This full text paper was peer-reviewed at the direction of IEEE Instrumentation and Measurement Society prior to the acceptance and publication.

DOSeqSLAM: Dynamic On-line Sequence Based Loop Closure Detection Algorithm for SLAM

Konstantinos A. Tsintotas, Loukas Bampis and Antonios Gasteratos

***Abstract*— Simultaneous Localization and Mapping (SLAM) is vital for modern autonomous robots. Visual place recognition of pre-visited areas, widely known as Loop Closure Detection (LCD), constitutes one of the most important SLAM components. In this paper a sequence-based LCD algorithm is proposed by evolving SeqSLAM method. Instead of using fixed-size sequences' length during search process, as in the original approach, we suggest a dynamical length definition based on the images' content proximity. Specifically, sequence segmentation is achieved through a feature matching technique applied on the incoming visual sensory information, while the mechanism operates on-line without the need of any pre-training procedure. Each sequence's similarity score, produced by a weighted average function, is utilized as a decision factor for the loop closing selection. Finally, an extended image-to-image search in the chosen group of images, avails the system of identifying the appropriate match. A temporal constraint prevents early pre-visited areas to be presented as false-positive matches. The method is evaluated on several outdoor publicly-available datasets, revealing a substantial improvement on SeqSLAM.**

I. INTRODUCTION

Within the scope of visual Simultaneous Localization and Mapping (SLAM) methods [1], [2] the robot should estimate its pose as it navigates through the working field. The importance of an efficient and robust estimation is vital as it allows for an accurate navigation. The ability to detect and recognize a place which has previously been observed, is referred to as Loop Closure Detection (LCD) [3], [4]. Due to noisy sensor measures or field abnormalities, drifts occur on the robot's map. Decreasing such cases through accurate LCDs provides an improved pose estimation. The detection of numerous fault-free loop closure events constitutes the goal of modern autonomous systems.

Owed to the increased availability of computational power during the last years, cameras have become the primary sensor unit in most of the robotic mechanisms. The richness of the acquired information guided the researchers in robotics community to produce appearance-based systems and specifically, methods which rely their functionality solely on the incoming visual information [5]. Appearance-based systems can be distinguished into two main categories, namely global and local ones. The global appearance-based systems, seek for the most similar image in the database in an exhaustive manner. Comparisons are performned using techniques such as Sum of Absolute Differences (SAD), Bag of Words (BoW) [6] histograms or features' vote density [7]. On the other hand, local appearance-based systems search for the best image match into a group of neighbooring images and utilize similar comparison techniques. SeqSLAM [8], one of the most renown algorithms in sequence-based visual place recognition [9], [10], [11], [12], [13], tackles the recognition task in an off-line procedure by comparing the query sequence against the database groups-of-images for the appropriate sequence match selection. However, the majority of the aforementioned approaches segments the incoming image stream into groups, in an arbitrary manner based on a predefined temporal threshold. As a result, the instances of the same group might not always exhibit the same semantics, in which case they may burden the LCD mechanism, or at least they do not facilitate it.

Having identified this drawback, in this paper, we present an online LCD method which improves SeqSLAM, being capable of providing more accurate detections in the robot's traversed path. During the on-line aquisition of the camera data a pre-processing stage is met, where images are converted to greyscale and downsampled. Comparisons with the database are performed via SAD technique. The proposed comparison method is generic and other techniques can also be used, viz. BoW histograms or CNN derived feature vectors comparisons [15], [16]. However, the original metric is preferred in accordance to the initial version of SeqSLAM.

