Image Features for Long-Term Mobile Robot Autonomy

Tomáš Krajník[1] Pablo Cristóforis[2] Jan Faigl[1] Hana Szücsová[1] Matías Nitsche[2] Libor Přeučil[1] Marta Mejail[2]

Abstract— **In this paper, we present an evaluation of standard feature descriptors in a scenario of outdoor vision-based long-term autonomous navigation. Although benchmarks of visual features can be found in computer vision literature, we propose an evaluation focused on navigational aspects, especially to achieve a long-term autonomy under seasonal changes. The considered dataset captures seasonal changes along a path used for an autonomous navigation in a park like environment over a whole year. The visual features are used to estimate the robot orientation and its correct estimation is considered as the primary measure of feature descriptor suitability for long-term autonomous navigation. The achieved results indicate that a suitability of the descriptor for the considered type of navigation does not necessary correlate with its robustness and invariance regarding regular quality metrics used in computer vision.**

Index Terms— **visual navigation, mobile robotics, long-term autonomy**

I. Introduction

Cameras are becoming a de-facto standard part of sensoric equipment of mobile robotic systems including field robots. It is due to improving ratio of their cost and information that can be extracted from the captured images. Moreover, and probably more importantly, the computational requirements needed by computer vision techniques are not a significant issue due to available computational hardware nowadays. Therefore, on-board cameras of mobile robots are used as the primary sensors to gather information about the robot's surroundings to establish the robot position.

Many visual localization and mapping methods in mobile robotics rely on the so-called local feature extractors [1]. In these methods, a feature extractor is used to decompose a captured image into regions that are then repeatably detected and matched despite of a particular viewpoint or illumination variations. Regarding a quality of feature extractors, a key paper of Mikojaczyk [2] introduced a methodology for evaluation of extractor invariance to image scale, rotation, exposure and camera viewpoint changes. However, the local feature extraction techniques are intended to be used in a broad area of problems and some of their invariant properties are not so important for purposes of mobile robot navigation [3].

Let us consider a mobile robot navigating along a known (previously mapped) path in an outdoor environment. In this case, it is not necessary to use a feature extractor highly invariant to large viewpoint changes since the robot keeps itself close to the intended path. In addition, the rotational invariance is also not crucial for the navigation, because one can assume that the robot moves on a locally planar ground. On the other hand, the map provided to the robot might be obsolete, because the environment appearance changes over time [4]. The map decay is caused mainly by illumination variations, current weather conditions and long term environment changes caused by seasonal factors.

These considerations about the visual features in the navigation task motivate us to analyze available feature extraction algorithms and their long-term performance in the autonomous navigation based on pre-learned map, e.g., used in [5], [6], [7]. The intention of this paper is to present our proposed evaluation methodology and achieved results using six feature extraction algorithms freely available as a part of open source implementations.

[1]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 166 27 Prague, Czech Republic {krajnt1,szucs1,faigl,preucil}@fel.cvut.cz

[2]Laboratory of Robotics and Embedded Systems, Faculty of Exact and Natural Sciences, University of Buenos Aires {pdecris,mnitsche,marta}@dc.uba.ar

II. Local image feature extractors

An image feature extractor consists of detection and description phases. The feature detector serves to locate salient areas of the image while the feature descriptor captures information about the local neighbourhood of the detected area. Six local image feature extractors have been evaluated.

SIFT – Scale Invariant Feature Transform [8] – an established feature detector with high precision and good robustness, which is known to be computationally demanding.

SURF – Speeded Up Robust Features [9] – similar to SIFT, but it is computationally less demanding due to approximations, which allow to use a so-called integral image.

STAR – combines the SURF descriptor with modified Center Surround Extremas [10] detector, which is fast and precise. Exploits the advantages of the integral image as well.

BRIEF – Binary Robust Independent Elementary Features [11], which describe image area by a number of random pairwise intensity comparisons and use STAR as a detector.

BRISK – Binary Robust Invariant Scalable Keypoints [12]. Scale and rotation invariant version of BRIEF. Unlike BRIEF, it uses a deterministic comparison pattern.

ORB - Oriented FAST and Rotated BRIEF. Another attempt to achieve a scale and rotation invariant BRIEF. Uses FAST (Features from Accel. Segment Test) [13] detector.

While SIFT and SURF descriptors are floating point vectors, the BRIEF descriptor is a binary string, which reduces the computational complexity of the subsequent matching. All the aforementioned detectors and descriptors are part of the Open Source Computer Vision (OpenCV) software library (version 2.4.3), which is used for the results presented in this paper.

III. THE DATASET

The dataset covers seasonal changes of the Stromovka forest park in Prague throughout an entire year. Each month, a robot was manually driven through a given closed path, while the robot captured images by its on-board camera, see Fig. 1. Although the path started and ended at an identical location, it has been slightly altered every time. Therefore, the first image of each traversed path is taken from exactly the same location, while the locations of other pictures vary up to ± 1 m.

