Towards Life-Long Visual Localization using an Efficient Matching of Binary Sequences from Images

Roberto Arroyo[1], Pablo F. Alcantarilla[2], Luis M. Bergasa[1] and Eduardo Romera[1]

***Abstract*— Life-long visual localization is one of the most challenging topics in robotics over the last few years. The difficulty of this task is in the strong appearance changes that a place suffers due to dynamic elements, illumination, weather or seasons. In this paper, we propose a novel method (ABLE-M) to cope with the main problems of carrying out a robust visual topological localization along time. The novelty of our approach resides in the description of sequences of monocular images as binary codes, which are extracted from a global LDB descriptor and efficiently matched using FLANN for fast nearest neighbor search. Besides, an illumination invariant technique is applied. The usage of the proposed binary description and matching method provides a reduction of memory and computational costs, which is necessary for long-term performance. Our proposal is evaluated in different life-long navigation scenarios, where ABLE-M outperforms some of the main state-of-the-art algorithms, such as WI-SURF, BRIEF-Gist, FAB-MAP or SeqSLAM. Tests are presented for four public datasets where a same route is traversed at different times of day or night, along the months or across all four seasons.**

I. Introduction

Navigation of autonomous vehicles in long-term periods has experienced a great interest by the robotics community in recent times. Due to this, solving the life-long visual localization problem for identifying where a robot is over time has become one of the main challenging areas of research. Unfortunately, this is not an easy task, because places have strongly different appearances at different times of day, along the months and especially along the seasons.

In the last years, vision has successfully demonstrated that it can be a complementary or alternative option for localization with respect to other sensing techniques such as range-based or GPS-based. According to this, FAB-MAP [1] can be considered as the milestone in visual topological localization methods for detecting loop closures. This algorithm allows to recognize previously visited places by only taking into account the space of appearance and individually matching the images. In subsequent work, FAB-MAP was tested over 1000 km [2], being one of the first approaches to life-long visual localization in the literature.

More recently, SeqSLAM [3] introduced the idea of matching places by considering sequences instead of single images like previous proposals such as FAB-MAP. The usage of sequences denoted a higher robustness to extreme perceptual changes. SeqSLAM was validated under challenging life-long visual localization conditions which were mainly based on comparing the images of a same route in a sunny summer day and a stormy winter night. However, in [4] some weaknesses of SeqSLAM were reported, such as the field of view dependence and the influence of parameters configuration. For these reasons, the community continues searching for new methods which can satisfy the high requirements needed to achieve a robust life-long visual localization.

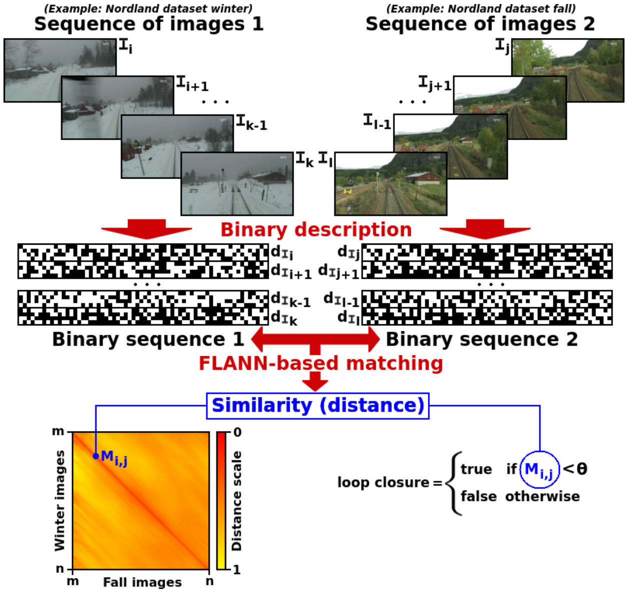


Fig. 1. General diagram of our life-long visual topological localization system using an efficient matching of binary sequences from images.

During the last year, some new methods using binary codes extracted from images and fast matching techniques have successfully contributed to the state of the art of visual topological localization, such as the different variants of ABLE [5], [6]. One of the main advantages of applying binary descriptors in life-long visual localization is the reduction of computational and memory costs without a significant loss of description power. The main goal of the present work is to design an improved approach in this line which can satisfactorily and efficiently operate in a life-long context.

Attending to the previous considerations, we propose a novel method named ABLE-M for achieving a life-long visual localization using an efficient matching of binary sequences from images, as detailed in the general diagram of our approach presented in Fig. 1.

*This work is funded by the UAH through a FPI grant, the Spanish MINECO through the project Smart Driving Applications (TEC2012-37104) and the CAM through the project RoboCity2030-III-CM (P2013/MiT2748).

[1] Department of Electronics, University of Alcalá (UAH), Alcalá de Henares, 28871, Madrid, Spain. {roberto.arroyo, bergasa, eduardo.romera}@depeca.uah.es

[2] Computer Vision Group, Toshiba Research Europe Ltd., Cambridge, CB4 0GZ, UK. pablo.alcantarilla@crl.toshiba.co.uk

The main contribution of this paper is the implementation of a robust visual topological localization system focused on helping robots and intelligent vehicles to perform a life-long navigation, which represents an innovative approach regarding the state of the art (see Section II). The novelty of our method is in the description of sequences of monocular images as binary codes, which are extracted from a global LDB [7] descriptor and efficiently matched using the Hamming distance jointly with approximated nearest neighbor search using the FLANN library [8]. Besides, an illumination invariant technique is also applied (see Section III). In addition, we contribute a wide set of experiments (see Section IV) and results (see Section V) that validate our approach in some of the most challenging publicly available datasets for life-long visual localization: the St Lucia dataset [9], the Alderley dataset [3], the CMU-CVG Visual Localization dataset [10] and the Nordland dataset [4]. Finally, we present the main conclusions and the future research lines (see Section VI).

