Features for Ground Texture Based Localization - A Survey

Supplementary Material

Jan Fabian Schmid[1,2]
SchmidJanFabian@gmail.com
Stephan F. Simon[1]
Stephan.Simon@de.bosch.com
Rudolf Mester[2]
mester@vsi.cs.uni-frankfurt.de

[1] Robert Bosch GmbH
Hildesheim, Germany

[2] VSI Lab
Goethe University
Frankfurt, Germany

1 Parameter optimization

In order to find the best parameter settings of the evaluated methods, we extract features from synthetically transformed images using varying parameter settings, and evaluate the obtained results based on a few selected key performance metrics.

A set of 3 non-overlapping images have been selected for each of the 6 examined texture types (fine asphalt, coarse asphalt, carpet, concrete, tiles, and wood). This results in a number of 18 evaluated images, which are selected from a training set that is not included during testing. Each image is synthetically transformed as for testing (overall 72 synthetic transformations are tested), which results in 1296 pairs of reference and transformed images.

For keypoint detector parameter optimization, we introduce another performance metric. *Adjusted repeatability* is a derived metric that combines the conventional repeatability metric of Mikolajczyk *et al*. [17] and our ambiguity score:

$$\text{adjusted repeatability} = \frac{\text{repeatability}}{\text{ambiguity}}. \tag{1}$$

The advantage of adjusted repeatability over conventional repeatability is that it does not reward clustering of keypoints. The following example illustrates the advantage of adjusted repeatability for parameter optimization. Repeatability increases if keypoints are assigned large associated regions, because it is more likely that a pair of keypoint objects from the reference image and the test image have an IoU greater 0.5. In the extreme case, if each keypoint region is as large as the image, repeatability takes the optimal value of 1.0. This behavior of the repeatability metric is misleading during parameter optimization. Ambiguity, on the other hand, increases as well if keypoint regions are increased. Therefore, adjusted repeatability takes a very small value if each keypoint object is as large as the whole image, which makes it a more suited performance metric for detector parameter optimization.

© 2019. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

Keypoint detector	Parameters
CenSurE [1]	Implemented in OpenCV as **StarDetector**. `maxSize`=11, `responseThreshold`=0, `lineThresholdProjected`=27, `lineThresholdBinarized`=24, `suppressNonmaxSize`=4
FAST [12]	`threshold`=5, `nonmaxSuppression`=true, `type`=TYPE_9_16
AGAST [10]	`threshold`=5, `nonmaxSuppression`=false, `type`=TYPE_7_12s
MSER [11]	`_delta`=0, `_min_area`=160, `_max_area`=14400, `_max_variation`=0.02
MSD [13]	`m_patch_radius`=3, `m_search_area_radius`=3, `m_nms_radius`=5, `m_nms_scale_radius`=0, `m_th_saliency`=30, `m_kNN`=50, `m_scale_factor`=4.5, `m_n_scales`=-1, `m_compute_orientation`=false
HarrisLaplace	`numOctaves`=6, `corn_thresh`=0.0008, `DOG_thresh`=0.001, `num_layers`=2
GFTT [16]	`qualityLevel`=0.01, `minDistance`=5, `blockSize`=5, `useHarrisDetector`=true, `k`=0.01

Table 1: Evaluated parameters for keypoint detection methods.

Feature descriptor	Parameters
DAISY [17]	`radius`=5, `q_radius`=3, `q_theta`=8, `q_hist`=10, `norm`=NRM_NONE, `interpolation`=true, `use_orientation`=true
BRIEF [3]	`bytes`=32, `use_orientation`=true
FREAK [2]	`orientationNormalized`=true, `scaleNormalized`=false, `patternScale`=17, `nOctaves`=2
LATCH [8]	`bytes`=64, `rotationInvariance`=true, `half_ssd_size`=6, `sigma`=3.6

Table 2: Evaluated parameters for feature description methods.

