Strip-Fusion: Spatiotemporal Fusion for Multispectral Pedestrian Detection - Multimedia Tables

Asiegbu Miracle Kanu-Asiegbu[1], Nitin Jotwani[2] and Xiaoxiao Du[3]

Table R1: **KAIST** MR-All results for Algorithm 1 post-processing. F is the number of frames and S_T the temporal stride. s^v and s^t denote visible and thermal confidence thresholds, and iou_{thres} the IoU threshold. Scores are fused by AVG, $(s^v + s^t)/2$, or MAX, $\max(s^v, s^t)$. KL divergence uses $\beta = 2$. The default setting in the main paper is $s^v = s^t = 0.1$, $iou_{thres} = 0.75$, AVG.

F	S_T	s^v	s^t	iou_{thres}	Mode	**No KL**	**With KL**
1	1	0.1	0.1	0.50	AVG	9.90	**7.40**
1	1	0.1	0.1	0.50	MAX	10.65	9.01
1	1	0.1	0.1	0.75	AVG	**9.69**	7.52
1	1	0.1	0.1	0.75	MAX	9.83	7.51
1	1	0.2	0.2	0.50	AVG	10.64	**7.40**
1	1	0.2	0.2	0.50	MAX	11.15	8.18
1	1	0.2	0.2	0.75	AVG	10.38	7.49
1	1	0.2	0.2	0.75	MAX	10.83	7.54
3	3	0.1	0.1	0.50	AVG	9.92	8.11
3	3	0.1	0.1	0.50	MAX	11.25	9.48
3	3	0.1	0.1	0.75	AVG	**9.84**	**8.01**
3	3	0.1	0.1	0.75	MAX	10.43	8.82
3	3	0.2	0.2	0.50	AVG	10.31	9.02
3	3	0.2	0.2	0.50	MAX	10.87	9.84
3	3	0.2	0.2	0.75	AVG	10.17	8.73
3	3	0.2	0.2	0.75	MAX	11.15	9.67
5	3	0.1	0.1	0.50	AVG	9.05	**9.74**
5	3	0.1	0.1	0.50	MAX	10.42	9.93
5	3	0.1	0.1	0.75	AVG	**8.84**	10.01
5	3	0.1	0.1	0.75	MAX	9.76	10.42
5	3	0.2	0.2	0.50	AVG	9.74	10.41
5	3	0.2	0.2	0.50	MAX	10.87	10.75
5	3	0.2	0.2	0.75	AVG	9.92	10.16
5	3	0.2	0.2	0.75	MAX	10.83	11.02
7	3	0.1	0.1	0.50	AVG	10.44	10.70
7	3	0.1	0.1	0.50	MAX	12.03	10.14
7	3	0.1	0.1	0.75	AVG	10.23	9.91
7	3	0.1	0.1	0.75	MAX	**9.76**	**9.36**
7	3	0.2	0.2	0.50	AVG	11.23	11.60
7	3	0.2	0.2	0.50	MAX	11.72	11.41
7	3	0.2	0.2	0.75	AVG	10.95	11.07
7	3	0.2	0.2	0.75	MAX	10.72	10.16
7	10	0.1	0.1	0.50	AVG	9.38	8.84
7	10	0.1	0.1	0.50	MAX	11.16	10.10
7	10	0.1	0.1	0.75	AVG	**9.33**	8.40
7	10	0.1	0.1	0.75	MAX	9.58	**8.39**
7	10	0.2	0.2	0.50	AVG	10.28	9.51
7	10	0.2	0.2	0.50	MAX	11.50	9.58
7	10	0.2	0.2	0.75	AVG	10.14	9.16
7	10	0.2	0.2	0.75	MAX	11.11	9.37

Table R2: **CVC-14** MR-All results for Algorithm 1 post-processing. F is the number of frames and S_T the temporal stride. s^v and s^t denote visible and thermal confidence thresholds, and iou_{thres} the IoU threshold. Scores are fused by AVG, $(s^v + s^t)/2$, or MAX, $\max(s^v, s^t)$. KL divergence uses $\beta = 1$. The default setting in the main paper is $s^v = s^t = 0.1$, $iou_{thres} = 0.75$, AVG.

F	S_T	s^v	s^t	iou_{thres}	Mode	**No KL**	**With KL**
1	1	0.1	0.1	0.50	AVG	17.90	18.94
1	1	0.1	0.1	0.50	MAX	21.17	20.46
1	1	0.1	0.1	0.75	AVG	**17.79**	**18.76**
1	1	0.1	0.1	0.75	MAX	20.02	20.56
1	1	0.2	0.2	0.50	AVG	17.88	18.96
1	1	0.2	0.2	0.50	MAX	19.67	20.89
1	1	0.2	0.2	0.75	AVG	17.81	18.76
1	1	0.2	0.2	0.75	MAX	19.30	20.56
3	3	0.1	0.1	0.50	AVG	17.93	16.72
3	3	0.1	0.1	0.50	MAX	18.83	18.27
3	3	0.1	0.1	0.75	AVG	**17.26**	16.53
3	3	0.1	0.1	0.75	MAX	18.53	17.61
3	3	0.2	0.2	0.50	AVG	18.12	16.63
3	3	0.2	0.2	0.50	MAX	19.40	17.67
3	3	0.2	0.2	0.75	AVG	17.48	**16.46**
3	3	0.2	0.2	0.75	MAX	18.70	17.39
5	3	0.1	0.1	0.50	AVG	16.89	16.41
5	3	0.1	0.1	0.50	MAX	18.42	17.99
5	3	0.1	0.1	0.75	AVG	16.76	**16.34**
5	3	0.1	0.1	0.75	MAX	17.14	17.37
5	3	0.2	0.2	0.50	AVG	16.69	16.47
5	3	0.2	0.2	0.50	MAX	18.28	17.67
5	3	0.2	0.2	0.75	AVG	**16.54**	**16.34**
5	3	0.2	0.2	0.75	MAX	16.98	17.34
7	3	0.1	0.1	0.50	AVG	16.85	19.03
7	3	0.1	0.1	0.50	MAX	18.37	20.81
7	3	0.1	0.1	0.75	AVG	17.09	**18.99**
7	3	0.1	0.1	0.75	MAX	17.78	19.82
7	3	0.2	0.2	0.50	AVG	**16.78**	18.97
7	3	0.2	0.2	0.50	MAX	17.81	20.41
7	3	0.2	0.2	0.75	AVG	16.97	19.01
7	3	0.2	0.2	0.75	MAX	17.90	19.32
7	5	0.1	0.1	0.50	AVG	16.73	19.11
7	5	0.1	0.1	0.50	MAX	17.87	19.93
7	5	0.1	0.1	0.75	AVG	16.80	18.82
7	5	0.1	0.1	0.75	MAX	17.05	19.65
7	5	0.2	0.2	0.50	AVG	**16.68**	18.88
7	5	0.2	0.2	0.50	MAX	18.40	20.46
7	5	0.2	0.2	0.75	AVG	16.85	**18.69**
7	5	0.2	0.2	0.75	MAX	17.87	**18.69**

