Neural Networks 25 (2012) 191–199

Contents lists available at SciVerse ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Real-time stylistic prediction for whole-body human motions

Takamitsu Matsubara [a,b,*], Sang-Ho Hyon [b,c], Jun Morimoto [b]

[a] Graduate School of Information Science, NAIST, 8916-5, Takayama-cho, Ikoma, Nara, 630-0101, Japan
[b] Department of Brain Robot Interface, ATR-CNS, 2-2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan
[c] Department of Robotics, Ritsumeikan University, 1-1-1, Nojihigashi, Kusatsu, Shiga, 525-8577, Japan

ARTICLE INFO

Article history:
Received 13 April 2011
Received in revised form 7 July 2011
Accepted 29 August 2011

Keywords:
Human motion prediction
Multifactor state-space models
Style-content separation
Stylistic prediction

ABSTRACT

The ability to predict human motion is crucial in several contexts such as human tracking by computer vision and the synthesis of human-like computer graphics. Previous work has focused on off-line processes with well-segmented data; however, many applications such as robotics require real-time control with efficient computation. In this paper, we propose a novel approach called *real-time stylistic prediction for whole-body human motions* to satisfy these requirements. This approach uses a novel generative model to represent a whole-body human motion including rhythmic motion (e.g., walking) and discrete motion (e.g., jumping). The generative model is composed of a low-dimensional state (phase) dynamics and a two-factor observation model, allowing it to capture the diversity of motion styles in humans. A real-time adaptation algorithm was derived to estimate both state variables and style parameter of the model from non-stationary unlabeled sequential observations. Moreover, with a simple modification, the algorithm allows real-time adaptation even from incomplete (partial) observations. Based on the estimated state and style, a future motion sequence can be accurately predicted. In our implementation, it takes less than 15ms for both adaptation and prediction at each observation. Our real-time stylistic prediction was evaluated for human walking, running, and jumping behaviors.

1. Introduction

Over the last decade, a considerable number of studies have been conducted on learning the generative models of human motion for modeling, prediction, and recognition (Howe, Leventon, & Freeman, 2000; Li, Wang, & Shum, 2002; Ormoneit, Sidenbladh, Blank & Hastie, 2001; Pavlovic, Rehg & MacCormick, 2000; Sidenbladh, Black & Fleet, 2000; Urtasun, Fleet, & Fua, 2006; Urtasun, Fleet, Hertzmann & Fua, 2005; Wang, Fleet, & Hertzmann, 2008). A significant limitation of these methodologies is that they cannot explicitly consider the natural variations of human motions in the generative model, widely referred to as *style* (Brand & Hertzmann, 2000; Grochow, Martin, Hertzmann & Popovic, 2004; Hsu, Pulli, & Popovic, 2005; Shapiro, Cao, & Faloutsos, 2006; Taylor & Hinton, 2009; Torresani, Hackney & Bregler, 2006; Wang, Fleet, & Hertzmann, 2007). For example, as illustrated in Fig. 1, even for an individual, each walking motion sequence has a distinct walking style. These differences can be much larger between different individuals. Therefore, to achieve highly accurate prediction for a newly observed motion sequence, adaptation of the generative model to the motion sequence by capturing the style of the sequence is necessary.

While most previous studies have focused on off-line processes with well-segmented data, many robotics applications (e.g., human–robot interaction (Onishi, Luo, Odashima, Hirano, Tahara & Mukai, 2007), imitation learning by humanoids (Ijspeert, Nakanishi, & Schaal, 2002; Inamura, Toshima, & Nakamura, 2002; Riley, Ude, Wada, & Atkeson, 2003) and powered suits (Fukuda, Tsuji, Kaneko, & Otsuka, 2003; Kawamoto, Kanbe, & Sankai, 2003)) require real-time control with high accuracy and efficient computation in the prediction procedure.

