# Backing Off: Hierarchical Decomposition of Activity for 3D Novel Pose Recovery

John Darby[1]
j.darby@mmu.ac.uk

Baihua Li[1]
b.li@mmu.ac.uk

Nicholas Costen[1]
n.costen@mmu.ac.uk

David Fleet[2]
fleet@cs.toronto.edu

Neil Lawrence[3]
neill@cs.man.ac.uk

[1] Department of Computing and Mathematics
Manchester Metropolitan University

[2] Department of Computer Science
University of Toronto

[3] School of Computer Science
University of Manchester

## Abstract

For model-based 3D human pose estimation, even simple models of the human body lead to high-dimensional state spaces. Where the class of activity is known a priori, low-dimensional activity models learned from training data make possible a thorough and efficient search for the best pose. Conversely, searching for solutions in the full state space places no restriction on the class of motion to be recovered, but is both difficult and expensive. This paper explores a potential middle ground between these approaches, using the hierarchical Gaussian process latent variable model to learn activity at different hierarchical scales within the human skeleton. We show that by training on full-body activity data then descending through the hierarchy in stages and exploring subtrees independently of one another, novel poses may be recovered. Experimental results on motion capture data and monocular video sequences demonstrate the utility of the approach, and comparisons are drawn with existing low-dimensional activity models.

## 1 Introduction

In recent years model-based analysis-by-synthesis has proven a popular and successful method for the recovery of 3D human body pose from video sequences [17]. Even simple models of the human body contain 30 or more degrees of freedom, resulting in a state space for which search is computationally expensive [3, 4]. Low-dimensional models of activity have been employed to effectively constrain this search task, using deterministic optimisation or particle filtering (*e.g.*, [14, 15, 19, 22, 25, 26]). Nevertheless, with such global, full-body motion models, a particular model can only be used to recover poses from its particular activity. Although generalisation to some intra-activity variations in style have been demonstrated [25, 26], recovering substantially novel poses has not so far been possible.

This paper shows that with a learned hierarchical model of body coordination for multiple activities, one can recover novel poses that comprise aspects of different activities. The idea

is well captured by the following pose estimation problem. Given training data for *(i)* a person walking and *(ii)* a person standing and waving, how can we construct and explore a model that can describe a person walking *whilst* waving? To solve this problem we adopt the hierarchical Gaussian process latent variable model (H-GPLVM) [12]. As discussed in § 3, the H-GPLVM is constructed by learning separate low-dimensional models for the variation in individual body parts. It is then augmented by further latent variable models capturing the coordination between these component parts. As such, one can view this model as a "part-based hierarchical model". With a form of annealed particle filtering (§ 4), we show that one can use the H-GPLVM learned from two or more activities to recover novel test poses.

The flexibility of the approach is demonstrated with the recovery of novel 3D poses from 3D motion capture (MoCap) data in § 5.1, and by estimating 3D human pose from 2D monocular data in § 5.2 and § 5.3. Although we test our approach on tracking problems, such as continuous video featuring people walking, what we present is a 3D pose estimation technique that is independent at each frame. We do not manually initialise the pose at the start frame, nor do we use the recovered pose at each frame to initialise the search at the next frame.

# 2 Background

There are many examples in the literature of model-based approaches to pose estimation that synthesise body model poses by sampling from a pre-learned low-dimensional latent space (*e.g.*, [9, 14, 15, 19, 20, 22, 24, 25, 26, 27]). Within this paradigm, principal components analysis (PCA) [20, 27] and a range of invertible non-linear dimensionality reduction techniques including Gaussian process latent variable models (GP-LVMs, including back-constrained and scaled variants) [9, 24, 25], Gaussian process dynamical models (GPDMs, including balanced variant) [19, 26], locally linear coordination [14] and Laplacian eigen-maps [15, 22] have been used to recover high-dimensional pose data. Such activity models have been explored using particle filtering techniques [14, 15, 20] and deterministic optimisation [25, 26, 27] to recover poses for accurate multiple camera tracking [14, 19], real-time tracking [9], tracking through occlusion and self occlusion [24, 26] and monocular and stereo camera tracking [20, 22, 25, 26, 27].

