

Modeling and Learning Contact Dynamics in Human Motion

Alessandro Bissacco and Stefano Soatto

UCLA CSD TR040048

Abstract

We propose a simple model of human motion as a switching dynamical system where the switches correspond to contact forces with the ground. This significantly improves the modeling performance when compared to simpler linear systems, with only marginal increase in complexity. We introduce a novel closed-form (non-iterative) algorithm to estimate the switches and learn the model parameters in between switches. We validate our model qualitatively by simulation, and quantitatively by computing prediction errors that show significant improvements over previous approaches using linear models.

1. Introduction

Detection, classification and recognition of human activity has become a topic of strategic importance in recent years, and the computer vision community is contributing significantly to develop technology to automate the process. “Activity” in this context indicates a distribution of statistics that can be inferred from a time series; in other words, the “information” resides in the temporal characteristics of the data. As in all inference problems, the choice of underlying model for the phenomenon of interest plays a crucial role in human motion analysis. In exercising the art of modeling, one can choose complex models that capture the global temporal statistics of the data, or choose a simple class of models and infer the local statistics, together with the “segments” of data where such a model is a good fit. For instance, each segment could represent a portion of the data that exhibit some form of (statistical) stationarity. The choice of model ultimately has to balance parsimony in the representation with fidelity in the inference task, in our case classification.

We are interested in developing local models of human motion that can support classification tasks: We restrict our attention to stationary segments, called *gaits*, and we aim at our models being *discriminative* in the sense of supporting statistical decision tasks. However, faithful reproduction of the statistics of the data is always a good measure of modeling performance, so we will use *synthesis* to validate our models using prediction-error criteria. We want our models to capture the underlying *dynamics* (high-order temporal statistics) of the data, in addition to the kinematics of the human body. Eventually, spatial and temporal models will have to be integrated into a higher-level discrete-event model to determine the segments and the transition between them.

While much of the attention of the human motion community has been on such higher-level (“activity”) models as of late, and some authors have suggested that low-level modeling of gaits is easily solved, we maintain that there are some fundamental issues that have not been adequately addressed, yielding overall models that do not support fine classification or recognition. Chief among these is the modeling of external forces, that are key in determining the dynamic properties of human motion. *We present the simplest possible statistical model that captures contact forces and, to the best of our knowledge, the only one so far.*

The reader should not let the word “forces” induce her to believe that we use physical models. Such models, often used in computer graphics, are too complex for inference (inverse dynamics), although their generative power is naturally superior. Our approach is purely statistical, as we aim at capturing the phenomenology of contact forces to the extent in which it affects the discriminative power of the model. In our models we emphasize the need to capture true *dynamics* (not just kinematics), as well as to make use of the wide body of training data available in the form of motion capture sequences. While some readers may object to that on the grounds that the actual “data” are images, not position of marker points, contact forces manifest themselves as discontinuities in the dynamics at every level of the modeling hierarchy, all the way to the images. Therefore, in order to not dilute our message, we use the simplest possible scenario to illustrate it, hence *we operate directly on motion capture data.*

1.1. State of the art and contributions

The literature on modeling human motion is too broad for us to review here. However, we provide a brief synopsis of alternative approaches that can guide the reader to place our contribution into context.

Computer graphics and physical simulations: generative approaches to realistic human motion synthesis include designing controllers driving physical simulations of the body dynamics [15, 12] and constrained optimization [32, 26]. In simulations contact forces are modeled with stiff springs, knowing the ground. This solution is not appropriate for vision applications, because these models cannot be efficiently “learned” (they contain parameters that cannot be identified from data, hence are not “observable”). Alternative methods are procedural models, that produce synthesis by concatenation and interpolation of measured data [17, 1]. This is not suitable for detection and recognition tasks.

Robotics literature: most of the work focus on physical models of biped mechanisms [31, 24, 9]. These models are simpler than human body because of rigidity of the links, fewer joint degrees of freedoms, and simplification on ground contact modeling. There, hybrid models have been in use for some time [7], although with an emphasis on control and stability. The model is identified separately in controlled experiments where all the external forces are known. Unfortunately in the case of the human body we do not have access to the mechanism to learn the model. Moreover we want to be able to handle variability due to different individuals and different gaits.

Vision literature: simple statistical models of stationary gaits have been proposed by several authors, for instance in the form of linear dynamical systems [3, 5, 18]. While these models capture the temporal statistics in between contacts with the ground, and their generative capability is remarkable considering their simplicity, they fail the perceptual test miserably at or around contacts (Figure 1).

We propose the next simplest thing, that is a *switching model where contact is represented as switch between two linear models, each of which models an inter-contact segment*.

Switching linear models have been used in the graphics literature to model and synthesize human motion [18], where each linear model represents an “action” or “movement” and switches represent the transition between actions. In such models the emphasis was synthesis, that was performed semi-procedurally by interpolating (through a controller) between data segments. Switching systems have also been used for tracking and classification [5].

