Appearance-Based Loop Closure Detection with Scale-Restrictive Visual Features

Konstantinos A. Tsintotas(✉), Panagiotis Giannis, Loukas Bampis, and Antonios Gasteratos

Laboratory of Robotics and Automation, School of Engineering, Department of Production and Management Engineering, Democritus University of Thrace, 67132 Xanthi, Greece
{ktsintot,panagian8,lbampis,agaster}@pme.duth.gr
https://robotics.pme.duth.gr/robotics/

Abstract. In this paper, an appearance-based loop closure detection pipeline for autonomous robots is presented. Our method uses scale-restrictive visual features for image representation with a view to reduce the computational cost. In order to achieve this, a training process is performed, where a feature matching technique indicates the features' repeatability with respect to scale. Votes are distributed into the database through a nearest neighbor method, while a binomial probability function is responsible for the selection of the most suitable loop closing pair. Subsequently, a geometrical consistency check on the chosen pair follows. The method is subjected into an extensive evaluation via a variety of outdoor, publicly-available datasets revealing high recall rates for 100% precision, as compared against its baseline version, as well as, other state-of-the-art approaches.

Keywords: Localization · Mapping · Visual-based navigation · Mobile robots

1 Introduction and Literature Review

Nowadays, a major research focus is devoted in robots' map formulation techniques via the utilization of several exteroceptive sensors, such as laser scanners, cameras, odometry and inertial measurement units [4,13,20,27]. An autonomous system in an unknown environment needs to construct a map of the working area in order to perform various tasks, such as path planning, exploration, and collision avoidance. Simultaneous Localization And Mapping (SLAM) is the process

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-00737). The paper was partially supported by project ETAA, DUTH Research Committee 81328.

D. Tzovaras et al. (Eds.): ICVS 2019, LNCS 11754, pp. 75–87, 2019.
https://doi.org/10.1007/978-3-030-34995-0_7

where the robot concurrently constructs a model of the environment (the map), while at the same time is able to estimate its position as moving within it [12,32]. Thus, SLAM is a sine qua non procedure in any modern autonomous system. One of the essential components for any SLAM architecture is place recognition, a process which allows an intelligent mechanism to realize if a location has been already visited, widely known as loop closure detection [1,33,34].

Owed to the increased availability of computational power during the last years, cameras overcame range type sensors, e.g., laser, ultrasonic, radar, due to the rich textural information provided by visual data. Appearance-based place recognition is the ability to trigger a loop closure in the environment using vision as the main sensory modality. Traditionally, place recognition is casted as an image retrieval task since the query instance (the current robot view) seeks for the most visually similar one, by searching into the database. Each database visual information is represented using invariant local features, such as SURF [7], SIFT [22], or binary equivalents like BRISK [21] and ORB [29]. Comparisons are performed via voting techniques [17,23], in order to highlight the proper candidate. Gehrig et al. [17] propose a loop closure framework depending on the votes' aggregation which each database instance pools through a k Nearest Neighbor (k-NN) scheme, while a probabilistic score highlights the proper pair. Accompanied with a geometrical verification step, outliers are avoided. The authors in [23] built a k-d tree from projected BRISK descriptors and via a similar NN method previsited locations are identified. Despite the fact that the outcome of a voting procedure is more robust, the searching process is computational costly.

To address the challenge of complexity, recent studies [1,6,11,15,26,36], adopted the Bag-of-Words (BoW) model [30] originally proposed for text retrieval tasks [3]. BoW approaches make use of a visual vocabulary, generated off-line through a training procedure, in order to represent images with visual word histograms. Owing to histogram comparisons the proper instance is selected. In [1], a BoW algorithm for scene recognition is proposed. Along the navigation, two parallel visual vocabularies (one for image-descriptors and one for color histograms) are constructed and combined. Candidate pairs are highlighted via a Bayesian filter and validated via a epipolar geometry constraint. In our previous work we propose the representation of a group of instances by a common visual word histogram [5,6]. Matches are indicated through these histogram comparisons and a quantitative interpretation of temporal consistency enhances the results. A probabilistic appearance-based pipeline based on a pretrained vocabulary of SIFT descriptors is proposed in [11]. In addition, this approach includes a Chow and Liu tree to learn the co-occurrence probabilities among visual words [9]. Similarly, a binary vocabulary accompanied with geometrical and temporal checks prevents the system from fault detections [15,26]. Providing the well known sequence-based place recognition algorithm SeqSLAM [25] with the BoW model a significant performance improvement is presented in [36].

