

# Fast and Effective Visual Place Recognition using Binary Codes and Disparity Information

Roberto Arroyo<sup>1</sup>, Pablo F. Alcantarilla<sup>2</sup>, Luis M. Bergasa<sup>1</sup>, J. Javier Yebes<sup>1</sup> and Sebastián Bronte<sup>1</sup>

**Abstract**—We present a novel approach for place recognition and loop closure detection based on binary codes and disparity information using stereo images. Our method (ABLE-S) applies the Local Difference Binary (LDB) descriptor in a global framework to obtain a robust global image description, which is initially based on intensity and gradient pairwise comparisons. LDB has a higher descriptiveness power than other popular alternatives such as BRIEF, which only relies on intensity. In addition, we integrate disparity information into the binary descriptor (D-LDB). Disparity provides valuable information which decreases the effect of some typical problems in place recognition such as perceptual aliasing.

The KITTI Odometry dataset is mainly used to test our approach due to its varied environments, challenging situations and length. Additionally, a loop closure ground-truth is introduced in this work for the KITTI Odometry benchmark with the aim of standardizing a robust evaluation methodology for comparing different previous algorithms against our method and for future benchmarking of new proposals. Attending to the presented results, our method allows a fast and more effective visual loop closure detection compared to state-of-the-art algorithms such as FAB-MAP, WI-SURF and BRIEF-Gist.

## I. INTRODUCTION

Visual navigation systems used by intelligent robots and vehicles need effective and efficient computer vision algorithms for robust and fast place recognition. Since the appearance of FAB-MAP [1], loop closure techniques based on visual information have experienced a significant growth and several works have contributed to this research line (see Section II). However, FAB-MAP has some drawbacks, such as the need of a previous training to build a visual vocabulary of the environment and the associated probabilistic approach, making the algorithm not suitable for real-time applications.

The usage of binary descriptors for place recognition has been recently proposed in some works in the literature [2], [3], [4], [5]. Binary descriptors provide a reduction in computational costs and memory resources compared to vector-based descriptors. In [2], the authors introduced a global binary descriptor named BRIEF-Gist for loop closure detection problems. BRIEF-Gist was shown to accomplish better detection rates than FAB-MAP, while being several orders of magnitude faster. Besides, BRIEF-Gist also has better performance than global vector-based descriptors such as WI-SURF [6], as demonstrated in our tests.

\*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 II (S2009/DPI-1559).

<sup>1</sup>Department of Electronics, University of Alcalá (UAH), Alcalá de Henares, 28871, Madrid, Spain. {roberto.arroyo, bergasa, javier.yebes, sebastian.bronte}@depeca.uah.es

<sup>2</sup>Toshiba Research Europe Ltd., Cambridge, United Kingdom. pablo.alcantarilla@crl.toshiba.co.uk

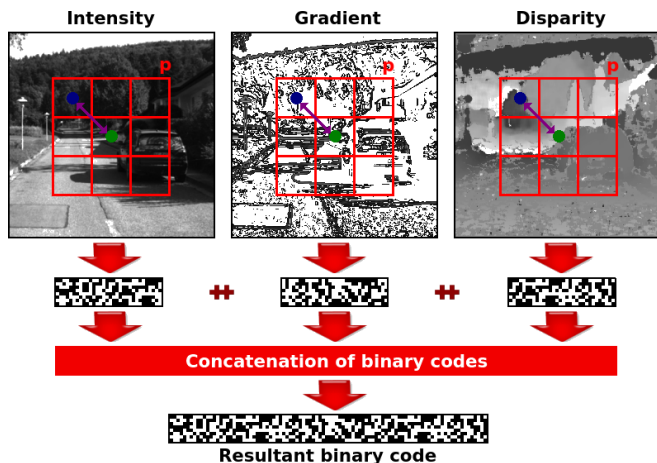


Fig. 1. Proposed visual place recognition method based on LDB codes and disparity information for loop closure detection. Given an image patch  $p$  and a sampling pattern, a binary test is applied as a comparison of the image intensity, gradient and disparity.

In this paper, we propose a novel approach for visual place recognition and loop closure detection which uses global binary descriptors that are built from intensity, gradient and disparity pairwise comparisons, as presented in Fig. 1. We extend the Local Difference Binary (LDB) [7] descriptor to incorporate disparity information (D-LDB). As will be shown in our experiments, the addition of disparity provides a more precise visual localization than only using intensity and gradient information. The image description applying disparity helps to reduce some typical problems related to visual place recognition, such as perceptual aliasing. Our method, named ABLE-S, improves the results obtained by other state-of-the-art algorithms such as FAB-MAP, WI-SURF and BRIEF-Gist and it also has a low computational cost.

