Bidirectional Loop Closure Detection on Panoramas for Visual Navigation

Roberto Arroyo[1], Pablo F. Alcantarilla[2], Luis M. Bergasa[1], J. Javier Yebes[1] and Sergio Gámez[1]

Abstract— Visual loop closure detection plays a key role in navigation systems for intelligent vehicles. Nowadays, state-of-the-art algorithms are focused on unidirectional loop closures, but there are situations where they are not sufficient for identifying previously visited places. Therefore, the detection of bidirectional loop closures when a place is revisited in a different direction provides a more robust visual navigation.

We propose a novel approach for identifying bidirectional loop closures on panoramic image sequences. Our proposal combines global binary descriptors and a matching strategy based on cross-correlation of sub-panoramas, which are defined as the different parts of a panorama. A set of experiments considering several binary descriptors (ORB, BRISK, FREAK, LDB) is provided, where LDB excels as the most suitable. The proposed matching proffers a reliable bidirectional loop closure detection, which is not efficiently solved in any other previous research. Our method is successfully validated and compared against FAB-MAP and BRIEF-Gist. The Ford Campus and the Oxford New College datasets are considered for evaluation.
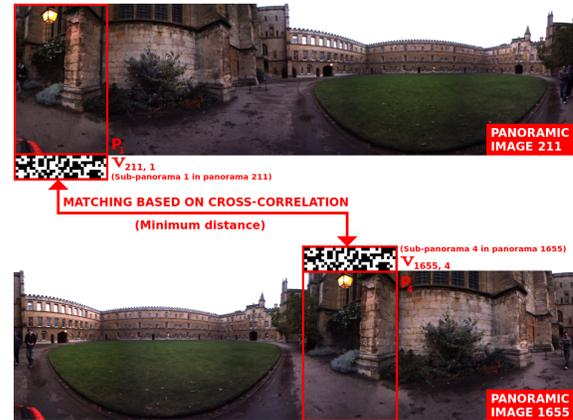
Fig. 1. Example of bidirectional loop closure detection using panoramas and matching of binary descriptors based on cross-correlation.

I. INTRODUCTION

Nowadays, autonomous vehicles rely on vision-based loop closure detection algorithms to recognize previously visited places. This task is essential to obtain a robust navigation in any intelligent transportation system. In fact, loop closure information is very effective to correct the accumulated error in vision-based navigation systems such as the employed for visual odometry [1] and SLAM [2].

FAB-MAP [3] has been popularized for distinguishing loop closures by using an appearance-only SLAM system. This method proposes a probabilistic approach for recognizing places based on a vocabulary tree. However, FAB-MAP requires a previous training and exhibits a high computational cost, which is due to its complex methodology for image description and matching.

In the last years, binary descriptors have become one of the main tendencies in image description and visual place recognition because of its low computational cost and effectiveness, which can be applied in devices with low processing and storage capacity, such as mobile phones. BRIEF [4] was one of the first binary descriptors broadly extended for image matching applications. Shortly after, BRIEF-Gist [5], a global image descriptor based on BRIEF, was defined for providing a visual loop closure detection more efficient than the proposed by FAB-MAP.

Nevertheless, an important weakness of BRIEF-Gist was noted by its authors in [5]: it can not detect bidirectional loop closures in any case. This kind of loop closures are considered when a place is traversed by a vehicle in a different direction, while unidirectional or standard loops appear when a place is revisited in the same direction. Other algorithms such as FAB-MAP can only theoretically identify bidirectional loop closures in panoramic image sequences. This is because FAB-MAP relies on the usage of bags of visual words, which are invariant to spatial location. However, as we will show in our experiments, the approach proposed by FAB-MAP does not work properly for bidirectional loop closures in practice and implies a high complexity.