Many of the above approaches show some capacity to generalise to variations in style, or unknown subjects. Hou *et al.* [9] tracked an unknown subject performing jumping jacks. Urtasun *et al.* were able to track unknown walking subjects [27], the golf swings of unknown subjects [25] and an exaggerated walking style with increased stride length and rigid limbs [26]. However, when the activities to be tracked deviate significantly from those in the training data, these global models are unable to cope and pose estimation fails (*e.g.*, [22]). We argue that for effective 3D pose estimation, some capacity to relax the constraints of these full-body models and exploit conditional independencies in the kinematic tree is desirable.

There are two common ways that one might try to cope with poses beyond those captured by activity-specific models, *i.e.*, with generic kinematic models but more efficient search methods, and with part-based models. Using a tree-based kinematic model Gavrila and Davis employed a hierarchical search that first localises the torso, which then constrains a subsequent search for arms and legs [8]. Similarly, partitioned sampling divides the full state space into a series of smaller subspaces which may be estimated with particle filters [16]. The annealed particle filter [5] does not explicitly partition the state space but employs "adaptive diffusion", reestimating its noise model from each set of resampled particles. This
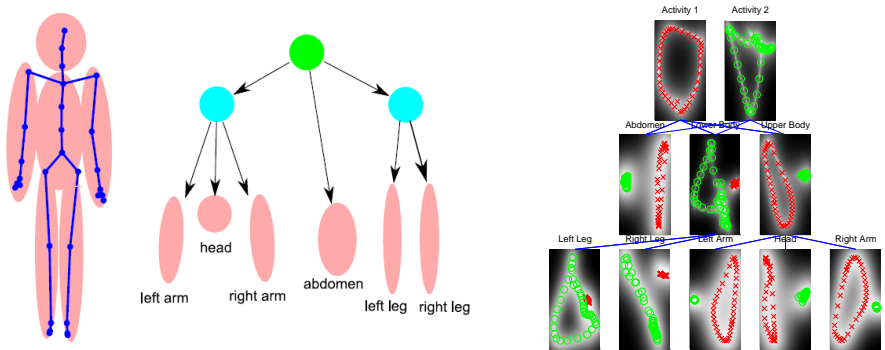
Figure 1: Skeleton and hierarachy of latent variables (from [12]) and a learned H-GPLVM for two motions (illustrated below in Fig. 2(a) and 2(b)), namely "swinging arms" and "walking with hurt stomach".

allows for the addition of reduced levels of noise to well localised parameters in a "soft partitioning" approach. This is similar to the use of larger noise terms for poorly localised parameters in covariance scaled sampling [23]. Part-based models typically detect possible body parts, and then search for poses using soft constraints based on the kinematic structure of the body. Tree-based models, in particular, facilitate efficient inference through belief propagation (*e.g.*, [6, 13]). Loopy graphical models allow for more expressive constraints on pose, but greatly increase the complexity of inference (*e.g.*, [21]). The H-GPLVM approach presented below is a compromise between the restrictions of a low-dimensional full-body activity model and the challenges of searching the high-dimensional state space of a generic kinematic model.

# 3 Hierarchical GP-LVM

The GP-LVM [11] represents high-dimensional data through a low-dimensional latent model, and a non-linear Gaussian Process (GP) mapping from the latent space to the data space. This makes it ideal for the representation of human motion. The GP-LVM exploits a probabilistic interpretation of PCA as a product of independent GP models over features, each with a linear covariance function [11]. By the consideration of non-linear covariance functions, such as a radial basis function kernel, non-linear latent variable models can be formulated. Optimising the latent variables (initialised with PCA) and kernel parameters given the set of high-dimensional training points results in a probabilistic model of the original data.

The H-GPLVM [12] is a form of GP-LVM with a hierarchical latent representation (*e.g.*, see Fig. 1). The leaves of the latent model comprise a latent model for each limb or distinct part of the body. That is, each node is a GP-LVM for a single body part. To capture the natural coordination of body parts one can then model the joint distribution over latent positions in a subset of leaf nodes with a GP from a parent latent variable. For example in Fig. 1 the *left leg* and *right leg* are coordinated by the *lower body latent variable*. Given a lower-body latent position, there is a GP mapping to latent positions for the left and right legs, from which there are GP mappings to the joint angles of the two legs.

