Human identification from human movements

Nikolaos Gkalelis[†⋆], Anastasios Tefas[⋆], Ioannis Pitas[†⋆]

[†]*Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece*
[⋆]*Department of Informatics, Aristotle University of Thessaloniki, Greece*
{galelis,tefas,pitas}@aiia.csd.auth.gr

***Abstract*—In this paper a multi-modal method for human identification that exploits the discrimination power of several movement types performed from the same human is proposed. Utilizing a fuzzy vector quantization (FVQ) and linear discriminant analysis (LDA) based algorithm, an unknown movement is first classified, and, then, the person performing the movement is recognized from a movement specific person classifier. In case that the unknown person performs more than one movements, a multi-modal algorithm combines the results of the individual classifiers to yield the final decision for the id of the unknown human. Using a publicly available database, we provide promising results regarding the discrimination power of the different movements for the human identification task, as well as we indicate that the combination of the individual classifiers may increase the robustness of the human recognition algorithm.**

I. Introduction

Identification of humans from video sources using gait has been recently attracted increasing attention in several application domains, e.g., for content-based video annotation and retrieval, video surveillance, and other applications, as this technology is the only one offering non-invasive, unobtrusive human identification [1], [2], [3], [4], [5]. However, the vast majority of the researchers in this topic concern only walk as human biometric, while only very few works have been reported that utilize run as a second gait biometric [6]. To the best of our knowledge, identification of humans from gait types or in general other human movements, different from walk or run, is still an unexplored topic.

Exploitation of more than one movements for the task of human identification may be realistic and beneficial in many applications for several reasons. First of all, some humans may not be considerable different from others in the way they walk but in the way they perform another movement, e.g., skip or jump. Moreover, in many applications the human that should be identified may perform more than one movements, where the movement of walk may not be even included. For instance, a criminal captured by a hidden camera during a robbery may be depicted to run, jump, or even skip. Similarly, in video annotation applications, movies to be annotated may show actors that walk, run, etc.

To allow the use of movement-based person classifiers, the different movements contained in a test video should be firstly extracted and recognized. Currently, many promising movement recognition algorithms have been proposed [7], a development, which can considerable advance as well as advocate the use of specific movement person classifiers and their combination for the task of human identification.

Motivated from the above discussion, we design a number of human identification models that recognize humans from different movements, and combine these models to provide a robust human identification system. The components of this system are presented in section II, while experimental results, regarding the discrimination power of the individual classifiers as well as the overall identification system are presented in section III. Finally, conclusions regarding the proposed approach are given in section IV.

II. Proposed method

In [8] we presented a method that combines FVQ and LDA to recognize a human movement depicted in a video stream. We have extended this method to build a multi-modal human identification algorithm, which is described in detail in the following paragraphs.

A. Preprocessing

In video based recognition systems a movement is represented by a sequence of frames, where at each frame, a distinct posture of the movement is depicted. A basic requirement of our system is that the binary body posture mask at each frame is obtainable. This requirement can be relatively easily satisfied in cases of static/constant background. From the body posture masks, the body posture regions of interest (ROIs) are extracted, centered in respect to the centroid of the body postures along the whole movement sequence, and scaled to the same dimension using bicubic interpolation. A ROI is scanned column wise to produce the so called posture vector $\mathbf{x} \in \Re^F$, where F is the number of pixels in the ROI.

B. FVQ plus LDA for video sequence classification

Let $\mathcal{W}$ be an annotated database of movement videos belonging to one of $q = 1, \ldots, Q$ different classes, where q declares a specific person or a specific movement, depending on the recognition task. We represent the $\imath$th video sequence of the qth class with length $L_\imath$ as a set of posture vectors $\{\mathbf{x}^{(q)}_{\imath,1}, \ldots, \mathbf{x}^{(q)}_{\imath,L_\imath}\}$, and, similarly, the whole database with the set $\{\mathbf{x}^{(1)}_{1,1}, \ldots, \mathbf{x}^{(1)}_{1,L_1}, \ldots\ldots, \mathbf{x}^{(Q)}_{O_Q,1}, \ldots, \mathbf{x}^{(Q)}_{O_Q,L_{O_Q}}\}$, where O_q is the number of sequences in the q-th class and $O = \sum_{q=1}^{Q} O_q$ is the total number of movement sequences in the database.