The most important contributions introduced in this work are the following:

- An innovative algorithm for visual place recognition and loop closure detection based on LDB global binary descriptors and disparity (see Section III and IV).
- A new ground-truth designed for loop closure detection in a challenging dataset as the KITTI Odometry [8] and an objective evaluation methodology (see Section V).
- An extensive comparative study about the experimental results obtained with our algorithm in its different variants against the principal state-of-the-art algorithms (see Section VI) and the main conclusions obtained jointly with some future research lines (see Section VII).

## II. RELATED WORK

BRIEF-Gist demonstrated that vision-based loop closure detection algorithms do not need a very complex visual description such as the one suggested by FAB-MAP. This theory was reinforced by [9], which experimented with low resolution images to make evident that a handful of bits is sufficient for conducting an effective vision-based navigation. Furthermore, matching of binary features can be extremely fast for associating views of the same place in different moments, as deduced from [10].

Local binary descriptors have improved significantly since the initial formulation of BRIEF [11]. Other approaches such as ORB [12] and BRISK [13] extend the BRIEF formulation to add invariance to rotation and scale changes. One of the most recent and attractive methods for image description based on local binary codes is LDB. This binary descriptor not only uses intensity information as BRIEF, but also gradient difference tests, dividing the image into several cells and applying a multiple gridding strategy with the aim of capturing distinct patterns at different spatial resolutions. This approach allows a high precision for image description and a faster matching speed compared to BRIEF. Furthermore, an evolution of the original algorithm called M-LDB [14] was recently presented, adding invariance to scale and rotation.

Apart from intensity and gradient information provided by LDB, disparity information of the environment can improve place recognition. Several works in the literature use 3D information to improve performance of vision-based navigation methods, such as [15], [16], [6]. In [15], FAB-MAP is extended to incorporate the spatial distribution of visual words in 3D. Similarly, a combination of visual words and 3D information from stereo sequences was used in [16] to perform robust place recognition. In [6], a global SURF [17] descriptor called WI-SURF was applied in combination with rough 3D information captured by a single line laser scanner to improve loop closure detection. However, the calculation and matching of combined visual features and 3D information is associated with a high computational cost. For this reason, a binary code that integrates disparity as the one proposed in our work (D-LDB) is a promising approach, since 3D information is efficiently integrated in the binary descriptor, improving place recognition performance.

## III. BINARY DESCRIPTORS

In this section, we briefly explain the basic principles of binary descriptors. In general, binary descriptors are built from a set of pairwise intensity comparisons from a sampling pattern centered in a point of interest, as shown in Fig. 1. Each bit in the descriptor is the result of exactly one comparison. It is important to note that the set of pairwise comparisons is not only limited to intensity, since gradient and other image cues can also be applied, increasing the discriminative power of the descriptors [7], [14]. Furthermore, most of the binary descriptors differ in the sampling pattern, which can be fixed or adapted to obtain descriptors invariant to scale and rotation.

One of the nicest properties of binary descriptors is that they can be matched very efficiently by using the Hamming distance. This is more efficient than the traditional way of matching descriptors with the  $L_2$ -norm.

Now, we will explain how these binary descriptors are constructed. Considering a smoothed image patch  $\mathbf{p}$  centered in the location of the point of interest  $\mathbf{x} = (x, y)$ , a binary test  $\tau$  is defined as:

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

where  $f(i)$  is a function that returns an image feature response for a certain pixel or cell in  $\mathbf{p}$ .  $f(i)$  can be simply the smoothed image intensity  $I$  at one pixel location  $\mathbf{x}_i = (x_i, y_i)$ , as proposed by binary descriptors such as BRIEF, ORB and BRISK:

$$f(i) = I(\mathbf{x}_i) \quad (2)$$

Additionally,  $f(i)$  can also be the concatenation of different binary comparisons such as averaged image intensities  $I_{avg}$  and image gradients  $G_x, G_y$  on a particular cell  $\mathbf{c}_i$  in  $\mathbf{p}$ , as proposed by binary descriptors such as LDB and M-LDB:

$$f(i) = \{I_{avg}(\mathbf{c}_i), G_x(\mathbf{c}_i), G_y(\mathbf{c}_i)\} \quad (3)$$