II. State of the Art

A. Life-long Visual Topological Localization

Apart from the previously referenced algorithms such as FAB-MAP or SeqSLAM, several proposals have appeared in the last years with the aim of achieving an effective life-long visual localization. Some of these recent works have studied the main problems that produce drastic changes in image appearance: environmental conditions such as illumination or weather [11], dynamic elements in a scene [12] or camera configuration [13].

One of the main tasks for being able to perform a life-long localization is to reduce the information storage and computational costs. Recently, some methods have been proposed based on a long-term memory methodology, such as RTAB-Map [14]. Another option lately used for decreasing memory costs is to reduce the weight of the stored descriptors and the computational costs of matching by applying a global binary description of images [5] [6]. Besides, in [15] a great diminution of processing costs for loop closure detection is achieved by using very simple binarized image representations tested in a map of 20 million key locations.

Nowadays, one of the most fashionable and challenging topics in life-long visual topological localization is to recognize previously visited places along the different seasons. The method presented in [16] is probably one of the first approaches to visual topological localization over seasonal conditions and it is mainly based in the detection, extraction and matching of SURF descriptors [17]. In [10], places are compared in different months of the year by using a global SURF descriptor called WI-SURF, which is combined with 3D laser information. Another very recent proposal is focused on the prediction of changes on places based on superpixel vocabularies [18], which has been tested for a 3000 km dataset across all four seasons [4]. Besides, changes in scene along the different hours of day and night have also been studied with a great interest in SeqSLAM tests and in other recent approaches based on co-occurrence maps [19].

B. Binary Descriptors and Matching of Images

Binary descriptors have recently been popularized in computer vision due to their simplicity and favorable conditions, such as the low memory requirements needed to store them or the possibility of carrying out a very fast matching by using the Hamming distance, as exposed in [8]. Furthermore, in works such as [6], it is demonstrated that they can be competitive with respect to vector-based descriptors such as WI-SURF in a global place recognition framework.

Some state-of-the-art methods have satisfactorily applied global binary descriptors for visual topological localization in short-term scenarios, such as Gabor-Gist [20] or BRIEF-Gist [21], which is based on the BRIEF descriptor [22]. One of the most recent approaches in this line is ABLE, which has achieved remarkable results in its two current versions: panoramic (ABLE-P) [5] and stereo (ABLE-S) [6]. Besides, in [5] several studies demonstrated that LDB [7] can be considered the most effective binary descriptor for visual topological localization with respect to the main state-of-the-art binary descriptors. For these reasons, in the present paper we follow some of these concepts about binary description and matching, but now improved for applying them in a life-long localization context.

III. Our Approach: ABLE-M

A. Single Images or Sequences?

Traditionally, visual topological localization has been performed by considering places as single images. In fact, there are several remarkable approaches which trust in this philosophy, such as WI-SURF, BRIEF-Gist or FAB-MAP. However, other more recent proposals such as SeqSLAM changed this concept and introduced the idea of recognizing places as sequences of images.

In the present paper, we also follow the idea of using sequences of images instead of single images for identifying places. This approach allows to achieve better results in life-long visual topological localization, as can be seen in Fig. 2, where some previous results obtained in the Nordland dataset between the sequences of winter and fall are presented.

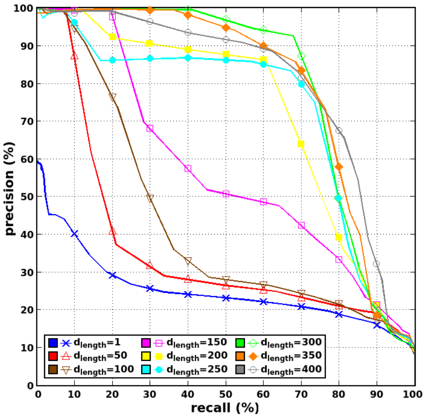


Fig. 2. An example of performance comparison of our proposal (ABLE-M) depending on the image sequence length ($\mathbf{d}_{length}$) in the challenging Nordland dataset (winter vs fall).

Attending to the precision-recall curves advanced in Fig. 2, the influence of the sequence length ($\mathbf{d}_{length}$) is decisive for improving the performance of visual topological localization in life-long conditions. Furthermore, there is a limit near to a length of 300 where results are not greatly enhanced. For this reason, in the rest of the experiments and results which will be presented in Sections IV and V, we will apply $\mathbf{d}_{length} = 300$.

With the aim of referring to our approach and following the nomenclature started in our previous works [5] [6], we name the method presented in this paper as **ABLE-M** (**A**ble for **B**inary-appearance **L**oop-closure **E**valuation - **M**onocular).