We propose a new method for efficiently detecting bidirectional loop closures consisting in the binary description of sub-panoramas, which are defined as the different parts of a panorama, and the cross-correlation between them, as shown in the example presented in Fig. 1. Due to the recent definition of several new binary descriptors, which represent an evolution with respect to BRIEF, we test the most relevant ones for refining the visual description of sub-panoramas in our approach: ORB [6], BRISK [7], FREAK [8], LDB [9]. Furthermore, our contribution for matching panoramas is remarkable, because cross-correlation of sub-panoramas allows to identify bidirectional loop closures with a high precision and a similar computational cost compared to BRIEF-Gist.

The proposal presented in this paper is validated and contrasted with two main state-of-the-art works. Two public datasets composed of several panoramic images with different loop closure situations are used in our tests: the Ford Campus [10] and the Oxford New College [11] datasets.

II. RELATED WORK

A. Loop closure detection based on visual appearance

Apart from FAB-MAP and BRIEF-Gist, there are other recent works which have contributed to the research line related to loop closure detection based on visual appearance. The research efforts are currently focused on reducing the processing time and the memory resources. Several works reinforce this theory, such as [12], which experiments with low resolution images to make evident that a handful of bits is sufficient for carrying out an effective visual navigation.

The previous ideas perfectly fit in with the characteristics provided by binary descriptors, which allow an image description efficient in speed and memory. Moreover, matching of binary features is extremely fast, as deduced from [13]. This speed in matching also supports the usage of bags of binary words in a faster way compared to the classic bags of words employed in place recognition, as proposed by [14]. Other approaches also use binary features for relocalization in visual odometry, such as the defined in [15].

The final objective of any loop closure detection algorithm is to provide a robust navigation over time. Recently, several works have been proposed for life-long visual navigation problems, such as [16]. These kind of works consider problems such as place recognition in dynamic urban environments [17], in different hours of day [18], in scenes with changeable illumination [19] or during all the seasons [20].

Bidirectional loop closure detection is a needed step in visual navigation systems which is not efficiently solved in previous works. However, the importance of this topic is crucial, because only unidirectional loop closure information can be insufficient in long itineraries. The proposal presented in this work solves this problem by contributing a new method based on the cross-correlation of panoramic views.

B. Binary descriptors

Local binary descriptors have evolved since the initial implementation of BRIEF, which uses pairwise intensity comparisons to generate every bit of the final binary string. Some of these recent local binary descriptors have never been applied in loop closure detection, but they can improve the results obtained by methods based on BRIEF, such as BRIEF-Gist, which computes a global binary descriptor for a query image. In this work, we use four of these binary descriptors in a global framework for testing them jointly with our matching based on cross-correlation of sub-panoramas:

- **ORB (Oriented FAST and Rotated BRIEF) [6]:** this method solves the lack of rotation invariance produced by BRIEF and improves the resistance to noise.
- **BRISK (Binary Robust Invariant Scalable Keypoints) [7]:** the methodology applied by this descriptor allows to obtain invariance to scale and rotation.
- **FREAK (Fast Retina Keypoint) [8]:** this algorithm computes a cascade of binary strings by comparing image intensities over a retinal sampling pattern.
- **LDB (Local Difference Binary) [9]:** this descriptor uses not only intensity information as BRIEF, but also gradient difference tests at several image granularities.

III. PROPOSED METHOD[1]

A. Image description

As stated in [12], high resolution images are not needed to accomplish an effective place recognition. For this reason and with the aim of reducing computational cost, we resize the sub-panoramas before applying the binary description by considering a patch of 64x64 pixels. According to our tests, a lower patch decreases the effectiveness of our loop closure detection algorithm, but a higher one does not substantially increase precision. In addition, the applied binary strings have a length of 32 bytes, as justified in Section V-A.

