Noname manuscript No.
(will be inserted by the editor)

Are you ABLE to perform a life-long visual topological localization?

Roberto Arroyo[1] · **Pablo F. Alcantarilla**[2] · **Luis M. Bergasa**[1] · **Eduardo Romera**[1]

Received: date / Accepted: date

Abstract Visual topological localization is a process typically required by varied mobile autonomous robots, but it is a complex task if long operating periods are considered. This is because of the appearance variations suffered in a place: dynamic elements, illumination or weather. Due to these problems, long-term visual place recognition across seasons has become a challenge for the robotics community. For this reason, we propose an innovative method for a robust and efficient life-long localization using cameras. In this paper, we describe our approach (ABLE), which includes three different versions depending on the type of images: monocular, stereo and panoramic. This distinction makes our proposal more adaptable and effective, because it allows to exploit the extra information that can be provided by each type of camera. Besides, we contribute a novel methodology for identifying places, which is based on a fast matching of global binary descriptors extracted from sequences of images. The presented results demonstrate the benefits of using ABLE, which is compared to the most representative state-of-the-art algorithms in long-term conditions.

Keywords Localization across seasons · Visual place recognition · Loop closure detection · Image matching · Binary descriptors

✉ Roberto Arroyo
roberto.arroyo@depeca.uah.es

Pablo F. Alcantarilla
palcantarilla@irobot.com

Luis M. Bergasa
luism.bergasa@uah.es

Eduardo Romera
eduardo.romera@depeca.uah.es

[1] Department of Electronics, University of Alcalá (UAH), Alcalá de Henares, 28871, Madrid, Spain.

[2] iRobot Corporation, 10 Greycoat Place, Victoria, London, UK.

1 Introduction

Autonomous robots and intelligent vehicles commonly need robust localization methods with the aim of correctly detecting its position in the real world and performing an accurate navigation based on this information. In the last decades, different methodologies derived from generic Simultaneous Localization and Mapping (SLAM) algorithms have been extensively studied in order to solve the important challenges of the localization problem (Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006). There are several sensing technologies traditionally applied in these localization systems, such as GPS-based or range-based, among others. However, the robotics community has also considered camera-based solutions as an interesting alternative in the last years. For this reason, visual localization systems have been broadly extended within the recent past due to the improvements in camera features, price and size, added to the progress in computer vision algorithms for techniques such as visual SLAM (Fuentes-Pacheco et al, 2012; Alcantarilla et al, 2013) or visual odometry (Scaramuzza and Fraundorfer, 2011; Fraundorfer and Scaramuzza, 2012).

Place recognition is commonly a key stage in different visual localization methods, because it provides valuable information about the situational awareness of the traversed environment. Besides, it is typically used for detecting loop closures and identifying revisited places in order to correct the accumulated localization error in vision-based navigation systems. As stated in Williams et al (2009), the algorithms for solving the loop closure problem in visual localization are divided into three groups: map-to-map (metric) (Clemente et al, 2007), image-to-map (topometric) (Williams et al, 2008) and image-to-image (topological) (Cummins and Newman, 2008).

Topological methods for visual localization have been popularized after the formulation of FAB-MAP (Cummins and Newman, 2008), which allows to recognize places by using only the space of visual appearance. However, FAB-MAP requires a prior training phase and applies a computationally expensive approach that requires feature extraction followed by probabilistic inference, which can make the proposal not suitable for real-time applications. Additionally, visual localization is complex in long operating periods due to the strong appearance changes that a place suffers by cause of dynamic elements, illumination, weather or seasons, as can be observed in the examples presented in Fig. 1. For this reason, life-long visual topological localization has been one of the most challenging topics in robotics over the last years, where proposed solutions not only need to solve the problems associated with changing environments, but also to apply efficient algorithms with low computational cost that can work in real scenarios. According to the described requirements, Fig. 1 depicts the general diagram of our proposal called ABLE (Able for Binary-appearance Loop-closure Evaluation).

We have already presented some preliminary studies related to our approach in different international conferences in robotics and autonomous vehicles (Arroyo et al, 2014a,b, 2015). In this paper, we describe our final work, where the complete proposal of ABLE is explained. In addition, we include new contributions and results to validate the versatility and effectiveness of our solution:

- Different versions of ABLE are proposed depending on the type of camera for providing a higher adaptability and taking advantage of the additional image information that can be obtained in each case: monocular (ABLE-M), stereo (ABLE-S) or panoramic (ABLE-P).
- Global binary features are applied in image description jointly with a matching based on the Hamming distance and an Approximate Nearest Neighbors (ANN) search, that provides both low processing times and high precision rates.
- Sequences of images are used instead of single images in all the cases with the aim of carrying out a better recognition of places in long-term scenarios, as introduced in some state-of-the-art works such as Milford and Wyeth (2012).
- An illumination invariant transformation is previously performed in order to minimize the problems related to changing lighting conditions and shadows in a visual place recognition context, inspired by innovative proposals such as Upcroft et al (2014); McManus et al (2014).

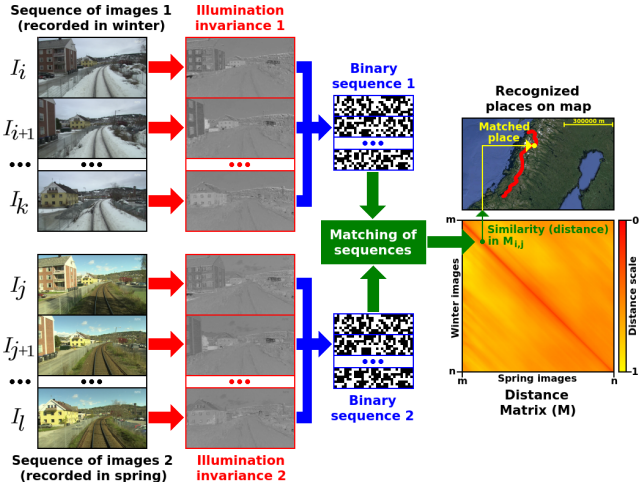


Fig. 1 General diagram of our life-long visual topological localization system. This representation shows monocular images (ABLE-M), but along this paper we also describe more detailed diagrams about our stereo (ABLE-S) and panoramic (ABLE-P) versions.

Apart from some of the previously mentioned approaches such as FAB-MAP, there are other remarkable proposals for visual topological localization in the state of the art, which are discussed along Section 2. In Section 3, we introduce and explain the concept of the binary descriptors applied by our algorithm for the description and matching of places. In Section 4, our final method and its different versions are extensively described. In Section 5, we define an objective evaluation methodology for visual place recognition and loop closure detection, where the tests are carried out in several public datasets with different characteristics recorded in varied long-term situations. In Section 6, a wide set of new results are presented and compared to the main state-of-the-art algorithms with the aim of validating our complete proposal. Finally, the main conclusions about this work and future research are discussed in Section 7.

2 Related Work

Although FAB-MAP can be considered as the milestone in works related to visual topological localization, the initial researches in these kinds of techniques were started some years before (Ulrich and Nourbakhsh, 2000). Furthermore, a great number of novel approaches have been presented following the research line started by FAB-MAP, as studied in surveys such as Garcia-Fidalgo and Ortiz (2015) or Lowry et al (2016). In fact, the authors of FAB-MAP also tested their algorithm over 1000 km in more recent papers (Cummins and Newman, 2010a,b), which is probably one of the first robust approaches to life-long visual topological localization in the literature. In addition, a 3D implementation of

FAB-MAP (Paul and Newman, 2010) was also contributed to incorporate geometric information, but in this case it was only tested in short-term localization.

One of the most relevant proposals recently contributed for visual localization in long-term scenarios is SeqSLAM (Milford and Wyeth, 2012), that introduced the idea of recognizing places as sequences instead of single images, in contrast to previous proposals such as FAB-MAP. SeqSLAM was satisfactorily evaluated in challenging life-long visual localization situations, where a same route was traversed in a sunny summer day and a stormy winter night. However, in Sünderhauf et al (2013) some drawbacks of SeqSLAM were revealed, such as the field of view dependence and the influence of parameters configuration. As will be explained along this paper, these problems have been ameliorated by ABLE and other recent approaches that specifically study the difficulties of a changing viewpoint (Pepperell et al, 2014; Lowry and Milford, 2015).

Nowadays, the recognition of previously visited places along the different seasons of the year is the most challenging topic in life-long visual localization due to the difficulties associated with this task: changes in vegetation, illumination, weather, dynamic elements, etc. For this reason, several proposals have been presented within the recent past for facing these problems related to seasonal changes. In Neubert et al (2015), a novel algorithm for place recognition based on appearance change prediction was tested in the Nordland dataset (Sünderhauf et al, 2013), where a same route of more than 750 km is traversed by a train four times (one in each season). This public dataset is one of the used in our tests, and it has been also employed for evaluation in other works, such as Mohan et al (2015), where co-occurrence matrices are computed with the aim of improving the precision for matching places in long-term scenarios. In fact, the effectiveness of co-occurrence for place recognition in dynamic scenes had been previously demonstrated by Johns and Yang (2014).

Additionally, other novel techniques have been applied for visual localization tasks in long-term mapping scenarios. Models based on bags-of-words for place recognition have been successfully applied in robotics, such as the designed in Gálvez-López and Tardós (2012). This has been recently employed for improving the performance in loop closure detection for correcting the drift in monocular SLAM systems (Mur-Artal et al, 2015). Other algorithms focused on seasonal changes are based on visual experiences (Dymczyk et al, 2015; Linegar et al, 2015), which are defined as appearance representations of an environment under certain conditions to obtain a visual memory. Besides, new tendencies propose to use pre-trained Convolutional Neural Networks (CNNs) for an accurate place recognition along the time (Sünderhauf et al, 2015). Nevertheless, supervised deep learning techniques require a large amount of manually annotated data for an specific problem at hand and they use to be computationally expensive, as studied in methods focused on topological place learning (Erkent and Bozma, 2015) or loop closure detection for visual SLAM based on CNNs (Gao and Zhang, 2017).

However, embedded robotic systems can reduce memory resources and computational costs using less complex solutions, such as the computation of simplified image representations. In this line, automatic image scaling can be an interesting idea for achieving more efficiency in place recognition for changing environments, as discussed in Pepperell et al (2015). Moreover, some works based on compact scene descriptors have obtained remarkable results in cross-season place recognition, such as Masatoshi et al (2015). Another technique typically used in the last years is the application of global image descriptors, in order to achieve an efficient long-term performance and try to obtain a real-time visual localization. Solutions similar to the proposed approach by BRIEF-Gist (Sünderhauf and Protzel, 2011) are the most representative of this tendency. Recently, a method based on global image signatures has also been published for visual loop closure detection (Negre-Carrasco et al, 2016), which demonstrates the proliferation of these techniques. In our case, ABLE applies a methodology based on a global image description using binary features, whose main characteristics will be explained in depth in Section 3.

Finally, although a great part of the algorithms in the state of the art are designed for monocular cameras, there are some specific approaches that are focused on stereo and panoramic images. On the one hand, stereo information allows to acquire a more complete description of the geometry of an environment, which is exploited in works such as Cadena et al (2010, 2012), where a bag-of-words model is combined with the application of stereo pairs to check a valid spatial transformation in place matching. In our approach, the method defined by ABLE-S computes disparity with a similar purpose. On the other hand, some state-of-the-art algorithms trust in panoramic images for localization across the different seasons of the year (Valgren and Lilienthal, 2010). Additionally, other works also use panoramic views for a more robust loop closure detection (Murillo et al, 2013; Korrapati et al, 2013; Korrapati and Mezouar, 2017). The main advantage of panoramas is that they allow a visual perception of the environment in all the possible orientations, which can be used for detecting places revisited in other direction. For this reason, panoramic images are also employed by ABLE-P to take advantage of the extra visual information provided by them in visual topological localization, as will be justified along some of the following sections of the paper.

3 Binary Descriptors for Visual Localization

The application of binary features for describing places is one of the main characteristics of our life-long visual localization proposal. Before starting the detailed explanation of ABLE, it is necessary to introduce the main properties of this type of descriptors and how they work, with the aim of understanding the main benefits of using them in our approach.