Each class is modelled as a mixture density, where the mixture components are represented by their centers, called dyneme vectors, $\{\mathbf{v}_1, \ldots, \mathbf{v}_C\}$. Considering unlabelled data and assuming that the number of dynemes C and the fuzzification parameter m are known, the fuzzy c-means (FCM) algorithm is used to compute the dynemes, and FVQ is applied to compute the quantized posture vectors $\boldsymbol{\phi}_\imath \in \Re^C$, $\boldsymbol{\phi}_\imath = [\phi_{c,\imath}]$, where,

$$\phi_{c,\imath} = \frac{(\| \mathbf{x}_\imath - \mathbf{v}_c \|_2)^{\frac{2}{1-m}}}{\sum_{\jmath=1}^{C} (\| \mathbf{x}_\imath - \mathbf{v}_\jmath \|_2)^{\frac{2}{1-m}}} \ . \tag{1}$$

After relabelling the posture vectors, the arithmetic mean of the quantized postures is taken to represent the $\imath$th video sequence of the qth class,

$$\mathbf{s}_\imath^{(q)} = \frac{1}{L_\imath} \sum_{\jmath=1}^{L_\imath} \boldsymbol{\phi}_{\imath,\jmath}^{(q)} \ . \tag{2}$$

Therefore, each video in the database is represented by a single vector, $\{\mathbf{s}_1^{(1)}, \ldots, \mathbf{s}_{O_1}^{(1)}, \ldots \ldots, \mathbf{s}_{O_Q}^{(Q)}\}$. The labelling information can be further exploited to reduce the dimensionality of the feature vectors using LDA. Assuming that $\boldsymbol{\Psi} \in \Re^{C \times Q-1}$ is the projection matrix computed using LDA the final representation of the video is $\mathbf{y}_\imath^{(q)} = \boldsymbol{\Psi}^T \mathbf{s}_\imath^{(q)}$. The qth class (movement type or human id) can then be represented by the mean of all feature vectors belonging to this class, i.e.,

$$\boldsymbol{\zeta}^{(q)} = \frac{1}{O_q} \sum_{\imath=1}^{O_q} \mathbf{y}_\imath^{(q)}, q = 1, \ldots, Q \ . \tag{3}$$

Hence, in order to classify a test movement video we first retrieve the feature vector $\mathbf{z}$ of the video, as explained above, and then we compute Q cosine similarity values,

$$g_q(\mathbf{z}) = \frac{\mathbf{z}^T \boldsymbol{\zeta}^{(q)}}{\| \mathbf{z} \| \| \boldsymbol{\zeta}^{(q)} \|} \ . \tag{4}$$

The test video is assigned to the class represented by the prototype that gave the maximum cosine value,

$$y(\mathbf{z}) = \underset{q \in [1,\ldots,Q]}{\operatorname{argmax}} \ (\, g_q(\mathbf{z}) \,) \ . \tag{5}$$

The number of dynemes C and the fuzzification parameter m are initially not known. The LOOCV procedure is combined with the global-to-local search strategy, e.g., similar to [9], in order to identify C, m.

C. Human identification from human movements

Let $\mathcal{U}$ be an annotated movement video database that contains P persons performing R different movements, i.e., each movement video $\{\mathbf{x}_{\imath,1}^{(r,p)}, \ldots, \mathbf{x}_{\imath,L_\imath}^{(r,p)}\}$ has two labels, $r \in [1, R]$ and $p \in [1, P]$ regarding the movement type and the person it belongs respectively. Our target is to devise an algorithm that recognizes a person from a video sequence depicting this person executing the R different movements.

Using all the movement videos of the database and utilizing only the movement type labelling information r, the procedure described in section II-B is used to train a movement type classifier $y()$. Then, we break the movement video database to R distinct subsets $\mathcal{U}_r$, $r = 1, \ldots, R$, i.e., $\mathcal{U}_r$ subset contains only movement videos of the rth movement type. Each subset, $\mathcal{U}_r$, is then used to train a movement specific, person classifier, $h_r()$. The training of each classifier within each subset is done using the algorithm of section II-B where now only the person specific labelling information p is exploited.

At the testing phase, assuming that a test video depicts the same person performing sequentially the R different movements, the test video is segmented to produce R different movement videos. Then, each movement video is classified from the movement classifier $y()$, and it is directed to the respective movement specific, person classifier $h_r()$. Therefore, for each classifier a different feature vector $\mathbf{z}_r$ is computed. The movement specific classifiers have been trained using different training sets, and, thus, we may assume that the feature vectors of the test movement videos, are conditionally statistically independent. In this case, the sum rule proposed in [10] can be applied to combine the results of the individual classifiers as we explain below.

