Appendix

In the following, we provide additional qualitative and quantitative results about the Local Invariance Selection at Runtime for Descriptors (LISRD). Section A gives details about the benchmark dataset that was created to evaluate the impact of rotation and illumination invariance. Additional evaluations on multiple kinds of keypoint are available in Section B. Section C shows the performance of LISRD with respect to the state of the art for varying time intervals between the matched images. We also provide in Section D an extended evaluation on the Aachen Day-Night dataset on the day split in addition to the night split. Finally, Section E displays qualitative matches and selected invariances for challenging scenarios.

Additionally, we provide a video (`demo_lisrd_sift.mp4`) showing the online invariance selection for LISRD-SIFT, the generalization of our method to SIFT and Upright SIFT selection, on a scene with both rotated and non rotated objects.

A A benchmark dataset for illumination and rotation invariances

The Day-Night Image Matching (DNIM) [7] dataset was originally released to evaluate the impact of day-night changes on local features matching. It consists of 1722 images grouped in 17 sequences of a fixed webcam taking pictures at regular time spans over 48h. In order to obtain pairs of images to match, a day and a night references are chosen for each sequence: the image with timestamp closest to noon is selected as day reference and similarly for the timestamp closest to midnight for the night reference. We then pair all the images in a sequence both with the corresponding day and night references, thus resulting in two benchmark datasets of 1722 pairs of images each. One dataset matches day references to all the DNIM images and is composed of day-day and day-night pairs, while the other dataset matches the night references to the DNIM images and displays night-night and night-day pairs. To simultaneously evaluate the robustness of our method to rotation and its discriminative power for non rotated images, we also warp the second image of each pair (i.e. the non reference image of the pair) with homographies. Similarly as in [2], these homographies are generated by combining random translations, rotations, scalings, and perspective distortions, with an equal distribution of rotated and non rotated images. Thus, we call this augmented dataset in the following RDNIM, for *Rotated* DNIM. Examples of the RDNIM image pairs are available in Figure 1.

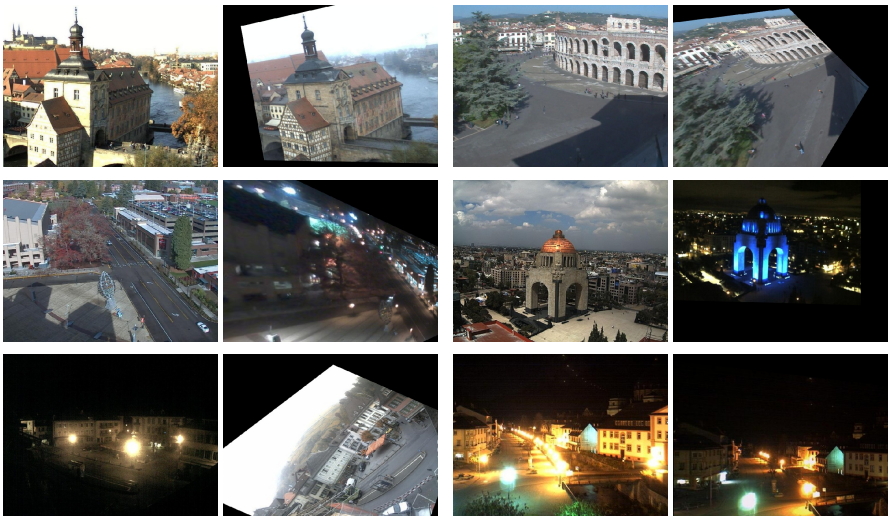


Fig. 1: **Sample images of the DNIM [7] dataset augmented with rotations.** All combinations among day-day, day-night and night-night pairs are available. Homographies are generated with random translations, rotations, scalings and perspective distortions, and images with and without rotation are equally distributed. When matching the images, the black artifacts created by the homography warping are masked out and ignored.

