Supplementary Materials of "Accurate Dynamic SLAM using CRF-based Long-term Consistency"

I. PARAMETERS CROSS VALIDATION

We have 12 parameters, i.e. $\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta, \mu_\gamma, \sigma_\gamma, \sigma_P, \sigma_p, w^1, w^2, t, c$, mainly used in the CRF inference of dynamic landmark detection (Section 3.3) in our LC-CRF SLAM system. To justify the optimized parameter configuration, we performed a parameter study with cross validation on these parameters. We first group these parameters into 6 pairs: $\{\mu_\alpha, \sigma_\alpha\}$, $\{\mu_\beta, \sigma_\beta\}$, $\{\mu_\gamma, \sigma_\gamma\}$, $\{\sigma_P, \sigma_p\}$, $\{w^1, w^2\}$, and $\{t, c\}$. The first five pairs are parameters used in the pairwise potentials, and the landmark static likelihood threshold t and static confidence threshold c are used in the unary potential. In the parameter study, we set the range of these parameters as: $\{\mu_\alpha, \sigma_\alpha\} \in [1.1, 2.0] \times [0.2, 2.0]$, $\{\mu_\beta, \sigma_\beta\} \in [3.8, 5.6] \times [1.3, 2.2]$ and $\{\mu_\gamma, \sigma_\gamma\} \in [0.1, 1.0] \times [0.1, 1.0]$, $\{\sigma_P, \sigma_p\} \in [0.1, 1.0] \times [10, 28]$, $\{w^1, w^2\} \in [2, 20] \times [10, 55]$, $\{t, c\} \in [0.1, 1.0] \times [0.55, 1.00]$.

Since the enumeration of all the parameter configuration is too large, to make a practical study, we choose to select the parameter configuration of the 6 parameter pairs one by one. Specifically, for a pair of parameters to be studied, we evenly sample n candidate values of the parameters in this parameter pair and randomly select m candidate values of the parameters from the other 5 parameter pairs. Then in total we get $6 \times n \times m$ parameter configuration in the parameter study. For each parameter configuration, we perform our LC-CRF SLAM on the TUM RGB-D dynamic dataset and calculate the average ATE as the accuracy measurement of this parameter configuration. Typically, we set $n = 10 \times 10$, $m = 10$ and the detailed configurations can be found in Table I. Fig. 1-6 shows the average ATE measurements of the total $6 \times n \times m$ parameter configuration in our parameter study.

II. MORE COMPARISON RESULTS ON BONN DATASET

We demonstrate more compassion results performed on the Bonn dataset, including 20 RGB-D scan sequences of dynamic scenes which are more challenging than the TUM dynamic dataset. In the comparison, we choose three state-of-the-art dynamic reconstruction methods, i.e. Refusion(RF), StaticFusion(SF) and MaskFusion(MF) to compare with our LC-CRF SLAM method. Fig. 7 and Fig. 8 show that estimated camera pose trajectories (along with the ground truth trajectories) for the four methods in each sequence of the Bonn dataset. The red line in each sequence shows the ATE error between the estimated camera pose trajectories and ground truth trajectories at every frame timestamp. As we can see, it is clear that the trajectories estimated by our LC-CRF SLAM are much more closer to the ground truth trajectories than other methods.

III. EVALUATION ON OUTDOOR ENVIRONMENT

We also perform an accuracy evaluation for camera pose estimation on the outdoor environment. Since there are not too many public release RGB-D datasets due to the missing of depth camera sensor for the outdoor environment, we choose KITTI odometry dataset to perform the evaluation, which contains stereo RGB data streams scan for a large scale outdoor environment. To make our system feasible on the KITTI dataset, we made a slight modification of our system, by using the estimated depth from stereo RGB image pair as the landmark's depth prior for the camera pose estimation and leaving the main part of our approach (the dynamic detection part in the camera tracking) unchanged. Fig.9 shows the estimated camera pose trajectories (blue) by our approach and the ground truth trajectories (red) of sequence 01-10 in the KIITI dataset. As we can see, the estimated camera pose trajectories are close to the ground truth trajectories, which means that our approach can also give reliable camera pose estimation in the outdoor environments.

TABLE I
PARAMETER CONFIGURATIONS.

