Fusion of movement specific human identification experts

Nikolaos Gkalelis[†‡], Anastasios Tefas[‡], and Ioannis Pitas[†‡]

[†]Informatics and Telematics Institute, CERTH, Greece
[‡]Department of Informatics, Aristotle University of Thessaloniki, Greece

Abstract. In this paper a multi-modal method for human identification that exploits the discriminant features derived from 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 recognition expert. In case that the unknown person performs more than one movements, a multi-modal algorithm combines the scores of the individual experts to yield the final decision for the identity of the unknown human. Using a publicly available database, we provide promising results regarding the human identification strength of movement specific experts, as well as we indicate that the combination of the outputs of the experts increases the robustness of the human recognition algorithm.

Key words: Multi-modal human identification, movement recognition, movement specific person recognition expert, fuzzy vector quantization, linear discriminant analysis.

1 Introduction

Identification of humans from video sources using gait has been recently attracted increasing attention in several application domains, e.g., for video surveillance, content-based video annotation and retrieval, and other applications, as this technology is the only one offering non-invasive, unobtrusive human identification [1–4, 6]. 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 [7]. 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 thief captured by a hidden camera during a pursuit may be depicted to run, jump, or performing some other movement. Similarly, in video retrieval application, it may be necessary to semantically label and retrieve videos of a specific actor performing a specific movement not necessarily the movement of walk.

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 [8], 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 propose the use of a number of human identification experts each of them trained to recognize a human from a specific movement type, and exploit a fusion algorithm in the score level to provide a robust human identification system. The components of this system are presented in section 2, while various experimental results, regarding the discrimination power of the individual classifiers as well as the overall identification system are presented in section 3. Finally, conclusions regarding the proposed approach are given in section 4.

2 Proposed method

In video based biometric systems a movement is represented by the so-called movement video, i.e., a video that depicts a person performing a single period of a movement, e.g., a step of walk. Therefore, at each frame of the video a unique posture of the movement is depicted. A basic requirement of our system is that the binary body mask at each frame has been retrieved. This requirement can be relatively easily satisfied in cases of static/constant background. From the body masks, the body regions of interest (ROIs) are extracted, centered in respect to the centroid of the bodies along the whole movement sequence, and scaled to the same dimension using bicubic interpolation. A ROI is scanned column wise to produce the respective vector $\mathbf{x} \in \Re^F$, where F is the number of pixels in the ROI, and, thus, a movement video is represented with a sequence of such vectors. We model the several movements as well as we design movement specific human recognition experts using a method that combines FVQ and LDA (FVQLDA). Finally we combine the output of the experts in order to provide a robust estimate for the identity of the human depicted in a test video. We briefly review FVQLDA as well as we describe the fusion strategy in the next subsections.

2.1 FVQ plus LDA for video classification

Let $\mathcal{U}$ be a training database of videos $\{\mathbf{x}_{i,j}\}$ belonging to one of R different classes, $\{\{\mathbf{x}_{i,j}\}, u_i \in \{1, \ldots, R\}\}$, where $\mathbf{x}_{i,j} \in \Re^F$ is the vector derived by scanning raster-wise the j-th frame of the i-th video, and u_i is its class label.

The task of a classification algorithm is using the above database to construct a classifier such that given a test video $\{\mathbf{z}_{t,j}\}$ to identify its class label u_t.

The intrinsic dimensionality of the biometric data is much lower than the dimension of the data in the image space $\Re^F$, and a dimensionality reduction method is commonly used to avoid the high computational cost of classification in the image space. Fisher's linear discriminant analysis (FLDA) [5], one of the most popular subspace techniques, cannot be directly used on most video-based recognition algorithms as usually the number of frames in the training database is mach smaller than the dimension of the data in the image space. To elevate this problem several LDA variants have been proposed [10], which are usually computationally expensive. Moreover, a computationally demanding distance metric, e.g., Hausdorff distance [4, 12], is commonly used to compare not aligned video sequences.

In [9] a method that combines FVQ and LDA has been used for movement recognition, which addresses the above issues providing a computationally efficient algorithm. In this method, the labelling of the data is initially ignored, and the FCM algorithm is used to provide C centroid vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_C\}$. Then, FVQ is applied to compute a membership vector for each frame, $\mathbf{x}_{i,j} \mapsto \boldsymbol{\phi}_{i,j} \in \Re^C$, $\boldsymbol{\phi}_{i,j} = [\phi_{i,j,c}]$, where the c-th component of the membership vector is given by

$$\phi_{i,j,c} = \frac{(\parallel \mathbf{x}_{i,j} - \mathbf{v}_c \parallel_2)^{\frac{2}{1-m}}}{\sum_{\ell=1}^{C} (\parallel \mathbf{x}_{i,j} - \mathbf{v}_\ell \parallel_2)^{\frac{2}{1-m}}} \ , \tag{1}$$

and, consequently, the i-th video with L_i frames is represented by the arithmetic mean of the membership vectors derived by its frames

$$\mathbf{s}_i = \frac{1}{L_i} \sum_{j=1}^{L_i} \boldsymbol{\phi}_{i,j} \ . \tag{2}$$