B Generalization to different keypoint detectors

LISRD was trained using SIFT keypoints [4], but it can be used at test time with any other keypoints. We demonstrate this by providing additional comparisons to the state of the art on the RDNIM dataset, using SIFT, SuperPoint [2] and R2D2 [5] keypoints. Table 1 presents the evaluation with homography estimation, precision and recall for an error threshold of 3 pixels and Figure 2 shows the precision curves at multiple error thresholds. Overall, LISRD is competitive with the state of the art (HardNet and SOSNet) on SIFT keypoints and ranks first with learned keypoints on most metrics. Note the improved homography estimation score with learned keypoints, probably because these keypoints are well spread across the image and the limitation of LISRD mentioned in the main paper is curtailed. Indeed, this limitation was due to RANSAC producing bad estimates when the invariance selection failed in some regions and all the matches became concentrated in a small area. This phenomenon is less likely to happen when the keypoints are covering the whole image, and LISRD is thus able to get a more accurate homography estimation. Note that for each keypoint, the associated descriptor is not necessarily performing better, except for R2D2 that gets a slight improvement in precision when evaluated on their own keypoints. This is due to the reliability map used during their training, which makes their descriptors more discriminative at their keypoint locations.

Table 1: **Evaluation with SIFT [4], SuperPoint (SP) [2] and R2D2 [5] keypoints on the RDNIM dataset.** Homography estimation, precision and recall are computed for an error threshold of 3 pixels. LISRD is not restricted to the SIFT keypoints that were used during its training, but can be generalized to any keypoints (KP). The best score is in bold and the second best one is underlined.

			Root SIFT	HardNet	SOSNet	SP	D2-Net	R2D2	GIFT	LISRD (Ours)
SIFT KP	Day ref	HEstimation	0.166	0.170	**0.215**	0.084	0.057	0.121	0.145	0.127
		Precision	0.220	0.200	**0.232**	0.150	0.144	0.140	0.126	0.226
		Recall	0.113	0.155	0.197	0.114	0.081	0.107	0.108	**0.212**
	Night ref	HEstimation	0.255	0.278	**0.307**	0.156	0.118	0.167	0.215	0.204
		Precision	0.368	0.394	**0.416**	0.254	0.231	0.228	0.246	0.357
		Recall	0.212	0.288	**0.316**	0.183	0.135	0.162	0.183	0.284
SP KP	Day ref	HEstimation	0.121	**0.199**	0.178	0.146	0.094	0.170	0.187	0.198
		Precision	0.188	0.232	0.228	0.195	0.195	0.175	0.152	**0.291**
		Recall	0.112	0.194	0.203	0.178	0.117	0.162	0.133	**0.317**
	Night ref	HEstimation	0.141	**0.262**	0.211	0.182	0.145	0.196	0.241	**0.262**
		Precision	0.238	0.366	0.297	0.264	0.259	0.237	0.236	**0.371**
		Recall	0.164	0.323	0.269	0.255	0.182	0.216	0.209	**0.384**
R2D2 KP	Day ref	HEstimation	0.107	0.187	0.181	0.140	0.093	0.135	0.157	**0.193**
		Precision	0.162	0.201	0.192	0.166	0.171	0.210	0.118	**0.237**
		Recall	0.093	0.167	0.172	0.168	0.101	0.076	0.102	**0.290**
	Night ref	HEstimation	0.135	**0.196**	0.168	0.145	0.101	0.132	0.183	0.189
		Precision	0.200	**0.302**	0.244	0.221	0.221	0.241	0.166	0.291
		Recall	0.132	0.260	0.215	0.230	0.149	0.110	0.147	**0.335**

As a feature direction of work, LISRD would benefit from learning its own keypoints with an additional head. This single head would predict invariant keypoints trained on images with multiple lightings and rotations and could be used with all descriptors - whether they are variant or not. This would ensure a better correlation between the keypoints and their descriptors and offer a faster prediction, instead of predicting separately keypoints and descriptors as is currently the case.

