Learning Probabilistic Non-Linear Models for Tracking Complex Activities: Extended Experimentation

Angela Yao[*] ETH Zurich
Juergen Gall ETH Zurich
Luc Van Gool ETH Zurich
Raquel Urtasun TTI Chicago
{yaoa, gall, vangool}@vision.ee.ethz.ch, rurtasun@ttic.edu

Abstract

We provide here details on the effect of different distance functions for determining the neighborhood and compare them to random sampling of points for estimating the gradient. In addition, we provide the complete set of experiments on monocular tracking on the CMU dataset.

1 Neighborhood type and size

We tested three different distance measures (*'xL2'* - Euclidean distance in the latent space, *'yL2'* - Euclidean distance in the data space and *'temp'* - temporal neighbors) for determining the neighborhood $\mathcal{R}$ and compared them to random sampling of points for estimating the gradients $\frac{\partial \mathcal{L}}{\partial \mathbf{X}_R}$ and $\frac{\partial \mathcal{L}}{\partial \beta_R}$. We also compare the effect of neighborhood size, as determined by the number of points R used to compute the gradients. Finally we look at the effect of randomly subsampling R from a larger selection $\kappa \cdot N$ neighbors. We apply the learning algorithm to 4 motion capture sequences of walking (740 frames total) from a single subject. For the same initialization, we repeat the learning process ten times for the differing neighborhood types and sizes and compare the minimum, mean and maximum negative-loglikelihoods over the repetitions in Figure 1.

When we select the nearest R neighbors, learning on the GPLVM is poor; *'xL2'* outperforms *'yL2'* and *temp*, but in all three cases, the gradient estimates are too local to capture the more global shape of the latent space. If we subsample R neighbors from a larger neighborhood of $\kappa \cdot N$ neighbors, however, we find that the resulting negative-loglikelihood is much lower and there is little difference between the different types of distance measures. Randomly sampling R neighbors, on the other hand, while sufficient for estimating the latent space, is not as successful as maintaining some form of neighborhood.

2 Neighborhood Sampling

We show here the complete set of monocular tracking results for within-subject 2 and cross-subject 3 experiments as described in Section 3.1 of the paper.

[*] *This research was supported by the Swiss National Foundation NCCR project IM2, NSERC Canada and NSF #1017626.* Source code is available at `www.vision.ee.ethz.ch/yaoa`

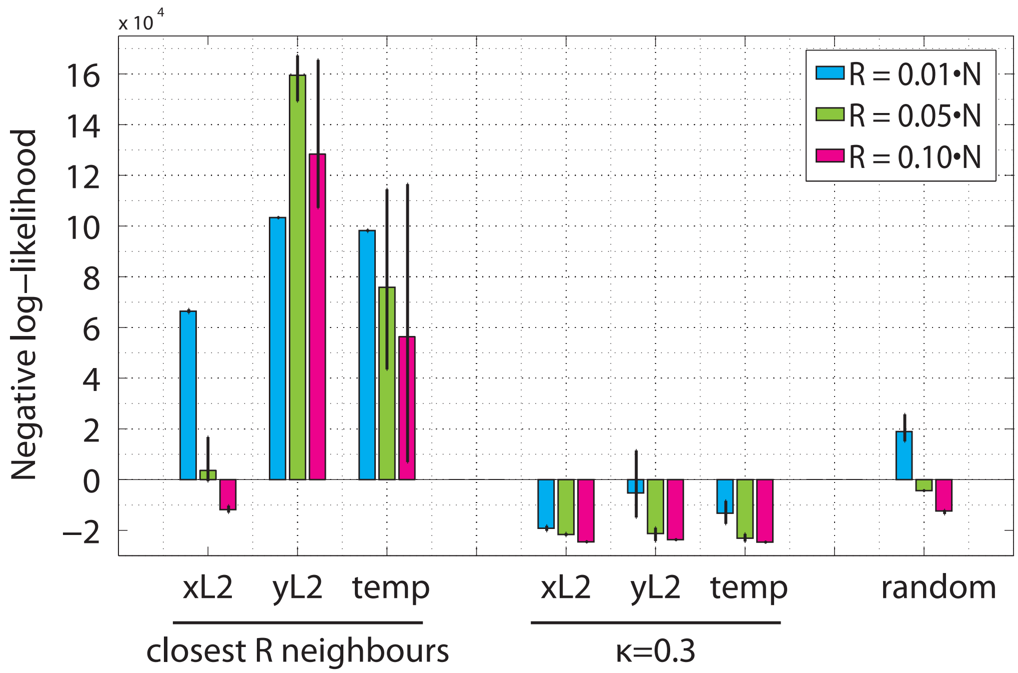

Figure 1: **Negative log-likelihood** (lower is better) of learning walking motions with different types and sizes of neighborhoods. Bars show the mean value over 10 runs, while the errorbars indicate the maximum and minimum value over runs. The different colored bars indicated differing sizes of R, where R is given as a fraction of the total number of training samples ($N = 740$). To maintain a fair comparison, we keep the number of iterations for learning, T, inversely proportional to R.

