Modeling Human Motion and Contact with Gaussian Processes

Olivier Rémillard* Paul Kry†

School of Computer Science, McGill University

Abstract

We introduce a simple approach for using Gaussian processes to model human motion involving contact. It comprises a low dimension latent space with dynamics augmented by switching variable. A Gaussian process models the time relationship of a motion while the switching variable helps model the discontinuities created by interaction with the environment.

1 Introduction

Generating realistic and lifelike animated characters from captured motion sequences is a hard and time-consuming task. The task is challenging due to the high dimensionality of human pose data and the complexity of the motion. Gaussian processes (GP) are useful for modeling the dynamics of human movement when combined with a latent variable model to approximate the lower dimensional manifold of human motion. These models alleviate the difficult problem of explicitly modeling physics and control, while providing a means of predicting behaviour, with applications in tracking and motion capture reuse.

The GP model will give good predictions if the latent trajectories are smooth. However, in many cases, the latent trajectories are not smooth because they include abrupt changes when the actor's body motion suddenly changes. It occurs often when a knee or an elbow joint reaches its full extent and locks or again when the actor interacts with the environment. These interactions can be as simple as stepping on the floor, pushing or pulling an object. Our solution borrows ideas from Switching Gaussian Process Dynamical Model (SGPDM) [1]. However, instead of using the switching variable to separate different motions, say walking and running, we use the switching variable to separate the nonsmooth dynamics acting on the motion into distinct sets. This separation further permits the use of simpler techniques such as principal component analysis (PCA) to reduce the dimension of the original problem.

2 Approach

We represent 3D human motion as a sequence of joints angles to describe how the pose changes over time, thus, a pose can be packaged in a vector, and a motion in a matrix. Assuming a first order Markov dynamic, modeling human motion is the task of computing

$$p(q_{t+1}|q_t),$$

where q_{t+1} denotes the next pose following the current pose, q_t. However, due to the high dimensionality of a pose, over 60 dimensions, most modeling techniques yield poor results. It is preferred to reduce the size of the input space.

We reduce the dimension of the space with a linear mapping

$$z_t = f(q_t),$$

where q is the pose and z its low dimensional latent coordinates. Since f is a linear transformation, we can chose f such that its inverse f^{-1} exists and use it to transform latent coordinate back to poses

$$q'_t = f^{-1}(z_t).$$

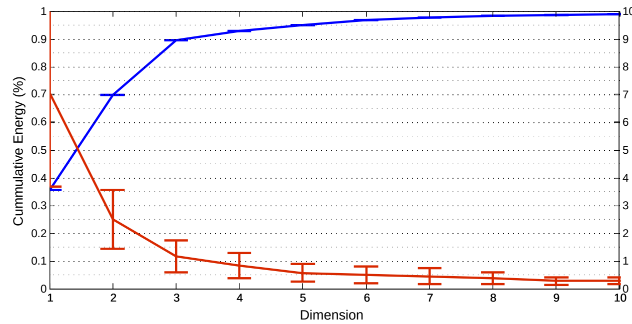

Figure 1: Effect of PCA Reduction on foot location over a walking sequence of 125 poses. The blue curve depicts the cumulative energy for a specific dimension. The red curve shows the mean error on the foot location of the reconstructed sequence in comparison to the original sequence. The error bars show 2 standard deviations.

The result, q'_t, is not equal to the original pose because f is a projection and the inverse does not reconstruct the original full space.

Figure 1 illustrates the trade off between the latent space dimension and the resulting reconstruction error at an end effectors of the body. Because of the hierarchical description of joints and angles of a pose, the error is accumulated as a body part is deeper in the hierarchy. Consequently, the error is greater at the feet and hands. As such, we use discrepancies of the foot location in q_t and q'_t as an indicator of the overall quality of the mapping. We choose to use the three first principal component of the observed pose data as transformation f because this captures 90% of the variation in the original motion (see Figure 1).

With dimension reduction, the model simplifies to

$$p(z_{t+1}|z_t).$$

We will model the time dependence between two consecutive poses with a GP. This is a non-parametric approach for solving regression problem.

Given a training motion sequence $Q \in R^{NxD}$, and its latent sequence $f(Q) = Z \in R^{Nxd}$, we can model the dynamics with

$$p(Z_+|Z_-,\theta) = \frac{\exp\left(-\frac{1}{2}\mathrm{trace}(Z_-^T K^{-1} Z_-)\right)}{\sqrt{(2\pi)^{(N-1)d}|K|^d}},$$

where K is the process kernel (covariance of the inputs), θ the kernel's hyper parameters, and $Z_+ = [z_2,...,z_N]^T$ is the vector of states that follow $Z_- = [z_1,...,z_{N-1}]^T$. The GP is maximized via

$$\theta = \arg\max_\theta p(Z_+|Z_-,\theta)$$

by optimizing the log likelihood of $p(Z_+|Z_-,\theta)$ using scaled conjugate gradient methods. Once trained we obtain

$$\begin{aligned} p(z_{t+1}|z_t) &= N(\mu(z_t),\sigma^2(z_t)),\\ \mu(z) &= Z_+^T K^{-1} k(z,Z_-),\\ \sigma^2(z) &= k(z,z) \; - \; k(z,Z_-)^T K^{-1} k(z,Z_-). \end{aligned}$$

