Paper 576: Supplemental Material
Hybrid Stochastic / Deterministic Optimization
for Tracking Sports Players and Pedestrians
Robert Collins and Peter Carr

This supplemental material contains an expanded table of evaluation measures that includes the intermediate counts used to compute the scores reported in the paper, additional evaluations that help better understand the performance of the algorithm, and videos showing the results of our algorithm on the test sequences in the paper.

Evaluation Scores

The evaluation measures shown in Table 1 of the paper are computed from other intermediate numbers like number of true positive detections, number of mismatch errors (ID swaps), and so on. Here we present an expanded table showing all numbers involved in the evaluation computations.

Doohan sequence												
Alg	**GT**	**TP↑**	**FN↓**	**FP↓**	**MME↓**	**MOTA↑**	**MOTP↓**	**Prec↑**	**Recall↑**	**ADlen↑**	**AGlen**	**AvgLen↑**
SF-1S	492	473	19	19	29	86.38	0.27	96.14	96.14	167	500	33.40
SF-DC	492	422	70	**13**	91	64.63	**0.23**	**97.01**	85.77	42	500	8.40
Full(ours)	492	**475**	**17**	16	**15**	**90.24**	0.26	96.74	**96.54**	**441**	500	**88.20**
APIDIS Camera 6 sequence												
Alg	**GT**	**TP↑**	**FN↓**	**FP↓**	**MME↓**	**MOTA↑**	**MOTP↓**	**Prec↑**	**Recall↑**	**ADlen↑**	**AGlen**	**AvgLen↑**
SF-1S	348	271	77	62	51	45.40	0.37	81.38	77.87	75	313	23.96
SF-DC	348	232	116	63	26	41.09	0.43	78.64	66.67	110	313	35.14
POM-1S	348	268	80	138	69	17.53	0.39	66.01	77.01	37	313	11.82
POM-DC	348	229	119	90	32	30.75	0.49	71.79	65.80	104	313	33.23
Full(ours)	348	**283**	**65**	**48**	**17**	**62.64**	**0.34**	**85.50**	**81.32**	**190**	313	**60.70**
Oxford Town Centre sequence												
Alg	**GT**	**TP↑**	**FN↓**	**FP↓**	**MME↓**	**MOTA↑**	**MOTP↓**	**Prec↑**	**Recall↑**	**ADlen↑**	**AGlen**	**AvgLen↑**
SF-1S	7004	5084	1920	2730	261	29.88	0.46	65.06	72.59	101	307	32.90
SF-DC	7004	4113	2891	5094	265	-17.79	0.59	44.67	58.72	163	307	53.09
POM-1S	7004	4930	2074	3210	365	19.35	**0.40**	60.57	70.39	62	307	20.20
POM-DC	7004	4159	2845	4296	249	-5.51	0.61	49.19	59.38	183	307	59.61
Full(ours)	7004	**5172**	**1832**	**2169**	**109**	**41.32**	0.45	**70.45**	**73.84**	**212**	307	**69.06**

Table 1: Expanded quantitative evaluation table. The match threshold for CLEAR MOT measures is 1 meter, applied in the ground plane. Up arrows denote columns where higher values are better; down arrows mark columns where lower values are better.

The meaning of each column is described below:
GT = number of ground truth detections
TP = number of true positive detections
FN = number of false negative detections
FP = number of false positive detections
MME = number of mismatch errors (aka ID swaps)
MOTA = [1-(FN+FP+MME)/GT]*100
MOTP = average distance error of true positive detections (in meters)
Prec = TP / (TP+FP)

Recall = TP / (TP+FN) = TP / GT
ADlen = average length of estimated trajectories (in frames)
AGlen = average length of ground truth trajectories (in frames)
AvgLen = (ADlen / AGlen)*100

As was noted in the paper, the MOTP numbers are generally worse for the Oxford Town Centre sequence because the "ground truth" x,y locations on the ground were inferred from labeled head locations in the image that were provided with the dataset, which introduces errors in the ground truth because people are not all the same height. Also noted in the paper was that trajectory smoothing within discrete-continuous (DC) data association appeared to do more harm than good in the fast moving sports sequences, and this is likely to have been further compounded by the sparse temporal sampling of frames used by the SF and POM detection baselines.

Additional Evaluations

We performed the CLEAR MOT evaluations using a 3D match measure of distance in the ground plane because that is the domain where the pedestrian locations estimated by our approach "live." However, another popular match measure is 2D bounding box overlap measured in the image, with a threshold of 50 percent determining a match. We therefore have also measured performance of our hybrid method using this 2D CLEAR measure. Furthermore, we have run additional experiments that explore the change in performance if appearance information (color histogram similarity) is removed from the edge costs in our data association model, leading to costs that are only based on distance. These additional evaluation numbers are presented in the table below, for the publicly available APIDIS and Oxford TownCentre sequences.

APIDIS sequence				
Algorithm	**MOTP(m)**	**MOTA(-)**	**Prec(%)**	**Recall(%)**
HybridFull(ours)	74.0	70.7	91.2	86.8
Ours, no color	73.0	70.1	90.1	86.5
TownCentre sequence				
Algorithm	**MOTP(m)**	**MOTA(-)**	**Prec(%)**	**Recall(%)**
HybridFull(ours)	70.5	63.4	85.8	77.6
Ours, no color	70.0	57.3	82.8	78.7

Table 2: Quantitative evaluation of our hybrid algorithm both with and without color appearance information, on APIDIS and Oxford Towncentre datasets, using 2D bounding box overlap of 50 percent as the CLEAR MOT match threshold. Higher values are better for all scores.

As expected, the 2D scores are higher, because image bounding box overlap can tolerate a fair amount of inaccuracy along the line of sight, leading to more matches being made than when an isotropic 3D ground plane distance threshold is used. Turning off color appearance causes the scores to decrease only slightly for APIDIS (many players

are dressed in the same colors, so color is not a very discriminative feature) but more so for the TownCentre dataset.

We also have run additional experiments to see how detection scores evolve as a function of number of MCMC iterations. To do this, we ran 10 trials on randomly sampled 100-frame subwindows of the Oxford TownCentre dataset, keeping track of the average evaluation scores as number of iterations per frame rises from 0 (oneshot initialization with no iterative improvement) to 1000, yielding the table below.

Iterations	0	100	500	1000
MOTP	68.6	68.8	68.7	68.8
TP	114.6	111.2	109.9	112.5
FN	44.2	47.6	48.9	46.3
FP	31.4	30.9	28.7	26.2
MME	6.4	3.4	1.5	0.5
MOTA	50.0	50.5	51.9	55.0
PREC	79.0	79.0	80.1	81.5
RECALL	73.1	71.1	70.1	71.5
AvgPathLen	52.7	60.0	66.8	68.8

Table 3: Performance measures as a function of number of MCMC iterations.

Steady reduction is seen in false positives and MME (ID swaps), leading to steady increase in MOTA and precision. Average path length also steadily increases. Scores with little or no improvement with respect to number of iterations include MOTP, false negatives, and recall.

Uploaded Videos

The videos ourDoohanResult.mp4, ourApidisResult.mp4, and ourOxfordResult.mp4 are provided to allow qualitative assessment of the tracking results of our proposed hybrid stochastic-deterministic optimization algorithm. The video codec is H264 using an mp4 container, generated by MPEG Streamclip on a Mac. The videos should play on any sufficiently modern platform; we have successfully viewed them in Quicktime 7, Quicktime X and VLC.