Hierarchical Annealed Particle Swarm Optimization for Articulated Object Tracking

Xuan Son Nguyen, Séverine Dubuisson, Christophe Gonzales

Laboratoire d'Informatique de Paris 6 (LIP6/UPMC)
4 place Jussieu, 75005 Paris, FRANCE

Abstract. In this paper, we propose a novel algorithm for articulated object tracking, based on a hierarchical search and particle swarm optimization. Our approach aims to reduce the complexity induced by the high dimensional state space in articulated object tracking by decomposing the search space into subspaces and then using particle swarms to optimize over these subspaces hierarchically. Moreover, the intelligent search strategy proposed in [20] is integrated into each optimization step to provide a robust tracking algorithm under noisy observation conditions. Our quantitative and qualitative analysis both on synthetic and real video sequences show the efficiency of the proposed approach compared to other existing competitive tracking methods.

Keywords: particle filter, articulated object tracking, PSO

1 Introduction

Tracking articulated structures with accuracy and within a reasonable time is challenging due to the high complexity of the problem to solve. For this purpose, various approaches based on particle filtering have been proposed. Among them, one class addresses the complexity issue by reducing the dimensionality of the state space. For instance, some methods add constraints (e.g., physical) to the mathematical models [4, 13], to the object priors [7] or to their interactions with the environment [11]. Relying on the basic assumption that some body part movements are mutually dependent, some learning-based approaches [16, 19] reduce the number of degrees of freedom of these movements.

Alternatively, a second class of methods has been proposed in the literature [5, 9, 12, 14, 17, 18] whose key idea is to decompose the state space into a set of small subspaces where particle filtering can be applied: by working on small subspaces, sampling is more efficient and, therefore, fewer particles are needed to achieve a good performance. Finally, in the class of the optimization-based methods, the approach is to optimize an objective function corresponding to the matching between the model and the observed image features [3, 6, 8]. Recently, Particle Swarm Optimization (PSO) has been reported to perform well on articulated human tracking [10, 20]. Its key idea is to apply evolutionary algorithms inspired from social behaviors observed in wildlife to make the particles evolve following their own experience and the experience of the global population.

In this paper, our approach consists in decomposing the search space into subspaces of smaller dimensions and, then, in exploiting the approach proposed in [20] to search within these subspaces in a hierarchical order. A hierarchical particle swarm optimization has also been introduced in [10]. The main difference between this approach and ours is that we incorporate the sampling covariance and the annealing factor into the update equation of PSO at each optimization step to tackle the problem of noisy observations and cluttered background.

The paper is organized as follows. In Section 2, we briefly recall PSO. Section 3 presents the proposed algorithm. Section 4 reports the results of our experimental evaluation. Finally, Section 5 gives some conclusions and perspectives.

2 Particle Swarm Optimization (PSO)

Let $\mathcal{X}$ denote the state space: our goal is to search for state $\mathbf{x} \in \mathcal{X}$ that maximizes a *cost function* $f : \mathcal{X} \to \mathbb{R}$, with $\mathbf{a} \leq \mathbf{x} \leq \mathbf{b}$. A swarm consists of N particles, each one representing a candidate state of the articulated object. Denote $\mathbf{x}_{(m)}^{(i)}$ the ith particle at the mth iteration. $\mathbf{x}_{(m)}^{(i)}$ is decomposed into K (object) parts, i.e., $\mathbf{x}_{(m)}^{(i)} = \{\mathbf{x}_{(m)}^{(i),1}, ..., \mathbf{x}_{(m)}^{(i),K}\} \in \mathcal{X}$. Unlike evolutionary algorithms, to each particle in PSO is assigned a velocity $\mathbf{v}_{(m)}^{(i)} = \{v_{(m)}^{(i),1}, ..., v_{(m)}^{(i),K}\} \in \mathcal{X}$ and each particle has the ability to memorize its best state computed so far $\mathbf{s}^{(i)} = \{\mathbf{s}^{(i),1}, ..., \mathbf{s}^{(i),K}\} \in \mathcal{X}$. Let $\mathbf{s}^g$ be the current global best state, i.e., $\mathbf{s}^g = \text{Argmax}\{f(\mathbf{s}^{(i)})\}_{i=1}^N$. The evolution of the particles in PSO is described by the following equations:

