THE COMPARISON OF POINT FEATURE DETECTORS AND DESCRIPTORS IN THE CONTEXT OF ROBOT NAVIGATION

Adam Schmidt, Marek Kraft, Michał Fularz, Zuzanna Domagała

Abstract:

This paper presents the evaluation of various contemporary interest points detectors and descriptors pairs in the context of robots navigation. The robustness of the detectors and descriptors was assessed using publicly available datasets: the first gathered from the camera mounted on the industrial robot [17] and the second gathered from the mobile robot [20]. The most efficient detectors and descriptors for the visual robot navigation were selected.

Keywords: point features, detectors, descriptors

1. Introduction

The detection, description and matching of point features plays a vital role in most of the contemporary algorithms for visual odometry [1] [2] or visual simultaneous localization and mapping [3] [4]. Over the last years several new fast detectors (FAST [5], SURF [6], CenSurE-based STAR [7]) and descriptors (SURF [6], BRIEF [8], ORB [9], FREAK [10]) have been proposed and successfully applied to the robot navigation tasks. As the processing speed is the key aspect in such tasks some of the detectors and descriptors were implemented in the FPGA [11] [12] or simplified [13].

At the moment, to the extent of authors knowledge there is no comparative study of the newest point detectors and descriptors with regard to their applicability in robot navigation. In [14] and [15] the authors describe the desired characteristics of the point detectors and descriptors, however they do not present any experimental results. The authors of [17] compared various interest point detectors using sequences recorded with the camera placed on the industrial robot. In the follow up research they compared the detector-descriptor pairs efficiency, however only cross-correlation, SIFT [19] and DAISY descriptors were considered. Another experimental study was presented in [16] where the detector-descriptor pairs were graded according to the number of feature matches supporting the 8-point algorithm solution found by the RANSAC algorithm. This indirect evaluation method was caused by difficulties in gathering the ground truth correspondence data for image pairs.

This papers present the evaluation of the detector-descriptor pairs in the context of robot navigation. The measure of the pair's efficiency was based on the reprojection error of point feature pairs matched on two images. The images used were selected from the publicly available datasets [17] [20]. The analysis allowed to select detector and descriptor pair most suitable for application in the robot navigation both in the context of reliability as well as the processing time.

The section 2 provides a short summary of the detectors and descriptors evaluated in the study. The datasets used in the experiments and the evaluation procedure is presented in the Section 3. The section 4 contains results, concluding remarks and the future work.

2. Detectors and descriptors

2.1. FAST feature detector

The FAST [5] (Features from Accelerated Segment Test) feature detector inspects the values of the intensity function of pixels in a circle of radius r around the candidate point p. The pixel on a circle is considered 'bright' if its intensity value is brighter by at least t, and 'dark' if its intensity value is darker by at least t than the intensity value of p, where t is some arbitrary threshold. The candidate pixel is classified as a feature on a basis of a segment test – if a contiguous, at least n pixels long arc of 'bright' or 'dark' pixels is found in the circle. The original solution uses $r = 3$ and $n = 9$. The ID3 algorithm is used to optimize the order in which pixels are tested, resulting in high computational efficiency. The segment test alone produces small sets of adjacent positive responses. To further refine the results, an additional cornerness measure is used for non-maximum suppression (NMS). As NMS is applied only to a small fraction of pixels that positively passed the segment test, the detector preserves its speed.

2.2. SURF feature detector

SURF [6] (Speeded Up Robust Features) is an image feature detector and descriptor, inspired by the SIFT detector/descriptor. The main motivation for the development of SURF was to overcome SIFT's main weakness – its computational complexity and hence also low execution speed. SURF is reported to be up to a few times faster than SIFT without compromising the performance. The detection step in SURF takes advantage of the use of Haar wavelet approximation of the blob detector based on the Hessian determinant. The approximations of Haar wavelets can be efficiently computed using integral images, regardless of the scale. Accurate localization of multiscale SURF features requires interpolation.

2.3. STAR feature detector

The STAR keypoint detector was implemented as a part of the OpenCV computer vision library. It is a derivative of the CenSurE (Center Surround Extrema) feature detector [7]. The authors of the solution aimed at the creation of a multiscale detector with full spatial resolution. As described in [7], the subsampling performed by SIFT and SURF affects the accuracy of feature localization. The detector uses a bi-level approximation of the Laplacian of Gaussians (LoG) filter. The circular shape of the mask is

replaced by an approximation that allows to preserve rotational invariance and enables the use of integral images for efficient computation. Scale-space is constructed without interpolation, by applying masks of different size.

2.4. SURF corner descriptor

The SURF [6] feature descriptor uses Haar wavelets in conjunction with integral images to encode the distribution of pixel intensity values in the neighborhood of the detected feature while accounting of the feature's scale. Computation of the descriptor for a given feature at the scale s begins with the assignment of the dominant orientation to make the descriptor rotation invariant. The process starts with computing the Haar wavelet responses in two dominant directions for every point within the radius of $6s$ from the feature. The size of the square wavelet masks is also adjusted according to the feature scale and set to $4s$. The responses are then weighted with a Gaussian centered at the feature point. Each one of the responses gives rise to a point in the vector space, with the x-responses along the abscissa and the y-responses along the ordinate. Afterwards, a circle segment covering an angle of $\frac{\pi}{3}$ is rotated around the origin (feature point). The responses under the segment are summed and form a resultant vector. The rotation angle corresponding to the longest resultant vector is selected as the dominant orientation of the feature descriptor. The computation of the descriptor itself starts with placing a square window with a side length of $20s$ centered on the feature point and oriented as it was computed in the previous step. The window is divided into 4×4 regular square subregions. Each subregion is divided into 5×5 uniformly distributed sample points. For each sample point, Haar wavelet responses for two principal directions are computed. Each subregion contributes to the descriptor with four components: the sums of the responses in the two principal directions and their absolute values. The responses from the 16 subregions are once again weighted with a Gaussian. For 16 subregions, the descriptor size is 64.

