Image-based modeling of human gaits with higher-order statistics

Payam Saisan
UCLA
Electrical Engineering
Los Angeles, CA 90095
saisan@ee.ucla.edu

Alessandro Bissacco
UCLA
Computer Science
Los Angeles, CA 90095
bissacco@cs.ucla.edu

Keywords: Image-based modeling and rendering, higher-order statistics, dynamic independent component analysis, human motion modeling, visual recognition, animation

Abstract

We present a novel approach to modeling human gaits such as walking and running. We represent the trajectories of a certain number of salient features on the human body as the output of a dynamical system driven by an unknown stochastic input. We present techniques for inferring model parameters and input signal distributions corresponding to different optimality criteria, and evaluate the corresponding models for accuracy and predictive power. In particular, we exploit the higher-order statistical information content in motion capture data to arrive at input signals with independent components. We show that human gaits synthesized from non-Gaussian inputs best capture the dynamic complexities of the original gait data.

1 Introduction

Human gaits can convey rich and subtle information. By looking at a person walking from afar, we often can tell whether she is happy, tired, or wounded. We can make accurate guesses as to individual traits such as the rough age or gender: we can even identify a known individual from her walk. It is clear that such information is encoded not in each individual static pose, but in the *dynamics* of the moving body. In Johansson's experiments [10] one cannot tell much from a single frame, but when the sequence is animated suddenly the event becomes easy to parse and identify[1].

Modeling the subtleties in human motion can play a crucial role in a number of applications ranging from security (recognizing individuals from their gaits) to entertainment and the arts (motion capture and synthesis). For the sake of concreteness, we take as our driving application image-based modeling for the purpose of motion synthesis. The idea is simple. We want to collect motion-capture data[2] for an individual, and from these data build a model that can be used to generate novel synthetic motions, for instance of an animated character. However, unlike most prior work, we want to be able to do so while retaining the "distinctive character" of the individual person in the training set. For instance, if we observe Mr. Jones walking around long enough, our long term goal is to develop a model that can then be used to synthesize novel motions that look like Mr. Jones'.

In order to do so, we define a gait as the output of a dynamical system driven by a stochastic process with an unknown distribution. While this was done in [9] for the case of Linear Gaussian models, in this work we generalize the model to arbitrary input distributions. Learning a model then amounts to inferring the model dynamics as well as the input distribution. Unlike the Gaussian case, no closed-form exists, and even the optimality criterion is somewhat open to discussion. In this work we explore alternatives, in particular non-Gaussian models obtained from a dynamic version of the independent component analysis (ICA) [7].

The quality of the model inferred can be measured by the size of the residual, the L^2 norm or alternatively the relative entropy (Kullbach-Leibler divergence) that measures how well the model fits the training data. However, more importantly for the driving application we are considering, one can simulate the model forward, and visually inspect whether the resulting simulation captures the "character" of the input set. More quantitatively, one can use a portion of the data to learn a model, use the model to predict future data, and then compare that to the real data acquired in subsequent times.

Before delving into the model, we discuss how our approach relates to the state of the art.

1.1 Relation to previous work and contribution of this paper

Modeling subtleties in human or animal motion has been subject of considerable attention lately. Bregler and coworkers [34] have proposed a variety of methods to model non-rigid motions in an attempt to capture subtleties. Such models are built as linear combinations of a collection of "key" poses, learned using principal component analysis from motion capture data. These approaches are also related to a linear-Gaussian model that is a special case of what we describe below. Bissacco et al. [9] used a similar model for the purpose of recognition, and defined a metric on the space of mod-

[1] Naturally there is a great deal of information in the photometry and geometry of the scene that can be conveyed in a single static frame. However, in this study we concentrate on the scene dynamics. Johansson's experiments are enlightening because they show that even after stripping a sequence of all of its pictorial content, the dynamics of moving dots still retains remarkable amount of information.