Binary descriptors are typically constructed from a set of pairwise comparisons from a sampling pattern which is normally centered in a point of interest of the image. The sampling pattern differs depending on the specific binary descriptor and it can be adapted for obtaining invariance to scale and rotation. When the descriptor is computed, each bit in the binary feature is the result of precisely one comparison.

Apart from the previous considerations, it must be explained how these binary features are formulated and built. If we define a smoothed image patch ($\mathbf{p}$) centered in the point of interest $\mathbf{x} = (x, y)$, a binary test (τ) is characterized as:

$$\tau(\mathbf{p}; f(i), f(j)) = \begin{cases} 1 & f(i) < f(j) \\ 0 & f(i) >= f(j) \end{cases}, \; i \neq j, \tag{1}$$

where $f(i)$ is a function that returns an image feature response for the point of interest, which is compared to other $f(j)$ for a certain pixel or cell in $\mathbf{p}$. According to this, $f(i)$ can simply be the smoothed image intensity (I) at one pixel location $\mathbf{x_i} = (x_i, y_i)$, as proposed by binary descriptors such as BRIEF (Calonder et al, 2012), which is probably the most popular approach:

$$f(i) = I(\mathbf{x_i}), \tag{2}$$

where $f(i)$ can also be the concatenation of other different binary comparisons, such as averaged image intensities (I_{avg}) and image gradients (G_x, G_y) on a specific cell ($\mathbf{c_i}$) in $\mathbf{p}$, as proposed by other binary features such as LDB (Yang and Cheng, 2014):

$$f(i) = \left\{ I_{avg}(\mathbf{c_i}), G_x(\mathbf{c_i}), G_y(\mathbf{c_i}) \right\}. \tag{3}$$

Furthermore, we defined a new binary descriptor called D-LDB (Arroyo et al, 2014a), which also computes features based on geometric characteristics of the environment in the binary description process. This novel strategy is designed to reduce the effect of different place recognition problems such as perceptual aliasing and to obtain better results in long-term situations. The initial proposal of LDB is improved by our D-LDB descriptor, where several binary comparisons are also applied for averaged disparity information (D_{avg}):

$$f(i) = \left\{ I_{avg}(\mathbf{c_i}), G_x(\mathbf{c_i}), G_y(\mathbf{c_i}), D_{avg}(\mathbf{c_i}) \right\}. \tag{4}$$

As a last step in the procedure for constructing the binary feature, the resulting descriptor ($\mathbf{d}$) is processed as a sequence of n binary tests, where n is also the final dimension of the resultant descriptor, which can be empirically adjusted depending on the system requirements or other constraints:

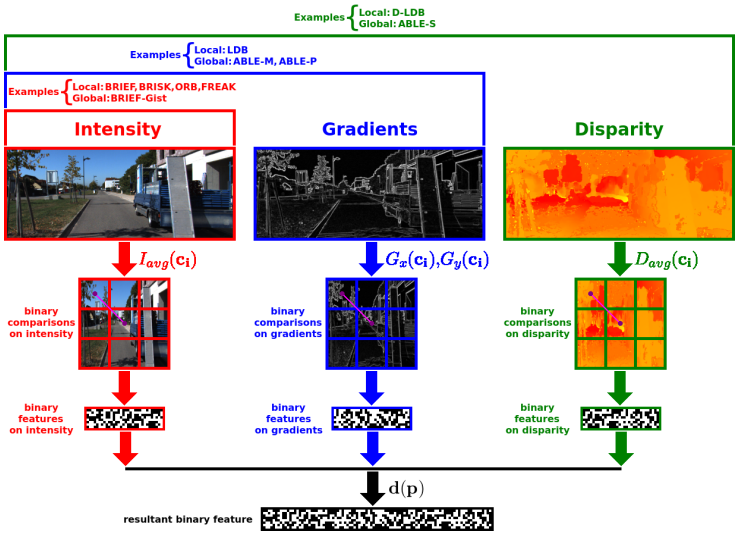
$$\mathbf{d}(\mathbf{p}) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(\mathbf{p}; f(i), f(j)). \tag{5}$$

The definitions previously contributed about the construction of binary features give an idea of their advantages for describing images in an efficient way. Firstly, these descriptors consist of a simple concatenation of bits, which involves a minor memory consumption in general terms, especially if it is compared to descriptors based on vectors of features, such as SIFT (Lowe, 2004) or SURF (Bay et al, 2008). In addition, binary features can be matched using a basic Hamming distance (Muja and Lowe, 2012), which is much more efficient than the traditional way of matching descriptors with the L_2-norm. This efficiency provided by the Hamming distance ($dist_H$) is due to the simplicity of the calculation needed to compute it, which consists on an elementary XOR operation ($\oplus$) and a basic sum of bits:

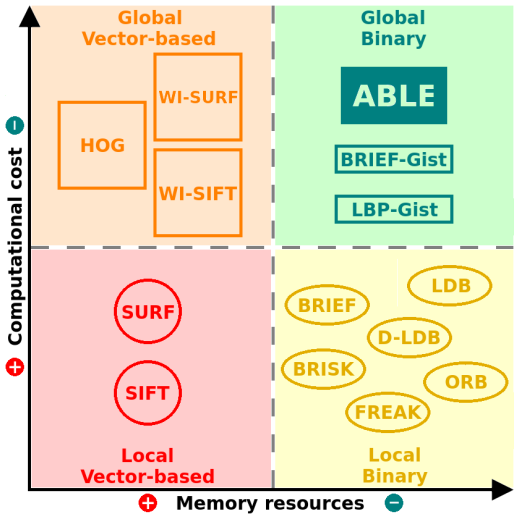
$$dist_H = \text{bitsum}(\mathbf{d}(\mathbf{p}_i) \oplus \mathbf{d}(\mathbf{p}_j)). \tag{6}$$

Due to the mentioned benefits, binary features have been used for describing images in visual localization in some state-of-the-art approaches. In this line, some experiments presented in (Milford, 2012) demonstrated that a handful of bits can be enough for correctly identifying places in a robust way. As a representative example, in works such as Gálvez-López and Tardós (2012), BRIEF is computed as a local binary descriptor for place recognition, where several points of interest are previously detected and the associated local features are extracted from image sequences.

Besides, there are other local binary descriptors typically applied in these kinds of works, such as BRISK (Leutenegger et al, 2011) or ORB (Rublee et al, 2011), which add invariance to rotation and scale to the initial BRIEF formulation, or FREAK (Alahi et al, 2012), which is a keypoint descriptor inspired by the human visual system and based on a retinal sampling pattern. All these local binary descriptors are only focused on the intensity information from images, which can be insufficient to carry out a robust life-long visual localization. For this reason, we use LDB in ABLE, because it also includes gradient comparisons that give a higher descriptiveness power. More specifically, LDB is used in the ABLE-M and ABLE-P versions, but in ABLE-S we compute our D-LDB descriptor in order to take advantage of the valuable information provided by the disparity obtained from stereo images, as shown in Fig. 2(a).



(a) Features used by the different versions of ABLE. Intensity and gradient information are computed by ABLE-M and ABLE-P. Besides, disparity is also included in ABLE-S.



(b) Qualitative classification of description methods typically used in visual localization applications. Vector-based vs Binary / Local vs Global.

Fig. 2 A visual representation about the global binary description performed by ABLE and the differences with respect to the state of the art.

Additionally, it must be noted that ABLE does not apply a local description model, LDB and D-LDB are computed as global binary descriptors in both cases. This approach is computationally more efficient than local methods. Moreover, some state-of-the-art algorithms for place recognition have obtained remarkable results using global description techniques, such as BRIEF-Gist (Sünderhauf and Protzel, 2011), which calculates a global BRIEF descriptor based on the Gist of scenes (Oliva and Torralba, 2006). Other similar proposals provide an acceptable performance too, such as LBP-Gist (Campos et al, 2013) (based on LBP features (Ojala et al, 1996)) or the Gabor-Gist algorithm (Liu and Zhang, 2012). Finally, there are also global descriptors that consist of float-based or vector-based representations of features, such as WI-SIFT and WI-SURF (Badino et al, 2012) or HOG (Dalal and Triggs, 2005), but similarly to the mentioned for the case of local descriptors, they have a less efficient performance compared to a global binary description, as we depict in Fig. 2(b).

4 ABLE

ABLE is a mature research project whose final goal is performing a life-long visual topological localization in a robust manner and trying to hold the maximum efficiency along the time. The progressive evolution achieved during the development of our proposal can be seen in our recent publications (Arroyo et al, 2014a,b, 2015). Hereafter, we will explain the full method including the last contributions, which will be validated with new tests.

In this final proposal, an illumination invariant transformation is applied in the three ABLE versions, with the aim of improving the results in long-term situations, where lighting conditions are extremely variable. Besides, sequences are now computed instead of single images in all the versions, because it allows to enhance the visual situational awareness of places. Apart from this, the matching of binary features has been enhanced for all the cases using an ANN search, which is more efficient for similarity computation. Additionally, the panoramic image matching is improved in ABLE-P with the application of an optimized cross-correlation to associate panoramas. Furthermore, the calculation of disparity is revised in ABLE-S in order to obtain better results when the global D-LDB descriptor is computed.

4.1 Monocular, Stereo or Panoramic Cameras?

As briefly introduced along Section 2, the visual topological localization methods in the literature employ different types of cameras for perceiving the environment: monocular, stereo or panoramic. Each approach has its pros and cons, but the use of one or another camera usually depends on the constraints of the specific application or problem. For this reason, we have developed a solution which can take advantage of the valuable information provided by the images of the different of cameras, with the aim of adapting our algorithm to the best conditions for each case: ABLE-M for

monocular images, ABLE-S for stereo images and ABLE-P for panoramic images. The main characteristics and differences among the three ABLE versions are detailed in Table 1.

Table 1 Differences among the properties of each ABLE version. Qualitative assessments are represented by ⋆⋆⋆ for the best effectiveness and the lowest memory consumption and computational costs.

	ABLE-M	**ABLE-S**	**ABLE-P**
Camera	Monocular	Stereo	Panoramic
Descriptor	Global LDB	Global D-LDB	Global LDB
Matching	Simple corr.	Simple corr.	Cross corr.
Loop closure	Unidirectional	Unidirectional	Bidirectional
Effectiveness	⋆	⋆⋆	⋆⋆⋆
Memory	⋆⋆⋆	⋆⋆	⋆
Computation	⋆⋆⋆	⋆⋆	⋆

The first notable difference is related to the image description methodology used in each case. Although the three versions use a global binary description, ABLE-M and ABLE-P apply a LDB descriptor, while ABLE-S computes D-LDB, with the aim of adding the helpful disparity information provided by stereo cameras to the description process. Disparity calculation was typically performed in the D-LDB descriptor using a stereo matching based on a standard Semi-Global Block Matching (SGBM) (Hirschmuller, 2008), but in this final approach we also implement a stereo matcher that applies ELAS (Geiger et al, 2010) to obtain more precise disparity maps, which improve the effectiveness of ABLE-S in place recognition, as we will demonstrate in the tests presented in Section 6.2.

In addition, there are differences in the way of processing the similarity between places depending on the ABLE version. More specifically, in ABLE-P each panoramic image is divided into subpanoramas, which are matched by using a cross-correlation technique, achieving more accurate similarity distances between a pair of panoramas. The main advantage of using this approach is that bidirectional loop closures can be detected over the panoramic images. This situation appears when a place is revisited in an opposite direction. In these cases, methods as BRIEF-Gist that computes a simple correlation of subpanoramas can not identify the bidirectional loop closures, because they do not take advantage of the association of the different views taking into account the visual perception in several directions provided by the panoramic images, which is solved using our cross-correlation of subpanoramas, as graphically illustrated in Fig. 3. These observations about the benefits of the approach proposed in ABLE-P for place recognition using panoramic images are reinforced with the results presented in Section 6.3.