2.5. BRIEF corner descriptor

The BRIEF [8] (Binary Robust Independent Elementary Features) descriptor proposed in [8] uses binary strings for feature description and subsequent matching. This enables the use of Hamming distance to compute the descriptor similarity. Such similarity measure can be computed very efficiently – much faster than the commonly used L_2 norm. Due to BRIEF's sensitivity to noise, the image is smoothed with a simple averaging filter before applying the actual descriptor. The value of each bit contributing to the descriptor depends on the result of a comparison between the intensity values of two points inside an image segment centered on the currently described feature. The bit corresponding to a given point pair is set to if the intensity value of the first point of this pair is higher than the intensity value of the second point, and reset otherwise. The sampling strategy for the selection of point for the pairs to be compared was selected based on experiments with uniform and Gaussian random sampling using different distribution parameters. The proposed descriptor is 512-bit long and computed over a 48×48 pixel image patch. The initial smoothing is performed with a 9×9 pixel rectangular averaging filter. The basic form of BRIEF is not invariant w.r.t. rotation.

2.6. ORB feature descriptor

The ORB [9] (Oriented FAST and Rotated BRIEF) descriptor extends the BRIEF descriptor by adding two important improvements. The first one is to augment the descriptor with orientation data from the FAST feature detector. This allows to make the descriptor robust to in-plane rotation. This is done by rotating the coordinates of the point pairs for binary tests around the described feature by the feature orientation angle. Second innovation is the selection scheme for point pairs whose comparisons contribute to the descriptor. The random sampling has been replaced with a sampling scheme that uses machine learning for de-correlating BRIEF features under rotational invariance. This makes the nearest neighbor search during matching less error-prone.

2.7. FREAK feature descriptor

The FREAK [10] (Fast Retina Keypoint) descriptor is another extension of the basic concepts of BRIEF [8]. It provides the descriptor with feature orientation by summing the estimated local gradients over selected point pairs. Using a specific point sampling pattern allows to apply more coarse discretization of rotation, allowing for savings in memory consumption. A special, biologically inspired sampling pattern is also used. While the resulting descriptor is still a binary string, the sampling pattern allows for the use of a 'coarse-to-fine' approach to feature description. Point pairs carrying the information on most distinctive characteristics of the feature neighborhood are compared in the first place. This allows for faster rejection of false matches and shortening of the computation time.

3. Experiments

3.1. Datasets

Two datasets were used in the experiments. The Robot Data Set [17] [18] consists of 60 scenes, registered from 119 positions under varying lighting conditions using high resolution camera mounted on the industrial robot. The 119 positions form 4 trajectories: 3 angular with constant distance from the scene and one linear with constant camera heading (Figure 1). Such a diverse dataset allows to evaluate the robustness of detector-descriptor pairs with regard to the scale, rotation and illumination changes. Exemplary images from the dataset are presented on the Figure 2.

However, such variety is rarely seen in the video sequences gathered by a mobile robot (especially in indoor environment). Therefore the detector-descriptor pairs were also tested on the sequences gathered with the Kinect sensor mounted on the wheeled robot [20]. The 'Pioneer SLAM' sequence consisting of 2921 images was used. The robot's trajectory during the sequence registration is presented on the Figure 4. Exemplary images from the second dataset are presented on the Figure 3.

Both the datasets contain images, ground truth data on the camera position, camera intrinsic parameters and distortion coefficients.

3.2. Evaluation

The following procedure was performed for every analyzed pair of images:

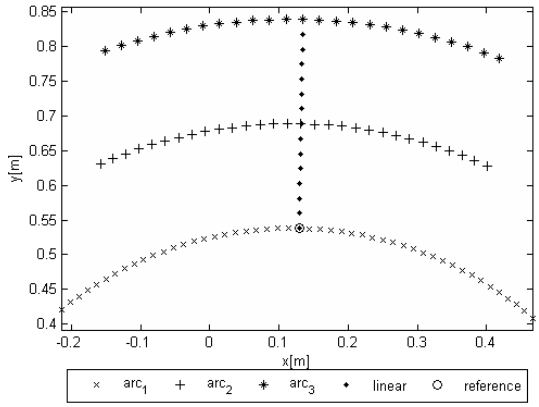

Fig. 1. The four trajectories of the Robot Data Set and the position of the reference frame

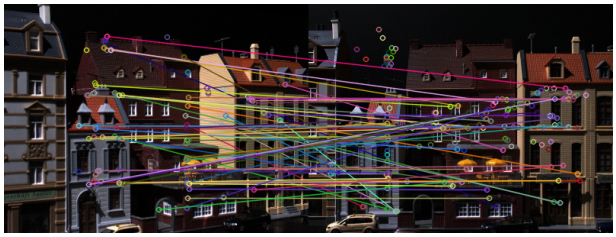

Fig. 2. Exemplary pair of images from the Robot Data Set with matches using the FAST detector and BRIEF descriptor

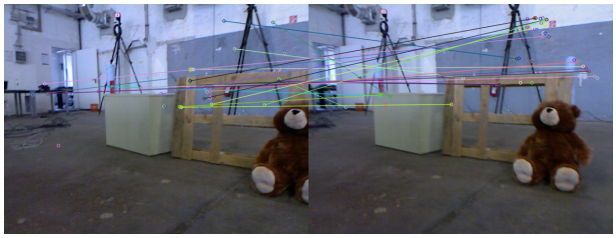

Fig. 3. Exemplary images from the Freiburg Data Set with matches using the FAST detector and BRIEF descriptor