For the purpose of this evaluation, images from five different locations have been selected, see Fig. 2. The radial distortion of the captured images has been corrected and the bottom half of each undistorted image has been removed as it does not provide useful information for the considered autonomous navigation. The resulting dataset captures one year of environment changes caused by seasonal factors, displaced objects, and weather (a detailed description is available at [14], [15]). Thus, the dataset consists of sixty 1024×384 color images with removed radial distortion.

Six persons have compared all the images independently and established particular camera rotations between them. The results were checked for outliers (these were removed) and the averaged estimations were used as a ground truth.

The dataset has been processed by the aforementioned feature extractors with ten different threshold settings. Each of the thresholds was chosen to extract a specific average number of features per image.

IV. EVALUATION

Regarding the navigational method considered, it is not necessary to perform a full 6DOF localization to reliably traverse the given path. It is just sufficient to correct the robot's heading to keep it on the intended path [5], [6], [7]. Therefore, the proposed evaluation focuses on the ability of the feature extraction and matching algorithm to establish heading of the robot relatively to the intended path. This corresponds to the relative rotation of the camera to the moment the particular images were captured.

Two methods have been considered for determining the relative rotation of the camera. The first method closely follows a classical approach presented by Hartley's book on Multiple View Geometry. In this method, known camera parameters and correspondences between extracted and mapped features are used to calculate the essential matrix, which is factorized to obtain the robot rotation. The second method, based on articles [5], [7], establishes the robot heading simply by finding a modus of horizontal (in image coordinates) displacements of the tentative correspondences. The modus is found simply by histogram voting. In both cases, the tentative correspondences were established as suggested in original articles on SIFT [8] and BRIEF [11].

The proposed evaluation is based on a measure of the feature utility for a long-term mobile robot navigation. The utility is computed as the "success rate" that is a ratio of the correct estimations to the total number of estimations performed. An estimation of the heading is considered as correct if it differs from the ground truth by less than two degrees.

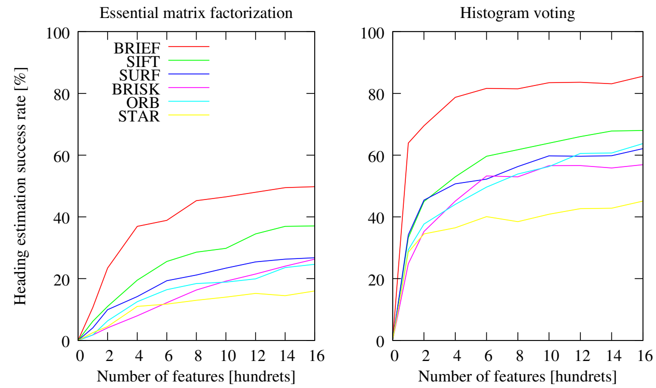


Fig. 3: Heading estimation success rate.

Since the success rate depends on the number of extracted features, we have performed ten evaluation trials, each with a different average amount of extracted features per image. Each trial consists of six evaluations, one for each feature extractor. The dataset contains 12 images from five locations, which means 660 comparisons for each evaluation, which allows to establish the success rate with a sufficient granularity. The dependence of the success rate on the number of extracted features is given in Fig. 3.

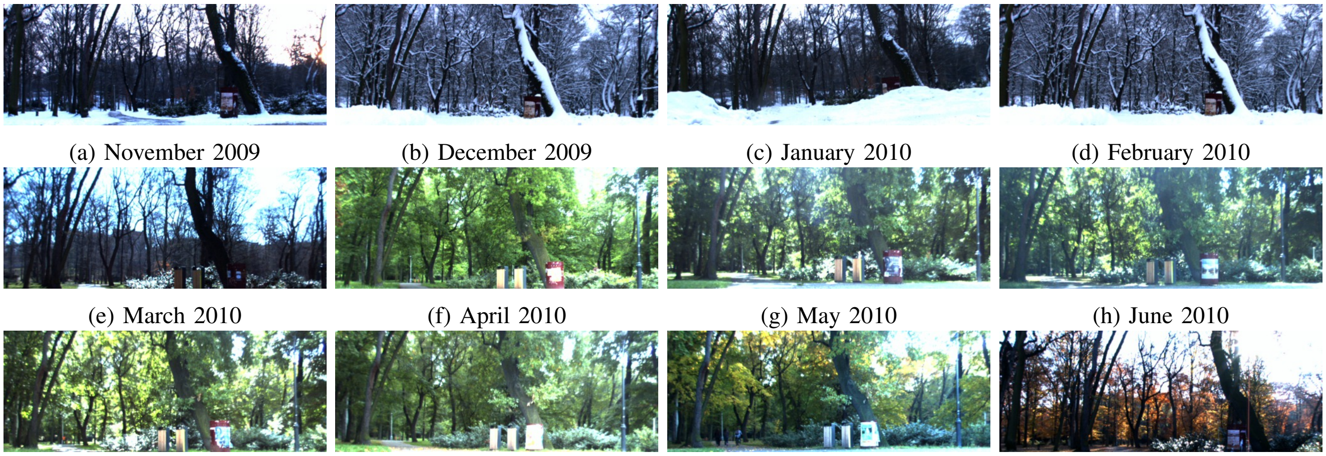
TABLE I: Success rates and real required computational times for 1000 extracted features

Feature Extractor	BRIEF	BRISK	ORB	SIFT	STAR	SURF
Required time [ms]	23	26	24	109	64	211
Histogram voting [%]	83	57	56	64	41	60
Hartley [%]	46	19	19	30	14	23

In addition, we also examined the real required computational time of the feature extractor that is established as the average runtime to extract one thousand of features per dataset image using Intel i5 PC running at 2.5 GHz. The results are depicted in Table I. Notice, SIFT seems to be faster than SURF, which is probably an implementation issue of the OpenCV library.

