In order to perform gender classification they formed a descriptor that combines (fuses) histograms of pairwise Euclidean distances of 2- Dimensional boundary points of an extracted silhouette and histograms of pairwise distances of 3 Dimensional point clouds obtained by stacking cropped depth silhouettes. The presence of random sampling points from the silhouette contour and the 3D point cloud introduces significant variability in the resulted classification rate. Furthermore their method does not solve the problems induced by changes in scale, view and carrying objects. Additionally, their dataset uses annotated gait cycles, which makes the problem not realistic due to the fact that in real case scenarios, estimation of gait cycle of a walking person is extremely difficult while might even be impossible if not prohibitive in real time scenarios.

In [18], authors using silhouette templates [3] extracted from monocular cameras, shown that several parts of the template image have different contribution to the overall performance of gender classification. Thus, in [10] authors proposed to use depth information in order to achieve a view invariant transformed Gait Energy Image (GEI) [3] in a form of an image histogram that would take advantage of distinct parts as proposed in [18]. While they surpass the problem of view angle by projecting the silhouette surface onto a PCA plane attached to the body, they still have to deal with the problem of clothing variability and carrying object occlusions. Furthermore scale invariance is not resolved as the silhouette is cropped using a predefined window size that is not adaptable to the distance from the sensor. Their evaluation is performed on strictly annotated walking cycles achieving significant improvement in contrast to the method they proposed in [9]. However, away from this ideal “gait cycle” annotated environment the classification accuracy degrades significantly. This is shown in the evaluation procedure, in the scenario where a window of 20 frames is used in order to approximate a walking cycle. The results received by this approach are significantly decreased.

IV. Proposed Method

Our method utilizes walking style information in order to recognize the gender of individuals by encoding into normalized histograms, information conveyed in a gait sequence. Compared to other methods requiring annotated gait cycles for train and test, our method is characterized by its ability to work with gait sequences of arbitrary length while also remains operational even in cases where only a few frames are available A skeletal representation consisting of 20 joints and estimated using Shotton’s technique [13], is employed to model the underlying kinematics. In addition an efficient data representation algorithm is incorporated to express the directions of several skeletal primitives corresponding to limbs. Furthermore a novel and efficient feature extraction approach, able of capturing the dynamics of human motion in the form of a normalized histogram is used. Only particular dynamic features are invoked while all static ones (height, length of legs, arms, etc) are discarded.

The whole procedure can be split in two parts. A more detailed overview of the method is given in figure2. The two parts include:

a) The on-line aggregation of limb direction information into histograms through skeletal primitives Euler angles voting and

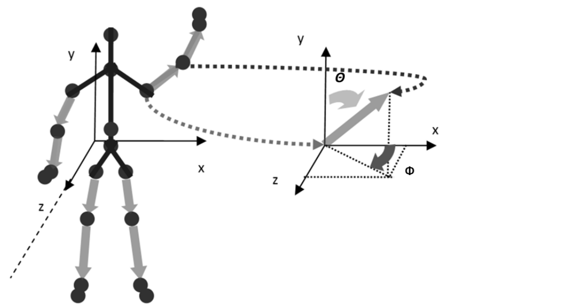

Figure 2. Left: Used skeletal Primitives. Right: Angular representation of the directionality for every skeletal primitive using Euler angles.

b) The concatenation and normalization of the histograms computed in the previous step. These are subcequently projected on a PCA basis in order to reduce the dimensionality. This final histogram will be the input vector to the classifier

Particularly, in the first part for every frame of a gait sequence, a skeletal representation is estimated and transformed to a coordinate system attached to the human body, in order to achieve view invariance. Then, the directions of several limbs are represented with a set of two Euler angles $\{\varphi, \theta\}$ as illustrated in figure 1.

These angles estimated for every limb cast votes into a corresponding histogram of bin size 40 $\{h\varphi, h\theta\}$ where $h\varphi, h\theta \in \mathbb{R}^{40}$, hence information of the directions of several limbs is on line aggregated into the corresponding histograms $\{h\varphi, h\theta\}$.