**Learning:**    The latent positions in the hierarchy are learned from training data through maximum likelihood. Following [12], initial estimates of the latent variables in the leaf nodes are made using PCA. Initial estimates for the parents of leaf nodes are found by applying PCA to the concatenated latent variables of their dependents. Bottom-up construction continues in this manner until the root nodes are reached. There is one root node for each activity modelled, and the latent model in each root node is a function of the latent variables of its dependents that belong to its specific activity only. In this work all latent spaces have two dimensions. To prevent overfitting, fixed dynamical priors (using periodic kernels) are added to the root nodes, and the noise variance of each GP (other than leaf nodes) is fixed. Finally, the parameters of each GP kernel matrix and all latent variables are jointly optimised. (See [12] for further details.)

**Pose Generation:**    The H-GPLVM defines a sequence of Gaussian conditional distributions as one descends from a node at the top level of the hierarchy down to the leaves, and then to the 3D joint angles for each part of the body. Accordingly, a natural way to formulate pose estimation is in terms of Monte Carlo search, beginning at the top level of the hierarchy. One can draw proposals for pose from the H-GPLVM by sampling the top level, and then drawing a sample from intermediate level latent nodes conditioned on the top level, and so on as one descends the hierarchy.

   However, it is the H-GPLVM's part-based decomposition of the body that makes it a strong candidate for the recovery of novel poses that are not present in the training set. That is, one could also estimate pose by concentrating search at the intermediate levels of the hierarchy, or even at the leaves. For example, the leaf nodes provide latent models for the 3D pose of individual body parts. Searching the leaf nodes, without conditioning on their parents in the H-GPLVM, amounts to searching a "flat" part-based model where limbs are independent. In doing so this might allow different body parts to generate joint angles consistent with different activities from the training set. This motivates the approach described below.

# 4   Pose Estimation with the H-GPLVM

To exploit the hierarchical structure of the model Lawrence and Moore [12] suggest that a "back off" method inspired by language modelling might be used for the recovery of poses not featured in the H-GPLVM's training set. The idea is to descend the hierarchy and search nodes at the next level *independently*; this concept forms the basis for inference in this work. By shifting search down one level in the hierarchy we can gradually relax the level of coordination amongst body parts. While we may be unable to recover a novel test pose by inspection of full-body models at the root nodes, a good fit might be obtained by backing off to the middle level nodes to optimise the abdomen, upper body and lower body independently.

   Given the non-linear form of the model, and the potential for ambiguity in pose estimation (*i.e.*, for multi-modality), we formulate pose estimation using a form of Monte Carlo inference. For efficiency, given the dimensions of the entire latent space and the pose space, we advocate the use of the annealed particle filter (APF) [5]. In particular, we use a form of coarse-to-fine search, descending through the model from rough full-body pose estimates at the top level nodes of the model, to the eventual refinement of partial pose parameters for each limb in the leaf nodes.

To cope with novel poses not explicitly present in the training data, following Deutscher and Reid [5], we introduce a form of *crossover operator* to recombine the *building blocks* of candidates (particle proposals) that have high likelihood. This type of mechanism is ideal for the exploitation of reduced levels of coordination between limbs. For novel poses it will be necessary to retain full-body poses from the top level nodes of the hierarchy that are somewhat *flawed*. That is, even poses that show comparatively poor agreement with observation data may prove valuable in lower levels of the hierarchy, since they fit some but not all parts of the body well. The annealing schedule of APF is able to support a wide range of hypotheses in the early annealing layers before eventually concentrating on a particular solution in later layers. The next two sections outline the form of the APF and the way in which search proceeds in more detail.