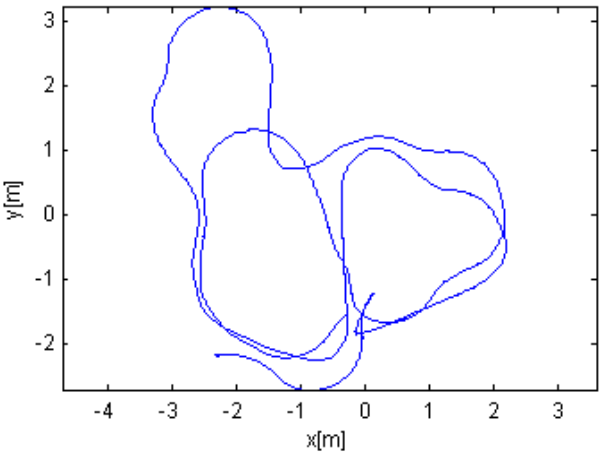

Fig. 4. The 'Pioneer SLAM' trajectory

1) the point features were detected on both images using the selected detector

Detector	Robot Image Dataset	Freiburg Dataset
FAST	1469	634
Pyramid FAST	2292	1074
GFTT	1469	868
PyramidGFTT	2984	773
SURF	4580	1421
StarKeypoint	221	101

Tab. 1. The average number of detected features.

2) the point features descriptors were calculated using the selected descriptor algorithm
3) the coordinates of the features were undistorted and normalized according to the camera parameters
4) the essential matrix E describing the images' epipolar geometry was calculated using the relative ground truth translation ($t = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$) and rotation ($R$) between the two camera positions:

$$E = R \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (1)$$

5) the features from both images were matched by minimizing the distance between their descriptors resulting in the set of quadruples (u_i, v_i, U_j, V_j) where (u_i, v_i) are the normalized coordinates of the feature extracted from the first image and (U_j, V_j) are the normalized coordinates of the best matching feature from the second image
6) for each quadruple the symmetric reprojection error was calculated according to:

$$err = MAX\left(|e_i, (U_j, V_j)|, |e_j, (u_i, v_i)|\right) \quad (2)$$

where e_i and e_j are the epipolar lines defined as:

$$\begin{aligned} a_i x + b_i y + c_i &= 0 && (3)\\ A_j x + B_j y + C_j &= 0 && (4)\\ \begin{bmatrix} a_i & b_i & c_i \end{bmatrix}^T &= E \begin{bmatrix} u_i, v_i, 1 \end{bmatrix}^T && (5)\\ \begin{bmatrix} A_j & B_j & C_j \end{bmatrix}^T &= \begin{bmatrix} U_j, V_j, 1 \end{bmatrix} E && (6) \end{aligned}$$

If the error was smaller than the threshold $thresh$ the match was considered an inlier

7) the ratio of the number of inliers to the number of matches was calculated

The final score of the detector-descriptor pair over a dataset was calculated as the mean of the inliers to matches ratios of all the image pairs in the dataset.

4. Results and conclusions

The various, contemporary point features detector and descriptor pairs were compared in order to determine the best combination for the task of robot visual navigation. The sequences chosen as a testbed displayed typical point feature distortions encountered during indoor mobile robot navigation – scaling and affine transformation with very little or none in-plane rotation. The experiments show, that binary vector based BRIEF and ORB descriptors