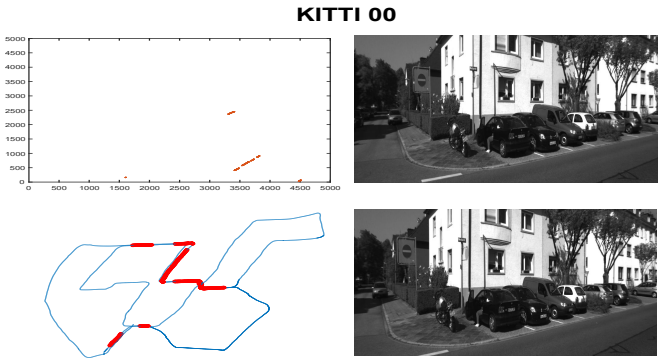


Fig. 1: Loop closure detections produced by the proposed technique on the KITTI 00 [14] dataset. Confident selection of the apropriate loop closing sequence candidate is achieved through a weighted average similarity score. An extended search in the chosen sequence's images is performed to identify the best match (right). The red stripes show the detected loop closures illustrated on the images' similarity matrix (left-top), while the red cycles indicate the corresponding detections in the robot's navigated path (left-down).

Authors are with the Department of Production and Management Engineering, Democritus University of Thrace, 12 Vas. Sophias, GR-671 32, Xanthi, Greece ktsintot@pme.duth.gr, lbampis@pme.duth.gr, agaster@pme.duth.gr

978-1-5386-6628-9/18/$31.00 ©2018 IEEE

A local feature detection and description method is applied to the incoming visual stream as well and through feature matching dynamic sequences are defined on the traversed route, resulting in camera measurements segmentation according to their visual information. During this procedure, we are interested on the local features' coherence rather than the exact semantics of the environment. The appropriate candidate sequence is highlighted thanks to the similarity score produced by a weighted average function. Furthermore, the most suitable loop closing instance is located into the selected sequence via an image-to-image search at the SAD measurements domain (Fig. 1).

The main contributions offered by the proposed paper are summarized as follows:

- An adaptive version of SeqSLAM algorithm, used for LCD, with robust results, dubbed as DOSeqSLAM. The approach relies on the efficient dynamic segmentation of the incoming image stream.
- A mechanism which exceeds the fixed-size sequence length, providing distinct dynamic areas that decrease the computational complexity at query time.
- An LCD framework able to recognize pre-visited places in an on-line scheme.

The rest of the paper is articulated as follows: Section II provides a brief discussion on LCD methods. In Section III the DOSeqSLAM is described in detail, while in Section IV the method's evaluation and comparative results are presented. Finally, conclusions are drawn in Section V.

II. RELATED WORK

A widely successful global appearance-based method depending on the BoW model is FAB-MAP 2.0 [17]. The incoming visual stream is represented by a set of Visual Words (VWs) belonging to a Visual Vocabulary (VV) constructed off-line. At query time, the most recent image in the pipeline is submitted to the database and loop closing candidates are detected through a Bayes filter. Sequence-based algorithms, such as [10] and [18], tackle the LCD task by grouping the images' extracted local features or VWs. Comparisons are performed between the constructed sequences. In [10], a subset of database candidate sequences is provided via a Bayes filter. Evaluation on the selected areas proved a faster implementation of SeqSLAM. Similarly, the authors in [18] group the camera measurements into fixed-size sequences of images represented by VWs. In the course of query sequences' VW histograms are compared, while temporal consistency checks are implemented for performance enhancement. In addition, the work of Wang et al. [19] improves the SeqSLAM providing a real-time algorithm by using a sliding window visual-inertial odometer. Utilizing odometry information along the incoming camera data, sequence matching is performed. Lastly, a multiscale search at the chosen sequence provides the correct loop closing image. Furthermore, Siam and Zhang [20] proposed the formulation of a more efficient version of SeqSLAM, titled as "Fast-SeqSLAM". The description of downsampled images is implemented by histogram of gradients [21], while a k-d tree [22] is constructed along to the robot's path. An approximative nearest neighbor technique is utilized on the dimensionally reduced descriptor space for an optimum sequence match, while a greedy image-to-image search is performed to achieve the final match.

III. METHODOLOGY

In this section, the DOSeqSLAM algorithm is presented in detail. A brief description of the initial SeqSLAM approach is provided as well. Since the main algorithm modification refers to the sequences' dynamical definition, in the following subsections the transition from the original approach to the proposed version is described.