## 4.1 Annealed Particle Filtering

A set of $N$ properly weighted particles, $\{(\mathbf{x}_t^{(1)}, \pi_t^{(1)}), ..., (\mathbf{x}_t^{(N)}, \pi_t^{(N)})\}$ is used to represent a Monte-Carlo approximation to a target (posterior) distribution over 3D pose and H-GPLVM latent positions. The distribution is propagated from one time to the next with a combination of *dispersion* by a model of temporal dynamics $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, *evaluation* against an observation $\mathbf{z}_t$ by a weighting function $w(\mathbf{z},\mathbf{x})$ based on the likelihood $p(\mathbf{z}|\mathbf{x})$, and *resampling* with probability proportional to the weights $\pi_t^{(n)}$ [7].

The APF [5] attempts to recover a maximally probable pose by cooling the target distribution before gradually allowing peaks in the distribution to emerge through resampling iterations, $r = R, R-1, ..., 1$ where

$$w_r(\mathbf{z},\mathbf{x}) = w(\mathbf{z},\mathbf{x})^{\beta_r}, \tag{1}$$

for $\beta_1 > \beta_2 > ... > \beta_R$. The value of $\beta_r$ is chosen to attain a desired particle survival rate $\alpha_r$, the proportion of particles that will be resampled at each layer. The survival rate is also used to rescale the addition of noise for dispersion, keeping perturbations inversely proportional with the particle set resolution. The effect is to gradually concentrate particles into the main mode of the target distribution. Although there are APF variants that try to maintain a proper representation of the posterior [7], in this work as in [5] the posterior distribution is no longer fully represented, constituting a departure from the formal Bayesian framework. In tracking applications APF often provides better pose estimates than simple particle filters [4].

## 4.2 Hierarchical Search

Our particle-based search proceeds from the top to the bottom of the H-GPLVM over a number of annealing layers, with back off occurring after each resampling stage. The reader may find it helpful to refer to the hierarchical decomposition of two activities used in this work and depicted in Fig. 1, but the method is also applicable to other decompositions. To aid our exposition of inference in the H-GPLVM we will refer to latent positions in non-leaf nodes as specifying *partial* or *full-body* poses in the original 3D pose space. Strictly speaking, there is no direct connection between the two, and implicit in these statements is the assumption that the probabilistic mappings between parent and child are used to fully descend the hierarchy through the leaf nodes to the pose space (see § 3).

Given an observation $\mathbf{z}_t$, the first annealing layer proceeds as follows. A set of $N$ particles is initialised by uniformly sampling from the latent representatives of the training set at the

root nodes. Each particle holds a latent position that may be used to completely descend the hierarchy and recover a full-body 3D pose. To allow these initial poses to depart from the training data, each particle's latent position is then perturbed with the addition of a zero mean Gaussian random variable with covariance $\Sigma$. Each corresponding pose $\mathbf{x}_t^{(n)}$ is then *evaluated* against the observation using $w_R(\mathbf{z}, \mathbf{x})$, and $N$ particles are *resampled* with likelihood in proportion to their weights and with replacement.

Given the latent position held by each resampled particle, the H-GPLVM defines Gaussian conditional distributions over the child nodes in the level below. To exploit the potential for independence between these latent spaces (and therefore body parts) in the *dispersion* step, a new particle set is constructed as follows. A single sample is drawn from the conditional distributions corresponding to each particle, yielding $N$ new latent positions in each child node. $N$ new particles, each holding a *set* of latent positions are then created by randomly sampling once from the new latent positions in each of the child nodes, without replacement. Subsequent annealing layers $r = R - 1, ..., 1$ proceed to back off down the hierarchy in just the same way, but are initialised with the new particle set from the previous layer.

Where a pose observation features occlusion or self occlusion it is desirable to infer the location of an occluded limb from visible limbs based on their correlations within the training data. Where $\mathbf{z}_t$ takes the form of a set of labelled 2D features *e.g.* [10, 13] and one or more are absent, the implication for the search strategy is as follows: where image evidence for a subtree of the skeleton is missing, do not descend below that subtree's parent node. In § 5.3 we use this principle to recover the occluded arm of a walking subject. Otherwise, back off ceases to take place only once the leaf nodes are reached.