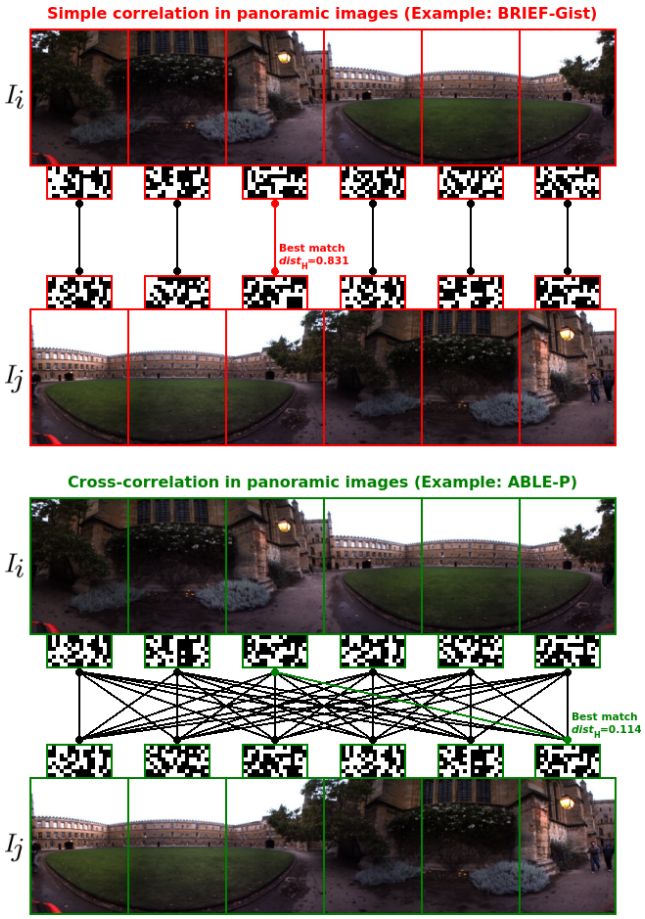


Fig. 3 Differences between a simple correlation and a cross-correlation to associate panoramas. The two panoramic images to be matched in these examples correspond to a place revisited in an opposite direction (a bidirectional loop closure). It can be seen how the best match between subpanoramas, obtained in the cross-correlation applied by ABLE-P, has a much lower distance and is much more similar than using a simple correlation.

Finally, there are other parameters related to the performance of each ABLE version that are also qualitatively described in Table 1. Obviously, the effectiveness of each version in visual topological localization has a great dependence on the amount of information provided by each type of camera: ABLE-P obtains the best precision because it exploits the visual data acquired in all the possible directions, ABLE-S also has a remarkable effectiveness because geometrical information given by stereo cameras is used, but ABLE-M is not as precise as the other two versions because it only employs monocular cameras, which provide a worse visual awareness of the environment. However, ABLE-M has a better efficiency in memory and computational costs due to the processing of a lower amount of data with respect to ABLE-S or ABLE-P. This can be an advantage in systems that require a localization for long operating periods, because in these cases a moderate consumption of resources

is wished. All these qualitative assessments will be quantitative demonstrated in the experiments explained along Section 6.

4.2 Illumination Invariant Transformation of Images

One of the main problems in life-long visual topological localization algorithms is the identification of places when there are important illumination changes on scene. This question has acquired a great interest in some works that try to solve these issues in several difficult situations: zones with shadows (Corke et al, 2013), dynamic lighting environments (Carlevaris-Bianco and Eustice, 2014) or night conditions (Nelson et al, 2015).

We also consider these lighting problems in our final solution. For this reason, ABLE transforms images into an illumination invariant color space, with the aim of refining the description process in these troublesome situations, as it was introduced in other previous approaches such as Upcroft et al (2014); McManus et al (2014), where place recognition is enhanced using this kind of transformation. According to this, our implementation includes an initial stage to obtain illumination invariance, which reduces the difficulties associated with changing lighting conditions, as exposed in Eq. 7:

$$\mathscr{I} = \log(G) - \alpha \cdot \log(B) - (1-\alpha) \cdot \log(R), \tag{7}$$

where R, G, B represent the different color channels of the computed image and $\mathscr{I}$ is the obtained illumination invariant image. As presented in Eq. 8, α is a parameter that is conditioned by the peak spectral responses of each color channel (λ_R, λ_G, λ_B), which are typically available in camera specifications:

$$\frac{1}{\lambda_G} = \frac{\alpha}{\lambda_B} + \frac{(1-\alpha)}{\lambda_R}, \tag{8}$$

where α is a parameter which can be simply determined, as explained in Eq. 9:

$$\alpha = \frac{(\frac{\lambda_B}{\lambda_G} - \frac{\lambda_B}{\lambda_R})}{(1 - \frac{\lambda_B}{\lambda_R})}. \tag{9}$$

For example, the PointGrey Flea2 camera used in datasets such as the KITTI Odometry (Geiger et al, 2012) has $\lambda_R = 610nm$, $\lambda_G = 535nm$, $\lambda_B = 470nm$, so in this case $\alpha = 0.47$. The calculation is analogue for other datasets.

As deduced from all the previous equations, illumination invariant transformation is not an arduous or computationally expensive process, but its application in the visual place recognition algorithm contributes an extra robustness to our method when lighting changes appear, as we shown in our previous work (Arroyo et al, 2015).

4.3 Sequences instead of Single Images

The major part of the traditional state-of-the-art algorithms proposed for visual topological localization are based on the typical assumption that places are defined by a single image. Actually, some of the most popular methods applied in visual loop closure detection follow this philosophy, such as WI-SURF, BRIEF-Gist or FAB-MAP.

Nevertheless, more recent algorithms such as SeqSLAM changed this assumption and considered the idea of identifying places as sequences of images, with the aim of enhancing the situational awareness in long-term conditions. For this reason, our current proposal also follows a similar methodology, because it is more accurate than using single images.

In this case, ABLE extracts binary codes as descriptors of each individual image, but they are concatenated ($+\!\!+$) to build the final binary sequence ($\mathbf{d}$), which corresponds to a sequence of images. This is formulated in Eq. 10, where $k-i$ is equal to $\mathbf{d}_{length}$, which is the length of the sequence considered by the algorithm:

$$\mathbf{d} = \mathbf{d}_{\mathscr{I}_i} +\!\!+ \mathbf{d}_{\mathscr{I}_{i+1}} +\!\!+ \mathbf{d}_{\mathscr{I}_{i+2}} +\!\!+ ... +\!\!+ \mathbf{d}_{\mathscr{I}_{k-2}} +\!\!+ \mathbf{d}_{\mathscr{I}_{k-1}} +\!\!+ \mathbf{d}_{\mathscr{I}_k}. \tag{10}$$

According to this, $\mathbf{d}_{length}$ is an adaptable parameter that mainly depends on the camera frame rate, as will be explained in the experiments carried out along Section 6.

Although some of our previous works (Arroyo et al, 2014a,b) do not take into account this concept, our final approach applies sequences instead of single images in all the versions of ABLE introduced for each type of camera. Due to this, the effectiveness of our method has been improved in life-long scenarios for the three versions, because sequences of images provide more robustness in localization across seasons, as we will show in the results presented in Section 6.1.

4.4 Extraction of Binary Codes

The binary codes that form the final binary sequence, previously introduced in Eq. 10, are extracted from each image using a global LDB descriptor (D-LDB in the case of the stereo images processed by ABLE-S).

The first step before starting the extraction of binary codes is to downsample the acquired images to 64x64 pixels. This size is also applied in the patch ($\mathbf{p}$) considered in the binary description. The reduction of the image size is performed because high resolution images are not required to carry out a robust visual topological localization, as it was evidenced in works such as Milford (2012). Besides, this strategy followed by ABLE allows to reduce memory and computational costs without decreasing precision. Additionally, downsampling the initial images implicates

smoothing and interpolation over neighboring regions that attenuate the negative influence of rotation and scale in place recognition, as stated in Sünderhauf and Protzel (2011).

After the computation of each image patch, the global binary descriptor is extracted by processing the center of the resized image patch as a keypoint without dominant rotation or scale. The resultant binary code is adjusted to a dimension (n) of 32 bytes, which is supported by previous works such as Arroyo et al (2014b). This value of n is fixed using the random bit selection algorithm proposed in Yang and Cheng (2014).

LDB and its derivatives are chosen as the core of our global binary description method because these features provide several advantages with respect to other descriptors. First of all, LDB is not only based on intensity comparisons, like other popular approaches such as BRIEF, as shown in Fig. 2(a). This technique gives more robustness to the description process thanks to the extra gradient information. Apart from this, one of the main benefits provided by LDB is that it computes the features using a multi-resolution scheme, where different grids (2x2, 3x3, 4x4...) are applied to capture information at different granularities, as explained in Yang and Cheng (2014). The application of this multi-resolution approach alleviates the dependence on the field of view suffered in visual place recognition by proposals such as SeqSLAM, as it was exhibited in Sünderhauf et al (2013) and will be confirmed by some of the tests presented in this paper in Section 6.1 (see Fig. 5).

4.5 Matching of Binary Sequences

The similarity or distance between the previously processed binary sequences is efficiently calculated by means of the Hamming distance. The obtained values can be included in a distance matrix (M) for analyzing them in loop closure detection or for evaluation purposes, as explained in Eq. 11:

$$M_{i,j} = M_{j,i} = \text{bitsum}(\mathbf{d}_i \oplus \mathbf{d}_j). \tag{11}$$

In addition, POPCNT is a machine SSE4.2 instruction which can be used for a faster matching of the binary sequences, since it allows to count the total number of bits that are set to one in a more efficient way, as exposed in Muja and Lowe (2012). We take advantage of this instruction for increasing the speed of the simple correlation performed by ABLE-M and ABLE-S in the calculation of similarity, as shown in Eq. 12:

$$M_{i,j} = M_{j,i} = \text{POPCNT}(\mathbf{d}_i \oplus \mathbf{d}_j). \tag{12}$$

Besides, ABLE-P applies the cross-correlation graphically described in Fig. 3, which allows to detect bidirectional loop closures in panoramic images. This is formulated in Eq. 13 and Eq. 14, where similarity is computed for each pair of sub-panoramas (m,n) corresponding respectively to the panoramic images (i,j). The distances between all the sub-panoramas are saved in a preliminary cross-correlation matrix (C), where a minimum is calculated to obtain the final value stored in M:

$$C_{k,l} = C_{l,k} = \text{POPCNT}(\mathbf{d}_{i,m} \oplus \mathbf{d}_{j,n}), \tag{13}$$

$$M_{i,j} = M_{j,i} = \min(C). \tag{14}$$

As a last contribution in our matching method, it must be noted that we apply an ANN (Approximate Nearest Neighbors) search based on the functionalities given by the FLANN library (Muja and Lowe, 2012, 2014). The index used in this search consists on a multi-probe LSH (Local Sensitive Hashing).

This hashing technique is used due to the characteristics of our binary codes, which are keys compatible with this method and give a fast performance for our Hamming matcher. The main idea behind the multi-probe LSH index used in our ANN search is focused on systematically testing multiple binary codes for the image queries in a hash table, whose hash keys may not necessarily be completely identical to the hash value of the query vector. If we consider an image query ($\mathbf{q}$), the applied hash function ($g(\mathbf{q})$) is denoted by the different hash slots (h) which are involved on it:

$$g(\mathbf{q}) = \{h_1(\mathbf{q}), h_2(\mathbf{q}), h_3(\mathbf{q}), ..., h_{k-2}(\mathbf{q}), h_{k-1}(\mathbf{q}), h_k(\mathbf{q})\}, \tag{15}$$

where multi-probe LSH searches for a sequence of a hash perturbation vector (δ_i), formulated as follows:

$$\delta_i = \{\delta_{i_1}, \delta_{i_2}, \delta_{i_3}, ..., \delta_{i_{k-2}}, \delta_{i_{k-1}}, \delta_{i_k}\}, \tag{16}$$

where the algorithm sequentially probes the different hash buckets $\{g(\mathbf{q}) + \delta_i\}$. Finally, a score ($s_i(\mathbf{q})$) is computed to sort the perturbation vectors with the aim of accessing the buckets in order of increasing scores and easily obtaining the searched hash codes. The score is calculated as shown in Eq. 17, where $x_j(\delta_{ij})$ is the distance of $\mathbf{q}$ from the boundary of the slot $h_j(\mathbf{q}) + \delta_j$:

$$s_i(\mathbf{q}) = \sum_{j=1}^{k} x_j^2(\delta_{ij}). \tag{17}$$

As demonstrated in Lv et al (2007), multi-probe LSH achieves the same search quality with a similar time consumption if it is compared to the conventional LSH. However, the difference resides in the number of hash tables, which is reduced in an order of magnitude by multi-probe LSH, and for this reason we chose this method as core of our ANN search.