$$\mathbf{v}_{(m)}^{(i)} = w\mathbf{v}_{(m-1)}^{(i)} + \beta_1 r_1(\mathbf{s}^{(i)} - \mathbf{x}_{(m-1)}^{(i)}) + \beta_2 r_2(\mathbf{s}^g - \mathbf{x}_{(m-1)}^{(i)}) \quad (1)$$

$$\mathbf{x}_{(m)}^{(i)} = \mathbf{x}_{(m-1)}^{(i)} + \mathbf{v}_{(m)}^{(i)} \quad (2)$$

where β_1, β_2 are constants, $r_1, r_2 \sim U(0,1)$ are random numbers drawn from a uniform distribution, w is the inertia weight and $w\mathbf{v}_{(m-1)}^{(i)}$ is the inertial velocity.

PSO has the ability to balance between the local and global search strategies of particles by setting the appropriate values for constants β_1, β_2 and inertia weight w. A large inertia weight results in an exploration of the search space (global search) while a small inertia weight limits the search around the globally best particle (local search). The value of the inertia weight can be fixed as a constant or adaptively changed throughout the search.

In the next section, we introduce our approach, inspired from PSO, and dedicated to articulated object tracking in cluttered environments.

3 Proposed approach

We propose to exploit the hierarchical nature of the kinematic structure of the articulated object to improve tracking. First, the state space of the target object is decomposed into lower dimensional subspaces. Then, optimal states are searched for in these subspaces in the hierarchical order of the kinematic structure using Partitioned Sampling (PS) [12]. These optimal states are then used to constrain the search in the next subspaces in the hierarchical order.

At time t, let $\mathbf{x}_t^{(i),k}$ (resp. $\mathbf{s}_t^{(i),k}$) denote the kth substate of the ith particle $\mathbf{x}_t^{(i)}$ (resp. the ith particle's best state $\mathbf{s}_t^{(i)}$) and let $\mathbf{s}_t^{\mathbf{g},k}$ be the kth substate of the global best state. Then, at the mth iteration, $\mathbf{x}_{t,(m)}^{(i)} = \{\mathbf{x}_{t,(m)}^{(i),1}, ..., \mathbf{x}_{t,(m)}^{(i),K}\}$, $\mathbf{v}_{t,(m)}^{(i)} = \{v_{t,(m)}^{(i),1}, ..., v_{t,(m)}^{(i),K}\}$ and $\mathbf{s}_{t,(m)}^{(i)} = \{\mathbf{s}_{t,(m)}^{(i),1}, ..., \mathbf{s}_{t,(m)}^{(i),K}\}$. We follow the approach proposed in [20], except that the state and velocity update equations for each subpart k are written as follows:

$$v_{t,(m)}^{(i),k} = r_0 P_{(m-1)} + \beta_1 r_1 (\mathbf{s}_t^{(i),k} - \mathbf{x}_{t,(m-1)}^{(i),k}) + \beta_2 r_2 (\mathbf{s}_t^{\mathbf{g},k} - \mathbf{x}_{t,(m-1)}^{(i),k}) \quad (3)$$

$$\mathbf{x}_{t,(m)}^{(i),k} = \mathbf{x}_{t,(m-1)}^{(i),k} + v_{t,(m)}^{(i),k} \quad (4)$$

$P_{(m-1)} = \alpha_0 * P_{(m-2)}$, $m \geq 2$, is the sampling covariance, with α_0 a constant, and $P_{(0)}$ is a covariance matrix whose diagonal elements are fixed with respect to the model configuration parameters. We propose to compute factors β_1 and β_2 at each iteration m using the annealing principle so that:

$$\beta_1 = \beta_2 = \beta_0 \beta_{max} \left(\frac{\beta_{max}}{\beta_{min}} \right)^{-\frac{m}{M}} \quad (5)$$

where $\beta_0, \beta_{max}, \beta_{min}$ are constants, $0 < \beta_0 \leq 1$, and M is the maximal number of iterations.