Inference of switching linear systems is a topic that has received considerable attention in the learning and system identification communities during the past few years, but all the algorithms proposed there either provide iterative approximated solutions or, when non-iterative [30], they only cover the auto-regressive case, which is not general enough to capture our model. Most attention in the identification literature has been devoted to Jump Markov Linear Systems, hybrid systems where the discrete state evolution is modeled by a Markov chain. For these models the common approach to the learning problem is the Expectation Maximization (E-M) algorithm, where the Expectation step amounts to computing the posterior probability of the states given measurements and parameters. A variety of solutions have been proposed for this NP-hard filtering problem, including approximating distributions [14, 4, 28, 21, 23], E-M [19], particle filters [22, 16] and other Monte Carlo sampling techniques [2, 11, 8]. We need a more efficient inference procedure (closed form, non-iterative), although we can afford to deviate from optimality, as long as our models have discriminative power, which can be measured in simulation.

Therefore, we propose *an efficient (non-iterative) learning algorithm to detect the switches and identify the model parameters*.

Our contribution is 3-fold: (a) we propose a very simple switching linear system to model human gaits (walking, running, hopping etc.); (b) we introduce a novel, closed-form algorithm to estimate the switches, as well as identify the models in between switches, (c) we validate our models empirically by generating simulations with the model and showing that humans cannot distinguish them from the original data, obtained with a motion capture system.

One limitation of our work is that we model motion capture data, which is relatively “clean.” In practice one must go from images to a representation that is suitable to represent dynamics. These can be joint positions or angles if an explicit skeletal model is present, or an articulated region-based model otherwise. For this, however, there are many valid approaches, and we refer the reader to the literature [25, 27, 6]. We concentrate on a specific, narrow but very important aspect of modeling the dynamics of human motion, namely external contact forces.

What we do not do is to test our model for classification. We are confident that this will significantly improve the discriminative power over simple linear system, but a thorough experimental test is well beyond the scope of this paper and will be presented separately. Instead, we perform simulation to validate the phenomenology of the model qualitatively (see uploaded movies) and quantitatively using prediction-error.

2. A simple model of human motion

The motion of biped organisms such as humans involves great dynamical complexity, even if it is idealized to a system of rigid levers with torques acting at the joints. If we want to use physical models to study the effect of contact forces on gait motions it is therefore necessary to consider simplified systems. The model must be simple enough for analytic treatment, yet able to produce motion that is perceptually close to a human gait.

Models with these properties have been proposed in the robotics literature. Passive walkers [20, 10, 13] are a special kind of unactuated biped robots that can be made to walk down a slight slope only by the effect of gravity and produce naturally looking motion. The simplest of these models is the one proposed in [13]. It is a 2D two-link model, with rigid massless legs hinged at the hip and infinitesimal point masses at the feet.

The general equations of one-step motion for these robots are:

$$\begin{cases} \dot{x}(t) = f(x(t)), & x = \begin{bmatrix} q^T & \dot{q}^T \end{bmatrix}^T \in \mathbb{R}^n, & x(0) = x_0 \\ y(t) = C_{\lambda(t)}x(t), & y \in \mathbb{R}^m, \lambda \in \{1, 2\} \end{cases} \quad (1)$$

where $q \in \mathbb{R}^m$ is the vector of generalized coordinates (for the simple walker the two angles between the legs and the vertical) and $C_1, C_2 \in \mathbb{R}^{m \times n}$ are permutations of the first m rows of the $n \times n$ identity matrix. Basically $\lambda(t)$ select at each step which ones are the swing and stance legs. The attentive reader will notice that if we exclude singular cases equation (1) is just a rewriting of the familiar equation of motion for rigid bodies $M(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) = 0$.

If we linearize (1) around an operating point x_0 and discretize with sampling time T_s , we obtain the following linear system:

$$\begin{cases} x_{k+1} = Ax_k + B, & x_k \in \mathbb{R}^n \\ y_k = Cx_k, & y_k \in \mathbb{R}^m \end{cases} \quad (2)$$

$$A = e^{J_0 T_s}, \quad J_0 = \frac{\partial f}{\partial x}(x_0)$$

$$B = (e^{J_0 T_s} - I)J_0^{-1}f(x_0) - e^{J_0 T_s}x_0 + x_0$$

A system of the form (2) can be used to approximate the evolution of the original nonlinear system (1), and experiments show that in the case of the simple walker the quality of the approximation is remarkably good (Figure 1).