5 Evaluation

The evaluation of our life-long visual topological localization method is mainly based on analyzing the performance of the different ABLE versions in datasets where place recognition and loop closure detection can be tested for monocular, stereo and panoramic images. We also compare our approach against the main state-of-the-art algorithms in long-term conditions, with the aim of validating the effectiveness of ABLE in an objective way and using a fair evaluation methodology.

5.1 Methodology

The designed methodology for testing ABLE performance is principally based on precision-recall curves, which are calculated from the distance matrices (M) obtained in each test. Before starting evaluation, the distance values contained in M must be normalized by following Eq. 18:

$$M_{i,j} = \frac{M_{i,j}}{\max(M)}. \tag{18}$$

After the previous step, M is thresholded for comparing it against the ground-truth matrix (G) associated with a specific dataset. In this respect, true positives (tp) are contemplated if a positive of the thresholded M matches with a positive of G in a temporal vicinity according to the frame rate. False positives (fp) are considered in the inverse situation, and false negatives (fn) if a negative is found in the thresholded M when a positive should appear. According to these considerations, the values of precision and recall can be calculated as exposed in Eq. 19 and Eq. 20:

$$precision = \frac{tp}{tp+fp}, \tag{19}$$

$$recall = \frac{tp}{tp+fn}. \tag{20}$$

The final precision-recall curve is processed by varying the threshold value (θ) in a uniform distribution between 0 and 1 and computing the associated values of precision and recall in each iteration. In our tests, 100 values of θ are taken in order to obtain well-defined curves.

Apart from evaluation purposes, the distance values registered in M can be used in real application for correcting SLAM or visual odometry errors based on loop closure detection (Caramazana et al, 2016). A threshold (θ) must be applied to discern if the similarity is sufficient to consider a loop closure between two places, as stated in Eq. 21:

$$loop\ \ closure = \begin{cases} true & \text{if } M_{i,j} < \theta \\ false & \text{otherwise.} \end{cases} \tag{21}$$

It must be noted that adaptive thresholds can be also an interesting option for adjusting them according to the evolution of the environment conditions, as studied in some approaches such as Lee and Pollefeys (2014).

5.2 Datasets

The validation of the different ABLE versions is carried out over several publicly available datasets. Concretely, three datasets are used in our tests, where experiments are focused on a specific ABLE variant in each case depending on the type of camera available: ABLE-M in the Nordland dataset (Sünderhauf et al, 2013), ABLE-S in the KITTI Odometry dataset (Geiger et al, 2012) and ABLE-P in the Oxford New College dataset (Smith et al, 2009). The characteristics of each dataset are summarized in Table 2, which presents information about the total length in kilometers for the complete trip, the type of camera used in recordings, image samples shown along the paper, the number of registered images or frames jointly with their properties and a final textual description about each dataset.

In this paper, ABLE is clearly validated in long-term conditions, especially if we consider that a distance near to 3,000 km is traversed over all the tests. Furthermore, several challenging situations appear in the evaluated datasets: seasonal changes, illumination problems, perceptual aliasing, dynamic objects or loop closures. In addition, it must be noted that other datasets focused on life-long localization in monocular images were considered for previous experiments, which can be revised in Arroyo et al (2015), such as the St Lucia dataset (Glover et al, 2010), the Alderley dataset (Milford and Wyeth, 2012) or the CMU-CVG Visual Localization dataset (Badino et al, 2012). However, we consider that in the present paper it is more relevant to provide new results not only focused on several datasets exclusively based on monocular visual localization, but also on the stereo and panoramic cases. Hence, one different dataset is chosen for each type of camera.

5.3 Experiments

The provided tests are focused on comparing the different ABLE variants against the main available state-of-the-art algorithms: WI-SURF, BRIEF-Gist, FAB-MAP and SeqSLAM. For evaluating WI-SURF and BRIEF-Gist, we developed implementations based on the SURF and BRIEF descriptors provided by the OpenCV libraries in version 3.0[1]. OpenFABMAP (Glover et al, 2012) is the toolbox chosen for testing FAB-MAP, which is applied in a standard configuration and properly trained. The experiments for SeqSLAM are performed with the source code provided by OpenSeqSLAM (Sünderhauf et al, 2013), which is implemented in Matlab, with respect to the other methods that are based on C/C++. Additionally, the parameters applied by the different ABLE versions in the tested datasets are summarized in Table 3.

[1] OpenCV is available from: http://opencv.org/

Table 2 Main characteristics of the datasets used in the experiments presented in this paper.

Name	Distance	Main camera	Samples	No. images	Description
Nordland	2,912 km	Monocular (1,920x1,080 px) (25 fps)	See Fig. 1	3,576,688 in 4 seqs.	A train trip across Norway is registered four times, once in each season. Video sequences are synchronized and field of view is always the same. Ground-truth is available from GPS readings.
KITTI Odometry	39.2 km	Stereo (1,226x370 px) (10 fps)	See Fig. 2	44,182 in 22 seqs.	22 sequences recorded across different car routes in Karlsruhe (Germany). Perceptual aliasing is considerable in this dataset. A ground-truth for loops is defined in Arroyo et al (2014a).
Oxford New College	2.2 km	Panoramic (2,048x618 px) (5 fps)	See Fig. 3	8,127 in 1 seqs.	A dataset captured by a robot at the University of Oxford (UK). It contains bidirectional loop closures. Stereo images are also available. A ground-truth matrix is provided in the dataset.

Table 3 Standard parameters applied in the experiments depending on the dataset and the type of camera used in each case.

	ABLE-M	ABLE-S	ABLE-P
Main dataset	Nordland	KITTI Odometry	Oxford New College
Descriptor (**d**)	LDB	D-LDB	LDB
Patch size (**p**)	64x64	64x64	64x64 (each panorama)
Descriptor size (n)	32 bytes	32 bytes	32 bytes (each panorama)
Sequence length ($\mathbf{d}_{length}$)	300 images	120 images	60 images
Alpha (α) $[\lambda_R, \lambda_G, \lambda_R]$	0.46 $[620nm, 540nm, 470nm]$	0.47 $[610nm, 535nm, 470nm]$	0.40 $[610nm, 540nm, 460nm]$
Threshold for loop closure detection (θ)	0.2	0.2	0.2

6 Results

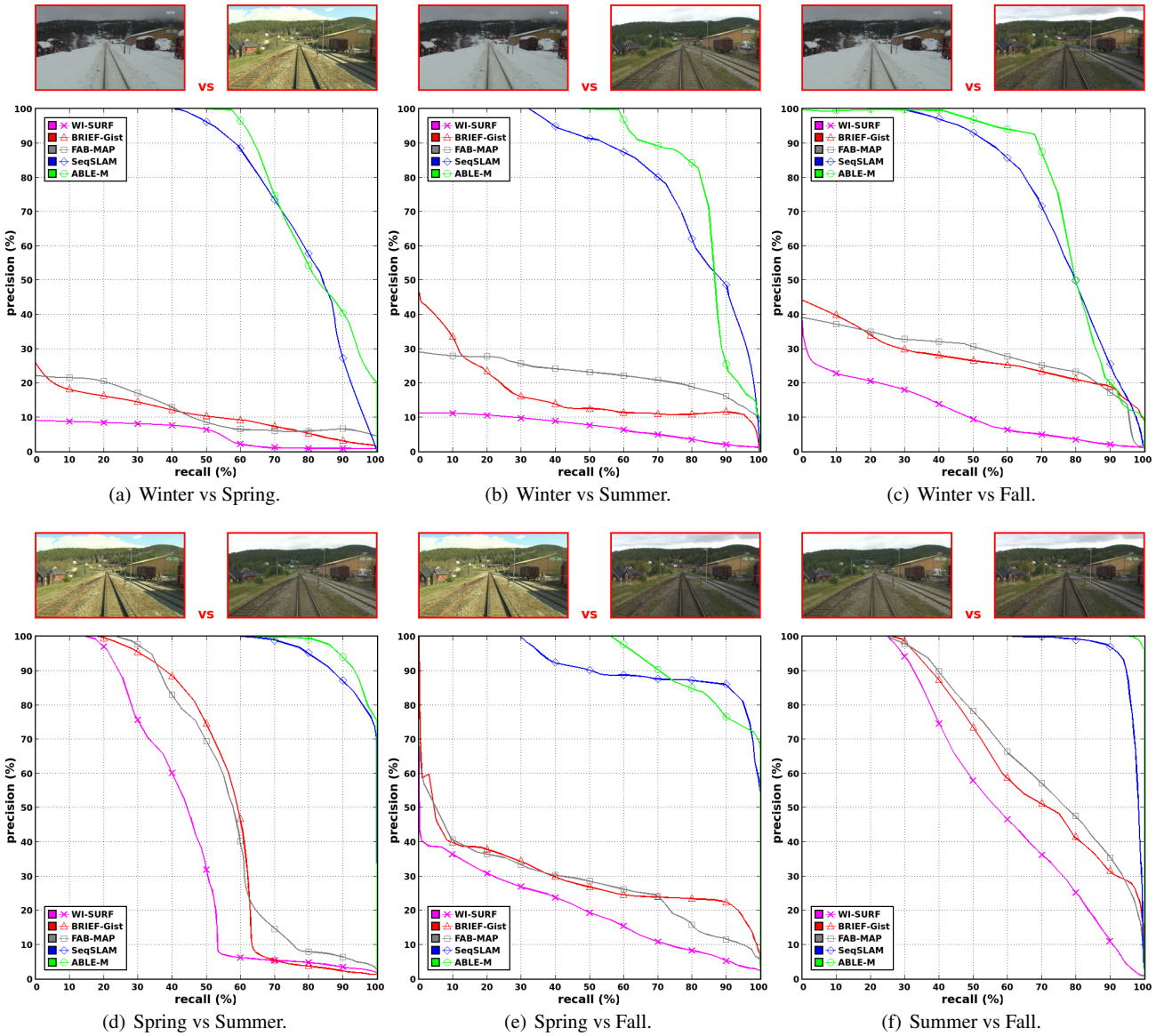
The wide set of results presented along this work is divided into three categories for each type of camera. Moreover, the computational costs are also discussed. These new tests in life-long visual topological localization will corroborate the satisfactory performance and efficiency of our method compared to the state of the art.

6.1 Results for Monocular Cameras

The first results introduced for monocular evaluation are depicted in Fig. 4. A comparison about the accuracy of ABLE-M with respect to the main state-of-the-art approaches is provided for the different seasonal changes contained in the Nordland dataset. Here, it is confirmed how the methods based on single images instead of sequences have a much lower precision in long-term conditions. For this reason, WI-SURF, BRIEF-Gist and FAB-MAP considerably reduce their performance in these cases, especially if their effectiveness is compared to the obtained by SeqSLAM and ABLE-M in the six tests among all the seasons. The differences between the precision of ABLE-M and SeqSLAM are not so significant, but it can be seen that our method has a slightly higher performance in the major part of the cases. Besides, it is also remarkable that the winter sequence is the most troublesome. The accuracy of the algorithms is reduced due to the extreme changes that the appearance of a place suffers when snow covers the scene and illumination is also substantially variable.

