Available online at www.sciencedirect.com

ScienceDirect

Neurocomputing 70 (2007) 1938–1942

NEUROCOMPUTING

www.elsevier.com/locate/neucom

Extraction of spatio-temporal primitives of emotional body expressions

Lars Omlor, Martin A. Giese*

Laboratory for Action Representation and Learning, Department of Cognitive Neurology, Hertie Institute for Clinical Brain Research, University of Tübingen, Germany

Available online 9 November 2006

Abstract

Experimental and computational studies suggest that complex motor behavior is based on simpler spatio-temporal primitives, or synergies. This has been demonstrated by application of dimensionality reduction techniques to signals obtained by electrophysiological and EMG recordings during the execution of limb movements. However, the existence of spatio-temporal primitives on the level of the joint angle trajectories of complex full-body movements remains less explored. Known blind source separation techniques, like PCA and ICA, tend to extract relatively large numbers of sources from such trajectories that are typically difficult to interpret. For the example of emotional human gait patterns, we present a new non-linear source separation technique that treats temporal delays of signals in an efficient manner. The method allows to approximate high-dimensional movement trajectories very accurately based on a small number of learned spatio-temporal primitives or source signals. It is demonstrated that the new method is significantly more accurate than other common techniques. Combining this method with sparse multivariate regression, we identified spatio-temporal primitives that are specific for different emotions in gait. The extracted emotion-specific features match closely features that have been shown to be critical for the perception of emotions from gait pattern in visual psychophysics studies. This suggests the existence of emotion-specific motor primitives in human gait.

Keywords: Blind source separation; ICA; Delayed mixing; Emotions; Movement primitives; Kinematics; Gait analysis

Human full-body movements are characterized by a large number of degrees of freedom. This makes the accurate synthesis of human trajectories for applications in computer graphics and robotics a challenging problem. The analysis of motor behavior suggests the existence of simple basis components, or spatio-temporal primitives, that form building blocks for the realization of more complex motor behavior [6,12]. Since such basic components cannot be directly observed, several studies have aimed at identifying spatio-temporal primitives by application of unsupervised learning techniques, like PCA or ICA [8,3,5], to data from electrophysiological and EMG recordings (e.g. [9,2]). The same methods can be applied directly to joint angle trajectories. However, this analysis of complex full-body movements typically results in the extraction of a relatively large number of basic components or source signals that are difficult to control and interpret (e.g. [11]). In our study we tried to learn movement primitives of emotional gaits from joint-angle trajectories. We present a new technique for blind source separation, which is based on a non-linear generative model that, opposed to normal PCA and ICA, can model time delays between source components and individual joint angles. Opposed to other existing algorithms for blind source separation with delays [4,15], our method scales up to large problems, it allows dimensionality reduction, and it requires no additional sparseness assumptions. It provides a much better approximation of gait data with few basic components than other common unsupervised learning methods.

By approximating the trajectories of emotional gaits by superpositions of the extracted component signals and applying a sparsifying regression algorithm to learn a model for the mixing matrix, we extracted emotion-specific spatio-temporal features from the trajectory data. A comparison with psychophysical studies on the perception

*Corresponding author.
E-mail address: martin.giese@uni-tuebingen.de (M.A. Giese).
URL: http://www.uni-tuebingen.de/uni/knv/arl/index.html.

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.10.100

of emotional gaits reveals that the emotion-specific components derived from our kinematic analysis match features that have been described as fundamental for the visual recognition of emotions from gait. This indicates that the novel algorithm is suitable for the extraction of biologically valid movement components.

1. Trajectory data

Using a VICON motion capture system with seven cameras, we recorded the gait trajectories from 13 lay actors executing walking with four basic emotional styles (happy, angry, sad and fear), and normal non-emotional walking. Each trajectory was executed three times by each actor, resulting in a data set with 195 gait trajectories. Approximating the marker trajectories with a hierarchical kinematic body model (skeleton) with 17 joints, we computed joint angle trajectories. Rotations between adjacent segments were described by Euler angles, defining flexion, abduction and rotations about the connecting joint. Data for the unsupervised learning procedure included only the flexion angles of the hip, knee, elbow, shoulder and the clavicle, since these showed the most reproducible variation.

2. Blind source separation

To establish a benchmark, we first applied three established methods for the estimation of source signals to our trajectory data: PCA, fast ICA and Bayesian ICA [7] with a positivity constraint for the elements of the mixing matrix. These methods required at least five sources for reconstructions of the original trajectories, in order to explain at least 90% of the variation of the data. We then performed separate ICAs for the individual joints, resulting in separate sets of source variables for each individual joint. By computing the cross-correlation functions between different sources, we found that sources derived from different joints were often astonishingly similar and differed only by additional time delays. This finding motivated us to develop a new source separation algorithm that takes this property of the data into account by explicit modeling of these delays.