B. Illumination Invariance

Some techniques can be applied before extracting the binary sequences from images to reduce the problems associated with illumination changes, especially along the different hours of day. As demonstrated in recent works such as [23] or [24], images can be previously transformed to an illuminant invariant color space with the aim of improving the performance of place recognition in changing illumination conditions, as presented in Eq. 1:

$$\mathcal{I} = \log(G) - \alpha \cdot \log(B) - (1-\alpha) \cdot \log(R) \qquad (1)$$

where R, G, B are the color channels of the processed image and $\mathcal{I}$ is the resultant illumination invariant image. As shown in Eq. 2, α is a parameter which depends on the peak spectral responses of each color channel (λ_R, λ_G, λ_B), which are commonly available in camera specifications.

$$\frac{1}{\lambda_G} = \frac{\alpha}{\lambda_B} + \frac{(1-\alpha)}{\lambda_R} \qquad (2)$$

Therefore, α can be easily calculated by considering the peak spectral responses, as exposed in Eq. 3. In addition, Fig. 3 shows the influence of applying the illumination invariant transformation.

$$\alpha = \frac{\left(\frac{\lambda_B}{\lambda_G} - \frac{\lambda_B}{\lambda_R}\right)}{\left(1 - \frac{\lambda_B}{\lambda_R}\right)} \qquad (3)$$

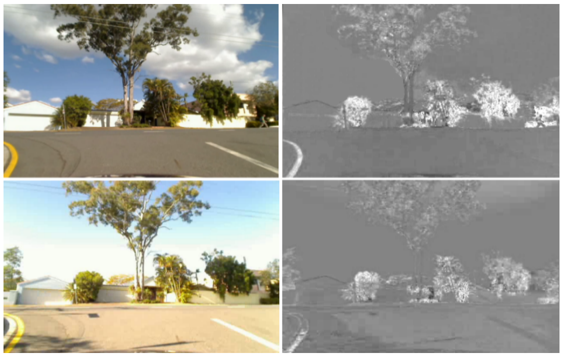


Fig. 3. An example of illumination invariance application in the St Lucia dataset between two places at different hours. The images on right are the illumination invariant transformation of the left images. It can be seen how this approach reduces the effects produced by sunlight and shadows.

The advantages of illumination invariance will be more detailed and demonstrated with several results for visual place recognition and loop closure detection at different hours of day in Section V-A.

C. Extracting Binary Sequences

According to works such as [25], high resolution images are not needed to perform an effective visual topological localization along time and only a handful of bits is sufficient. We agree with this concept because it can decrease computational cost without a robustness reduction. For this reason, as demonstrated in [5], we previously downsample the processed images to 64x64 pixels, which is also the patch used for the subsequent binary description of each image.

We use a global LDB descriptor for extracting the binary codes of each image because LDB improves the performance of binary description for place recognition by adding gradient information jointly with intensity comparisons, which is more robust than the approaches presented by descriptors such as BRIEF, where only intensity is processed. Besides, we compute the global binary descriptor by taking the center of the downsampled images as a keypoint without dominant rotation or scale. The resultant binary codes computed for each image ($\mathbf{d}_{\mathcal{I}}$) are adjusted to a length of 32 bytes, as justified in [5].

Finally, the binary codes extracted from each image are concatenated ($+\!\!\!+$) to form the final binary sequence ($\mathbf{d}$) corresponding to a sequence of images, as exposed in Eq. 4, where $k - i$ is equal to the $\mathbf{d}_{length}$ chosen.

$$\mathbf{d} = \mathbf{d}_{\mathcal{I}_i} +\!\!\!+ \mathbf{d}_{\mathcal{I}_{i+1}} +\!\!\!+ \mathbf{d}_{\mathcal{I}_{i+2}} +\!\!\!+ ... +\!\!\!+ \mathbf{d}_{\mathcal{I}_{k-2}} +\!\!\!+ \mathbf{d}_{\mathcal{I}_{k-1}} +\!\!\!+ \mathbf{d}_{\mathcal{I}_k} \qquad (4)$$

D. Efficient Matching of Binary Sequences

Binary descriptors can be efficiently matched by computing the Hamming distance, which is faster than the traditional way of matching descriptors with the L_2-norm. As exposed in Eq. 5, similarity between binary sequences is computed for loop closure detection by using the Hamming distance, which is based on a simple XOR operation ($\oplus$) and a sum of bits. The similarity values can be stored on a distance matrix (M) for evaluation purposes.

$$M_{i,j} = M_{j,i} = \text{bitsum}(\mathbf{d}_i \oplus \mathbf{d}_j) \qquad (5)$$

Approximated nearest neighbor search using the FLANN library can be employed for reducing the computational cost of matching over time. We use a hashing method for fast matching of binary features [8], which is conveniently implemented in the last versions of the OpenCV libraries [26]. The index applied in search is based on a multi-probe LSH (Local Sensitive Hashing), which is described in detail in [27].

IV. Experiments in Life-long Visual Localization

A. Datasets and Ground-truths

With the aim of carrying out a robust evaluation of ABLE-M performance in life-long scenarios and comparing it against some of the main state-of-the-art methods, we use four publicly available datasets with different conditions, which are more detailed in Table I.

TABLE I

DESCRIPTION OF THE MAIN CHARACTERISTICS OF THE DATASETS EMPLOYED IN THE EXPERIMENTS.