A. SeqSLAM

As the incoming image stream is acquired by the system, the data are converted to grayscale and then downsampled to a vector of dimensionality χ, which equals the total multitude of image pixels. Each downsampled image is normalized per pixel in an N size neighboring area and subsequently the images are compared. Each normalized image i (I_i) is compared to all the images in the database by means of SAD:

$$D_{ij} = \frac{1}{R_x R_y} \sum_{x=0}^{R_x} \sum_{y=0}^{R_y} |\rho_{x,y}^i - \rho_{x,y}^j| \tag{1}$$

R_x and R_y represent the reduced dimensions of the images, while ρ denotes the intensity value of each pixel. A difference vector D_i against each database image j (I_j) is produced. Since the aim of SeqSLAM focuses on searching neighboring images, instead of single ones, a contrast enhancement process is performed on every element of D_i, which is proportional to 1D patch-normalization in a local area of ε pixels:

$$\hat{D}_{i,\mu} = \frac{D_{i,\mu} - \overline{D_\varepsilon}}{\sigma_\varepsilon} \tag{2}$$

where $\overline{D_\varepsilon}$ denotes the local mean and σ_ε the local standard deviation around element μ.

The recognition process is established on the produced difference-enhanced matrix $\hat{D}$. At query time, a number of trajectories is projected on $\hat{D}$ for each database image j. The trajectories' length depends on SeqSLAM pre-defined sequence size. Each trajectory represents a possible velocity assumption corresponding to different vehicle's velocities V on the navigated path. For each trajectory, a score s is calculated:

$$s = \sum_{z=i-w}^{i} \hat{D}_{z,k} \tag{3}$$

where i represents the query index and w is the sequence's length. The different velocity assumptions k are computed for each $\hat{D}_{i,j}$ as:

$$k = j + V(w - i + t) \tag{4}$$

V is assigned with multiple values within the range of $[V_{min}, V_{max}]$ (advancing by V_{step} each time step t). The minimum score s is selected for the particular database image

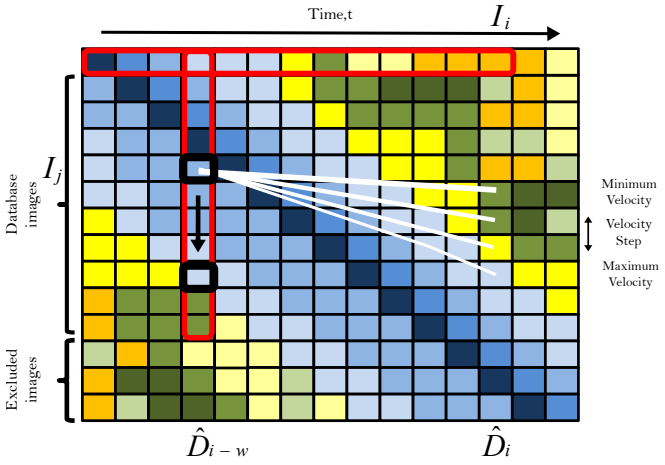


Fig. 2: The searching procedure of SeqSLAM [8] algorithm. In the course of query image (I_i), a sliding window technique scans the database for the proper image match by searching into groups of images. The white stripes indicate the proposed trajectories which are evaluated for the database image I_j. The trajectory with the minimum accumulative distance is propagated to the query's i score vector S. The horizontal red rectangle illustrates the query's evolution process, while the vertical one indicates the sliding scan of the database.

I_j. The possible trajectory assumptions for image pair $I_{i,j}$ are illustrated in Fig. 2.

Finally, the computed scores s of each database image I_j for query I_i are arranged into a vector format S_i [$s_{i,j}$, $s_{i,j+1}$, $s_{i,j+2}$, ... $s_{i,j+n}$]. The minimum distance is located and selected. Subsequently the chosen score is normalized with respect to the second lowest value, outside the corresponding w range. In order for an image pair to be considered as loop closure a decision factor θ has to be met.