We decide to optimize keypoint detector parameters for: 1. $<$ 100 KPs, 2. adjusted repeatability, 3. detection time. In order to obtain the best parameter settings, we adapt the parameters in such a way that we reach optimal performance for each detector. Table 1 presents the finally used parameters of the methods taken from OpenCV [5] and the evaluated ORB implementation from ORB-SLAM2 [14]. If methods allowed to specify the number of keypoints to retrieve, we put this value to 1000. Otherwise, non-maximum suppression was used to reduce the number of keypoints to 1000.

For optimization of the parameters of feature descriptor methods, we chose SURF to provide the keypoints (besides for AKAZE, where it had to be the AKAZE detector) and evaluate pose estimation performance on synthetically transformed images. The parameters shown in Table 2 have been optimized for 1. pose estimation success rate, 2. matching precision, 3. number of correct feature matches.

Parameters for feature extractors that perform keypoint detection as well as feature description are presented in Table 3. We optimized these methods first for keypoint detection and then for feature description, using the same optimization strategies previously described.

Feature extractor	Parameters
SIFT [9]	`nOctaveLayers`=12, `contrastThreshold`=0.003, `edgeThreshold`=9, `sigma`=8.7
SURF [4]	`hessianThreshold`=20, `nOctaves`=1, `nOctaveLayers`=2, `extended`=false, `upright`=false
ORB [15, 18]	`scaleFactor`=1.0, `nlevels`=1, `iniThFAST`=29, `minThFAST`=3
BRISK [7]	`thresh`=6, `octaves`=1, `patternScale`=1.45
AKAZE [6]	`descriptor_type`=DESCRIPTOR_MLDB, `descriptor_size`=486, `descriptor_channels`=3, `threshold`=0.0001, `nOctaves`=2, `nOctaveLayers`=2, `diffusivity`=DIFF_CHARBONNIER

Table 3: Evaluated parameters for feature extraction methods that combine keypoint detection and feature description.

2 Detailed results

We present more detailed results of the conducted experiments. First, texture dependent and transformation dependent repeatability scores evaluated on synthetic transformations are summarized. Then, we examine texture dependent pose estimation success rates on pairs of separately taken images as well as transformation dependent pose estimation success rate on synthetically transformed images.

2.1 Texture dependent repeatability

Figure 1 shows repeatability scores for varying types of texture. SIFT presents itself as the most robust keypoint detector for different texture types as it has the lowest amount of texture dependent performance variance. We identify wood to be the most difficult type of texture. All detectors achieve lowest repeatability on images of the wooden floor. AKAZE, SURF and CenSurE, the three detectors we identified as the best performing ones, suffer from low performance on wood, as well. While they achieve about 75% to 90% repeatability on asphalt, carpet, concrete, and tiles, they only have about 60% to 65% repeatability on wood.

2.2 Transformation dependent repeatability

Individual repeatability scores for the four different types of synthetic transformations are presented in Figure 2. Most keypoint detectors perform well on translated images and worse on images with synthetic noise. SIFT, on the other hand, is the worst performing detector for synthetic translations, but the best performing one for noise. Again, SIFT has the lowest amount of performance variance. It becomes clear that differences in overall repeatability are mostly determined by the detector performance on noisy images, as the differences on other synthetic transformations are less severe.

Figure 3 (left) presents repeatability scores for different degrees of synthetic rotation. SIFT exhibits dominant performance in regard to synthetic rotations, its repeatability does not depend on the rotation angle. The other detectors, due to their architectures, show certain performance patterns. SURF, for example, approximates the difference of Gaussian pyramid