By combining PSO and hierarchical search, our approach aims to increase the tracking accuracy and to reduce the computational cost of the tracking algorithm by integrating the benefits of both methods. First, the search efficiency is improved by performing PSO within lower dimensional subspaces, thereby increasing tracking accuracy. Second, since the search is performed in the same way as PS, the number of particles required and thus the computational cost of the tracking algorithm is greatly reduced. Our proposed Hierarchical Annealed based Particle Swarm Optimization Particle Filter (HAPSOPF) is described in Algorithm 1, where $\bar{\mathbf{x}}$ is the estimated state of the object at time slice t, $w(., \mathbf{y})$ is the cost function to be optimized by PSO, and $\mathbf{y}$ is the current observation.

4 Experimental results

We compare our approach with APF [6], PSAPF [2], APSOPF [20] and HPSO [10]. The *cost function* $w(\mathbf{x}_{t,(m)}^{(i),k}, \mathbf{y})$ to be optimized by PSO measures how well a state hypothesis $\mathbf{x}_{t,(m)}^{(i),k}$ matches the true state w.r.t. the observed image $\mathbf{y}$, and is constructed using histogram and foreground silhouette [6]. An articulated object is described by a hierarchy (a tree) of parts, each part being linked to its parent in the tree by an articulation point. For instance, in the top row of Fig. 1, the blue polygonal parts are the root of the tree and the colored rectangles are

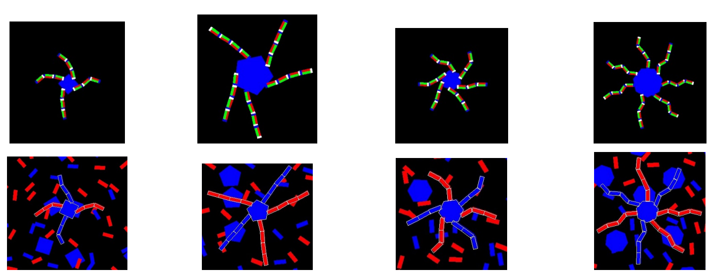
Input: $\{\mathbf{s}_{t-1}^{(i)}\}_{i=1}^{N}$, α_0, β_0, β_{max}, β_{min}, $P_{(0)}$, M (number of iterations)
Output: $\{\mathbf{s}_t^{(i)}\}_{i=1}^{N}$
for $k = 1$ **to** K **do**
  Sample: $\mathbf{x}_{t,(0)}^{(i),k} \sim \mathcal{N}(\mathbf{s}_{t-1}^{(i),k}, P_{(0)})$, $i = 1, \ldots, N$
  for $m = 0$ **to** M **do**
    if $m \geq 1$ **then**
      Compute $P_{(m)}$ and update β_1, β_2
      Carry out the PSO iteration based on Eq. (3) and (4)
    Evaluate: $f(\mathbf{x}_{t,(m)}^{(i),k}) = w(\mathbf{x}_{t,(m)}^{(i),k}, \mathbf{y})$, $i = 1, \ldots, N$
    Update $\{\mathbf{s}_t^{(i),k}\}_{i=1}^{N}$ and the k-th part of the global best state $\mathbf{s}_t^{\mathbf{g},k}$
  Evaluate and normalize the particle weights: $\pi_t^{(i)} = w(\mathbf{s}_t^{(i),k}, \mathbf{y})$, $\bar{\pi}_t^{(i)} = \frac{\pi_t^{(i)}}{\sum_{j=1}^{N} \pi_t^{(j)}}$
return $\{\mathbf{s}_t^{(i)}\}_{i=1}^{N}, \bar{\mathbf{x}} = \sum_{i=1}^{N} \bar{\pi}_t^{(i)} \mathbf{s}_t^{(i)}$

