Sparse-to-Dense Hypercolumn Matching for Long-Term Visual Localization

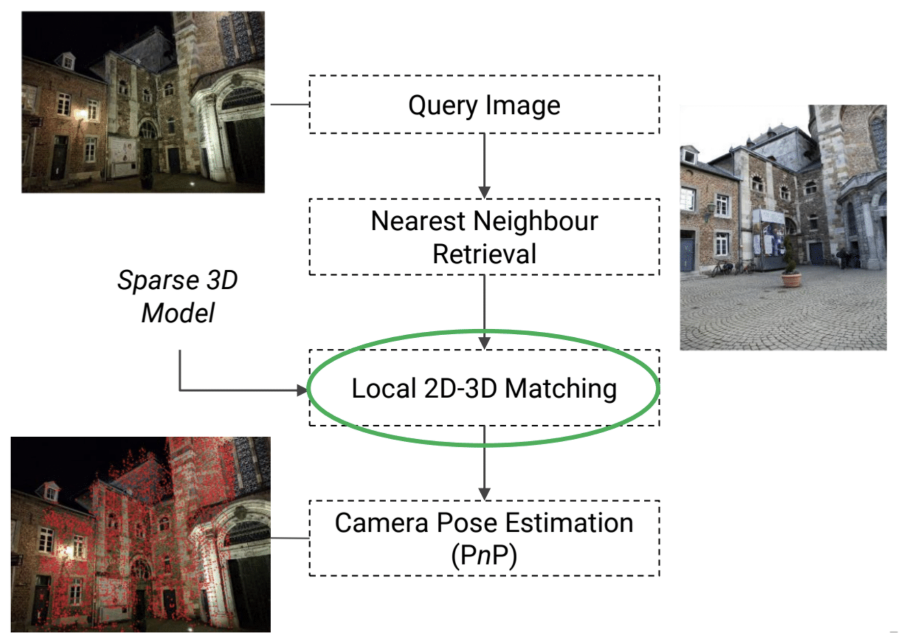
Hugo Germain[1] Guillaume Bourmaud[2] Vincent Lepetit[1]

[1]Laboratoire Bordelais de Recherche en Informatique, Université de Bordeaux, France
[2]Laboratoire IMS, Université de Bordeaux, France

Contribution

We show that exhaustive search for keypoint correspondents (**sparse-to-dense matching**) outperforms traditional keypoint matching in challenging conditions.

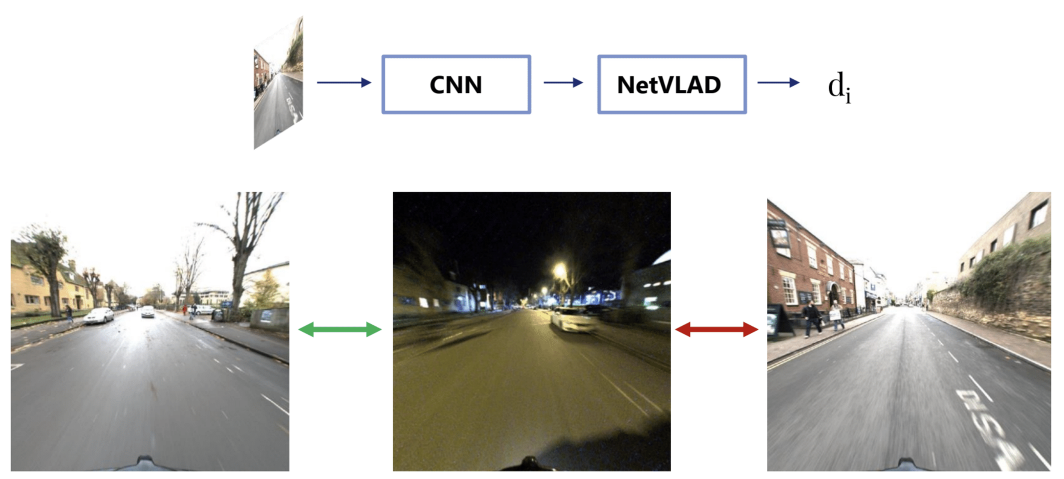
Pipeline



We consider a hierarchical localization pipeline, similar to HF-Net:Given a query image, we first identify a set of prior locations using image retrieval. We then match feature points across the 2D query image and the retrieved local 3D point cloud. This step, however, is prone to fail as it is still very difficult to detect and match sparse feature points across very different conditions.