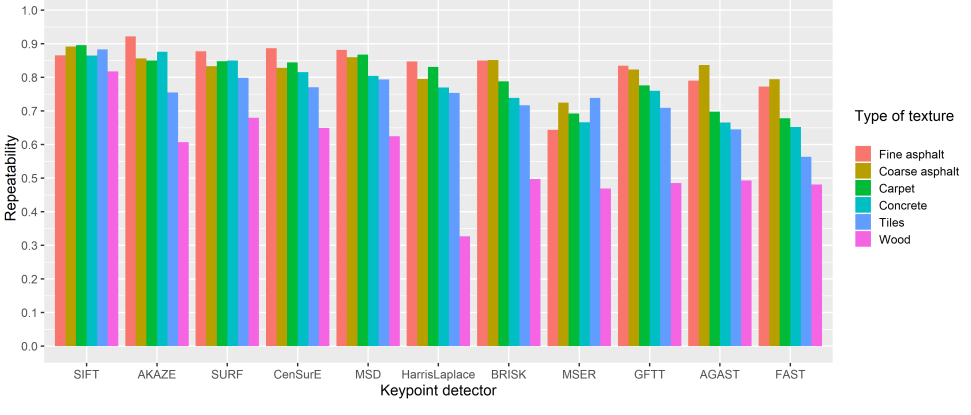

Figure 1: Repeatability for different types of texture.

of SIFT using difference of boxes. The response of box filters, however, only is invariant to rotation for rotations in 90 degree steps.

Figure 3 (right), illustrates repeatability for different amounts of overlap of reference and test image after synthetic translation. SIFT's weakness in regard to synthetic translation stands out. We examine SIFT's behavior for synthetic transformation in Figure 4. The repeatability for the presented image pair is 42.9%. We expect to extract the same keypoint objects (blue circles) from the intersecting area (green box) of reference and test mask (red boxes). However, while most keypoint objects extracted from this intersection in the second image also occur in the intersection in the first image, there are many additional keypoint objects in the first image that are not extracted in the second image. This behavior decreases repeatability of SIFT and might be the reason why on real image sequences most descriptors have lower pose estimation success rates using SIFT keypoints compared to CenSurE and SURF, even though we found SIFT to be an otherwise robust keypoint detector.

Repeatability for increasing levels of Gaussian noise is depicted on the left of Figure 5. For increasing amounts of noise detector performance has an exponential decay. We notice

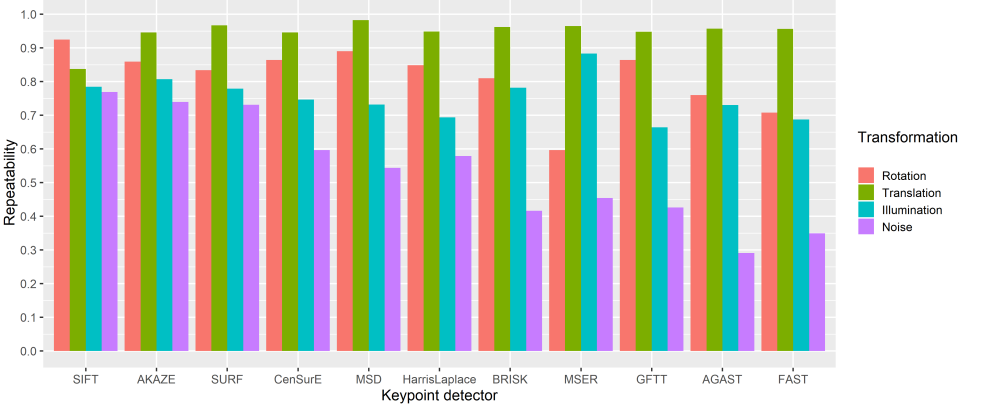

Figure 2: Repeatability for different types of synthetic transformation.

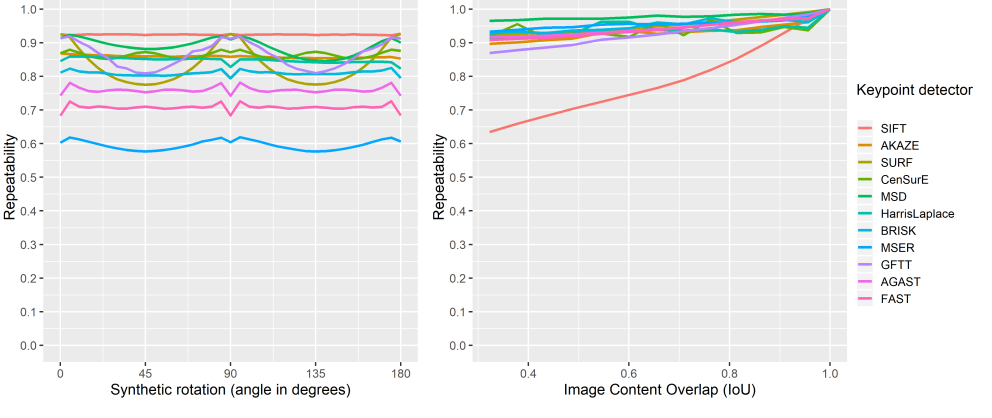

Figure 3: Repeatability for rotated **(left)** and translated images **(right)**.