The system evolves according to (1) until landing time τ , when the swing leg touches the ground. At τ a switch occurs: the role of swing and stance leg are swapped, and there is a discontinuity in the velocity. The state is mapped according to a linear function g :

$$x(\tau^+) = g(x(\tau^-)) \quad , \quad \lambda(\tau^+) = \{1, 2\} \setminus \lambda(\tau^-) \quad (3)$$

Equations (1) and (3) describe the dynamics of a class of mechanisms more general than the simple passive walker considered here. The walking motion of any legged robot with point foot-ground contacts, such as [24, 9], can be written in this way. The approximating linearization of the hybrid system (1), (3) is the switching model:

$$\begin{cases} x_{k+1} = A_1x_k + B_1, & x_0 = x_{0,1} & k < \tau - 1 \\ y_k = C_1x_k \\ x_{k+1} = A_2x_k + B_2, & x_\tau = x_{0,2} & k \geq \tau \\ y_k = C_2x_k \end{cases}$$

Based on these considerations, we propose a novel switching system for representing walking gaits. The model is composed of two linear systems (2) representing the alternating stance and swing role assumed by the legs, the initial states of the systems and a condition for the model switch. Distinctive properties of such hybrid models are the possibility to have unstable systems as components and the ability to produce periodic motion without the need of constraining the poles of the systems to the unit circle.

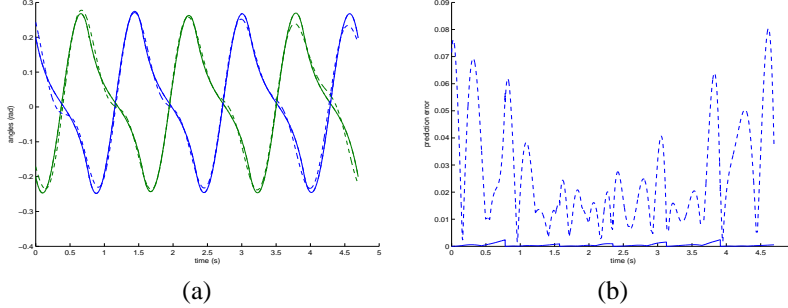


Figure 1: **Linear models fail to capture perceptually important details:** (a) The joint trajectories of synthetic data used to learn a linear model (solid) superimposed with the trajectory simulated from the model (dashed). Although the coarse structure of the data is captured, the small deviations around the contacts causes the simulation to fail the perceptual test, as the simulated walk appears to penetrate the ground and slip (see uploaded movie `walk-linear.mov`). The switching linear model captures these details quite faithfully: (b) shows the prediction error residual for the linear model (dashed) and for the switching linear model (solid).

3. Learning

As anticipated in the previous sections, exact inference of a general linear switching system of the form (2) is NP-complete, and all approximated solutions proposed in the literature are iterative.

In this work we present a closed-form suboptimal identification algorithm. We separate the learning process in two independent steps: detection of the switch instants and estimation of the parameters of the two linear systems from the segmented data.

For switch detection we propose a novel algorithm based on subspace identification concepts [29]. The main idea can be explained in four steps:

- (a) For linear systems, the hypothesis is that future data are a linear function of past data. This hypothesis is violated at switches.
- (b) We can therefore pose the problem of switch detection as a hypothesis testing problem.
- (c) However, we have to do so without knowing the model parameters. We will implicitly eliminate the model parameters by linear projections, which yields a simple thresholding solution to detect the switches.
- (d) Once switches are detected, the data can be arranged so that the linear assumption is satisfied, at which point the problem becomes standard identification which we solve with established methods.

3.1. Switch Detection

For ease of notation and in order to simplify the derivations, in what follows we consider systems which do not include an explicit noise model. We will tacitly assume that the noise is Gaussian and is, in L_2 norm, significantly smaller than the state switch. An explicit noise model will be introduced later in the identification step.

Consider a system of the form (2) where the pair (A, C) is observable. Let us construct the following block Hankel matrices:

$$\begin{aligned}
Y_p &= \begin{bmatrix} y_0 & y_1 & y_2 & \vdots & y_{j-1} \\ y_1 & y_2 & y_3 & \vdots & y_j \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{i-1} & y_i & y_{i+1} & \vdots & y_{i+j-2} \end{bmatrix} \\
Y_f &= \begin{bmatrix} y_i & y_{i+1} & y_{i+2} & \vdots & y_{i+j-1} \\ y_{i+1} & y_{i+2} & y_{i+3} & \vdots & y_{i+j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{2i-1} & y_{2i} & y_{2i+1} & \vdots & y_{2i+j-2} \end{bmatrix} \\
X_p &= [x_0 \ x_1 \ x_2 \ \cdots \ x_{j-1}] \\
X_f &= [x_i \ x_{i+1} \ x_{i+2} \ \cdots \ x_{i+j-1}] \\
\Gamma_i &= \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{i-1} \end{bmatrix}, \quad J_i = \begin{bmatrix} 0 \\ CB \\ CB + CAB \\ \vdots \\ \sum_{k=0}^{i-2} CA^k B \end{bmatrix} \\
H_i &= \sum_{k=0}^{i-1} A^k B
\end{aligned} \tag{4}$$

