# Dimensionality reduction through visual data resampling for low-storage loop-closure detection

Konstantinos A. Tsintotas[1], Shan An[2], Ioannis Tsampikos Papapetros[1], Fotios K. Konstantinidis[1], Georgios Ch. Sirakoulis[3], and Antonios Gasteratos[1]

2022 IEEE International Conference on Imaging Systems and Techniques (IST) | 978-1-6654-8102-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/IST55454.2022.9827748

***Abstract*— As loop-closure detection plays a fundamental role in any simultaneous localization and mapping (SLAM) system, through its ability to recognize previously visited locations, one of its main objectives is to permit consistent map generation for an extended period. Within large-scale SLAM autonomy, the scalability in terms of timing needed for database search and the storage requirements has to be addressed. In this paper, a low-storage visual loop-closure detection technique is proposed. Our system is based on the incremental bag-of-tracked-words scheme for the trajectory mapping still, the generated visual representations are reduced to lower dimensions through a resampling process. This way, we achieve to shorten the overall database size and searching time, while at the same time preserving the high performance. The evaluation, which took place on different well-known datasets, exhibits the system's low-storage requirements and high recall scores compared to the baseline version and other state-of-the-art approaches.**

## I. Introduction and related work

As an autonomous robot navigates in an unknown environment, it constructs an internal map based on its sensors' incoming measurements [1]. This process is widely known as simultaneous localization and mapping (SLAM) and has evolved as the most precious asset for autonomous navigation over the last three decades [2]. Nevertheless, the sensors' measurements that are subject to noise as well as the absence of global positioning measurements cause an accumulated drift to the robot's estimated pose (position and orientation). Due to this fact, identifying previously visited locations, *i.e.,* loop-closures [3], along the robot's path is necessary as a rectification on the internal map is performed and a consistent map is built.

Nowadays, cameras have become the central perception unit [4]–[7], for modern autonomous systems, compared to the early years, when range and bearing sensors were adopted [8], [9]. This shift is owed to the camera's ability to provide rich textural information while at the same time is a cheap sensor that can be adopted by platforms with limited computational capabilities, such as unmanned aerial vehicles [10]–[13] and space exploration rovers [14]–[18], with high efficiency.

[1]Authors are with the Department of Production and Management Engineering, Democritus University of Thrace, 12 Vas. Sophias, GR-671 32, Xanthi, Greece {ktsintot, ipapapet, fokonsta, agaster}@pme.duth.gr

[2]Shan An is with the with AR/VR department, JD.com, Beijing, 100191, China, anshan@jd.com

[3]Georgios Ch. Sirakoulis is with the Department of Electrical and Computer Engineering, Democritus University of Thrace, Kimmeria. GR-67 100, Xanthi, Greece, gsirak@ee.duth.gr

A system's architecture for recognizing previously seen areas and performing visual loop-closure detection includes the *feature extraction* process, the *environment representation*, and the *decision-making* module [19]. The former two are related to the image processing needed to represent the incoming sensory data and formulate the robot's internal map, *i.e.,* the database, while the latter measures the system's confidence about a loop event.

As one of the main objectives of any loop-closure detection pipeline is to facilitate robust navigation for an extended period, scalability in terms of storage requirements and database search timing are issues that every SLAM system must address [20]. Both constitute demanding tasks from the aspect of the system's computational requirements and achieved performance because, in dense maps, in which every incoming image is considered a node in the topological graph, the generated database increases linearly to the map's size [21]. Consequently, there has been much interest in developing compact appearance representations or mapping techniques to demonstrate sub-linear scaling in computational complexity and memory demands [22].

Regarding the former, compact representations and increased computational efficiency during database search are achieved using global visual features [23], [24], which use techniques that describe the image's appearance using a single vector [25]–[27]. However, robustness against transformations like scale and rotation are achieved via local visual features extracted through salient regions-of-interest [28]–[34]. Even if higher performance can be reached when local visual features are adopted [35], increased computational complexity is observed during the database search [36]. With the aim to exploit the advantages of both schemes, the robotics scholars borrowed the model of bag-of-words (BoW) from text retrieval [37], to effectively address the visual loop-closure detection task. In particular, each extracted local visual feature is assigned to a visual word belonging to a previously generated visual vocabulary [38]. This way, the environment is represented by visual word histograms [39]–[41]. Nevertheless, these approaches are dependent on the quality of their vocabulary and, in turn, their training data. As a result, this yields a trade-off between memory usage with detection performance and computational efficiency [42]. To address this drawback, visual words generated incrementally, typically produced by clustering similar description vectors along the navigation course, are proposed to improve the system's accuracy [43]–[48]. It is worth noting, when global features are used, comparisons are made based on

their descriptors' distance [49], *e.g.,* Euclidean or cosine [50]. However, for methods based on local features, voting schemes are adopted [51]. As a final note, many pipelines are based on GPU-enabled techniques to close loops in real-time [52].