After processing the patch, we calculate a global binary descriptor by taking the center of the resized sub-panorama as a keypoint without dominant rotation or scale. Besides, we also test other approach, which consists in splitting the sub-panorama up into several grids and computing the binary descriptor by adopting the centers of the grids as keypoints. Better results are obtained by this second approach, but obviously there is an increase in computational cost and memory. Fig. 2 is presented for comparing both approaches.
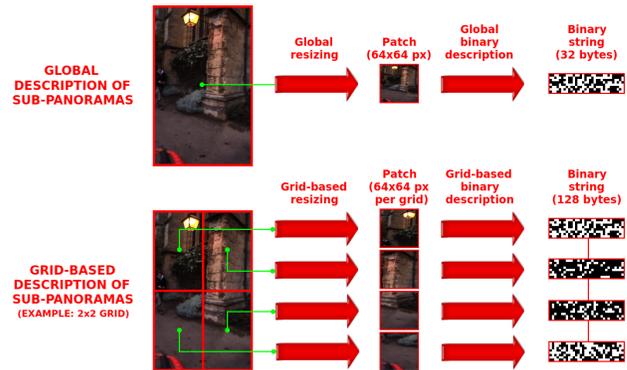

Fig. 2. Global and grid-based sub-panoramas description approaches.

For each sub-panorama (P_k) which composes each panorama (P), a binary string $(\mathbf{p}_k)$ is obtained. Finally, the strings computed for the total number of sub-panoramas (n) are concatenated $(+\!\!+)$ to form the final binary descriptor $(\mathbf{v}_i)$, as explained in Eq. 1. $\mathbf{v}$ is a vector which stores the final binary descriptors processed for each complete panorama.

$$\mathbf{v}_i = \mathbf{p}_1 +\!\!+ \mathbf{p}_2 +\!\!+ ... +\!\!+ \mathbf{p}_n \qquad (1)$$

B. Image matching for unidirectional loop closures

Loop closures can be identified by correlating the elements of $\mathbf{v}$ generated for each analyzed panoramic image and composing a distance matrix (M). For detecting only unidirectional loop closures as BRIEF-Gist does, an standard correlation based on a XOR operation $(\oplus)$ and a sum of bits is sufficient, as shown in Eq. 2.

$$M_{i,j} = M_{j,i} = \text{bitsum}(\mathbf{v}_i \oplus \mathbf{v}_j) \qquad (2)$$

However, the goal of this work is detecting not only unidirectional, but also bidirectional loop closures. Hence, we propose a new matching approach based on cross-correlation.

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

C. Image matching for bidirectional loop closures

Bidirectional loop closures provide a valuable information which can be employed to strengthen visual navigation. Taking this into account, we have developed an algorithm for identifying this kind of situations on panoramas, which give knowledge about numerous views of a place and procure a visual perception in different directions.

Our method is based on the cross-correlation of the binary strings obtained for each sub-panorama contained in a pair of panoramic images to be matched (i,j). The first step is to fill a cross-correlation matrix (C), whose elements correspond to the distances obtained for each matching between the different binary strings associated to each pair of sub-panoramas (k,l), as presented in Eq. 3. After that, the minimum value of C is calculated and stored in M as the final distance between i and j, as exposed in Eq. 4.

$$C_{k,l} = C_{l,k} = \text{bitsum}(\mathbf{v}_{i,k} \oplus \mathbf{v}_{j,l}) \tag{3}$$

$$M_{i,j} = M_{j,i} = \min(C) \tag{4}$$

For each new panoramic image processed in the sequence, Algorithm 1 is iterated to calculate the distances with respect to the previously analyzed images. When the sequence of panoramic images is completely processed, the different unidirectional loop closures appear clearly emphasized in M as right-side diagonals ($\searrow$) and the bidirectional ones as left-side diagonals ($\swarrow$).

Algorithm 1 Bidirectional loop closure detection on panoramas by applying a matching of binary descriptors based on cross-correlation.