Each classifier $h_r()$ according to equation (4) produces P score values $g_{r,p}$, $p = 1, \ldots, P$, where each value is in the interval [-1,1] and refers to a specific person in the database. Let ω_p be the class representing the pth person. In order to produce estimates of the a posteriori probability $P(\omega_p|\mathbf{z}_r)$ of the pth class regarding measurement vector $\mathbf{z}_r$, the score values are shifted to the interval [0,1] and normalized using

$$P(\omega_p|\mathbf{z}_r) = \frac{g_{r,p}}{\sum_{i=1}^{P} g_{r,i}} \ . \tag{6}$$

Then, assuming equiprobable priors for each person class, the sum rule can be directly applied to yield the identity d of the person in the test video

$$d = \underset{p \in [1,\ldots,P]}{\operatorname{argmax}} \frac{1}{R} \sum_{r=1}^{R} P(\omega_p|\mathbf{z}_r) \ . \tag{7}$$

The algorithm is summarized in Figure 1. We should also note that the same algorithm can be applied in the case that the test video depicts the same person performing a fraction of the R movement types and not necessarily all of them.

III. Experimental results

In this section we present experimental results on the database reported in [11]. From this database we used low resolution videos (180×144 pixels resolution at 25 fps), depicting nine persons, namely, Daria (dar), Denis (den), Ido (ido), Ira (ira), Lena (len), Lyova (lyo), Moshe (mos), Shahar (sha), performing seven movements, i.e., walk (wk), run (rn), skip (sp), gallop sideways (sd), jump jack (jk), jump forward (jf) and jump in place (jp). A few body masks from 6 movement types in the database are depicted in Figure 2. Some videos depict a person performing more than one cycles of a specific movement, e.g., the videos of walk. We break such videos to their constituting single period movement videos to create a database of 193 movement videos in total. Each video sequence is labelled according to the person and movement that belongs to, and preprocessed as described in section II to yield 3072-dimensional posture vector sequences, where the

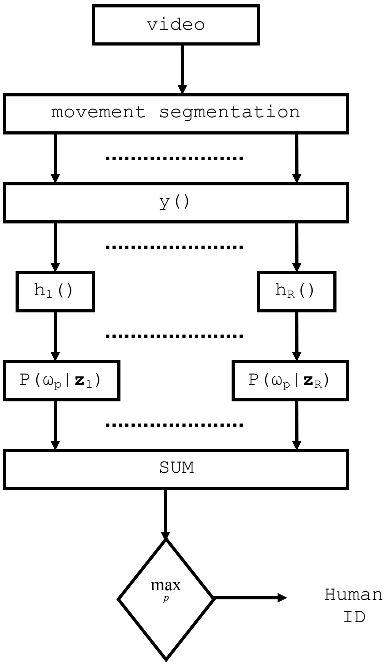

Fig. 1. *Recognition of humans from their movements.*

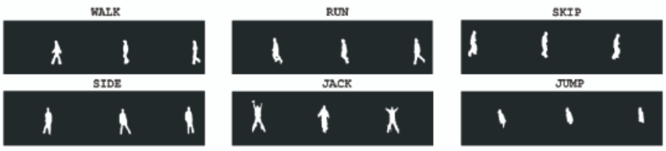

Fig. 2. *Three binary body masks for 6 human movements.*

posture vectors are formed by scanning column wise 64×48 pixel size ROIs.

A. Discriminative power of single movements

In order to assess the discriminative power of each movement type in the database, we create seven disjoint datasets, one for each movement type and then we apply the procedure described in section II-B to train seven movement specific, person classifiers, $h_{\text{wk}}(), h_{\text{rn}}(), \ldots, h_{\text{jp}}()$. In general, during the design of the classifiers we found that the optimal range for the fuzzification parameter was $m \in [1.1, 1.2]$, while the number of dynemes varied depending on the movement type, i.e, from $C = 20$ for jump forward to $C = 49$ for run. The correct classification rates (CCR) for each classifier is shown in Table I, while the confusion matrix regarding classifier $h_{\text{sp}}()$ is shown in Table II. Surprisingly, we see that the worst CCR was given from the individual classifiers based on the movements of walk and jack, while a CCR above 90% was obtained using the person classifiers based on the movements of skip, run and jump in place.

