21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Fusion of Low-and High-Dimensional Approaches by Trackers Sampling for Generic Human Motion Tracking

Ye Liu, Jinshi Cui, Huijing Zhao and Hongbin Zha
Key Lab of Machine Perception (MOE), Peking University, China
{liuye,cjs,zhaohj,zha}@cis.pku.edu.cn

Abstract

Recently, fusion of low- and high-dimensional approaches shows its success in the generic human motion tracking. However, how to choose the trackers adaptively according to the motion types is still a challenging problem. This paper presents a trackers sampling approach for generic human motion tracking using both low- and high-dimensional trackers. Gaussian Process Dynamical Model(GPDM) is trained to learn the motion model of low-dimensional tracker, and it performs better on specific motion types. Annealed Particle Filtering(APF) shows its advantage in the tracking without limitation on motion types. We combine both of the two methods and automatically sample trackers according to the motion types that it is tracking on. To improve performance, trackers communication is adopt to keep the better state of trackers. The approach facilitates tracking of generic motions with low particle numbers.

1. Introduction

Techniques that employ smart sampling in the high-dimensional state space have been widely used in human motion tracking, this kind of approach does not have limitations on motion types, but it lacks stability and has high computational cost. In this paper, we adopt the Annealed Particle Filter (APF) as the high-dimensional tracking tracker.

Another kind of approaches employs the learned motion models to deal with the high dimensional problems of human motion tracking. They perform better on specific motion types, but cannot extend to other generic motions. Gaussian Process Dynamical Model(GPDM) [6] is one of the most effective approaches to learn the motion model. For tracking, we adopted a similar framework as the GP-APF [4], and tracked in the low-dimensional space.

Recently, fusion approaches [2][7] which integrate low- and high-dimensional approaches for tracking generic human motions have gained researchers' attentions. [2] defined three activity models(i.e. unknown activities, known activities, known activity transitions), designed different tracking approaches for each of them and combined these approaches into a multiple activity model APF(MAM-APF) framework. And APF is employed for searching the optimal mode. It obtains a relative good performance on tracking generic motion sequences. But when the number of "known" activities increases, the dimensionality of the joint-activity space must grow and its computational cost will increase at the same time. Instead of modeling this joint activity space, [7] proposed a method to integrate the model learning approach (GPDM-APF) with the standard APF into one framework. The two parallel trackers run separately, and they are fused by a set of criteria. This makes the system choose the tracker which achieved better performance as the output. By fusion of the two trackers, the tracking system outperforms any system that uses the single approach. But it costs a lot of time when the two trackers run in parallel. To save time and resources, we propose a trackers sampling method, which could automatically sample tracker according to the motion types that it is tracking on.

Inspired by earlier work [7], we extend a previous approach by using a trackers sampling scheme. In this framework, system tracks human motions robustly by searching for the appropriate trackers in each frame, and the number of particles for each tracker varies according to its tracking performance. By trackers sampling, when tracking on trained activities, the system automatically switches to the low-dimensional approaches (GPDM-APF), and most of particles are propagated by GPDM-APF. While tracking on un-trained motions, the high-dimensional approaches(standard APF) takes over the tracking, then more particles are assigned to APF. By the sampling procedure, system can "sample the best trackers adaptively from a tracker space to the current situation" [3]. Fig.1 illustrates the trackers sampling