B. From constant to dynamic sequence length

While the original approach of SeqSLAM utilizes a fixed-size sliding window scheme to examine the database images,

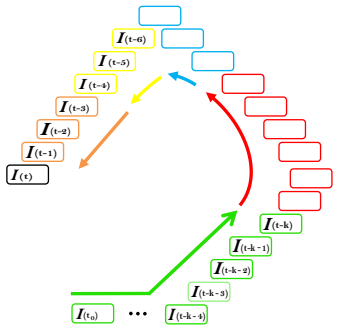


Fig. 3: As the incoming image $I_{(t)}$ (black outline) enters the pipeline, the query procedure starts after a time constant κ is satisfied. Sequences with orange, yellow and blue outline participate in the query process. The search area is determined by the green outline, while the red ones are the excluded images.

DOSeqSLAM focuses on searching in distinct dynamically defined group of instances. Towards this goal, the incoming image stream is converted to greyscale and firstly passes through a local feature extraction module, while the preprocessing stage, mentioned in Section III-A, follows. SURF [23] method is used for feature detection and description, whilst a descriptor database is temporary retained along with an inverted index until the sequence's construction. Dynamic sequences are determined through a feature matching coherence check. Particularly, at time t, the incoming image stream $I(t-n), ..., I(t-2), I(t-1), I(t)$ is segmented once the correlation between the last n image's descriptors cease to exist:

$$\left| \bigcap_{i=0}^{i=n} d_{I(t-i)} \right| \leq 1 \tag{5}$$

where d_I denotes the image's descriptors and $|\mathbb{X}|$ the cardinality of set $\mathbb{X}$.

C. DOSeqSLAM

Following the extraction of SURF local features from the input camera data, SeqSLAM is applied. For each image $I_{(t)}$ comparisons are performed against the database using the SAD metric (eq. 1). Finally, contrast enhancement (eq. 2) is applied on the difference matrix (D) entries. The LCD mechanism is not activated unless a time constant κ is satisfied. This factor is used to protect the system from early observed detections which are highly probable to occur when two images are captured under low time-proximity. Such cases can reduce the system's performance by producing false-positive identifications. The searching area spans between the first acquired image $I_{(t_0)}$ and the κ-th one $I_{(t_0-\kappa)}$, as depicted in Fig. 3.

At time t, when the latest sequence Seq_N is created, the query process is implemented for the previously one constructed Seq_{N-1}. Since DOSeqSLAM avoids the sliding window scheme of the initial approach, the possible trajectory assumptions' scores s_{do} are calculated only for the first entry $I_1^{seq_Q}$ of the query sequence Seq_{N-1}, resulting to a reduced computational cost. Thus, every instance $[I_1^{seq_Q} ... I_{end}^{seq_Q}]$ belonging to Seq_{N-1} gets the same trajectory score s_{do} per database image I_j. An illustration of DOSeqSLAM's functionality is shown in Fig. 4.

Similarly to SeqSLAM, trajectories' assumptions depend on the velocity parameters ($V_{max}, V_{min}, V_{step}$). Since the proposed method adopts a dynamic window scheme, each score s_{do} is adjusted to the average of the accumulated values for each trajectory path:

$$s_{do} = \frac{1}{seq_{Len}^Q} \sum_{I_1^Q}^{I_{end}^Q} \hat{D}_k \tag{6}$$

where I_1^Q and I_{end}^Q are the first and last image timestamps of the query sequence, respectively; seq_{Len}^Q is the query's length and k denotes the velocity assumption paths (eq. 4).