In practice, the covariance of the GP mappings from parent to child in the H-GPLVM are often relatively small. This is due in part to the regularisation conditions (see § 3) and in part to our use of only one training subject. To encourage individual body parts to depart from the training data and increase pose diversity, we inflate the covariance to be equal to $\Sigma$, rescaled by $\alpha_r$ at each annealing layer. We also apply this weighted noise term to latent positions in nodes where back off has ceased to take place.

# 5 Experiments

In each experiment we train H-GPLVMs from 50D joint rotation vectors recovered from MoCap [1] and decomposed as shown in Fig. 1. In § 5.1 we also use MoCap test data to investigate the performance of the H-GPLVM using a simple, well defined weighting function. The weight of each particle is calculated from the sum of the squared 3D Euclidean distances between a set of 15 markers on the wrists, elbows, shoulders, feet, knees, hips, head, neck and pelvis of the hypothesised skeleton and the test skeleton. The skeletons are identical in size, estimated from the MoCap data of CMU subject 35 [1]. In § 5.2 and § 5.3 we obtain a set of 2D feature tracks for a subset of these 15 joint locations for unknown subjects in monocular video sequences. We again compare these with the skeleton of CMU subject 35, this time using the sum of squared 2D Euclidean distances. To facilitate this comparison we presume an orthographic camera projection and estimate by hand a single constant scaling factor to give reasonable agreement in the height of the subject and skeleton.

To provide a baseline for comparison, we also train a GPDM [28] on each training data sequence and perform APF in the resulting latent spaces in an approach similar to [19]. GPDMs are an extension of the GP-LVM that incorporate dynamics [28]. An extra GP is
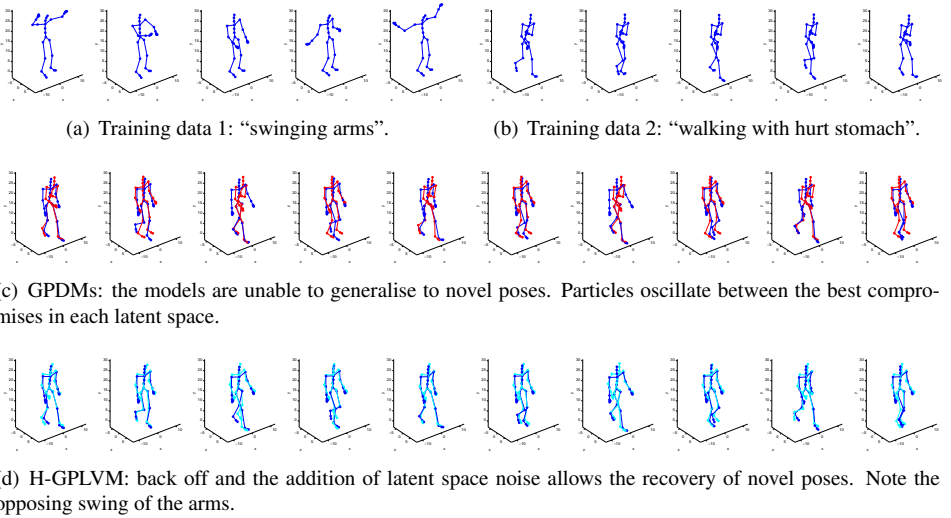
(a) Training data 1: "swinging arms".      (b) Training data 2: "walking with hurt stomach".



(c) GPDMs: the models are unable to generalise to novel poses. Particles oscillate between the best compromises in each latent space.



(d) H-GPLVM: back off and the addition of latent space noise allows the recovery of novel poses. Note the opposing swing of the arms.

Figure 2: MoCap training data [11] and resulting pose estimation results for a "walking" sequence using GPDMs and H-GPLVM.
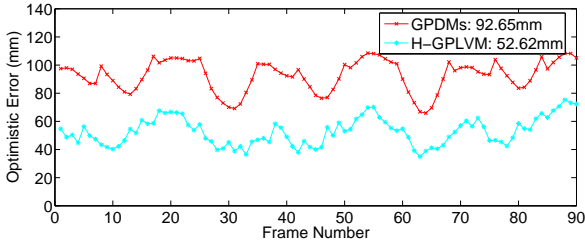


Figure 3: Optimistic errors for the "walking" sequence.