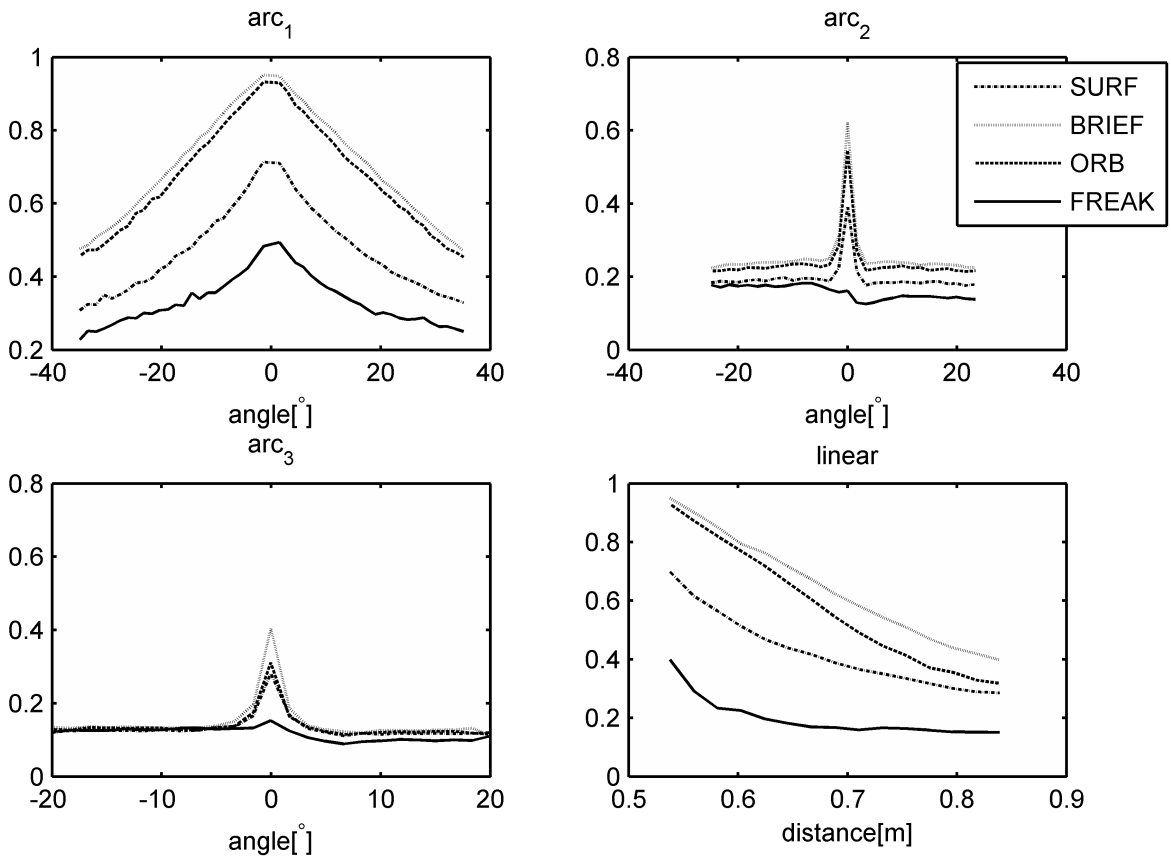

Fig. 5. Matching ratio of points detected with the FAST detector. Robot Image Dataset

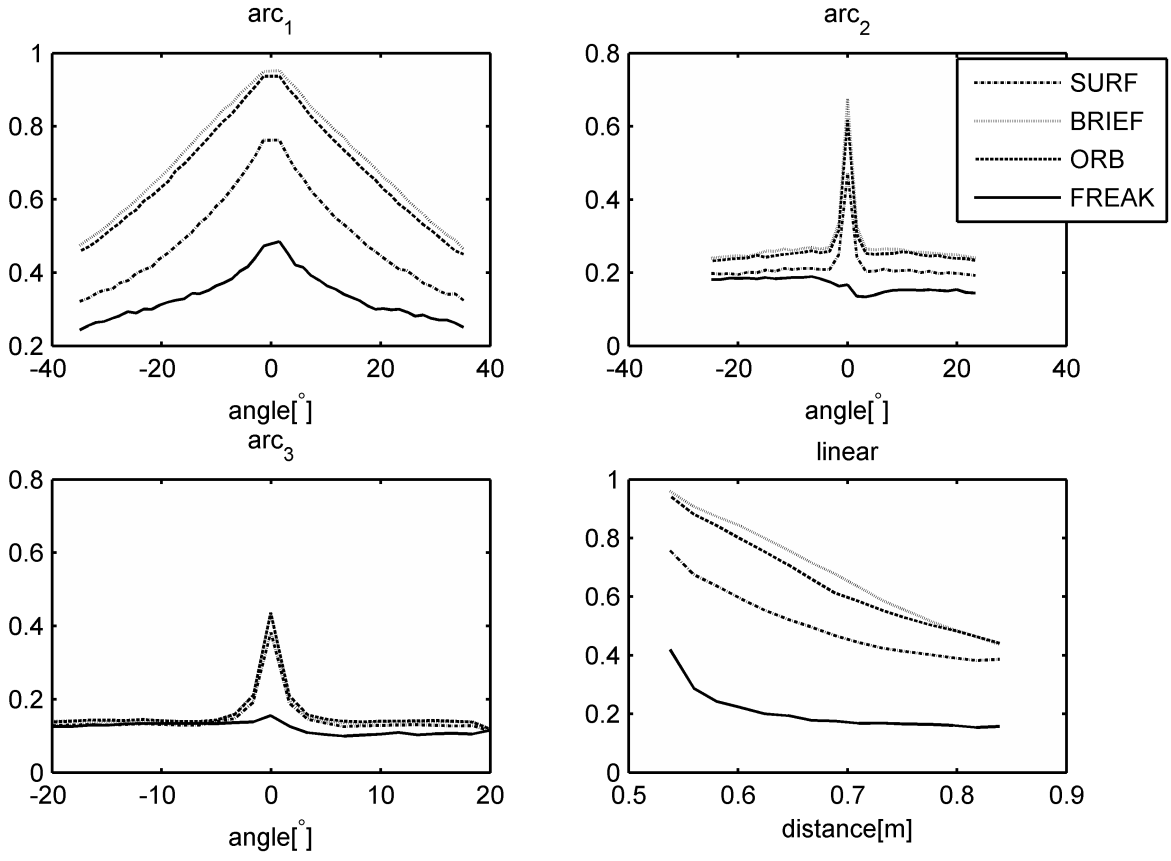

Fig. 6. Matching ratio of points detected with the Pyramid FAST detector. Robot Image Dataset