Although, BoW has shown efficient performance in loop closure detection frameworks, it has a key drawback which related to the training procedure. The

visual vocabulary is generated offline via a set of descriptors extracted from a generic environment. This practically means that the system may not be able to represent the incoming images appropriately and false detections may appear due to perceptual aliasing (high similarity between different locations). In order to avoid such situations, incremental dictionaries which are build on-line appeared in the robotics community [10,16,19,33,35,37]. Cieslewski et al. [10] proposed a voting scheme where a search into the database is performed by the usage of an incremental vocabulary tree, in order to retrieve the appropriate match. In [16], an incremental visual dictionary is built on a hierarchical structure of visual words. Similarly, in [19,35] visual words are generated on-line through a local feature-tracking, while voting techniques are responsible for the detection of loop closures. The authors in [37] propose a binary codebook with perspective invariance to the camera's motion, while unique visual words are generated on-line and assigned to dynamic places of the traversed map in [33].

Convolutional Neural Networks-based approaches were recently introduced with a view to solve the place recognition task [2,28,31]. Convolutional or fully connected layers are used as image descriptors and comparisons are performed among them. Despite their highly efficient performance, these frameworks are known for their excess demand in computational resources [24].

This paper presents a straightforward appearance-based loop closure pipeline, which relies on the images' description by a selective subset of raw SURF visual features, with the aim to reduce the computational complexity. The selected keypoints are chosen with respect to the scale which are extracted, similarly to BoRF [38]. In order to define the most informative visual features, we examine their repeatability among consecutive images though a features' matching technique. This decision is based on the observation that a feature's extracted scale, whose repeatability is strong, is not affected by variations in velocity or view point. Following this pipeline a vast database reduction is achieved, minimizing the computational complexity of a large feature multitude. Subsequently, a voting technique is performed to the descriptors' space via a k-NN technique, and a binomial probability function determines the proper candidates. Finally, a geometrical verification check between the chosen pair suppresses the system's false detections. Due to the careful features' selection, our method is capable of achieving high recall rates with 100% precision, as evaluated on three outdoor, community datasets.

The rest of the paper is organized as follows: Sect. 2 contains the formulation of the method in detail. The experimental results demonstrating the feasibility of the proposed pipeline are in Sect. 3, while in Sect. 4, the conclusion and future work are discussed.

2 Methodology

In this section an extended description of the proposed pipeline is presented in detail. The core algorithm of the system is based on the usage of scale-restrictive visual features in order to represent the incoming visual sensory information for

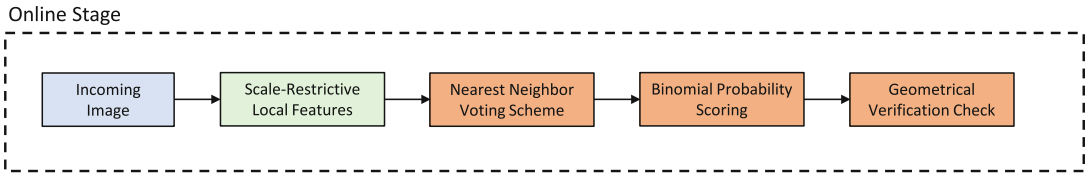

Fig. 1. Method's overview.

loop closure detection. A k-NN scheme follows, aiming to implement the votes' pooling, while a binomial distribution function is adopted with a view to classify pre-visited and non-visited locations, as proposed in [35]. An overview of the method is illustrated in Fig. 1.