Input: P {Panoramic image to be processed.}
           n {Number of sub-panoramas in P.}
           M {Accumulated distance matrix.}
           $\mathbf{v}$ {Accumulated vector of binary descriptors.}

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

Algorithm:

  $i \leftarrow \text{Length}(\mathbf{v}) + 1$; {Current panoramic image index.}

  for $k \leftarrow 1$ **to** n **do**
    $\mathbf{p_k} \leftarrow$ Calculate the binary string associated to P_k;
    $\mathbf{v}_i \leftarrow \mathbf{v}_i \mathbin{+\!\!+} \mathbf{p}_k$; {See Eq. 1.}
  end for

  for $j \leftarrow 1$ **to** i **do**
    for $k \leftarrow 1$ **to** n **do**
      for $l \leftarrow k$ **to** n **do**
        $C_{k,l} \leftarrow C_{l,k} \leftarrow \text{bitsum}(\mathbf{v}_{i,k} \oplus \mathbf{v}_{j,l})$; {See Eq. 3.}
      end for
    end for
    $M_{i,j} \leftarrow M_{j,i} \leftarrow \min(C)$; {See Eq. 4.}
  end for

IV. EVALUATION

A. Datasets

The datasets selected for evaluating our visual loop closure detection method are the Ford Campus and the Oxford New College. These datasets are both captured using panoramic cameras and have challenging situations for these kind of visual navigation problems.

The Ford Campus dataset comprises 7789 panoramic images (5,1 km) collected by an autonomous car. It contains several loop closures, but they are only unidirectional. For this reason, this dataset is employed only for testing the effectiveness of the different binary descriptors applied in this work.

The Oxford New College dataset comprises 8127 panoramic images (2.2 km) collected by a mobile robot. It contains several unidirectional and bidirectional loop closures. In consequence, this dataset can be used for evaluating our complete approach, including our matching based on cross-correlation for detecting bidirectional loop closures.

B. Methodology

We propose an evaluation technique common to all the tests presented in this paper with the aim of validating our proposal in an objective way. This methodology is employed to process the results obtained not only with our loop closure detection method, but also with the main state-of-the-art algorithms that we compare to ours, which are FAB-MAP and BRIEF-Gist. For testing FAB-MAP, we use an open source implementation called OpenFABMAP [21], which is applied in a standard configuration and conveniently trained in the datasets chosen. BRIEF-Gist is implemented using the BRIEF descriptor provided by OpenCV [22].

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 panoramic images. First of all, M is normalized by following Eq. 5, with the aim of comparing the results of the different algorithms in a suitable way. After that, M is thresholded to compare it to the ground-truth matrix (G) of each specific dataset. In this case, true positives are considered when a positive of the thresholded M coincides with a positive of G in a temporal vecinity of 7 panoramic images. The final precision-recall curve is computed by varying the threshold value (θ) in a linear distribution between 0 and 1 and calculating the corresponding values of precision and recall in each iteration. In this work, 100 values of θ are processed for each evaluation to obtain well-defined curves.

$$M_{i,j} = \frac{M_{i,j}}{\max(M)} \tag{5}$$

V. RESULTS

A. Results for only unidirectional loop closures

These first results are exclusively for situations with only unidirectional loop closures to analyze if in this basic case our algorithm improves the results achieved by FAB-MAP and BRIEF-Gist. Besides, we test the main recently appeared binary descriptors (defined in Section II-B) for determining which is the most suitable for our image description. With the aim of better referring to our approach, we make allusion to this method in the different tests as **ABLE** (**A**ble for **B**inary-appearance **L**oop-closure **E**valuation). Furthermore, in this case, it is applied on panoramas, so this specific approach is called **ABLE-P**.