Concerning the mapping techniques, different map management methods are adopted, including sparse topological maps or key-frames [53], hierarchical mapping via the description of a group-of-sequential-images [54], and the map's size reduction through memory scale discretization [55]. Key-frame selection is performed by utilizing techniques that detect changes in the incoming visual scenes. Using different decision metrics, such as distance and angle between observations in space [56], specific time intervals [57], and a minimum number of tracked landmarks [58], the system's complexity is reduced.

As the storage requirements needed to map the whole environment in long-term applications constitute a crucial factor, this paper proposes the compact representation of the incoming image stream for a low-storage loop-closure detection system. Our method is inspired by the work of Liu and Zhang, who detected loops using the most discriminant information extracted through a PCA technique [59]. This way, the authors achieved to reduce the descriptor space from 960 dimensions to 60 ones while maintaining high accuracy. However, in our framework, we rely on a data resampling process to reduce the descriptors' dimensions. Our system is built upon the bag-of-tracked-words scheme, which incrementally constructs a visual vocabulary and decides about familiar locations based on a probabilistic model. The main contributions of the proposed work are summarized as follows:

- A low-storage and low-complexity visual loop-closure technique based on downsized descriptor vectors, generated through a data resampling process.
- An extended experimental evaluation upon the database's size, the storage requirements, and the query timings.

The rest of the paper is organized as follows. Section II describes the proposed system. Section III evaluates its performance and, finally, Section IV concludes our work.

## II. Methodology

This section starts by briefly describing the bag-of-tracked-words model. In particular, the two primary operations, *viz.,* the database's build and search, are presented, while the proposed dimensionality reduction technique follows.

### A. Building the database

The incremental visual bag-of-tracked-words model maps the traversed environment based on three steps. In particular, the first regards the key-points extraction and tracking, the second the guided feature selection, and the last one the visual words' generation. In particular, as the extracted speeded-up robust features' [29] key-points ($\mathrm{P}_{t-1} = \{\mathrm{p}^1_{t-1}, \mathrm{p}^2_{t-1}, ..., \mathrm{p}^\nu_{t-1}\}$) enter the Kanade-Lucas-Tomasi tracker [60], their projected location between the previous image $I_{t-1}$ and the current visual data $I_t$ is obtained. This set of points is referred to as tracked points ($\mathrm{TP}_t = \{\mathrm{tp}^1_t, \mathrm{tp}^2_t, ..., \mathrm{tp}^\nu_t\}$). Using the nearest neighbor ($NN$) scheme among the tracked points' coordinate space, the ones ($\mathrm{TP}_t$) projected in image $I_t$ and the ones in $\mathrm{P}_t$ are matched. The nearest point $\mathrm{p}^{NN}_t \in \mathrm{P}_t$, for each tracked point $\mathrm{tp}^i_t$, is selected as a track-member when their points' and descriptors' distance satisfy a threshold [48]. If one of the above conditions is not met, the corresponding track point stops existing and another one is chosen in $I_t$ takes its place. Finally, a new tracked word is generated when a specific key-point's length $\tau$ is reached ($\tau > \rho$). The representative visual word is the average of tracked descriptors:

$$\mathrm{TW}[i] = \frac{1}{\tau}\sum_{j=1}^{\tau} \mathrm{d}_j[i] \tag{1}$$

where $\mathrm{d}_j[i]$ denotes the element in the $i$-th (SURF: $i \in [1, 64]$) dimension of the $j$-th ($j \in [1, \tau]$) description vector.

### B. Searching the database

A probabilistic voting scheme is used as a *decision-making* module. More specifically, during the query time $I_Q$, the features formulated by guided feature selection are matched with the generated tracked words in the database through a $k$-NN ($k$=1) technique. Next, votes are distributed into the traversed map to the corresponding locations and a database vote counter $x_l(t)$ in generated. Using a binomial density function, a probabilistic score is assigned to every database location:

$$X_l(t) \sim Bin(n, p), n = N(t), p = \frac{\lambda_l}{\Lambda(t)}. \tag{2}$$

$X_l(t)$ corresponds to the random variable concerning each database location's $l$ number of aggregated votes at time $t$. $N$ is the multitude of query's tracked words, *i.e.,* $\mathrm{TP}_Q$, $\lambda$ represents the number of visual elements included in $l$, *i.e.,* $\mathrm{TW}_l$, and $\Lambda(t)$ corresponds to the size of the searched database. If a location's binomial value satisfies a loop-closure threshold $th$, it is accepted:

$$Pr(X_l(t) = x_l(t)) < th < 1, \tag{3}$$

where $x_l(t)$ corresponds to the respective location's aggregated votes. Still, to avoid cases where unexpectedly few votes are aggregated by a location, the following condition should also hold:

$$x_l(t) > E[X_l(t)]. \tag{4}$$

Finally, the chosen pair of images is verified through a geometrical check. More specifically, a fundamental matrix based on RANSAC (RANSAC stands for random sample consensus [61]) should be estimated.

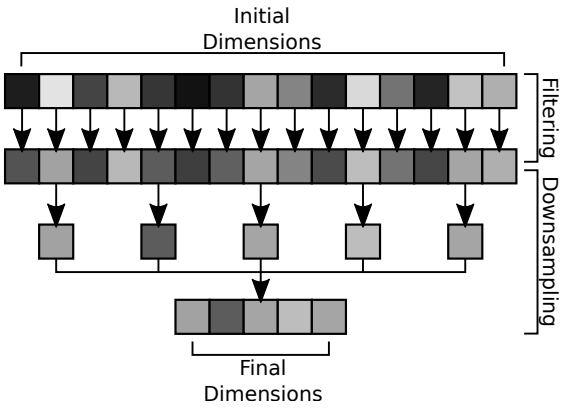


Fig. 1: An overview of the dimensionality reduction process. First, a lowpass filter is applied to the description vector, aiming to remove high frequency components. Then, samples are selected from the filtered signal with a constant interval, in order to form the final descriptor.

### C. Dimensional reduction of the visual data

Intending to accelerate the execution time needed for the proposed pipeline, we reduce the dimensionality of the descriptors' space for both query and database observations. Treating every vector as a one-dimensional signal, we utilize a re-sampling process to modify its length. While a direct value sampling using a constant interval can reduce the signal's length, embedded high-frequency components can distort the encoded information due to aliasing. Those frequencies are owed to texture-rich patches around the detected features and the descriptor method's layout, causing rapid changes in the vector's values. Therefore, we employ a lowpass filter [62] prior the periodic sampling process, to discard those frequencies and thus conserve the visual representation's integrity [63] (see Fig. 1).

## III. Experimental evaluation

In this section, the proposed technique is evaluated. At first, the datasets used for our experiments are described. Subsequently, the evaluation metrics adopted for assessing the system's performance are given, while the system's complexity and comparative results follow. All experiments were performed on an Intel i7-6700HQ 2.6 GHz processor with 8 GB RAM, while the bag-of-tracked-words parameterization is based on the open-source implementation[1].

### A. Evaluation datasets

Four different publicly-available datasets are selected for our experiments. Aiming to show our method's adaptability over several operational conditions, these are chosen because they present a variety of conditions, *i.e.,* camera properties, images per image-sequence, trajectory size, and sensor frequency. An overview of the used datasets is given in Table I. Using the visual stream recorded through a stereo camera rig mounted on a forward-moving car, sequences 00 and 05 are selected from the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) vision suite [64]. They provide substantial loop-closure examples and long-distance trajectories. Malaga 2009 parking 6L [65], obtained via the stereo vision system of an electric buggy-typed vehicle, constitutes the third scenario, while the incoming image stream in the EuRoC machine hall (EuRoC MH 05) 05 dataset is retrieved by cameras mounted on a hex-rotor helicopter.

[1]https://github.com/ktsintotas/Bag-of-Tracked-Words

TABLE I: Evaluation datasets.

| Name | Camera properties | # Images | Distance |
|---|---|---|---|
| KITTI 00 [64] | 1241 × 376, 10 Hz | 4551 | 12.5 Km |
| KITTI 05 [64] | 1241 × 376, 10 Hz | 2761 | 7.5 Km |
| Malaga 2009 6L [65] | 1024 × 768, 7.5 Hz | 3474 | 1.2 Km |
| EuRoC MH 05 [67] | 752 × 480, 20 Hz | 2273 | 0.1 Km |

The ground truth information for each dataset is employed to measure the proposed framework's performance. It is a boolean matrix with elements set to 1 when an actual loop-closure event happens and 0 otherwise. Concerning the data used for Malaga 2009 parking 6L, the authors in [52] formulated the ground truth information manually, while for the KITTI courses and EuRoC MH 05, we were based on the information employed in [66] and [22].