Paramters	μ_α	σ_α	μ_β	σ_β	μ_γ	σ_γ	σ_P	σ_p	w^1	w^2	t	c
	-	-	5.37	1.96	0.89	0.7	0.29	25.52	14.19	18.13	0.93	0.86
	-	-	4.39	2.1	0.61	0.59	0.73	19.43	7.87	36.9	0.89	0.66
	-	-	4.57	2.2	0.6	0.93	0.16	12.88	11.1	10.51	0.56	0.67
	-	-	4.45	1.74	0.68	0.95	0.56	24.15	16.83	15.69	0.74	0.69
$\mu_\alpha, \sigma_\alpha$	-	-	4.42	1.33	0.13	0.57	1.0	18.5	14.47	53.67	0.64	0.9
	-	-	4.93	2.06	0.44	0.19	0.64	13.95	12.71	41.84	0.41	0.87
	-	-	4.12	1.94	0.76	0.74	0.49	25.76	4.12	22.59	0.67	0.61
	-	-	4.06	1.77	0.81	0.88	0.18	14.89	12.26	29.55	0.49	0.66
	-	-	3.98	2.19	0.24	0.58	0.53	13.37	19.53	42.31	0.11	0.95
	-	-	3.95	1.49	0.51	0.39	0.34	12.28	15.32	37.8	0.49	0.73
	1.61	0.51	-	-	0.97	0.62	0.41	16.62	11.11	30.34	0.97	0.69
	1.35	1.05	-	-	0.16	0.51	0.32	17.91	15.57	49.03	0.61	0.98
	1.6	1.63	-	-	0.59	0.87	0.88	17.33	14.92	43.75	0.36	0.95
	1.67	0.48	-	-	0.75	0.34	0.41	18.45	3.56	51.29	0.49	0.69
	1.78	0.81	-	-	0.13	0.3	0.15	13.4	11.7	44.58	0.69	0.95
μ_β, σ_β	1.52	0.71	-	-	0.32	0.79	0.78	18.35	12.37	10.91	0.86	0.8
	1.13	0.62	-	-	0.53	0.92	0.44	16.16	14.23	29.37	0.13	0.72
	1.58	1.46	-	-	0.73	0.43	0.68	13.92	2.91	27.83	0.88	0.67
	1.18	0.78	-	-	0.77	0.66	0.37	23.73	9.25	24.67	0.71	0.76
	1.63	1.11	-	-	0.51	0.4	0.9	26.27	11.35	48.44	0.67	0.67
	1.73	0.38	5.09	1.65	-	-	0.12	10.95	14.48	43.94	0.63	0.68
	1.15	1.63	5.23	1.9	-	-	0.66	15.03	10.12	36.22	0.91	0.93
	1.57	0.31	5.45	2.07	-	-	0.74	26.63	15.76	48.56	0.74	0.85
	1.99	1.32	4.16	1.89	-	-	0.74	22.04	3.47	15.29	0.56	0.95
	1.73	1.77	4.74	1.89	-	-	1.0	17.57	15.48	29.27	0.92	0.77
$\mu_\gamma, \sigma_\gamma$	1.77	1.38	4.44	1.55	-	-	0.26	17.25	15.53	45.19	0.37	0.88
	1.81	0.97	5.12	1.67	-	-	0.73	25.92	9.43	34.41	0.18	0.93
	1.25	1.51	4.57	1.65	-	-	0.2	23.64	18.36	14.69	1.0	0.81
	1.59	0.65	3.95	1.51	-	-	0.3	25.75	12.38	42.59	0.35	0.84
	1.52	0.3	4.57	1.58	-	-	0.12	23.27	13.43	31.91	0.49	0.69
	1.82	0.21	3.99	2.14	0.71	0.6	-	-	13.0	45.86	0.69	0.55
	1.91	1.82	5.39	2.13	0.16	0.46	-	-	15.01	16.1	0.63	0.64
	1.58	1.18	4.32	1.48	0.77	0.4	-	-	7.39	49.24	0.14	0.91
	1.11	1.93	5.17	2.13	0.69	0.53	-	-	9.51	18.03	0.64	0.78
	1.56	0.51	5.6	2.18	0.82	0.64	-	-	11.25	47.47	0.81	0.76
σ_P, σ_p	1.61	0.54	4.97	2.06	0.81	0.54	-	-	19.66	39.25	0.25	0.74
	1.92	0.23	5.58	1.42	0.89	0.17	-	-	15.19	48.57	0.12	0.63
	1.57	0.34	5.32	2.19	0.96	0.58	-	-	12.34	39.68	0.26	0.91
	1.39	0.79	5.2	1.39	0.3	0.8	-	-	10.35	15.44	0.2	0.72
	1.86	0.53	4.01	2.09	0.33	0.55	-	-	9.59	43.95	0.85	0.6
	1.5	1.19	4.79	1.9	0.41	0.74	0.52	16.44	-	-	0.37	0.81
	1.93	0.84	4.69	2.08	0.51	0.47	0.94	21.84	-	-	0.86	0.92
	1.52	0.85	4.18	1.66	0.61	0.34	0.44	18.27	-	-	0.32	0.99
	1.14	0.76	5.06	2.01	0.79	0.44	0.75	23.0	-	-	0.88	0.75
w^1, w^2	1.65	1.62	5.45	1.58	0.14	0.42	0.37	10.21	-	-	1.0	0.74
	1.52	1.34	4.76	1.57	0.3	0.82	0.93	10.69	-	-	0.16	0.64
	1.87	1.93	5.59	1.57	0.74	0.23	0.19	19.93	-	-	0.97	0.7
	1.4	1.06	4.89	1.99	0.15	0.45	0.14	27.43	-	-	0.41	0.6
	1.88	0.49	4.44	2.09	0.71	0.35	0.67	22.2	-	-	0.64	0.99
	1.6	0.58	5.08	1.55	0.4	0.57	0.28	13.23	-	-	0.29	0.7
	1.53	1.61	5.28	1.59	0.26	0.42	0.58	26.71	9.47	40.2	-	-
	1.91	0.52	3.95	1.79	0.22	0.64	0.55	11.15	16.71	50.42	-	-
	1.4	0.65	3.87	2.09	0.97	0.83	0.69	26.09	9.94	19.24	-	-
	1.89	0.73	5.57	1.74	0.59	0.78	0.67	24.74	17.13	28.9	-	-
t, c	1.2	0.95	3.83	1.69	0.85	0.91	0.74	14.8	11.9	32.33	-	-
	1.52	1.54	5.47	1.87	0.74	0.31	0.52	19.47	18.74	51.4	-	-
	1.52	0.91	5.14	1.81	0.27	0.85	0.83	12.62	6.82	16.94	-	-
	1.2	1.4	5.55	1.53	0.14	0.43	0.13	15.95	8.04	30.01	-	-
	1.34	1.56	4.91	1.49	0.23	0.99	0.49	11.71	5.29	29.1	-	-
	1.8	0.7	4.36	1.98	0.67	0.18	0.35	21.52	12.27	15.64	-	-