Dataset	Lenght	No. Images	General comments	Image samples of revisited places	Map route
St Lucia [9]	10x12 km	10x21814 (640x480 px) (15 fps)	This dataset is collected in the Brisbane suburb of St Lucia, in Australia. A route of about 20-25 minutes is traversed by a car ten times at different day hours. GPS positions are logged during each journey.		
Alderley [3]	2x8 km	2x17000 (640x256 px) (25 fps)	Two car rides recorded in a same route in the Brisbane suburb of Alderley, in Australia. One of them takes place in a sunny summer day and the other one in a stormy winter night. Correspondences are manually labeled.		
CMU-CVG Visual Loc. [10]	5x8 km	5x13357 (1024x768 px) (15 fps)	A same route is traversed five times in Pittsburgh, PA, USA. The sequences are recorded with two cameras in different months under varying environmental and climatological conditions. GPS and LiDAR information are registered.		
Nordland [4]	4x728 km	4x894172 (1920x1080 px) (25 fps)	A ten hour train ride in northern Norway is recorded four times, once in every season. Video sequences are synchronized and the camera position and field of view are always the same. GPS readings are available.		

The datasets presented in Table I allow us to test our proposal for the problematic visual changes that a place suffers along different periods of time: along the day (St Lucia dataset), along the day and night (Alderley dataset), along the months (CMU-CVG Visual Localization dataset) and along the seasons (Nordland dataset). For each dataset, Table I shows the number of recorded sequences and its length in kilometers, the number of images or frames which compose the dataset and their characteristics, a brief description of the dataset jointly with general comments about it, some image samples of revisited places in different environmental conditions and the map route followed in all the sequences of each dataset. The routes showed for each dataset map are represented by processing the available GPS measurements, which can be also used for generating the ground-truths used for evaluation.

B. Evaluation

The results presented for our visual topological localization approach are processed by following the objective evaluation methodology detailed in [6], which is mainly based on precision-recall curves obtained from the computed distance matrices.

It must be noted that distance matrices are used for evaluating the global performance of our proposal. In real application, for purposes such as correcting SLAM or visual odometry errors, some kind of threshold (θ) must be applied to determine if the similarity between a pair of sequences of images corresponds to a loop closure or not, as exposed in Eq 6. In this line, empirical thresholds can be used or also adaptive thresholds, as has been studied in some recent state-of-the-art works such as [28].

$$\text{loop closure} = \begin{cases} true & \text{if } M_{i,j} < \theta \\ false & \text{otherwise} \end{cases} \tag{6}$$

V. RESULTS IN LIFE-LONG VISUAL LOCALIZATION

A. Along the day: St Lucia dataset

This dataset allows us to evaluate the improvements provided by our illumination invariant proposal in a visual topological localization along the day. This is because the St Lucia dataset contains several video sequences recorded for a same route where varied illumination changes and shadowing effects appear when a place is traversed at different times of day.

Apart from the qualitative results previously showed for illumination invariance in Fig. 3, we present now precision-recall curves where the advantages of applying the illumination invariant transformation in ABLE-M are evidenced by comparing two sequences of the St Lucia dataset recorded at two different hours of a same day, as depicted in Fig. 4.

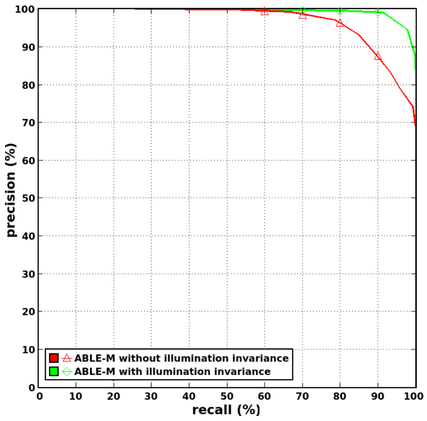


Fig. 4. Precision-recall curves comparing the performance of ABLE-M with and without using the illumination invariant technique between the sequences of the St Lucia dataset corresponding to the car rides recorded on 10/09/2009 at 8:45 am and at 2:10 pm.

In Fig. 4, it is demonstrated that the usage of illumination invariance improves the general performance of ABLE-M in visual localization when a route is traversed along the day. The description of the sequences of images is more robust in this case due to the reduction of the effects produced by sunlight and shadows when illumination invariance is applied by our algorithm.

B. Along the day and night: Alderley dataset

The Alderley dataset contains two video sequences which are very challenging for life-long visual localization: one of them in a sunny summer day and the other one in a stormy winter night. Obviously, in these conditions is much more difficult to match places and ABLE-M achieves worse results than for the tests carried out in the St Lucia dataset, as corroborated if the precision-recall curves presented in Fig. 5 are compared to the previously depicted in Fig. 4.

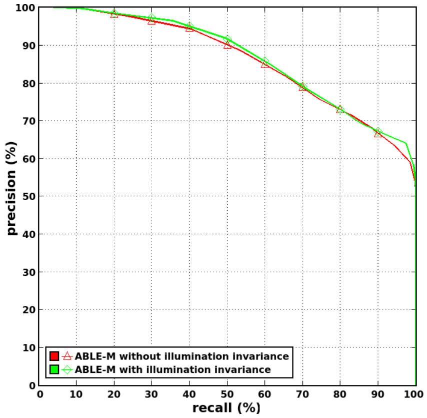


Fig. 5. Precision-recall curves comparing the performance of ABLE-M with and without using the illumination invariant technique between the day and night sequences of the Alderley dataset.

