Visual Story Writing of SNSM for VPR

March 31, 2025

- Mobile robot autonomy
-

1 Visual Story Writing of SNSM

Visual Place Recognition (VPR) has many challenges in practice due to rotation, viewpoint, scale changes, occlusion, weather artefacts, etc., between reference and query images (Fig. 1).

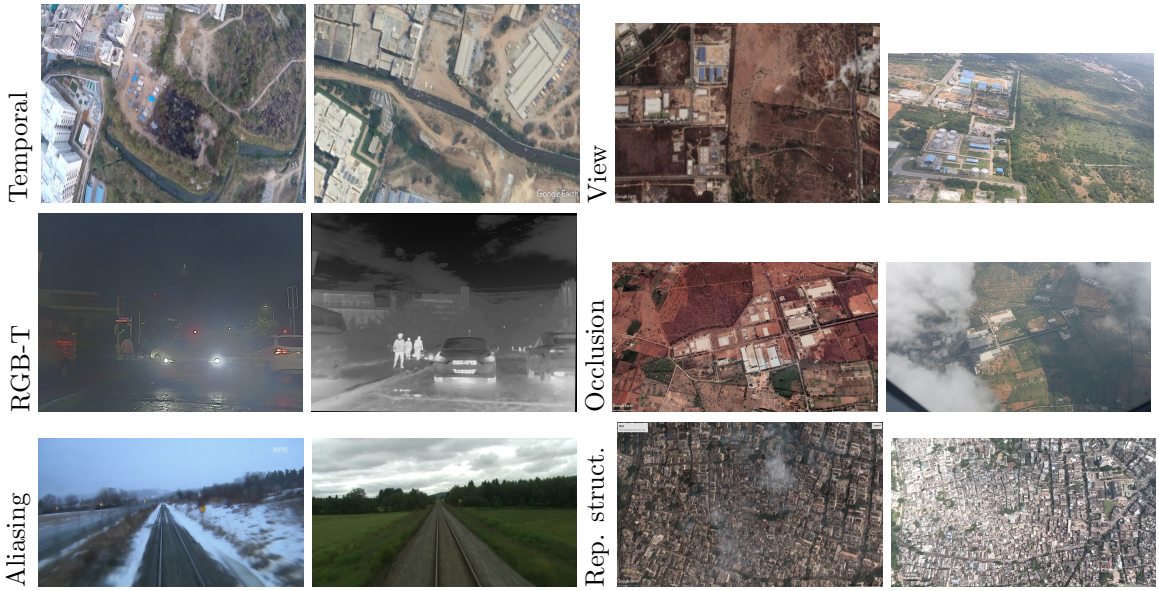

Figure 1: Sample images in different scenarios

In this work, we investigate the illumination variation, which is crucial for an all-day VPR algorithm. In general, RGB cameras are sensitive to light, where they fail to capture the scene in poorly lit scenarios (Fig. 2).

Figure 2: Night scene using RGB camera

Alternatively, thermal sensors are used.

Figure 3: Night scene in Fig. 2 using thermal camera

However, due to the distinct sensing mechanisms of RGB and thermal cameras, significant appearance differences are introduced in the captured images. In practice, there could be a scenario where a query could contain a combination of all of these distortions, making it highly complex to develop a universal VPR algorithm. As most VPR techniques address the VPR problem in the image retrieval framework, this work studies the performance of various VPR-specific methods for RGB-thermal retrieval. We consider a few publicly available RGB-T datasets originally introduced for different vision applications such as image fusion and tracking. Representative pairs of these datasets are illustrated below (Fig. 4).