The Nordland dataset is chosen for monocular tests because it is the longest ($\approx$ 3,000 km) and most challenging dataset currently available to the best of our knowledge. However, it has a characteristic that can be advantageous for the precision of some place recognition algorithms: images are recorded with a static camera mounted on a train, which always offers the same point of view. This property makes this dataset propitious for the adequate performance of methods such as SeqSLAM, which has a great sensitivity to changes on the field of view. This was proven in Sünderhauf et al (2013), where artificial changes in the field of view of images are introduced in some tests described in that paper. We decide to process new experiments supporting this assumption, with the aim of conducting an evaluation as fair as possible of our method. The associated results are presented in Fig. 5, where the following changes on the field of view are tested: a translation of a 10%, a rotation of a 10% and a combination of both. We perform these evaluations between the sequences of winter and spring, because this is the worst case for all the algorithms, as shown in Fig. 4(a).

According to Fig. 5, the changes on the field of view (especially rotations) negatively affect to SeqSLAM and ABLE-M, mainly if their curves are compared to the obtained in Fig. 4(a). However, it is also evident that the performance of SeqSLAM decreases much more than the achieved by our approach in this case. The best behavior



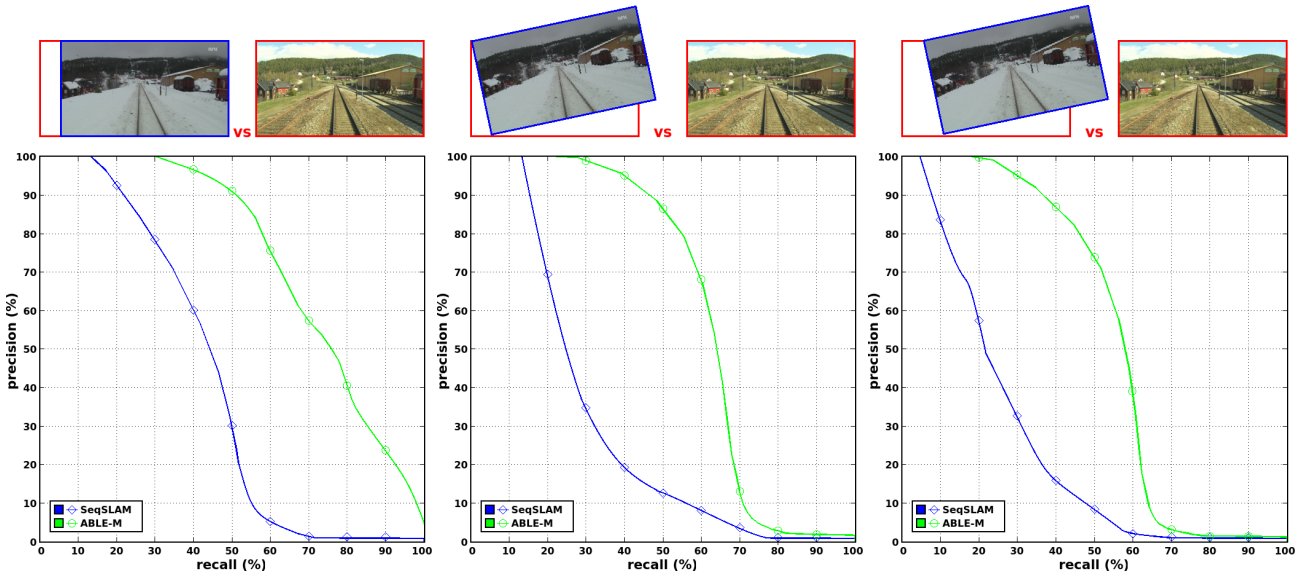
(a) Winter vs Spring. (b) Winter vs Summer. (c) Winter vs Fall.

(d) Spring vs Summer. (e) Spring vs Fall. (f) Summer vs Fall.

Fig. 4 Precision-recall curves for comparing the performance of ABLE-M against the main state-of-the-art methods in the complete Nordland dataset. The four video sequences corresponding to each season are evaluated between them for all the algorithms. An illustrative frame from the two video sequences matched in each test is shown above the precision-recall graphics, with the aim of visually understanding the complexity of place recognition in each case.

of ABLE-M in these conditions is associated with two of the properties about its description method, introduced in Section 4.4. On the one hand, the multi-resolution scheme applied by the global LDB features used in ABLE-M mitigates the dependence on the field of view, with respect to the approach considered by SeqSLAM, based on simple image difference vectors. On the other hand, the negative effects produced by scale and rotation are also slightly alleviated with the initial image downsampling computed by ABLE-M, due to the benefits provided by smoothing and interpolation over neighboring zones. Finally, it must be noted that the approaches based on single images instead of sequences (WI-SURF, BRIEF-Gist, FAB-MAP) are not represented in Fig. 5, because their low accuracy in long-term conditions was sufficiently evidenced in Fig. 4.

Before ending this section about monocular results, it must be noted that the length of the sequence of images applied by ABLE-M in the Nordland dataset tests is adjusted to $\mathbf{d}_{length} = 300$, which is a value contrasted in experiments performed in our previous work (Arroyo et al, 2015). We use an analogue value for the length of the sequences employed by SeqSLAM in order to carry out an objective comparison. This parameter is proportionally adaptable to the frame rate for other datasets, as deduced from Table 3.



(a) Winter vs Spring (with translation). (b) Winter vs Spring (with rotation). (c) Winter vs Spring (with rotation and translation).

Fig. 5 Precision-recall curves for comparing the performance of ABLE-M against SeqSLAM when translation and rotation effects are introduced in the Nordland dataset (between the sequences of winter and spring). These new tests demonstrate that the proposal presented by ABLE outperforms other algorithms based on sequences of images such as SeqSLAM, especially when the field of view is changing. It must be also noted that both methods decrease its precision with respect to the case of an invariant field of view, but this is much more accentuated in the SeqSLAM results than in the obtained by our final approach.

6.2 Results for Stereo Cameras

The experiments presented for stereo cameras are focused on the KITTI Odometry dataset, which is selected for these tests because it is a consolidated benchmark commonly used in autonomous driving and robotics. In this case, this dataset is not as long as the Nordland dataset chosen for monocular tests, but it contains several challenging situations for place recognition in long-term conditions, such as perceptual aliasing between scenes, dynamic objects in places and a considerable amount of loop closures in the different recorded sequences (defined in the ground-truth presented in Arroyo et al (2014a)).

The first test carried out in the sequence 06 of the KITTI Odometry dataset is related to the performance of our D-LDB features applied in the global description approach proposed by ABLE-S. We check out the precision of other descriptors as core of our method compared to the application of D-LDB on it, as shown in Fig. 6. The descriptors used in this comparison were described in Section 3. They can be grouped into two main categories: vector-based (SIFT, SURF, HOG) and binary (BRIEF, BRISK, ORB, FREAK, LDB, D-LDB).

As deduced from the precision-recall curves presented in Fig. 6, we decided to use LDB and D-LDB as core of our description approach, because they achieve a higher accuracy for solving localization problems, especially if they are compared to other state-of-the-art descriptors. In these results, it is also remarked that our D-LDB features are more effective than LDB for stereo place recognition. This is due to the addition of disparity information in the global description process provided by D-LDB. In this way, spatial information about scene can be better captured, with the aim of solving common life-long localization difficulties, such as perceptual aliasing.

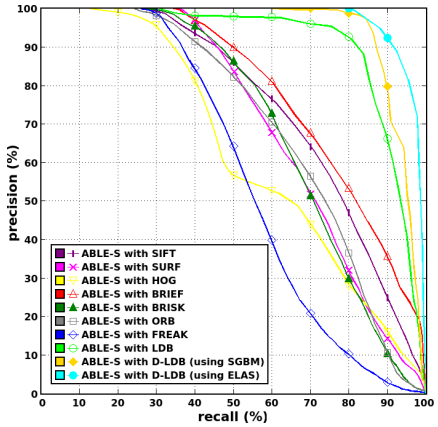
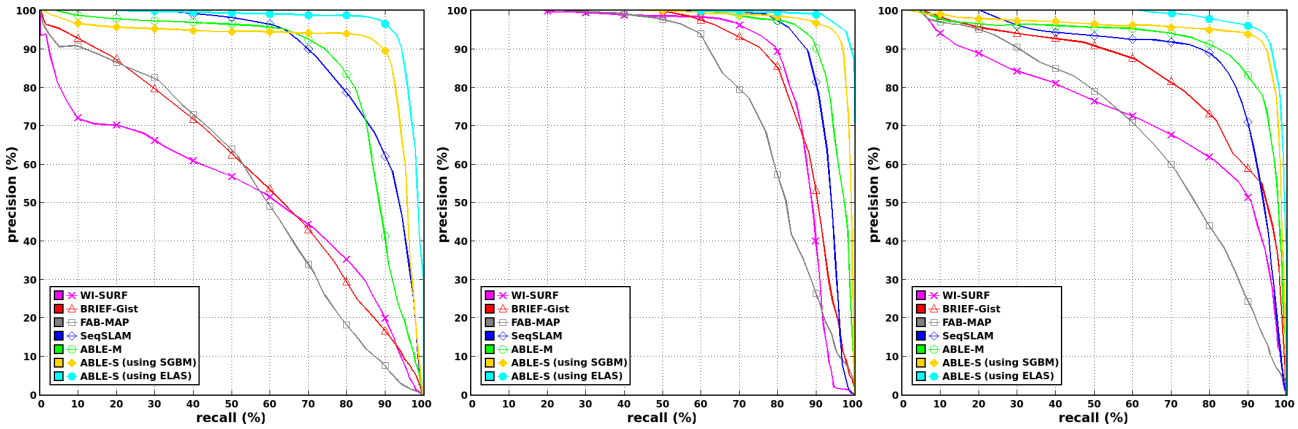
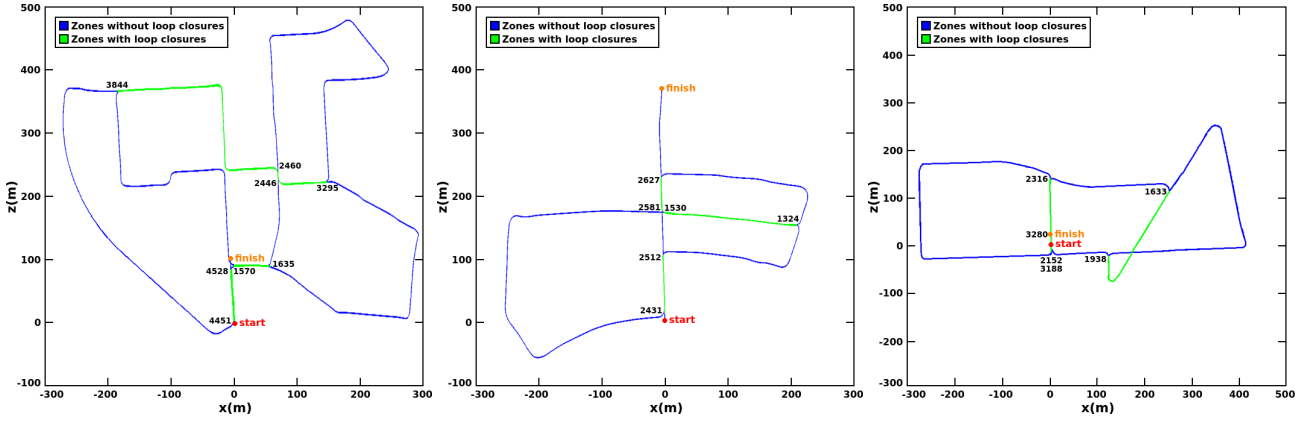


Fig. 6 Precision-recall curves for comparing the performance of ABLE-S using different features as core for global description in the sequence 06 of the KITTI Odometry dataset. D-LDB precision is also evaluated using the different stereo matching methods implemented for disparity calculation.



(a) Sequence 00. (b) Sequence 05. (c) Sequence 13.

Fig. 7 Precision-recall curves for comparing the performance of ABLE-S against the main state-of-the-art methods in the most representative sequences of the KITTI Odometry dataset. ABLE-S is tested using the different implementations of D-LBD depending on the stereo matching method: SGBM or ELAS.



(a) Sequence 00. (b) Sequence 05. (c) Sequence 13.