C Evaluation across a full day

The evaluation on the RDNIM dataset shows the global performance across a mix of day-day and day-night, or night-night and night-day images. But it is also interesting to study the performance at various times during the day. Figure 3 displays the precision and recall curves on the RDNIM dataset along a full day. For every image in the second pair, we extract the hour at which the picture was taken from the timestamp, and round it to the closest integer. For each hour, the precision and recall are then computed and averaged across all images corresponding to this time and these averaged numbers are then plotted over the twenty-four hours of a day. We naturally get two peak curves, one centered at noon for the pairs with the day reference and the other centered at midnight

Table 2: **Visual localization benchmark on the Aachen Day-Night dataset [6].** We report the percentage of correctly localized queries on both day and night query images for various distance and orientation error thresholds for SIFT, SuperPoint and D2-Net single scale (SS). LISRD leverages illumination variance on the day part and light invariance when querying night images and is thus able to improve the performance of various descriptors on multiple keypoints.

	Error threshold	SIFT KP		SuperPoint KP		D2-Net KP	
		Upright Root SIFT	LISRD	SuperPoint	LISRD	D2-Net (SS)	LISRD (SS)
Day	$0.5m, 2^\circ$	80.8	**82.4**	80.5	**85.6**	77.8	**81.7**
	$1m, 5^\circ$	89.0	**89.9**	88.3	**91.3**	**89.6**	**89.6**
	$5m, 10^\circ$	93.2	**94.2**	93.2	**95.6**	**95.6**	94.4
Night	$0.5m, 2^\circ$	51.0	**68.4**	65.3	**78.6**	**78.6**	71.4
	$1m, 5^\circ$	61.2	**79.6**	73.5	**87.8**	**87.8**	**87.8**
	$5m, 10^\circ$	69.4	**91.8**	86.7	**95.9**	**95.9**	94.9

for the night reference. LISRD is overall better than the other descriptors and, interestingly, the largest improvements come from the time intervals with day-night illumination changes. Thus, LISRD leverages its illumination variant and more discriminative descriptors when the timestamp of both images of the pair are close, and it switches to the invariant and more general ones when the images are taken at different times of the day.

D Evaluation on Aachen Day-Night dataset

The local features benchmark on the Aachen dataset [6] is restricted to the night part of the dataset. In order to have a more extensive evaluation, we also report the results for the visual localization benchmark on both day and night splits of the dataset. For each query image, the 20 best candidate images are retrieved from the database using NetVlad [1] and the official pipeline of the benchmark[1] is then used to estimate the pose of the query images.

Table 2 shows that LISRD is able to improve over several state-of-the-art descriptors and can generalize to different keypoints. It can indeed leverage variant descriptors for day-day image pairs of the day part and the invariant descriptors on the night part for day-night image pairs.

As the ground truth poses of the Aachen Day-Night dataset have been updated between submission and acceptance of this paper, we additionally show in Table 3 the results of the local features benchmark on the night part of the Aachen dataset with the old numbers for legacy and easier comparison with previous methods.

[1] `https://github.com/tsattler/visuallocalizationbenchmark`

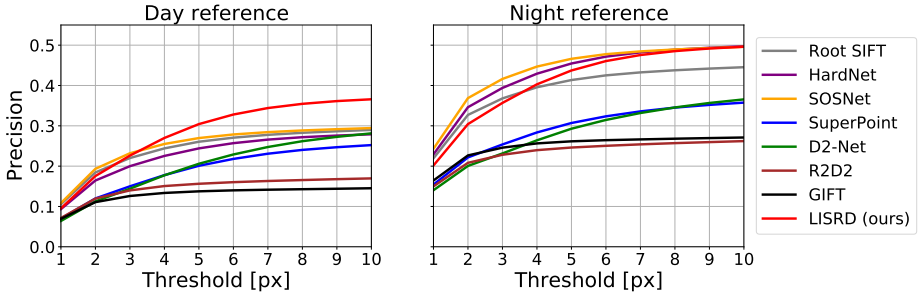
Table 3: **Legacy results of the local features benchmark on the Aachen Day-Night dataset [6].** We report the percentage of correctly localized queries for various distance and orientation error thresholds for SIFT, SuperPoint and D2-Net multi-scale (MS). Our method shows a good generalization when evaluated on different keypoints (KP) and can improve the original descriptor performance.

