

Swarm Intelligence Based Searching Schemes for Articulated 3D Body Motion Tracking

Bogdan Kwolek, Tomasz Krzeszowski, and Konrad Wojciechowski

Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warszawa
<http://www.pjwstk.edu.pl>

Abstract. We investigate swarm intelligence based searching schemes for effective articulated human body tracking. The fitness function is smoothed in an annealing scheme and then quantized. This allows us to extract a pool of candidate best particles. The algorithm selects a global best from such a pool. We propose a global-local annealed particle swarm optimization to alleviate the inconsistencies between the observed human pose and the estimated configuration of the 3D model. At the beginning of each optimization cycle, estimation of the pose of the whole body takes place and then the limb poses are refined locally using smaller number of particles. The investigated searching schemes were compared by analyses carried out both through qualitative visual evaluations as well as quantitatively through the use of the motion capture data as ground truth. The experimental results show that our algorithm outperforms the other swarm intelligence searching schemes. The images were captured using multi-camera system consisting of calibrated and synchronized cameras.

1 Introduction

Vision-based tracking of human bodies is a significant problem due to various potential applications like user friendly interfaces, virtual reality, surveillance, clinical analysis and sport. The aim of articulated body tracking is to estimate the joint angles of the human body at any time. It is one of the most challenging problems in computer vision due to body self-occlusions, high dimensional and nonlinear state space and large variability in human appearance. To alleviate some of the difficulties, much previous work has investigated the use of 3D human body models of various complexity to recover the position, orientation and joint angles from 2D image sequences [3][4][11][13][15]. An articulated human body can be considered as a kinematic chain consisting of at least eleven parts, corresponding naturally to body parts. This means that around twenty six parameters might be needed to describe the full body pose. The state vectors describing the human pose are computed by fitting the articulated body model to the observed person's silhouette. Thus, the 3D model-based approaches rely on seeking the pose space to find the geometrical configuration of 3D model that matches best the current image observations [12]. In order to cope with practical difficulties arising due to occlusions and depth ambiguities, multiple cameras and simplified background are typically used by the research community [13][15].

2 Searching schemes for articulated human body tracking

In articulated 3D human body tracking the techniques based in particle filtering are widely used. Particle filters [5] are recursive Bayesian filters that are based on Monte Carlo simulations. They approximate a posterior distribution for the configuration of a human body given a series of observations. The high dimensionality of articulated body motion requires huge number of particles to represent well the posterior probability of the states. In such spaces, sample impoverishment may prevent particle filters from maintaining multimodal distribution for long periods of time. Therefore, many efforts have been spent in developing methods for confining the search space to promising regions with true body pose. Deutscher and Reid [3] proposed an annealed particle filter, which adopts an annealing scheme with the stochastic sampling to concentrate the particle spread near the global maximum. In the discussed approach the fitness function is smoothed using annealing factors $0 = \alpha_1 < \alpha_2, \dots, < \alpha_n = 1$, and the particles migrate towards the global maximum without getting stuck in local minima. Additionally, a crossover operation is utilized in order to maintain the diversity of the particles.

The configuration space can also be constrained using a hierarchical search. In such an approach, a part of the articulated model is localized independently in advance, and then its location is used to constrain the search for the remaining limbs. In [6], an approach called search space decomposition is proposed, where on the basis of color cues the torso is localized first and then it is used to confine the search for the limbs. However, in realistic scenarios, among others due to occlusions, it is not easy to localize the torso and to extract reliably such a good starting guess for the search.

Compared with the ordinary particle filter, the annealed particle filter greatly improves the tracking performance. However, it still requires a considerable number of particles. Since the particles do not exchange information and do not communicate with each other, they have reduced capability of focusing the search on some regions of interest in dependency on the previous visited values. In contrast, the particle swarm optimization (PSO) [7], which is population-based searching technique, has high search efficiency by combining local search (by self experience) and global one (by neighboring experience). In particular, a few simple rules result in high effectiveness of exploration of the search space.