### B. Evaluation metrics

Using the most frequently evaluation metrics, *viz.* precision and recall [68], we evaluated our method. More specifically, precision is the ratio between the correct detections provided by the proposed system, *i.e.,* true-positives, over the framework's total identifications. The recall score corresponds to the number of true-positives over the total of actual loop events in the ground truth. As false-negatives are defined the cases that ought to be detected, but the pipeline failed to. The recall at 100% precision ($R_{P100}$) is utilized as a single evaluation metric for the system's performance, which demonstrates the highest achieved recall without false-positive detections.

### C. Performance evaluation

In Fig. 2, the system's overall performance is provided. The precision-recall curves are generated by varying the loop-closure threshold over the binomial probabilistic score. Adopting the same parameters used in the baseline version [48], we evaluate the impact of our filtering scheme. Our first remark is that each resulting curve presents a high recall rate on each evaluation datasets. As shown, the proposed visual reduction of the database permits the pipeline to successfully detect loops through a recall score ranging from 80.1% (EuRoC MH 05 - 32D) to 93.5% (KITTI 00 - 32D). The best results at 100% of precision are illustrated by the colored bars. It is worth noting that high performances are achieved when 32 dimensions are used; however, the recall drops as we decrease the size of visual data representations. This is owed to the fact that the distinctiveness of visual cues is reduced following the database size.

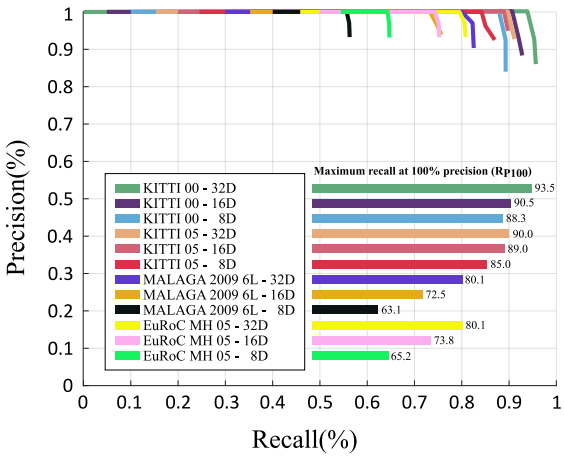


Fig. 2: Precision and recall curves as generated by the proposed pipeline. Colored bars at the bottom of each curve indicate the recall rates achieved from the system for each dataset. The 100% precision ($R_{P100}$) is also presented as an evaluation metric.

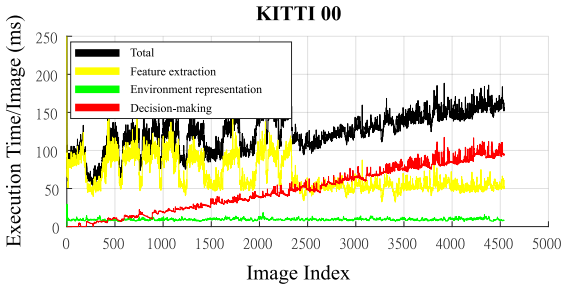


Fig. 3: System's response over the KITTI 00 dataset [64] for each of the main processing stages of the proposed algorithm.

### D. System's complexity

The average response timings per image produced by our system are illustrated in Fig. 3. The proposed technique is evaluated on the KITTI 00 image-sequence as it is the longest one while, at the same time, exhibiting a remarkable amount of loop-closure events. As a result, 4551 images are processed. An average timing of 124.4 ms per query image is achieved. Table II provides extensive timing documentation for each stage. The feature extraction process involves the computation of SURF key-points detection and description, while the incremental bag-of-tracked-words-based database is included in the environment representation. The decision-making step corresponds to the time needed for the visual vocabulary search via $k$-NN, as well as the time required for the geometrical verification of the candidate pair.

As presented in Table II, the system achieves low computational complexity. Except for the feature extraction process, which is high as expected for every approach using local visual features, the other components need a short time. Finally, the database search, which the most costly component in any visual loop-closure detection pipeline, reaches 45.2 ms on average due to the low dimensional space of the database.