[2] In particular, trajectories of a collection of marker positions in space.

els that was used for classification. To the best of our knowledge, dynamical systems with arbitrary input distributions have never been used to model human gaits before.

The literature on modeling human motion is sizeable and growing (see [11] for a survey). A common approach consists of extracting low-level features by local spatio-temporal filtering on the images and using hidden Markov models (HMM) on the collection of sequences of points thus obtained for recognition and classification tasks [12, 13]. In [14, 15] Bayesian Networks are used for recognition tasks. Local representations of motion based on optical flow have been exploited in [17, 20, 21], and view-based methods are proposed in [18, 16, 19]. Other approaches are based on principal component analysis [22], parameterization of the motion on joint angles [23] and snake fitting [24]. Estimation of motion from stereo [25] and multiple-view systems [26] have also been investigated. In [27] a mixed-state statistical model for the representation of motion has been proposed. In this Switching Linear Dynamic Model a stochastic finite-state automata at the highest level switches between local linear Gaussian models. Estimation and recognition is performed with expectation maximization approaches using particle filters [28, 29] or structured variational inference techniques [30].

Our models are discrete-time, continuous-state dynamical systems, and the action is coded in the dynamical model (i.e. the system parameters) as well as the input distribution. Our treatment of the problem comes at a level of abstraction higher than typical motion detection and tracking algorithms. It can use motion tracker data of varying modalities to infer a global model of the dynamics for a human gait, using dynamical systems with unknown input distributions.

2 Modeling subtleties

We assume that we are given the trajectory of a number of distinctive feature points on the human body. For instance, when a motion capture system is used for data acquisition, these points correspond to the position of a number of markers placed ad-hoc on the human subject, for instance at her joints. We denote the position of each marker i at time instant t by $y_i(t) \in \mathbb{R}^3$. Alternatively, one may consider the joint angles corresponding to a skeletal model of the subject. In that case, one may call the joint angle i at time t $y_i(t) \in [0,\ C_i)$, where C_i is a constant that can be either π or 2π depending upon the joint. In any case, the dataset consists of a trajectory in some M-dimensional space; for simplicity we consider the first case where

$$y(t) \in \mathbb{R}^{3M},\ t \in [0,\ t_f] \tag{1}$$

However, the considerations developed here can be transposed to any other representation of the input data one chooses.

As we anticipated, we model $y(t)$ as the output of a dynamical system driven by a stochastic input. That is, we assume that at each instant of time there exists a vector $x(t) \in \mathbb{R}^N$, and suitable functions f and h such that

$$\begin{cases} x(t+1) &= f(x(t)) + Bv(t); \quad x(0) = x_0 \\ y(t) &= h(x(t)) + w(t), \end{cases} \tag{2}$$

where x_0 is an unknown but constant vector (the initial condition), and $v(t)$ and $w(t)$ are white, zero-mean, stationary stochastic processes. While we assume that $w(t)$ is normally distributed, $w(t) \sim \mathcal{N}(0, R)$, we allow $v(t)$ to be a sample from an unknown probability density q:

$$v(t) \overset{IID}{\sim} q(v). \tag{3}$$

Our goal, then, is to infer the input density q, the initial condition x_0, the order of the model N, the dynamics f and the output map h from a time series $\{y(t)\}_{t\in[0,\ t_f]}$. Naturally, there are many ways of doing so, depending on what inference criterion one chooses. In the following section we describe several criteria that we have evaluated, and the resulting learning algorithms. For the sake of simplicity we restrict our attention to linear dynamics $f(x) = Ax$ and linear output maps $h(x) = Cx$, shifting the emphasis of the modeling power to the input distribution q. Some of the techniques outlined below can be extended to arbitrary nonlinear models, although this is beyond the scope of this paper.