that SIFT, AKAZE, and SURF are significantly more robust to noise than other detectors.

Results for synthetic illumination changes are presented on the right of Figure 5. MSER is most robust to illumination changes as it identifies keypoint objects as image regions with similar brightness. Since gamma correction is applied to each pixel, these regions are likely to still have similar brightness after the transformation.

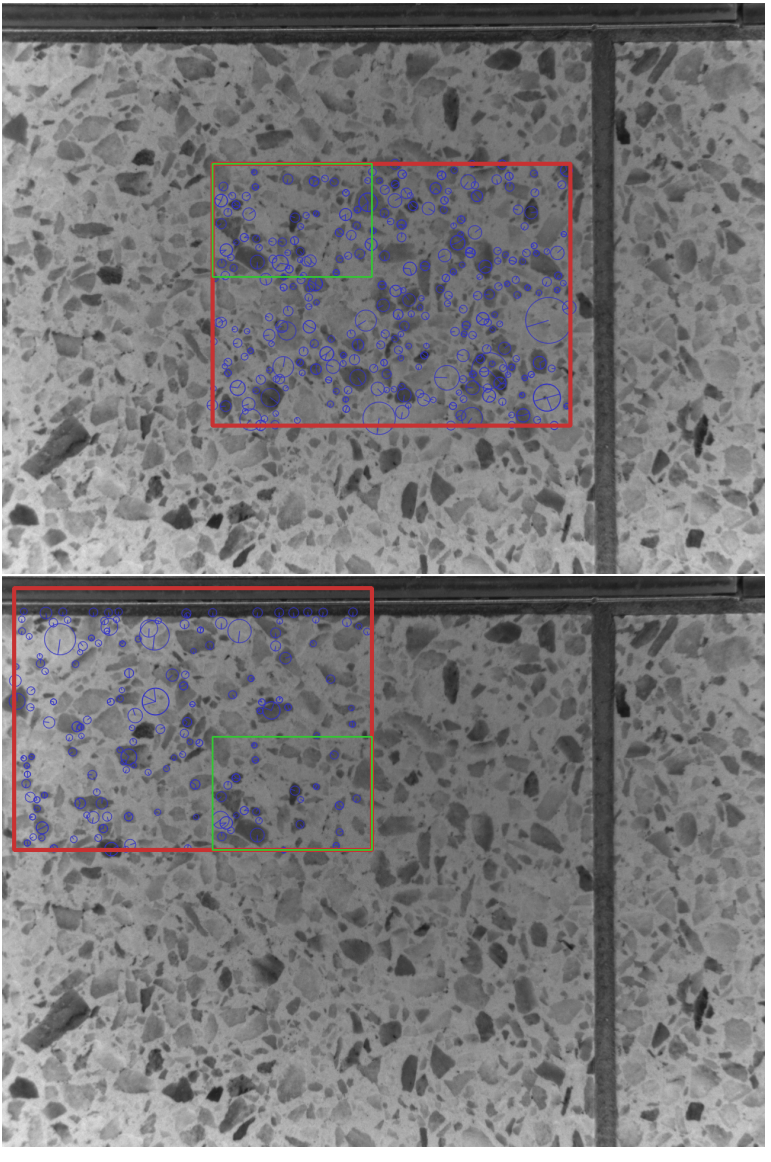

Figure 4: Visualization of SIFT keypoint objects for synthetic translation. SIFT keypoint objects are depicted as blue circles, the image sections from which keypoints are extracted are bounded by red rectangles, and the intersection of reference image section and translated image section is depicted as green rectangle (IoU of 0.1925).