Algorithm 1: Our HAPSOPF algorithm.

the other nodes of the tree. The root is described by its center (x, y) and its orientation θ whereas the other parts are only characterized by their angle θ. For all algorithms, particles are propagated using a random walk with standard deviations fixed to $\sigma_x = 2$, $\sigma_y = 2$ and $\sigma_\theta = 0.05$. For APSOPF and HAPSOPF, $P_{(0)}$ is a diagonal matrix with the values of σ_x, σ_y and σ_θ. Our comparisons are based on two criteria: estimation errors and computation times.

4.1 Tests on synthetic sequences

Video sequences. We have generated two sets of various synthetic video sequences composed of 200 frames of 640×480 pixels (with ground truth). The video sequences in the first set contain no noise while, in the second set, cluttered background was generated to demonstrate the robustness of the proposed approach. The clutter is made up of polygons and rectangles randomly positioned



(a)

(b)

$L_a = 3$, $N_a = 4$ $L_a = 4$, $N_a = 5$ $L_a = 3$, $N_a = 6$ $L_a = 4$, $N_a = 7$

Fig. 1: Synthetic video sequences used for quantitative evaluation (number of arms N_a, length of arms L_a): (a) without clutter and (b) with clutter.

in the image. An articulated object is defined by its number N_a of arms, and their length L_a: some examples are given in Fig. 1.

Quantitative tracking results. The tracking errors are given by the sum of the Euclidean distances between each corner of the estimated parts and their corresponding corner in the ground truth. We used $M = 3$ layers for PSAPF and APF since it produces stable results for both algorithms, and $M = 3$ maximal iterations for HAPSOPF, HPSO and APSOPF. Table 1 gives the performances of the tested algorithms for sequences without or with noise (cluttered background). In our experiments, tracking in noisy sequences is challenging due to the background. In such cases, the annealing factor helps the particle swarm to follow its own searching strategy without being affected by any wrong guide of the local or global best states. On the contrary, the annealing process of PSAPF forces the particle set to represent one of the modes of the cost function, which causes some parts of the object to get stuck in wrong locations. This problem of annealing approaches was reported in [1]. Moreover, the use of the sampling covariance instead of the inertial velocity of Eq. (1) leads to an efficient exploration of the search space without losing the searching power of PSO. This is validated by our experiments on sequences without cluttered background, where our approach outperforms all the other ones. Fig. 2 gives comparative convergence results (error depending on the number N of particles) and computation times for a synthetic sequence (behaviors are similar for other sequences). Note that our approach converges better and faster than the other methods.

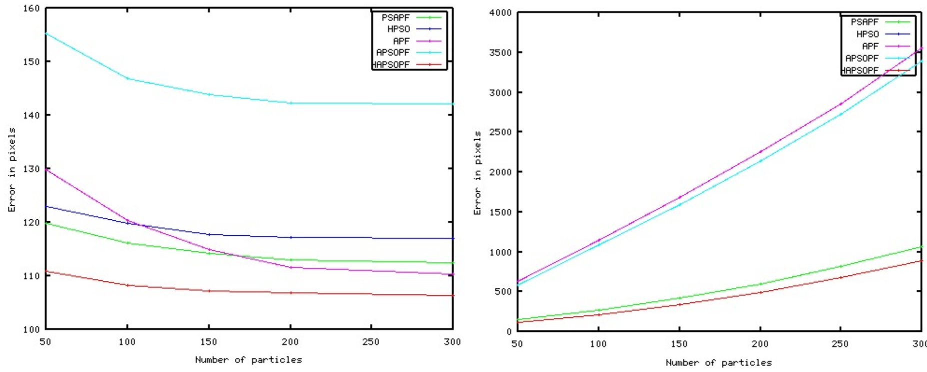
4.2 Tests on real sequences

Dataset. We used sequences *S1 Gesture* and *S2 Throwcatch* of the HumanEva-I dataset [15] that include ground truths, thus allowing to evaluate quantitatively our approach. For both sequences, the lower right hands of the subject move quickly, which makes them difficult to track. Moreover, *S2 Throwcatch* contains self-occlusions (hands and torso, left and right hands, left and right legs).