On the one hand, Fig. 3 depicts the precision-recall curves obtained in the Ford Campus dataset, which has only unidirectional loop closures. On the other hand, Fig. 4 shows the results achieved in the Oxford New College dataset. Although this last dataset contains unidirectional and bidirectional loop closures, we consider only the unidirectional ones in this first evaluation.
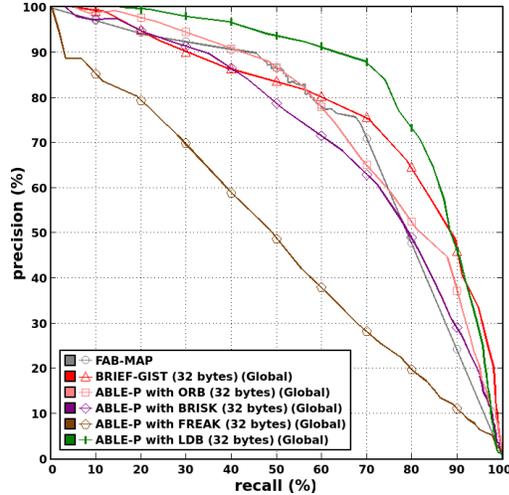

Fig. 3. Precision-recall curves for unidirectional loop closures in the Ford Campus dataset.
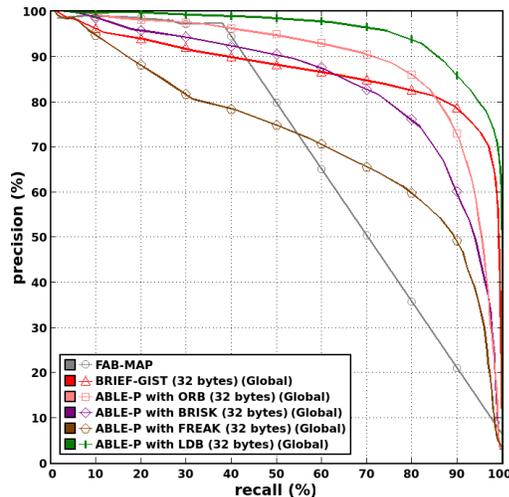
Table I confirms the successful results achieved by our method with LDB, which are supported by the comparison between the average precision of the different algorithms tested. In this case, average precision is defined as the percentage of total area under the precision-recall curve. Furthermore, the average processing time for each image description and matching is comparable to the attained by BRIEF-Gist. All these performance parameters have been obtained in the experiments carried out in the Oxford New College by utilizing a computer with the following features: an Intel Core i7 2,40 GHz processor and a 8 GB RAM.

In addition, Fig. 5 supports the decision of using binary strings of 32 bytes. A lower length slightly decreases the results, but a higher one does not increase precision.

TABLE I

COMPARISON OF SOME PERFORMANCE PARAMETERS BETWEEN THE
APPROACHES TESTED IN UNIDIRECTIONAL LOOP CLOSURE DETECTION.

		Average precision	Time per description	Time per matching
FAB-MAP		70.90 %	167.49 ms	$1.31 \cdot 10^{-2}$ ms
BRIEF-Gist		87.58 %	0.04 ms	$2.35 \cdot 10^{-5}$ ms
ABLE-P	**ORB**	89.33 %	0.17 ms	$2.33 \cdot 10^{-5}$ ms
	BRISK	84.38 %	0.18 ms	$2.61 \cdot 10^{-5}$ ms
	FREAK	72.36 %	0.38 ms	$2.66 \cdot 10^{-5}$ ms
	LDB	95.74 %	0.11 ms	$2.29 \cdot 10^{-5}$ ms

