Available online at www.sciencedirect.com
SCIENCE DIRECT
NEUROCOMPUTING
ELSEVIER
Neurocomputing 69 (2005) 250–256
www.elsevier.com/locate/neucom

Letters

View-independent person identification from human gait

Zonghua Zhang*, Nikolaus F Troje

BioMotionLab, Department of Psychology, Queen's University, Canada

Received 17 May 2005; received in revised form 30 May 2005; accepted 2 June 2005
Available online 31 August 2005
Communicated by R.W. Neucomb

Abstract

Based on a three-dimensional (3D) linear model and the Bayesian rule, a method is explored to identify human walkers from two-dimensional (2D) motion sequences taken from different viewpoints. Principal component analysis constructs the 3D linear model from a set of Fourier represented examples. The sets of coefficients derived from projecting 2D motion sequences onto the 3D model by means of a maximum *a posterior* estimate is used as a signature of a walker. Simulating an identification experiment on a set of walking data we show that these signatures show invariance across viewpoints and can be used for viewpoint-independent person identification.

Keywords: View-independent gait identification; Linear model; Principal component analysis (PCA)

1. Introduction

Human gait contains biometric signatures that can be used for person identification [2]. Most practical application fields cannot rely on constant viewing conditions but require viewpoint independent approaches. A number of researchers have addressed the problem. For instance, Shakhnarovich et al. [4] developed a

*Corresponding author. Present address: Mechanical Engineering Heriot-Watt University Edinburgh EH14 4AS, UK. Tel.: +44 131 4514305; fax: +44 131 4513129.
E-mail addresses: z.zhang@hw.ac.uk, zhzhangtju@hotmail.com (Z. Zhang).

0925-2312/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2005.06.002

view-normalization recognition algorithm by synthesizing virtual sequences rendered from canonical viewpoints. Their recognition is based on image sequences from multiple cameras, instead of one image sequence. Grauman et al. [3] reconstructed a visual hull from a contour-based representation of human gait. Then, they used the visual hull to infer the three-dimensional (3D) structure of the human body in terms of its major joints. Even though 3D structure in terms of 19 joint locations could be reconstructed from pedestrian sequences, it had to be inferred from the viewpoint of four known training positions, not from an unknown viewpoint. Furthermore, it is not clear how well the algorithm would perform in a person identification task as this was not a subject of the study.

Here, we explore a view-independent gait identification method, which is motivated by a linear model of human gait [5,6]. Human gait is a cyclic motion and can be approximated by a Fourier expansion. Applying principal components analysis (PCA) to the Fourier representation we obtain a low-dimensional basis for our linear model. A gait signature of a two-dimensional (2D) motion sequence is defined as its projective coefficients, calculated using a Bayesian approach to obtain a maximum *a posterior* (MAP) estimate. The performance of identification of 2D projections from different viewpoints is quantitatively evaluated by means of a cross-validation procedure.

2. Linear model

Using an optical motion capture system (Vicon, Oxford Metrics, 120 Hz), we acquire human walking data as a time series of postures $s_n(t) : t = 1, 2, \ldots, T_n,\ n = 1, 2, \ldots, N$ specified by the locations of 15 discrete markers representing the main joints of the human body [5]. N is the number of walkers and T_n is the number of postures for walker n. Because each joint has three coordinates in 3D space, a posture s is a 45-dimensional vector. Using a second-order Fourier expansion, we can represent a time series $s(t)$ of postures obtained from walking in terms of its average posture p_0, the characteristic postures of the fundamental frequency (p_1, q_1) and second harmonic (p_2, q_2), and the fundamental angular frequency ω [6]

$$s(t) = p_0 + p_1 \sin(\omega t) + q_1 \cos(\omega t) + p_2 \sin(2\omega t) + q_2 \cos(2\omega t). \tag{1}$$

Using this transformation for a specific walking pattern $s_n(t)$ we obtain a 3D Fourier representation w_n by concatenating the five components into a single 225-dimensional vector:

$$w_n = (p_{0,n}, p_{1,n}, q_{1,n}, p_{2,n}, q_{2,n}). \tag{2}$$

We then applied PCA to a set of Fourier representations to learn the main motion variations such that we can now approximate a walker in terms of a sum of the average walker $\bar{w}$ and a linear combination of eigenwalkers $e_1, \ldots, e_J$:

$$w_n = \bar{w} + \sum_{j=1}^{J} k_{j,n} e_j. \tag{3}$$

$\bar{w} = (1/N)\sum_{n=1}^{N} w_n$ denotes the average walker and $k_{j,n}$ is the projective coefficient of eigenwalker e_j for walker n.