TABLE II: System's timing (ms/query) for the KITTI 00 dataset [64].

| | | 32D | 16D | 8D |
|---|---|---|---|---|
| Feature extraction | Key-points detection | 42.6 | 42.6 | 42.6 |
| | Key-points description | 24.8 | 24.8 | 24.8 |
| Environment representation | Kanade-Lucas-Tomasi | 7.5 | 7.5 | 7.5 |
| | Guided feature selection | 1.8 | 1.8 | 1.8 |
| Decision-making | Database search | 45.2 | 23.1 | 16.9 |
| | Binomial scoring | 0.7 | 0.7 | 0.7 |
| | Geometrical verification | 1.8 | 1.8 | 1.8 |
| Total pipeline | | 124.4 | 102.3 | 96.1 |

TABLE III: In depth comparison with the baseline version of bag-of-tracked-words [48].

| Method | Baseline [48] | | | Proposed | | |
|---|---|---|---|---|---|---|
| Dataset | SURF (#) | Size (Mb) | Time (ms) | SURF (#) | Size (Mb) | Time (ms) |
| KITTI 00 [64] | 51K | 12.4 | 173.5 | 51K | **6.2** | **124.4** |
| KITTI 05 [64] | 21K | 7.0 | 130.1 | 21K | **3.5** | **102.1** |
| Malaga 6L [65] | 41K | 10.0 | 171.8 | 41K | **5.0** | **144.3** |
| EuRoC 05 [67] | 20K | 4.8 | 90.8 | 20K | **2.4** | **78.5** |

TABLE IV: In depth comparison with our previous work BoTW-LCD [22].

| Method | BoTW-LCD [22] | | | Proposed | | |
|---|---|---|---|---|---|---|
| Dataset | SURF (#) | Size (Mb) | Time (ms) | SURF (#) | Size (Mb) | Time (ms) |
| KITTI 00 [64] | 34K | 8.3 | 126.2 | 51K | **6.2** | **124.3** |
| KITTI 05 [64] | 20K | 4.8 | 105.3 | 21K | **3.5** | **102.1** |
| Malaga 6L [65] | 28K | 6.8 | 146.7 | 41K | **5.0** | **144.3** |
| EuRoC 05 [67] | 13K | 3.1 | 82.6 | 20K | **2.4** | **78.5** |

TABLE V: In depth comparison with the state-of-the-art iBoW-LCD [47].

| Method | iBoW-LCD [47] | | | Proposed | | |
|---|---|---|---|---|---|---|
| Dataset | ORB (#) | Size (Mb) | Time (ms) | SURF (#) | Size (Mb) | Time (ms) |
| KITTI 00 [64] | 958K | 29.2 | 400.2 | 51K | **6.2** | **124.3** |
| KITTI 05 [64] | 556K | 16.9 | 366.5 | 21K | **3.5** | **102.1** |
| Malaga 6L [65] | 806K | 24.5 | 440.8 | 41K | **5.0** | **144.3** |
| EuRoC 05 [67] | 443K | 13.5 | 383.7 | 20K | **2.4** | **78.5** |

### E. Comparative results

This section compares the proposed method against other state-of-the-art frameworks in incremental visual vocabulary building, namely BoTW-LCD[2], iBoW-LCD[3], as well as the baseline version of bag-of-tracked-words. Note that a comparison with off-line BoW schemes regarding their respective complexities is not presented since a direct analogy with methods based on a pre-trained vocabulary would not be meaningful. In Tables III, IV, and V, we exhaustively compare the memory requirements (Size) needed for each visual vocabulary and the corresponding time (Time) for

[2]The BoTW-LCD [22] open-source implementation can be found at https://https://github.com/ktsintotas/BoTW-LCD.

[3]The iBoW-LCD [47] open-source implementation can be found at https://github.com/emiliofidalgo/ibow-lcd.

TABLE VI: Comparisons against other state-of-the-art methods using the recall scores for 100% precision ($R_{P100}$).

| Dataset | Baseline [48] | BoTW-LCD [22] | iBoW-LCD [47] | Ours |
|---|---|---|---|---|
| KITTI 00 [64] | 97.5% | **97.7%** | 76.5% | 93.5% |
| KITTI 05 [64] | 92.6% | **94.0%** | 53.0% | 90.0% |
| Malaga 6L [65] | 85.0% | **85.2%** | 57.4% | 80.1% |
| EuRoC 05 [67] | 83.7% | **85.0%** | 25.6% | 80.1% |

each method to detect loop-closures. Note that in most cases, the proposed technique maps the robot's traversed path with a noticeably lower amount of memory consumption, which also permits lessened timings during database search.