In this paper, we propose a novel approach called *real-time stylistic prediction for whole-body human motions*. Unlike previous studies (Brand & Hertzmann, 2000; Taylor & Hinton, 2009; Wang et al., 2007), as illustrated in Fig. 2, in our approach the generative model adapts to a newly observed motion sequence by estimating its style by a real-time process. Being able to perform this process in real-time is based on (1) the simple structure of the generative model and (2) the adaptation algorithm which requires small computational effort. We propose a generative model for whole-body human motion that is composed of a low-dimensional state (phase) dynamics and a two-factor (phase dependent observation bases and style parameter) observation model to capture the diversity of motion styles in humans. We also present a learning procedure to acquire the model from a variety of motion sequences

* Corresponding author at: Graduate School of Information Science, NAIST, 8916-5, Takayama-cho, Ikoma, Nara, 630-0101, Japan.
E-mail address: takam-m@is.naist.jp (T. Matsubara).

0893-6080/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2011.08.008

Fig. 1. Illustration of *style* in human motion sequences. Ten walking phase-aligned sequences by two individuals are overwritten in order of phase. The style in walking behavior is considered as a control variable for the spatial variations.

including a diversity of motion styles. A real-time adaptation algorithm was derived using an on-line Expectation-Maximization (EM) algorithm for computationally efficient inference of both the corresponding state variables and the style parameter from non-stationary unlabeled sequential observations. Moreover, with a simple modification, the algorithm allows real-time adaptation even from incomplete (partial) observations. Such applicability of the adaptation algorithm for partial observations is very important in a practical sense because we often meet situations where some elements of the observations are missing due to the limited number of sensors available or occlusions (Chai & Hodgins, 2005). Based on the estimated state and style, the generative model can accurately predict a future motion sequence.

On the other hand, most of the existing models that explicitly estimate the style of motion can achieve neither real-time adaptation nor non-stationary motion estimation since the inference algorithm requires large computational effort (Brand & Hertzmann, 2000; Ormoneit et al., 2001; Sidenbladh et al., 2000; Taylor & Hinton, 2009; Urtasun & Fua, 2004; Wang et al., 2007).

The organization of this paper is as follows. In Section 2, we present a novel generative model to represent a whole-body human motion including rhythmic motion (e.g., walking) and discrete motion (e.g., jumping). We also present a learning procedure to acquire the model from a variety of motion sequences including a diversity of motion styles. In Section 3, a real-time adaptation algorithm is derived by applying an approximated EM algorithm to the generative model. Moreover, we present a simple modification in the adaptation algorithm, which allows real-time adaptation even from incomplete (partial) observations. A real-time prediction method of future motion sequence based on the estimated state and style is also presented. In Section 4, the effectiveness of our real-time stylistic prediction is validated for human walking, running, and jumping behaviors with motion capture data. Section 5 concludes this paper.

2. Learning generative model

This section describes the proposed generative model and a learning procedure for the model with a stylistic data set.

2.1. Generative model for whole-body human motions

We first define the notation of the proposed generative model. $\mathbf{x} \in \mathbb{R}^d$ is the state variable, $\mathbf{y} \in \mathbb{R}^D$ is the observation and the probability distribution $p(\mathbf{y}_t|\mathbf{x}_t; \mathbf{w})$ is the observation model. $p(\mathbf{x}_{t+1}|\mathbf{x}_t)$ is the state-transition probability distribution. The parameter vector $\mathbf{w} \in \mathbb{R}^J$ is an additional latent variable that controls the spatial variation of observations. We call this the style parameter. Its graphical model is depicted in Fig. 3. For periodic and discrete motions, we explicitly define the state variable $\mathbf{x}$ as

$$\mathbf{x}_t = [\phi_t \ \omega_t]^T = \begin{cases} [\psi_t \ \dot{\psi}_t]^T & \text{(Rhythmic)} \\ [p_t \ \dot{p}_t]^T & \text{(Discrete).} \end{cases} \tag{1}$$

That is, we define the state variable $\mathbf{x}$ by phase ϕ as a point on a one dimensional sphere in two dimensional Euclidean space $\phi \equiv \psi \in \mathbb{S} \subset \mathbb{R}^2$ and its velocity $\omega \equiv \dot{\psi}$ to represent its periodicity of rhythmic motions, similar to Ormoneit et al. (2001) and Urtasun and Fua (2004). We also define the phase ϕ as a point on a one dimensional closed line segment $\phi \equiv p \in \mathbb{L}$ for discrete motions to represent its non-periodicity (discreteness). The explicit use of these assumptions in the generative model yields the low-dimensional state variable $\mathbf{x}$. Moreover, as presented in the next section, it allows a simple learning algorithm for the generative model from data.