The PSO is initialized with a group of random particles (hypothetical solutions) and then it searches hyperspace (i.e. R^n) of a problem for optima. Particles move through the solution space, and undergo evaluation according to some fitness function $f()$. Much of the success of PSO algorithms comes from the fact that individual particles have tendency to diverge from the best known position in any given iteration, enabling them to ignore local optima, while the swarm as a whole gravitates towards the global extremum. If the optimization problem is dynamic, the aim is no more to seek the extrema, but to follow their progression through the space as closely as possible. Since the object tracking process is a dynamic optimization problem, the tracking can be achieved through incorporating the temporal continuity information into the traditional PSO al-

gorithm. This means, that the tracking can be accomplished by a sequence of static PSO-based optimizations to determine the best person's pose, followed by re-diversification of the particles to cover the possible states in the next time step. In the simplest case, the re-diversification of the particle i can be realized as follows:

$$x_t^{(i)} \leftarrow \mathcal{N}(\hat{x}_{t-1}, \Sigma) \quad (1)$$

where \hat{x}_{t-1} is the estimate of the state in time $t-1$.

In order to improve the convergence speed, Clerc and Kennedy [2] proposed to use the constriction factor ω in the following form of the equation for the calculation of the i -th particle's velocity:

$$v^{i,k+1} = \omega[v^{i,k} + c_1 r_1 (p^i - x^{i,k}) + c_2 r_2 (g - x^{i,k})] \quad (2)$$

where constants c_1 and c_2 are used to balance the influence of the individual's knowledge and that of the group, respectively, r_1 and r_2 are uniformly distributed random numbers, x^i is position of the i -th particle, p^i is the local best position of particle, whereas g stands for the global best position.

In our approach the value of ω depends on annealing factor α in the following manner:

$$\omega = -0.8\alpha + 1.4 \quad (3)$$

where $\alpha = 0.1 + \frac{k}{K+1}$, $k = 0, 1, \dots, K$, and K is the number of iterations. The annealing factor is also used to smooth the objective function. The larger the iteration number is, the smaller is the smoothing. In consequence, in the last iteration the algorithm utilizes the non-smoothed function. The algorithm, which we term as annealed particle swarm optimization (APSO) can be expressed in the following pseudo-code:

1. For each particle i
2. initialize $v_t^{i,0}$
3. $x_t^{i,0} \sim \mathcal{N}(g_{t-1}, \Sigma_0)$
4. $p_t^i = x_t^{i,0}$, $f_t^i = f(x_t^{i,0})$
5. $u_t^i = f_t^i$, $\tilde{u}_t^i = (u_t^i)^{\alpha_0}$
6. $i^* = \arg \min_i \tilde{u}_t^i$, $g_t = p_t^{i^*}$, $w_t = u_t^{i^*}$
7. For $k = 0, 1, \dots, K$
8. update ω_α on the basis of (3)
9. $G = \arg \min_i \text{round}(\text{num_bins} \cdot \tilde{u}_t^i)$
10. For each particle i
11. Select a particle from $\{G \cup g_t\}$ and assign it to g_t^i
12. $v_t^{i,k+1} = \omega_\alpha [v_t^{i,k} + c_1 r_1 (p_t^i - x_t^{i,k}) + c_2 r_2 (g_t^i - x_t^{i,k})]$
13. $x_t^{i,k+1} = x_t^{i,k} + v_t^{i,k+1}$
14. $f_t^i = f(x_t^{i,k+1})$
15. if $f_t^i < u_t^i$ then $p_t^i = x_t^i$, $u_t^i = f_t^i$, $\tilde{u}_t^i = (u_t^i)^{\alpha_k}$
16. if $f_t^i < w_t$ then $g_t = x_t^i$, $w_t = f_t^i$

The smoothed objective functions undergo quantization, which constrains the real numbers to relatively small discrete set of bin values (integers), see 9th line in the pseudo-code. Thanks to such an operation the similar function values are clustered into the same bins. In each iteration the algorithm determines the set G of the particles, which after the quantization of the smoothed fitness function from the previous iteration, assumed the smallest values (the best fitness scores), see 9th line in the pseudo-code. For each particle i the algorithm selects the global best particle g_t^i from $\{G \cup g_t\}$, where g_t determines the current global best particle of the swarm. That means that the whole swarm selects the global best location from a set of candidate best locations. Figure 1 depicts the number of particles in each bin, which has been determined in one of the experiments, where 200 particles, 10 iterations and a quantization into 30 bins were employed in estimating the human pose.