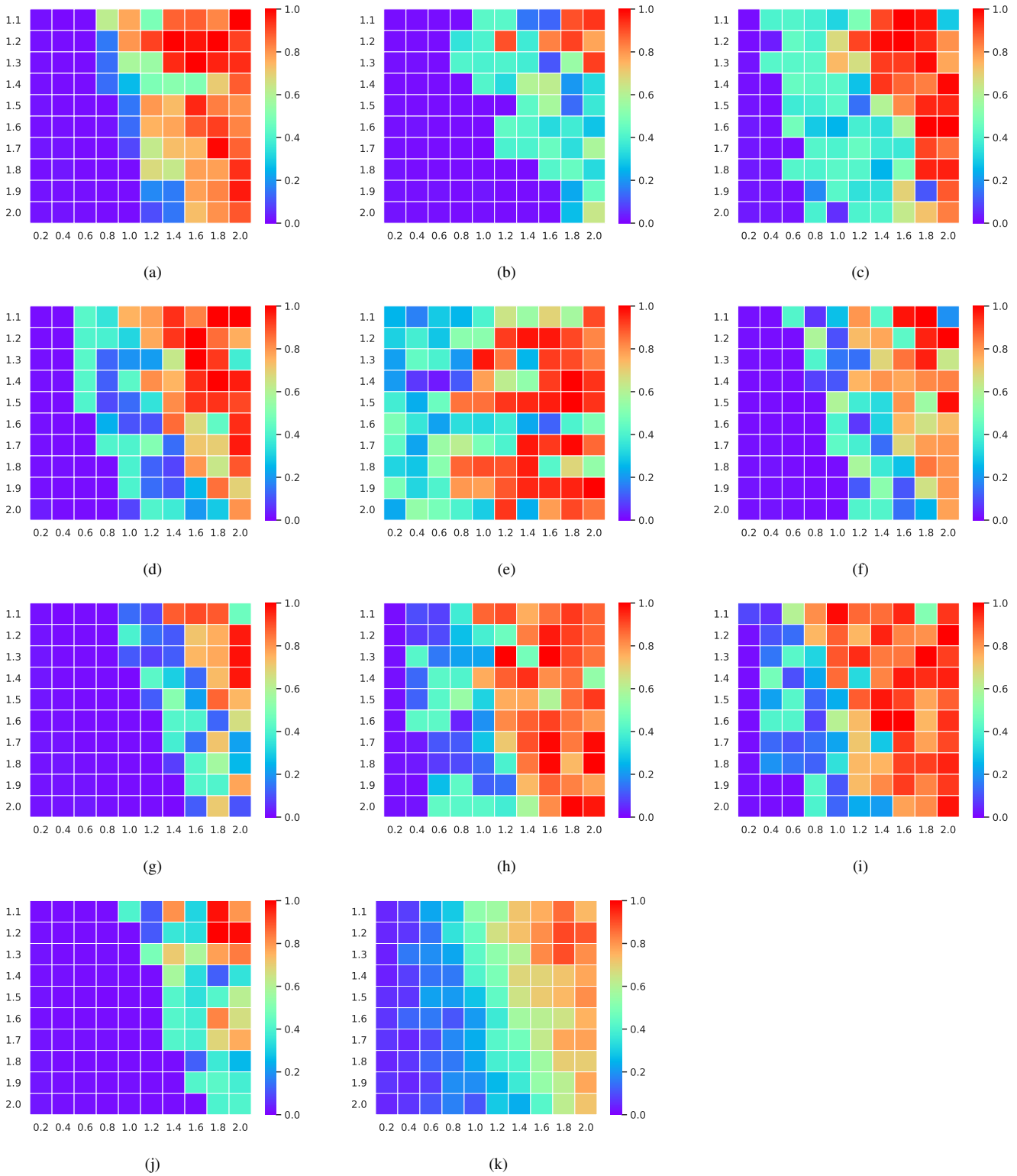


Fig. 1. The visualization of ATE accuracy measurement of parameter configurations for parameter pair $\{\mu_\alpha, \sigma_\alpha\}$. Each sub-figure from (a)-(j) represents the ATE accuracy measurement of the $n = 10 \times 10$ evenly sampled parameters in $\{\mu_\alpha, \sigma_\alpha\}$ with the parameters from the other 5 pairs randomly selected; sub-figure (k) is the average ATE. Red presents higher error, blue lower.