Error threshold	SIFT KP		SuperPoint KP		D2-Net KP	
	Upright Root SIFT	LISRD	SuperPoint	LISRD	D2-Net (MS)	LISRD (MS)
$0.5m, 2^\circ$	33.7	**43.9**	42.9	**44.9**	44.9	**45.9**
$1m, 5^\circ$	52.0	**62.2**	57.1	**65.3**	64.3	**66.3**
$5m, 10^\circ$	65.3	**82.7**	77.6	**84.7**	**88.8**	87.8

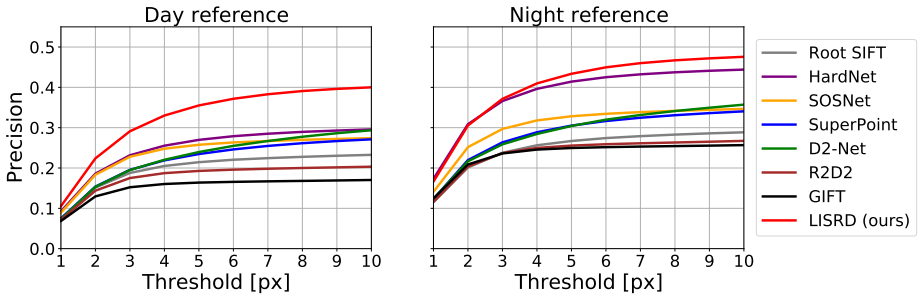
E Qualitative examples

We provide additional qualitative examples of matches based on SIFT keypoints and LISRD descriptors. All matches are filtered with mutual nearest neighbor, followed by a homography fitting with RANSAC [3]. Figure 4 brings a visualization of the invariance selection, with a different color for each kind of invariance that was selected. Since the selection is in practice based on a soft weighting, we only show the color of the learned descriptors that contributed the most in the matching decision. These sample images show that in some situations, a single invariance is sufficient for the full image, but in other cases multiple invariances can be leveraged within the same image, demonstrating the need of tiled meta descriptors. This is for example useful when the overall illumination is constant in a pair of images, but one part an image (e.g. a building) is overexposed or in the shadow.

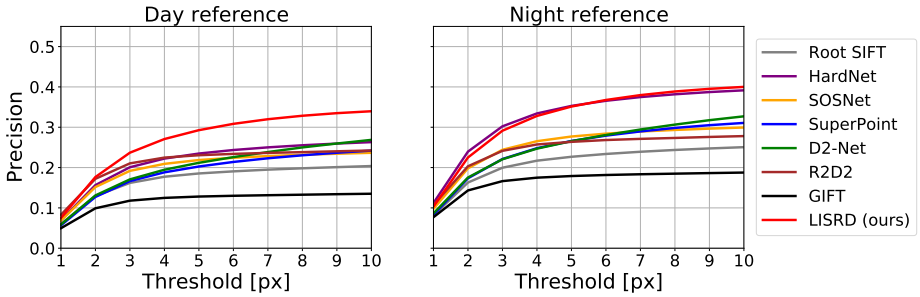
Finally, Figure 5 displays a selection of matches in challenging scenarios, for example with day-night and/or with strongly rotated images.



(a) SIFT keypoints.



(b) SuperPoint keypoints.



(c) R2D2 keypoints.

Fig. 2: **Precision curves with SIFT [4], SuperPoint [2] and R2D2 [5] keypoints on the RDNIM dataset.** The discriminative power of LISRD descriptors is not limited to SIFT keypoint locations, but also shows a high precision compared to the state of the art on other keypoints.