3. Gait signature

Analogous to the 3D Fourier representation, a 2D motion sequence $\hat{s}(t) : t = 1, 2, \ldots, T$ has the following 2D Fourier representation:

$$\hat{w} = (\hat{p}_0, \hat{p}_1, \hat{q}_1, \hat{p}_2, \hat{q}_2). \tag{4}$$

The average posture $\hat{p}_0$, and the characteristic postures $\hat{p}_1$, $\hat{q}_1$, $\hat{p}_2$, and $\hat{q}_2$ are 30-dimensional vectors containing 2D coordinates of the 15 markers, so the 2D Fourier representation $\hat{w}$ is a 150-dimensional vector.

Approximating a 3D Fourier representation of a walker by a linear combination of J eigenwalkers $e_1, \ldots, e_J$, the corresponding coefficients characterize each walker and represent a potential signature for gait identification. If the most probable 3D explanation of $\hat{w}$ is a linear combination of a set of eigenwalkers, the problem is to determine a set of coefficients $k_j, j = 1, \ldots, J$ for $\hat{w}$ with maximum probability in the 3D linear space.

A projection matrix $C(\alpha, \beta)$ relates $\hat{w}$ and its 3D Fourier representation:

$$\hat{w} = C(\alpha, \beta) w, \tag{5}$$

where $C(\alpha, \beta) : \Re^{225} \mapsto \Re^{150}$ is the projection matrix with an unknown viewpoint defined by azimuth α and elevation β (we assume an upright camera and orthographic projection).

We assume that each dimension of $\hat{w}$ is subject to uncorrelated Gaussian noise with variance σ^2. A matrix notation E denotes all eigenwalkers. Then the likelihood of measuring $\hat{w}$ is given by

$$p(\hat{w}/k) \propto \exp\left(-\|\hat{w} - \hat{\bar{w}}(\alpha, \beta) - \hat{E}(\alpha, \beta) k\|^2 / (2\sigma^2)\right), \tag{6}$$

where $\hat{\bar{w}}(\alpha, \beta) = C(\alpha, \beta)\bar{w}$ and $\hat{E}(\alpha, \beta) = C(\alpha, \beta) E$ denote the projections of the average walker and the eigenwalkers, respectively.

Assuming a normal distribution in each direction of the eigenspace, the prior probability is

$$p(k) \propto \prod_{j=1}^{J} \exp\left(-k_j^2 \Big/ (2\lambda_j)\right) = \exp\left(-\sum_{j=1}^{J} k_j^2 \Big/ (2\lambda_j)\right), \tag{7}$$

where λ_j is the eigenvalue corresponding to eigenwalker e_j.

According to the Bayesian rule, the posterior probability is

$$\begin{aligned} p(k/\hat{w}) &\propto p(\hat{w}/k) p(k) \\ &= \exp\left(-\|\hat{w} - \hat{\bar{w}}(\alpha, \beta) - \hat{E}(\alpha, \beta) k\|^2 / (2\sigma^2) - \sum_{j=1}^{J} k_j^2 \Big/ (2\lambda_j)\right). \end{aligned} \tag{8}$$

The optimal estimate k_{opt}, corresponding to a MAP estimate is then calculated by

$$k_{\mathrm{opt}} = k_{\mathrm{MAP}} = \left[diag\left(\frac{\sigma^2}{\lambda}\right) + \hat{E}(\alpha,\beta)^{\mathrm{T}}\hat{E}(\alpha,\beta)\right]^{-1} \times \hat{E}(\alpha,\beta)^{\mathrm{T}}\left(\hat{w} - \hat{\bar{w}}(\alpha,\beta)\right). \tag{9}$$

This is a nonlinear over-determined system and the optimal solutions for view angles α and β, and coefficients k_{opt} can be calculated by using a nonlinear least square fitting.

4. Experiment

Using walking data acquired with an optical marker-based motion capture system (see [5] for details) we conducted cross-validated simulations on a data set of 80 walkers using six different learning views and a large number of testing views. All 2D views were orthographic projections of the 3D walkers. Because human walking is a symmetric motion and the elevation angle is not very large in practical applications, we chose all views within a range of 0° (frontal) to 90° for azimuth angle α and from 0° (horizontal) to 30° for elevation angle β. Azimuth and elevation angle of the learning views were $(0,0)$, $(45,0)$, $(90,0)$, $(0,30)$, $(45,30)$ and $(90,30)$ deg, respectively. Testing views covered the whole range from 0° to 90° for azimuth and from 0° to 30° for elevation in 1° increments resulting in a total of 2821 probes for each walker.