2.1 Scale-Restrictive Visual Features Projection

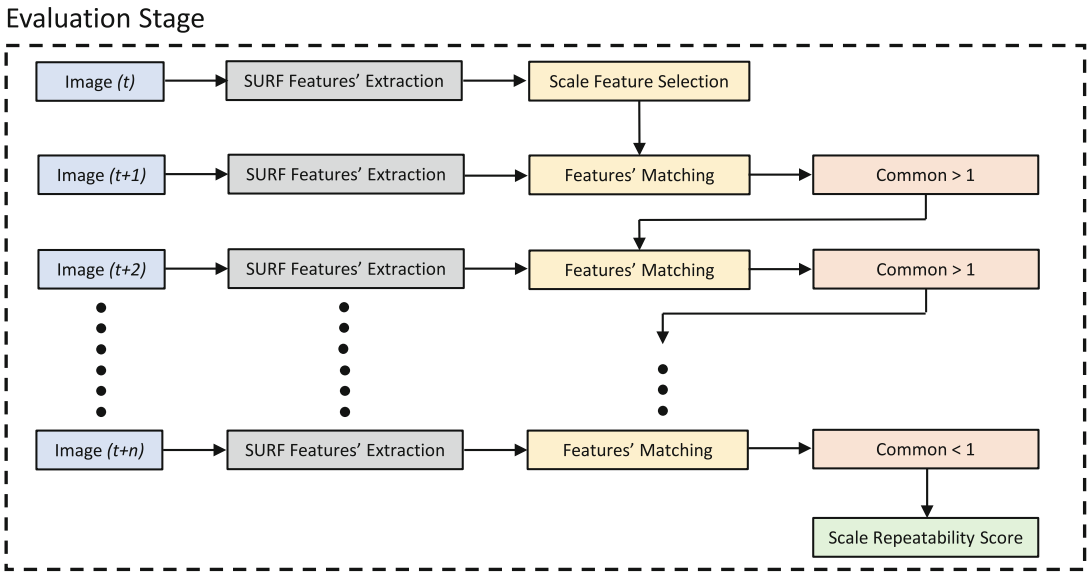

Fig. 2. Proposed scheme for measuring the features' scaling repeatability.

The features extraction procedure finds points of interest or keypoints in the incoming visual data that are distinct in a certain way. Beyond the achieved compression in the amount of data, these features have various invariance properties which define them, viz. scale, rotation, intensity, giving them the capability for images' comparison. Although robust results can be provided, the searching process is computational costly, especially in cases where the robot encounters a highly informative texture. Aiming to a system with improved matching success and operational frequency, we examine the SURF visual features' repeatability with respect to scale along consecutive images. In this off-line process, we are able to selectively detect these scale-restrictive keypoints which are meant to

be used during the course of the main procedure. More specifically, at time t, a low quantity of scale-restrictive features $D_{S(t)}$ are extracted from the incoming visual sensory information. These features are matched with the total of extracted descriptors $D_{(t+1)}$ in the following image $I_{(t+1)}$:

$$\left| D_{S(t)} \bigcap D_{(t+1)} \right|, \tag{1}$$

where $|X|$ denotes the cardinality of set X. Common features are maintained and subsequently matched to the next ones $D_{(t+2)}$ until the correlation between the images' descriptors cease to exit:

$$\left| D_{S(t)} \bigcap_{i=t+1}^{i=t+n} D_{(i)} \right| \leq 1, \tag{2}$$

whilst a counter representing the passed images is retained. In this process, we are not interested about the number of frames each feature is presented along the incorporated instances, but only for the duration of feature matching. Our scaling repeatability algorithm is shown in Fig. 2.

2.2 Nearest Neighbor Vote Assignment

When a query image I_Q is captured, we retain only the local features detected in scales with the highest repeatability, as measured from the procedure presented in Sect. 2.1. Then, a searching scheme is performed with all the pre-visited locations in the database, seeking for the most suitable loop closing candidates. The proposed framework utilizes a voting mechanism between the query's generated scale-restrictive features and the database's ones. A k-NN ($k = 1$) search defines the most similar descriptors in the traversed path and votes are distributed into the corresponding locations. The vote density $x_l(t)$ of each visited location l constitutes the primary factor for the binomial probability function. Subsequently, in order to avoid erroneous detections in cases where the robot' velocity decreases or the platform remains still, the proposed pipeline seeks into the database frames which are recorded prior to a temporal constant of 40 s [33]. This way, the system maintains the certainty that I_Q does not share common features with early visited locations.