used to give a first order model of data dynamics in the latent space. The smooth latent space recovered by a GPDM is suitable for exploration with particle filtering techniques. We present qualitative results showing the expected joint angle pose

$$E\left[\mathbf{x}_t\right] = \sum_{n=1}^{N} \pi_t^{(n)} \mathbf{x}_t^{(n)} \qquad (2)$$

recovered by the particle set versus the ground truth test pose. In § 5.1 we also give quantitative results using optimistic error. Optimistic error was proposed for the cross comparison of particle-based techniques [4]. It gives a lower bound on error by finding the particle (pose) with the lowest average Euclidean distance between its 15 joint centres and those of the test skeleton. We also quote average weighted error scores, given by the error of every particle weighted by its normalised likelihood. All experiments used 100 particles and 10 annealing layers with a constant survival rate of $\alpha_r = 0.5$ and noise covariance of $\Sigma = \begin{pmatrix} 0.25 & 0 \\ 0 & 0.25 \end{pmatrix}$.
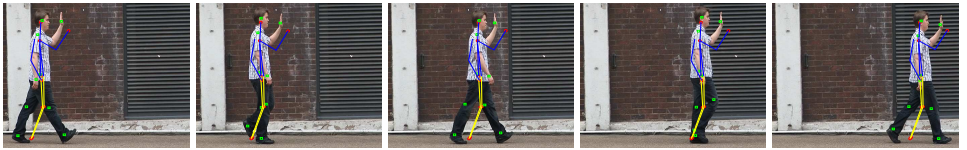
(a) GPDMs: neither latent space contains the pose. The waving hand is recovered at the expense of the legs.



(b) H-GPLVM: back off allows the combination of training data for accurate pose recovery.
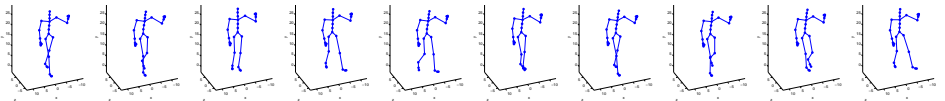


(c) H-GPLVM: inferred 3D poses from a different view point.

Figure 4: Pose estimation results using 2D WSL feature tracks from a monocular "walking whilst waving" sequence. Training data is "slow walk/stride" and "stand and wave".

## 5.1  MoCap Data: Novel Poses

We trained an H-GPLVM (shown in Fig. 1) using single 40 frame cycles of "swinging arms" (CMU file 86_07.amc) and "walking with hurt stomach" (CMU file 91_26.amc) activity sequences, before attempting to recover novel poses from a walking subject (CMU file 35_01.amc, 90 frames) using the 3D Euclidean distance weighting function. The required departure from the training data is quite considerable (see Fig. 2). The GPDMs were unable to recover the walking poses with the particle set oscillating between the latent spaces of the two activities with constant frequency, jumping from the least worst "swinging arms" pose to the least worst "walking with hurt stomach" pose, see Fig. 2(c).

In contrast, the H-GPLVM was able to optimise limbs independently recovering good pose estimates at every frame, see Fig. 2(d). The required subdivision of the skeleton operates at two scales. The lower body is recovered from "walking with hurt stomach" pose data and the upper body from "swinging arms". The upper body is then further subdivided between the two arms. While the arms swing *in phase* in the training data, they are uncoupled to give the *out of phase* opposing swing seen in the walking data. Error values for pose estimation are shown in Fig. 3. The H-GPLVM consistently outperforms the GPDMs with average optimistic error across the sequence of 52.6mm versus 92.7mm for the GPDMs (or in terms of weighted error, 62.0mm versus 93.3mm).