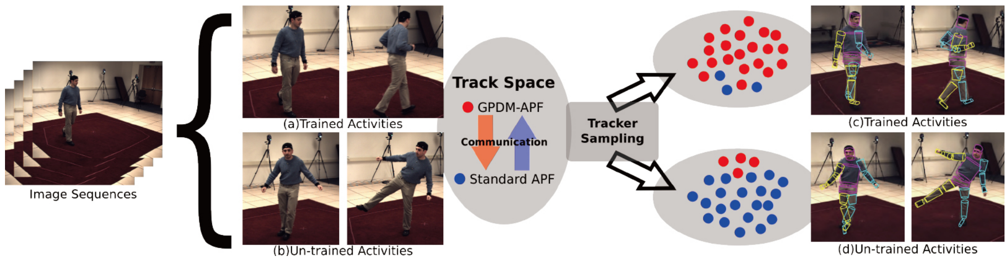

Figure 1. Illustration of trackers sampling procedure. For pre-trained activities in *HumanEva-II S2 Combo* **((a) Walking or Jogging), sampling methods will assign more particles for GPDM-APF. While for untrained motions ((b) Balancing), more particles are assigned to APF and it will take over tracking. Sampling procedure is based on tracker's performance.**

procedure.

Our main contributions are

- A joint state space is adopted to integrate both low- and high-dimensional states in order to maintain particles in both state spaces.
- A joint dynamic model is utilized to sample the trackers in both low and high dimensional spaces, and this joint dynamic model allows both types of particles' propagations.
- The trackers communicate with each other in a probabilistic way to utilize particles from the other tracker.

2. Tracking Model

2.1 Tracker Space

The trackers space we used in this paper contains two independent trackers. The first tracker (denoted by GPDM-APF [4]) employs the learned motion models to track human motions. The second tracker takes the standard APF [1] algorithm to recover human poses in the high-dimensional pose space.

2.2 Joint State Space

In this trackers sampling framework, a joint state space is adopted. In this joint space, at time step t, the state X_t can be decomposed as: $\{\{X_t^l, Z_t^l, A_t^l\}, \{X_t^h\}\}$. Where $\{X_t^l, X_t^l, A_t^l\}$ is the state for the low-dimensional approach, and $\{X_t^h\}$ is the state for high-dimensional approach. Given the state X_t at time t, and the observation $Y_{1:t}$ up to time t, our goal is to estimate the posteriori probability $(X_t|Y_{1:t})$.

2.3 Joint Dynamic model

In this joint state space, the posteriori probability $p(X_t|Y_{1:t})$ updates with the following formula:

$$p(X_t|Y_{1:t}) \propto p(Y_t|X_t) \int p(X_t|X_{t-1})p(X_{t-1}|Y_{1:t-1})dX_{t-1} \quad (1)$$

To propagate the particles in the joint state space, we propose a probabilistic fusion method to integrate the two kinds of dynamic models into one. Given the result of tracker selection at time t, T_t^i(where $T_t^i = 0, 1$ when $i = h, l$), the joint dynamic model $p(X_t|X_{t-1})$ can be decomposed as

$$\begin{aligned} p(X_t|X_{t-1}) = p(X_t|T_t^h, X_{t-1})p(T_t^h|X_{t-1}) + \\ p(X_t, Z_t|T_t^l, X_{t-1}, Z_{t-1})p(T_t^l|X_{t-1}) \end{aligned} \quad (2)$$

where $p(X_t|T_t^i, X_{t-1})$ represents the i-th decomposed dynamic model and $p(T_t^i|X_{t-1})$ devotes to the trackers' probability. In this work, the dynamic model of standard APF $p(X_t|T_t^h, X_{t-1})$ uses the addition of Gaussian noise with covariance σ_{h_dyn} to approximate.

$$p(X_t|T_t^h, X_{t-1}) = N(X_t - X_{t-1}, \sigma_{h\_dyn}) \quad (3)$$

The dynamic model of GPDM-APF can be represented as

$$p(X_t, Z_t|T_t^l, X_{t-1}, Z_{t-1}) = p(X_t|Z_t)p(Z_t|Z_{t-1}) \quad (4)$$

where $p(X_t|Z_t)$ is the mapping from the latent variable space to high-dimensional pose space in (5)

$$X = f_{gp}(Z) \quad (5)$$

and $p(Z_t|Z_{t-1})$ is the temporal dynamics in the latent variable space in (6), which are both learned from GPDM [6].

$$p(Z_t|Z_{t-1}) = N(Z_{t-1} + (f_{gp\_dyn}(Z_{t-1}) - Z_{t-1})\Delta_T, \sigma_{l\_dyn}) \quad (6)$$

For the trackers' probability $p(T_t^i|X_{t-1})$, we use the average of particle weights to approximate it.

$$p(T_t^i|X_{t-1}) \approx \sum_j (\omega_j)/N_i \quad (7)$$

where ω_j is the weight of the j-th particle for tracker T_t^i in time step t, and N_i is the number of the particles for tracker T_t^i.