2.3 Database Probabilistic Assignment

After votes' aggregation, each location in the database receive a matching score, obtained via a binomial probability function [17], which evaluates the similarity with the query image. Using the probabilistic score, the naïve approach of selecting a heuristic threshold over the aggregated votes is avoided. The proposed score examines the rareness of an event and is based on the assumption that if a robot visits a new location, which has never encountered before, votes should be distributed randomly over the total of the traversed map. Accordingly, in

case where a location has been seen in the past, the corresponding votes' density should be high indicating the existence of a loop closure candidate:

$$X_l(t) \sim Bin(n, p), n = N(t), p = \frac{\lambda_i}{\Lambda(t)}, \tag{3}$$

where $X_l(t)$ represents the random variable for the number of aggregated votes of the pre-visited location l at time t, N denotes the multitude of query's projected scale-restrictive features, λ is the total of features in l, and Λ corresponds to the size of database searching area ($\Lambda = \sum_{i=1}^{i=t-40s} D_{(i)}$).

The probabilistic score is calculated for each traversed location which gathers more than one votes, in order for a time saving to be performed, while two conditions have to be satisfied before an instance is indicated as loop closure candidate. First, a binomial probability threshold θ needs to be met:

$$Pr(X_l(t) = x_l(t)) < \theta < 1, \tag{4}$$

and additionally the number of accumulated votes has to be greater than the distribution's expected value:

$$x_l(t) > E[Xl(t)], \tag{5}$$

such that the system being able to discard cases with low votes' pooling.

2.4 Candidate Selection and Geometrical Verification Approvement

Up to this point, the proposed pipeline is capable of highlighting a set of pre-visited locations in the navigated map as candidates for loop closure events. Since the decision threshold may provide more than one detections, the algorithm selects as proper the one with the highest votes' accumulation (I_L). The chosen image is then processed for further validation. Aiming to a robust system without any false detection, the matched pair (I_Q, I_L) is subjected to a geometrical check through a RANSAC scheme [14] for the estimation of a representative fundamental matrix. If the computation of this matrix fails or the number of inlier points to the corresponding transformation is lower than a factor ($\varphi < 9$), the candidate loop closure detection is ignored. The parameterization of the applied RANSAC method is based on [35].

3 Experimental Validation

In this section, the experimental protocol followed by this work is described in detail. A total of three outdoor image-sequences is selected and used for the method's assessment. The system's evaluation is presented and comparisons against its baseline version and one state-of-the-art method are also reported and discussed. Experiments were performed on an Intel i7-6700HQ 2.6 GHz processor with 8 GB of RAM.

Table 1. Datasets' synopsis

Name	Description	Images	Frequency	Resolution
KITTI 05 [18]	Dynamic, urban area observing mostly buildings and cars	2761	10 Hz	1241×371
Lip 6 Outdoor [1]	Highly dynamic, urban dataset of crowed street recorded from a hand-held camera	301	1 Hz	240×192
Malaga 2009 6L [8]	University parking containing cars and trees	3474	7.5 Hz	1024×768

3.1 Datasets

Publicly-available outdoor datasets are chosen with a view to achieve a variety of different data characteristics e.g., vehicle's velocity, camera's frame-rate. Table 1 provides a summary of each image-sequence utilized. In the case of KITTI 05 dataset [18], the incoming visual stream is obtained through a stereo camera system mounted on a car, with high resolution instances depicting houses, cars and trees. Several loop closure events are performed under different platform velocity. The second data-sequence belongs to the Lip 6 Outdoor environment [1]. The visual information is provided through a hand-held camera with low acquisition frequency and resolution, recording an urban environment with many buildings. A high amount of loop closure events are encountered along the navigated path. In Malaga 2009 6L [8], the recorded data offers high resolution images with accurate odometry information from a University parking area. Plenty loop closing examples are accounted from a stereo vision system mounted on an electric buggy-typed vehicle. Since the proposed approach aims to an appearance-based pipeline, only the right visual information was utilized for the system's evaluation.