Fig. 4. Precision-recall curves for only unidirectional loop closures in the Oxford New College dataset.
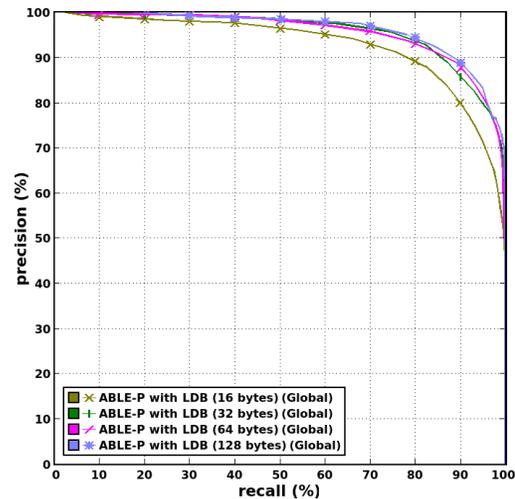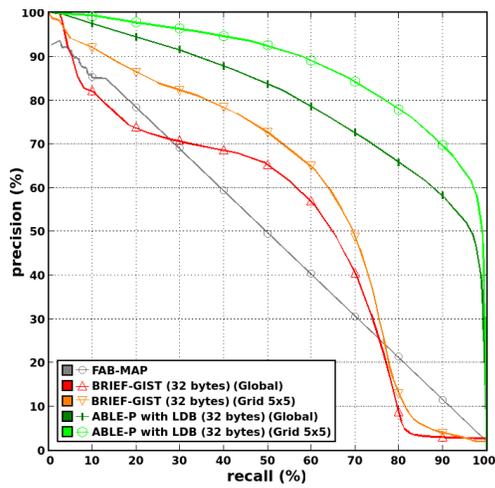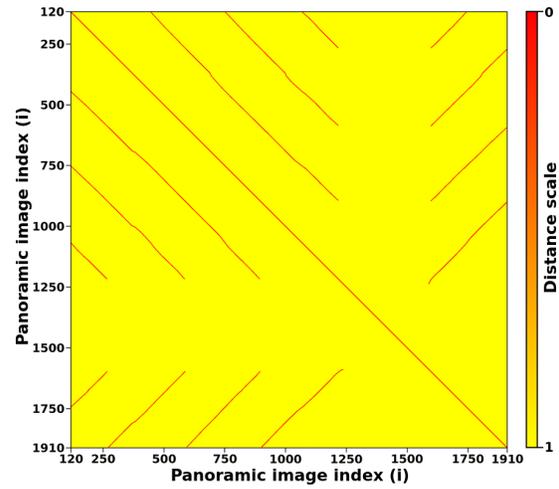

Fig. 5. Precision-recall curves for only unidirectional loop closures in the Oxford New College dataset using LDB binary strings of different lengths.

The results obtained in both cases demonstrate that the best performance is achieved by our method and the usage of LDB, probably due to the addition of gradient information that this descriptor incorporates with respect to ORB, BRISK or FREAK, which are mainly based on intensity. Besides, the invariance to rotation or scale does not affect excessively by following our global description approach. For this reason, the results attained by employing our method with ORB or BRISK are similar to the obtained by BRIEF-Gist, if only unidirectional loop closures are considered.

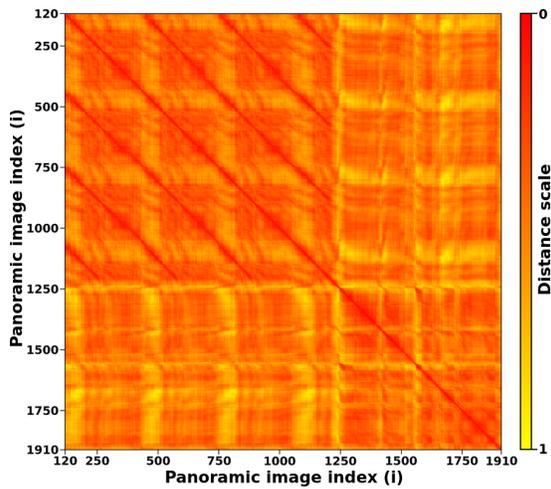### B. Results for unidirectional and bidirectional loop closures