Computing the s_{do} score for every database image results into the vector S_{do}. The minimum value of S_{do} is selected

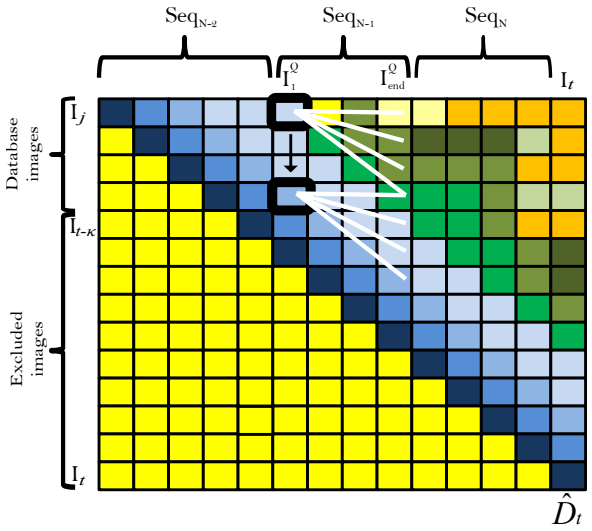


Fig. 4: The proposed DOSeqSLAM algorithm. As image I_t enters the pipeline (concluding sequence Seq_N) a searching procedure is performed for sequence Seq_{N-1}. Each instance $[I_1^{seq_Q}...I_{end}^{seq_Q}]$ belonging on Seq_{N-1} gets the same trajectory score s for database image I_j. The system scans the database till the instance I_{t-k}, which is determined by the time constant κ. Yellow bins represent the images which have not been observed by the system since the mechanism performs on-line. The white stripes indicate the possible trajectories' paths for different vehicle's velocities.

for further processing. Since the original approach searches for the second smallest value outside w, the proposed method defines the W_{DOSeq} range using the lengths of Seq_{N-2} and Seq_N (Fig. 5):

$$W_{DOSeq} = Seq_{N-2} * 2 + Seq_N * 2 \tag{7}$$

The minimum score is then normalized over the second lowest value resulting to γ. The final decision depends on a weighted average function:

$$\mathbf{f}(\gamma) = \frac{1}{6}\gamma_{(Seq_{N-2})} + \frac{4}{6}\gamma_{(Seq_{N-1})} + \frac{1}{6}\gamma_{(Seq_N)} \tag{8}$$

Through the satisfaction of factor th a loop closure event is determined when:

$$\mathbf{f}(\gamma) < th \tag{9}$$

Finally, an image-to-image association is performed by a greedy search in the enhanced difference sub-matrix $[\hat{D}_{I_1^Q,I_j}...\hat{D}_{I_{end}^Q,I_{j+Seq^Q_{Len}}}]$ for the minimum distance of each image in query sequence.

IV. Experimental Evaluation

The proposed DOSeqSLAM algorithm is evaluated on several outdoor publicly-available datasets, summarized in Table I, and compared against the initial approach. A brief description of the experimental procedure also follows.

A. Procedure

1) Datasets: Four different datasets are selected representing dynamic outdoor environments with many loop closure events in their trajectory. Moreover, a variety in frame rate is achieved. The incoming visual data are obtained by stereo camera systems in most datasets, nevertheless for the purposes of our method only the right input stream is used. KITTI 00, KITTI 05 [14] and Malaga 2009 Parking6L [24] (Mlg6L) are recorded from cameras mounted on a moving car, while in Lip6 Outdoor [25] (Lip6O) the incoming data is retrieved from a hand-held sensor. KITTI 05 and Mlg6L are preferred as evaluation datasets since the recorded data provide considerable loop closure events, while KITTI 00 and Lip6O are chosen as the testing ones.

TABLE I: Datasets' description

Dataset	Description	FPS
KITTI 00 [14]	Outdoor, dynamic	10
KITTI 05 [14]	Outdoor, dynamic	10
Malaga 2009 Parking 6L [24]	Outdoor, slightly dynamic	7
Lip6 Outdoor [25]	Outdoor, slightly dynamic	1

2) Ground Truth: As Ground Truth (GT) is defined the information data which indicate the actual loop closure events occuring in a dataset. GT is structured as a binary matrix with image-to-image correspondences, where the ones ($GT_{ij} = 1$) denote the existance of a loop closure event. For KITTI 00, KITTI 05 and Mlg6L, the used GT were constructed manualy in our previous work [26]. For Lip6O, GT is available by the respective authors.