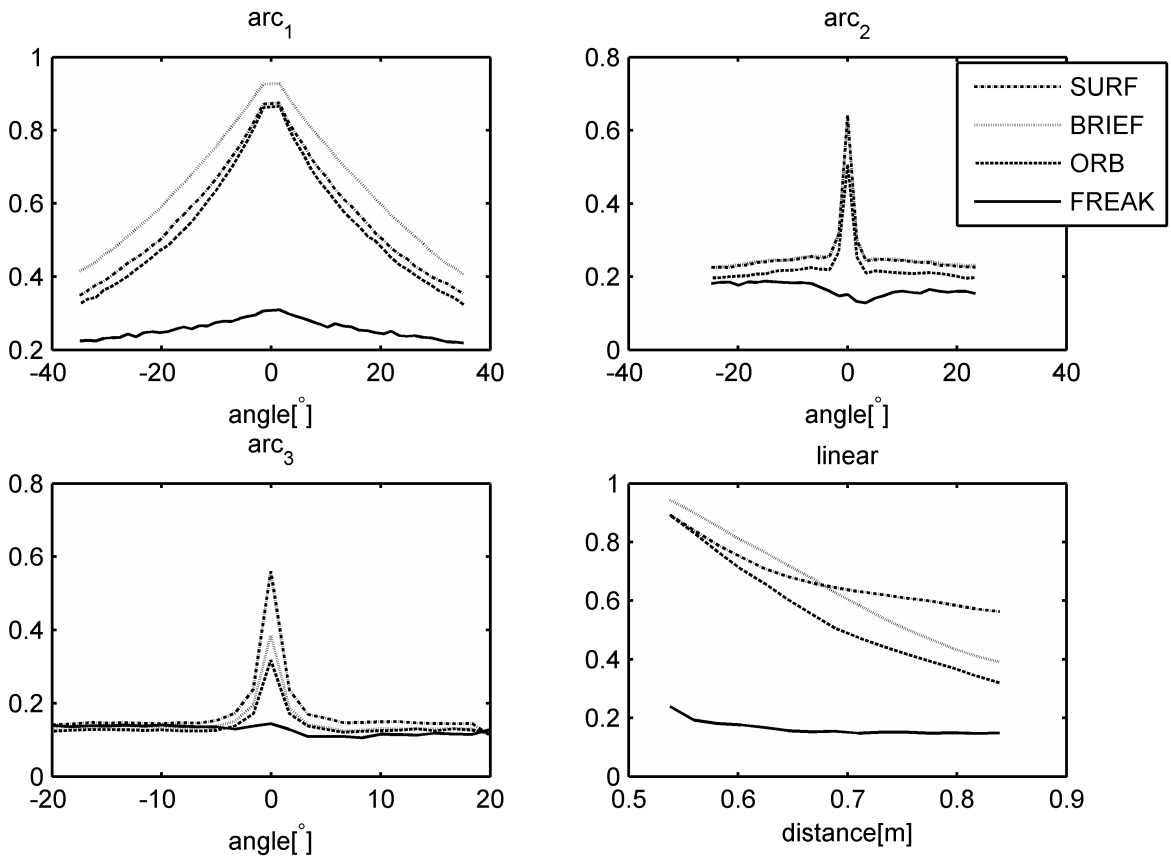

Fig. 7. Matching ratio of points detected with the GFTT detector. Robot Image Dataset

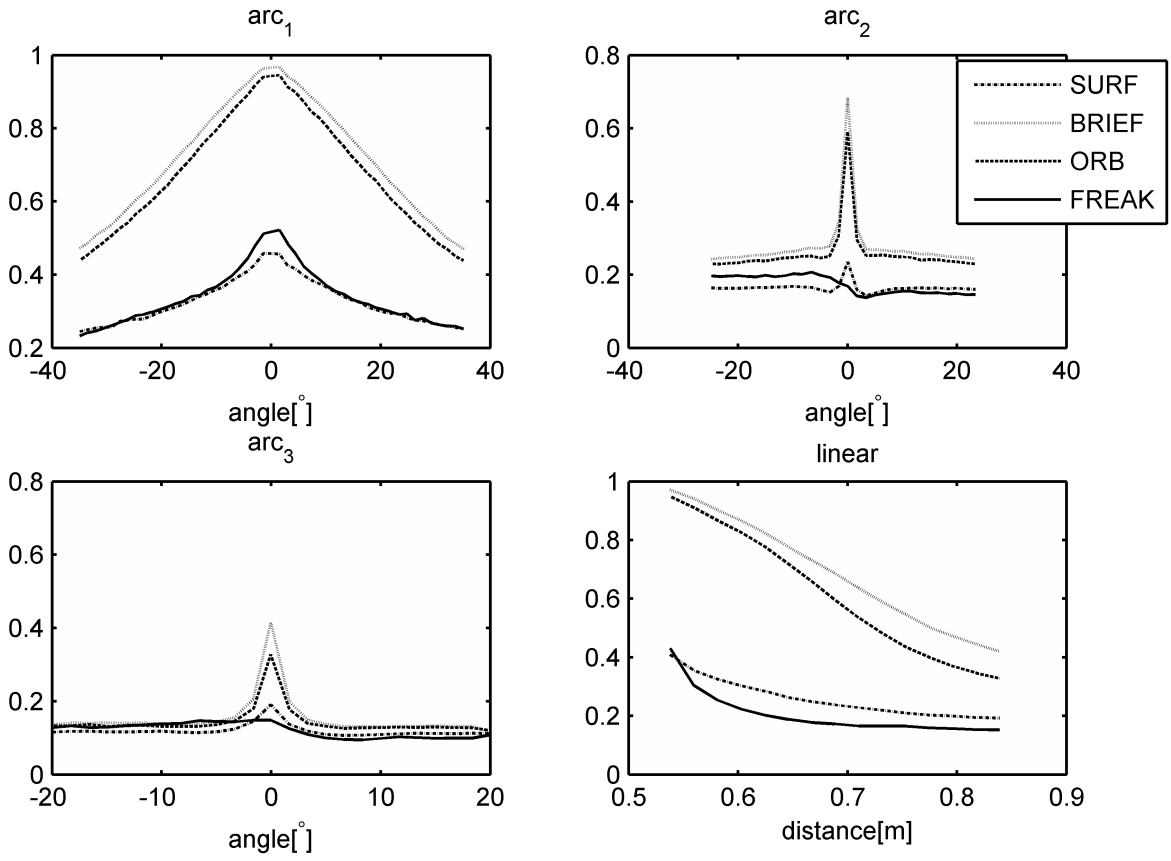

Fig. 8. Matching ratio of points detected with the Pyramid GFTT detector. Robot Image Dataset