Table R3: **KAIST** comparison of inference time in frames-per-second for each sequence. Recall the Number of Frames (F) and stride (S_T).

F	$\mathrm{S_T}$	**No KL** ($\beta = 0$)			**With KL** ($\beta = 2$)		
		Algo. 1	VIS	IR	Algo. 1	VIS	IR
1	1	3.6599	3.3662	3.8625	3.8023	3.3331	3.9262
3	3	2.4697	2.5685	2.6519	2.7125	2.7384	2.7442
5	3	1.9156	2.0040	2.0191	2.1030	2.1130	2.1295
7	3	1.5997	1.6653	1.6835	1.7666	1.7644	1.7935
7	10	1.7508	1.7954	1.8178	1.8312	1.8269	1.8525

Table R4: **CVC-14** comparison of inference time in frames-per-second for each sequence. Recall the Number of Frames (F) and stride (S_T).

F	$\mathrm{S_T}$	**No KL** ($\beta = 0$)			**With KL** ($\beta = 1$)		
		Algo. 1	VIS	IR	Algo. 1	VIS	IR
1	1	3.7279	3.8658	4.1278	3.4043	4.1462	4.1729
3	3	2.8153	2.8694	2.8597	2.7421	2.8676	2.8647
5	3	2.1206	2.1933	2.2105	2.1850	2.2049	2.2044
7	3	1.8601	1.8709	1.8659	1.8623	1.8709	1.8723
7	10	1.8491	1.8766	1.8775	1.8523	1.8684	1.8689

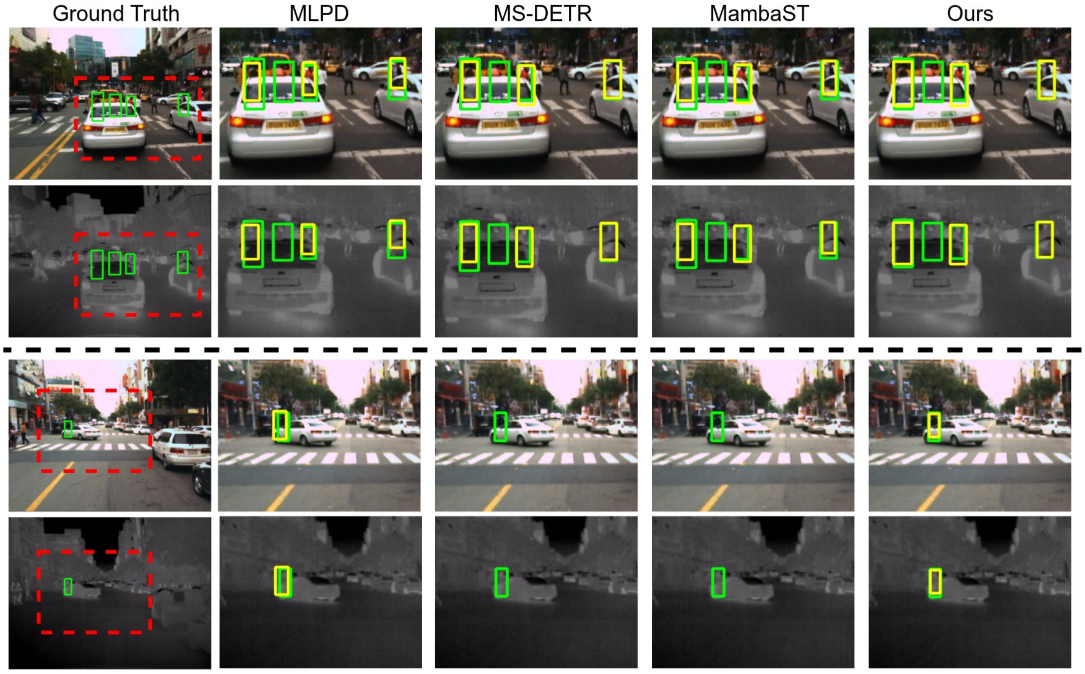

Figure R1: Extension of Fig. 4 from our paper, we have included the thermal images. Visual examples of KAIST detection results for heavily occluded pedestrians. Green and yellow bounding boxes correspond to ground truth and detection results, respectively. Ours show better detection on occluded/small pedestrians.