Classifier	CCR (%)
wk	78
rn	92
sp	93
sd	81
jk	77
jf	89
jp	92

TABLE I
CCR for each movement specific, person classifier.

	dar	den	eli	ido	ira	len	lyo	mos	sha
dar	4								
den		3							
eli			3						
ido								2	
ira					2				
len						7			
lyo							2		
mos								3	
sha									3

TABLE II
Identification of nine persons from the way they skip.

B. Human identification using several movements

The seven classifiers computed above can be combined using the framework presented in section II-C. To evaluate this algorithm we removed seven movement videos refering to object Daria, where each video depicts Daria to perform a different movement type. Using the recognition algorithm presented in [8] all the movement types in the test video were recognized correctly and each segmented movement video was passed to the respective person classifier. The similarity scores produced from the individual classifiers were combined using the sum rule to yield the final decision for the person depicted in the video. All the individual classifiers identified correctly the person in the videos, except from the classifier that uses the movement walk for human identification. However, using the sum rule the test person identified correctly with high confidence over the median, relatively to the scores of the sum rule regarding the other persons, as shown in Table III.

	dar	den	eli	ido	ira
$\frac{1}{R}\sum_{r=1}^{R} P(\omega_p\|\mathbf{z}_r)$	**0,83**	0,78	0,75	0,74	0,79
	len	lyo	mos	sha	—
$\frac{1}{R}\sum_{r=1}^{R} P(\omega_p\|\mathbf{z}_r)$	0,78	0,77	0,77	0,77	0.77

TABLE III
Recognition of human subject Daria using discriminant information from seven movements. In each column the score for the respective person in the database is computed using the sum rule. In the last column the median score value is given.

IV. Conclusions

We pursue the idea of movement identification using not only the walk gait but also other movement types. Assuming that videos depicting a person performing more than one movements are provided, we train a movement type classifier as well as movement specific person classifiers. At the testing phase, the different movement sequences of the test video are extracted and handed to the individual classifiers. Consequently, the scores yielded from each individual classifier are combined using the sum rule, to yield the final person identification decision. Due to the absence of an adequate database for evaluating our method, i.e., a database with a large number of both movements and persons, we used a dataset designed mainly for the purpose of movement type recognition, and, thus, our results are not fully representable of the capability of the proposed method, and can only be considered preliminary. However, the fact that the walk-based classifier provided the second worst performance, while other classifiers provided quite good results, indicates that movement types, other than walk, may contain considerable discriminant information for the task of human identification, which may be further exploited to design suitable combination schemes as we suggested here.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 211471 (i3DPost).

REFERENCES

[1] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa: Identification of Humans Using Gait. IEEE Trans. Image Process., vol. 13, no. 9, pp. 1163–1173 (2004)
[2] N.V. Boulgouris, D. Hatzinakos, K.N. Plataniotis: Gait recognition: a challenging signal processing technology for biometric identification. IEEE Signal Processing Magazine, vol. 22, no. 6, pp. 78–90 (2005)
[3] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, K.W. Bowyer: The humanID gait challenge problem: data sets, performance, and analysis. IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 2, pp. 162–177 (2005)
[4] D. Xu, S. Yan; D. Tao, S. Lin, H. J. Zhang: Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval. IEEE Trans. Image Process., vol. 16, no. 11, pp. 2811–2821 (2007)
[5] A.W.M. Smeulders, M. Worring,S. Santini, A. Gupta, R. Jain: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1349–1380 (2000)
[6] C. Y. Yam , M. S. Nixon, J. N. Carter: Gait Recognition By Walking and Running: A Model-Based Approach. In: Proceedings Asian Conference on Computer Vision, ACCV, (2002)
[7] P. Turaga, R. Chellappa, V. S. Subrahmanian, O. Udrea: Machine recognition of human activities: A survey. IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 11, pp. 1473–1488 (2008)
[8] N. Gkalelis, A. Tefas, I. Pitas: Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition. IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 11, pp. 1511–1521 (2008)
[9] J. Yang, A.F. Frangi, J. Y. Yang, D. Zhang, Z. Jin: KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 2, pp. 230–244 (2005)
[10] J. Kittler, M. Hatef, R. P. W. Duin, J. Matas: On combining classifiers. IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 3, pp. 226–239 (1998)
[11] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. Basri: Actions as Space-Time Shapes. IEEE Trans. Pattern Anal. Mach. Intell.,vol. 29, no. 12, pp. 2247–2253 (2007)