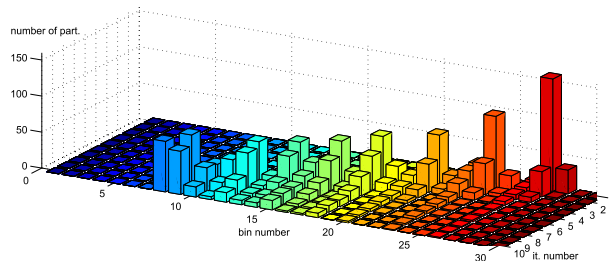


Fig. 1: Number of the particles in each bin in iterations 1,...,10.

In [15], an annealed PSO based particle filter has been proposed and evaluated in tracking articulated 3D human body. Our approach is different from the discussed algorithm, in particular, it relies on annealing scheme that is based on the smoothing and the quantization of the fitness function. Additionally, the constriction factor that controls the convergence rate depends on the annealing factor.

In the global-local particle swarm optimization (GLPSO) algorithm [9], at the beginning of each frame we estimate the pose of the whole body using PSO. Given the pose of the whole body, we construct state vectors consisting of the estimated state variables for pelvis and torso/head, the arms and the legs. At this stage the state variables describing the pose of the legs are perturbed by normally distributed motion. Afterwards, we execute particle swarm optimization in order to calculate the refined estimate of the legs pose. Such refined state variables are then placed in the state vector of the whole body. The state variables describing the hands are refined analogously. Our global-local annealed particle swarm optimization algorithm (GLAPSO) operates analogously, but instead of the ordinary PSO algorithm we employ APSO optimizations.

3 Tracking framework

The skeleton of the human body is modeled as a kinematic tree. The articulated 3D model consists of eleven segments with limbs represented by the truncated cones, which model the pelvis, torso/head, upper and lower arm and legs. The configuration of the model is defined by 26 DOF. It is parameterized by the position and the orientation of the pelvis in the global coordinate system and the relative angles between the connected limbs. In order to obtain the 3D human pose each truncated cone is projected into 2D image plane via perspective projection. In such a way we obtain an image with the rendered model in a given configuration. Such image features are then matched to the person extracted by image analysis.

In most of the approaches to articulated 3D human body tracking, the cameras are static and background subtraction algorithms are utilized to extract the object of interest [14]. In addition, image cues like edges, ridges, color are used frequently to get better delineation of the person [11]. In [10], face detection, head-shoulders contour matching and elliptical skin-blob detection techniques were used in estimating the 3D human poses in static images. In our approach, the background subtraction algorithm [1] is used to extract the binary image of the person, see Fig. 2b). It is then used to calculate a silhouette-overlap term.

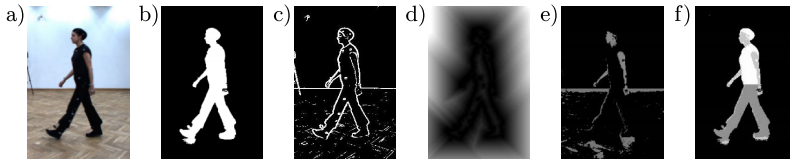


Fig. 2: Person segmentation. Input image a), foreground b), gradient magnitude c), edge distance map d), skin color patches e), extracted body segments f).

Image edges complement silhouette features and contribute toward precise aligning the body limbs. The most common type of edge detection process uses a gradient operator. Gradient features share many properties with optical flow. In particular, they do not depend on background subtraction. Gradient angle is invariant to global changes of image intensities. In contrast to optical flow, gradients features are discriminative for both moving and non-moving body parts. In our approach, the gradient magnitude, see Fig. 2c), is masked with the closed image of the foreground and then used to generate the edge distance map, see also Fig. 2d). It assigns each pixel a value that is the distance between that pixel and the nearest nonzero edge pixel. In our implementation we employ chessboard distance and limit the number of iterations on the chain propagation to three. Additionally, in the GLPSO and GLAPSO algorithms we perform the segmentation of the person's shape into individual body parts. To accomplish this we model the distribution of skin color using 16×16 histogram in rg color space.

The histogram back-projection is employed to identify the skin patches and to extract the skin binary masks, see Fig. 2e. Such masks are then used to delineate the skin segments in the person binary images. Taking into account the height of the extracted person we perform rough segmentation of the legs and feet, see Fig. 2f).