In the second part, histograms of all limbs are finally normalized and concatenated into a histogram $H \in \mathbb{R}^{640}$. This final histogram deriving from normalized aggregated histograms is then projected on a basis computed by PCA in order to form the final descriptor for every sequence. PCA basis is computed using histogram descriptors from the training set. A Gaussian SVM trained with histogram descriptors of the train set, is used in order to classify the computed histogram descriptor. In the following paragraphs we give a detailed description of our method, starting with the limb representation through Euler angles and then continue with the feature extraction procedure.

A. Data representation

Each captured gait sequence is modeled using a 20 joint skeletal representation as in [11, 4]. Initially, skeletal joint positions are provided according to the sensor coordinate system. In order to achieve view invariance, in every frame we computed a coordinate system that is attached to the human body according to the torso-PCA framework proposed in [4]. Thus, the 20 skeletal joints are projected on the three eigenvectors computed by PCA on seven “torso” joints consisted of the two shoulders, the center of the shoulders, the waist, the right and left hip. The new coordinate system is attached to the human body thus providing us view invariance. Then, we select eight skeletal primitives corresponding to Humerus, Radius, Femur and Tibia bones from both sides of

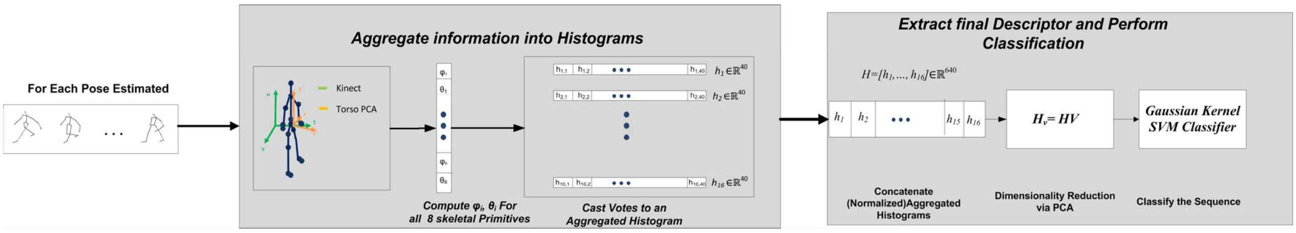

Figure 2. Overview of our method.

human body, considered to be the most informative in the context of human motion. In this transformed space, every primitive is represented by a feature vector consisting of two Euler angles {φ, θ} as illustrated in figure 1. This implies that for every frame of a gait sequence we have eight pairs of angle $\{\varphi_i, \theta_i\}$, $i = 1,..,8$.

B. *Feature Extraction and Classification*

We designed our descriptor in order to compactly encode the gait motion information independently of the number of available frames. To this purpose, normalized histograms of individual angles are computed on-line along a sequence of frames, and then concatenated in order to form the final descriptor. Specifically, each one of the 16 angles casts a vote to its corresponding histogram h_i and thus given a gait sequence, for every feature φ_i , θ_i we compute a normalized histogram of 40 bins $h_i \in \mathbb{R}^{40}$. Hence we have 16 histograms of 40 bins each, that we concatenate into a feature vector $H \in \mathbb{R}^{640}$ as follows:

$$H = [h_{\varphi 1,} h_{\theta 1}, ..., h_{\varphi 16,} h_{\theta 16}] \quad (2)$$

In order to reduce the dimensions of our descriptor we performed dimensionality reduction using PCA, by projecting on the 90 eigenvectors corresponding to the largest eigenvalues, resulting a 90 dimensional descriptor $H_v \in \mathbb{R}^{90}$

$$H_v = HV \in \mathbb{R}^{90} \qquad (3)$$

Where $V \in \mathbb{R}^{640 \times 90}$ is the matrix holding the 90 principal directions in the form of column vectors, computed each time on a subset of the training sample.

In the final stage of our method a Gaussian Kernel SVM classifier is employed in order to classify gait sequences to corresponding gender categories {male, female}. Kernel parameters optimized using a subset of the available set.