2.3 Texture dependent pose estimation success rate (separately taken images)

For the best performing detector-descriptor pairings, Figure 6 presents pose estimation success rates for the different texture types on image pairs that have been taken in direct sequence, as it is done for incremental localization tasks. Surprisingly to us, all detector-descriptor pairings enable almost perfect pose estimation on fine asphalt and carpet. These textures are particularly well suited for ground texture based localization, they reveal structure that can be exploited by any descriptor method to extract distinctive feature descriptors. Wood texture poses significant difficulties on some detector-descriptors pairings, like SIFT-SIFT and also for SURF-SURF, this is where CenSurE-ORB, CenSurE-BRIEF, and CenSurE-LATCH have an advantage.

For image pairs with larger overlap, but more severe rotation and photometric transformations, as they occur for absolute localization, Figure 7 presents the corresponding pose estimation success rates for the best performing pairings on this task. Again the most chal-

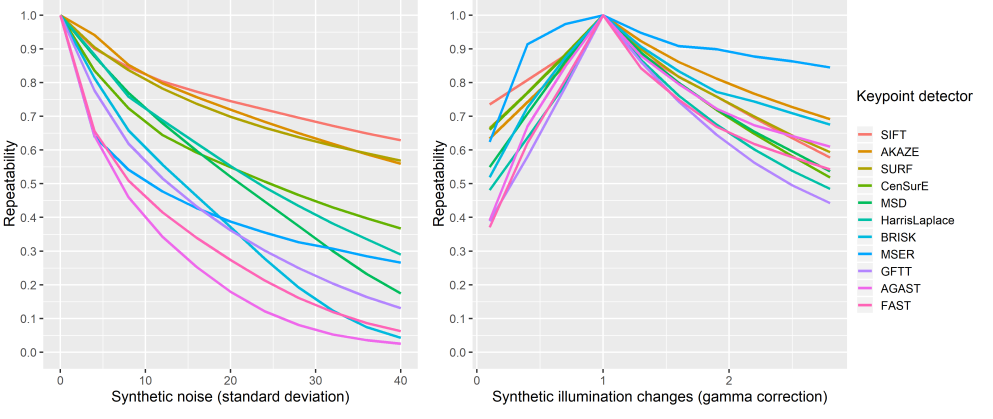

Figure 5: Repeatability for images with added noise **(left)** and changed illumination **(right)**.

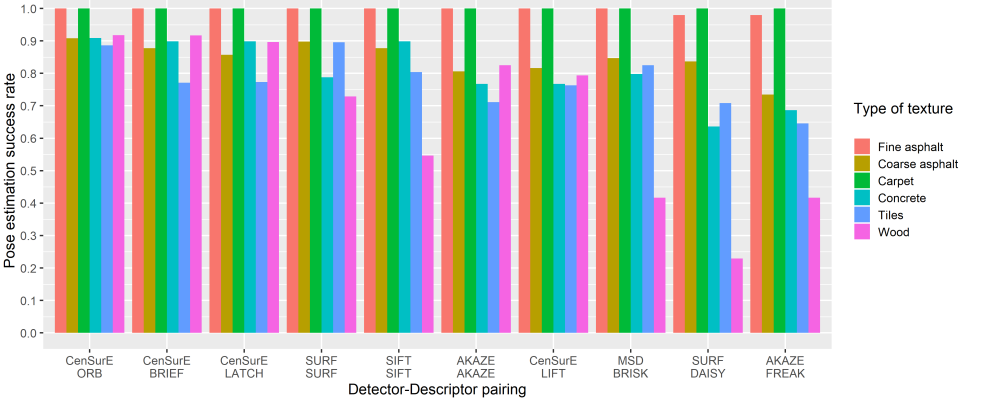

Figure 6: Texture dependent pose estimation success rate for incremental localization tasks.

lenging texture is wood, while transformations between the image pairs of the other textures can be estimated with great success by most of the presented detector-descriptor pairings.