3.2 Visual Features Selection

The scale-restrictive visual features has to be sufficiently reliable so as to accurately describe the incoming camera measurements. Towards this aim, KITTI 05 is selected as the evaluation dataset since it includes high resolution images with rich texture and considerable frame-rate along the traversed path. For our experiments two of the longest routes in the dataset are used with the SURF [7] detector and descriptor being adopted for keypoint extraction. To evaluate the features' repeatability with respect to keypoint scale, a total of four value ranges cases wherein the scaling value ranges cases was assessed ($\sigma = \{(0,3],(3,6],(6,9],(6,9)\}$). The specific scales are selected experimentally through a quantitative estimation of the extracted visual features. Throughout

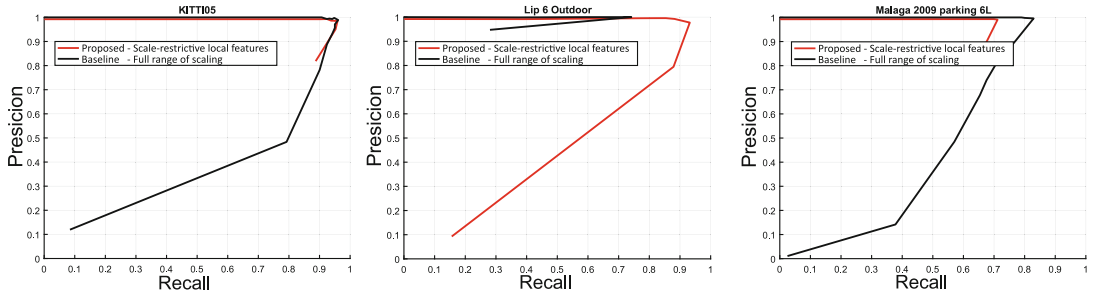

Fig. 3. Precision and recall curves evaluating the utilized scale-restrictive local features. The proposed approach (red lines) performs comparably, reaching high recall rates for perfect precision, as compared to the baseline version (black lines) where extracted keypoints are used arbitrarily. (Color figure online)

the experiments, features which belong to $\sigma = (0, 3]$ showed a dominance in repeatability against the higher ranges, and thus they are selected for the main pipeline.

3.3 Loop Closure Performance Evaluation Protocol

Precision-Recall Metric: In order to evaluate the system's performance against the chosen datasets, the precision-recall metrics are illustrated. Precision is defined as the ratio of the system's correct detected loop closure matches over the total method's identifications:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}. \tag{6}$$

Recall is the ratio between the detected true positive events and the actual loop closures declared through the ground truth:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}. \tag{7}$$

A true positive match is considered to be any database association occurring within a small distance radius of 10 locations from the query image, while a false positive corresponds to any match lies outside this area. On the contrary, a false-negative is considered the incoming image that ought to be matched but the system was unable to achieve any detection. Ground Truth (GT) information is defined as the binary matrix (GT) whose elements correspond to the absence (GT_{ij} = false) or existence (GT_{ij} = true) of a loop closure event. In the cases of KITTI 05 and Malaga 2009 6L datasets, the used GTs were constructed manually within the scope of our previous work [33]. The evaluation of Lip 6 Outdoor is established through the GT data offered by the authors [1].