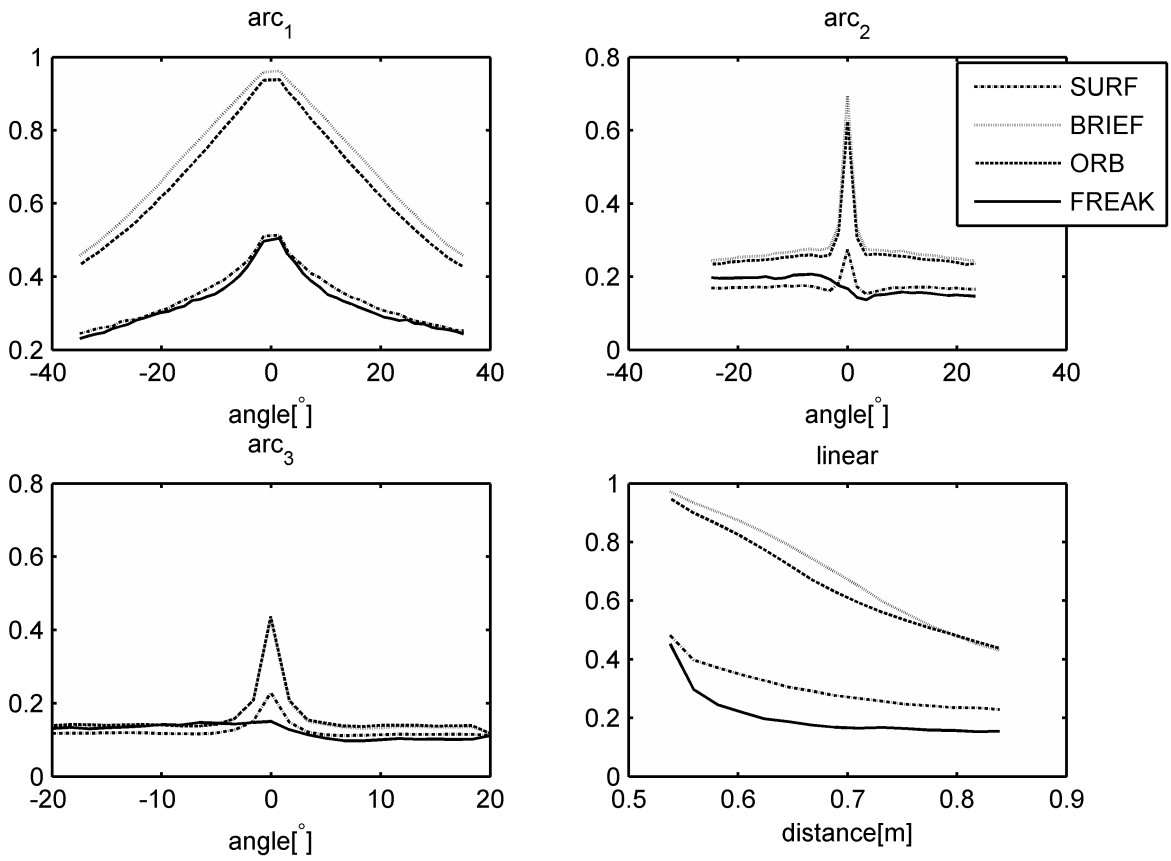

Fig. 9. Matching ratio of points detected with the SURF detector. Robot Image Dataset