We then divided the 80 walkers into four groups of 20 walkers each. Sixty walkers always served to construct the linear model while the remaining 20 walkers were used to test identification performance. The 20 learning views of these walkers are called a gallery. The signature of a walker in the gallery $K_{\mathrm{g}}(n)$, $n = 1, 2, \ldots, 20$ is the set of coefficients revealed from projecting a 2D Fourier representation with the known viewpoint onto the linear model. The coefficients $K_{\mathrm{g}}(n)$ are obtained by solving the nonlinear over-determined equation system (Eq. (9)) with known azimuth and elevation angles. The value of the equivalent intrinsic noise was set to $\sigma^2 = 70\,\mathrm{mm}^2$. A probe's signature K_{p} is the set of coefficients revealed from projecting a 2D Fourier representation onto the model, but with an unknown viewpoint. Again, the coefficients are obtained by solving Eq. (9), this time also optimizing for the unknown viewing angles. Because our interest is to evaluate the proposed view-independent method, a simple similarity measure $d(n) = \|K_{\mathrm{g}}(n) - K_{\mathrm{p}}\|^2$ (the Euclidean distance of two projective coefficient vectors) was defined to identify a walker according to a nearest neighbor rule. We can classify a novel probe into subject c that minimizes the similarity measure by

$$c = \arg\min_{n} d(n). \tag{10}$$

We calculated the identification rate from all testing viewpoints for the six learning views. The results are shown in Fig. 1. The dark and light parts show high and low identification rates. On average, identification rates are very high compared to