Method's Evaluation: In order to monitor the system's performance through precision-recall metrics, a variety of binomial probability thresholds θ were

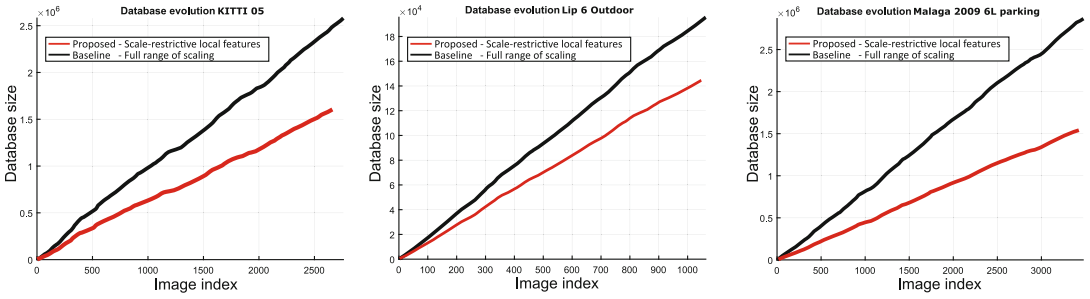

Fig. 4. Database evolution of the proposed pipeline for the scale-restrictive (red line) version, as well as for the unscaled one (black line). The generated visual database is about half in the proposed approach, resulting into a significant data reduction for each assessed dataset. (Color figure online)

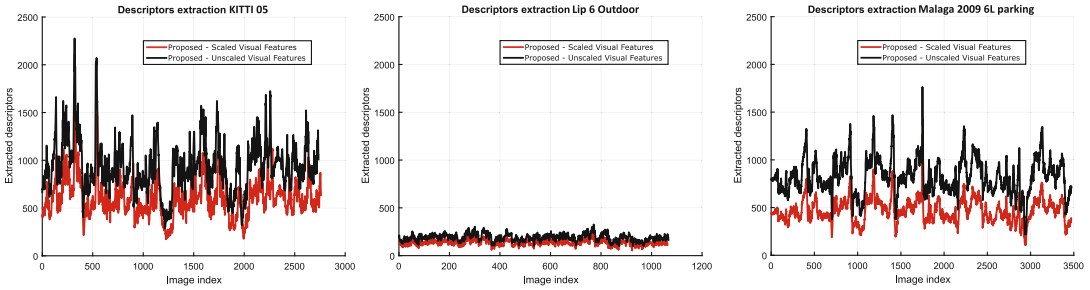

Fig. 5. Number of extracted SURF [7] features per incoming camera measurement. Red lines depict the scale-restrictive approach which achieves great reduction of the utilized keypoints. This effect is mostly highlighted in the case of Malaga 2009 6L parking. (Color figure online)

tested. In Fig. 3, the illustrated scores prove the ability of the proposed scale-restrictive pipeline to perform comparably as compared to the baseline method. In the KITTI 05 dataset, the system reaches a recall rate over 90% for perfect precision in both cases. In case of Lip 6 Outdoor environment the proposed version shows a superiority over the baseline reaching nearly 90% recall, while the precision remains at 100%. This is owed to the fact that the images' acquisition frequency is low and the camera's orientation changes along the navigated path, resulting into a robust database construction of scale-restrictive with distinct sets of votes during the query process. In the Malaga 2009 6L data-sequence, the performance in the baseline version is increased due to the high perception aliasing of the dataset, which allows the system to identify more loop closing events through the more information available.

Descriptors and Database Evolution: Aiming to analyze the computational complexity of the proposed pipeline, in Fig. 4, the system's overall database evolution is illustrated, while Fig. 5 shows the number of features extracted per image. It is noteworthy that for each tested dataset, the number of

Table 2. Comparative results

Dataset	Metric (%)	Proposed restrictive	Baseline	Tsintotas et al. [35]	Gehring et al. [17]
KITTI 05 [18]	Recall	92	91	92.6	**94**
	Precision	100	100	100	**100**
Lip 6 Outdoor [1]	Recall	**73**	67	50	Not available
	Precision	**100**	100	100	
Malaga 6L [8]	Recall	70	75	**85**	Not available
	Precision	100	100	**100**	

scale-restrictive visual features is about half of the baseline version, resulting into a significant reduction of computational complexity.