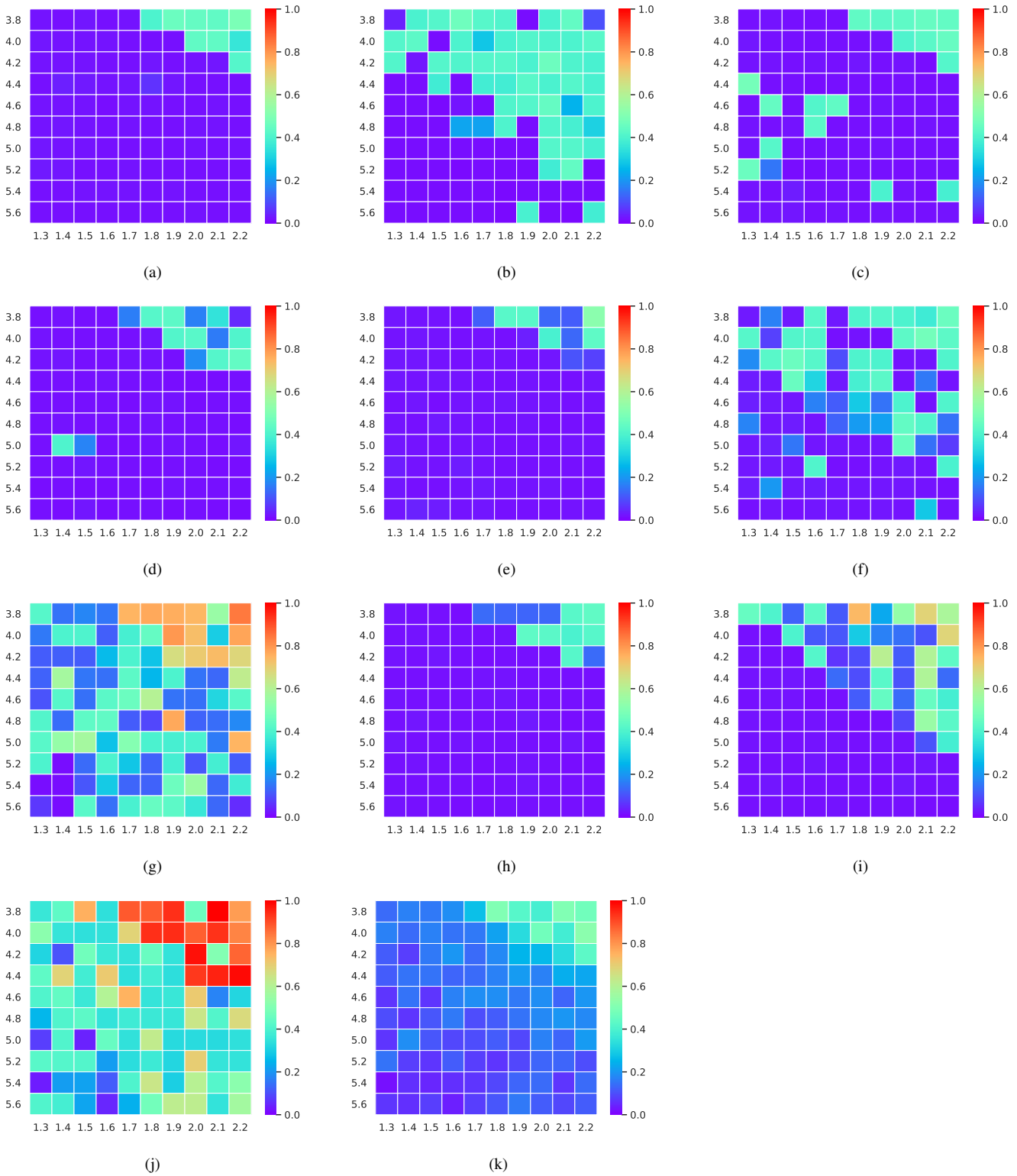


Fig. 2. The visualization of ATE accuracy measurement of parameter configurations for parameter pair $\{\mu_\beta, \sigma_\beta\}$. Each sub-figure from (a)-(j) represents the ATE accuracy measurement of the $n = 10 \times 10$ evenly sampled parameters in $\{\mu_\beta, \sigma_\beta\}$ with the parameters from the other 5 pairs randomly selected; sub-figure (k) is the average ATE. Red presents higher error, blue lower.

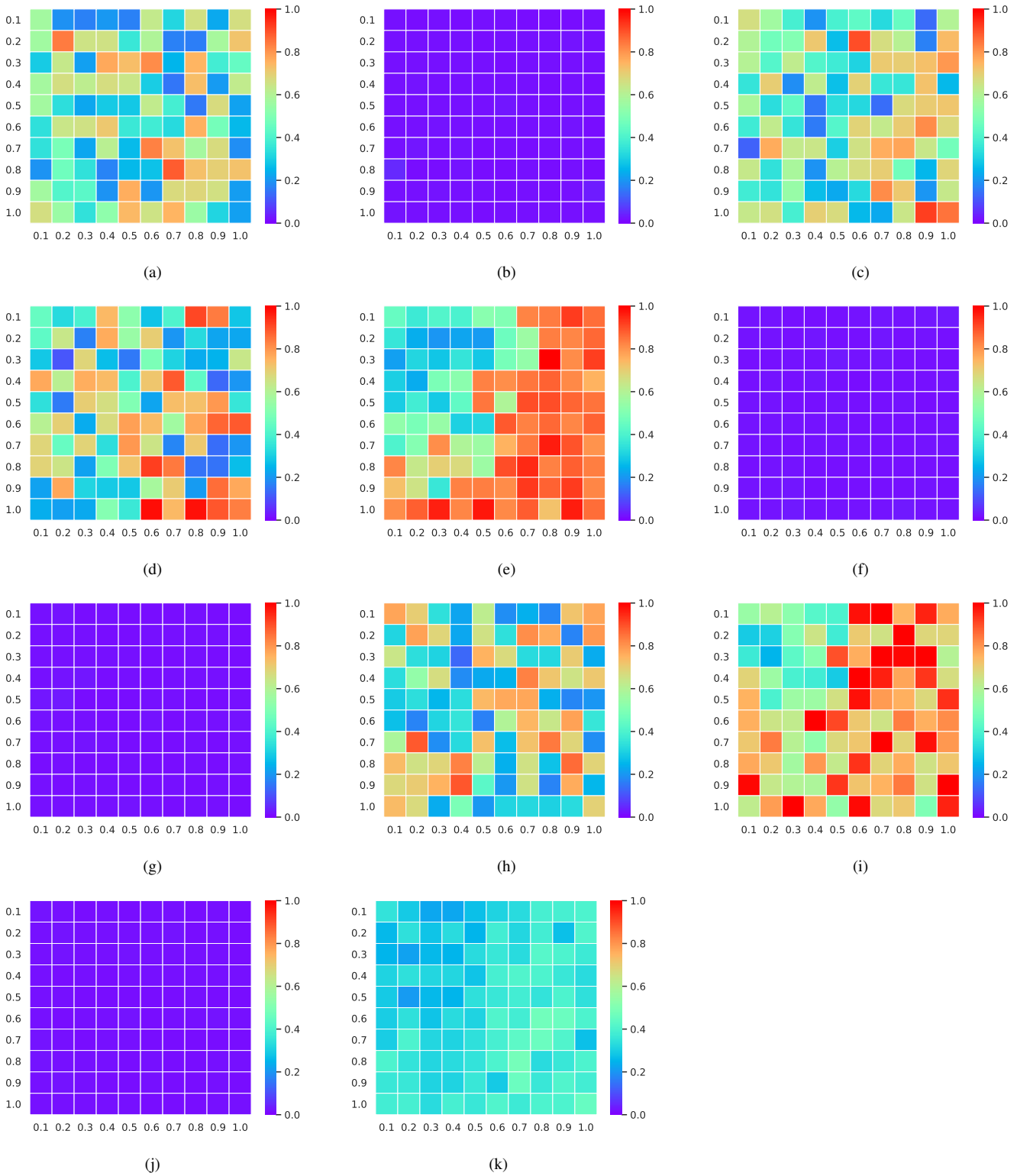


Fig. 3. The visualization of ATE accuracy measurement of parameter configurations for parameter pair $\{\mu_\gamma, \sigma_\gamma\}$. Each sub-figure from (a)-(j) represents the ATE accuracy measurement of the $n = 10 \times 10$ evenly sampled parameters in $\{\mu_\gamma, \sigma_\gamma\}$ with the parameters from the other 5 pairs randomly selected; sub-figure (k) is the average ATE. Red presents higher error, blue lower.