The fitness score is calculated on the basis of following expression: $f(x) = 1 - (f_1(x)^{w_1} \cdot f_2(x)^{w_2})$, where w denotes weighting coefficients that were determined experimentally. The function $f_1(x)$ reflects the degree of overlap between the segmented body parts and the projected model's parts into 2D image. The overlap degree is calculated through checking the overlap from the binary image to the considered rasterized image of the model and vice versa. The larger the degree of overlap is, the larger is the fitness value. In GLPSO and GLAPSO algorithms the silhouette-overlap term is calculated with consideration of the distinguished body parts. The second function reflects the degree of overlap between model edges and image edges. At this stage the above mentioned edge-proximity term is utilized.

4 Experimental results

The algorithms were compared by analyses carried out both through qualitative visual evaluations as well as quantitatively through the use of the motion capture data as ground truth. The images were captured using multi-camera system consisting of four calibrated and synchronized cameras. The system acquires color images of size 1920×1080 with rate 24 fps. Each pair of the cameras is approximately perpendicular to the other two. The placement of the video cameras in our laboratory is shown in Fig. 3. A commercial motion capture

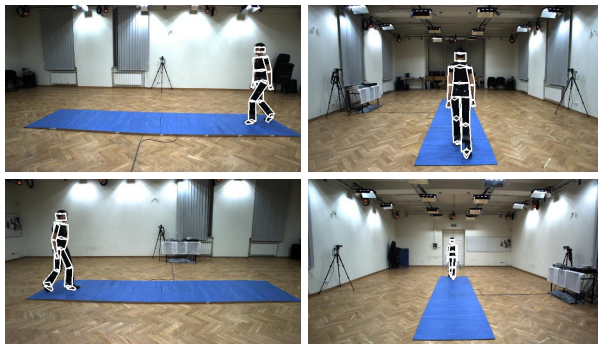


Fig. 3: Layout of the laboratory and camera setup. The images illustrate human motion tracking in frame #20 seen in view 1 and 2 (upper row), and in view 3 and 4 (bottom row).

(MoCap) system from Vicon Nexus provides ground truth motion of the body at rate of 100 Hz. It utilizes reflective markers and sixteen cameras to recover the 3D location of such markers. The cameras are all digital and are capable to differentiate overlapping markers from each camera's view. The synchronization between the MoCap and multi-camera system is done through hardware from Vicon Gigaset Lab.

Figure 4 demonstrates some tracking results that were obtained using particle swarm optimization (PSO), see images in the first row, global-local PSO, see images in the second row, annealed PSO, see images in the third row, and global-local annealed PSO, see images in the last row. Each image consists of two sub-images, where the left sub-image contains the model overlaid on the images from the view 1, whereas the second one illustrates the model overlaid on the images from the view 4. In the experiments presented below we focused on analyses of motion of walking people with bared and freely swinging arms. The analysis of the human way of walking, termed gait, can be utilized in several applications ranging from medical applications to surveillance. This topic is now a very active research area in the vision community.

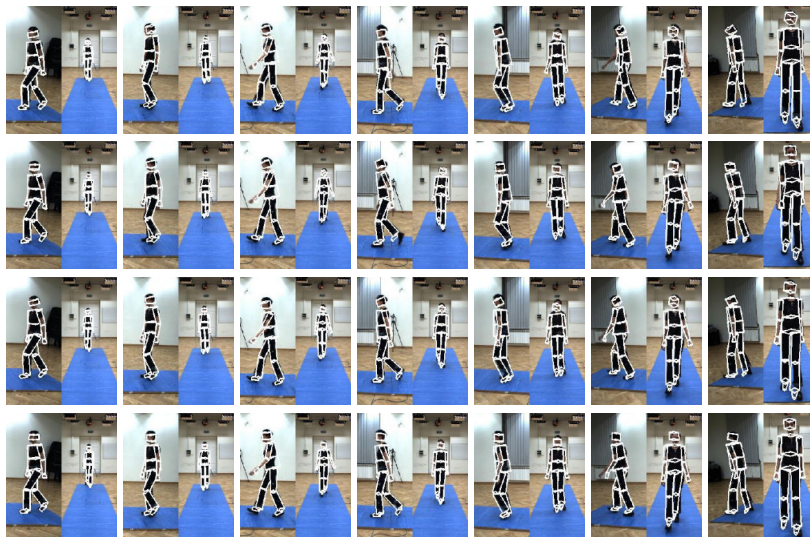


Fig. 4: Articulated 3D human body tracking. Shown are results in frames #20, 40, 60, 80, 100, 120, 140, obtained by PSO (1st row), GLPSO (2nd row), APSO (3rd row), GLAPSO (4th row), respectively. The left sub-images are seen from view 1, whereas the right ones are seen from view 4.