3.4 Comparative Results

In Table 2, the algorithm's obtained results for each tested dataset are presented. With a view to carry out a fair comparison between the two evaluated versions, as well as against other state-of-the-art methods, the binomial probability threshold was selected as the single value which performs the highest recall for 100% precision. The chosen decision values, $\theta_{restrictive} = 1e^{-11}$ and $\theta_{baseline} = 1e^{-16}$, remain the same for every image-sequence. As it can be seen, the proposed scale-restrictive pipeline can achieve remarkable recall scores for perfect precision in every tested dataset. When comparing the KITTI 05 image-sequence, the proposed version exhibits over 90% recall score performing comparably to the rest of the approaches. In Lip 6 Outdoor sequence, high recall rates are achieved, while in Malaga 6L the proposed framework performs unfavorable against the other methods. This in mainly due to low-texture environment that the robot encounters, as well as the perceptual aliasing occurring, which prevents the scale-restrictive features to reach high recall scores.

4 Conclusion

The paper in hand presented a scale-restrictive visual loop closure detection framework. Through an evaluation procedure, local features are assessed for their repeatability with respect to their scale. The mechanism represents each visited location through the scale-specific local features extracted via SURF detector in order to reduce the computational cost. At query time, a probability score is generated for every instance in the database, based on the votes' aggregation which are collected via a k-NN technique. The system's loop closure belief generator is based on a probabilistic framework, while a geometrical check is also adopted in order to reduce possible false detections. The proposed method is tested on several environments demonstrating a substantial performance as compared to

Fig. 6. Different local features extracted for the proposed method (top) and its baseline version (bottom) on KITTI 05 [18] (left), Lip 6 Outdoor [1] (center) and Malaga 2009 Parking 6L [8] (right) datasets.

its baseline version, as well as to other state-of-the-art approaches, offering high recall rates for perfect precision. Future work will focus on a more extensive evaluation and the utilization of the same informative scales for the generation of an on-line incremental vocabulary. Examples of extracted scale-restrictive visual features are illustrated in Fig. 6.

References

1. Angeli, A., Filliat, D., Doncieux, S., Meyer, J.A.: A fast and incremental method for loop-closure detection using bags of visual words. IEEE Trans. Robot. 1027–1037 (2008)
2. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J.: NetVLAD: CNN architecture for weakly supervised place recognition. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 5297–5307 (2016)
3. Baeza-Yates, R., Ribeiro-Neto, B., et al.: Modern Information Retrieval, vol. 463. ACM Press, New York (1999)
4. Balaska, V., Bampis, L., Gasteratos, A.: Graph-based semantic segmentation. In: Proceedings of International Conference on Robotics in Alpe-Adria Danube Region, pp. 572–579 (2018)
5. Bampis, L., Amanatiadis, A., Gasteratos, A.: Encoding the description of image sequences: a two-layered pipeline for loop closure detection. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4530–4536 (2016)
6. Bampis, L., Amanatiadis, A., Gasteratos, A.: Fast loop-closure detection using visual-word-vectors from image sequences. Int. J. Robot. Res. **37**(1), 62–82 (2018)

7. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: speeded-up robust features. In: Proceedings of European Conference on Computer Vision, pp. 404–417 (2006)
8. Blanco, J.L., Moreno, F.A., Gonzalez, J.: A collection of outdoor robotic datasets with centimeter-accuracy ground truth. Auton. Robots **27**(4), 327 (2009)
9. Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. IEEE Trans. Inf. Theory **14**(3), 462–467 (1968)
10. Cieslewski, T., Stumm, E., Gawel, A., Bosse, M., Lynen, S., Siegwart, R.: Point cloud descriptors for place recognition using sparse visual information. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 4830–4836 (2016)
11. Cummins, M., Newman, P.: Appearance-only SLAM at large scale with FAB-MAP 2.0. Int. J. Robot. Res. **30**(9), 1100–1123 (2011)
12. Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: Part I. IEEE Robot. Autom. Mag. **13**(2), 99–110 (2006)
13. Erkent, Ö., Bozma, H.I.: Bubble space and place representation in topological maps. Int. J. Robot. Res. **32**(6), 672–689 (2013)
14. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
15. Gálvez-López, D., Tardos, J.D.: Bags of binary words for fast place recognition in image sequences. IEEE Trans. Robot. **28**(5), 1188–1197 (2012)
16. Garcia-Fidalgo, E., Ortiz, A.: iBoW-LCD: an appearance-based loop-closure detection approach using incremental bags of binary words. IEEE Robot. Autom. Lett. **3**(4), 3051–3057 (2018)
17. Gehrig, M., Stumm, E., Hinzmann, T., Siegwart, R.: Visual place recognition with probabilistic voting. In: Proceedings of IEEE International Conference on Robotics and Automation, Singapore, pp. 3192–3199, May 2017
18. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Proceedings of Conference on Computer Vision and Pattern Recognition (2012)
19. Khan, S., Wollherr, D.: IBuILD: incremental bag of binary words for appearance based loop closure detection. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 5441–5447 (2015)
20. Kostavelis, I., Gasteratos, A.: Semantic mapping for mobile robotics tasks: a survey. Robot. Auton. Syst. **66**, 86–103 (2015)
21. Leutenegger, S., Chli, M., Siegwart, R.: BRISK: binary robust invariant scalable keypoints. In: Proceedings of IEEE International Conference on Computer Vision, pp. 2548–2555 (2011)
22. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
23. Lynen, S., Bosse, M., Furgale, P.T., Siegwart, R.: Placeless place-recognition. In: Proceedings of IEEE International Conference on 3D Vision, pp. 303–310 (2014)
24. Maffra, F., Chen, Z., Chli, M.: Tolerant place recognition combining 2D and 3D information for UAV navigation. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 2542–2549 (2018)
25. Milford, M.J., Wyeth, G.F.: SeqSLAM: visual route-based navigation for sunny summer days and stormy winter nights. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 1643–1649 (2012)
26. Mur-Artal, R., Tardós, J.D.: Fast relocalisation and loop closing in keyframe-based SLAM. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 846–853 (2014)

27. Newman, P., Cole, D., Ho, K.: Outdoor SLAM using visual appearance and laser ranging. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 1180–1187 (2006)
28. Radenović, F., Tolias, G., Chum, O.: CNN image retrieval learns from BoW: unsupervised fine-tuning with hard examples. In: Proceedings of European Conference on Computer Vision, pp. 3–20 (2016)
29. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Proceedings of IEEE International Conference on Computer Vision, pp. 2564–2571, November 2011
30. Sivic, J., Zisserman, A.: Video Google: a text retrieval approach to object matching in videos, p. 1470 (2003)
31. Sünderhauf, N., Dayoub, F., Shirazi, S., Upcroft, B., Milford, M.: On the performance of convnet features for place recognition. arXiv preprint arXiv:1501.04158 (2015)
32. Thrun, S., Leonard, J.J.: Simultaneous localization and mapping. In: Siciliano, B., Khatib, O. (eds.) Springer Handbook of Robotics, pp. 871–889. Springer, Heidelberg (2008)
33. Tsintotas, K.A., Bampis, L., Gasteratos, A.: Assigning visual words to places for loop closure detection. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 1–7 (2018)
34. Tsintotas, K.A., Bampis, L., Gasteratos, A.: DOSeqSLAM: dynamic on-line sequence based loop closure detection algorithm for SLAM. In: Proceedings of IEEE International Conference on Imaging Systems and Techniques, pp. 1–6 (2018)
35. Tsintotas, K.A., Bampis, L., Gasteratos, A.: Probabilistic appearance-based place recognition through bag of tracked words. IEEE Robot. Autom. Lett. **4**(2), 1737–1744 (2019)
36. Tsintotas, K.A., Bampis, L., Rallis, S., Gasteratos, A.: SeqSLAM with bag of visual words for appearance based loop closure detection. In: Proceedings of International Conference on Robotics in Alpe-Adria Danube Region, pp. 580–587 (2018)
37. Zhang, G., Lilly, M.J., Vela, P.A.: Learning binary features online from motion dynamics for incremental loop-closure detection and place recognition. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 765–772. IEEE (2016)
38. Zhang, H.: BoRF: loop-closure detection with scale invariant visual features. In: Proceedings of IEEE International Conference on Robotics and Automation, pp. 3125–3130 (2011)