## 5.2  Monocular Image Data: Combined Poses

In order to test the H-GPLVM's ability to recover combined poses from 2D feature points we used the WSL tracker [10] to track 9 feature points on the body of a subject "walking whilst waving". These comprised the hands, feet, knees, head, right shoulder and pelvis locations at each frame (see green squares in Fig. 4). We trained an H-GPLVM using single cycles of "slow walk/stride" (CMU file 08_11.amc, 45 frames) and "stand and wave" (CMU file
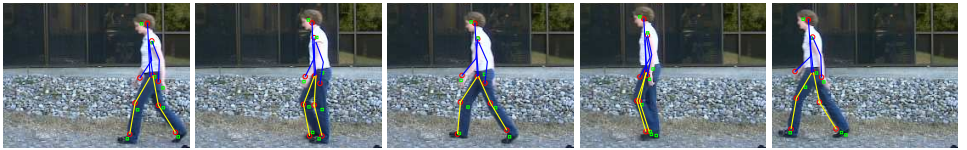
Figure 5: H-GPLVM: pose estimation for a walking sequence [20]. Position of occluded right arm is inferred from the visible upper body.

143_25.amc, 40 frames) activity sequences and attempted to recover the test poses using the 2D Euclidean distance weighting function. The tracking skeleton's pelvis was also allowed to translate horizontally to allow for a moving subject, and the extra particle parameter was dispersed with Gaussian noise and preserved between frames. Results for the H-GPLVM and the GPDM baseline are shown in Fig. 4. The baseline recovers the best possible candidate from the two GPDMs at each frame, this is a "stand and wave" pose at every instant. The H-GPLVM is able to combine walking poses for the lower body and right arm with a waving pose for the left arm to give good 3D pose reconstruction throughout, see Fig. 4(c).

## 5.3   Monocular Image Data: Poses with Occlusion

One advantage of learning global latent models of activity at the full-body scale is the ability to recover *known* poses given limited image evidence. For example, given a latent variable model learned from walking poses, walking sequences featuring occluded limbs have been reconstructed from a small set of 2D feature points [9, 25, 26]. In this section we show that our use of the H-GPLVM is "back compatible" with this kind of 2D to 3D reconstruction of partially occluded known poses performed by an unknown subject.

We trained an H-GPLVM using a single 45 frame cycle of "slow walk/stride" data (1 root node only) and attempted to reconstruct poses from the 2D WSL tracker data [10] used by Urtasun *et al.* [25] (see green squares in Fig. 5). In contrast to § 5.2, there is no data for the right arm and the right knee track is lost about half way through the sequence resulting in a challenging reconstruction problem. The placement of occluded limbs must be inferred from higher level correlations in the training data. In the case of a missing right arm, back off is terminated in the "upper body" node (see § 4.2). While the legs are independently optimised in the leaf nodes, the left arm, right arm and head are *jointly* optimised in the "upper body" node. The result is a right arm that is necessarily coordinated with the visible upper body.

Pose estimation results are shown in Fig. 5. To account for marker loss, we used a piecewise weighting function to give no further increase in the contributions of markers separated by 30cm or more to $w(\mathbf{z}, \mathbf{x})$. Despite an absence of image evidence for the right arm, well coordinated walking poses are recovered at each frame with the occluded right arm oscillating out of phase with the visible left arm. This long-range skeletal correlation is a benefit of the hierarchical approach; in a "flat" part-based model, an occluded limb would be free to randomly flail.

## 6   Future Work and Conclusions

This paper has outlined a middle ground between searching low-dimensional global pose models and searching in the original high-dimensional state space. This is achieved by de-

scending through a hierarchical part based model: the H-GPLVM. A final dispersion and resampling stage could be conducted in the full state space to reduce remaining pose errors, as suggested by Raskin *et al*. [19]. Future work will concentrate on implementing this final step to complete a principled route from low-dimensional global model (at the root nodes) to the full state space, and on compiling a compact set of complimentary "basis activities" that combine to give a single H-GPLVM capable of recreating a range of human poses.

At its top level the H-GPLVM is akin to a set of GPDMs, one for each activity. But by "backing off" to benefit from progressively greater independence between body parts, and by making increasingly discerning comparisons with image evidence, our stochastic search algorithm is able to recover novel pose configurations. We have shown how H-GPLVMs can be used to recover poses that are beyond the scope of other widely used global latent variable models such as the GPDM, using both 3D and 2D image evidence.

# 7 Acknowledgements

# References

[1] CMU graphics lab motion capture database. http://mocap.cs.cmu.edu/.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. on Sig. Proc.*, 50 (2):174–188, 2002.