For fairness, in all experiments we use the identical particle configurations. For the global-local PSO and global-local annealed PSO the sum of particles responsible for tracking the whole body, arms and legs corresponds to the number

of the particles in the PSO and APSO. For instance, the use of 300 particles in PSO or APSO is equivalent to the use of 200 particles for tracking the full body, 50 particles for tracking the arms and 50 particles for tracking both legs in GLPSO or GLAPSO. The use of 200 in PSO and APSO corresponds to the exploitation of 150, 25 and 25 particles, respectively, whereas the use of 100 particles equals to utilization 80 particles for tracking the global configuration of the body, along with 10 and 10 particles for tracking hands and legs, respectively. In Tab. 1 we can see some quantitative results that were obtained using two image sequences. For each sequence the results were averaged over ten runs with unlike initializations. They were achieved using image sequences consisting of 150 and 180 frames. In the quantization the number of bins was set to 30. Figure 4 demonstrates the images from the sequence two.

Table 1: Average errors for $M = 39$ markers in two image sequences.

				Seq. 1		Seq. 2	
		#particles	it.	error [mm]	std. dev. [mm]	error [mm]	std. dev. [mm]
PSO	100	10	86.17	50.58	73.89	35.98	
	100	20	77.71	39.36	67.58	32.15	
	300	10	75.31	41.50	65.56	30.26	
	300	20	75.11	38.42	63.43	28.63	
GLPSO	100	10	80.95	42.69	68.50	32.00	
	100	20	67.66	27.15	67.17	30.08	
	300	10	68.58	30.98	64.40	28.01	
	300	20	67.96	30.03	62.87	26.00	
APSO	100	10	71.56	36.26	65.04	29.74	
	100	20	68.81	31.87	61.29	26.86	
	300	10	66.51	29.63	61.78	26.69	
	300	20	64.63	28.91	59.70	24.98	
GLAPSO	100	10	69.44	31.21	63.37	30.74	
	100	20	63.71	28.79	60.42	26.72	
	300	10	60.07	21.07	60.71	24.41	
	300	20	58.96	19.43	57.62	22.49	

The pose error in each frame was calculated using $M = 39$ markers $m_i(x)$, $i = 1, \dots, M$, where $m_i \in R^3$ represents the location of the i -th marker in the world coordinates. It was expressed as the average Euclidean distance:

$$E(x, \hat{x}) = \frac{1}{M} \sum_{i=1}^M \|m_i(x) - m_i(\hat{x})\| \quad (4)$$

where $m_i(x)$ stands for marker's position that was calculated using the estimated pose, whereas $m_i(\hat{x})$ denotes the position, which has been determined using data

from our motion capture system. From the above set of markers, four markers were placed on the head, seven markers on each arm, 6 on the legs, 5 on the torso and 4 markers were attached to the pelvis. For the discussed placement of the markers on the human body the corresponding marker's assignment on the 3D model was established. Given the estimated human pose the corresponding 3D positions of virtual markers were determined. On the basis of data stored in c3d files the ground truth was extracted and then utilized in calculating the average Euclidean distance given by (4).

The errors that are shown in Tab. 1 were calculated on the basis of the following equation:

$$Err(x, \hat{x}) = \frac{1}{LM} \sum_{k=1}^L \sum_{i=1}^M \|m_i(x) - m_i(\hat{x})\| \quad (5)$$

where L denotes the number of frames in the utilized test sequences. As Tab. 1 shows, the particle swarm optimization algorithm allows us to obtain quite good results. The GLPSO outperforms the PSO algorithm. However, such considerably better results can only be obtained if skin and legs segmentation is used in GLPSO. Due to global-local searching strategy the GLPSO algorithm is superior in utilizing the information about the location of the hands and the legs. The APSO algorithm tracks better in comparison to GLPSO and PSO. Moreover, it requires no segmentation of skin patches or body parts. The GLAPSO takes advantages of both algorithms and achieves the best results.