In the C-dimensional space conventional LDA can be applied to further reduce the dimensionality, as in most cases the dimensionality of the membership vectors will be smaller than the number of training videos. Thus, the final representation of the video will be $\mathbf{y}_i = \mathbf{W}^T \mathbf{s}_i$, where $\mathbf{W} \in \Re^{C \times R-1}$ is the projection matrix computed with LDA. The r-th class with cardinality N_r can then be represented by the mean of all its feature vectors

$$\boldsymbol{\zeta}_r = \frac{1}{N_r} \sum_{\mathbf{y}_i \in U_r} \mathbf{y}_i \ , \tag{3}$$

Assuming equiprobable priors as well as a common covariance matrix $\boldsymbol{\Sigma}$ for all classes, the feature vector $\mathbf{z}_t$ of a test movement video is first retrieved and then R Mahalanobis distance values are computed

$$g_r(\mathbf{z}_t) = (\mathbf{z}_t - \boldsymbol{\zeta}_r)^T \boldsymbol{\Sigma}^{-1} (\mathbf{z}_t - \boldsymbol{\zeta}_r) \ , r = 1, \ldots, R \ . \tag{4}$$

The test video is assigned to the class according to the following rule

$$u_t = y(\mathbf{z}_t) = \underset{r \in [1,\ldots,R]}{\operatorname{argmin}} \left(g_r(\mathbf{z}_t) \right) \ . \tag{5}$$

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 [10], in order to identify the optimal parameters C and m that best discriminate the R classes.

2.2 Fusion of movement specific human recognition experts

Let $\mathcal{U}$ be an annotated movement video database that contains P persons and R movement types, $\{\{\mathbf{x}_{i,j}\}, u_i \in \{1,\ldots,R\}, k_i \in \{1,\ldots,P\}\}$, i.e., each movement video has two labels, u_i and k_i, according to the movement type and the person it belongs respectively. Our target is to devise an algorithm that recognizes a person using a number of movement videos, where each movement video depicts the same person performing one of the R different movement types.

Using all the movement videos of the database and utilizing only the movement type labelling information u_i, the FVQLDA method is used to train a movement type classifier $y()$. Then, we break the database to R distinct subsets $\mathcal{U}_r$, $r = 1,\ldots,R$, i.e., $\mathcal{U}_r$ subset contains only movement videos of the r-th movement type, e.g. of the movement walk. Each subset is then used to train a movement specific person recognition expert $h_r()$. The training of each expert is done using again FVQLDA, where now only the person specific labelling information k_i is exploited.

At the testing phase, it is assumed that R movement videos, each of them depicting the same unknown person performing a different movement type, are available. This can be done for example, by temporarily segmenting a test video that depicts the same person performing sequentially the R different movements. Each movement video is classified from the movement classifier $y()$, and it is routed to the respective expert $h_r()$. Therefore, for each expert $h_r()$ a feature vector $\mathbf{z}_{t_r}$ is computed, while all feature vectors, $\mathbf{z}_{t_r}$, $r = 1,\ldots,R$, possess the same but unknown label k_t. The training and test videos of the movement specific person recognition experts are different, and, thus, it can be assumed that the feature vectors of the test movement videos, are conditionally statistically independent. In this case, the sum rule proposed in [11] can be applied to fuse the matching scores produced from each expert as we explain below.

Each expert $h_r()$ according to equation (4) produces P Mahalanobis distance values, $g_{r,p} = g_{r,p}(\mathbf{z}_{t_r})$, $p = 1,\ldots,P$, each referring to one of the P persons in the database. The distance values are transformed to matching scores by taking their reciprocal, and then normalized to produce an estimate of the a posteriori probability that the test video belongs to the class ω_p given the feature vector $\mathbf{z}_{t_r}$

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

where ω_p is the class representing the p-th person. Then, assuming that the posterior probabilities do not deviate dramatically from the prior probabilities, the sum rule can be applied to yield the identity k_t of the person in the test

video

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

The algorithm is summarized in Figure 1. We should note that the same al-

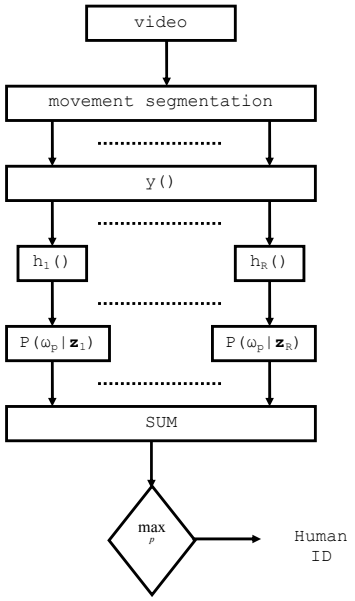

Fig. 1. *Recognition of humans from their movements.*

gorithm 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.