With the aim of reducing the effect of place recognition problems such as perceptual aliasing, the initial proposal of LDB is extended in our method by also applying binary comparisons for averaged disparity information  $D_{avg}$ :

$$f(i) = \{I_{avg}(\mathbf{c}_i), G_x(\mathbf{c}_i), G_y(\mathbf{c}_i), D_{avg}(\mathbf{c}_i)\} \quad (4)$$

Finally, the resulting descriptor  $d_n(\mathbf{p})$  is computed as a vector of  $n$  binary tests, where  $n$  is also the final dimension of the descriptor, which is usually empirically adjusted:

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

## IV. PROPOSED METHOD

### A. Binary code calculation

We have chosen LDB instead of other binary descriptors for our place recognition proposal because LDB improves the performance of BRIEF by adding the output of binary tests based on gradient comparisons, apart from intensity. Besides, we have extended the original formulation of LDB to incorporate disparity information (D-LDB), as exposed in Eq. 4. Disparity is calculated by following the SGBM (Semi Global Block Matching) algorithm proposed in [18], which is applied by using its OpenCV [19] implementation.

As stated in [9], high resolution images are not needed to accomplish an effective visual navigation. According to this, our image patch  $\mathbf{p}$  is sized at 64x64 pixels and images are downsampled to this resolution before applying the global binary descriptor. A lower size of  $\mathbf{p}$  decreases our place recognition effectiveness and a higher one does not yield increased precision. In addition, we fix the binary codes to a dimension  $n$  of 256 bits (32 bytes), which is supported by works such as [2], [7]. The code is adjusted to the fixed  $n$  by using the random bit selection method proposed in [7].

The global binary descriptor is computed by taking the center of the resized image patch as a keypoint without dominant rotation or scale. However, other alternative can be applied, which consists in dividing the image into grids, calculating a descriptor for each one and concatenating them in a final binary code, by adopting the centers of the resized grids as keypoints. This approach can consider different numbers of grids in width and height ( $g_w \times g_h$ ).

### B. Binary codes matching

Loop closures can be detected in our place recognition method by correlating the elements of  $\mathbf{v}$ , which is a vector that contains the binary codes obtained for the total number of analyzed scenes  $m$ , as defined in Eq. 6. Similarity between binary codes is calculated by using the Hamming distance exposed in Eq. 7 and computing a distance matrix  $M$ . POPCNT is a machine SSE4.2 instruction that is applied for fast matching of binary codes, since it allows to efficiently count the total number of bits which are set to one in a binary string.

$$\mathbf{v} = (d_n(\mathbf{p}_1), d_n(\mathbf{p}_2), d_n(\mathbf{p}_3), \dots, d_n(\mathbf{p}_m)) \quad (6)$$

$$M_{i,j} = M_{j,i} = \text{POPCNT}(\mathbf{v}_i \oplus \mathbf{v}_j) \quad (7)$$

### C. Algorithmic proposal for visual loop closure detection

Algorithm 1 is presented for a better definition of the proposed method for loop closure detection, which includes our binary code calculation using disparity for an input stereo image and the computation of its similarity with respect to the previously analyzed scenes in a matching process.

---

**Algorithm 1** Loop closure detection on stereo images by using binary codes and disparity information.

---

**Input:**  $L, R$  {Left and right stereo camera images.}  
 $g_w, g_h$  {Number of grids in width and height ( $(1 \times 1) \equiv \text{global}$ ).}  
 $M$  {Accumulated distance matrix.}  
 $\mathbf{v}$  {Accumulated vector of binary codes.}

**Output:**  $M$  {Updated distance matrix.}  
 $\mathbf{v}$  {Updated vector of binary codes.}

**Algorithm:**

```

i ← Length( $\mathbf{v}$ ) + 1; {Current stereo image index.}
for k ← 1 to  $g_w$  do
  for l ← 1 to  $g_h$  do
     $d_n(\mathbf{p}_{i(k,l)})$  ← Calculate the binary code for stereo image ( $L, R$ )
      in the grid ( $k, l$ ); {See Eq. 1-5.}
     $\mathbf{v}_i \leftarrow \mathbf{v}_i \parallel d_n(\mathbf{p}_{i(k,l)})$ ; {Concatenation of each grid code.}
  end for
end for
for j ← 1 to i do
   $M_{i,j} \leftarrow M_{j,i} \leftarrow \text{POPCNT}(\mathbf{v}_i \oplus \mathbf{v}_j)$ ; {See Eq. 7.}
end for

```

---

## V. EVALUATION

### A. Datasets

In our previous work [5], we used the Oxford New College dataset [20], which is often employed in the literature [1], [2]. We observed that our previously presented approach improved the results obtained by algorithms such as FAB-MAP or BRIEF-Gist. However, this dataset is not representative

enough, since it only has 8127 scenes (2,2 km) and there are not enough varied environments. Due to this, in this paper we decide to use the popular KITTI benchmark dataset [8] for testing our new proposal.