3) Precision-Recall: Comparisons between algorithms is achieved through the precision and recall metric. As precision is defined the ratio between true-positive identifications (image-to-image) and the total system's detections. Recall is defined by the number of detected loop closure events over the actual events appeared in the GT.

B. Evaluation

The DOSeqSLAM's parameters are shown in Table II. Most of these parameters are defined similarly to the initial

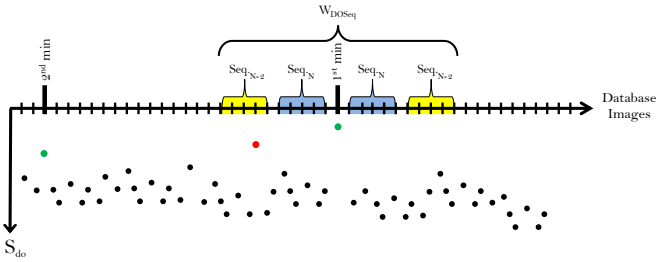


Fig. 5: The proposed algorithm searches for the second lowest value outside of a window range W_{DOSeq}. The yellow and blue areas indicate sequences' Seq_{N-2} and Seq_N lengths, respectively. The dots are the produced scores for each database image. The green ones illustrate the minimum scores, while the red dot is rejected due to its position within the window range.

TABLE II: Parameters

Downsampled image pixels, χ	:	2048
Pixel normalization side length, N	:	8
Reduced image side length , R_x, R_y	:	32, 64
Contrast enhancement local area, ε	:	10
Time constant, κ	:	40s
Maximum trajectory velocity, Vel_{max}	:	1.3
Minimum trajectory velocity, Vel_{min}	:	0.7
Trajectory velocity step, Vel_{step}	:	1
DOSeqSLAM decision factor th	:	0.75

version, i.e. downsampled image size, normalization area and contrast enhancement neighborhood length. The rest of them (SeqSLAM's sequence length w and velocity's properties $[V_{max}, V_{min}, V_{step}]$) rely on the implementation of OpenSeqSLAM[1].

Among the selected testing cases, KITTI 05 is the one exhibited the largest time periods in which the camera remains still. Based on this dataset, the value of κ was selected as $40s$, though, different operating conditions may result into different conclusions without affecting the system's accuracy.

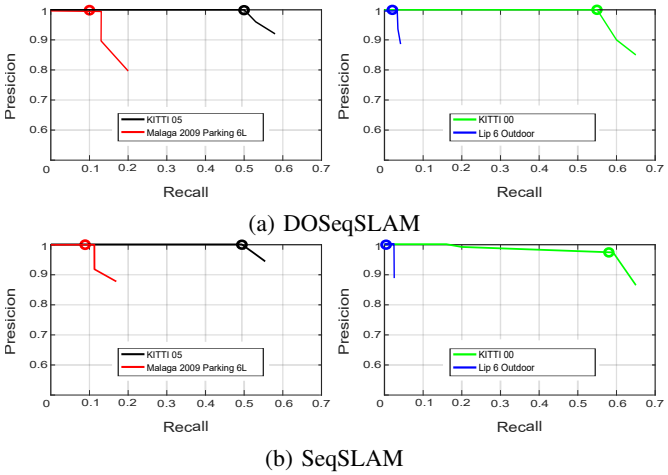
An evaluation mechanism is constructed to measure the effect of the system's parameters on the achieved performance. A variety of different th are selected so as to produce the precision-recall curves. The estimation of proper th parameter is accomplished through KITTI 05 and Mlg6L (Fig.6a-left). In order to provide a robust and efficient system without the existence of any false-positive LCD the decision factor is selected to $th = 0.75$. Using the same set of parameters we assess the impact of our method on the two standalone testing datasets. The achieved performance is shown in Fiq. 6a-right. We observe that DOSeqSLAM executes high recall rates for 100% precision.