where j is large (ideally $j \rightarrow \infty$) and i is not smaller than the observability index, i.e. $\text{rank}(\Gamma_i) = n$. The system (2) can be rewritten as:

$$\begin{aligned}
Y_f &= \Gamma_i X_f + J_i \mathbf{1}, \quad Y_p = \Gamma_i X_p + J_i \mathbf{1} \\
X_f &= A^i X_p + H_i \mathbf{1}
\end{aligned} \tag{5}$$

where $\mathbf{1} = [1 \ 1 \ 1 \ \cdots \ 1]$ is a row vector of j elements. From subspace identification, we know that we can express the future outputs Y_f as a linear combination of past outputs Y_p and constant inputs:

$$\begin{aligned}
Y_f &= \Gamma_i X_f + J_i \mathbf{1} = \Gamma_i A^i X_p + \Gamma_i H_i \mathbf{1} + J_i \mathbf{1} = \\
&= \Gamma_i A^i \Gamma_i^\dagger Y_p - \Gamma_i A^i \Gamma_i^\dagger J_i \mathbf{1} + \Gamma_i H_i \mathbf{1} + J_i \mathbf{1} = \\
&= M W_p \\
M &= [\Gamma_i A^i \Gamma_i^\dagger \quad \Gamma_i H_i - \Gamma_i A^i \Gamma_i^\dagger J_i + J_i], \\
W_p &= \begin{bmatrix} Y_p \\ \mathbf{1} \end{bmatrix}
\end{aligned} \tag{6}$$

Consider the model (2) with a switch occurring at time τ . This hybrid system is equivalent to a single linear system with a state jump at τ :

$$\begin{aligned}
x_{k+1}^s &= A x_k^s + B + S \delta(k+1-\tau) \\
y_k^s &= C x_k^s \\
A &= \begin{bmatrix} A_1 & 0 & -B_1 \\ 0 & A_2 & B_2 \\ 0 & 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} B_1 \\ 0 \\ 0 \end{bmatrix} \\
C &= [C_1 \ C_2 \ 0]
\end{aligned} \tag{7}$$

The output matrices Y_f^s, Y_p^s of system (7) are related to the matrices Y_f, Y_p of the system (2) by:

$$\begin{aligned}
Y_f^s &= Y_f + \Delta_f & Y_p^s &= Y_p + \Delta_p \\
\Delta_f &= \begin{bmatrix} 0 & \vdots & 0 & 0 & \vdots & | & | & \vdots & | & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \Gamma_i S & \Gamma_i A S & \vdots & \Gamma_i A^{i-1} S \\ 0 & \vdots & 0 & \Gamma_2 S & \vdots & | & | & \vdots & | & \vdots \\ 0 & \vdots & \Gamma_1 S & | & \vdots & | & | & \vdots & | & \vdots \end{bmatrix} \\
\Delta_p &= \begin{bmatrix} 0 & \vdots & 0 & 0 & \vdots & 0 & 0 & \vdots & 0 & \vdots \\ 0 & \vdots & 0 & 0 & \vdots & 0 & 0 & \vdots & | & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \Gamma_{i-1} S \\ 0 & \vdots & 0 & 0 & \vdots & 0 & \Gamma_1 S & \vdots & | & \vdots \end{bmatrix}
\end{aligned}$$

Let $E = Y_f^s - MW_p^s$. We have:

$$\begin{aligned}
E &= \begin{bmatrix} 0 & \vdots & 0 & 0 & \vdots & | & | & \vdots & | & 0 & \vdots \\ 0 & \vdots & 0 & 0 & \vdots & | & | & \vdots & | & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \Gamma_i S & G_1 S & \vdots & G_{i-1} S & 0 & \vdots \\ 0 & \vdots & 0 & \Gamma_2 S & \vdots & | & | & \vdots & | & 0 & \vdots \\ 0 & \vdots & \Gamma_1 S & | & \vdots & | & | & \vdots & | & 0 & \vdots \end{bmatrix} \\
G_j &= \Gamma_i A^j (I_n - \Gamma_j^\dagger \Gamma_j)
\end{aligned}$$

E has $2i - 1$ nonzero columns from $\tau - 2i + 1$ to $\tau - 1$. Their norm increases from $\tau - 2i + 1$ to $\tau - i$. If we assume A stable and exclude pathological cases, the norm of $G_j S$ decreases with increasing j , and is null when j reaches the observability index. Therefore the norm of the columns of E has an impulsive character, with peak at $\tau - i$ and impulse width smaller than $2i$.