The KITTI Odometry has 22 sequences containing a total of 44182 stereo images (39,2 km). These sequences include environments with different characteristics and challenging situations such as perceptual aliasing, changes on scene, etc. Nevertheless, there is not any specific ground-truth for loop closure detection in this dataset. For this reason, we have created this ground-truth, which is described in Table I.

There are 12 sequences with loop closures, 21 of them unidirectional and 4 bidirectional in total. Bidirectional loop closures appear when a same place is traversed in a different direction. As stated in [2], BRIEF-Gist can not detect this kind of loop closures and, although FAB-MAP is theoretically invariant to spatial location, in practical situations with stereo images it also can not identify bidirectional loop closures, because stereo images do not provide a visual perception in all directions. Currently, our approach also can not perceive bidirectional loop closures by only using stereo images. Therefore, they are not evaluated in this paper, although they could be considered in future works or if panoramic images would be used, as proposed in our previous work [5]. Additionally, the defined ground-truth for the KITTI Odometry also takes into account long stops of vehicle for ignoring these frames in the evaluation.

### B. Methodology

We propose an evaluation technique common to all the tests presented in this paper for objectively comparing our proposal against the state-of-the-art algorithms. The KITTI Odometry dataset offers a methodology for evaluating translational and rotational errors in visual odometry. Unfortunately, there is not any specific benchmarking for loop closure detection algorithms, so we will introduce it.

Our evaluation methodology is based on precision-recall curves obtained from  $M$ , which in our case is achieved after the iteration of Algorithm 1 in a complete sequence of stereo images.  $M$  is normalized by following Eq. 8. Later, a threshold  $\theta$  is applied to  $M$  for comparing it to the ground-truth matrix  $G$  of each specific sequence. True positives are considered when a positive of the thresholded  $M$  coincides with a positive of  $G$  in a temporal vicinity according to the frame rate. The final curve is computed by varying  $\theta$  in a linear distribution between 0 and 1 and calculating the corresponding values of precision and recall. 100 values of  $\theta$  are processed to obtain well-defined curves.

$$M_{i,j} = \frac{M_{i,j}}{\max(M)} \quad (8)$$

For a better comprehension of the proposed benchmarking methodology, an example of evaluation for the sequence 06 of the KITTI Odometry is presented in Fig. 2. With the aim of referring to our loop closure detection algorithm, we make allusion to this method in the tests as **ABLE** (Able for **B**inary-appearance **L**oop-closure **E**valuation). As it is applied on stereo images, this specific approach is called **ABLE-S**.

TABLE I

DESCRIPTION OF THE GROUND-TRUTH CREATED FOR VISUAL LOOP CLOSURE DETECTION IN THE KITTI ODOMETRY DATASET.