Fig. 8 Loop closures detected by ABLE-S over the metric maps obtained from the most representative sequences contained in the KITTI Odometry dataset. Image indexes are shown in the zones where loops start and finish to be detected, with the aim of allowing a comparison to the ground-truth presented in Arroyo et al (2014a).

In the evaluations depicted in Fig. 7, the performance of ABLE-S is compared against the main state-of-the-art algorithms for visual topological localization. In this case, we use the three sequences from the KITTI Odometry dataset that contain a higher number of loop closures included in the traversed route, which are the sequences 00, 05 and 13.

The methods based on single images (BRIEF-Gist, WI-SURF, FAB-MAP) obtain better results in Fig. 7 than in the monocular tests showed in Fig. 4, which is due to the minor distance traveled in the sequences of the KITTI Odometry dataset with respect to the Nordland dataset. This is because the influence of using sequences instead of single images is much more evident when a very large amount of kilometers is processed, such as in the case of the Nordland dataset. Apart from this, ABLE also outperforms the SeqSLAM results in all the cases. More specifically, the stereo version (ABLE-S) is more precise than the monocular version (ABLE-M), because of the application of disparity. As said in Section 4.1, a more sophisticated stereo matcher based on ELAS is implemented in our final approach for calculating disparity, which slightly improves the effectiveness of ABLE-S with respect to the precision obtained using the traditional SGBM stereo matcher. These tests demonstrate the importance of the chosen algorithm for disparity computation in D-LDB description, because the usefulness of the spatial information captured by the features has a great dependence on the quality of the applied stereo matching method.

As a last contribution to the stereo tests, Fig. 8 presents the loop closures detected by ABLE-S over the maps of the KITTI Odometry sequences analyzed in Fig. 7. These results support one of the practical applications of our method,

where loop closures can help to identify and correct errors in visual localization tasks. In Fig. 8, loops are depicted when the matched scenes exceed a certain similarity value in the distance matrices, as deduced from Eq. 21 and the explanations about loop closure thresholding given in Section 5. It can be seen how all the loop closures are correctly identified in the maps.

6.3 Results for Panoramic Cameras

The Oxford New College dataset is selected for the last experiments carried out in this paper, which are focused on ABLE-P. This dataset is not very long with respect to the used in the previous tests, but it has been chosen because it allows us to compare ABLE-P against the other two versions (ABLE-M and ABLE-S). This is possible because the dataset is not only recorded with panoramic cameras, but also with stereo.

In Fig. 9, results about the effectiveness of each final ABLE version are contrasted against the obtained by the main state-of-the-art proposals in the Oxford New College dataset. Again, ABLE clearly outperforms the precision yielded by algorithms such as WI-SURF, BRIEF-Gist, FAB-MAP or SeqSLAM. However, the most important conclusions are focused on the comparison between the three ABLE versions. First of all, the difference between the precision-recall curves obtained by ABLE-M and ABLE-S is appreciable. Similarly to the results presented in Section 6.2, in this case the application of stereo information is also decisive to improve the performance achieved by ABLE-M, due to the exploitation of the disparity integrated in D-LDB, which in these tests is definitively calculated using ELAS because of its demonstrated better accuracy. Even so, the most significant results are the provided by ABLE-P. The main reason of its improved performance with respect to the rest of the methods is that ABLE-P is the only algorithm that can detect the bidirectional loop closures appeared along the route, which are one of the most challenging characteristics of the Oxford New College dataset. This is due to the usage of cross-correlation for matching the subpanoramas contained in the panoramic images, which allows to clearly identify the bidirectional loop closures in these cases, as justified in some of the explanations given along Section 4.1 and in Fig. 3.

With the aim of demonstrating how ABLE-P detects the bidirectional loop closures contained in the Oxford New College dataset, Fig. 10 shows a representative part of the distance matrices (M) obtained by the different ABLE versions and the two best state-of-the-art methods in this case (BRIEF-Gist and SeqSLAM), jointly with the ground-truth matrix (G). In the different matrices, unidirectional loop closures appear emphasized as right-side diagonals ($\searrow$) and the bidirectional ones as left-side diagonals ($\swarrow$). According to this, it can be seen that ABLE-P is the only method that represents the bidirectional loop closures in M (see the inferior zone depicted in Fig. 10(f) to check it out). Additionally, it is also appreciable that the distance matrices obtained by ABLE have less noise in the similarity measurements, because perceptual aliasing is reduced thanks to the explained contributions of our proposal.

Finally, the loops identified in Fig. 10(f) by ABLE-P are marked over the corresponding part of the map of the Oxford New College dataset in Fig. 11. M is thresholded by following Eq. 21 to clearly distinguish the similarities associated with a loop closure (see Fig. 11(a)). Progressive representations of the map are shown each time that one of the loops is completed in the route. This valuable information can be applied in localization tasks to correct the accumulated drift commonly appeared in odometry-based systems along the time.

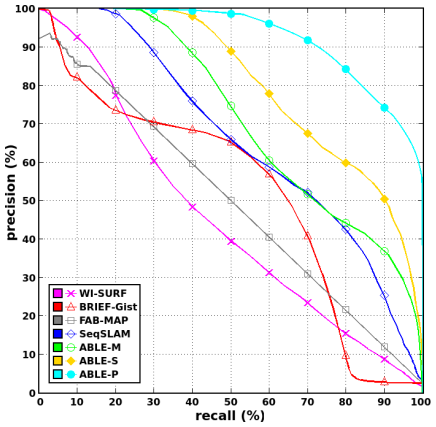
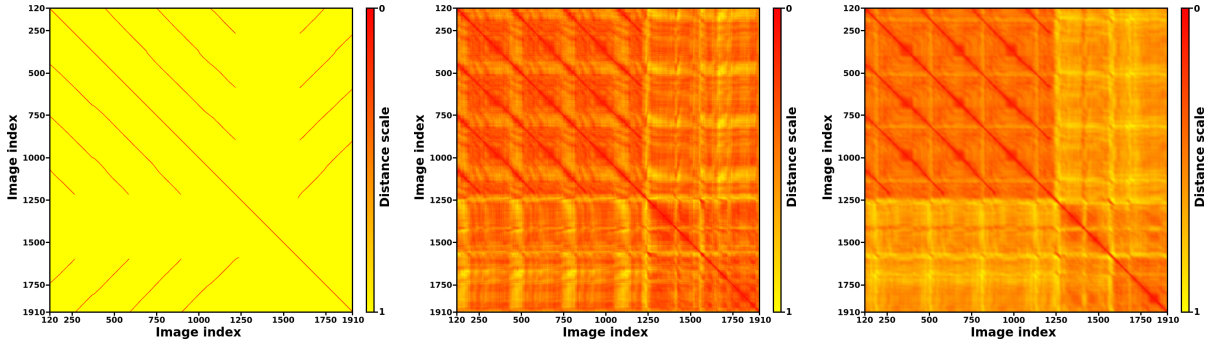
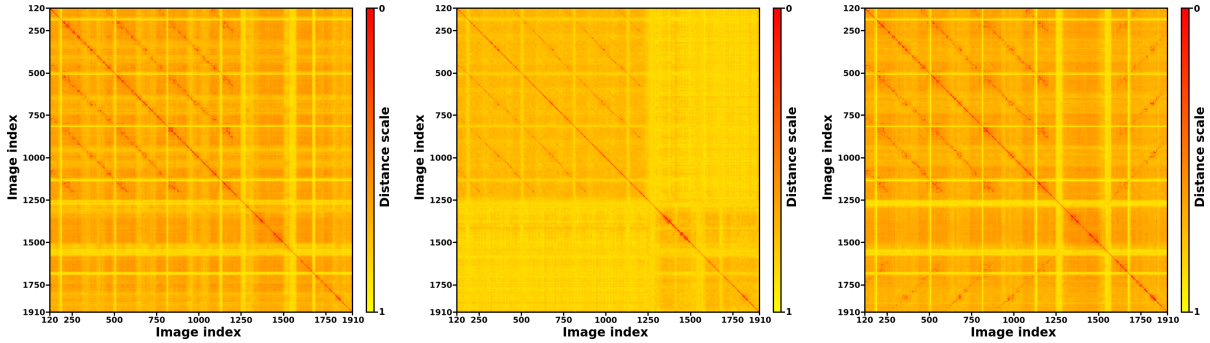


Fig. 9 Precision-recall curves for comparing the performance of the three ABLE versions (ABLE-M, ABLE-S, ABLE-P) against some of the main state-of-the-art methods in the Oxford New College dataset.

Although the results presented in this paper compare the three versions of ABLE in the Oxford New College dataset, it must be noted that we have also carried out some additional tests for evaluating the specific performance of ABLE-P in other datasets based on panoramic images. More specifically, in our previous work (Arroyo et al, 2014b), we conducted some experiments over the Ford Campus dataset (Pandey et al, 2011), where our method also demonstrated its satisfactory performance using panoramas in an environment different to the provided in the Oxford New College dataset.

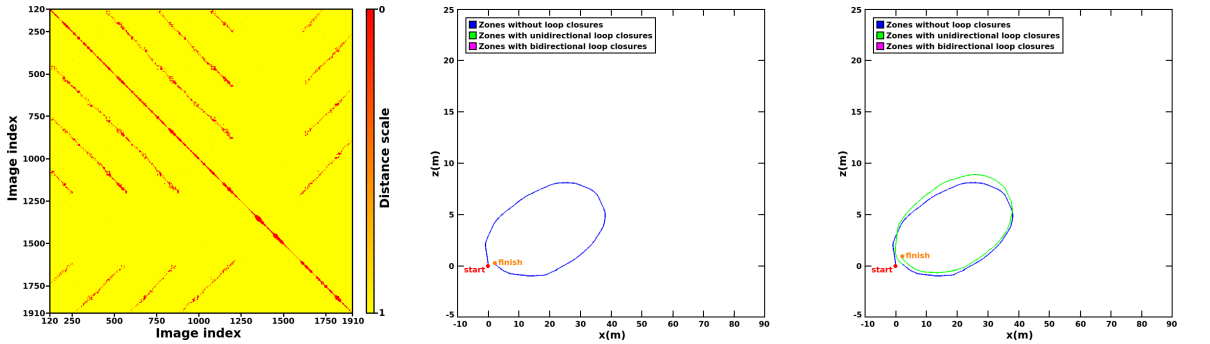


(a) M Ground-truth (G). (b) M using BRIEF-Gist. (c) M using SeqSLAM.

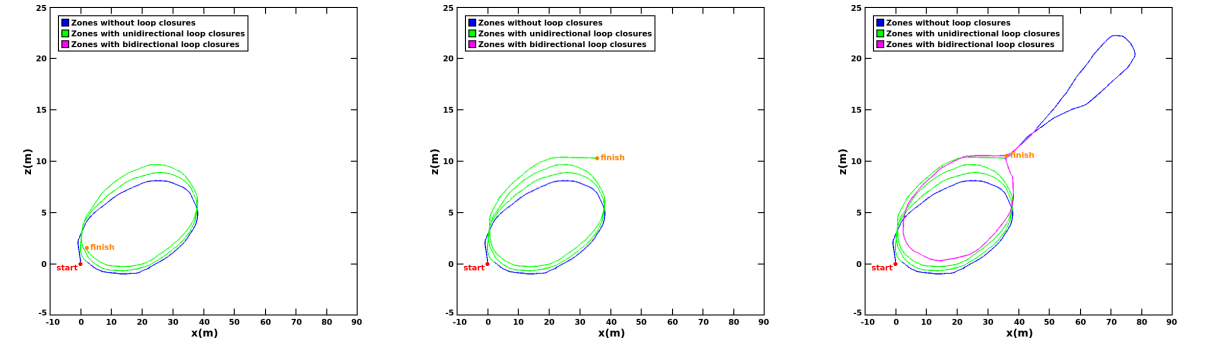


(d) M using ABLE-M. (e) M using ABLE-S. (f) M using ABLE-P.

Fig. 10 Distance matrices for comparing loop closure detection between ABLE versions and state-of-the-art methods in the Oxford New College dataset. We only show a part of M for a representative subset between images 120 and 1,910, because of the limitations of paper format.