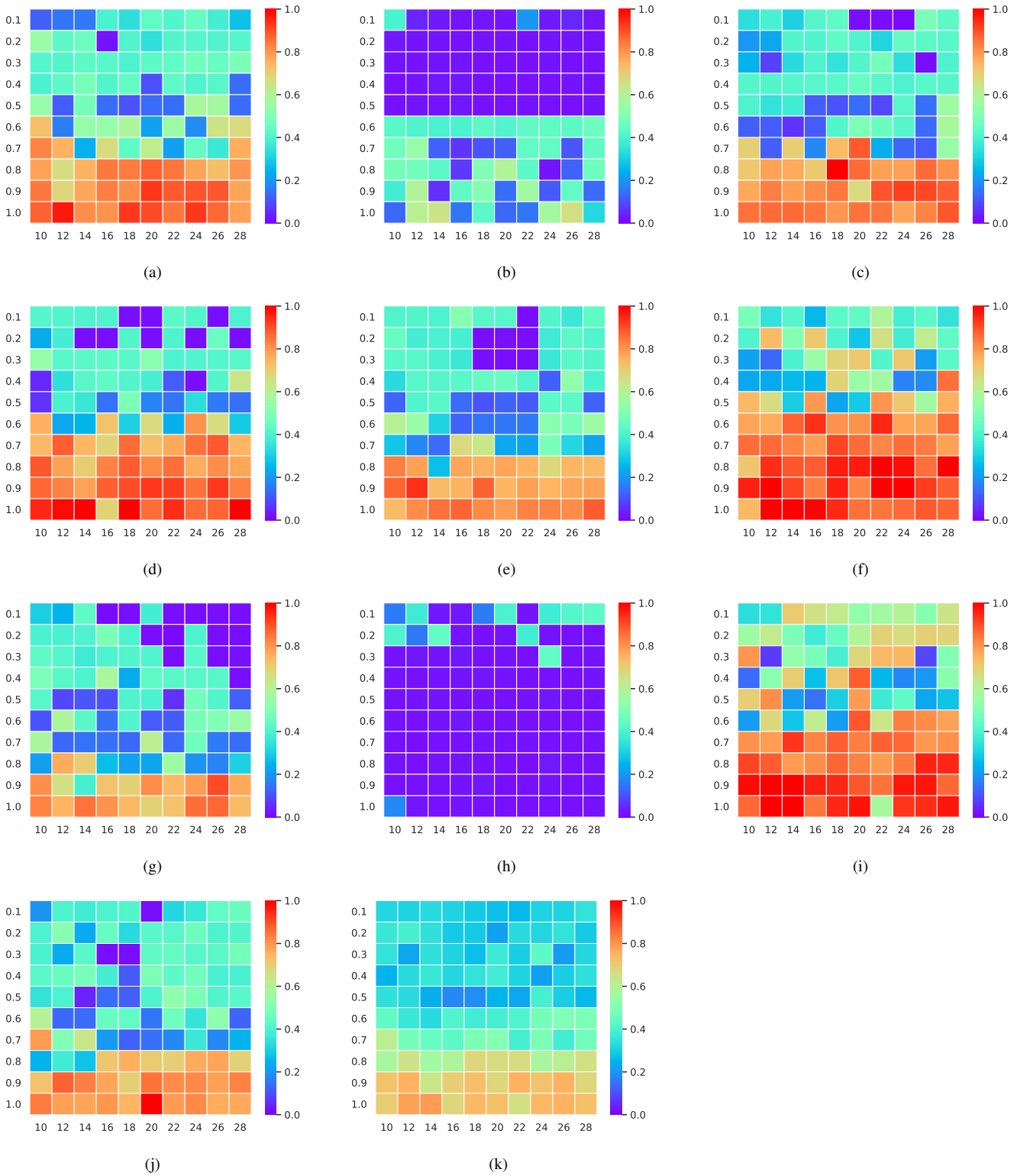


Fig. 4. The visualization of ATE accuracy measurement of parameter configurations for parameter pair $\{\sigma_P, \sigma_p\}$. Each sub-figure from (a)-(k) represents the ATE accuracy measurement of the $n = 10 \times 10$ evenly sampled parameters in $\{\sigma_P, \sigma_p\}$ with the parameters from the other 5 pairs randomly selected; sub-figure (k) is the average ATE. Red presents higher error, blue lower.