With a view to carry out a fair comparison between the two algorithms, a similar procedure is followed for SeqSLAM. Since the method has not been tested at the aforementioned environments the precision-recall curves are constructed on the evaluation datasets for the estimation of θ (Fig.6b-left). Subsequently, the algorithm was tested on KITTI 00 and Lip6O, as illustrated in Fig.6b-right.

C. Comparative Results

Table III shows the obtained results for every tested dataset in a similar manner to [25], [27], where the biggest number of true-positive detections are counted, with the fewest possible false-positive events. In addition to this metric, precision and recall have also been considered. It can be seen that the proposed DOSeqSLAM method is capable of detecting more true-positive loop closure events with 100% precision in comparison to the initial version. In KITTI 05 and Mlg6L, our algorithm performs favorably, while in KITTI 00 the LCDs are higher in SeqSLAM, nevertheless its precision is reduced to 97%. In Lip6O where SeqSLAM totally fails, the proposed algorithm exhibits good performance improvement.

[1]The OpenSeqSLAM algorithm can be found online in : https://openslam.org/openseqslam.html



(a) DOSeqSLAM

(b) SeqSLAM

Fig. 6: Precision-recall curves of (a) the proposed algorithm and (b) the original approach. On the left side, the produced curves from the evaluation datasets KITTI 05 [14] and Mlg6L [24] were used for the estimation of th and θ. On the right side, the produced curves illustrate each system's performance on the tested environments, KITTI 00 [14] and Lip6O [25]. Color markers (cycles) on the top of each graph indicate the achieved recall rate using the chosen parameters.

V. CONCLUSIONS

In this paper, a modified version of the widely known SeqSLAM algorithm is presented for LCD, dubbed as DOSeqSLAM. The proposed method dynamically segments the incoming visual sensory information for sequences' definition through a local feature matching technique. Based on the query's sequence size, the system seeks into the database for the most appropriate match, while the final decision about a loop closure event is obtained via a weighted average function. By converting the fixed-size sliding window search of SeqSLAM into a dynamic one, we construct a system capable of performing on-line with lower computational cost. The proposed method is evaluated on several environments demonstrating an improvement against the original. Examples of LCDs are illustrated in Fig 7.

REFERENCES

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous Localization and Mapping: part I," *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[2] S. Thrun and J. J. Leonard, "Simultaneous Localization and Mapping," in *Springer Handbook of Robotics*, 2008, pp. 871–889.

[3] K. L. Ho and P. Newman, "Loop closure detection in SLAM by combining visual and spatial appearance," *Robotics and Autonomous Systems*, vol. 54, no. 9, pp. 740–749, 2006.

[4] H. Zhang, "BoRF: Loop-Closure Detection with Scale Invariant Visual Features," in *Proc. IEEE International Conference on Robotics and Automation*, 2011, pp. 3125–3130.

[5] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual Place Recognition: A survey," *IEEE Transactions on Robotics*, vol. 32, no. 1, pp. 1–19, 2016.

[6] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos," in *Proc. International Confernece on Computer Vision*, 2003, p. 1470.

TABLE III: Comparative results

Dataset	Approach	True Positives	False Positives	Ground Truth Loop Closure Events	Precision (%)	Recall (%)
KITTI 00 [14]	SeqSLAM [8]	424	11	733	97	58
	DOSeqSLAM	**406**	**0**	**733**	**100**	**55**
KITTI 05 [14]	SeqSLAM [8]	186	0	379	100	49
	DOSeqSLAM	**188**	**0**	**379**	**100**	**50**
Malaga 2009 Parking 6L [24]	SeqSLAM [8]	32	0	382	100	8
	DOSeqSLAM	**38**	**0**	**382**	**100**	**10**
Lip6 Outdoor [25]	SeqSLAM [8]	0	0	301	100	0
	DOSeqSLAM	**5**	**0**	**301**	**100**	**2**

Fig. 7: Different loop closure examples detected by DOSeqSLAM, on KITTI 05 [14] (left), Lip6 Outdoor [25] (center) and Malaga 2009 Parking 6L [24] (right) datasets.