The application of the illumination invariant technique does not improve the results at night because the source illuminant cannot be modeled as a black-body radiator and the obtained appearance does not completely match the illumination invariance captured during the day. In addition, in the specific circumstances of the Alderley dataset, raindrops and humidity in camera also negatively affect to the effectiveness of this method, as can be observed in Fig. 6.

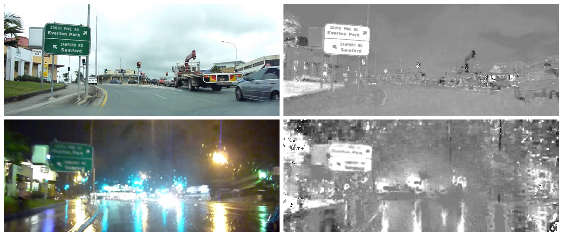


Fig. 6. An example of illumination invariance application in the Alderley dataset between two places in a sunny summer day and in a stormy winter night. The images on right are the illumination invariant transformation of the left images.

Similar problems due to night illumination conditions were reported in [24], where their approach combined with an illumination invariant transformation also has worse results for a same route traversed between 8:00 pm and 7:00 am. In these conditions, it can be very difficult to perform a robust visual localization, as demonstrated along this section.

C. Along the months: CMU-CVG Visual Localization dataset

The tests carried out in this dataset reveal the effectiveness of our method for visual topological localization along longer periods of time, where seasonal changes in places and other challenging environmental conditions such as the described in Fig. 7 must be taken into account.

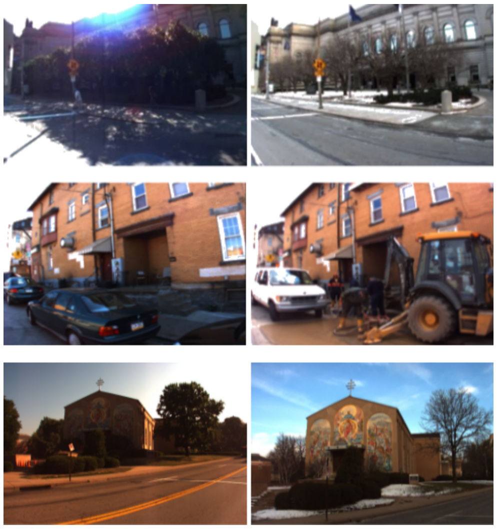


Fig. 7. Three representative examples of difficult situations for place recognition in the CMU-CVG Visual Localization dataset. From top to bottom, each pair of images represents: a) Seasonal changes produced by illumination, vegetation or snow. b) Problems of urban visual localization such as constructions or dynamic elements. c) Changes on the field of view.

We test ABLE-M in this dataset compared to the following methods: WI-SURF, BRIEF-Gist, FAB-MAP and SeqSLAM. For testing WI-SURF and BRIEF-Gist, we have implemented them using the SURF and BRIEF descriptors provided by OpenCV. FAB-MAP is evaluated using the OpenFABMAP implementation presented in [29], which is applied in a standard configuration and conveniently trained. The evaluation of SeqSLAM is carried out by employing the source code provided by OpenSeqSLAM [4].

The results presented in this section correspond to two sequences recorded with more of three months of difference. Only the left camera images of the dataset are employed in the test. As can be seen in Fig. 8, WI-SURF, BRIEF-Gist and FAB-MAP do not achieve great results in this life-long visual localization context because these methods follow the philosophy of considering places as single images, which does not have an effective performance for the environmental conditions produced by the strong changes that a place suffers along the months. The algorithms based on sequences of images achieve much better results, as evidenced by the precision-recall curves obtained by SeqSLAM and ABLE-M.

In this case, SeqSLAM has a slightly worse precision than ABLE-M, which is probably due to the changes on the field of view that this dataset has between the different sequences. This is because the performance of the image description method applied by of SeqSLAM has a certain dependence on the field of view, as demonstrated in [4]. However, ABLE-M does not have this problem due to the characteristics of the LDB descriptor, which applies a multi-resolution description that alleviates this dependence on the field of view.

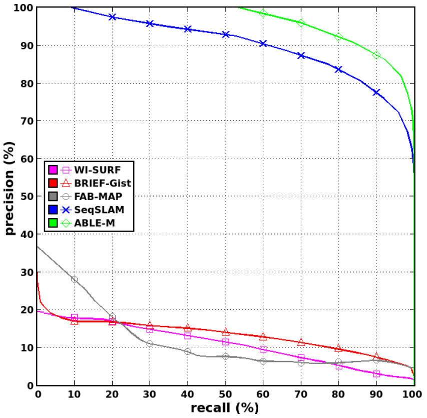


Fig. 8. Precision-recall curves comparing the performance of ABLE-M against some of the main state-of-the-art algorithms between the sequences of the CMU-CVG Visual Localization dataset corresponding to the car rides registered on 01/09/2010 and 21/12/2010.