Let \hat{E} be the projection of the rows of Y_f^s onto the space orthogonal to the rows of W_p^s :

$$\hat{E} = Y_f^s / W_p^{s\perp} \triangleq Y_f^s - Y_f Y_p^\dagger Y_p \quad (8)$$

We have:

$$\hat{E} = Y_f^s - \hat{M}W_p^s = (E + MW_p^s) / W_p^{s\perp} = E / W_p^{s\perp} \quad (9)$$

Since a basis of W_p^s is given by the modes of the system (2) which are exponentially damped oscillations, the projection of an impulse at $\tau - i$ onto the space orthogonal to W_p^s will still have an impulsive form with the same peak location. Superimposed to the impulse there will be small oscillations whose magnitude is proportional to the fraction of energy of the modes concentrated in $\tau - i$.

Based on the previous considerations we propose the following state switch detection algorithm:

- Construct the matrices Y_f and Y_p from the output $y_k, t = 0 \dots T$, with an appropriate choice of i .
- Compute \hat{E} as in (8).
- Apply a threshold on the norm of the columns of \hat{E} . Each peak t identifies a state jump occurring at time $t + i$.

The choice of i is important for the success for the algorithm. If i is smaller than the observability index then all the columns of \hat{E} will be nonzero and the proposed algorithm will fail. If i is large, the width of the impulse on the columns of \hat{E} will be large and \hat{E} may not have clear peaks.

3.2. Identification of the linear systems

Once we have segmented the input by detecting the switch instants, we identify the parameters of the two linear systems that model the dynamics of the walking gait. The segments are grouped in two classes, one for each system, with every other segment belonging to the same class.

The identification problem is, given a number of segments $y(t), \tau_i \leq t < \tau_{i+1}$ generated by the linear system M_λ :

$$\begin{cases} x_{k+1} &= A_\lambda x_k + B_\lambda + K_\lambda e(t) \\ y_k &= C_\lambda x_k + e(t) \end{cases}, \quad e \in \mathcal{N}(0, R_\lambda) \quad (10)$$

estimate the values of the parameters $A_\lambda, B_\lambda, C_\lambda, K_\lambda, R_\lambda$ and the initial conditions x_{τ_i} that maximize the likelihood of the errors $e(t)$. A closed-form solution to this problem is provided by subspace identification techniques. The main idea of these algorithms, used also in our switch detection method, is to project future outputs onto the space spanned by past inputs and past outputs. These algorithms are asymptotically optimal, in the sense that they give unbiased estimates of the maximum likelihood values of the system parameters as the number of observations goes to infinity. Unfortunately the segments that we extract from motion data are relatively short, typically less than 100 frames. Therefore, in order to refine the quality of the estimates, the subspace identification procedure is followed by a standard iterative optimization algorithm based on prediction error minimization. The algorithm minimizes:

$$V_N(\theta) = \frac{1}{T} \sum_{k=1}^T y_k - \hat{y}_{k|k-1} \quad (11)$$

where $\hat{y}_{k|k-1}$ is the prediction of the output at time k , given information up to time $k - 1$.

4. Experimental validation

We applied the proposed modeling and learning framework to the identification of switching models from motion capture data. Our goal is to learn models that can capture the details of human gait motions and that can be used for tracking, detection, recognition and synthesis.

In this work we focus on experimentally validating the modeling power of the proposed systems by using them to synthesize new motion and comparing the results with the original data. In future work we plan to investigate the possibility of applying these models to the problem of recognition and classification of human gaits.

As first preliminary experiment, we learned a switching model from synthetic motion data of the simple passive walker proposed in [13]. Figure 1 shows the joint angle data and the prediction errors for switching and linear models. We see how the switching system models with high accuracy the signal, while the linear system deviates sensibly from the original trajectories. In Figure 2 we show the norm of the residual of the switch detection algorithm. The peaks have conspicuous impulsive character and are located exactly i steps before the actual switch, where i is the parameter of the switch detection algorithm and defines the number of measurements stacked in the observation matrices (4).

In the experiments with real data we used the motion database freely distributed by the Carnegie Mellon Graphics group. The data consists of short clips of walking and running gaits, collected at 60 Hz with the VICON optical motion capture systems. The motion is represented as joint angle trajectories of a skeletal model of the performer.

For each frame, the pose is specified by position and orientation of the body reference frame and 56 angles describing the configuration of the 29 joints of the skeletal model.

As first preprocessing step, we removed the linear trend in the coordinates by subtracting from the (x, y, z) position of the body reference frame the least squares linear fit.

Then we selected a number of joint angles to detect the ground contact instants. The more informative angles are hip and knee joint rotation in the sagittal plane. We did not use the angle of the feet for switch detection because these measurements appear to be very noisy in the available sequences, in particular for the swinging foot. Figure 3 displays the evolution of the leg joint angles for a sample gait together with the residual of the switch detection algorithm. The switches are detected as the maxima on windows of 11 frames that are larger than a predefined threshold.

Once switches are detected, the data is segmented and the two linear systems M_λ , $\lambda \in \{1, 2\}$ of the form (10) are identified.