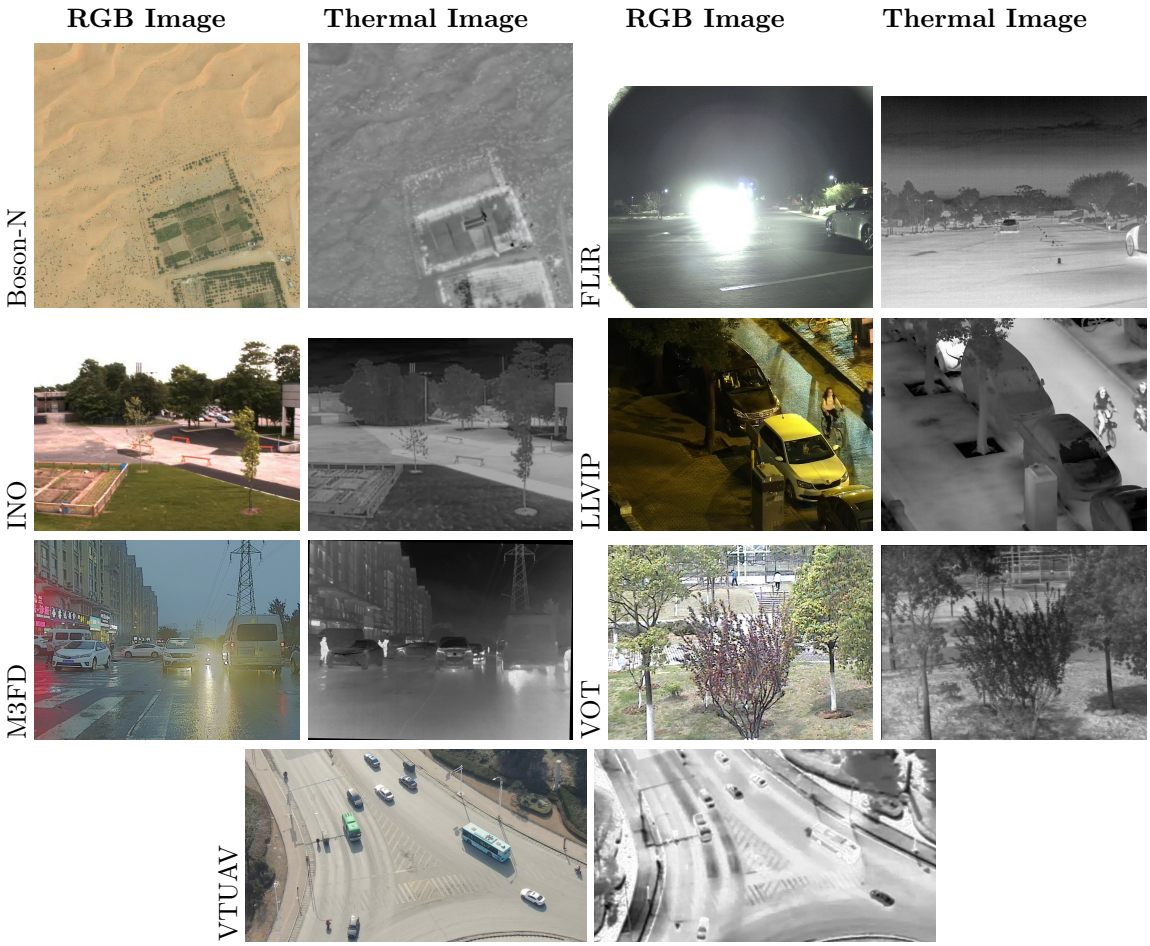

Figure 4: Sample RGB-T image pairs of different datasets used in this work

Characteristics of these datasets are presented in Table 1.

Table 1: Details of the test datasets. The OTDD is relatively large for all the datasets that contain urban scenes. This points to a large domain gap.

Dataset	Number of RGB-T pairs	Number of cities	Static backgound	Primary task	Location labels	Terrain type	OTDD
Boson-N	266	1	✗	RGB-T localization	✓	Desert	1577.88
FLIR	375	8	✗	RGB-T fusion for object detection	✗	Urban	4035.85
INO	428	10	✓	RGB-T registration for fusion	✗	-	3188.33
LLVIP	347	6	✓	Fusion, pedestrian detection, I2I translation	✗	Urban	3151.61
M3FD	300	-	✗	Object detection	✗	Urban	2561.3
VOT	363	16	✓	Object tracking	✗	Urban	4552.82
VTUAV	344	-	✗	Aerial tracking	✗	Urban	2687.91

Similarly, the recall performance of VPR-specific models on the above RGB-T datasets is presented in Table 2.

Table 2: Recalls of the popular VPR models on RGB-T datasets.

	MixVPR		SALAD		CricaVPR		SNSM	
	R@1	R@5	R@1	R@5	R@1	R@5	R@1	R@5
BOSON-N	18.05	38.35	9.4	21.05	9.4	22.56	**38.35**	**58.65**
FLIR	13.07	44.53	**18.4**	**53.6**	14.93	48.53	17.60	39.73
INO	8.64	27.1	4.44	17.06	5.61	20.33	**39.72**	**65.19**
LLVIP	9.8	27.09	5.76	17.87	5.76	17.58	**66.57**	**82.13**
M3FD	69	94.67	70.67	**97.33**	72.67	96.33	**89**	96
VOT	14.88	49.86	9.37	30.03	10.19	36.36	**66.94**	**85.40**
VTUAV	73.84	91.57	37.21	69.19	34.3	66.86	**97.38**	**99.42**

2 EXperiments

- Visual retrieval
- Similarity plot for a common query; This is if models can capture any trend
- t-SNEs
- Activation maps of the backbone and the SNSM maps