The difference of results between our method with LDB and the state-of-the-art algorithms is completely corroborated if bidirectional loop closures are also considered. In this case, the precision-recall curves presented in Fig. 6 (a) reveal a higher contrast between our approach and FAB-MAP or BRIEF-Gist than the initially obtained in the Oxford New College if only unidirectional loop closures are taken into account (see also Fig 4 to perceive these differences).
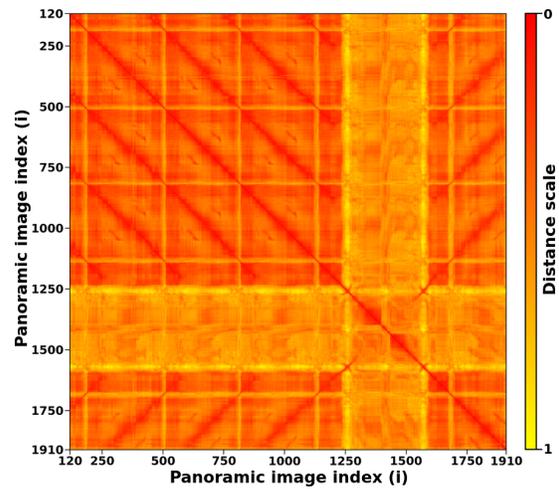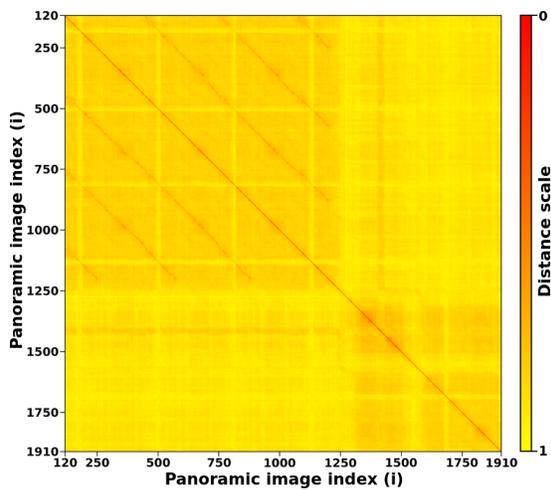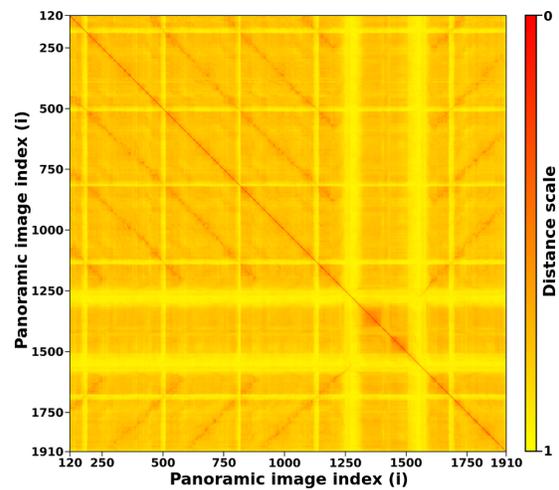
(a) Precision-recall curves.

(b) Ground-truth (G).

(c) M using BRIEF-Gist (Global).

(d) M using ABLE-P with LDB (Global).

(e) M using BRIEF-Gist (Grid 5x5).

(f) M using ABLE-P with LDB (Grid 5x5).

Fig. 6. Precision-recall curves for unidirectional and bidirectional loop closures in the Oxford New College dataset. Ground-truth and distance matrices between images 120 and 1910 are also presented, where several unidirectional and bidirectional loop closures appear represented in red (distance ≈ 0). Distance matrices demonstrate that our method can detect bidirectional loop closures while BRIEF-Gist can not detect them. FAB-MAP generates a confusion matrix instead of a distance matrix, which is not presented because of the worse results obtained for it in our evaluation, that are due to the low recall values achieved when $\theta \approx 1$.