(a) Thresholded M using ABLE-P. (b) Map after 1st lap (img. idx. = 449). (c) Map after 1st loop (img. idx. = 752).



(d) Map after 2nd loop (img. idx. = 1,066). (e) Map after 3rd loop (img. idx. = 1,219). (f) Map after 4th loop (img. idx. = 1,910).

Fig. 11 Unidirectional and bidirectional loops progressively detected by ABLE-P over a part of the map in the Oxford New College dataset. We only show a part of the thresholded M for a representative subset between images 120 and 1,910, because of the limitations of paper format.

6.4 Results about the efficiency of ABLE

Apart from the advantages related to the effectiveness of ABLE in visual topological localization across seasons, the efficiency yielded by our method is also important. For this reason, in Fig. 12 we provide a graph where the evolution of the processing times consumed by ABLE-M to match a determined number of images is compared to the achieved by some of the state-of-the-art algorithms. These times have been obtained in tests over images of the Nordland dataset using a standard computer with an Intel Core i7 2.40 GHz processor and a 8 GB RAM.

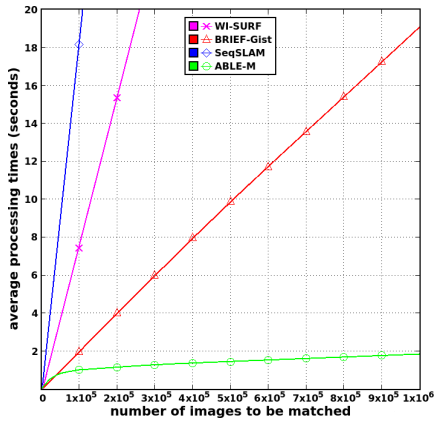


Fig. 12 Comparison between the average processing times of ABLE-M and some of the main state-of-the-art methods for image matching along the time.

In Fig. 12, the evaluation considers a large-scale matching of places until an amount of a million of images. Due to the graph scale requirements, the average processing times are only shown until a maximum of 20 seconds. The curves obtained by SeqSLAM and WI-SURF exceed this time limitation before arriving to the million of images, because these methods have a much higher memory and computational costs in image matching with respect to the other methods, which are based on binary descriptors that can be much faster matched by means of the Hamming distance. In the case of the approaches that apply a binary matching, BRIEF-Gist has a progressive increment of the average processing times, because it is based on a linear search. However, ABLE-M applies an ANN search using a multi-probe LSH index, which decreases the accumulated computational cost of matching the binary sequences with the previously processed to a sublinear time in a large-scale context. For instance, when 100,000 images are matched, the average processing time for BRIEF-Gist is 1.98 *s*, and ABLE-M obtains 0.93 *s*. Nevertheless, in the case of 1,000,000 images, BRIEF-Gist achieves about 19 *s*, while ABLE-M only needs 1.87 *s*, which clearly evidences the high influence in the efficiency of our sublinear search. ANN is used in the three ABLE versions, so the average processing times presented by ABLE-M will be also sublinear for the cases of ABLE-S and ABLE-P. Even so, there are slight differences in the individual processing times depending on the extra information that must be computed by each method. In order to understand these differences, we show Table 4, where the average times per individual image description and matching in each version are included. The description process has a higher computational cost in ABLE-S because of the extra effort that requires the calculation of the disparity. However, the matching costs are the bottleneck of our system. In this sense, the individual costs of a matching between two images are more critical in ABLE-P, because the cross-correlation of panoramas adds an extra computation, as deduced from the times presented in Table 4.

Table 4 Comparison between the average processing times in milliseconds (*ms*) of each ABLE version for describing and matching an individual image in the Oxford New College dataset. We also present times for some of the main state-of-the-art works in order to compare them against our proposals.

	WI-SURF	**BRIEF-Gist**	**SeqSLAM**
Description	0.23 *ms*	0.04 *ms*	0.48 *ms*
Matching	$1.74 \cdot 10^{-3}$ *ms*	$2.35 \cdot 10^{-5}$ *ms*	$3.63 \cdot 10^{-3}$ *ms*
	ABLE-M	**ABLE-S**	**ABLE-P**
Description	0.11 *ms*	7.11 *ms*	0.54 *ms*
Matching	$2.29 \cdot 10^{-5}$ *ms*	$4.51 \cdot 10^{-5}$ *ms*	$5.71 \cdot 10^{-4}$ *ms*

7 Conclusions and Future Works

Along this paper, our proposal for life-long visual topological localization (ABLE[2]) has been extensively justified, jointly with the description of its main contributions, which are validated by a wide set of experiments. The different final versions (ABLE-M, ABLE-S, ABLE-P) constitute a relevant innovation in visual place recognition and loop closure detection fields, which are completely adaptable to several types of cameras and can take advantage of the information acquired by monocular, stereo and panoramic images in each case. Besides, the three versions present in this paper several new characteristics and ideas that have enhanced the performance of our final system with respect to the preliminary versions reported in some of our previous

[2] More information, extra material, videos and open code (Arroyo et al, 2016b) about ABLE are available from the website of the project: http://www.robesafe.com/personal/roberto.arroyo/openable.html

works (Arroyo et al, 2014a,b, 2015). These novelties are the representation of places as sequences of images instead of single images, the illumination invariance, the matching based on an ANN search jointly with LSH or the improvements in the disparity calculation of D-LDB, among others.

Due to the described contributions, ABLE achieves a satisfactory precision in long-term conditions. This is corroborated by the exhibited practical applications and exhaustive results, especially if our method is compared to the main state-of-the-art algorithms, such as WI-SURF, BRIEF-Gist, FAB-MAP or SeqSLAM. The efficiency of our approach is also significant, mainly because of the application of global binary features, which supply an image description methodology with a low computational cost and a fast matching capacity.

In future works, our research line will be extensible to new concepts that could improve even more the accuracy of these techniques in long-term situations, such as localization across seasons. New alternatives recently followed in visual place recognition could be applied, such as the usage of CNNs (Sünderhauf et al, 2015; Arroyo et al, 2016a) or semantic information (Drouilly et al, 2015; Mousavian et al, 2015). In this sense, large amounts of data will be essential to test these approaches, where very recent large-scale datasets (Carlevaris-Bianco et al, 2016) could be interesting to perform new experiments. In addition, more datasets based on fish-eye and panoramic images could be also evaluated in future research, such as the IPDS (Korrapati et al, 2013) and Rawseeds (Ceriani et al, 2009) datasets.

Apart from this, the application of our topological localization proposal in new robotics trends is another area of future interest. More specifically, geometric change detection is a recent topic where these methods could help to improve the current performance (Alcantarilla et al, 2016). In any case, the future of the research line studied along this paper is promising in several robotics fields.

Acknowledgements This work has been funded in part from the Spanish MINECO through the SmartElderlyCar project (TRA2015-70501-C2-1-R) and from the RoboCity2030-III-CM project (Robotica aplicada a la mejora de la calidad de vida de los ciudadanos. fase III; S2013/MIT-2748), funded by Programas de actividades I+D (CAM) and cofunded by EU Structural Funds.

References

Alahi A, Ortiz R, Vandergheynst P (2012) FREAK: Fast retina keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol 2, pp 510–517, DOI 10.1109/CVPR.2012.6247715

Alcantarilla PF, Stasse O, Druon S, Bergasa LM, Dellaert F (2013) How to localize humanoids with a single camera? Autonomous Robots 34(1):47–71, DOI 10.1007/s10514-012-9312-1

Alcantarilla PF, Stent S, Ros G, Arroyo R, Gherardi R (2016) Street-view change detection with deconvolutional networks. In: Robotics Science and Systems Conference (RSS), pp 1–10, DOI 10.15607/RSS.2016.XII.044

Arroyo R, Alcantarilla PF, Bergasa LM, Yebes JJ, Bronte S (2014a) Fast and effective visual place recognition using binary codes and disparity information. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 3089–3094, DOI 10.1109/IROS.2014.6942989

Arroyo R, Alcantarilla PF, Bergasa LM, Yebes JJ, Gámez S (2014b) Bidirectional loop closure detection on panoramas for visual navigation. In: IEEE Intelligent Vehicles Symposium (IV), pp 1378–1383, DOI 10.1109/IVS.2014.6856457

Arroyo R, Alcantarilla PF, Bergasa LM, Romera E (2015) Towards life-long visual localization using an efficient matching of binary sequences from images. In: IEEE International Conference on Robotics and Automation (ICRA), pp 6328–6335, DOI 10.1109/ICRA.2015.7140088

Arroyo R, Alcantarilla PF, Bergasa LM, Romera E (2016a) Fusion and binarization of CNN features for robust topological localization across seasons. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 4656–4663, DOI 10.1109/IROS.2016.7759685

Arroyo R, Alcantarilla PF, Bergasa LM, Romera E (2016b) OpenABLE: An open-source toolbox for application in life-long visual localization of autonomous vehicles. In: IEEE Intelligent Transportation Systems Conference (ITSC), pp 965–970, DOI 10.1109/ITSC.2016.7795672

Badino H, Huber DF, Kanade T (2012) Real-time topometric localization. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1635–1642, DOI 10.1109/ICRA.2012.6224716

Bailey T, Durrant-Whyte H (2006) Simultaneous localisation and mapping (SLAM): Part II State of the art. IEEE Robotics and Automation Magazine (RAM) 13(3):108–117, DOI 10.1109/MRA.2006.1678144

Bay H, Ess A, Tuytelaars T, van Gool L (2008) Speeded-up robust features (SURF). Computer Vision and Image Understanding (CVIU) 110(3):346–359, DOI 10.1016/j.cviu.2007.09.014

Cadena C, Gálvez-López D, Ramos F, Tardós JD, Neira J (2010) Robust place recognition with stereo cameras. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 5182–5189, DOI 10.1109/IROS.2010.5650234

Cadena C, Gálvez-López D, Tardós JD, Neira J (2012) Robust place recognition with stereo sequences. IEEE

Transactions on Robotics (TRO) 28(4):871–885, DOI 10.1109/TRO.2012.2189497

Calonder M, Lepetit V, Özuysal M, Trzcinski T, Strecha C, Fua P (2012) BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 34(7):1281–1298, DOI 10.1109/TPAMI.2011.222

Campos FM, Correia L, Calado JMF (2013) Loop closure detection with a holistic image feature. In: Portuguese Conference on Artificial Intelligence (EPIA), vol 8154, pp 247–258, DOI 10.1007/978-3-642-40669-0_22

Caramazana L, Arroyo R, Bergasa LM (2016) Visual odometry correction based on loop closure detection. In: Open Conference on Future Trends in Robotics (RoboCity16), pp 97–104

Carlevaris-Bianco N, Eustice RM (2014) Learning visual feature descriptors for dynamic lighting conditions. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 2769–2776, DOI 10.1109/IROS.2014.6942941

Carlevaris-Bianco N, Ushani AK, Eustice RM (2016) University of Michigan North Campus long-term vision and lidar dataset. International Journal of Robotics Research (IJRR) 35(9):1023–1035, DOI 10.1177/0278364915614638

Ceriani S, Fontana G, Giusti A, Marzorati D, Matteucci M, Migliore D, Rizzi D, Sorrenti DG, Taddei P (2009) Rawseeds ground truth collection systems for indoor self-localization and mapping. Autonomous Robots 27(4):353–371, DOI 10.1007/s10514-009-9156-5

Clemente LA, Davison AJ, Reid ID, Neira J, Tardós JD (2007) Mapping large loops with a single hand-held camera. In: Robotics Science and Systems Conference (RSS), pp 297–304, DOI 10.15607/RSS.2007.III.038

Corke P, Paul R, Churchill W, Newman P (2013) Dealing with shadows: Capturing intrinsic scene appearance for image-based outdoor localisation. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 2085–2092, DOI 10.1109/IROS.2013.6696648

Cummins M, Newman P (2008) FAB-MAP: Probabilistic localization and mapping in the space of appearance. International Journal of Robotics Research (IJRR) 27(6):647–665, DOI 10.1177/0278364908090961