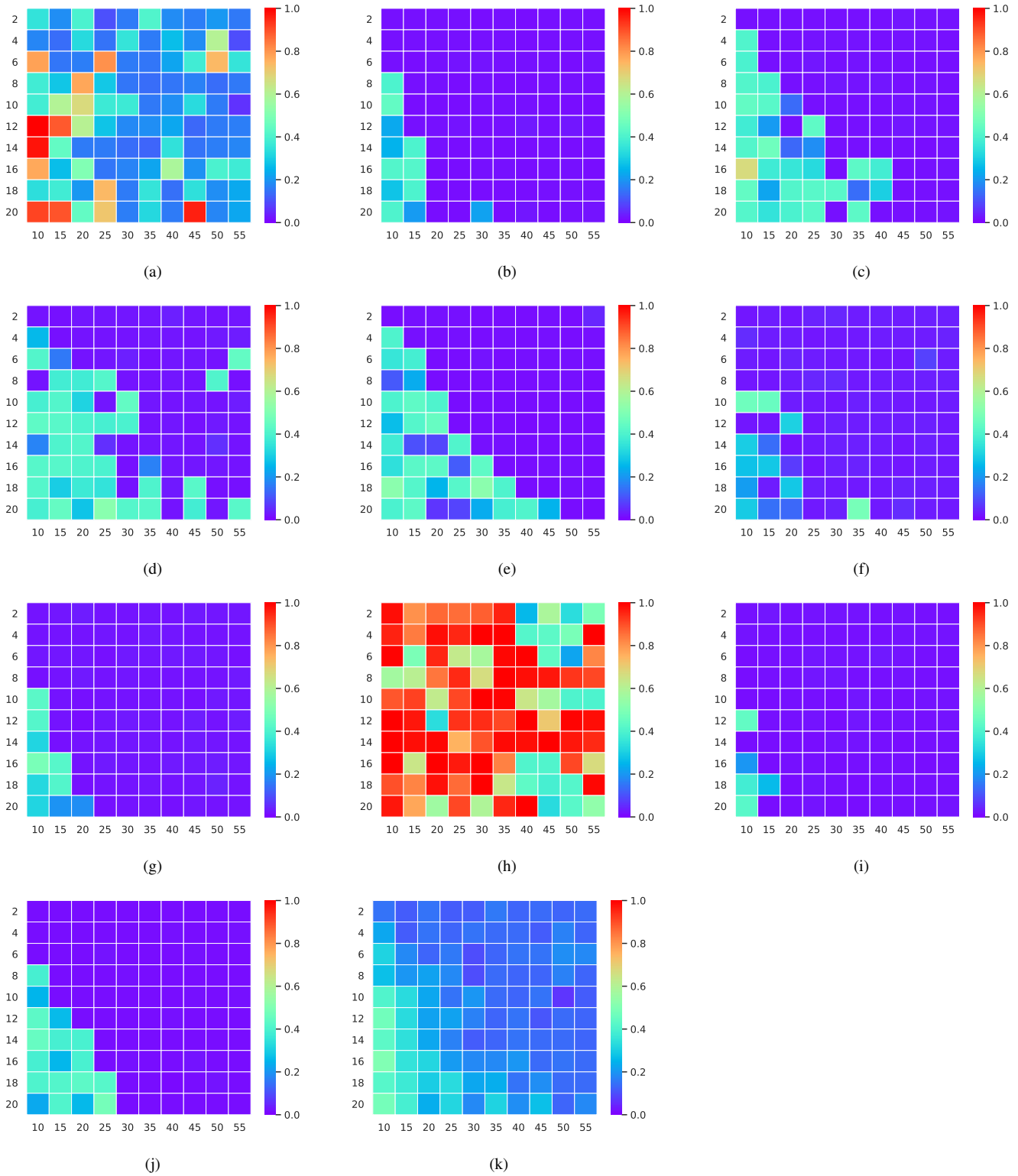


Fig. 5. The visualization of ATE accuracy measurement of parameter configurations for parameter pair $\{w^1, w^2\}$. Each sub-figure from (a)-(k) represents the ATE accuracy measurement of the $n = 10 \times 10$ evenly sampled parameters in $\{w^1, w^2\}$ with the parameters from the other 5 pairs randomly selected; sub-figure (k) is the average ATE. Red presents higher error, blue lower.

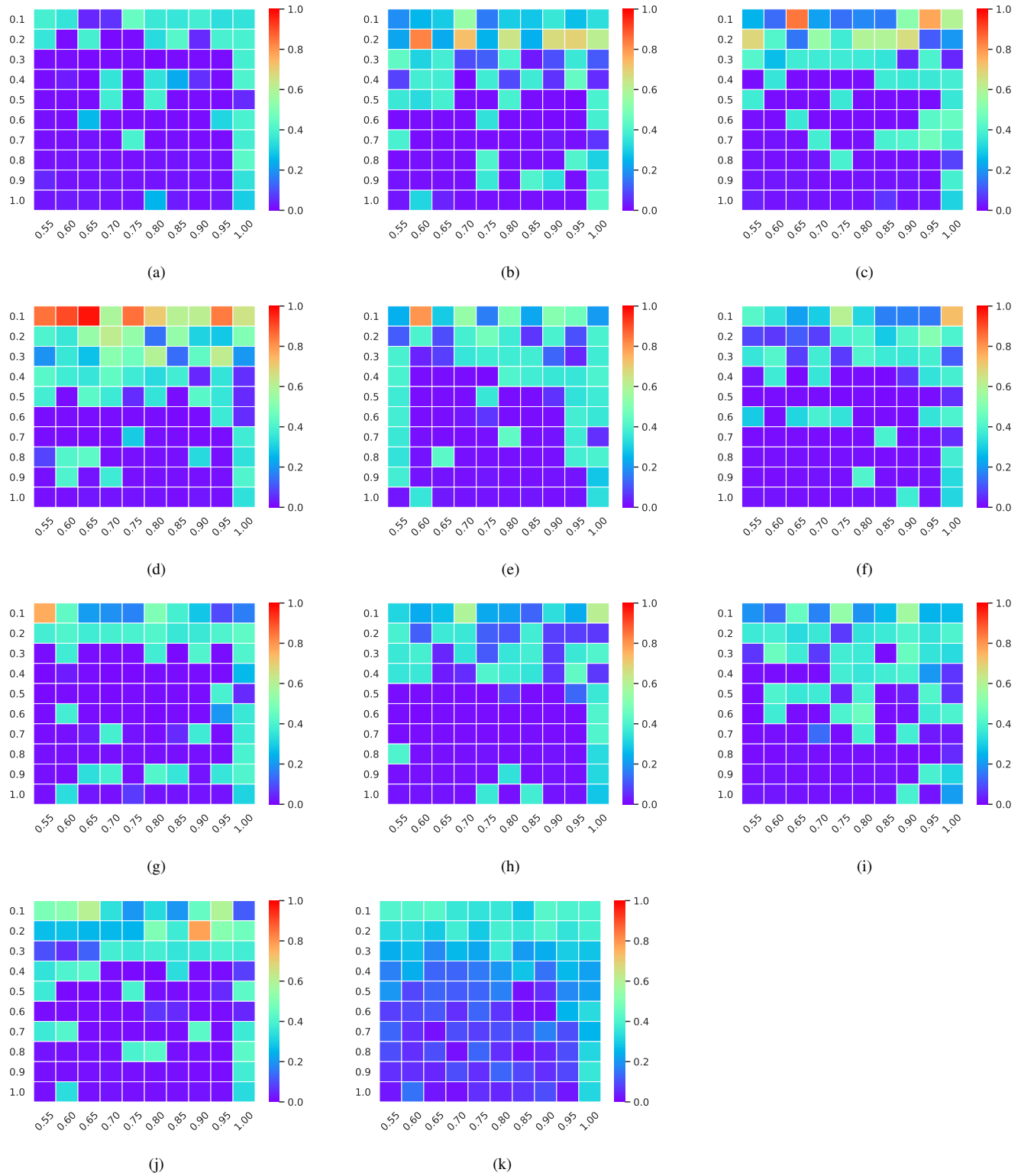


Fig. 6. The visualization of ATE accuracy measurement of parameter configurations for parameter pair $\{t, c\}$. Each sub-figure from (a)-(j) represents the ATE accuracy measurement of the $n = 10 \times 10$ evenly sampled parameters in $\{t, c\}$ with the parameters from the other 5 pairs randomly selected; sub-figure (k) is the average ATE. Red presents higher error, blue lower.