2.4 Trackers Sampling

The trackers sampling scheme contains obtaining the samples of trackers and then, given the sampled trackers, getting the samples for states. At each time step t, trackers are sampled according to trackers' probability $p(T_t^i|X_{t-1})$, the number of particles that assigned to each tracker is proportional to the trackers' probability $p(T_t^i|X_{t-1})$.

For the validation step, each particle is evaluated by the weighting function (8) that defined by Sigal et al. [5].

$$weight = exp - (\frac{1}{N}\sum_n(\frac{\sum_p(F(p)(1-M(p)))}{\sum_p(F(p))} + \frac{\sum_p(M(p)(1-F(p)))}{\sum_p(M(p))})). \tag{8}$$

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

With the updated particle sets, the trackers' sampling probability $p(T_t^i|X_t)$ at next time step is recalculated by (7). The number of particles that assigned to each tracker is redistributed. Suppose the number of particles in the joint space is N_s, then the number of particles assigned for each tracker in the next time step is updated with the following formula

$$N_i = p(T_t^i|X_t) * N_s \tag{9}$$

In this way, the tracker is sampled according to the trackers' probability, which is measured by tracking performance, and the system automatically assigns more particles for the tracker that has has better performance. So this could save a lot of time and resources for the tracking.

2.5 Trackers Communication

The trackers sampling mechanism can perform well in generic motion tracking, especially in switching between different kind of motions. However, when switches from trained motions to un-trained motions, or verse, some better states of trackers may be lost due to the redistribution of particles. Inspired by [3], we adopt a trackers communication mechanism. There are two stages for this communication step: parallel and interaction. At parallel stage, the tracker keeps its own states, while at interaction stage, the tracker updates its states with the other tracker's. And in each time step, which stage to stay depends on a probability. Here we use $p(T_t^i|X_t)$ to approximate this probability. The tracker's state S_i accepts the other's state S_j as its own state with

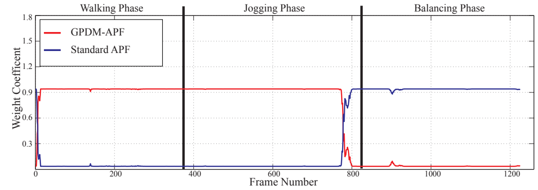

Figure 2. Weight coefficients of two trackers without communication on tracking *HumanEva-II S2*.

the following probability

$$\pi_i = \frac{p(T_t^i|X_t)}{\sum_{j=l,h}^{2} p(T_t^j|X_t)} \tag{10}$$

The details of the algorithm are stated as follows

Algorithm 1 Trackers Communication

Require: Particle Set of $Tracker_l$ S_l, Particle Set of $Tracker_h$ S_h
Ensure: Final Particle set S_l', S_h'
In each frame
Compute π_l, π_h
for each particle in particle set S_l **do**
  $r = Rand()$;
  if $r < \pi_l$ **then**
    keep its own particle (parallel);
  else
    update its particle with mean value of the particle set S_h (interaction);
  end if
end for
for each particle in particle set S_h **do**
  $r = Rand()$;
  if $r < \pi_h$ **then**
    keep its own particle (parallel);
  else
    update its particle with mean value of the particle set S_l (interaction);
  end if
end for

3 Experiments

In order to investigate the performance of our trackers sampling method in generic motion tracking, we designed two experiments to show the effectiveness of our approach.