3 Inference criteria and learning algorithms

Given a sequence of output data $\{y(t)\}_{t\in[0,\ t_f]}$ generated by a model of the following form with negligible output noise $w(t) = 0$

$$\begin{cases} x(t+1) &= Ax(t) + Bv(t); \quad x(0) = x_0 \\ y(t) &= Cx(t), \end{cases} \tag{4}$$

with

$$v(t) \overset{IID}{\sim} q(v). \tag{5}$$

our goal is that of finding, in some sense optimal, estimates of the unknowns

$$\hat{A}, \hat{B}, \hat{C}, \hat{q}(\cdot), \hat{x}(t). \tag{6}$$

although we consider suboptimal solutions as well. We can choose from several optimality criteria, depending on whether we seek the maximum likelihood parameters, the maximum a-posteriori (given a prior distribution on the unknowns), the parameters that result in a maximally independent estimated input sequence $\hat{v}(t)$, or the parameters that result in the minimum variance of the output error $y(t) - \hat{y}(t)$. There is no right or wrong criterion; the best way to decide is to test several, and compare the results. Comparison can be performed by verifying how well the model captures the training data (residuals), or how well the model can predict future data (prediction error) or simply by simulating the model and visually inspecting the results to verify the success of one criterion over another.

3.1 Linear Gaussian models and maximum likelihood

If we assume that the input distribution is Gaussian, $q(\cdot) \in \mathcal{N}(0, Q)$, then there are closed-form solutions for the identification of the model parameters A, C, Q and the state sequence $x(t)$ minimizing the likelihood of the output error. In particular, subspace identification algorithms can be used for this purpose. Since the emphasis of this paper is on non-Gaussian modeling, we refer the reader to [33] for details, and to the experimental section for experiments.

3.2 Dynam-ICA: dynamic independent component analysis

A criterion for inference of model parameters and input distributions can be constructed by requiring the estimated input sequence $\hat{v}(t)$ to be a realization from a stochastic process that has maximally independent components. Independence can be expressed in terms of information theoretic quantities, such as the mutual information among input components, or relative entropy (Kullback-Leibler divergence) between probability density functions.

This approach results in a semi-parametric statistical inference problem, where one has to simultaneously infer the finite-dimensional model parameters as well as the infinite-dimensional input distribution q. This is essentially an independent component analysis (ICA) problem.

In its conventional static form, ICA attempts to decompose a random vector into a linear combination of statistically independent components. If we call $y \in \mathbb{R}^m$ the random vector, then ICA looks for a matrix $C \in \mathbb{R}^{m \times n}$ with $n \leq m$, and a random vector $x \in \mathbb{R}^n$ with independent components, $p_{\mathbf{x}(t)}(x_1, ..., x_n) = \prod_{i=1}^n p_{x_i(t)}(x_i)$ such that

$$y = Cx. \tag{7}$$

The unknowns C and p_{x_i} can be estimated by minimizing the Kullback-Leibler divergence between the corresponding density functions. For the most common case of invertible square mixing matrix C ($c_i^{-T} \doteq$ i'th row of C^{-1}) this is formalized as follows:

$$(\hat{C}, \hat{p}_x) = \arg\min_{C, p_x} K(p_y(y) || \prod_{i=1}^n p_{xi}(c_i^{-T} y) \left| C^{-1} \right|) \tag{8}$$

with

$$K(p(x)||q(x)) \doteq \int p(x) \log \frac{p(x)}{q(x)} dx \tag{9}$$

computed or approximated using a number of independent and identically distributed (IID) samples from $p_{\mathbf{y}}$: $y(1), \ldots, y(t) \overset{IID}{\sim} p_{\mathbf{y}}$. Typically the process y is assumed to be ergodic, and therefore a time series is used in lieu of a fair sample.

What we have here, however, is a dynamic ICA problem of separating independent components of a signal $v(t)$ mixed by way of input into a linear dynamical system of form (4).