In addition, in Table VI, we compare the proposed technique with the pipelines mentioned above regarding their achieved performance. The recall score for flawless precision ($R_{P100}$) is provided. The cited methods' performance is acquired from our previous work [22], wherein each method was evaluated based on the same ground truth. As can be observed, the proposed system can achieve high recall rates in most environments compared to the state-of-the-art. Notably, in terms of recall, it is quite similar to the baseline while outperforming iBoW-LCD. However, it performs unfavorably compared to BoTW-LCD, which constitutes a method that evolves the original concept of the bag-of-tracked-words using visual vocabulary management techniques and more sophisticated decision methods. As a final note, using data resampling as a means of lowering the descriptors' dimensionality, can consistently reduce the computational times and the size of the storage requirements, yet does not always imply higher recall values.

## IV. Conclusions

In this work, a low-storage visual loop-closure detection framework is proposed. Based on the bag-of-tracked-words model, a resampling technique comprised of an anti-aliasing lowpass filter and a data selection mechanism is applied over the extracted visual local features to reduce the memory footprint. This way, an incremental visual vocabulary is constructed, offering low complexity and competitive accuracy as evidenced by its extensive evaluation on four different datasets. Finally, our future plans include the study of different indexing techniques to further reduce the system's timings.

## Acknowledgment

This work has been implemented within the projects "Wearable systems for the safety and wellbeing applied in security guards - SafeIT" and "Study, design, development and implementation of a holistic system for upgrading the quality of life and activity of the elderly" which have been financially supported by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the calls "Research - create - innovate" grant number [T2EDK-01862] and "Support for regional excellence" grant number [MIS 5047294].

## References

[1] I. Kostavelis and A. Gasteratos, "Semantic mapping for mobile robotics tasks: A survey," *Robotics and Autonomous Systems*, vol. 66, pp. 86–103, 2015.

[2] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Trans. Robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[3] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "The revisiting problem in simultaneous localization and mapping: A survey on visual loop closure detection," *Trans. Intelligent Transportation Systems*, 2022.

[4] I. Kostavelis, A. Gasteratos, E. Boukas, and L. Nalpantidis, "Learning the terrain and planning a collision-free trajectory for indoor post-disaster environments," in *Proc. IEEE Int. Symp. Safety, Security, and Rescue Robotics*, pp. 1–6, 2012.

[5] V. Balaska, L. Bampis, M. Boudourides, and A. Gasteratos, "Unsupervised semantic clustering and localization for mobile robotics tasks," *Robotics and Autonomous Systems*, vol. 131, p. 103567, 2020.

[6] F. K. Konstantinidis, I. Kansizoglou, K. A. Tsintotas, S. G. Mouroutsos, and A. Gasteratos, "The role of machine vision in industry 4.0: A textile manufacturing perspective," in *Proc. IEEE Int. Conf. Imaging Systems and Techniques*, pp. 1–6, 2021.

[7] V. Balaska, L. Bampis, and A. Gasteratos, "Self-localization based on terrestrial and satellite semantics," *Eng. Appl. Artificial Intelligence*, vol. 111, p. 104824, 2022.

[8] W. Burgard, C. Stachniss, and D. Hähnel, "Mobile robot map learning from range data in dynamic environments," in *Autonomous Navigation in Dynamic Environments*, pp. 3–28, 2007.

[9] K. A. Tsintotas, L. Bampis, A. Taitzoglou, I. Kansizoglou, and A. Gasteratos, "Safe UAV landing: A low-complexity pipeline for surface conditions recognition," in *Proc. IEEE Int. Conf. Imaging Systems and Techniques*, pp. 1–6, 2021.

[10] V. A. Feraru, R. E. Andersen, and E. Boukas, "Towards an autonomous UAV-based system to assist search and rescue operations in man overboard incidents," in *Pro. IEEE Int. Symp. Safety, Security, and Rescue Robotics*, pp. 57–64, 2020.

[11] R. Y. Brogaard, M. Zajaczkowski, L. Kovac, O. Ravn, and E. Boukas, "Towards UAV-based absolute hierarchical localization in confined spaces," in *Proc. IEEE Int. Symp. Safety, Security, and Rescue Robotics*, pp. 182–188, 2020.

[12] I. T. Papapetros, V. Balaska, and A. Gasteratos, "Multi-layer map: Augmenting semantic visual memory," in *Proc. Int. Conf. Unmanned Aircraft Systems*, pp. 1206–1212, 2020.

[13] R. Y. Brogaard, R. E. Andersen, L. Kovac, M. Zajaczkowski, and E. Boukas, "Towards an autonomous, visual inspection-aware 3d exploration and mapping system for water ballast tanks of marine vessels," in *Proc. IEEE Int. Conf. on Imaging Systems and Techniques*, pp. 1–6, 2021.