Retraining NetVLAD

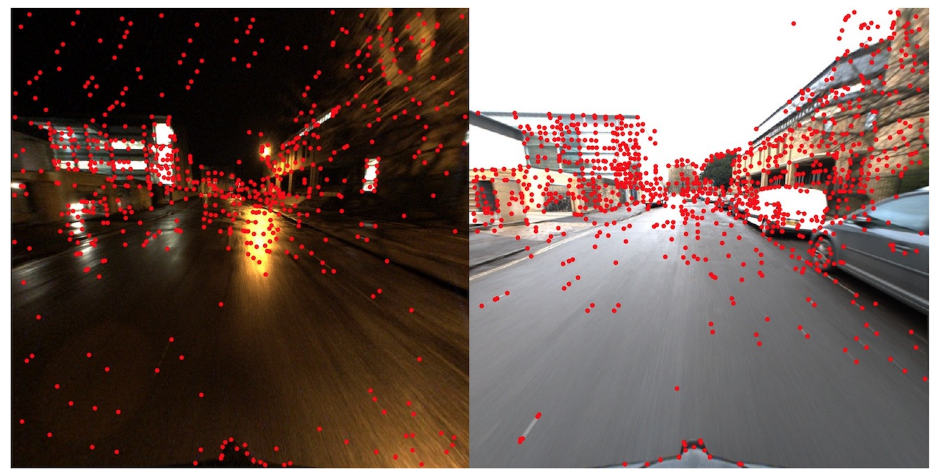
We train our image-retrieval pipeline using the popular pooling layer NetVLAD.



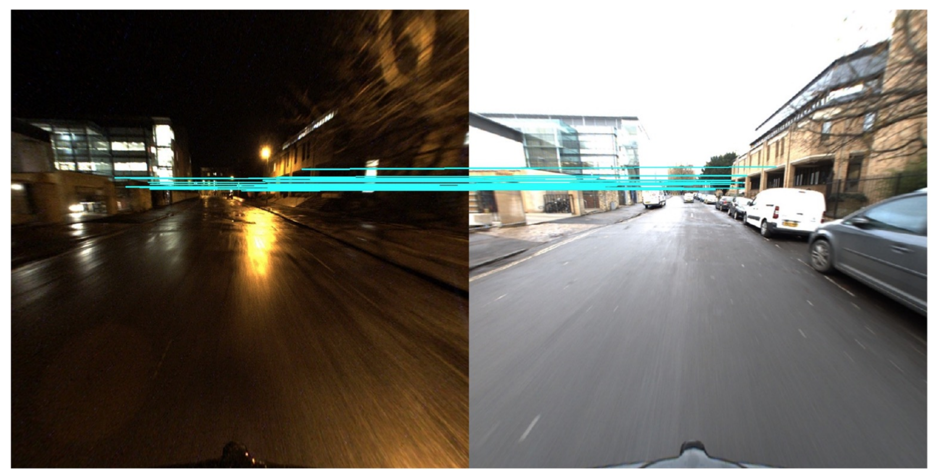
Traditional Matching

Traditional approaches for feature point matching consist of:

- Sparse feature detection in both images:



- Feature matching using generic feature descriptors:

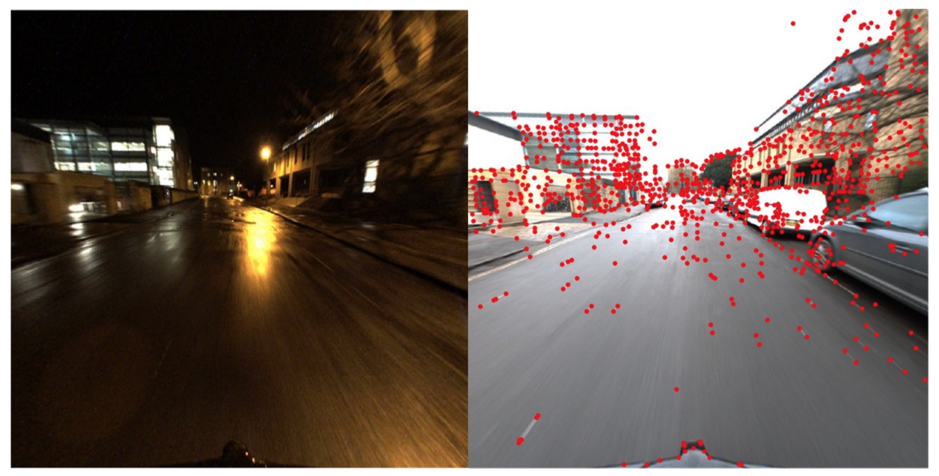


But in a long-term scenario, **consistent detection and matching is hard**.

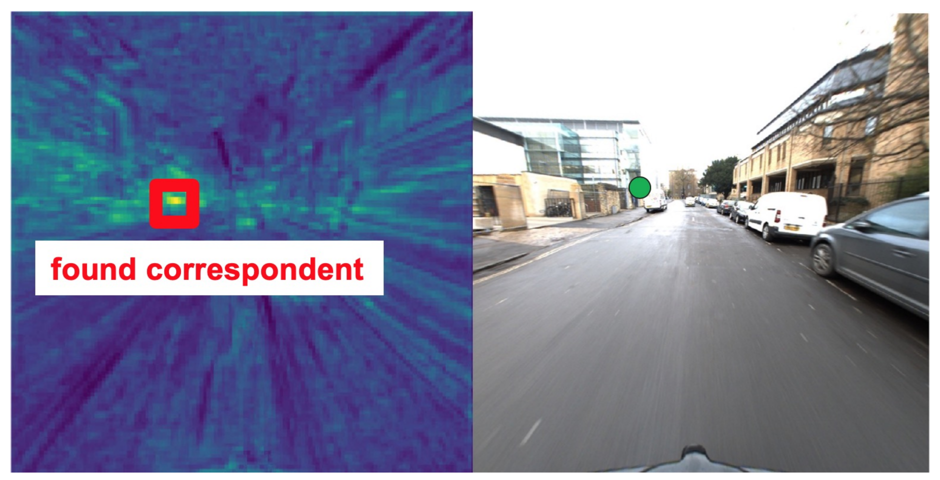
Sparse-to-Dense Matching **(Ours)**

We argue that:

- Detection should only be performed in the reference image:

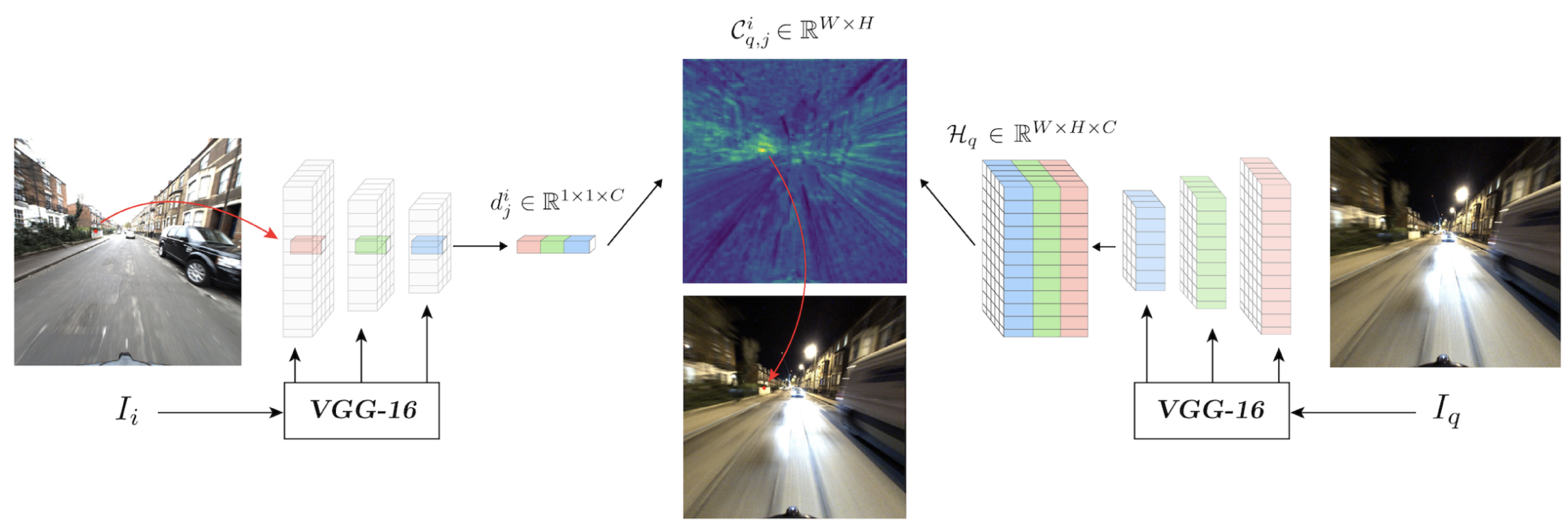


- The query correspondent should be searched *exhaustively* in the query:



Here make use of appropriate feature descriptors, coming from the retrieval network.