The plots shown in Fig. 5 illustrate the pose error versus frame number that was obtained by each algorithm. We can see that the PSO-based algorithm is able

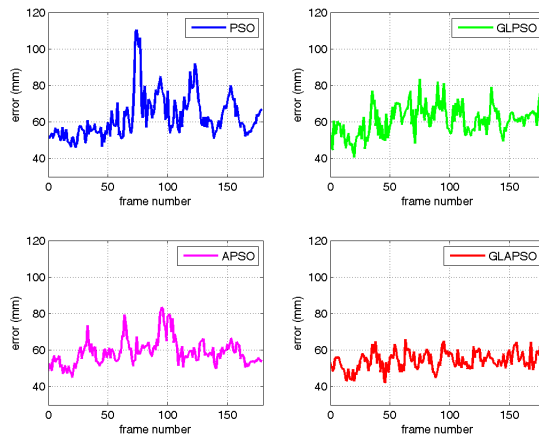


Fig. 5: Pose error [mm] versus frame number.

to provide estimates with errors that are sporadically larger than 100 mm. The average error of the GLAPSO is far below 60 mm. In the GLPSO algorithm, at the beginning of each PSO cycle, the estimation of the pose of the whole body takes place and then the poses of the limbs are refined locally using smaller number of particles. In the algorithms based on the annealing the particles are beforehand weighted by smoothed versions of the weighting function, where the influence of the local minima is weakened first but increases gradually. This leads to consistent tracking of the 3D articulated body motion. In consequence, the errors in our algorithm, which takes the advantages the two different strategies for exploration of the search space are far smaller. The discussed algorithm is capable of achieving better results because of its ability to thinking globally and acting locally. As a result, the GLAPSO algorithm outperforms GLPSO and APSO algorithms both in terms of the tracking accuracy as well as consistency in tracking of the human motion. In particular, the standard deviation is far smaller in comparison to the standard deviation of the other investigated algorithms.

Figure 6 presents the pose estimation errors for particular body parts. The results were obtained in 20 iterations using the GLAPSO algorithm with 200 particles for the whole body tracking and 2×50 particles for tracking the arms and legs. As we can observe, our algorithm can track body limbs with lower errors and it is robust to ambiguous configurations such as self occlusion. In more detail, from the discussed plot, we can see that for all body parts except the left forearm the maximal error does not exceed 100 mm.

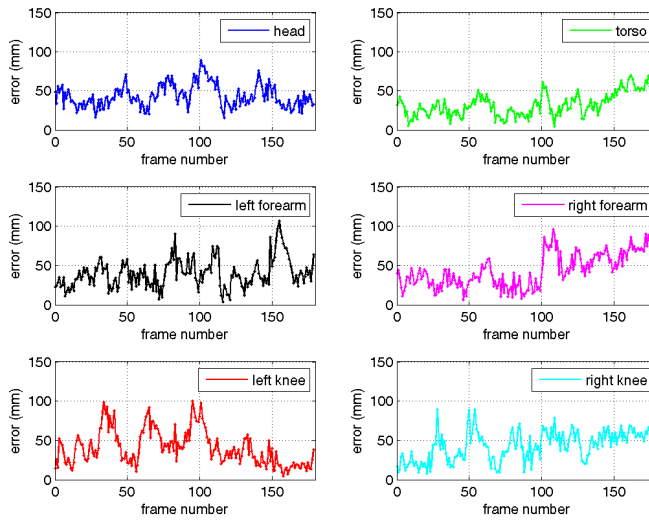


Fig. 6: Tracking errors [mm] versus frame number.

The complete human motion capture system was written in C/C++. The system runs on Windows and Linux in both 32 bit and 64 bit modes. It operates on color images with spatial resolution of 960×540 . The entire tracking process takes approximately 2.1 sec. per frame on Intel Core i5 2.8 GHz using a configuration with 100 particles 10 iterations. If Open Multi-Processing (OpenMP) is employed the tracking is completed in 1.12 sec. The image processing and analysis takes about 0.45 sec. One of the future research directions of the presented approach is to explore the CUDA/GPU in order to speed-up the computations [8].