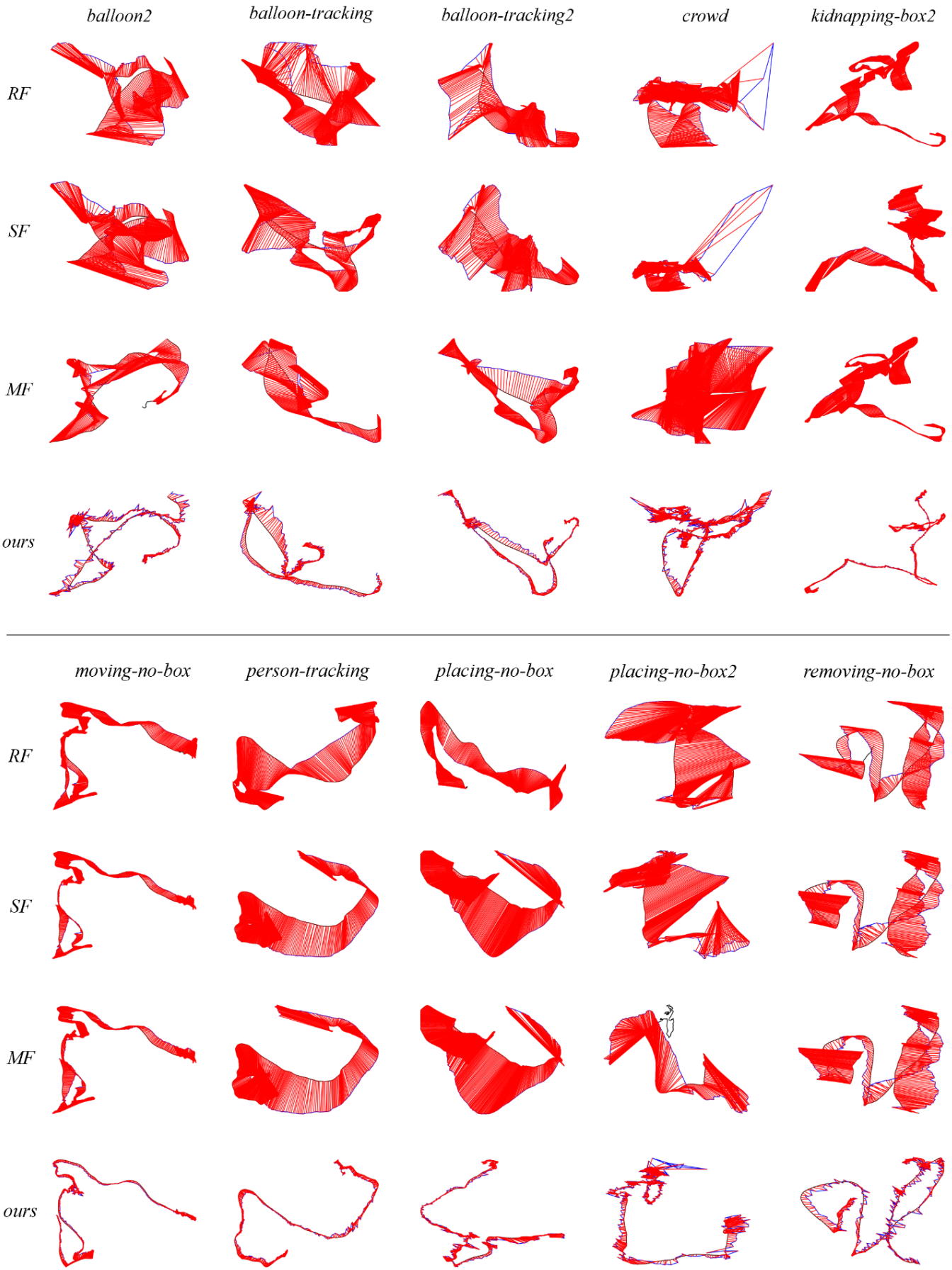


Fig. 7. The visualization of ATE errors with Refusion(RF),StaticFuson(SF), MaskFusion(MF) and our method on the Bonn dataset.