Let us rewrite the output of the model at time t:

$$y(t) = \left[CA^t,\ CA^{t-1}B, \ldots, CB\right] \begin{bmatrix} x(0) \\ v(0) \\ \vdots \\ v(t-1) \end{bmatrix} \doteq \tilde{\mathcal{C}}^t \tilde{\mathbf{V}} \tag{10}$$

with

$$p_{\mathbf{v}(t)}(v_1, \ldots, v_n) = \prod_{i=1}^n q(v_i) \ \forall\ t. \tag{11}$$

Stack the observations $y(1), \ldots, y(t)$ into a vector $\mathbf{Y}^t$ to obtain

$$\mathbf{Y}^t = \tilde{\mathcal{C}}^t \tilde{\mathbf{V}}. \tag{12}$$

One may be tempted to invoke the independence of the components of $\mathbf{V}$ – based on the assumptions that $v(t)$ is white (time samples are independent) and has independent components – and use standard ICA to estimate the mixing matrix $\tilde{\mathcal{C}}^t$. This, however, does not work because of several immediate difficulties. It is not possible to use time realizations as independent samples of $\mathbf{Y}$ due to the initial condition x_0. The problem is exacerbated by the ambiguity introduced by ICA in the order of the components of $\tilde{\mathbf{V}}$.

In what follows we make the simplifying assumption that the initial condition x_0 is zero. One may conjecture that if t is large enough and A is stable the effect of the initial condition will wane; therefore, the assumption may not be as restrictive. Under this assumption, the problem of dynamic ICA can be posed as follows. Consider $\mathbf{Y}^t(k) = [y((k-1)t)^T, \ldots, y(kt-1)^T]^T$, and similarly for $\mathbf{V}^t(k)$. Note that $\mathbf{Y}^t(i)$ and $\mathbf{Y}^t(j)$ do not share components $y(k)$. Furthermore, let $\mathcal{C}^t$ be the matrix obtained by completing, in the sense of Toeplitz, the following matrix

$$\begin{bmatrix} CB & & & \\ CAB & CB & & \\ \vdots & \vdots & \ddots & \\ CA^{t-1}B & CA^{t-2}B & \ldots & CB \end{bmatrix}. \tag{13}$$

Then $\hat{A}, \hat{B}, \hat{C}$ may be found sub-optimally by estimating the mixing matrix $\mathcal{C}^t$ having the particular structure above from a set of independent samples $\mathbf{Y}^t(1), \ldots, \mathbf{Y}^t(k)$.

Under the assumption of stationarity, the model above can be thought of an input-output form as $\mathbf{Y}^t(i) = \mathcal{C}^t \mathbf{V}^t(i)$. This is in the more familiar form of a blind deconvolution problem which maybe solved using Amari's natural gradient flow [2].

To the best of our knowledge, the formulation of the problem in the form and level of generality presented in this section is new, although variants have surfaced in literature under disguise of system identification using higher order statistics [4] and blind deconvolution [5]. In the course of developing optimal and efficient solution(s) to the bigger problem we are assessing simplifying assumptions like the one just presented. In the next section we describe yet a different, further simplified formulation and solution to the problem. We show that even suboptimal approximations leading to non-Gaussian input models produce visibly superior synthesized gaits, over their Gaussian input counterparts.

3.3 A suboptimal solution

We considered, as our starting point, a simplified analog and solved the problem of estimating system parameters and non-Gaussian input in two suboptimal stages. We show results indicating visible improvement despite the limitations of even this simplest approximate paradigm.