*e-mail: remillardo@msn.com
†e-mail: kry@cs.mcgill.ca

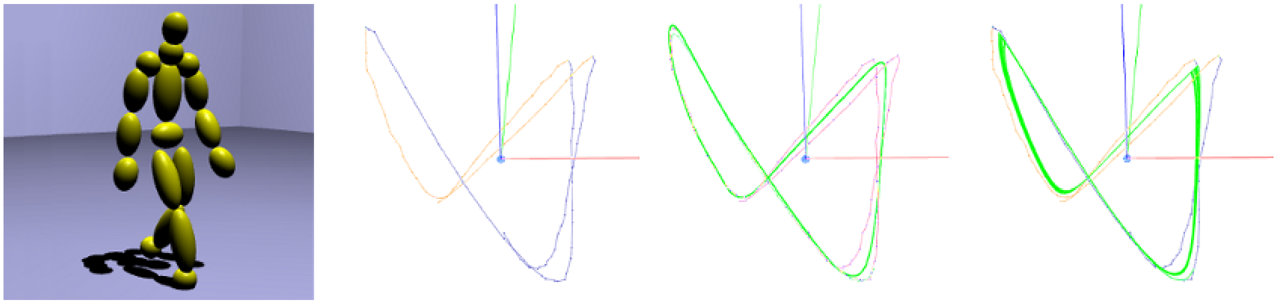

Figure 2: Walk. from left to right: a pose, the trajectory in latent space, inference of the model with no switching, inference with switching. The sequence consists of one and half cycles of walking for a total of 125 poses. The sequence is separated in 2 states S:contact with left foot (blue) and contact with the right foot(yellow).

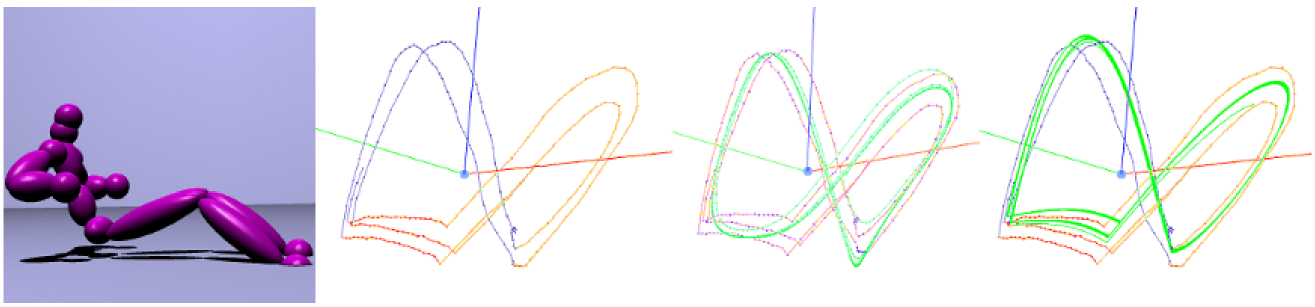

Figure 3: Rowing machine. from left to right: a pose, the trajectory in latent space, inference of the model with no switching, inference with switching. The sequence consists of two and half cycles of paddling for a total of more than 200 poses. The sequence is divided into 3 states S: pushing on the oars (yellow), pulling on the oars (blue) and pause between pushing and pulling (red).

where $k(z, Z_-)$ is the covariance function applied to the input z and the training set inputs Z_-. Details of each step are found in [2, 3].

The linear transformation and the Gaussian Process described so far are sufficient to model some motions. Given an initial pose q_0, we seek q_t the tth pose following q_0. We start with $z_0 = f(q_0)$ and iteratively use the mean of the Gaussian process to obtain q_t by

$$q_t = f^{-1}(z_t),$$
$$z_t = \mu(z_{t-1}).$$

2.1 Switching Models at Contact

The problem with the usual formulation of the GP model is that it tends to smooth sharp turns in the trajectory. These discontinuities are characteristic to the contact forces acting on the human in motion, and those contacts should also be correctly modeled.

To cope with this predicament, we will add to the model a switching variable $s \in S$ to divide the motion into smaller subsets. Each variable should describe a situation where specific contact forces are in action. For example, in the walking situation we could have 4 values, $S = \{\text{no contact, left foot, right foot, both}\}$.

The switching variable permits the decomposition of $p(z_{t+1}|z_t)$ along the values of S, and the training of $|S|$ individual GP models [1]. Besides, a mapping from latent space coordinates to switching values can be expressed as a GP classification problem as explained in [2].

We can infer in this model the same way we did with the previous formulation. The main difference being the use of more than one Gaussian process. When stepping in the latent space $z_t = \mu(z_{t-1})$, we should only use the mean function of the GP trained along the switching value of z_{t-1}.

3 Results

Figures 2 and 3 show two examples where model switching is of benefit (a walking motion, and paddling motion on a rowing machine). The figures show the inference of 1000 poses using a model without the switching variable, and 1000 poses using the switching variable. The differences of both models reside in the ability to model the discontinuities of the trajectory. The main consequence of smoothing the discontinuities for the walking sequence is the accentuation of foot skating. For the paddling sequence, this consequence is reflected in the elimination of the pause (red) between the pushing and pulling movements.

4 Conclusion

We presented a way to model human motion and contact using Gaussian processes and a switching variable. The models we produce are useful for generating arbitrary length sequences of cyclic motion that can be adjusted to fit specific situations, and we have the added benefit that the switching variable helps model discontinuities due to contacts. One limitation is that we must label the switches in the training data. As future work, it would be interesting to use an unsupervised framework to choose labels that optimize the fit of the model.

References

[1] J. Chen, M. Kim, Y. Wang, and Q. Ji. Switching Gaussian process dynamic models for simultaneous composite motion tracking and recognition. In *Computer Vision and Pattern Recognition*, pages 2655–2662, 2009.

[2] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[3] J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models for human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:283–298, 2008.