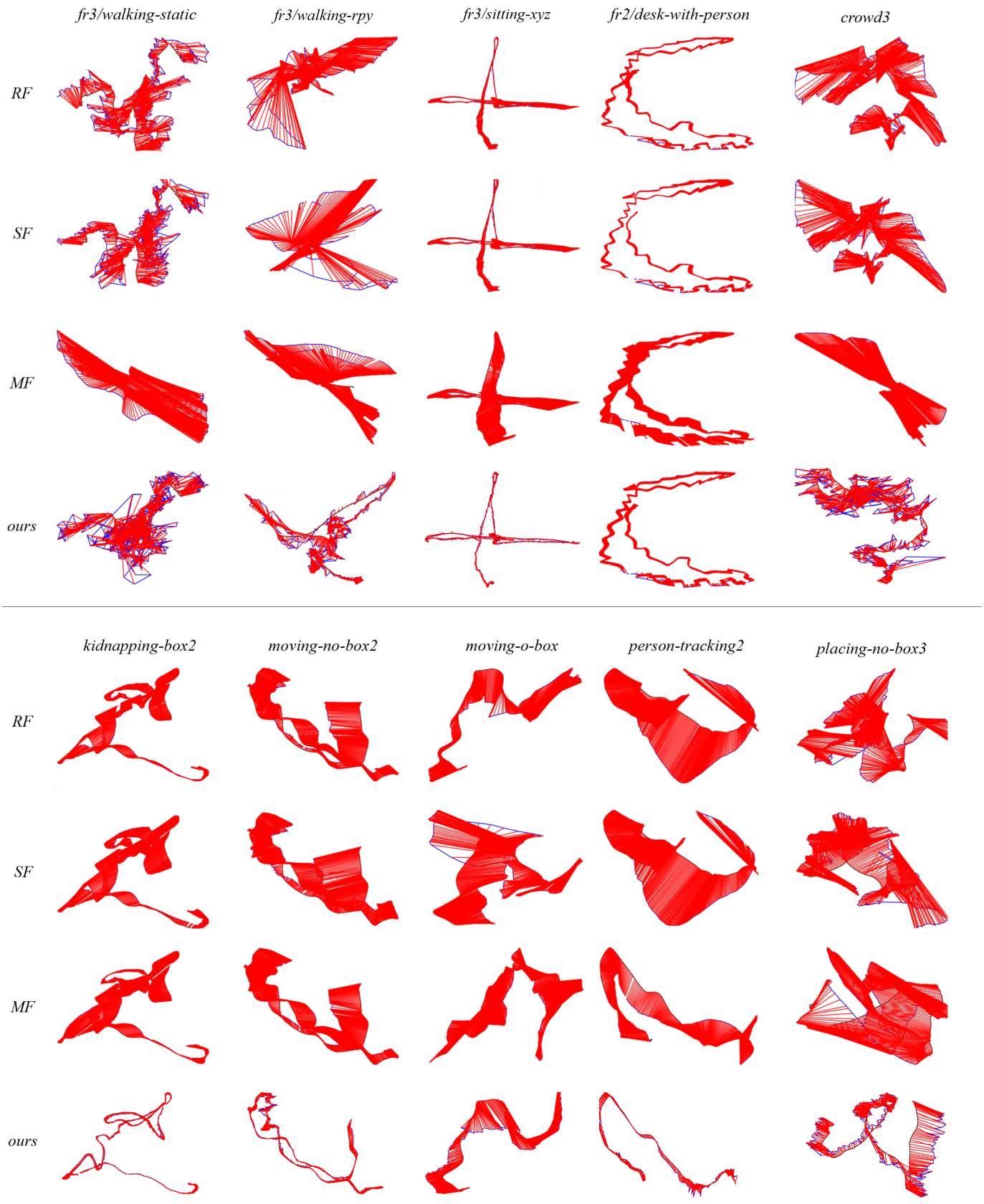
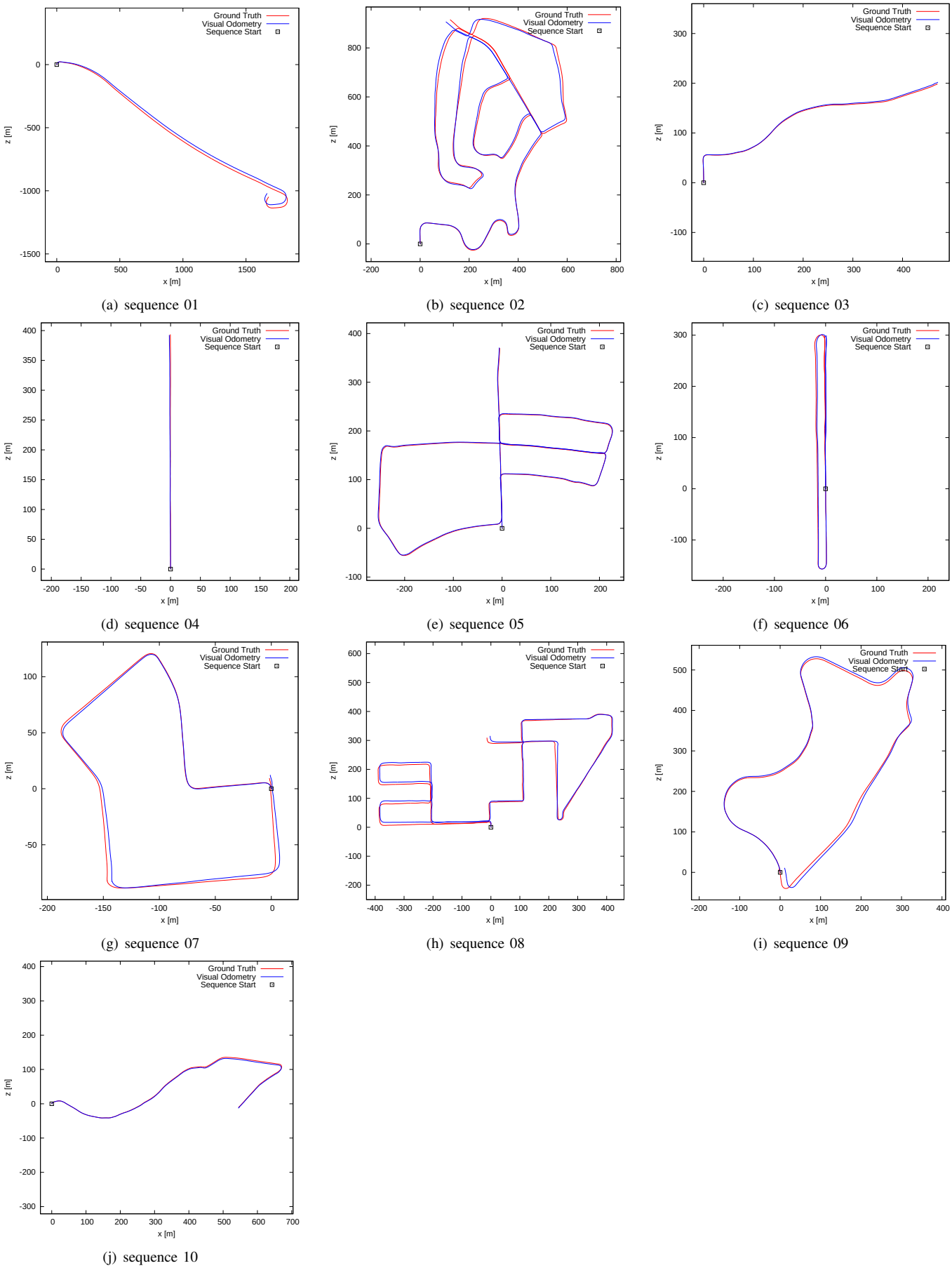


Fig. 8. The visualization of ATE errors with Refusion(RF),StaticFuson(SF), MaskFusion(MF) and our method on the Bonn dataset.



(a) sequence 01 (b) sequence 02 (c) sequence 03

(d) sequence 04 (e) sequence 05 (f) sequence 06

(g) sequence 07 (h) sequence 08 (i) sequence 09

(j) sequence 10

Fig. 9. The estimated camera pose trajectories by our approach and the ground truth trajectories of sequence 01-10 in KITTI dataset. Red: ground truth trajectories, Blue: estimated camera pose trajectories by our approach.