| Sequence | No. of frames | Unidirectional loop closures |                |             | Bidirectional loop closures |                |             | Long stops |             |
|----------|---------------|------------------------------|----------------|-------------|-----------------------------|----------------|-------------|------------|-------------|
|          |               | No.                          | Initial frames | Loop frames | No.                         | Initial frames | Loop frames | No.        | Stop frames |
| 00       | 4541          | 5                            | 0000 - 0099    | 4451 - 4528 | 0                           |                |             | 1          | 0534 - 0569 |
|          |               |                              | 0122 - 0196    | 1570 - 1635 |                             |                |             |            |             |
|          |               |                              | 0392 - 0412    | 2446 - 2460 |                             |                |             |            |             |
|          |               |                              | 0392 - 0941    | 3398 - 3844 |                             |                |             |            |             |
|          |               |                              | 2354 - 2460    | 3295 - 3418 |                             |                |             |            |             |
| 02       | 4661          | 2                            | 0933 - 1026    | 4205 - 4266 | 1                           | 3332 - 3397    | 4566 - 4620 | 0          |             |
|          |               |                              | 1810 - 1997    | 4404 - 4569 |                             |                |             |            |             |
| 05       | 2761          | 3                            | 0031 - 0121    | 2431 - 2512 | 0                           |                |             | 1          | 2321 - 2402 |
|          |               |                              | 0565 - 0787    | 1324 - 1530 |                             |                |             |            |             |
|          |               |                              | 0819 - 0885    | 2581 - 2627 |                             |                |             |            |             |
| 06       | 1101          | 1                            | 0000 - 0280    | 0835 - 1093 | 0                           |                |             | 0          |             |
| 07       | 1101          | 1                            | 0000 - 0013    | 1060 - 1067 | 0                           |                |             | 1          | 0657 - 0732 |
| 08       | 4071          | 0                            |                |             | 2                           | 0075 - 0227    | 1640 - 1796 | 1          | 4000 - 4031 |
|          |               |                              |                |             |                             | 0726 - 0765    | 1422 - 1464 |            |             |
| 09       | 1591          | 1                            | 0000 - 0023    | 1578 - 1590 | 0                           |                |             | 0          |             |
| 13       | 3281          | 4                            | 0000 - 0138    | 2152 - 2316 | 0                           |                |             | 0          |             |
|          |               |                              | 0000 - 0089    | 3188 - 3280 |                             |                |             |            |             |
|          |               |                              | 0553 - 0839    | 1633 - 1938 |                             |                |             |            |             |
|          |               |                              | 2152 - 2264    | 3188 - 3280 |                             |                |             |            |             |
| 15       | 1901          | 1                            | 0000 - 0086    | 1808 - 1900 | 0                           |                |             | 0          |             |
| 16       | 1731          | 1                            | 0000 - 0146    | 1614 - 1730 | 1                           | 0023 - 0079    | 0798 - 0843 | 0          |             |
| 18       | 1801          | 1                            | 0322 - 0493    | 1616 - 1800 | 0                           |                |             | 0          |             |
| 19       | 4981          | 1                            | 4246 - 4390    | 4812 - 4943 | 0                           |                |             | 0          |             |