V. DATASET

To the best of author's knowledge, there is no publicly available dataset for pose based gait recognition captured by Microsoft Kinect. Thus, in order to evaluate our method, we created the new dataset named –UPCVgait- that we made publicly available at [19]. Our dataset consists of pose data captured from 30 subjects (15 males and 15 females). Every subject was instructed to walk freely in a corridor, capturing five sequences from each one in three sessions. Kinect Sensor was placed at two meters from the ground, at about 30- 40 degrees. It should be noticed that for a proper performance evaluation our dataset is balanced in terms of gender of the subjects. In every frame an estimated pose is represented using the 20 point model depicted in figure 1. Figure 3 illustrates the RGB as well as the corresponding estimated skeletal pose from three frames of a gait sequence.

In order to approximate a realistic scenario of gender recognition, gait information is acquired from sequences of arbitrary length that might even be non-continuous. Thus, for every one of the 150 gait sequences a categorical label male or female is assigned. Our dataset is publicly available and is suggested as a benchmark for evaluating algorithms designed for pose based gait gender classification.

Figure 3. A subject performing one of the five gait sequences. Top: RGB images. Bottom: Skeletal

VI. EXPERIMENTAL RESULTS

A. *Classification Accuracy*

In general, given a set of gait sequences as training data $X_{\text{train}} \subseteq X = [H_{v1}, \ldots, H_{v150}]$ where H_i denotes the descriptor computed for sequence i, and categorical labels $Y_{\text{train}} \subseteq Y$, where $Y = [y_1, \ldots y_{150}]$ and $y =$ {male,female}, SVM solves a constraint optimization problem in order to maximize the margin between the two classes. In the first protocol called leave one sequence out, we trained a classifier using all gait sequences except one which is used as test. Therefore, every one of the 150 train samples is used as a test sample $X_{\text{test}} = H_{\text{i,}}$ where i=[1,..,150] and $X_{\text{test}} \nexists X_{\text{train}}$ and the rest as train. The presence of samples belonging to the test subject in the train set, might benefit the task of identification. Hence, in the second protocol, we removed all sequences of test subject which allowed us to get a more reliable evaluation of our algorithm. Therefore, all 5 sequences recorder by every person are used as test samples while the rest of the data (29 people) are used as train. Thus, from all the persons recorded $X_p=[X_{p1},..X_{p30}]$, where for every person i there are 5 gait sequences,

$$X_{pi} = \left[H_{v[(i-1)*5+1]}, \ldots, H_{v[(i-1)*5+5]} \right] \in \mathbb{R}^{90\times 5} \quad , \qquad (4)$$

$X_{test}=X_{pi}$ and $X_{pi} \nexists X_{train}$. The overall accuracy in both protocols is computed as the mean classification rate of all test samples.

In [12] authors proposed the use of some classical static and dynamic features for gait analysis for the task of human identification. Here, using our dataset we compare the results of our method with the method proposed in [12]. Results in Table-I indicate that regardless the simplicity of the proposed method, aggregated histograms are able to encode all the necessary information in order to perform gender recognition very efficiently. Moreover, the proposed method outperforms the method proposed in [12] in both experimental protocols. In particular, results using the "Leave One Sequence Out", protocol indicate that when sequences of the test person are present in the training set, classification accuracy is significantly improved. This implies that our features are more appropriate into capturing the unique style of several individuals as opposed to features proposed by [12], which is important for future work in the context of human identification.

TABLE I. CLASSIFICATION ACCURACY

	Leave one Sequence Out	**Leave one Person Out**
Our method	96.67	86.67
Preiss et al.[12]	88.00	83.33

Regarding other experimental protocols, authors in [9] using depth silhouettes, proposed a leave one subject out evaluation protocol which is based on a majority voting scheme. We tested our algorithm on our dataset using this protocol and the obtained classification accuracy was *100%*, although we had only 5 gait sequences for every person as opposed to their case where 11 sequences were available for each of the 53 persons.