Cummins M, Newman P (2010a) Accelerating FAB-MAP with concentration inequalities. IEEE Transactions on Robotics (TRO) 26(6):1042–1050, DOI 10.1109/TRO.2010.2080390

Cummins M, Newman P (2010b) Appearance-only SLAM at large scale with FAB-MAP 2.0. International Journal of Robotics Research (IJRR) 30(9):1100–1123, DOI 10.1177/0278364910385483

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol 2, pp 886–893, DOI 10.1109/CVPR.2005.177

Drouilly R, Rives P, Morisset B (2015) Semantic representation for navigation in large-scale environments. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1106–1111, DOI 10.1109/ICRA.2015.7139314

Durrant-Whyte H, Bailey T (2006) Simultaneous localisation and mapping (SLAM): Part I The essential algorithms. IEEE Robotics and Automation Magazine (RAM) 13(2):99–110, DOI 10.1109/MRA.2006.1638022

Dymczyk M, Lynen S, Cieslewski T, Bosse M, Siegwart R, Furgale P (2015) The gist of maps - Summarizing experience for lifelong localization. In: IEEE International Conference on Robotics and Automation (ICRA), pp 2767–2773, DOI 10.1109/ICRA.2015.7139575

Erkent O, Bozma HI (2015) Long-term topological place learning. In: IEEE International Conference on Robotics and Automation (ICRA), pp 5462–5467, DOI 10.1109/ICRA.2015.7139962

Fraundorfer F, Scaramuzza D (2012) Visual Odometry - Part II: Matching, robustness, and applications. IEEE Robotics and Automation Magazine (RAM) 19(2):78–90, DOI 10.1109/MRA.2012.2182810

Fuentes-Pacheco J, Ruiz-Ascencio J, Rendón-Mancha JM (2012) Visual simultaneous localization and mapping: A survey. Artificial Intelligence Review (AIR) pp 1–27, DOI 10.1007/s10462-012-9365-8

Gálvez-López D, Tardós JD (2012) Bags of binary words for fast place recognition in image sequences. IEEE Transactions on Robotics (TRO) 28(5):1188–1197, DOI 10.1109/TRO.2012.2197158

Gao X, Zhang T (2017) Unsupervised learning to detect loops using deep neural networks for visual SLAM system. Autonomous Robots 41(1):1–18, DOI 10.1007/s10514-015-9516-2

Garcia-Fidalgo E, Ortiz A (2015) Vision-based topological mapping and localization methods: A survey. Robotics and Autonomous Systems (RAS) 64:1–20, DOI 10.1016/j.robot.2014.11.009

Geiger A, Roser M, Urtasun R (2010) Efficient large-scale stereo matching. In: Asian Conference on Computer Vision (ACCV), vol 6492, pp 25–38, DOI 10.1007/978-3-642-19315-6_3

Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? the KITTI vision benchmark suite. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 3354–3361, DOI 10.1109/CVPR.2012.6248074

Glover AJ, Maddern W, Milford M, Wyeth GF (2010) FAB-MAP + RatSLAM: Appearance-based SLAM for multi-

ple times of day. In: IEEE International Conference on Robotics and Automation (ICRA), pp 3507–3512, DOI 10.1109/ROBOT.2010.5509547

Glover AJ, Maddern W, Warren M, Reid S, Milford M, Wyeth GF (2012) OpenFABMAP: An open source toolbox for appearance-based loop closure detection. In: IEEE International Conference on Robotics and Automation (ICRA), pp 4730–4735, DOI 10.1109/ICRA.2012.6224843

Hirschmuller H (2008) Stereo processing by semiglobal matching and mutual information. IEEE Trans on Pattern Analysis and Machine Intelligence (TPAMI) 30(2):328–341, DOI 10.1109/TPAMI.2007.1166

Johns E, Yang G (2014) Generative methods for long-term place recognition in dynamic scenes. International Journal of Computer Vision (IJCV) 106(3):297–314, DOI 10.1007/s11263-013-0648-6

Korrapati H, Mezouar Y (2017) Multi-resolution map building and loop closure with omnidirectional images. Autonomous Robots 41(4):967–987, DOI 10.1007/s10514-016-9560-6

Korrapati H, Uzer F, Mezouar Y (2013) Hierarchical visual mapping with omnidirectional images. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 3684–3690, DOI 10.1109/IROS.2013.6696882

Lee GH, Pollefeys M (2014) Unsupervised learning of threshold for geometric verification in visual-based loop-closure. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1510–1516, DOI 10.1109/ICRA.2014.6907052

Leutenegger S, Chli M, Siegwart RY (2011) BRISK: Binary robust invariant scalable keypoints. In: International Conference on Computer Vision (ICCV), pp 2548–2555, DOI 10.1109/ICCV.2011.6126542

Linegar C, Churchill W, Newman P (2015) Work smart, not hard: Recalling relevant experiences for vast-scale but time-constrained localisation. In: IEEE International Conference on Robotics and Automation (ICRA), pp 90–97, DOI 10.1109/ICRA.2015.7138985

Liu Y, Zhang H (2012) Visual loop closure detection with a compact image descriptor. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 1051–1056, DOI 10.1109/IROS.2012.6386145

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (IJCV) 60(2):91–110, DOI 10.1023/B:VISI.0000029664.99615.94

Lowry S, Milford M (2015) Change removal: Robust online learning for changing appearance and changing viewpoint. In: Workshop on Visual Place Recognition in Changing Environments at the IEEE International Conference on Robotics and Automation (W-ICRA)

Lowry S, Sünderhauf N, Newman P, Leonard JJ, Cox D, Corke P, Milford M (2016) Visual place recognition: A survey. IEEE Transactions on Robotics (TRO) 32(1):1–19, DOI 10.1109/TRO.2015.2496823

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

Masatoshi A, Yuuto C, Kanji T, Kentaro Y (2015) Leveraging image-based prior in cross-season place recognition. In: IEEE International Conference on Robotics and Automation (ICRA), pp 5455–5461, DOI 10.1109/ICRA.2015.7139961

McManus C, Churchill W, Maddern W, Stewart A, Newman P (2014) Shady dealings: Robust, long-term visual localisation using illumination invariance. In: IEEE International Conference on Robotics and Automation (ICRA), pp 901–906, DOI 10.1109/ICRA.2014.6906961

Milford M (2012) Visual route recognition with a handful of bits. In: Robotics Science and Systems Conference (RSS), pp 297–304, DOI 10.15607/RSS.2012.VIII.038

Milford M, Wyeth GF (2012) SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1643–1649, DOI 10.1109/ICRA.2012.6224623

Mohan M, Gálvez-López D, Monteleoni C, Sibley G (2015) Environment selection and hierarchical place recognition. In: IEEE International Conference on Robotics and Automation (ICRA), pp 5487–5494, DOI 10.1109/ICRA.2015.7139966

Mousavian A, Kosecká J, Lien J (2015) Semantically guided location recognition for outdoors scenes. In: IEEE International Conference on Robotics and Automation (ICRA), pp 4882–4889, DOI 10.1109/ICRA.2015.7139877

Muja M, Lowe DG (2012) Fast matching of binary features. In: Canadian Conference on Computer and Robot Vision (CRV), pp 404–410, DOI 10.1109/CRV.2012.60

Muja M, Lowe DG (2014) Scalable nearest neighbor algorithms for high dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 36(11):2227–2240, DOI 10.1109/TPAMI.2014.2321376

Mur-Artal R, Montiel JMM, Tardós JD (2015) ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE Transactions on Robotics (TRO) 31(5):1147–1163, DOI 10.1109/TRO.2015.2463671

Murillo AC, Singh G, Kosecká J, Guerrero JJ (2013) Localization in urban environments using a panoramic gist descriptor. IEEE Transactions on Robotics (TRO) 29(1):146–160, DOI 10.1109/TRO.2012.2220211

Negre-Carrasco PL, Bonin-Font F, Oliver-Codina G (2016) Global image signature for visual loop-closure de-

tection. Autonomous Robots 40(8):1403–1417, DOI 10.1007/s10514-015-9522-4

Nelson P, Churchill W, Posner I, Newman P (2015) From dusk till dawn: Localisation at night using artificial light sources. In: IEEE International Conference on Robotics and Automation (ICRA), pp 5245–5252, DOI 10.1109/ICRA.2015.7139930

Neubert P, Sünderhauf N, Protzel P (2015) Superpixel-based appearance change prediction for long-term navigation across seasons. Robotics and Autonomous Systems (RAS) 69(7):15–27, DOI 10.1016/j.robot.2014.08.005

Ojala T, Pietikäinen M, Harwood D (1996) A comparative study of texture measures with classification based on featured distributions. Pattern Recognition (PR) 29(1):51–59, DOI 10.1016/0031-3203(95)00067-4

Oliva A, Torralba A (2006) Building the gist of a scene: The role of global image features in recognition. Visual Perception, Progress in Brain Research (PBR) 155(B):23–36, DOI 10.1016/S0079-6123(06)55002-2

Pandey G, McBride JR, Eustice R (2011) Ford Campus vision and lidar data set. International Journal of Robotics Research (IJRR) 30(13):1543–1552, DOI 10.1177/0278364911400640

Paul R, Newman P (2010) FAB-MAP 3D: Topological mapping with spatial and visual appearance. In: IEEE International Conference on Robotics and Automation (ICRA), pp 2649–2656, DOI 10.1109/ROBOT.2010.5509587

Pepperell E, Corke P, Milford M (2014) All-environment visual place recognition with SMART. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1612–1618, DOI 10.1109/ICRA.2014.6907067

Pepperell E, Corke P, Milford M (2015) Automatic image scaling for place recognition in changing environments. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1118–1124, DOI 10.1109/ICRA.2015.7139316

Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: An efficient alternative to SIFT or SURF. In: International Conference on Computer Vision (ICCV), pp 2564–2571, DOI 10.1109/ICCV.2011.6126544

Scaramuzza D, Fraundorfer F (2011) Visual Odometry - Part I: The first 30 years and fundamentals. IEEE Robotics and Automation Magazine (RAM) 18(4):80–92, DOI 10.1109/MRA.2011.943233

Smith M, Baldwin I, Churchill W, Paul R, Newman P (2009) The New College vision and laser data set. International Journal of Robotics Research (IJRR) 28(5):595–599, DOI 10.1177/0278364909103911

Sünderhauf N, Protzel P (2011) BRIEF-Gist - Closing the loop by simple means. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 1234–1241, DOI 10.1109/IROS.2011.6094921

Sünderhauf N, Neubert P, Protzel P (2013) 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)

Sünderhauf N, Shirazi S, Jacobson A, Dayoub F, Pepperell E, Upcroft B, Milford M (2015) Place recognition with ConvNet landmarks: Viewpoint-robust, condition-robust, training-free. In: Robotics Science and Systems Conference (RSS), pp 1–10, DOI 10.15607/RSS.2015.XI.022

Ulrich I, Nourbakhsh IR (2000) Appearance-based place recognition for topological localization. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1023–1029, DOI 10.1109/ROBOT.2000.844734

Upcroft B, McManus C, Churchill W, Maddern W, Newman P (2014) Lighting invariant urban street classification. In: IEEE International Conference on Robotics and Automation (ICRA), pp 1712–1718, DOI 10.1109/ICRA.2014.6907082

Valgren C, Lilienthal AJ (2010) SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments. Robotics and Autonomous Systems (RAS) 58(2):149–156, DOI 10.1016/j.robot.2009.09.010

Williams B, Cummins M, Neira J, Newman P, Reid ID, Tardós JD (2008) An image-to-map loop closing method for monocular SLAM. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 2053–2059, DOI 10.1109/IROS.2008.4650996

Williams B, Cummins M, Neira J, Newman P, Reid ID, Tardós JD (2009) A comparison of loop closing techniques in monocular SLAM. Robotics and Autonomous Systems (RAS) 57(12):1188–1197, DOI 10.1016/j.robot.2009.06.010

Yang X, Cheng KT (2014) Local difference binary for ultrafast and distinctive feature description. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 36(1):188–194, DOI 10.1109/TPAMI.2013.150