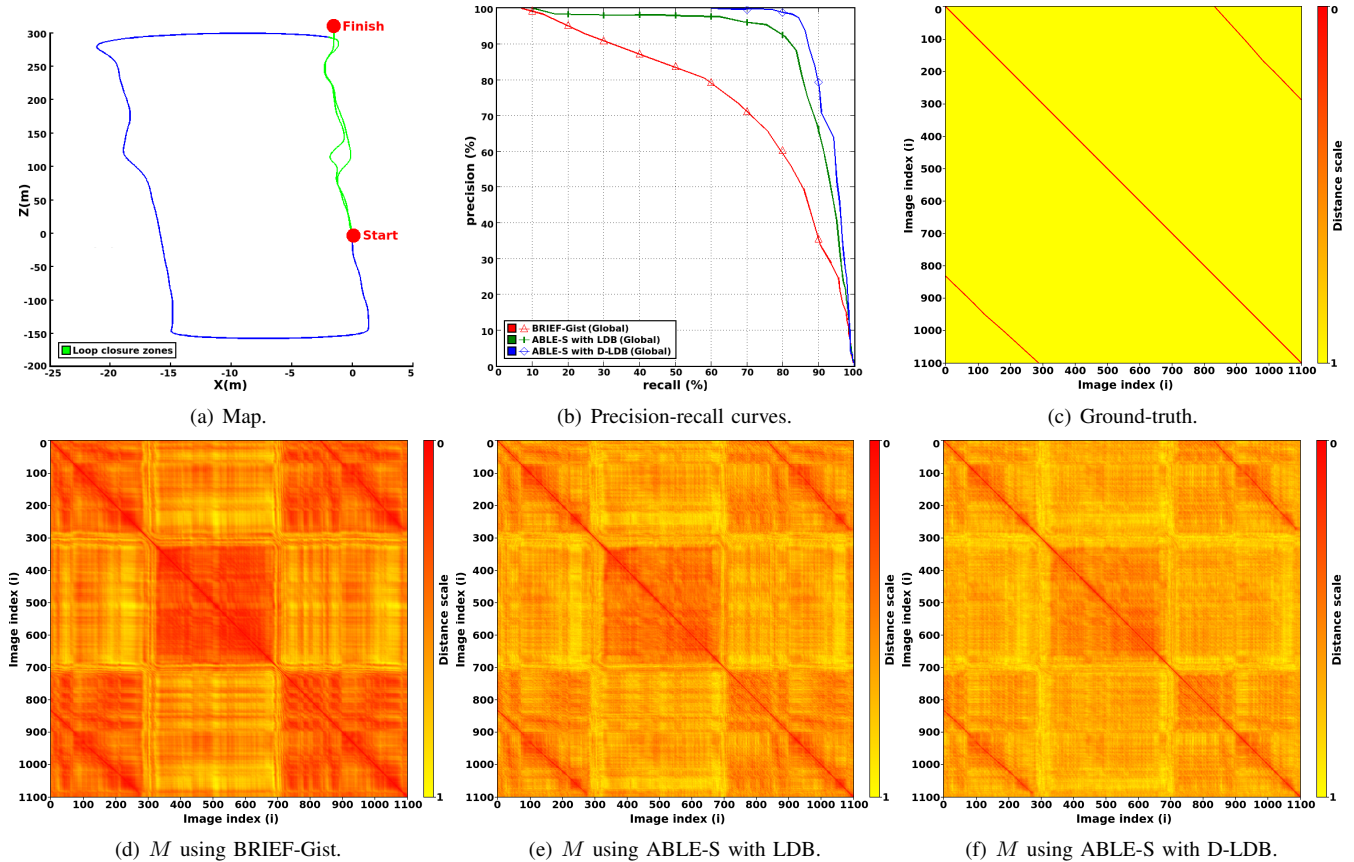


Fig. 2. An example of evaluation for loop closure detection in the sequence 06 of the KITTI Odometry dataset. The distance matrices show how the addition of gradient information (e) reduces the noise appeared around loop closure lines if only intensity is used (d). Besides, if disparity information is added (f), perceptual aliasing is clearly decreased, as confirmed by the precision-recall curves (b).

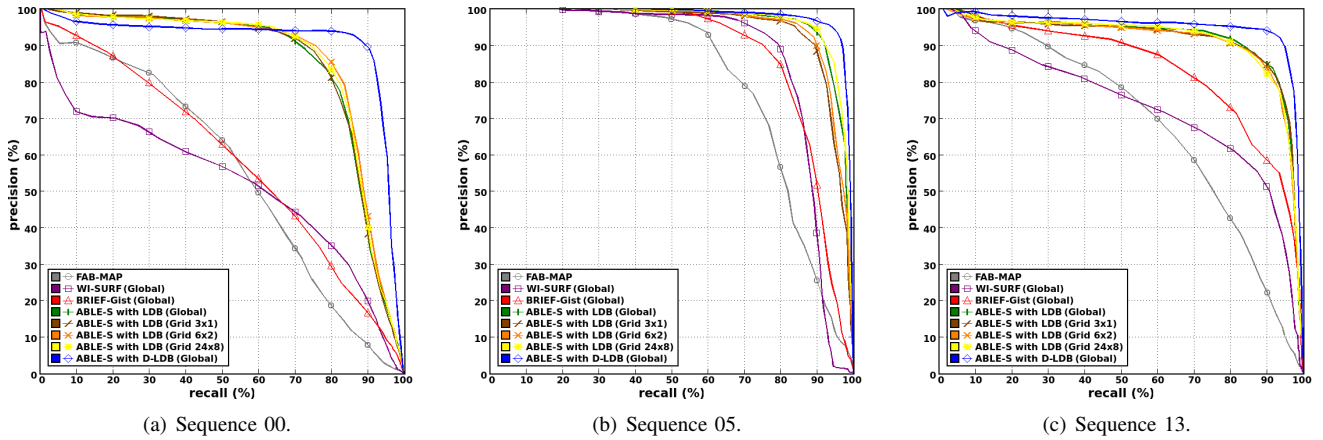


Fig. 3. Results obtained for loop closure detection in some of the most representative sequences of the KITTI Odometry dataset.

TABLE II

COMPARISON OF SOME PERFORMANCE PARAMETERS BETWEEN THE PLACE RECOGNITION APPROACHES TESTED IN THE PRESENTED RESULTS.