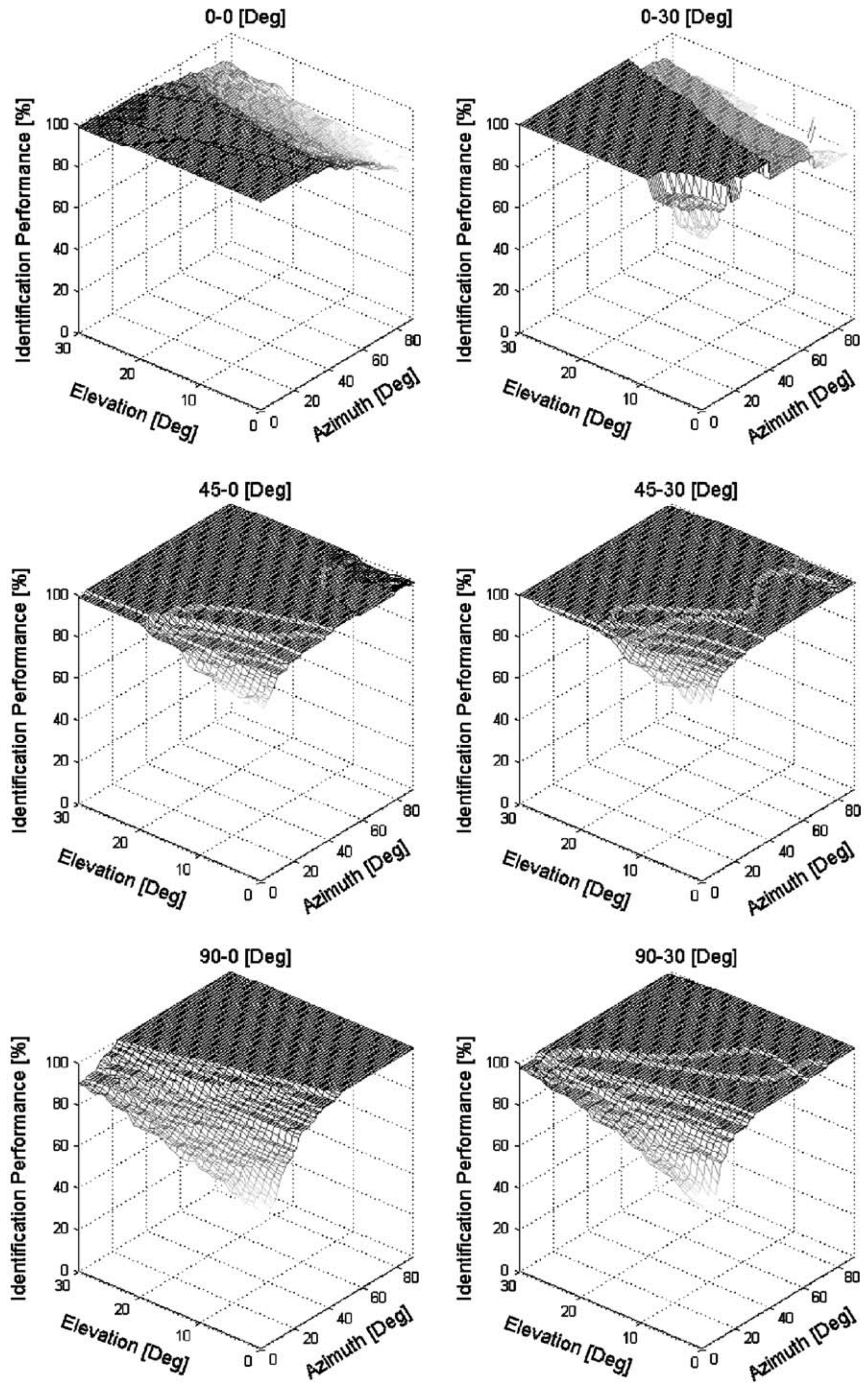

Fig. 1. Identification performance from different viewpoints for the six simulations Azimuth angle varied from 0° to 90° (rows) and elevation angle from 0° to 30° (columns).

Table 1
Average identification rate for six learning views

Gallery (α, β) (deg.)	0, 0	45, 0	90, 0	0, 30	45, 30	90, 30
Average rate (%)	91.8	98.8	95.6	93.9	98.8	96.9

chance level (5%). As expected, identification rates with viewing angles close to the learning view reach 100%.

The average performance of identification for each simulation is

$$\overline{p} = \frac{1}{2821} \sum_{\alpha=0}^{90} \sum_{\beta=0}^{30} p_{\alpha,\beta}. \tag{11}$$

The average performances for the six different learning views are presented in Table 1. The dependence of performance on learning view is relatively small. Yet, there is a clear advantage for oblique views (45, 0) and (45, 30) while performance with frontal views (0, 0) and (0, 30) is the worst.

5. Conclusions

This paper proposes a view-independent gait identification method, based on a linear model and a Bayesian approach. The model is constructed by submitting a set of Fourier representations of human walking to a principal component analysis. The calculation of gait signatures uses prior information to reduce ambiguity in a Bayesian framework. We evaluate this method quantitatively using walking data from different viewpoints. Results show generally very high performance. Furthermore, they suggest that the best viewpoint for 2D (e.g. video) recording of human motion for identification purposes is the $\frac{3}{4}$ view. Maybe not surprisingly, this finding is in accordance with work in the area of human face recognition [1].

References

[1] V. Bruce, T. Valentine, A.D. Baddeley, The basis of the 3/4 view advantage in face recognition, Appl. Cognitive Psychol. 1 (2) (1987) 109–120.

[2] D. Cunado, M. Nixon, J. Carter, Automatic extraction and description of human gait models for recognition purposes, Comput. Vis. Image Understand. 90 (1) (2003) 1–41.

[3] K. Grauman, G. Shakhnarovich, T. Darrell, A Bayesian approach to image-based visual hull reconstruction, in: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, Wisconsin, USA, June 2003, pp. 187–194.

[4] G. Shakhnarovich, L. Lee, T. Darrell, Integrated face and gait recognition from multiple views, in: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, Hawaii, USA, December 2001, pp. 439–446.

[5] N.F. Troje, Decomposing biological motion: a framework for analysis and synthesis of human gait patterns, J. Vis. 2 (5) (2002) 371–387.

[6] N.F. Troje, The little difference: Fourier based gender classification from biological motion, in: R.P. Würtz, M. Lappe (Eds.), Dynamic Perception, Aka Verlag, Berlin, 2002, pp. 115–120.

Zonghua Zhang received his B.S., M.S. and Ph.D. degrees from Tianjin University in 1992, 1998 and 2000, respectively. Then, he was a post-doctorate at the Mechanical Engineering Department of Tianjin University from December 2000 to October 2002. During April 2002 and September 2002, he was a research assistant at the Department of Computing of Hong Kong Polytechnic University. Supported by the Volkswagen Foundation, he was a post doctoral researcher in the BioMotionLab at Ruhr University in Bochum, Germany and later at Queen's University, Ontario, Canada. In 2005 he came to Heriot-Watt University where he is a research associate now. His research experiences include 3D imaging and modeling, human motion analysis, digital signal and image processing, and computer graphics.

Nikolaus Troje received his Ph.D. from the University of Freiburg, Germany, in 1994. Between 1994 and 1997 he worked as a researcher at the Max Planck Institute for Biological Cybernetics in Tübingen where he studied human face recognition and established the Max Planck Face Database. After 2 years at Queen's University in Kingston, Ontario, he recieved the prestigious Volkswagen Young Researcher's Award and set up the BioMotionLab at Ruhr University in Bochum, Germany. In 2003 he moved back to Queen's University where he is now a professor of psychology and Computer Science and a Canada Research Chair in Vision and Behavioural Sciences.