Our data consists of time trajectories of joint angles. We are set to be observing the state directly, thus assuming the trivial case of $y(t) = x(t)$. The system is, therefore, reduced to an autoregressive (AR) model:

$$y(t+1) = Ay(t) + Bv(t). \tag{14}$$

A can be estimated suboptimally by minimizing variance of the norm of the residuals $e(t) = y(t+1) - Ay(t)$. Given data samples $\{y(t)\}_{t=1}^{N}$, the simple closed form least squares solution is given as :

$$\begin{aligned} \hat{A} &= \arg\min_{A} \sum ||y(t+1) - Ay(t)||^2 \\ &= \left(\sum_{1}^{N-1} y(t+1)y(t)^T \right) \left(\sum_{1}^{N-1} y(t)y(t)^T \right)^{-1} \end{aligned} \tag{15}$$

Given an estimate of A, we can compute the residuals $e(t) = Bv(t)$ from the data. Invoking independence of components for $v(t)$, we can use static ICA techniques to decompose the residuals into a mixing matrix part, B and $v(t)$ with independent components. We have broken down the dynamic ICA problem into a sub-optimal but simple parameter estimation and static ICA stages.

Once a model and independent components have been inferred, synthesis can be performed by simulating the model forward in time, that is by sampling an input from the estimated non-Gaussian distribution $\hat{q}$, and using it to compute the one-step increment of the state $\hat{x}(t+1)$ and hence the output trajectory $\hat{y}(t)$. In some cases, an explicit expression of the density q may be available. In some other cases, when the order of the model N is very high, sampling may be non trivial, and techniques such as Gibbs sampling may be necessary [36]. The histogram of the values assumed by the elements v_i can be used as discrete approximation of the input density $\hat{q}_i$. Then synthesis can be easily performed by drawing samples from the estimated distributions $\hat{q}_i$.

Sampling techniques can also be used to perform inference, in a particle filtering framework, as we describe below.

3.4 Particle filtering and identification

Particle filtering techniques, of the kind discussed in [32, 31], can be employed to filter the state $\hat{x}(t)$ given (linear or non-linear) model parameters. On the other hand, model parameters can be inferred using maximum likelihood criteria once the best current estimate of the state is available. This strategy for filtering and identification of non-Gaussian models has been proposed by North and Blake [28]. Although in this preliminary study we have not implemented this technique, it is viable for the task we describe, and it is therefore our intention to test it against our current best system in the course of forthcoming experiments.

4 Experiments

In this first round of experiments we learned and synthesized from models of the form (4) with non-Gaussian inputs, $v(t)$, following the outline at the end of section 3.3 . The data sequences used in learning correspond to different subjects' walking and running actions. They consist of trajectories of four joint angles, the shoulder, the elbow, the hip and the knee. A typical sequence is 48 to 78 frames long, obtained by applying standard image-based tracking techniques [35] to video streams taken with a 30Hz camera.

For comparison, we also learned and synthesized from Gaussian models using subspace identification techniques discussed in section 3.1. Thus, we were able to compare optimally estimated Gaussian input models with their sub-optimally estimated non-Gaussian counterpart. Although our data set was small and restrictive assumptions were made in the implementation of the dynamic ICA paradigm, we noticed visible improvement in the quality of synthesis with non-Gaussian input models. Figures 1 and 2 show synthesized sequences versus the original gait data for walking and running. We have relied mostly on visual inspection in our interpretation of the two results at this point. Static figures may be insufficient in conveying complex dynamic behavior. We have, therefore, included movie clips accessible through our web site at http://vision.cs.ucla.edu. Collection of an extensive gait data set for more definitive results and comparisons is the goal of our future research.

5 Discussion

We described a novel approach for modeling of human gaits using linear dynamical systems driven by stochastic inputs. We put forth a formulation for the dynamic independent component analysis problem where the goal is estimation of system parameters as to allow recovery of input sequences with maximally independent components. Using a simplified sub-optimal reformulation we estimated system parameters and independent input components. We used this model to synthesize novel instances of a person's walking and running actions. We also learned and synthesized from optimally estimated Gaussian input models. Despite the preliminary nature of our experiments and the simplifying assumptions, we showed synthesized gaits that are more truthful, visibly, to the original gaits when generated using non-Gaussian inputs.