[14] K. Siozios, D. Diamantopoulos, I. Kostavelis, E. Boukas, L. Nalpantidis, D. Soudris, A. Gasteratos, M. Avilés, and I. Anagnostopoulos, "SPARTAN project: Efficient implementation of computer vision algorithms onto reconfigurable platform targeting to space applications," in *Proc. IEEE Int. Workshop on Reconfigurable Communication-Centric Systems-on-Chip*, pp. 1–9, 2011.

[15] I. Kostavelis, E. Boukas, L. Nalpantidis, A. Gasteratos, and M. A. Rodrigalvarez, "SPARTAN system: Towards a low-cost and high-performance vision architecture for space exploratory rovers," in *Proc. IEEE Int. Conf. Computer Vision Workshops*, pp. 1994–2001, 2011.

[16] E. Boukas and A. Gasteratos, "Modeling regions of interest on orbital and rover imagery for planetary exploration missions," *Cybernetics and Systems*, vol. 47, no. 3, pp. 180–205, 2016.

[17] E. Boukas, A. Gasteratos, and G. Visentin, "Introducing a globally consistent orbital-based localization system," *J. Field Robotics*, vol. 35, no. 2, pp. 275–298, 2018.

[18] V. Balaska, L. Bampis, I. Kansizoglou, and A. Gasteratos, "Enhancing satellite semantic maps with ground-level imagery," *Robot. Auton. Systems*, vol. 139, p. 103760, 2021.

[19] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Trans. Robotics*, vol. 32, no. 1, pp. 1–19, 2015.

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

[21] S. An, G. Che, F. Zhou, X. Liu, X. Ma, and Y. Chen, "Fast and incremental loop closure detection using proximity graphs," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 378–385, 2019.

[22] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "Modest-vocabulary loop-closure detection with incremental bag of tracked words," *Robotics and Autonomous Systems*, p. 103782, 2021.

[23] B. Schiele and J. L. Crowley, "Object recognition using multidimensional receptive field histograms," in *Eur. Conf. Computer Vision*, pp. 610–619, 1996.

[24] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," *Prog. Brain Research*, vol. 155, pp. 23–36, 2006.

[25] N. Sünderhauf and P. Protzel, "BRIEF-gist-closing the loop by simple means," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 1234–1241, 2011.

[26] E. Garcia-Fidalgo and A. Ortiz, "Vision-based topological mapping and localization methods: A survey," *Robotics and Autonomous Systems*, vol. 64, pp. 1–20, 2015.

[27] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "DOSeqSLAM: Dynamic on-line sequence based loop closure detection algorithm for SLAM," in *Proc. IEEE Int. Conf. Imaging Systems and Techniques*, pp. 1–6, 2018.

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

[29] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded-up robust features," in *Eur. Conf. Computer Vision*, pp. 404–417, May 2006.

[30] M. Agrawal, K. Konolige, and M. R. Blas, "CENSURE: Center surround extremas for realtime feature detection and matching," in *Proc. Eur. Conf. Computer Vision*, pp. 102–115, 2008.

[31] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proc. Int. Conf. Computer Vision*, pp. 2564–2571, 2011.

[32] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Computer Vision*, pp. 2548–2555, 2011.

[33] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Proc. Eur. Conf. Computer Vision*, pp. 214–227, 2012.

[34] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 510–517, 2012.

[35] J. P. Company-Corcoles, E. Garcia-Fidalgo, and A. Ortiz, "Towards robust loop closure detection in weakly textured environments using points and lines," in *Proc. IEEE Int. Conf. Emerging Technologies and Factory Automation*, pp. 1313–1316, 2020.

[36] K. A. Tsintotas, P. Giannis, L. Bampis, and A. Gasteratos, "Appearance-based loop closure detection with scale-restrictive visual features," in *Proc. Int. Conf. Computer Vision Systems*, pp. 75–87, 2019.

[37] R. Baeza-Yates, B. Ribeiro-Neto, *et al.*, *Modern information retrieval*, vol. 463. ACM press New York, 1999.

[38] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Computer Vision*, vol. 3, pp. 1470–1470, 2003.

[39] R. Mur-Artal and J. D. Tardós, "Fast relocalisation and loop closing in keyframe-based slam," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 846–853, 2014.

[40] V. Balaska, L. Bampis, and A. Gasteratos, "Graph-based semantic segmentation," in *Proc. Int. Conf. Robotics in Alpe-Adria Danube Region*, pp. 572–579, 2018.