[3] J. Bandouch, F. Engstler, and M. Beetz. Evaluation of hierarchical sampling strategies in 3D human pose estimation. In *BMVC*, 2008.

[4] A. O. Bălan, L. Sigal, and M. J. Black. A quantitative evaluation of video-based 3D person tracking. In *VS-PETS*, pages 349–356, 2005.

[5] J. Deutscher and I. Reid. Articulated body motion capture by stochastic search. *IJCV*, 61(2):185–205, 2005.

[6] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *IJCV*, 61(1):55–79, 2005.

[7] J. Gall, J. Potthoff, C. Schnörr, B. Rosenhahn, and H.-P. Seidel. Interacting and annealing particle filters: Mathematics and a recipe for applications. *JMIV*, 28(1):1–18, 2007.

[8] D. Gavrila and L. S. Davis. 3D model-based tracking of humans in action: a multi-view approach. In *CVPR*, pages 73–80, 1996.

[9] S. Hou, A. Galata, F. Caillette, N. Thacker, and P. Bromiley. Real-time body tracking using a Gaussian process latent variable model. In *ICCV*, 2007.

[10] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi. Robust online appearance models for visual tracking. *PAMI*, 25(10):1296–1311, 2003.

[11] N. D. Lawrence. Probabilistic non-linear principal component analysis with Gaussian process latent variable models. *JMLR*, 6:1783–1816, 2005.

[12] N. D. Lawrence and A. J. Moore. Hierarchical Gaussian process latent variable models. In *ICML*, 2007.

[13] M. Lee and I. Cohen. Proposal maps driven MCMC for estimating human body pose in static images. In *CVPR*, volume 2, pages 334–341, 2004.

[14] R. Li, M-H. Yang, S. Sclaroff, and T-P. Tian. Monocular tracking of 3D human motion with a coordinated mixture of factor analyzers. In *ECCV*, pages 137–150, 2006.

[15] Z. Lu, M. C. Perpinan, and C. Sminchisescu. People tracking with the Laplacian eigenmaps latent variable model. In *NIPS*, 2007.

[16] J. MacCormick and M. Isard. Partitioned sampling, articulated objects, and interface-quality hand tracking. In *ECCV*, pages 3–19, 2000.

[17] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *CVIU*, 104(2):90–126, 2006.

[18] D. Ramanan. Learning to parse images of articulated bodies. In *NIPS*, 2006.

[19] L. Raskin, E. Rivlin, and M. Rudzsky. Using Gaussian process annealing particle filter for 3D human tracking. *EURASIP J. on Adv. in Sig. Proc.*, 2008.

[20] H. Sidenbladh, M. J. Black, and L. Sigal. Implicit probabilistic models of human motion for synthesis and tracking. In *ECCV*, pages 784–800, 2002.

[21] L. Sigal, S. Bhatia, S. Roth, M. J. Black, and M. Isard. Tracking loose-limbed people. In *CVPR*, pages 421–428, 2004.

[22] C. Sminchisescu and A. Jepson. Generative modeling for continuous non-linearly embedded visual inference. In *ICML*, pages 759–766, 2004.

[23] C. Sminchisescu and B. Triggs. Estimating articulated human motion with covariance scaled sampling. *IJRR*, 22(6):371–391, June 2003.

[24] T. P. Tian, R. Li, and S. Sclaroff. Articulated pose estimation in a learned smooth space of feasible solutions. In *CVPR Learning Workshop*, 2005.

[25] R. Urtasun, D. J. Fleet, A. Hertzmann, and P. Fua. Priors for people tracking from small training sets. In *ICCV*, pages 403–410, 2005.

[26] R. Urtasun, D. J. Fleet, and P. Fua. 3D people tracking with Gaussian process dynamical models. In *CVPR*, pages 238–245, 2006.

[27] R. Urtasun, D. J. Fleet, and P. Fua. Temporal motion models for monocular and multi-view 3D human body tracking. *CVIU*, 104(2):157–177, 2006.

[28] J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models for human motion. *PAMI*, 30(2):283–298, 2008.