The searching order for PSAPF, HPSO, and HAPSOPF is: torso, head, left thigh, right thigh, left upper arm, right upper arm, left leg, right leg, left forearm,

Table 1: Tracking errors in pixels (average over 30 runs) and standard deviations for synthetic video sequences, N is the number of particles used per filter.

		$N_a = 4, K = 3$		$N_a = 5, L_a = 4$		$N_a = 6, L_a = 3$		$N_a = 7, L_a = 4$	
	N	50	200	50	200	50	200	50	200
HAPSOPF	without noise	110(2)	106(1)	214(5)	195(2)	243(11)	211(9)	312(7)	271(4)
	noise	204(39)	143(10)	227(56)	175(30)	322(67)	295(60)	553(194)	516(180)
PSAPF	without noise	120(2)	114(1)	238(6)	208(4)	251(7)	218(3)	319(8)	278(4)
	noise	309(109)	221(94)	281(78)	219(48)	432(86)	388(75)	1008(232)	914(213)
HPSO	without noise	125(5)	119(2)	252(9)	227(5)	254(11)	213(6)	382(5)	315(3)
	noise	277(78)	194(65)	245(42)	201(26)	345(27)	295(10)	922(334)	731(259)
APSOPF	without noise	184(3)	169(2)	260(12)	241(10)	265(15)	257(12)	471(30)	439(21)
	noise	254(16)	227(8)	308(33)	291(25)	490(68)	474(47)	817(223)	785(169)
APF	without noise	128(3)	109(2)	246(11)	221(9)	270(13)	236(11)	487(35)	412(24)
	noise	272(9)	258(5)	322(29)	309(18)	440(51)	429(40)	613(174)	592(156)



(**a**) (**b**)

Fig. 2: Comparison tests for convergence and computation time when tracking the object $N_a = 4, L_a = 3$: (a) convergence and (b) computation times (HPSO and our approach give same curves) in seconds.

right forearm. For a fair comparison, we fixed the number of evaluations of the weighting function at each frame for all the algorithms to 2000, and tuned parameters {N,M} for each method so that they achieve the best performance while satisfying the above constraint: $\{400, 5\}$ for APF, $\{40, 5\}$ for PSAPF, $\{200, 10\}$ for APSOPF and $\{20, 10\}$ for HPSO and HAPSOPF.

Quantitative tracking results. We used the evaluation measure proposed in [15], which is based on Euclidean distances between 15 virtual markers on the human body. Table 2 provides tracking errors and computation times. As can be observed, our approach has the same computation time as HPSO but reduces the estimation error and it outperforms the other approaches on both criteria.

Fig. 3 provides qualitative tracking results. Our approach always outperforms PSAPF and HPSO in cases of self-occlusions (frames 275, 523) or quick movements (frames 160, 387), showing its robustness. Because our approach incorporates the annealing into each searching stage of the hierarchical search, the problem of noisy observations is effectively alleviated. This makes our approach more robust to self-occlusions. The sampling covariance also helps to improve the searching effectiveness by shifting the particle swarm toward more promising regions.

Table 2: Tracking errors for full body in pixels (average over 30 runs).

	HAPSOPF		PSAPF		HPSO		APSOPF		APF	
	Error	Time	Error	Time	Error	Time	Error	Time	Error	Time
`S1 Gesture`	95(6)	287	99(11)	293	101(9)	287	102(4)	1348	105(2)	1412
`S2 Throwcatch`	212(10)	557	227(19)	579	232(12)	557	235(7)	2070	240(5)	2184

Fig. 3: Tracking results for frames 123,160,275,387,488,523: HPSO (first row), PSAPF (second row), HAPSOPF (third row). The tracking results for the other approaches as well as those for the sequence *S1 Gesture* are not presented due to space constraint.