Because of the short duration of the segments, identification of the system is possible only after dimensionality reduction. For this purpose we applied PCA, and experimentally determined that 8 components are sufficient to reproduce motion that

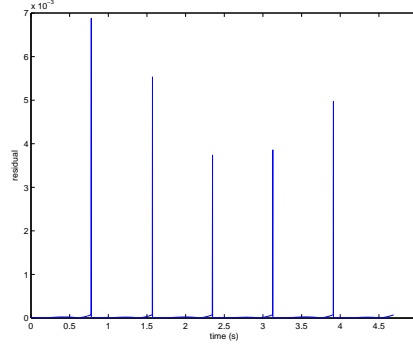


Figure 2: **Switch detection on synthetic data:** Results of the switch detection algorithm applied on simple walker data of figure 1. Plotted the switch detection residuals (norm of columns of \hat{E} , $i = 3$). Switches are very prominent and easy to detect by simple thresholding. Their peak is exactly i steps before the actual switch instants.

is perceptually indistinguishable from the original. From the reduced data, we learned the linear system M_λ as described in section (3.2), using the implementations of the identification algorithms provided in the Matlab System Identification Toolbox. In the all the simulations with motion capture data, we used models with dimension of the state $n = 12$.

Together with the model parameters, we estimated the initial state $x_{0,\lambda}^i$ for each segment i , mean μ_λ and variance σ_λ of the segment lengths.

In order to generate new motion, we repeat the following procedure, starting with $\lambda = 1$:

- select an initial state $x_0 \in \{x_{0,\lambda}^i\}$.
- run the linear system M_λ for T steps to generate the motion segment, with $T \sim \mathcal{N}(\mu_\lambda, \sigma_\lambda)$.
- switch to the other linear system, $\lambda = \{1, 2\} \setminus \lambda$.

The synthetic motion is obtained as linear combination of the PCA basis, where the coefficients are given by the switching system output. Finally, the estimated linear trend is added to the position coordinates.

In figure 4 we show the results of the synthesis with models learned from walking and running motions. Figure 5 shows the prediction errors for these models on the training data. On these plots we can also see the prediction errors obtained with standard linear systems of the same order. It is clear how the proposed switching systems capture distinctive traits of the signal that cannot be modeled by simple linear systems. For qualitative comparison between original and synthesized motion see also the uploaded movies.

5. Conclusions

In this work we have proposed a novel switching system for modeling human gaits that captures the effect of ground contact forces on the motion dynamics. We have introduced a novel closed form identification algorithm for detecting system switches corresponding to ground contact forces and derived an algorithm for learning the model parameters. We have validated the modeling power of our system empirically by measuring the prediction error on the training data and showing that motion synthesized by this system is perceptually equivalent to the original learning sequence.

References

- [1] O. Arikan and D. Forsyth. Interactive motion generation from examples. In *Proc. of ACM SIGGRAPH 02*, 2002.
- [2] M. Billio, A. Momfort and C.P. Robert. Bayesian estimation of switching ARMA models. In *Journal of Econometrics*, 93, 229-255, 1999.
- [3] A. Bissacco, A. Chiuso, Y. Ma and S. Soatto Recognition of human gaits. In *Proc. of Computer Vision and Pattern Recognition*, vol II, pages 52-58, 2001.
- [4] H. A. P. Blom and Y. Bar-Shalom. The interactive multiple model algorithm for systems with Markovian switching coefficients. In *IEEE Transactions on Automatic Control*, vol. 33, pp. 780-783, August 1988.

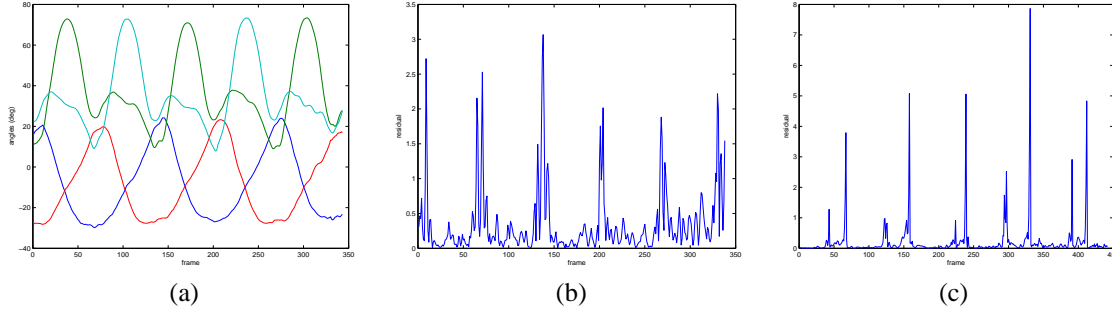


Figure 3: **Switch detection on human motion data:** Results of the switch detection algorithm applied on walking gait data from motion capture. (a) original trajectories of the 4 leg joint angles used for switch detection in a walking gait. (b) norm of the columns of \hat{E} , $i = 3$. Clear peaks appear at ground contact instants. Since for real data the discontinuity occurs at knee lock and foot ground contact, the peak is often multimodal, as in (c), – the first peak is the knee lock, the second peak is ground contact yet easy to detect by thresholding. Notice from the joint plots (a) how the quality of the data degrades towards the end of the sequence, as does the quality of the peaks (b) in the switch detection residual.