[7] K. Grauman and T. Darrell, "Efficient image matching with distributions of local invariant features," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition,*.

[8] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *Proc. IEEE International Conference on Robotics and Automation*, 2012, pp. 1643–1649.

[9] M. Milford, H. Kim, M. Mangan, S. Leutenegger, T. Stone, B. Webb, and A. Davison, "Place Recognition with Event-based Cameras and a Neural Implementation of SeqSLAM," *arXiv preprint arXiv:1505.04548*, 2015.

[10] Y. Liu and H. Zhang, "Towards Improving the Efficiency of Sequence-Based SLAM," in *Proc. IEEE International Conference on Mechatronics and Automation*, 2013, pp. 1261–1266.

[11] N. Sünderhauf, P. Neubert, and P. Protzel, "Predicting the change - a step towards life-long operation in everyday environments," *Robotics Challenges and Vision*, p. 17, 2014.

[12] P. Neubert, N. Sunderhauf, and P. Protzel, "Appearance change prediction for long-term navigation across seasons," in *Proc. IEEE European Conference on Mobile Robots*, 2013, pp. 198–203.

[13] K. A. Tsintotas, L. Bampis, S. Rallis, and A. Gasteratos, " SeqSLAM with Bag of Visual Words for Appearance Based Loop Closure Detection," in *Proc. The 27th International Conference on Robotics in Alpe-Adria-Danube Region*, 2018.

[14] J. Fritsch, T. Kuehnl, and A. Geiger, "A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms," in *Proc. IEEE International Conference on Intelligent Transportation Systems*, 2013, pp. 1693–1700.

[15] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of ConvNet features for place recognition," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2015, pp. 4297–4304.

[16] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, and E. Romera, "Fusion and binarization of CNN features for robust topological localization across seasons," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016, pp. 4656–4663.

[17] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *The International Journal of Robotics Research*, vol. 30, no. 9, pp. 1100–1123, 2011.

[18] L. Bampis, A. Amanatiadis, and A. Gasteratos, "Encoding the Description of Image Sequences: A Two-Layered Pipeline for Loop Closure Detection," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016, pp. 4530–4536.

[19] Y. Wang, X. Hu, J. Lian, L. Zhang, and X. Kong, "Improved SeqSLAM for Real-Time Place Recognition and Navigation Error Correction," in *Proc. IEEE Conference on Intelligent Human-Machine Systems and Cybernetics*, vol. 1, 2015, pp. 260–264.

[20] S. M. Siam and H. Zhang, "Fast-SeqSLAM: A Fast Appearance Based Place Recognition Algorithm," in *Proc. IEEE International Conference on Robotics and Automation*, 2017, pp. 5702–5708.

[21] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[22] C. Silpa-Anan and R. Hartley, "Optimised KD-trees for fast image descriptor matching," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition on*, 2008, pp. 1–8.

[23] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proc. European Conference on Computer Vision*, 2006, pp. 404–417.

[24] J.-L. Blanco, F.-A. Moreno, and J. Gonzalez, "A collection of outdoor robotic datasets with centimeter-accuracy ground truth," *Autonomous Robots*, vol. 27, no. 4, pp. 327–351, 2009.

[25] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1027–1037, 2008.

[26] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "Assigning Visual Words to Places for Loop Closure Detection," in *Proc. International Confernece on Robotics and Automation*, 2018, pp. 5979–5985.

[27] S. Khan and D. Wollherr, "IBuILD: Incremental Bag of Binary Words for Appearance Based Loop Closure Detection," in *Proc. IEEE International Conference on Robotics and Automation*, 2015, pp. 5441–5447.