Fast Articulated Motion Tracking using a Sums of Gaussians Body Model Supplementary Material

Carsten Stoll MPI Informatik
Nils Hasler MPI Informatik
Juergen Gall ETH Zurich
Hans-Peter Seidel MPI Informatik
Christian Theobalt MPI Informatik

A. Evaluating tracking quality with varying number of cameras

To evaluate the stability of our tracking method compared to the number of used cameras, we tracked the "Cartwheel" and "Walk 1" sequences with a reduced number of cameras. We used the tracking result of the full camera set as baseline. We measure the joint-position error and standard deviation of the tracked motion. As can be seen in Table 1, tracking results are reliable down to 6 cameras. First obvious tracking errors become visible with 5 cameras. Combining the local optimization with a global optimization in case of tracking failures would lead to more robust results with less cameras.

nr. cameras	12	10	8	6	5	4
Cartwheel error (mm)	0.0 ± 0.0	13.16 ± 9.06	19.51 ± 13.22	32.28 ± 21.17	34.85 ± 24.37	67.72 ± 97.12
Walk 1 error (mm)	0.0 ± 0.0	11.40 ± 8.37	17.97 ± 12.51	23.69 ± 22.39	72.15 ± 112.16	70.16 ± 107.299

Table 1. Influence of the number of cameras on tracking accuracy. Using the full camera set as baseline, we measure joint-position error and standard deviation. The tracking results for 5 and 4 cameras contain some obvious tracking errors.

B. Evaluating tracking quality with varying image SoG accuracy

We analyzed the influence of the maximum SoG image resolution on the tracking result by creating representations in several resolutions and comparing the tracking result for the "Cartwheel" sequence. The tracking result of the highest resolution SoG image representation is used as baseline tracking result. The error measure is the deviation in joint position in the tracked motion. Table 2 shows that even at a very low resolution the performer can still be tracked successfully. The most obvious differences are the tracking accuracy of extremities (i.e. feet and hands), which are more accurately positioned with higher resolutions. The quadtree depth is an intuitive control to trade tracking accuracy for processing speed.

max resolution / quadtree depth	324×240 / 9	162×120 / 8	81×60 / 7
error (mm)	0.0 ± 0.0	11.25 ± 6.85	15.29 ± 9.28
frames per second	1.4	5.59	12.71

Table 2. Influence of the camera resolution (quadtree depth) on tracking accuracy. Using the highest resolution as baseline, we measure joint-position error and standard deviation. Increased resolution results in higher tracking quality, but sacrifices processing speed.

Sequence	**Frames**	ϵ	**avg. iter. per frame**	**Time**	**FPS**
Cartwheel	597	0.0025	34.56	106.89s	5.59
Chair 1	1347	0.0025	24.97	175.68s	7.67
Chair 2	1830	0.0025	26.49	295.49s	6.19
Handstand	582	0.0025	24.61	74.85s	7.78
Talking	924	0.0020	32.73	180.75s	5.11
Table 1	527	0.0020	41.64	128.67s	4.10
Table 2	586	0.0025	19.86	67.71s	8.65
Table 3	556	0.0025	22.90	62.19s	8.94
Table 4	510	0.0025	35.42	108.72s	4.69
Throw	780	0.0025	26.61	117.33s	6.65
Walk 1	739	0.0025	35.44	150.72s	4.90
Walk 2	912	0.0025	35.12	189.67s	4.81
Fight	384	0.0015	58.34	138.75s	2.77
Dance	728	0.0020	19.30	78.77s	9.24
Shake	228	0.0025	14.06	15.61s	14.61
ShakeHug	375	0.0015	27.94	55.35s	6.78
Evaluation	501	0.0015	77.26	216.33s	2.32
Total	12106		31.35	2163.48s	5.60

Table 3. Overview of all tracked sequences. For all sequences $w_a = 0.005$, $w_l = 1.0$. Tracking speed is dependent on motion speed and complexity, as well as on the optimization threshold ϵ.