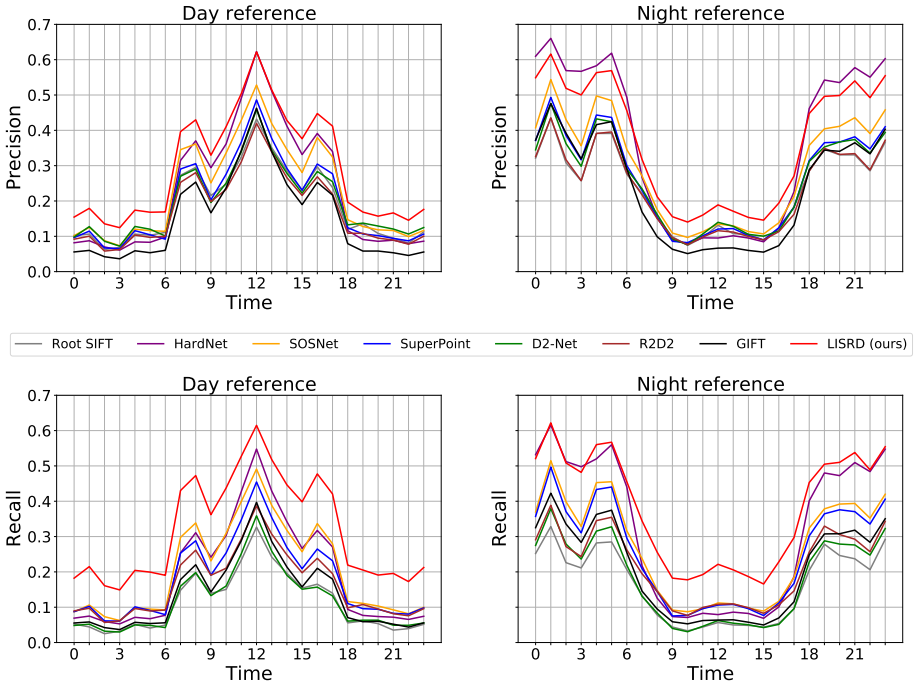


Fig. 3: **Precision and recall across the day.** Precision and recall are computed on the RDNIM dataset with SuperPoint keypoints and an error threshold of 3 pixels. They are averaged for each hour of the day, based on the timestamp of the second image. The performance gradually degrades when the timestamp of the second image moves away from the reference time (noon for the day reference and midnight for the night reference). For close timestamps, LISRD leverages its illumination variant descriptors, but switches to the invariant ones when the timestamps differ too much. Thus, LISRD remains competitive with state-of-the-art descriptors for close timestamps and outperforms them when significant illumination changes are present.