Sparse-to-Dense Hypercolumn Matching **(Ours)**



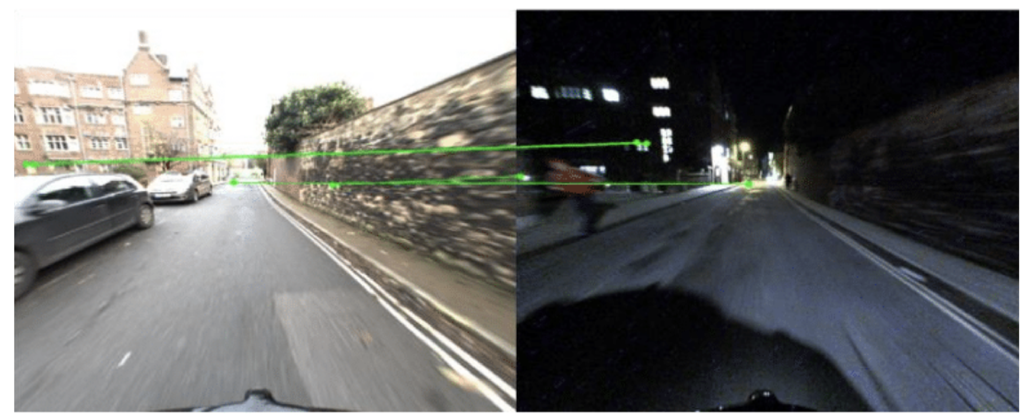
https://github.com/germain-hug/S2DHM

Ablation Study

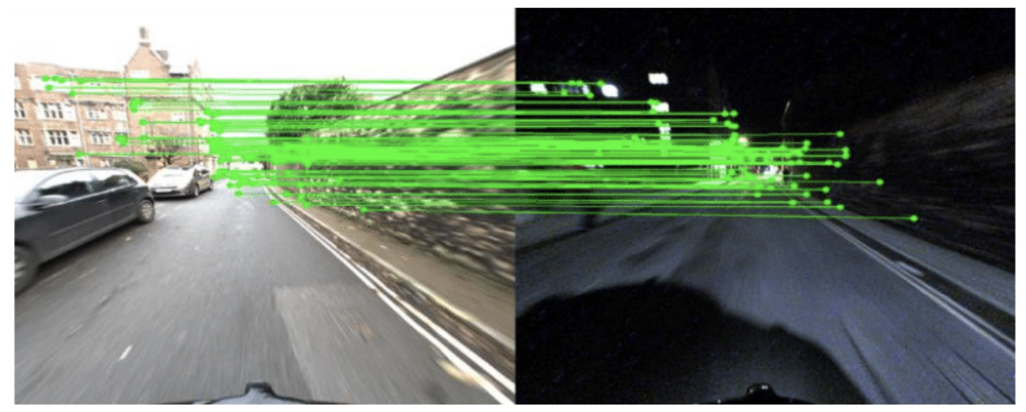
Method	*Day-All* Threshold Accuracy 0.25m 2°	0.5m 5°	5m 10°	*Night-All* Threshold Accuracy 0.25m 2°	0.5m 5°	5m 10°
NV (pre-trained)	6.4	26.3	90.9	0.3	2.3	15.9
NV-r (re-trained)	4.1	17.8	86.9	2.4	11.4	84.6
NV-r + S-S + SP	**52.9**	**78.5**	93.8	10.9	32.7	87.4
NV-r + S-S + H	49.0	77.9	93.6	14.8	44.5	89.7
NV-r + S-D + SP	50.3	77.5	92.9	14.4	43.2	87.8
NV-r + S-D + H	45.7	78.0	**95.1**	**22.3**	**61.8**	**94.5**

Ablation study run on RobotCar Seasons nighttime images

Qualitative Results

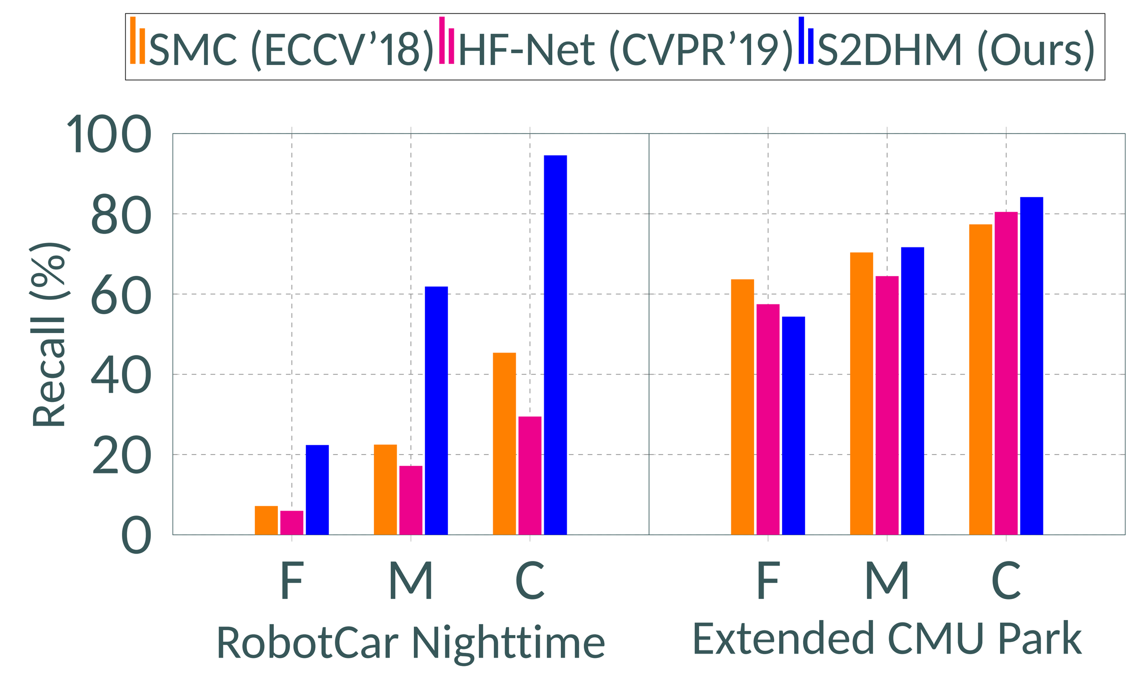


Sparse-to-Sparse Superpoint matching [4 inliers]



Sparse-to-Dense Hypercolumn matching [87 inliers]

Quantitative Results



Localization recall for Fine (F), Medium (M) and Coarse(C) thresholds.