|                            | Average precision (Seq. 00) | Average precision (Seq. 05) | Average precision (Seq. 13) | Average precision (All Seq.) | Average time per description (All Seq.) | Average time per matching (All Seq.) |   |
|----------------------------|-----------------------------|-----------------------------|-----------------------------|------------------------------|---|--------------------------------------|---|
| <b>FAB-MAP</b>             | 55.50 %                     | 80.19 %                     | 69.04 %                     | 70.24 %                      | 195.27 ms                               | $2.67 \cdot 10^{-2}$ ms              |   |
| <b>WI-SURF (Global)</b>    | 52.40 %                     | 86.75 %                     | 73.47 %                     | 72.59 %                      | 0.23 ms                                 | $1.74 \cdot 10^{-3}$ ms              |   |
| <b>BRIEF-Gist (Global)</b> | 58.76 %                     | 87.69 %                     | 83.85 %                     | 78.86 %                      | <b>0.04 ms</b>                          | $2.35 \cdot 10^{-5}$ ms              |   |
| <b>ABLE-S</b>              | <b>with LDB (Global)</b>    | 85.19 %                     | 96.35 %                     | 92.20 %                      | 90.98 %                                 | 0.11 ms                              | <b><math>2.29 \cdot 10^{-5}</math> ms</b> |
|                            | <b>with LDB (Grid 3x1)</b>  | 85.22 %                     | 94.78 %                     | 91.76 %                      | 90.85 %                                 | 0.35 ms                              | $8.54 \cdot 10^{-5}$ ms                   |
|                            | <b>with LDB (Grid 6x2)</b>  | 86.04 %                     | 95.43 %                     | 91.42 %                      | 91.05 %                                 | 1.01 ms                              | $2.68 \cdot 10^{-4}$ ms                   |
|                            | <b>with LDB (Grid 24x8)</b> | 85.78 %                     | 96.86 %                     | 91.75 %                      | 91.36 %                                 | 7.29 ms                              | $1.93 \cdot 10^{-3}$ ms                   |
|                            | <b>with D-LDB (Global)</b>  | <b>91.06%</b>               | <b>98.43%</b>               | <b>95.38%</b>                | <b>95.35%</b>                           | 7.11 ms                              | <b><math>2.29 \cdot 10^{-5}</math> ms</b> |

## VI. RESULTS

The addition of disparity in our proposal clearly improves the results achieved in place recognition by other similar methods based on binary codes. This can be observed if the distance matrices presented in the evaluation shown in Fig. 2 are compared. In the distance matrix provided by our method using the D-LDB descriptor (see Fig. 2 (f)), it can be shown how the noisy squares which appear around the loop closure lines due to perceptual aliasing are clearly reduced with respect to other global binary description approaches such as BRIEF-Gist (see Fig. 2 (d)), which is only based on intensity information. Our approach adding disparity also improves the results obtained by employing the original LDB (see Fig. 2 (e)), which only applies intensity and gradient comparisons. All the previous affirmations are supported by the precision-recall curves depicted in Fig. 2 (b).

Apart from the previous evaluation for the sequence 06 of the KITTI Odometry, we show precision-recall curves for the sequences 00, 05 and 13, which are the most representative because they contain the highest number of loop closures. We have obtained very satisfactory results for the different variants of our method compared to FAB-MAP, WI-SURF and BRIEF-Gist algorithms, as presented in Fig. 3.

For testing FAB-MAP, we use its open source implementation called OpenFABMAP [21], which is applied in a standard configuration and conveniently trained in the

KITTI Odometry dataset. WI-SURF and BRIEF-Gist are implemented by using the SURF and BRIEF descriptors provided by OpenCV. Our WI-SURF implementation do not add the range mean and standard deviation to the descriptor as proposed in [6], because we want to objectively compare our global binary description proposal against a standard global vector-based descriptor. In addition, we do not process results for global binary descriptors which can be proposed by applying local binary descriptors invariant to rotation and/or scale such as ORB, BRISK or M-LDB. This is because, as exposed in [2], downsampling the input images involves smoothing and interpolation over neighboring regions that reduces the effects of rotation and scale, so descriptors invariant to these circumstances do not provide better results in our case, as previously demonstrated in [5].

Table II confirms the successful results achieved by our place recognition method in loop closure detection using binary codes and disparity information, displaying the comparison between the average precision of the different algorithms tested in the KITTI Odometry sequences 00, 05, 13 and its mean value for all sequences. Average precision is defined as the percentage of total area under the precision-recall curve. Average processing times for each image description and matching are also presented. These performance parameters have been obtained using a computer with an Intel Core i7 2,40 GHz processor and a 8 GB RAM.

Attending to the results depicted in Fig. 3 and Table II, the usage of grids instead of a global binary description does not substantially improve precision in this case and adds computational cost. The KITTI Odometry images have a resolution of 1226x370 pixels, so we use a proportional number of grids such as 3x1, 6x2 and 24x8, obtaining the best results with 24x8. Besides, global binary descriptors such as BRIEF-Gist or the one proposed by our method achieve better results and have lower processing times in matching than global vector-based descriptors such as WI-SURF, due to the usage of the Hamming distance instead of a matching based on  $L_2$ -norm. FAB-MAP also reaches worse results and extremely higher processing times.