2.4 Transformation dependent pose estimation success rate (synthetic transformations)

Figure 8 presents (synthetic) transformation dependent localization success rates. All of the presented pairings can deal well with synthetic rotations and translations, which we confirm the detailed results in Figure 9. However, as shown in Figure 10, some of the presented pairings are outperformed for photometric transformations. Particularly, the pairing of DAISY descriptors with AKAZE keypoint objects has difficulties with more severe photometric transformations.

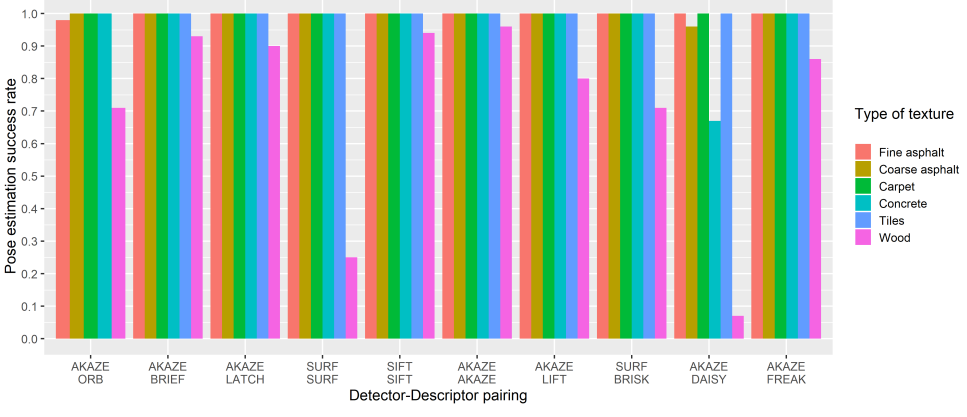

Figure 7: Texture dependent pose estimation success rate for absolute localization tasks.

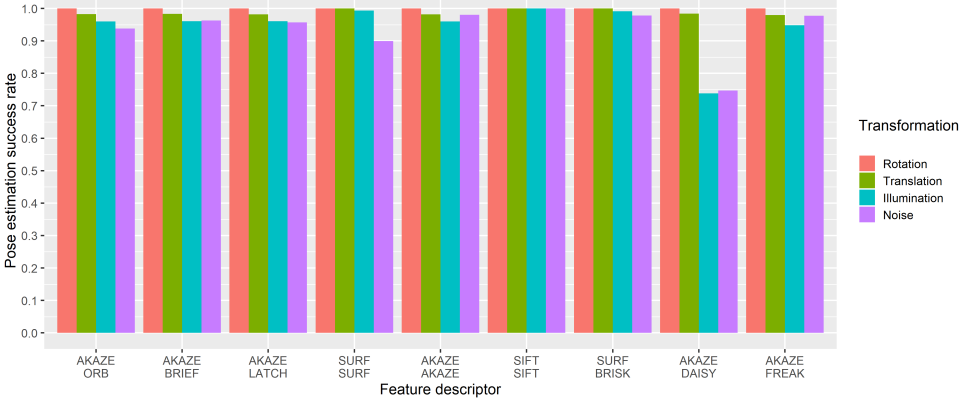

Figure 8: Pose estimation success rate for different types of synthetic transformation.