5 Conclusions and Future work

In this paper, we have introduced a new algorithm for articulated object tracking based on particle swarm optimization and hierarchical search. We addressed the problem of articulated object tracking in high dimensional spaces by employing a hierarchical search to improve search efficiency. Furthermore, the problem of noisy observation has been alleviated by incorporating the annealing factor terms into the velocity updating equation of PSO. Our experiments on synthetic and real video sequences demonstrate the efficiency and effectiveness of our approach compared to other common approaches, both in terms of tracking accuracy and computation time. Our future work will focus on evaluating the proposed approach in multi-view environments.

References

[1] Balan, A.O., Sigal, L., Black, M.J.: A quantitative evaluation of video-based 3d person tracking. In: PETS. (2005) 349–356

[2] Bandouch, J., Engstler, F., Beetz, M.: Evaluation of Hierarchical Sampling Strategies in 3D Human Pose Estimation. In: BMVC. (2008) 925–934

[3] Bray, M., Kollermeier, E., Vangool, L.: Smart particle filtering for high-dimensional tracking. Computer Vision and Image Understanding **106**(1) (2007) 116–129

[4] Brubaker, M., Fleet, D., Hertzmann, A.: Physics-based person tracking using the anthropomorphic walker. International Journal of Computer Vision **87**(1-2) (2009) 140–155

[5] Chang, I.C., Lin, S.Y.: 3D human motion tracking based on a progressive particle filter. Pattern Recognition **43**(10) (June 2010) 3621–3635

[6] Deutscher, J., Reid, I.: Articulated body motion capture by stochastic search. International Journal of Computer Vision **61**(2) (2005) 185–205

[7] Hauberg, S., Pedersen, K.S.: Stick it! articulated tracking using spatial rigid object priors. In: ACCV. (2010) 758–769

[8] Hofmann, M., Gavrila, D.: 3D human model adaptation by frame selection and shape-texture optimization. Computer Vision and Image Understanding **115**(11) (2011) 1559–1570

[9] Isard, M.: PAMPAS: real-valued graphical models for computer vision. In: CVPR. (2003) 613–620

[10] John, V., Trucco, E., Ivekovic, S.: Markerless human articulated tracking using hierarchical particle swarm optimization. Image and Vision Computing **28**(11) (April 2010) 1530–1547

[11] Kjellstrom, H., Kragic, D., Black, M.: Tracking people interacting with objects. In: CVPR. (2010) 747–754

[12] MacCormick, J., Isard, M.: Partitioned sampling, articulated objects, and interface-quality hand tracking. In: ECCV. (2000) 3–19

[13] Oikonomidis, I., Kyriazis, N.: Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints. In: ICCV. (2011) 2088–2095

[14] Rose, C., Saboune, J., Charpillet, F.: Reducing particle filtering complexity for 3D motion capture using dynamic Bayesian networks. AAAI (2008) 1396–1401

[15] Sigal, L., Balan, R.: Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. Technical report (2009)

[16] Urtasun, R., Fleet, D., Hertzmann, A., Fua, P.: Priors for people tracking from small training sets. In: ICCV. (2005) 403–410

[17] Wu, Y., Hua, G., Yu, T.: Tracking articulated body by dynamic Markov network. In: ICCV. (2003) 1094–1101

[18] Xinyu, X., Baoxin, L.: Learning Motion Correlation for Tracking Articulated Human Body with a Rao-Blackwellised Particle Filter. In: ICCV. (2007) 1–8

[19] Yao, A., Gall, J., Gool, L., Urtasun, R.: Learning probabilistic non-linear latent variable models for tracking complex activities. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K., eds.: Advances in Neural Information Processing Systems 24. (2011) 1359–1367

[20] Zhang, X., Hu, W., Wang, X., Kong, Y., Xie, N., Wang, H., Ling, H., Maybank, S.: A swarm intelligence based searching strategy for articulated 3D human body tracking. In: CVPRW. (2010) 45–50