Signifying by x_i the ith trajectory and by s_j the jth unknown source signal, the data are modeled by the following non-linear generative model:

$$x_i(t) = \sum_{j=1}^{n} \alpha_{ij} s_j(t - \tau_{ij}). \tag{1}$$

The matrix $\mathbf{A} = (\alpha_{ij})_{ij}$ is called the mixing matrix. A low dimensional non-linearity becomes obvious in the frequency domain

$$\mathscr{F}x_i(\omega) = \sum_{j=1}^{n} \alpha_{ij} \mathrm{e}^{-2\pi \mathrm{i} \tau_{ij}\omega} \mathscr{F}s_j(\omega) = \mathbf{A}(\omega) \cdot \hat{\mathbf{S}}(\omega), \tag{2}$$

where the matrix $\mathbf{A}(\omega)$ is dependent on the frequency variable, and where the vector $\hat{\mathbf{S}}(\omega)$ signifies the Fourier transform of the source signals. $\mathscr{F}$ denotes the Fourier transform.

The model is specified by the linear mixing coefficients α_{ij} and the time delays τ_{ij} between source signals and trajectory components. The problem of blind source separation with time delays has been treated only rarely in the literature (e.g. [4,15,14]). The existing algorithms were not applicable to our problem because they either require positive signals or were not suitable for dimensionality reduction (assuming more sources than signals).

An efficient algorithm for the solution of this blind source separation problem, which scales up to higher-dimensional problems, was obtained by representing the signals in time–frequency domain using the Wigner–Ville transform [10,1], which is defined by

$$Wf(x,\omega) := \int \mathrm{E}\left\{ f\left(x + \frac{t}{2}\right) \overline{f\left(x - \frac{t}{2}\right)} \right\} \mathrm{e}^{-2\pi \mathrm{i}\omega t}\, \mathrm{d}t, \tag{3}$$

where E denotes the expected value. Applying this integral transformation to Eq. (1) one obtains

$$\begin{aligned} Wx_i(\eta,\omega) &= \int \mathrm{E}\left\{ \sum_{j=1}^{n}\sum_{k=1}^{n} \alpha_{ij}\overline{\alpha_{ik}} s_j\left(\eta + \frac{t}{2} - \tau_{ij}\right) \right. \\ &\qquad \left. \times \overline{s_k}\left(\eta - \frac{t}{2} - \tau_{ik}\right) \right\} \mathrm{e}^{-2\pi \mathrm{i}\omega t}\, \mathrm{d}t \\ &\approx \sum_{j}^{n} |\alpha|_{ij}^2 W s_j(\eta - \tau_{ij}, \omega). \end{aligned} \tag{4}$$

The last term is derived exploiting the (approximate) independence of the sources. With the additional assumption that the data coincide with the mean of its distribution ($x_j \approx E(x_j)$) one can compute the first moment of Eq. (4) in η, defined as

$$\begin{aligned} &\int \eta \cdot Wx_i(\eta,\omega)\, \mathrm{d}\eta \\ &= |\mathscr{F}x_i(\omega)|^2 \cdot \frac{\partial}{\partial \omega} \arg\{\mathscr{F}x_i\} \\ &= \sum_{j}^{n} |\alpha|_{ij}^2 \int \eta \cdot W s_j(\eta - \tau_{ij}, \omega)\, \mathrm{d}\eta \\ &= \sum_{j}^{n} |\alpha|_{ij}^2 \cdot |\mathscr{F}s_i|^2 \cdot \left[\frac{\partial}{\partial \omega} \arg\{\mathscr{F}s_j\} + \tau_{ij} \right]. \end{aligned}$$

Analogously, the zero-order moment can be computed, yielding the following two equations:

$$|\mathscr{F}x_i|^2(\omega) = \sum_{j}^{n} |\alpha|_{ij}^2 |\mathscr{F}s_j|^2(\omega), \tag{5}$$

$$\begin{aligned} &|\mathscr{F}x_i(\omega)|^2 \cdot \frac{\partial}{\partial \omega} \arg\{\mathscr{F}x_i\} \\ &= \sum_{j}^{n} |\alpha|_{ij}^2 \cdot |\mathscr{F}s_i|^2 \cdot \left[\frac{\partial}{\partial \omega} \arg\{\mathscr{F}s_j\} + \tau_{ij} \right]. \end{aligned} \tag{6}$$

From these equations the unknowns can be estimated. To recover the unknown sources s_j, mixing coefficients α_{ij} and