D. Along the seasons: Nordland dataset

The Nordland dataset is probably the longest (≈ 3000 km) and one of the most challenging datasets that can be currently used for life-long visual topological localization evaluation. It contains four videos with very strong seasonal appearance changes in places for a same train ride, apart from other problematic situations, such as the depicted in Fig. 9.

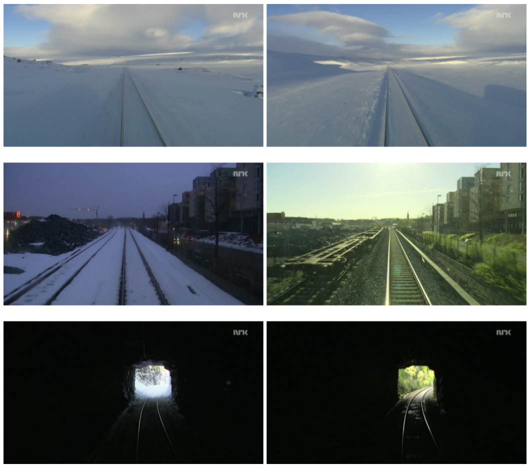


Fig. 9. Three representative examples of difficult situations for place recognition in the Nordland dataset. From top to bottom, each pair of images represents: a) Perceptual aliasing between different places with similar appearances. b) Sunlight conditions at the same hour depending on the season. c) Not much appearance information in places such as tunnels.

In Section V-C, we showed comparative results between ABLE-M and some of the main state-of-the-art methods for the CMU-CVG Visual Localization dataset in two sequences recorded on 01/09/2010 (summer) and 21/12/2010 (winter). For this reason, we present now a comparative between the two videos of the Nordland dataset corresponding to the remaining seasons (spring and fall). As depicted in Fig. 10, these results corroborate again the better effectiveness of the approaches based on sequences of images. Besides, the difference between the results of SeqSLAM and ABLE-M is reduced with respect to the obtained in Section V-C. This is due to the influence of the static field of view in the Nordland dataset, which is beneficial for SeqSLAM.

In addition, Figs. 11 and 12 show precision-recall curves and examples of distance matrices obtained by ABLE-M for the six possible combinations between the different video sequences of the Nordland dataset. As expected, when the winter sequence is evaluated, the effectiveness of our method decreases due to the extreme changes that this season causes in places appearance because of environmental conditions such as snow, illumination, vegetation changes, etc.

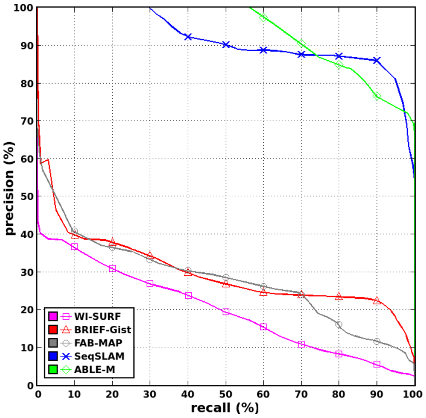


Fig. 10. Precision-recall curves comparing the performance of ABLE-M against some of the main state-of-the-art algorithms in the Nordland dataset (spring vs fall).

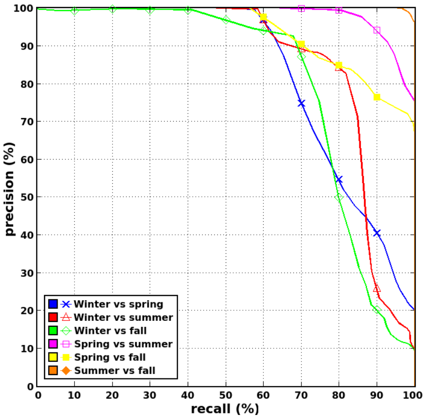
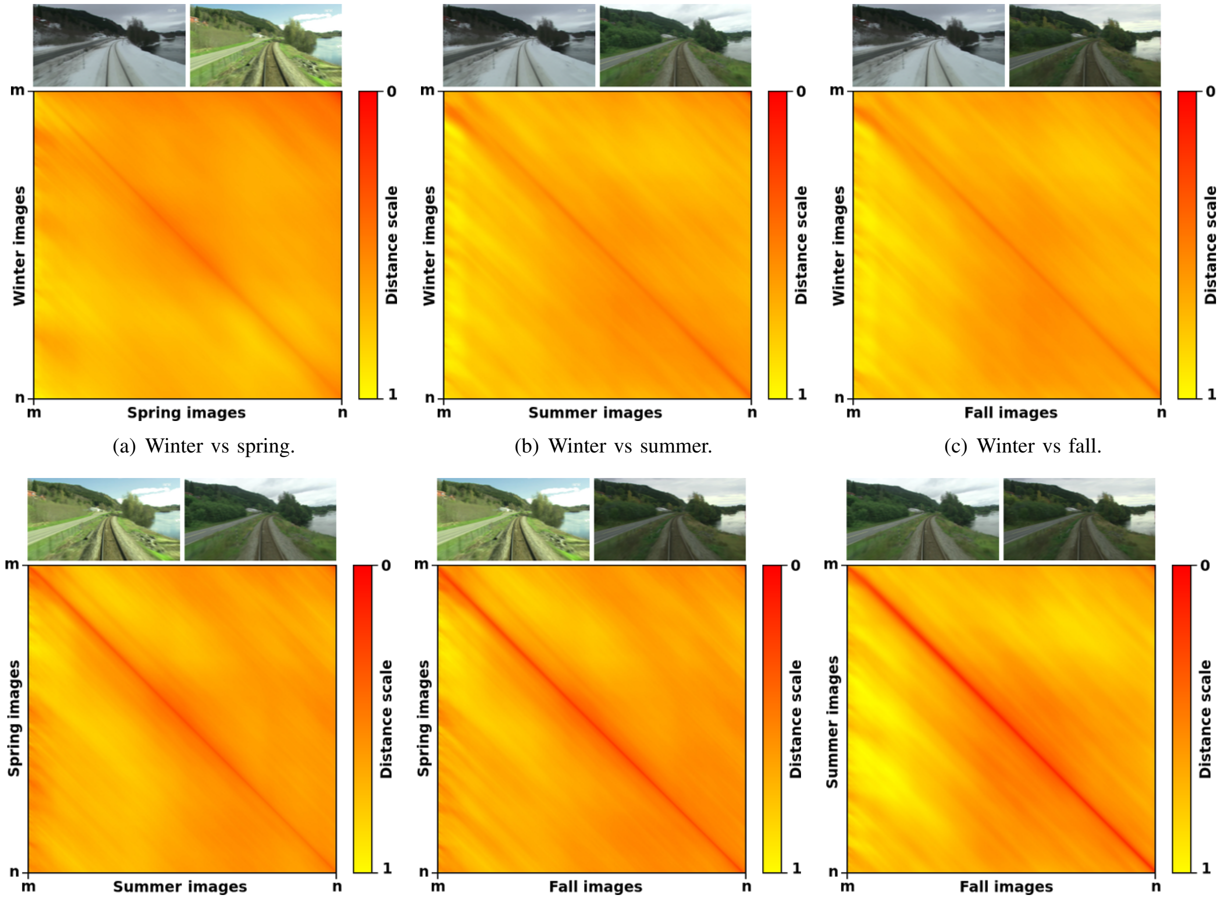