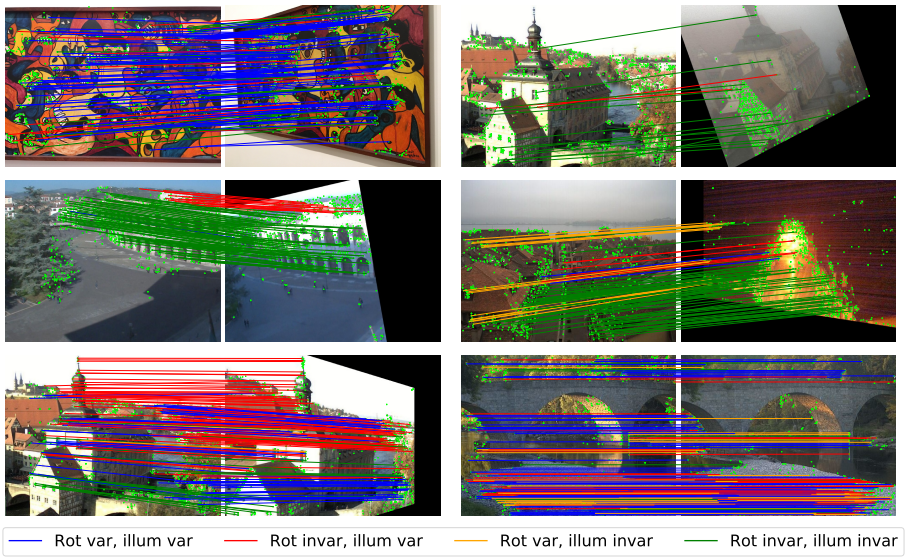


Fig. 4: **Visualization of the selected invariance.** Matches of SIFT keypoints with LISRD descriptors are filtered with mutual nearest neighbor and RANSAC [3]. Since our method uses a soft weighting of the invariances, each color corresponds only to the invariance that contributed the most to validate the match. First line: one type of invariance predominates in the whole image. Second line: two invariances are relevant in the same image (on the left, rotation invariance is needed, but the building in the top right corner is overexposed in both images and illumination invariance is not needed in this area ; on the right, illumination invariance is needed, but the image is upright on the left side, while the distortion creates a rotation on the central part and rotation invariance becomes necessary). Third line: multiple different invariances can be leveraged in the same image (on the left image, the right part of the image is mainly upright and with constant illumination, while the house in the lower left corner is overexposed and rotated, hence the fully invariant descriptor is selected ; on the right image, most of the selected descriptors are rotation variant as the viewpoint is fixed, but the left pier of the bridge has a constant illumination while the right pier has a different illumination and the illumination invariant descriptor predominates).

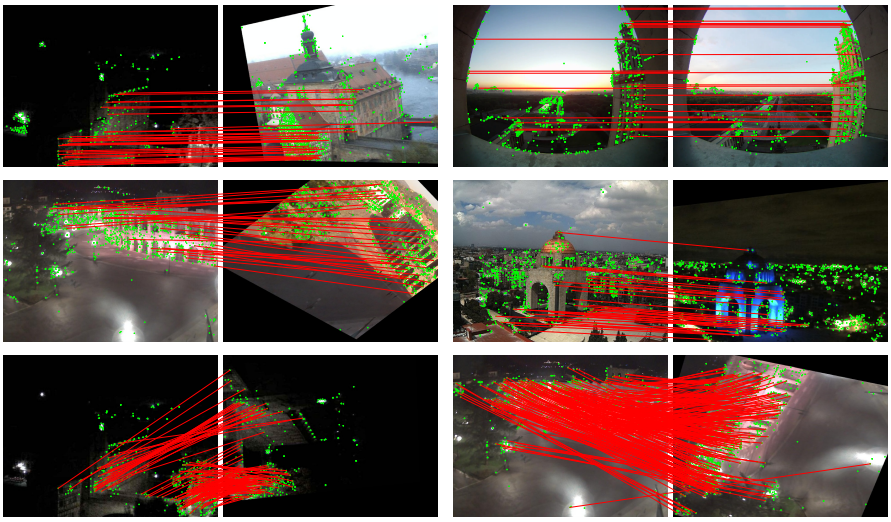


Fig. 5: **Matches in challenging situations.** SIFT keypoints are detected, matched with the LISRD descriptors, and mutual nearest neighbor and RANSAC [3] are used to filter out wrong matches. A single red color is used for all the inlier matches, regardless of the chosen invariance. Matches based on LISRD descriptors are able to handle strong illumination changes such as day-night, inter-image illumination variations in day-day and night-night pairs, and small as well as strong rotations.

References

1. Arandjelović, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J.: NetVLAD: CNN architecture for weakly supervised place recognition. In: Computer Vision and Pattern Recognition (CVPR) (2016)
2. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: Computer Vision and Pattern Recognition Workshops (CVPRW) (2018)
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the Association for Computing Machinery (ACM) **24** (1981)
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (IJCV) **60** (2004)
5. Revaud, J., Weinzaepfel, P., de Souza, C.R., Humenberger, M.: R2D2: repeatable and reliable detector and descriptor. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
6. Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., Kahl, F., Pajdla, T.: Benchmarking 6dof outdoor visual localization in changing conditions. In: Computer Vision and Pattern Recognition (CVPR) (2018)
7. Zhou, H., Sattler, T., Jacobs, D.W.: Evaluating local features for day-night matching. In: European Conference on Computer Vision Workshops (ECCVW) (2016)