Since the proposed method has no intrinsic restrictions in the number of frames required to function properly and is designed to operate in real-time, in order to compare our method with other based on Depth information we provide experimental results using the two different evaluation procedures presented in [10] in the "Experiments on the video sequences" section. According to this experimental protocol, in the first procedure, given a gait sequence, every frame is described by the features computed in a window of 20 past frames that is assumed to approximate a walking cycle. We refer to this procedure as "Results on Frames". In a rather different procedure, the same authors used a threshold of 60% of correctly classified individual "frames" in order to classify a subject to a category. We refer to this procedure as "Majority on Frames". Results shown in Table- II indicate that the proposed method based on a compact aggregated histogram description outperforms the results presented in [10] in both different protocols. This states that our method is more appropriate for real time applications as well as the fact that pose estimation could benefit the gender recognition task as opposed to silhouette extracted from depth images.

TABLE II. COMPARISON WITH METHOD [10]

	Our Method	**Igual et.al. [10]**
Results on Frames	79.00	74.70
Majority on Frames	93.33	92.31

B. Classification Rate- Number of Frames- Missing Frames

Tracking people in real word -especially in crowded scenes- still remains a challenging task due to the presence of significant occlusions and/ or restricted visual field caused by stationary (e.g. walls) or moving objects (e.g. cars). In the case of gait based gender recognition one has to track a walking person through a number of frames. These facts raise the question of how many frames are actually needed in order to accurately perform gender classification. In a similar problem authors [20] noticed that it is possible to recognize an action from video data in small time snippets 5- 7 video frames.

Since our method relies on the construction of histograms derived from accumulation of information from a number of frames of a particular gait sequence, it is important to study the classification accuracy with respect to the number of frames. Results presented in figure 3 indicate that our method is able to infer gender information even with a small number of frames. More precisely, the following figure shows that with approximately 20 frames our method will be able to adequately classify an incoming gait sequence with ~86% accuracy. Also it is worth noticing that the performance reaches a plateau after a certain number of frames.

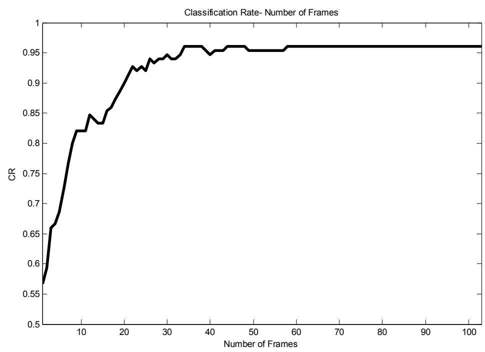

Figure 4. Classification Rate (CR)- Number of Frames for the leave one sequence out protocol.

VII. DISCUSSION

In this work we present a method for gait based gender recognition using pose estimation derived from depth images. Also a totally balanced and publicly available dataset UPCVgait is introduced, consisting of 150 gait sequences from 15 males and 15 females captured using Microsoft Kinect. Depth imaging exhibits several advantages due to the absence of color and its invariance to lighting conditions. These advantages have proved beneficial into solving the problem of human pose estimation and recently algorithms able to accurately estimate a human in real time have been proposed. In this context, our method is based on the utilization of the dynamics of gait motion by using dynamic features extracted by several limbs. In every frame of a given gait sequence, the directions of several limbs are represented by the Euler angles in a coordinate system attached to the human body. The computed angles are used to encode the dynamics by casting a vote into corresponding histograms, calculated in an on-line fashion. These histograms are concatenated and projected onto a subspace computed by PCA, resulting in a 90 dimensional feature descriptor. Finally, a non-linear SVM with RBF kernel is used for the gender classification. We evaluated the proposed scheme using several experimental protocols and the obtained results indicate high-level performance in terms of accuracy and robustness. We compared the results to other similar techniques proposed in the literature, and we showed that the proposed method outperforms the other schemes under all the evaluation protocols.

Beyond classification rates though, the proposed system is characterized by several other advantages. In particular, our method can operate independently of the view angle, the number of frames (with respect to a walking cycle), the scale, clothing variances, and possible carrying objects. Another significant advantage of our method is that does not rely on an analysis based on exact gait cycles, the detection of which is an extremely challenging task that usually prohibits methods from operating in real time. For example authors in [10] noticed that, good performance is achieved only if they utilize pre-annotated gait cycles. Without the exact information of gait cycles they observed significant reduction of performance.