[41] K. A. Tsintotas, L. Bampis, S. Rallis, and A. Gasteratos, "SeqSLAM with bag of visual words for appearance based loop closure detection," in *Proc. Int. Conf. Robotics in Alpe-Adria Danube Region*, pp. 580–587, 2018.

[42] I. T. Papapetros, V. Balaska, and A. Gasteratos, "Visual loop-closure detection via prominent feature tracking," *J. Intelligent & Robotic Systems*, vol. 104, no. 3, pp. 1–13, 2022.

[43] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Fast and incremental method for loop-closure detection using bags of visual words," *IEEE Trans. Robotics*, vol. 24, no. 5, pp. 1027–1037, 2008.

[44] T. Nicosevici and R. Garcia, "Automatic visual bag-of-words for online robot navigation and mapping," *IEEE Trans. Robotics*, vol. 28, no. 4, pp. 886–898, 2012.

[45] S. Khan and D. Wollherr, "IBuILD: Incremental bag of binary words for appearance based loop closure detection," in *Proc. IEEE Int. Conf. Robotics and Automation (ICRA)*, pp. 5441–5447, 2015.

[46] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "Assigning visual words to places for loop closure detection," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 5979–5985, May 2018.

[47] E. Garcia-Fidalgo and A. Ortiz, "iBOW-LCD: An appearance-based loop-closure detection approach using incremental bags of binary words," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3051–3057, 2018.

[48] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "Probabilistic appearance-based place recognition through bag of tracked words," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1737–1744, 2019.

[49] I. Kansizoglou, L. Bampis, and A. Gasteratos, "Deep feature space: A geometrical perspective," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2021.

[50] I. Kansizoglou, L. Bampis, and A. Gasteratos, "Do neural network weights account for classes centers?," *IEEE Trans. Neural Networks and Learning Systems*, 2022.

[51] M. Gehrig, E. Stumm, T. Hinzmann, and R. Siegwart, "Visual place recognition with probabilistic voting," in *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 3192–3199, 2017.

[52] S. An, H. Zhu, D. Wei, K. A. Tsintotas, and A. Gasteratos, "Fast and incremental loop closure detection with deep features and proximity graphs," *J. Field Robotics*, vol. 39, no. 4, pp. 473–493, 2022.

[53] E. Garcia-Fidalgo and A. Ortiz, "Hierarchical place recognition for topological mapping," *IEEE Trans. Robotics*, vol. 33, no. 5, pp. 1061–1074, 2017.

[54] K. A. Tsintotas, L. Bampis, S. An, G. F. Fragulis, S. G. Mouroutsos, and A. Gasteratos, "Sequence-based mapping for probabilistic visual loop-closure detection," in *Proc. IEEE Int. Conf. Imaging Systems and Techniques*, pp. 1–6, 2021.

[55] M. Labbe and F. Michaud, "Appearance-based loop closure detection for online large-scale and long-term operation," *IEEE Trans. Robotics*, vol. 29, no. 3, pp. 734–745.

[56] L. Bampis and A. Gasteratos, "Revisiting the bag-of-visual-words model: A hierarchical localization architecture for mobile systems," *Robotics and Autonomous Systems*, vol. 113, pp. 104–119, 2019.

[57] 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 Int. Conf. Intelligent Robots and Systems*, pp. 4530–4536, 2016.

[58] K. A. Tsintotas, L. Bampis, and A. Gasteratos, "Tracking-DOSeqSLAM: A dynamic sequence-based visual place recognition paradigm," *IET Computer Vision*, 2021.

[59] Y. Liu and H. Zhang, "Visual loop closure detection with a compact image descriptor," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 1051–1056, 2012.

[60] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artificial Intelligence*, pp. 674–679, 1981.

[61] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[62] T. W. Parks and C. S. Burrus, *Digital Filter Design*, p. 54–83. Wiley, 1987.

[63] R. E. Crochiere and L. R. Rabiner, *Basic Principles of Sampling and Sampling Rate Conversion*. Prentice-Hall, 1996.

[64] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 3354–3361, 2012.

[65] 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.

[66] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, J. J. Yebes, and S. Bronte, "Fast and effective visual place recognition using binary codes and disparity information," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp. 3089–3094, 2014.

[67] M. Burri, J. Nikolic, P. Gohl, T. Schneider, J. Rehder, S. Omari, M. W. Achtelik, and R. Siegwart, "The EuRoC micro aerial vehicle datasets," *Int. J. Robot. Res.*, vol. 35, no. 10, pp. 1157–1163, 2016.

[68] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *Int. J. Robotics Research*, vol. 27, no. 6, pp. 647–665, 2008.