3.1 Evaluations on Trackers Sampling Method

We tested the trackers sampling method with and without communication on the *HumanEva-II S2 Combo*

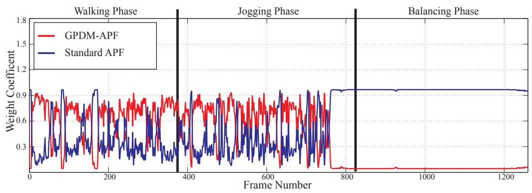

Figure 3. Weight coefficients of two trackers with communication on tracking ***HumanEva-II S2*****.**

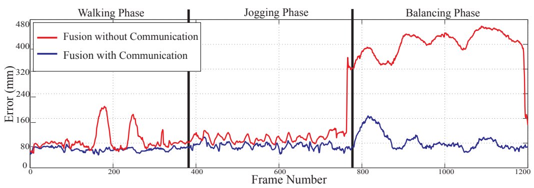

Figure 4. Performance comparison between the trackers sampling method with and without communication on tracking ***HumanEva-II S2*****.**

motion sequence, and 4 cameras are used for computing likelihood. We assigned 160 particles for the joint space and 3 annealing layers for the joint dynamic model. From Fig.2, we can see that the system can automatically sample trackers from Jogging to Balancing. But with less communication between trackers, the APF could not achieve a desired tracking performance. It was because that APF outperformed GPDM-APF in this switching phrase, but it still had a relative higher tracking error. Fig.3 showed weight coefficients of the two trackers with communication and Fig.4 illustrated performance comparison between trackers sampling with and without communication. As we can see from Fig.4, overall performance was improved, especially for Balancing phrase.

3.2 Comparison with Fusion Method [7]

We compared our approach with fusion method in [7] on tracking *HumanEva-II S2*. Both two methods were assigned with 160 particles, 3 annealing layers. Fig. 5 showed the performance comparison of the two methods. As it illustrated, the trackers sampling method outperformed fusion method in [7], especially in the Balancing phrase. This is because in the Balancing phrase, the sampling approach assigned nearly all particles to APF, while in [7], the number was fixed and was just assigned half of particles. That is to say, our method needed less particles in order to get same performance. So it saved time and resources.

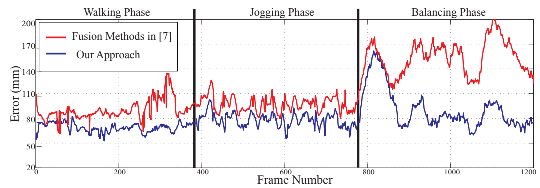

Figure 5. Performance comparison between our approach and fusion methods in [7] on tracking ***HumanEva-II S2*****.**

4 Conclusions

In this paper, we proposed a trackers sampling framework for the generic human motion tracking. The system can sample trackers automatically according to the motion types that is tracking on. To improve overall performance, a communication mechanism was also utilized, with which each tracker updated the other's state with a probability and then its own performance was improved. Experimental results showed that our approach can adapt itself automatically to motion types and outperformed other fusion approaches.

5 Acknowledgement

This research was funded by National Basic Research Program of China (973 Program) 2011CB302202, NSFC Grant (No.91120004) and NSFC-ANR Grants (No.61161130528).

References

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

[2] B. L. John Darby and N. Costen. Tracking human pose with multiple activity models. *Pattern Recognition*, pages 3042–3058, 2010.

[3] J. Kwon and K. M. Lee. Tracking by sampling trackers. In *ICCV*, 2011.

[4] L. Raskin, E. Rivlin, and M. Rudzsky. Using gaussian process annealing particle filter for 3d human tracking, 2008.

[5] L. Sigal, A. Balan, and M. Black. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *IJCV*, 87:4–27, 2010.

[6] J. M. Wang, D. J. Fleet, and A. Hertzmann. Gaussian process dynamical models. In *In Adv. Neural Info. Processing Systems*, pages 1441–1448.

[7] Y. Xu, J. Cui, H. Zhao, and H. Zha. Tracking generic human motion via fusion of low- and high-dimensional approaches. In *British Machine Vision Conference (BMVC)*, 2011.