TABLE II: Success rates [%] between individual months (Histogram voting, 1000 BRIEF features per image)

			No foliage						*Foliage*			
	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Nov	—	100	100	100	100	100	100	80	100	80	60	100
Dec	100	—	100	100	100	100	60	60	80	80	60	100
Jan	100	100	—	100	100	100	80	40	60	40	60	80
Feb	100	100	80	—	100	100	80	40	60	100	80	60
Mar	100	100	100	100	—	100	80	20	60	40	40	60
Apr	100	100	100	100	100	—	80	80	60	80	60	80
May	100	60	60	80	80	100	—	100	100	100	100	100
Jun	80	20	60	40	40	100	100	—	100	100	100	100
Jul	80	80	60	80	40	80	100	100	—	100	100	100
Aug	80	100	60	100	40	60	100	100	100	—	100	100
Sep	60	80	40	80	40	40	100	100	100	100	—	100
Oct	100	100	100	80	60	80	100	100	100	100	100	—



(a) November 2009 (b) December 2009 (c) January 2010 (d) February 2010

(e) March 2010 (f) April 2010 (g) May 2010 (h) June 2010

(i) July 2010 (j) August 2010 (k) September 2010 (l) October 2010

Fig. 1: Captured seasonal changes at location II.

(a) Location I (b) Location III (c) Location IV (d) Location V

Fig. 2: View from the robot camera at different locations.

The matching success rate between individual months is shown in Table II to illustrate that the "map decay" is caused mainly by seasonal factors. The results also indicates that the matching reliability is decreased in cases when images with foliage are matched to images without foliage. Therefore, we performed two additional tests during early (no foliage) and late (foliage) April of 2011 to show that the seasonal factors are mainly periodical. In both cases, the P3AT robot has repeatably traversed the testing path using maps from the original dataset (captured in 2009 and 2010).

V. Conclusion

In this short paper, we report our results on the evaluation of image feature extractors to long-term environment changes caused by seasonal factors. The considered dataset captures variations of a park like environment throughout one year, and thus, allow us to consider suitability of image feature extractor methods for the long-term autonomous navigation using vision based estimation of the robot heading.

Regarding the results, the best performing method is BRIEF, which outperforms SIFT by more than thirty percentage points. In addition, the BRIEF extractor is also less computationally demanding and thus it seems to be the most suitable feature descriptor for the navigational purposes.

Besides, we found out that the histogram voting exhibits a better robustness than the classical method based on advanced algebraic approaches. It is also worth to mention that the histogram voting does not rely on an identification of the camera parameters, which can be considered as an additional advantage.

The results also suggest that just two maps (with and without foliage) are sufficient for a reliable outdoor navigation.

Acknowledgments

The work has been supported by EU projects FP7-ICT-600623, FP7-ICT-216240, FP7-ICT-216342, Argentinean project ARC/11/11, Czech MSMT projects 7AMB12AR022, 7E08006, 7E11017 and Czech Science Foundation project 13-18316P.

References

[1] J. Li and N. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, 2008.

[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[3] A. Schmidt, M. Kraft, and A. Kasiński, "An evaluation of image feature detectors and descriptors for robot navigation," in *International Conference on Computer Vision and Graphics*, 2010.

[4] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 157–165, February 28 2010.

[5] Z. Chen and S. T. Birchfield, "Qualitative vision-based path following," *IEEE Transactions on Robotics and Automation*, 2009.

[6] E. Royer et al., "Monocular vision for mobile robot localization and autonomous navigation," *Int. Journal of Computer Vision*, 2007.

[7] T. Krajník, J. Faigl, V. Vonásek et al., "Simple, yet Stable Bearing-only Navigation," *Journal of Field Robotics*, 2010.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, 2008.

[10] M. Agrawal et al., "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching," in *Proceedings of the ECCV*, 2008.

[11] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: binary robust independent elementary features," in *Proceedings of the ICCV*, 2010.

[12] S. Leutenegger, M. Chli, and R. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proceedings of the ICCV*, 2011.

[13] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European Conference on Computer Vision*, 2006.

[14] "Stromovka dataset," [Cit: 2013-03-25]. [Online]. Available: http://purl.org/robotics/stromovka_dataset

[15] H. Szücsová, "Computer vision-based mobile robot navigation," Master's thesis, Czech Technical University, Prague, Czech republic, 2011.