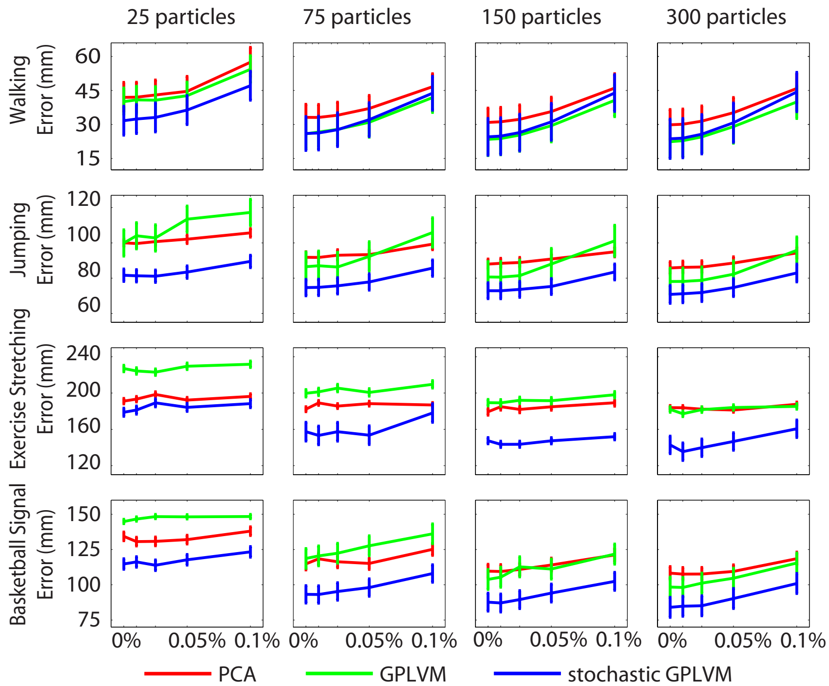

Figure 2: **Within-subject 3D tracking errors** for each type of activity sequence with respect to amount of additive noise for different number of particles, where error bars represent the standard deviation from repetitions runs.

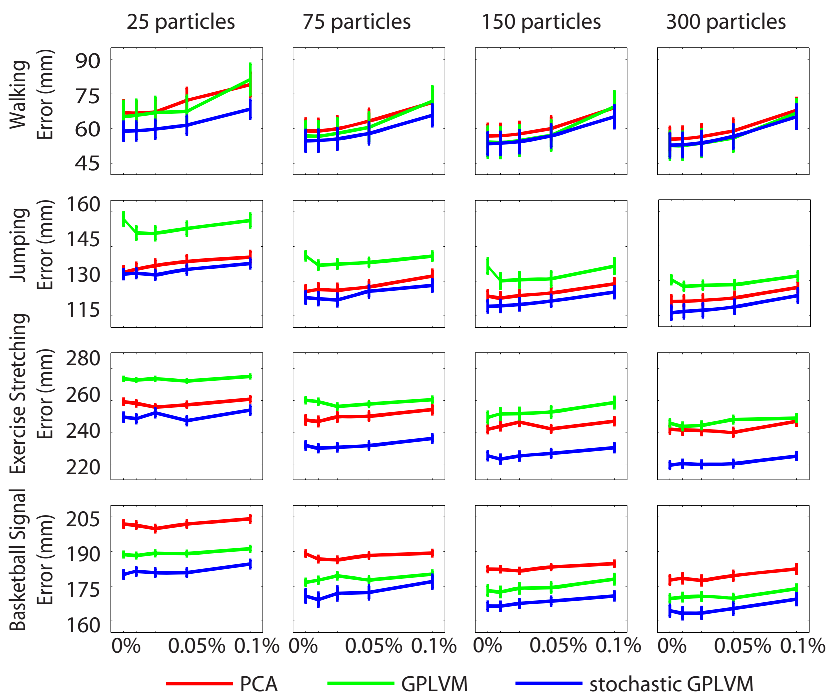

Figure 3: **Cross-subject 3D tracking errors** for each type of activity sequence with respect to amount of additive noise for different number of particles, where error bars represent the standard deviation from repetitions runs.