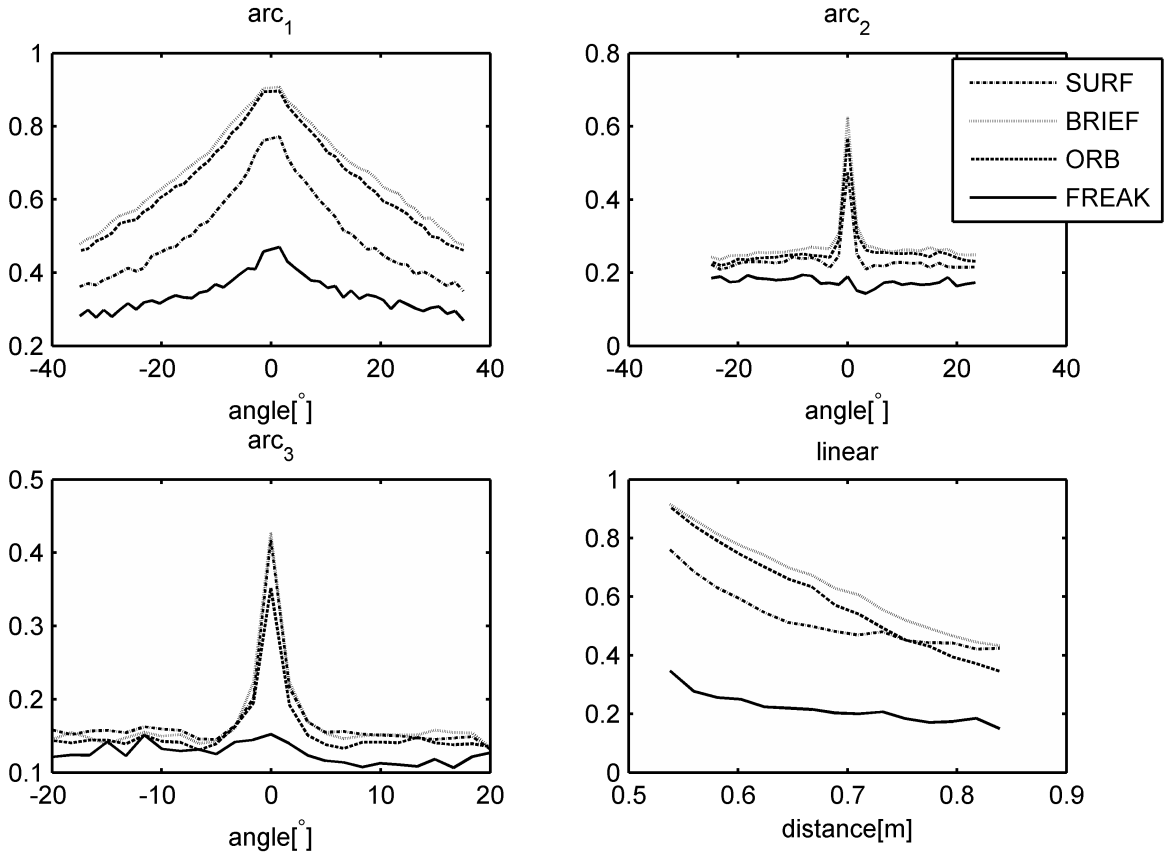

Fig. 10. Matching ratio of points detected with the StarKeypoint detector. Robot Image Dataset

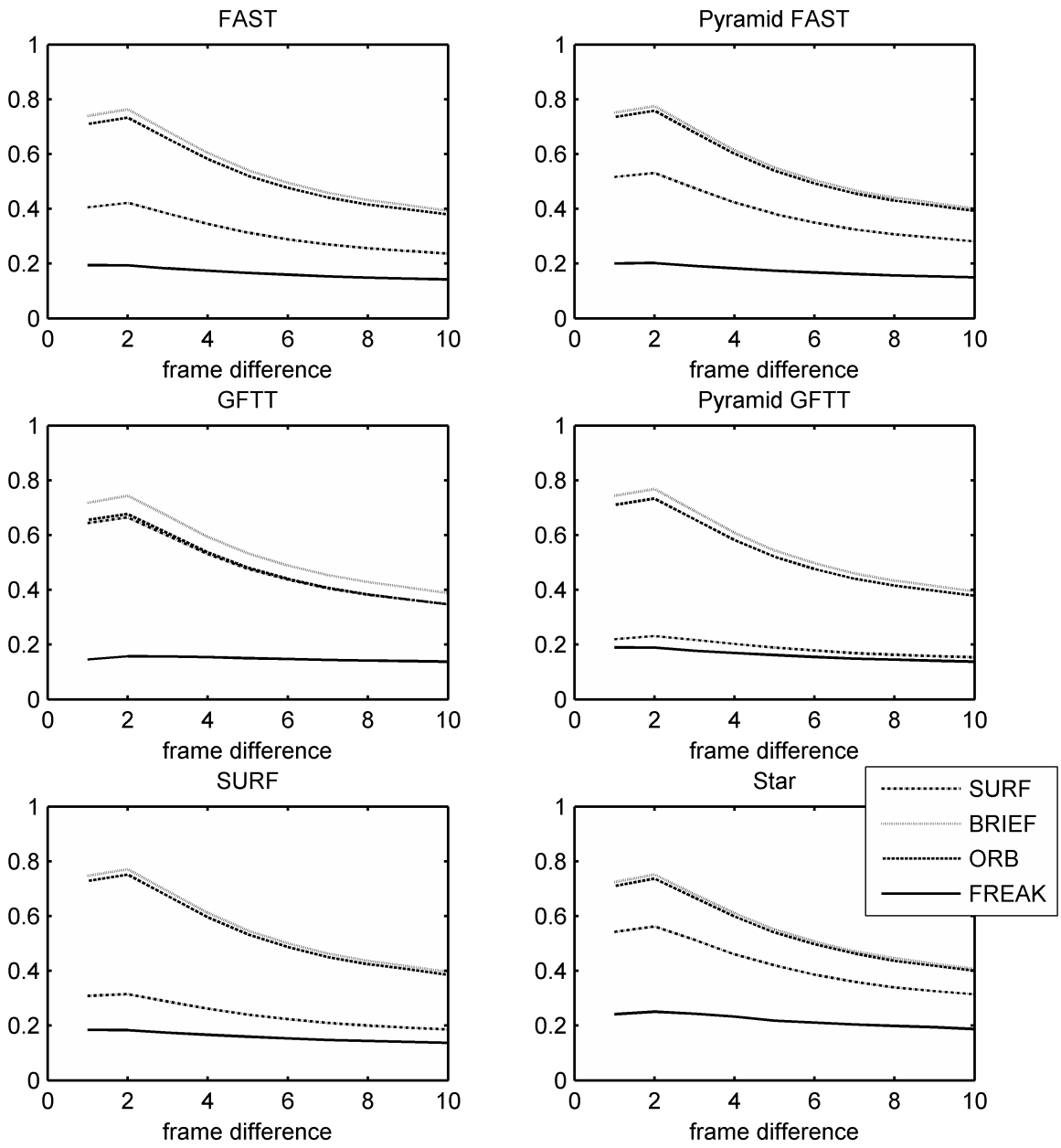

Fig. 11. Matching ratio of points detected with the analyzed detectors. Freiburg Dataset

clearly outperform the other tested solutions. Interestingly, the third binary vector descriptor - FREAK - displayed a significantly lower matching accuracy than the other tested solutions, contrary to the claims in [10]. The popular SURF feature descriptor performed relatively well, but since it offers lower matching accuracy than BRIEF and ORB and is slower to compute and match, there are strong arguments against using it. SURF is also the descriptor being most influenced by the type of the detector it is paired with. This is visible especially in the results from the Freiburg dataset (figure 11). Interestingly, the use of pyramid FAST and GFTT detectors with a multi-scale SURF and ORB descriptor does not increase the matching performance when using a scale-aware descriptor. This is rather surprising, and is probably caused by the necessity of performing interpolation to determine the location of the feature, which negatively impacts the feature location accuracy in higher scales. The Star keypoint detector is not burdened by this additional inaccuracy, as it offers full location accuracy across all scales. The results for the linear sequence shown in figure 10 support this claim. It must be noted however, that the Star descriptor requires a relatively feature-rich environment, as the average feature count it returns is relatively low as shown in table 1. The tests performed on the Freiburg dataset show also, that the computation of the visual odometry parameters under too small displacement of features between two consecutive images is inaccurate. In this dataset, the samples (robot-