Another advantage of using orientational features is the intrinsic scale invariance, which makes the representation invariant to the distance between the subject and the sensor. On the other hand, the performance of applications relying on view invariance, achieved by projecting a whole depth silhouette onto a body attached plane, might degrade due to change of scale; a complication that e.g. authors in [10] have not considered.

Results using the leave one sequence out indicate that our features are able to encode style information for several individuals. Furthermore, compared to the results obtained by [12], our histogram encoding scheme of human gait seems more promising for the task of gait based human identification. In the future, we plan to study the performance of the proposed scheme for encoding human motion on the task of real-time human identification. Also, we plan to expand our dataset by adding more subjects as well as providing also depth and RGB information.

ACKNOWLEDGMENT

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

REFERENCES

[1] M. Livne, et al., "Human attributes from 3D pose tracking," Computer Vision and Image Understanding, vol. 116, pp. 648-660, 2012.

[2] http://www.microsoft.com/en-us/kinectforwindows/

[3] H. Ju, "Individual Recognition Using Gait Energy Image," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 316-322, 2006.

[4] M. Raptis, et al., "Real-time classification of dance gestures from skeleton animation," presented at the Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, Vancouver, British Columbia, Canada, 2011.

[5] W. Jin, et al., "A Review of Vision-Based Gait Recognition Methods for Human Identification," in Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on, 2010, pp. 320-327.

[6] G. Junxia, et al., "Action and Gait Recognition From Recovered 3-D Human Joints," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 40, pp. 1021-1033, 2010.

[7] Sivapalan, Sabesan, Chen, Daniel, Denman, Simon, Sridharan, Sridha, & Fookes, Clinton B. (2011) Gait energy volumes and frontal gait recognition using depth images. In International Joint Conference on Biometrics, IEEE, Washington DC, USA. (In Press).

[8] M. Hofmann, et al., "2.5D gait biometrics using the Depth Gradient Histogram Energy Image," in Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on, 2012, pp. 399-403.

[9] R. Borràs, et al., "Depth Information in Human Gait Analysis: An Experimental Study on Gender Recognition," in Image Analysis and Recognition. vol. 7325, 2012, pp. 98-105

[10] Laura Igual, Àgata Lapedriza, Ricard Borràs, "Robust gait-based gender classification using depth cameras", , to appear in EURASIP Journal on Image and Video Processing 2013, 2013:1.

[11] X. Lu, et al., "View invariant human action recognition using histograms of 3D joints," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, pp. 20-27.

[12] Johannes Preis, Moritz Kessel, Martin Werner, and Claudia Linnhoff-Popien, "Gait recognition with kinect", In 1st International Workshop on Kinect in Pervasive Computing, 2012.

[13] J. Shotton, et al., "Real-time human pose recognition in parts from single depth images," presented at the Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[14] Moshe Gabel, Erin Renshaw, Assaf Schuster, and Ran Gilad-Bachrach, Full Body Gait Analysis with Kinect, in Proceedings of EMBC 2012, Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), August 2012.

[15] A. Yao, J. Gall, G. Fanelli and L.V. Gool. "Does Human Action Recognition Benefit from Pose Estimation?," In Jesse Hoey, Stephen McKenna and Emanuele Trucco, Proceedings of the British Machine Vision Conference, pp. 1- 11. BMVA Press, September 2011.

[16] L. Wanqing, et al., "Action recognition based on a bag of 3D points," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, 2010, pp. 9-14.

[17] R. Osada, et al., "Shape distributions," ACM Trans. Graph., vol. 21, pp. 807-832, 2002.

[18] Y. Shiqi, et al., "A Study on Gait-Based Gender Classification," Image Processing, IEEE Transactions on, vol. 18, pp. 1905-1910, 2009.

[19] http://www.upcv.upatras.gr/projects/UPCVgait_dsp/

[20] K. Schindler and L. Van Gool, "Action snippets: How many frames does human action recognition require?," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.