Fig. 11. Precision-recall curves comparing the performance of ABLE-M between the Nordland sequences corresponding to each season.



(a) Winter vs spring.

(b) Winter vs summer.

(c) Winter vs fall.

(d) Spring vs summer.

(e) Spring vs fall.

(f) Summer vs fall.

Fig. 12. An example of the distance matrices obtained by ABLE-M between the Nordland sequences corresponding to each season, jointly with image samples of the loop closures detected. It must be noted that the distance matrices correspond to the representative frames between m=200000 and n=201000, because due to the limitations of paper format we can not conveniently show the matrices for the full dataset. As can be seen, the winter sequence is the most problematic because of the strong appearance changes suffered by the places in this season, such as snow or low illumination. For this reason, the loop closure diagonal which appears in the distance matrices is not so clear when the winter sequence is evaluated against the other ones.

E. Discussion

In previous results sections we corroborated in four different datasets how ABLE-M can cope with the main problems of carrying out a robust visual topological localization along time. Furthermore, our proposal also obtains remarkable results compared to state-of-the-art algorithms such as WI-SURF, BRIEF-Gist, FAB-MAP or SeqSLAM.

As main comments about the performance of ABLE-M, in Sections V-C and V-D is evidenced that it achieves great results for place recognition and loop closure detection along the months and seasons, which nowadays is one of the main goals of the robotics community in this research line. In spite of the difficult conditions for recognizing places between day and night that reduce the effectiveness of our method with respect to other life-long situations, it can be observed how in the precision-recall curves shown in Section V-B we obtain acceptable results, which are comparable to the presented by SeqSLAM in [3] for the Alderley dataset. In fact, the usage of illumination invariance clearly improves ABLE-M results for visual localization at different sun hours, as exposed in Section V-A.

However, the advantages of our proposal are not only in its effectiveness, but also in the efficiency provided by ABLE-M. As explained in Section III, the usage of binary sequences reduces the memory resources and the processing time needed for carrying out a robust place description and matching compared to other alternatives such as vector-based descriptors or the method based on image difference vectors applied by SeqSLAM. Additionally, in our case the application of FLANN using a multi-probe LSH instead of a linear search also decreases the accumulated computational cost of matching the sequences of images with the previously processed to a sublinear time, as deduced from Table II. In tests, we use a standard Intel Core i7 2,40 GHz PC.

TABLE II

COMPARISON OF AVERAGE PROCESSING TIMES FOR IMAGE MATCHING.

No. images to match	**SeqSLAM**	**ABLE-M (without FLANN)**	**ABLE-M (with FLANN)**
1000	0.177 s	0.023 s	0.0093 s
10000	1.81 s	0.25 s	0.17 s
100000	18.23 s	2.53 s	0.42 s

VI. Conclusions

The approach presented in this paper (ABLE-M) has proved that it can successfully accomplish a life-long visual localization based on an efficient matching of binary sequences from monocular images. The performance of our method has been satisfactorily tested in extensive evaluations carried out in four challenging datasets where different situations are analyzed, such as recognizing places at different times of day or night, along the months or across all four seasons. Furthermore, our proposal has confirmed that is able to perform a robust life-long visual topological localization compared to some of the main state-of-the-art methods, such as WI-SURF, BRIEF-Gist, FAB-MAP or SeqSLAM.