5 Conclusions

We have presented a vision system that effectively utilizes swarm intelligence searching schemes to achieve better articulated 3D human body tracking. By combining two searching strategies, namely, annealed and global-local, the proposed method can tackle the inconsistency between the observed body pose and the estimated model configurations. Due to better capability of exploring the search space, the combination of above-mentioned searching strategies leads to superior tracking the articulated 3D human motion. Our global-local annealed (GLAPSO) algorithm is able to track the articulated 3D human motion reliably in multi-camera image sequences. In particular, the resulting algorithm is robust to ambiguous body configurations such as self occlusion. Moreover, it performs satisfactory even when small number of particles is employed, say 100 particles and 10 iterations. The fitness function is smoothed in an annealing scheme and then quantized. This allows us to maintain a pool of candidate best particles. Furthermore, the constriction factor that controls the convergence rate depends on the annealing factor. To show the advantages of our algorithm, we have conducted several experiments on walking sequences and investigated global-local and annealed searching strategies. The algorithms were compared by analyses carried out both through qualitative visual evaluations as well as quantitatively through the use of the motion capture data as ground truth.

Acknowledgment

This paper has been supported by the project “System with a library of modules for advanced analysis and an interactive synthesis of human motion” co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme - Priority Axis 1.

References

1. Arsic, D., Lyutskanov, A., Rigoll, G., Kwolek, B.: Multi camera person tracking applying a graph-cuts based foreground segmentation in a homography framework. In: IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance. pp. 30–37. IEEE Press, Piscataway, NJ (2009)

2. Clerc, M., Kennedy, J.: The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. on Evolutionary Computation* 6(1), 58–73 (2002)
3. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: *IEEE Int. Conf. on Pattern Recognition*. pp. 126–133 (2000)
4. Deutscher, J., Reid, I.: Articulated body motion capture by stochastic search. *Int. J. Comput. Vision* 61(2), 185–205 (2005)
5. Doucet, A., Godsill, S., Andrieu, C.: On sequential Monte Carlo sampling methods for bayesian filtering. *Statistics and Computing* 10(1), 197–208 (2000)
6. Gavrilu, D.M., Davis, L.S.: 3-D model-based tracking of humans in action: a multi-view approach. In: *Proc. of the Conf. on Computer Vision and Pattern Recognition (CVPR '96)*. pp. 73–80. IEEE Computer Society, Washington, DC, USA (1996)
7. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: *Proc. of IEEE Int. Conf. on Neural Networks*. pp. 1942–1948. IEEE Press, Piscataway, NJ (1995)
8. Krzeczowski, T., Kwolek, B., Wojciechowski, K.: GPU-accelerated tracking of the motion of 3D articulated figure. In: *Proc. of Int. Conf. on Computer Vision and Graphics: Part I*. pp. 155–162. ICCVG'10, Springer-Verlag, Berlin, Heidelberg (2010)
9. Krzeczowski, T., Kwolek, B., Wojciechowski, K.: Model-based 3D human motion capture using global-local particle swarm optimizations. In: *Computer Recognition Systems 4, Advances in Intelligent and Soft Computing*, vol. 95, pp. 297–306. Springer Berlin / Heidelberg (2011)
10. Lee, M.W., Cohen, I.: A model-based approach for estimating human 3D poses in static images. *IEEE Trans. Pattern Anal. Mach. Intell.* 28, 905–916 (June 2006)
11. Schmidt, J., Fritsch, J., Kwolek, B.: Kernel particle filter for real-time 3D body tracking in monocular color images. In: *IEEE Int. Conf. on Face and Gesture Rec., Southampton, UK*. pp. 567–572. IEEE Computer Society Press (2006)
12. Sidenbladh, H., Black, M., Fleet, D.: Stochastic tracking of 3D human figures using 2d image motion. In: *European Conf. on Computer Vision*. pp. 702–718 (2000)
13. Sigal, L., Balan, A., Black, M.: HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *Int. Journal of Computer Vision* 87, 4–27 (2010)
14. Sminchisescu, C., Kanaujia, A., Li, Z., Metaxas, D.: Discriminative density propagation for 3D human motion estimation. In: *IEEE Int. Conf. on Computer Vision and Pattern Recognition*. pp. I:390–397 (2005)
15. 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: *IEEE Workshop on 3D Information Extraction for Video Analysis and Mining in conjunction with CVPR*. pp. 45–50. IEEE (2010)