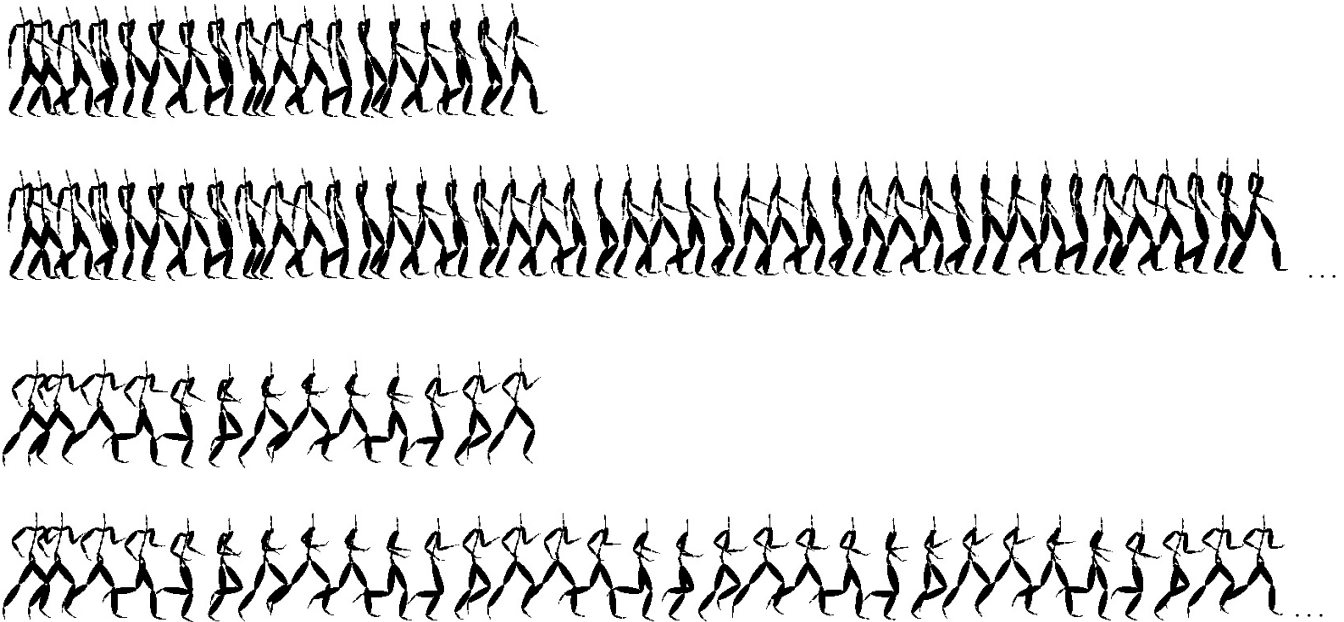


Figure 4: **Learning walking and running gaits (motion capture data):** First row shows the original walking motion data (see uploaded movie `walk1orig.mov`). Second row depicts the synthesized sequence (`walk1synth.mov`). Third row original running motion (`run1orig.mov`), fourth row synthetic running (`run1synth.mov`). See also the other uploaded movies `walk2orig`, `walk2synth`, `run2orig`, `run2synth`.

- [5] C. Bregler. Learning and Recognizing Human Dynamics in Video Sequences. In *Proc. IEEE Conf. Comp. Vision and Pattern Recognition*, June 1997.
- [6] C. Bregler and J. Malik. Tracking people with twists and exponential maps. In *Proc. IEEE Conf. Comp. Vision and Pattern Recognition*, 1998.
- [7] M. Buehler, D. E. Koditscheck and P. J. Kindlmann. A Family of Robot Control Strategies for Intermittent Dynamical Environments. In *IEEE Control System Magazine*, Vol 10(2), p 16-22, Feb. 1990.
- [8] R. Chen and J. S. Liu. Mixture Kalman Filters. In *Journal of the Royal Statistical Society - Series B*, 62:493-508, 2000.
- [9] C. Chevallereau and Y. Aoustin. Optimal reference trajectories for walking and running of a biped robot. In *Robotica*, No. 19, pp 557-569. 2001.