One of the main contributions of this work resides in the implementation of an image description method based on binary strings extracted from sequences of images instead of single images, which is a concept that decisively improves the effectiveness of ABLE-M in life-long visual localization, as supported by the results presented along this paper. Besides, our work also contributes other interesting and useful ideas, such as the application of an illumination invariant transformation of images before performing the binary description or the efficient matching of binary sequences based on the Hamming distance and the usage of FLANN for fast nearest neighbor search.

As an extra contribution to the computer vision and robotics communities, an open version of the code implemented for our proposal will be downloadable after publication for free use. This open source toolbox for life-long visual topological localization will be referred with the name OpenABLE[1].

References

[1] M. Cummins and P. Newman, "FAB-MAP: Probabilistic localization and mapping in the space of appearance," *International Journal of Robotics Research (IJRR)*, vol. 27, no. 6, pp. 647–665, June 2008.

[2] ——, "Highly scalable appearance-only SLAM - FAB-MAP 2.0." in *Robotics Science and Systems Conference (RSS)*, June 2009.

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

[4] N. Sünderhauf, P. Neubert, and P. Protzel, "Are we there yet? Challenging SeqSLAM on a 3000 km journey across all four seasons," in *Workshop on Long-Term Autonomy at the IEEE International Conference on Robotics and Automation (W-ICRA)*, May 2013.

[5] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, J. J. Yebes, and S. Gámez, "Bidirectional loop closure detection on panoramas for visual navigation," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2014, pp. 1378–1383.

[6] 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 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 2014, pp. 3089–3094.

[7] X. Yang and K. T. Cheng, "Local difference binary for ultrafast and distinctive feature description," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, no. 1, pp. 188–194, January 2014.

[8] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 36, no. 11, November 2014.

[9] A. J. Glover, W. Maddern, M. Milford, and G. F. Wyeth, "FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2010, pp. 3507–3512.

[10] H. Badino, D. F. Huber, and T. Kanade, "Real-time topometric localization," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2012, pp. 1635–1642.

[11] W. Churchill and P. Newman, "Experience-based navigation for long-term localisation," *International Journal of Robotics Research (IJRR)*, vol. 32, no. 14, pp. 1645–1661, December 2013.

[12] E. Johns and G. Yang, "Dynamic scene models for incremental, long-term, appearance-based localisation," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 2731–2736.

[13] A. Bansal, H. Badino, and D. Huber, "Understanding how camera configuration and environmental conditions affect appearance-based localization," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2014, pp. 800–807.

[14] M. Labbé and F. Michaud, "Appearance-based loop closure detection for online large-scale and long-term operation," *IEEE Transactions on Robotics (TRO)*, vol. 29, no. 3, pp. 734–745, June 2013.

[15] J. Wu, H. Zhang, and Y. Guan, "An efficient visual loop closure detection method in a map of 20 million key locations," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 861–866.

[16] C. Valgren and A. J. Lilienthal, "Incremental spectral clustering and seasons: Appearance-based localization in outdoor environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2008, pp. 1856–1861.

[17] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, June 2008.

[18] P. Neubert, N. Sünderhauf, and P. Protzel, "Appearance change prediction for long-term navigation across seasons," in *European Conference on Mobile Robotics (ECMR)*, September 2013, pp. 198–203.

[19] E. Johns and G. Yang, "Feature co-occurrence maps: Appearance-based localisation throughout the day," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 3212–3218.

[20] Y. Liu and H. Zhang, "Visual loop closure detection with a compact image descriptor," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2012, pp. 1051–1056.

[21] N. Sünderhauf and P. Protzel, "BRIEF-Gist - Closing the loop by simple means," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, September 2011, pp. 1234–1241.

[22] M. Calonder, V. Lepetit, M. Özuysal, T. Trzcinski, C. Strecha, and P. Fua, "BRIEF: Computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 7, pp. 1281–1298, July 2012.

[23] C. McManus, W. Churchill, W. Maddern, A. Stewart, and P. Newman, "Shady dealings: Robust, long-term visual localisation using illumination invariance," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 901–906.

[24] W. Maddern, A. Stewart, and P. Newman, "LAPS-II: 6-DoF day and night visual localisation with prior 3D structure for autonomous road vehicles," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2014, pp. 330–337.

[25] M. Milford, "Visual route recognition with a handful of bits," in *Robotics Science and Systems Conference (RSS)*, July 2012.

[26] G. Bradski, "The OpenCV library," *Dr. Dobb's Journal of Software Tools (DDJ)*, vol. 25, no. 11, pp. 122–125, November 2000. [Online]. Available: http://opencv.org

[27] Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li, "Multi-probe LSH: Efficient indexing for high-dimensional similarity search," in *International Conference on Very Large Data Bases (VLDB)*, September 2007, pp. 950–961.

[28] G. H. Lee and M. Pollefeys, "Unsupervised learning of threshold for geometric verification in visual-based loop-closure," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 1510–1516.

[29] A. J. Glover, W. Maddern, M. Warren, S. Reid, M. Milford, and G. F. Wyeth, "OpenFABMAP: An open source toolbox for appearance-based loop closure detection," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2012, pp. 4730–4735.

[1] Available from: http://www.robesafe.com/personal/roberto.arroyo/