References

[1] S. Amari and F. Cardoso. Blind source separation–semiparametric statistical approach. *IEEE Trans. Signal Processing*, 45(11):2692–2700, 1997.

[2] S. Amari Natural gradient works efficiently in learning. *Neural Computaion*, 10:251–276, 1998.

[3] Various authors. http://www.salk.edu. 2000.

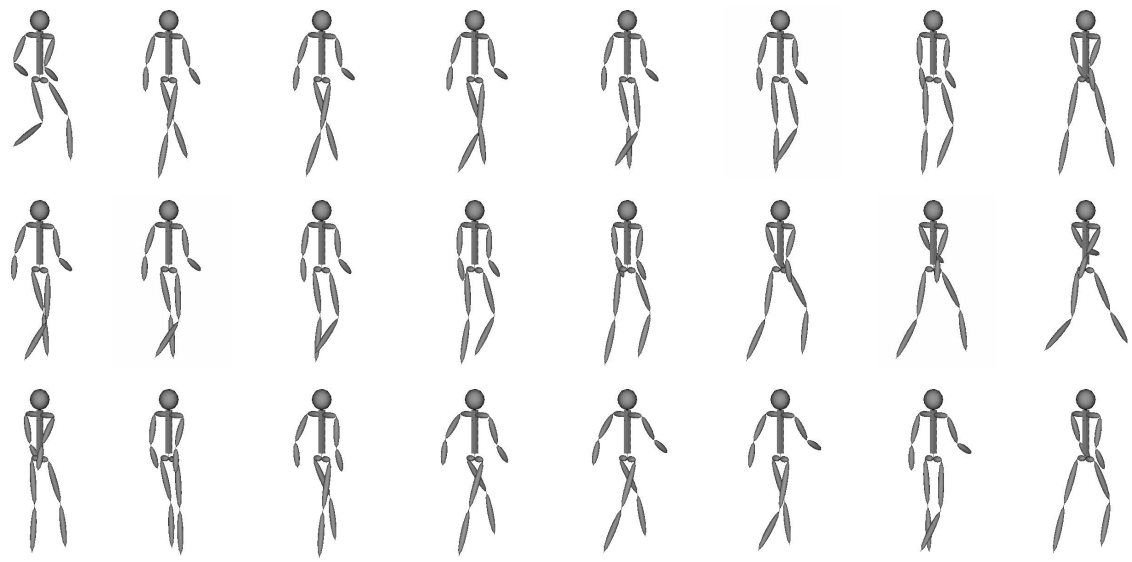

Figure 1: *Motion sequences for walking. Top row is the original walking data, second row is the synthesized sequence using Gaussian input and third row is the ICA based non-Gaussian input driven sequence.*

Figure 2: *Motion sequences for running. Top row is the original running data, second row is the synthesized sequence using Gaussian input and third row is the ICA based non-Gaussian input driven sequence.*

[4] B. Giannakis. and J. Mendel. Identification of non-minimum phase systems using higher order statistics. *IEEE Trans. Acoustic Speech and Signal Processing*, 37(3):360–377, 1989.

[5] L. Zhang. and A. Cichocki Blind deconvolution of Dynamical Systems : A State-Space Approach. *Proceedings of the IEEE. Workshop on NNSP'98*, 123–131, 1998.

[6] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[7] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36:287–314, 1994.

[8] A. Hyvärinen. Independent component analysis for time-dependent stochastic processes. 1998.

[9] A. Bissacco, A. Chiuso, Y. Ma and S. Soatto. Recognition of Human Gaits. In Proc. of the IEEE Intl. Conf. on Comp. Vision and Patt. Recog., pages 401-417, December 2001.

[10] G. Johansson. Visual perception of biological motion and a model for its analysis. Perception and Psychophysics, 14(2):201–211, 1973.

[11] D. M. Gavrila. The visual analysis of human movement: A survey. In *Computer Vision and Image Understanding*, volume 73, pages 82–98, 1999.