3 Experimental results

In this section we present experimental results on the database reported in [12]. From this database we used low resolution videos (180×144 pixels size, 25 fps), containing nine persons, namely, Daria (dar), Denis (den), Ido (ido), Ira (ira), Lena (len), Lyova (lyo), Moshe (mos), Shahar (sha), and seven movements, i.e., walk (wk), run (rn), skip (sp), gallop sideways (sd), jump jack (jk), jump forward (jf) and jump in place (jp).

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. Each video is labelled according to the person and movement that belongs to, and preprocessed as described in the beginning of section 2 to yield 3072-dimensional

vector sequences, where the vectors are formed by scanning raster-wise 64×48 pixel size ROIs.

3.1 Human recognition from a single movement

In order to assess the human characterization ability of each movement type in the database, we created seven disjoint datasets, one for each movement type and then we applied the procedure described in section 2.1 to train seven movement specific human recognition experts, $h_{\mathrm{wk}}(), h_{\mathrm{rn}}(), \ldots, h_{\mathrm{jp}}()$. During the design of the experts we found that the optimal range for the fuzzification parameter was $m \in [1.1, 1.2]$, while the optimum number of dynemes varied depending on the movement type, i.e, from $C = 20$ for jump forward to $C = 49$ for run. The

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

Table 1. *CCR for each movement specific person classifier.*

correct classification rates (CCR) for each expert is shown in Table 1, while the confusion matrix regarding $h_{\mathrm{sp}}()$ expert is shown in Table 2. Surprisingly, we see that the worst CCR was given from the experts based on the movements of walk and jack, while a CCR above 90% was obtained using the experts based on the movements of skip, run and jump in place.

	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 2. *Identification of nine persons from the way they skip.*

3.2 Human recognition from multiple movements

The experts computed above can be combined using the framework presented in section 2.2. To evaluate this algorithm we performed 26 experiments. At each

experiment we removed from the database five to seven movement videos to form a test case for the fusion algorithm, i.e., each video depicted the same person performing one different movement from the movements in the database. The movement types in the test videos were recognized correctly, and the respective test videos were routed correctly to the corresponding experts. For each experiment, the score values derived from the experts were fused using the sum rule. In some cases, one or more experts misclassified a test person. However, using the sum rule, in all 26 test cases the depicted human was identified correctly with high confidence over the median of the scores. In Table 3 we present the recognition results for nine test cases, one for each person in the database. Each row of the table corresponds to a specific evaluation of the fusion algorithm. The first column depicts the actual identity of the person for the specific test case, the next nine columns provide the confidence score for the respective person in the database, computed using the fusion algorithm, and the last column provides the median of the score values for the current evaluation.

	dar	den	eli	ido	ira	len	lyo	mos	sha	median
dar	**4.6455**	0.2932	0.2607	0.2517	0.3040	0.3308	0.2929	0.3101	0.3112	0.3040
den	0.3168	**3.5587**	0.2700	0.2538	0.3934	0.3162	0.2845	0.3275	0.2791	0.3162
eli	0.4148	0.3387	**2.1828**	0.2670	0.4354	0.3452	0.3402	0.3623	0.3137	0.3452
ido	0.3757	0.3379	0.3409	**1.6952**	0.3946	0.4051	0.3020	0.7964	0.3523	0.3757
ira	0.3596	0.4019	0.3705	0.2694	**4.2343**	0.4130	0.2849	0.3271	0.3392	0.3596
len	0.2538	0.2142	0.2123	0.2583	0.2752	**4.1482**	0.2197	0.2285	1.1898	0.2538
lyo	0.3646	0.4095	0.3263	0.3211	0.4245	1.0701	**3.3649**	0.3598	0.3591	0.3646
mos	0.5434	0.5527	0.5500	0.4781	0.6905	0.5825	0.5085	**2.5942**	0.5000	0.5500
sha	0.5784	0.5373	0.3941	0.4598	0.4764	0.5784	0.4240	0.5627	**2.9889**	0.5373

Table 3. *Evaluation of the fusion algorithm.*

4 Conclusions

We pursue the idea of movement identification using not only walk but also other movement types. Using a video database containing a number of persons and movement types, we train a movement type classifier as well as a number of movement-specific person recognition experts. At the testing phase, the movement type depicted on each video within a set of movement videos performed from the same test person is recognized, and each video is routed to the respective expert. Consequently, the scores yielded from each expert are combined using the sum rule, to yield the final decision for the identity of the test person. Several experiments have been performed, indicating 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 fu-

sion algorithms to enhance the accuracy and robustness of human recognition systems.

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) and COST Action 2101 on Biometrics for Identity Documents and Smart Cards.

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. K. Funkunaka: Statistical Pattern Recognition. Academic, San Diego, CA,(1990)
6. 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)
7. 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)
8. 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)
9. 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)
10. 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)
11. 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)
12. 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)