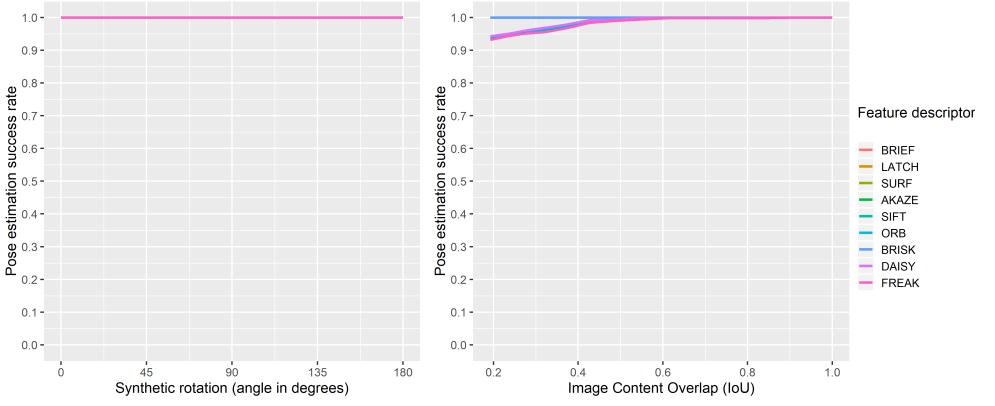

Figure 9: Pose estimation success rate for rotated **(left)** and translated images **(right)**.

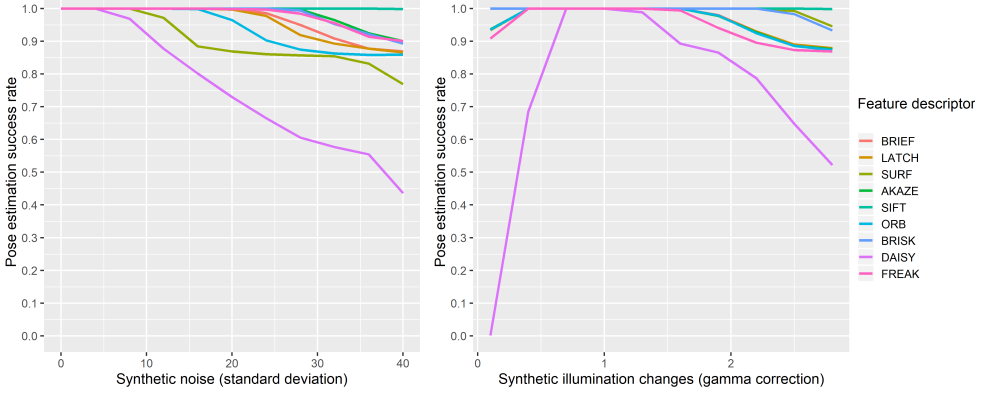

Figure 10: Pose estimation success rate for rotated **(left)** and translated images **(right)**.

References

[1] M. Agrawal, K. Konolige, and M. R. Blas. CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching. In *IEEE European Conference on Computer Vision (ECCV)*, pages 102–115, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[2] A. Alahi, R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 510–517, June 2012.

[3] P. F. Alcantarilla, J. Nuevo, and A. Bartoli. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2013.

[4] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded Up Robust Features. In *IEEE European Conference on Computer Vision (ECCV)*, pages 404–417, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[5] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[6] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary Robust Independent Elementary Features. In *IEEE European Conference on Computer Vision (ECCV)*, pages 778–792, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[7] S. Leutenegger, M. Chli, and R. Y. Siegwart. BRISK: Binary Robust invariant scalable keypoints. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2548–2555, Nov 2011.

[8] G. Levi and T. Hassner. LATCH: Learned arrangements of three patch codes. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, 2016.

[9] David G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, Nov 2004.

[10] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger. Adaptive and Generic Corner Detection Based on the Accelerated Segment Test. In *IEEE European Conference on Computer Vision (ECCV)*, pages 183–196, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761 – 767, 2004.

[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A Comparison of Affine Region Detectors. *International Journal of Computer Vision (IJCV)*, 65(1):43–72, Nov 2005.

[13] R. Mur-Artal and J. D. Tardós. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, Oct 2017.

[14] E. Rosten and T. Drummond. Machine Learning for High-Speed Corner Detection. In *IEEE European Conference on Computer Vision (ECCV)*, pages 430–443, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[15] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2564–2571, Nov 2011.

[16] J. Shi and C. Tomasi. Good Features to Track. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 593 – 600, January 1994.

[17] E. Tola, V. Lepetit, and P. Fua. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 32(5):815–830, May 2010.

[18] F. Tombari and L. Di Stefano. Interest Points via Maximal Self-Dissimilarities. In *Asian Conference on Computer Vision (ACCV)*, pages 586–600, Cham, 2014. Springer International Publishing.