mounted camera images) are taken fairly densely, and frame-to-frame odometry computation returns less inliers than computation using every 2^{nd} frame (see figure 11). As all descriptors perform well when paired with the FAST corner detector, the FAST-BRIEF pair is a good choice when processing speed is a concern. Under the camera movement conditions featured in both of the test sequences used, the additional computational cost to bear when using descriptors and detectors robust to in-plane rotation and large scaling seems to be unjustified. This also confirms the need for testing the detector-descriptor pairs in the context of the application, as the requirements raised by it may differ significantly from the requirements raised by typical, commonly used benchmarking image sequences. In the future it is planned to compare the efficiency of the detector-descriptor pairs in the monocular SLAM system.

AUTHORS

Adam Schmidt* – Poznań University of Technology, Institute of Control and Information Engineering, ul. Piotrowo 3A, 60-965 Poznań, Poland, e-mail: Adam.Schmidt@put.poznan.pl

Marek Kraft – Poznań University of Technology, Institute of Control and Information Engineering, ul. Piotrowo 3A, 60-965 Poznań, Poland, e-mail: Marek.Kraft@put.poznan.pl

Michał Fularz – Poznań University of Technology, Institute of Control and Information Engineering, ul. Piotrowo 3A, 60-965 Poznań, Poland, e-mail: Michal.Fularz@put.poznan.pl

Zuzanna Domagała – Poznań University of Technology, Institute of Control and Information Engineering, ul. Piotrowo 3A, 60-965 Poznań, Poland, e-mail: Zuzanna.Domagala@cie.put.poznan.pl

* Corrresponding author

References

[1] D. Scaramuzza, F. Fraundorfer, Visual Odometry: Part I - The First 30 Years and Fundamentals, IEEE Robotics and Automation Magazine, vol. 18(4), 2011, pp. 80–92

[2] F. Fraundorfer, D. Scaramuzza, Visual Odometry: Part II - Matching, Robustness and Applications, IEEE Robotics and Automation Magazine, vol. 19(2), 2012, pp. 78–90

[3] A. J. Davison, I. Reid, N. Molton and O. Stasse, MonoSLAM: Real-Time Single Camera SLAM, IEEE Trans. PAMI, vol. 29(6), 2007, pp. 1052¸-1067

[4] A. Schmidt, A. Kasiński, The Visual SLAM System for a Hexapod Robot, Lecture Notes in Computer Science, vol. 6375, 2010, pp. 260–267

[5] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, in Proc. of European Conf. on Computer Vision, 2006 pp. 430–443

[6] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding, vol. 110(3), 2008, pp. 346–359

[7] M. Agrawal, K. Konolige, M.R. Blas, CenSurE: Center surround extremas for realtime feature detection and matching, Lecture Notes in Computer Science, vol. 5305, 2008, pp. 102–115

[8] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, BRIEF: Binary Robust Independent Elementary Features, in Proceedings of ECCV 2010, pp. 778–792.

[9] E. Rublee, V. Rabaud, K. Konolige, G. R. Bradski, ORB: An efficient alternative to SIFT or SURF, in Proc. ICCV, 2011, pp. 2564–2571

[10] A. Alahi, R. Ortiz, P. Vandergheynst, FREAK: Fast Retina Keypoint. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2012

[11] M. Kraft, A. Schmidt, A. Kasinski, High-speed image feature detection using FPGA implementation of FAST algorithm, in Proc. VISAPP, 2008, pp. 174–179

[12] M. Kraft, M. Fularz, A. Kasiński, System on chip coprocessors for high speed image feature detection and matching, in Proc. of Advances Concepts for Intelligent Vision Systems, 2011, pp. 599–610

[13] M. Kraft, A. Schmidt, Simplifying SURF feature descriptor to achieve real-time performance, in Proc. Computer Recognition Systems, 2011, pp. 431–440

[14] Ó. Martínez, A. Gil, M. Ballesta, O. Reinoso, Interest Point Detectors for Visual SLAM, in Proc. of the Conference of the Spanish Association for Artificial Intelligence, 2007

[15] M. Ballesta, A. Gil, Ó. Martínez, O. Reinoso, Local Descriptors for Visual SLAM, in Proc. Workshop on Robotics and Mathematics, 2007

[16] A. Schmidt, M. Kraft, A. Kasiński, An evaluation of image feature detectors and descriptors for robot navigation, Lecture Notes in Computer Science, vol. 6375, 2010, pp. 251-259

[17] H. Aanas, A. L. Dahl, K. S. Pedersen, Interesting Interest Points - A Comparative Study of Interest Point Performance on a Unique data set, International Journal of Computer Vision, vol. 97, 2011, pp. 18–35

[18] A. L. Dahl, H. Aanas, K. S. Pedersen, Finding the Best Feature Detector-Descriptor Combination. inProc. of 3DIMPVT, 2011

[19] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, vol.60(2), 2004, pp. 91–110

[20] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremer, Towards a benchmark for RGB-D SLAM evaluation, in Proc. of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf. (RSS), 2011

[21] J. Shi, C. Tomasi, Good Features to Track, in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 1994, pp. 593–600