[12] T.Starner and A. Pentland. Real-time american

sign language recognition from video using hmm. In *Proc. of ISCV 95*, volume 29, pages 213–244, 1997.

[13] A. D. Wilson and A. F. Bobick. Parametric hidden markov models for gesture recognition. In *IEEE Trans. on Pattern Analysis and Machine Intelligence*, volume 21(9), pages 884–900, Sept. 1999.

[14] A. Madabhushi and J. K. Aggarwal. A bayesian approach to human activity recognition. In *Proc. of the 2nd International Workshop on Visual Surveillance*, pages 25–30, June 1999.

[15] J. Binder, D. Koeller, S. Russell, and K. Kanazawa. Adaptive probabilistic networks with hidden variables. In *Machine Learning*, volume 29, pages 213–244, 1997.

[16] M. J. Black. Eigentracking: robust matching and tracking of articulated objects. 1996.

[17] M. J. Black. Explaining optical flow events with parameterized spatio-temporal models. In *Proc. of Conference on Computer Vision and Pattern Recognition*, volume 1, pages 326–332, 1999.

[18] A. F. Bobick. Appearance-based representation of action. 1996.

[19] M. A. Giese and T. Poggio. Morphable models for the analysis and synthesis of complex motion patterns. In *International Journal of Computer Vision*, volume 38(1), pages 1264–1274, 2000.

[20] J. Hoey and J. J. Little. Representation and recognition of complex human motion. In *Proc. of the Conference on Computer Vision and Pattern Recognition*, volume 1, pages 752–759, 2000.

[21] J. J. Little and J. E. Boyd. Recognizing people by their gait: the shape of motion. 1996.

[22] Y. Yacoob and M. J. Black. Parameterized modeling and recognition of activities. In *Computer Vision and Image Understanding*, volume 73(2), pages 232–247, 1999.

[23] L. Campbell and A. Bobick. Recognition of human body motion using phase space constraints. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, page 8, 1995.

[24] A. A. Niyogi. Analyzing and recognizing walking figures in xyt. In *Proc. IEEE Conf. on Comp. Vision and Pattern Recogn.*, pages 469–474, Seattle, June, 1994.

[25] C. Wren. Dynamic models of human motion. 1998.

[26] D. M. Gavrila and L. S. Davis. Tracking of humans in action: a 3-d model-based approach. 1996.

[27] C. Bregler. Learning and recognizing human dynamics in video sequences. In *Proc. of the Conference on Computer Vision and Pattern Recognition*, pages 568–574, 1997.

[28] B. North and A. Blake and M. Isard and J. Rittscher. Learning and classification of complex dynamics. In *IEEE Transaction on Pattern Analysis and Machine Intelligence*, volume 22(9), pages 1016-34, 2000.

[29] M. J. Black and A. D. Jepson. A Probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In *Proc. of European Conference on Computer Vision*, volume 1, pages 909-24, 1998.

[30] V. Pavlovic and J. Rehg and J. MacCormick. Impact of Dynamic Model Learning on Classification of Human Motion In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2000.

[31] M. Isard and A. Blake. Condensation - conditional density propagation for visual tracking International Journal of Computer Vision 29(1), pp. 5–28, 1998.

[32] J. Liu and M. West Combined parameter and state estimation in simulation-based filtering In Sequential Monte Carlo Methods in Practice. New York: Springer-Verlag New York, 2000.

[33] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29:649–660, 1993.

[34] L. Torresani and D. Yang and G. Alexander and C. Bregler Tracking and Modelling Non-Rigid Objects with Rank Constraints In *Proc. International Conference on Computer Vision and Pattern Recognition*, 2001.

[35] C. Bregler Tracking people with twists and exponential maps In *Proc. International Conference on Computer Vision and Pattern Recognition*, 1998.

[36] S. Geman and D. Geman Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI–6, pp. 721–741, Nov. 1984