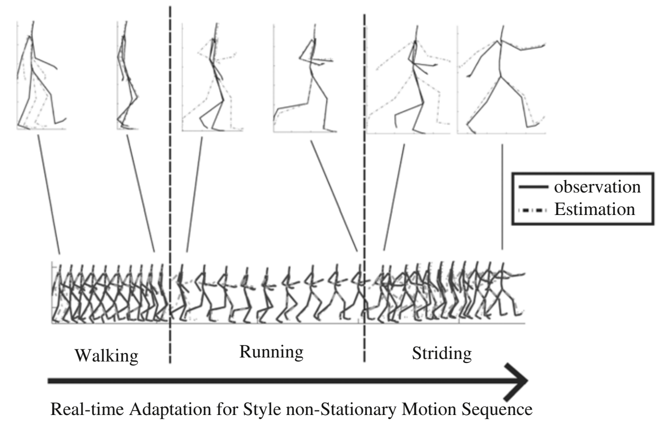

Fig. 2. Illustration of the real-time adaptation and prediction of the generative model for a non-stationary motion sequence with styles (walking behaviors). The test sequence consists of three motions, walking, running and striding generated by different individuals. The solid human figure is as observed and the dashed one is the predicted motion as a result of adaptation of the generative model to the observation sequence. The adaptation is achieved by on-line EM incrementally with little computation at each observation. For all motions, the model is rapidly adapted to the style of the recent test sequence since the time-forgetting factor effectively forgets past observations.

Based on the above assumptions, we conclude that the state-transition model and the observation model are modeled by Gaussian distributions as:

$$p(\mathbf{x}_{t+1}|\mathbf{x}_t) = \mathcal{N}(\mu_x(\mathbf{x}_t), \Sigma_x(\mathbf{x}_t)), \tag{2}$$

$$p(\mathbf{y}_t|\mathbf{z}_t; \mathbf{w}) = \mathcal{N}(\mu_y(\mathbf{z}_t; \mathbf{w}), \Sigma_y(\mathbf{z}_t; \mathbf{w})), \tag{3}$$

where

$$\mathbf{z}_t = g(\mathbf{x}_t) = \begin{cases} [\cos(\phi_t) \ \sin(\phi_t)]^T & \text{(Rhythmic)} \\ \phi_t & \text{(Discrete)} \end{cases} \tag{4}$$

and the observation model is defined as a probabilistic mapping from a phase ϕ_t (as $\mathbf{z}_t$) to an observation $\mathbf{y}_t$. The velocity of phase ω_t governs the temporal variation of the time-series, that is, it controls the velocity of human motions generated by the model. For rhythmic motions, $\mathbf{z}_t = g(\phi_t) \in \mathbb{R}^2$ represents a point on a manifold $\mathbb{S}$ in $\mathbb{R}^2$ where the radius is $r = 1$ and the angle is ϕ_t. This state representation allows us to approximately measure the geodesic distance between points on $\mathbb{S}$ as the Euclidean distance in $\mathbb{R}^2$. For discrete motions, $\mathbf{z}_t$ is the equivalent of ϕ_t.[1]

2.2. Learning procedure with a stylistic data set

The learning procedure assumes we have multiple human motion sequences including a diversity of motion styles. Let $\mathbf{Y}^s = [\mathbf{y}_1^s \cdots \mathbf{y}_{C(s)}^s]^T \in \mathbb{R}^{C(s)\times D}$ denote a time-invariant motion sequence with a distinct style, where $s \in \{1, 2, \dots, S\}$ is the style index in which each value indicates a corresponding distinct style, $c \in \{1, 2, \dots, C(s)\}$ is the content index that corresponds to the phase

[1] For the case of a discrete motion, $\mathbf{z}_t$ is a scalar; however, it is kept as a vector notation for simplicity of the overall description.