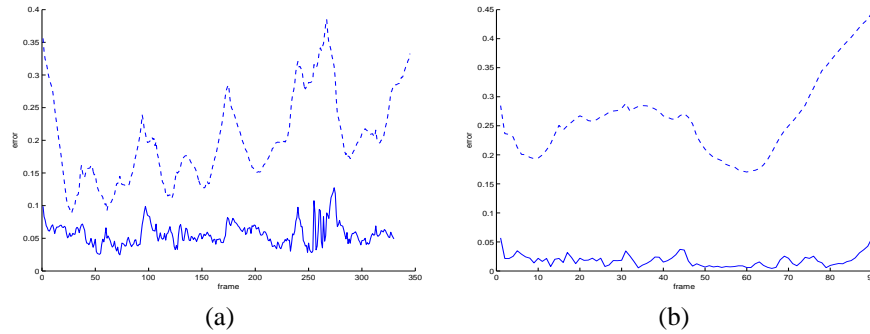


Figure 5: **Quantitative evaluation of the models (prediction error residuals):** Plot of the prediction error for the switching model (solid line) and linear system (dashed line), for walking (a) and running (b) motion data.

- [10] S.H. Collins, M. Wisse and A. Ruina. A 3-D passive-dynamic walking robot with two legs and knees. In *International Journal of Robotics Research*, April 2001.
- [11] A. Doucet and C. Andrieu. Iterative algorithms for state estimation of Jump Markov linear systems. In *IEEE Transactions on Signal Processing*, 49(6):1216:1227, June 2001.
- [12] P. Faloutsos, M. van de Panne, and D. Terzopoulos. Composable controllers for physics-based character animation. In *Proc. of ACM SIGGRAPH*, 2001.
- [13] M. Garcia, A. Chatterjee and A. Ruina. The simplest walking model: stability, complexity and scaling. In *Journal of Biomechanical Engineering*, 120, pp. 281-288, 1998.
- [14] Z. Ghahramani and G. E. Hinton. Variational learning for switching state-space models. In *Neural Computation*, 12(4):963-996, 1998.
- [15] J. K. Hodgins, W. L. Wooten, D. C. Brogan and J. F. O'Brien. Animating Human Athletics. In *Proc. of ACM SIGGRAPH*, Vol 29, pp. 71-78, 1995.
- [16] X. Koutsoukos, J. Kurien and F. Zhao. Monitoring and diagnosis of hybrid systems using particle filtering methods. In *Proc. of Int. Symposium on Mathematical Theory of Networks and Systems*, August 2002.
- [17] L. Kovar, M. Gleicher and F. Pighin. Motion graphs. In *Proc. of ACM SIGGRAPH*, 2002.
- [18] Y. Li, T. Wang and H.-Y. Shum. Motion texture: a two-level statistical model for character motion synthesis. In *Proc. SIGGRAPH*, 2002.
- [19] A. Logothetis and V. Krishnamurthy. Expectation maximization algorithms for MAP estimation of jump Markov linear systems. In *IEEE Transactions on Signal Processing*, 47(8):2139-2156, August 1999.
- [20] T. McGeer. Passive dynamic walking. In *International Journal of Robotics Research*, Vol. 9, No., 2, pp. 62-82, 1990.
- [21] K. Murphy. Learning Switching Kalman Filter Models. Technical Report 98-10, Compaq Cambridge Research, 1998.
- [22] B. North, A. Blake, M. Isard and J. Rittscher. Learning and Classification of Complex Dynamics. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 9, Sept. 2000.
- [23] V. Pavlovic, J. M. Rehg and J. MacCormick. Learning Switching Linear Models of Human Motion. In *Neural Information Processing Systems*, 2000.
- [24] F. Plestan, J. W. Grizzle, E. R. Westervelt, and G. Abba. Stable Walking of a 7-DOF Biped Robot. In *IEEE Transactions Robotics and Automation*, Vol. 19, No. 4, pp. 653-668, Aug. 2003.
- [25] C. Rasmussen and G. D. Hager. Probabilistic Data Association Methods for Tracking Complex Visual Objects. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, June 2001.
- [26] A. Safonova, J. K. Hodgins and N. S. Pollard. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. In *Proc. of ACM SIGGRAPH*, Aug. 2004.
- [27] K. Toyama and A. Blake. Probabilistic Tracking with Exemplars in a Metric Space. In *Intl. Journal of Computer Vision*, 48(1), 9-19, 2002.
- [28] J. K. Tugnait. Detection and Estimation for Abruptly Changing Systems. In *Automatica*, Vol. 18, No. 5, pp. 607-615, 1982.
- [29] P. Van Overschee and B. De Moor. Subspace Identification for Linear Systems: Theory, Implementation, Applications. Kluwer Academic Publishers, 1996.

- [30] R. Vidal, S. Soatto and S. Sastry. An algebraic geometric approach to the identification of linear hybrid systems. In *Proc. of the IEEE Conf. on Decision and Control*, 2003.
- [31] M. Vukobratovic, B. Borovac, D. Surla and D. Stokic. Biped Locomotion. Communications and Control Engineering Series, Springer-Verlag, 1990.
- [32] A. Witkin and M. Kass. Spacetime Constraints. In *Proc. of ACM SIGGRAPH*, v.22 n.4, p.159-168, Aug. 1988.