Finally, we must note that our method has a slightly higher processing time in description using D-LDB than applying a standard LDB or other global image description approaches such as BRIEF-Gist or WI-SURF. This is due to the cost of computing and processing the disparity information. However, other stereo matching algorithms faster than SGBM can be used for speeding-up the D-LDB descriptor. In addition, some commercial stereo cameras that already incorporate disparity estimation on the chip can also be used, such as the Bumblebee2. In any case, our approach using SGBM allows to describe 141 stereo images per second, which is more than sufficient for real-time applications. Furthermore, the bottleneck in place recognition is associated to the processing time in matching, which is greatly increased for each new computed image when the number of previously analyzed scenes is high because more matches are needed, while the processing time in description is constant.

## VII. CONCLUSIONS AND FUTURE WORKS

Along this paper, we have presented ABLE-S, demonstrating how this approach for place recognition and loop closure detection based on binary codes and disparity information provides a very effective performance, specially if it is compared to other state-of-the-art algorithms. We have achieved an average precision superior in about 25% to FAB-MAP, 23% to WI-SURF, 16% to BRIEF-Gist and 5% to our own method if LDB is considered in its standard version without using disparity. In addition, the computational cost of our method is low due to employment of binary codes, which allows a fast place recognition.

The challenging KITTI Odometry dataset has been employed in our tests, contributing a new ground-truth for loop closure detection and an objective evaluation methodology. The ground-truth matrices generated for each KITTI Odometry sequence and the code for benchmarking will be published with the aim of standardizing it, jointly with our place recognition source code and extra video material<sup>1</sup>.

The future objective of our research is to achieve a robust life-long visual localization. Recently, this topic has acquired a great importance in visual navigation, and several interesting works have appeared in the last year considering problems such as visual place recognition at different hours of day [22] or during all the seasons [23].

<sup>1</sup>All available from <http://www.robosafe.com/personal/roberto.arroyo/>

## 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] 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.
- [3] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics (TRO)*, vol. 28, no. 5, pp. 1188–1197, October 2012.
- [4] J. Straub, S. Hilsenbeck, G. Schroth, R. Huitl, A. Möller, and E. Steinbach, "Fast relocalization for visual odometry using binary features," in *International Conference on Image Processing (ICIP)*, September 2013, pp. 2548–2552.
- [5] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, J. J. Yeles, and S. Gámez, "Bidirectional loop closure detection on panoramas for visual navigation," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2014.
- [6] 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.
- [7] X. Yang and K. T. Cheng, "LDB: An ultra-fast feature for scalable augmented reality on mobile devices," in *International Symposium on Mixed and Augmented Reality (ISMAR)*, November 2012, pp. 49–57.
- [8] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the KITTI vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012, pp. 3354–3361.
- [9] M. Milford, "Visual route recognition with a handful of bits," in *Robotics Science and Systems Conference (RSS)*, July 2012.
- [10] M. Muja and D. G. Lowe, "Fast matching of binary features," in *Canadian Conference on Computer and Robot Vision (CRV)*, May 2012, pp. 404–410.
- [11] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *European Conference on Computer Vision (ECCV)*, September 2010, pp. 778–792.
- [12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *International Conference on Computer Vision (ICCV)*, November 2011, pp. 2564–2571.
- [13] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *International Conference on Computer Vision (ICCV)*, November 2011, pp. 2548–2555.
- [14] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conference (BMVC)*, September 2013.
- [15] R. Paul and P. Newman, "FAB-MAP 3D: Topological mapping with spatial and visual appearance," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2010, pp. 2649–2656.
- [16] C. Cadena, D. Gálvez-López, J. D. Tardós, and J. Neira, "Robust place recognition with stereo sequences," *IEEE Transactions on Robotics (TRO)*, vol. 28, no. 4, pp. 871–885, August 2012.
- [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] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 30, no. 2, pp. 328–341, February 2008.
- [19] G. Bradski, "The OpenCV library," *Dr. Dobbs' Journal of Software Tools (DDJ)*, vol. 25, no. 11, pp. 122–125, November 2000. [Online]. Available: <http://opencv.org>
- [20] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman, "The New College vision and laser data set," *International Journal of Robotics Research (IJRR)*, vol. 28, no. 5, pp. 595–599, May 2009.
- [21] 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.
- [22] 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.
- [23] 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.