Tracking Generic Human Motion via Fusion of Low- and High-Dimensional Approaches

Yuandong Xu
xuyd@cis.pku.edu.cn
Jinshi Cui
http://www.cis.pku.edu.cn/vision/Visual&Robot/people/CuiJinshi
Huijing Zhao
http://www.cis.pku.edu.cn/faculty/vision/zhaohj
Hongbin Zha
http://www.cis.pku.edu.cn/vision/Visual&Robot/people/zha

Key Laboratory of Machine Perception (MOE),
Peking University, Beijing, China

Background. The algorithms for tracking generic human motion should be able to cope with the high-dimensional state space as well as to recover complex postures with various motion types and styles. Many approaches have been proposed to address these problems [1, 3, 5]. One kind of low-dimensional approaches that learn motion models by dimensionality reduction can successfully deal with the high-dimensional problem, but it only works on specific motion types with available training data. Other approaches which employ smart sampling directly on high-dimensional pose space don't have that limitation. However, this kind of methods is lack of robustness, with high computational cost, and hard to recover from failures.

Fusion Framework. In order to solve the aforementioned problems simultaneously, we propose a fusion formulation to integrate the two kinds of tracking approaches into one framework. Within the framework, two independent trackers with different algorithms proceed in parallel in different state spaces, and are fused according to a set of criteria at each time step. The fusion criteria ensure that the overall tracking performance is improved by concentrating the advantages of the two approaches and avoiding their weak points.

An overview of our fusion framework is shown in Figure 1.

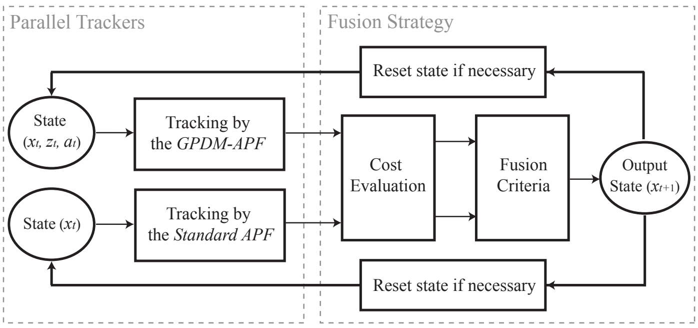

Figure 1: An overview of the fusion framework.

The Low- and High-Dimensional Approaches. In this fusion framework, we employ the standard annealed particle filtering (APF) [1] as the high-dimensional approach to track human motion in the original pose state space, because it proved to be relatively effective for human motion tracking and is often used as a baseline algorithm.

For the low-dimensional approach, we first learn the Gaussian Process Dynamical Models (GPDM)[5] for specific activities, and then track human motion by the APF algorithm in the low-dimensional spaces learned from the GPDM. The mixed-state CONDENSATION[2] is also used to provide a model-switching mechanism for tracking with multiple dynamic models. The low-dimensional approach is denoted as the GPDM-APF tracker.

Implementation of the two trackers is described in the paper as the details.

Fusion Criteria. The cost function is used to evaluate the performance of each tracker at each time step by comparing its expectation output with the image observation. We take the definition of the cost function from the symmetrical silhouette likelihood used in [4], as

$$Cost = \frac{1}{N}\sum_n \left(\frac{\sum_p (F(p)(1-M(p)))}{\sum_p (F(p))} + \frac{\sum_p (M(p)(1-F(p)))}{\sum_p (M(p))}\right). \quad (1)$$

where $F(p)$ denotes the observation foreground, $M(p)$ the silhouette of projection model, and N the number of camera views.

We denote $cost_{(SAPF)}$ as the cost of the standard APF tracker, and $cost_{(GPDM-APF)}$ as the cost of the GPDM-APF tracker. The fusion criteria for choosing the output and updating the trackers is set as follows:

The output is set as the expectation pose of the tracker with lower cost.

If $cost_{(SAPF)} - cost_{(GPDM-APF)}$ exceeds a threshold δ, the state of the particles of the standard APF tracker is updated by the expectation pose of the GPDM-APF tracker directly.

If $cost_{(GPDM-APF)} - cost_{(SAPF)}$ exceeds a threshold δ', a new state set, which contain poses, latent variables and activity labels, are generated by selecting the k nearest neighbor of the pose of the standard APF from the training dataset of each activity. Then, the state of the particles of the GPDM-APF tracker is reset with the new generated state set.

Results and Conclusions. In order to investigate the performance of our technical fusion approach to generic motion tracking, The proposed approach is tested on the *HumanEva-II Combo* dataset [4]. The experiments also conduct quantitative comparisons with the methods using only the standard APF or the GPDM-APF. Figure 2 provides an example of the results for the testing data produced by the three methods.

	Walking	Jogging	Balancing	Overall
Standard APF	86 ± 26 mm	91 ± 22 mm	105 ± 19 mm	94 ± 23 mm
GPDM-APF	72 ± 16 mm	82 ± 25 mm	180 ± 38 mm	112 ± 27 mm
Fusion	62 ± 10 mm	71 ± 9 mm	86 ± 9 mm	81 ± 15 mm

Figure 2: **The average error and standard deviation for the testing data produced by the standard APF, the GPDM-APF and the Fusion approach.**

As the results illustrate, the fusion approach not only incorporates the respective advantages of the low- and high-dimensional approaches, but also overcome their weakness. Therefore, the overall performance is better than any single approach. The fusion strategy is quit heuristic, but very easy to implement, extended and embedded into existing tracking systems.

[1] Jonathan Deutscher and Ian Reid. Articulated body motion capture by stochastic search. *International Journal of Computer Vision*, 61: 185–205, 2005.

[2] Michael Isard and Andrew Blake. A mixed-state condensation tracker with automatic model-switching. In *International Conference on Computer Vision (ICCV)*, 1998.

[3] Hedvig Sidenbladh, Michael J. Black, and D. J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *European Conference on Computer Vision (ECCV)*, pages 702–718, 2000.

[4] Leonid Sigal, Alexandru Balan, and Michael Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, 87:4–27, 2010.

[5] Raquel Urtasun. 3d people tracking with gaussian process dynamical models. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2006.