These satisfactory results achieved by our method are due to the applied matching based in cross-correlation, which allows to correctly detect bidirectional loop closures. This can be observed if M obtained by BRIEF-Gist (Fig 6 (c)) and our method with LDB (Fig 6 (d)) are compared to G (Fig 6 (b)). Bidirectional loop closures approximately appeared between images 1590 and 1910 can not be identified by BRIEF-Gist, but our method clearly defines them in M. We show only a part of M for a representative subset between images 120 and 1910 because of the limitations of paper format, but results are similar for the complete Oxford New College, as deduced from the precision-recall curves presented in Fig. 6 (a).

Finally, we also depict the results attained by employing a grid of 5x5 cells in image description with our method and LDB. As can be observed in the precision-recall curves provided in Fig. 6 (a), our method experiences a slight improvement if this grid is applied instead of the global image description initially tested. This is also perceived if Fig. 6 (d) and Fig. 6 (f) are compared, due to the finer definition of loop closures which can be noted if a grid is used. According to our tests, a grid based on 5x5 cells is sufficient to improve results, because a higher grid does not practically improve precision any more and the computational cost is increased.

VI. CONCLUSIONS AND FUTURE WORKS

Along this paper, we have proved how our approach based on binary descriptors and cross-correlation of panoramas improves the results achieved in visual loop closure detection by other state-of-the-art methods such as FAB-MAP or BRIEF-Gist, specially for bidirectional loop closures. We refer to our specific approach as ABLE-P.

Binary descriptors are an excellent tool for describing places in visual loop closure detection algorithms because of its great precision and low computational cost. In this work, several binary descriptors are tested and LDB is chosen as the best option. In the near future, other description methods could be contemplated to improve the obtained results, such as the usage of 3D information in an efficient way with the aim of reaching a more complete environment perception.

The efficient detection of bidirectional loop closures in panoramas is a novel contribution due to the implementation of a matching strategy based on the cross-correlation of sub-panoramas. This is a successful, fast and easy solution to the described problem that can be applied to improve the performance of intelligent vehicles in visual navigation. In future works, similar approaches could be suggested for testing in recent datasets which are more complete and longer than the employed in this work, such as the KITTI Odometry [23] or the proposed in [20], which is focused in a life-long navigation during the seasons. The reason why our method is not evaluated in these datasets is because they use monocular or stereo cameras and, unfortunately, the approach proposed in this paper works only on panoramas. According to this, a question which should be solved in a near future is the next: It would be possible detecting bidirectional loop closures with the information provided by only monocular or stereo images? We expect to answer it soon.

REFERENCES

[1] D. Nistér, O. Naroditsky, and J. R. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics (JFR)*, vol. 23, no. 1, pp. 3–20, January 2006.

[2] C. Mei, G. Sibley, M. Cummins, P. Newman, and I. D. Reid, "RSLAM: A system for large-scale mapping in constant-time using stereo," *International Journal of Computer Vision (IJCV)*, vol. 94, no. 2, pp. 198–214, September 2011.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] A. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK: Fast retina keypoint," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, June 2012, pp. 510–517.

[9] 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.

[10] G. Pandey, J. R. McBride, and R. Eustice, "Ford campus vision and lidar data set," *International Journal of Robotics Research (IJRR)*, vol. 30, no. 13, pp. 1543–1552, November 2011.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] H. Badino, D. F. Huber, and T. Kanade, "Visual topometric localization," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2011, pp. 794–799.

[17] W. Churchill and P. Newman, "Continually improving large scale long term visual navigation of a vehicle in dynamic urban environments," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, September 2012, pp. 1371–1376.

[18] ——, "Practice makes perfect? Managing and leveraging visual experiences for lifelong navigation," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2012, pp. 4525–4532.

[19] H. Lategahn, J. Beck, B. Kitt, and C. Stiller, "How to learn an illumination robust image feature for place recognition," in *IEEE Intelligent Vehicles Symposium (IV)*, June 2013, pp. 285–291.

[20] 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.

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

[23] 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.