Indexing Invariant Features for Topological Mapping and Localization*

Emilio Garcia-Fidalgo and Alberto Ortiz[1]

Abstract— **We propose an appearance-based approach for topological visual mapping and localization using local invariant features. To optimize running times, matchings between the current image and previously visited places are determined using an index based on a set of randomized kd-trees. We use a discrete Bayes filter for predicting loop candidates, whose observation model is a novel approach based on an efficient matching scheme between features. We assess our approach with several datasets obtained from indoor and outdoor environments under different weather conditions.**

I. INTRODUCTION

In Simultaneous Localization and Mapping (SLAM), loop closure detection is a key problem to overcome. It entails the correct detection of previously seen places from sensor data. In the last decades, there has been a significant increase in the number of visual solutions for SLAM and loop closure detection because of the low cost of the cameras and the richness of the sensor data provided. This naturally guides us to an appearance-based SLAM, where the environment is represented using a topological map.

One of the most used techniques in appearance-based SLAM is the Bag-Of-Words (BoW) approach [1], [2], [3]. However, this method presents some drawbacks: it is more affected by the perceptual aliasing effect and typically an off-line training phase is needed. Other approaches make use of global descriptors, such as Gist [4], [5], [6], BRIEF-Gist [7] or WGOH [8]. The main drawback of these descriptors is that they are not descriptive enough, and thus they are more sensitive to noise, which leads to a larger number of incorrect detections. Rather than BoW or global descriptors, some authors used local invariant features for visual localization and mapping as well as for loop closure detection.

In this paper we present a complete visual mapping and localization framework based on raw local invariant features. Our framework was assessed using multiple indoor and outdoor datasets captured under different weather conditions and illumination changes. As main contributions, we present a Bayesian framework for visual loop closure detection which uses constellations of local invariant features as image descriptors. It comprises a novel observation model which allows us to succeed in challenging loop closure situations such as camera rotations, occlusions and changes in illumination. Using this algorithm as a key component, we also propose a topological mapping and localization framework. Our approach is independent of the robotic platform and can be used in several kinds of vehicles, e.g. ground, underwater or aerial vehicles. For further information about our approach, please see [9].

*This work is supported by the European Social Fund through the grant FPI11-43123621R (Conselleria d'Educacio, Cultura i Universitats, Govern de les Illes Balears) and by the FP7 project INCASS (GA 605200).

[1]Emilio Garcia-Fidalgo and Alberto Ortiz are with the Department of Mathematics and Computer Science, University of the Balearic Islands, 07122 Palma de Mallorca, Spain {emilio.garcia,alberto.ortiz}@uib.es.

II. TOPOLOGICAL MAPPING

Given an input image sequence, our approach is based on a subset of these images called *keyframes*. In our map, each node represents a keyframe image, and each keyframe is represented by its corresponding SIFT [10] features. In order to select these keyframes, we discard: (a) images similar to the current location of the robot (keyframe); and (b) robot camera turns. For the first case, SIFT features of the current image are matched applying the ratio test [10] to the features of the current location keyframe. If the number of matched features is higher than a threshold, the image is considered similar to the current location. The same matching step is applied between the current image and the last received image in the sequence: if it is not possible to match a certain number of features, the image is classified as a turn. In these two cases, the image is discarded. Otherwise, it is considered useful and needs to be processed in order to determine whether it is a loop closure or a new keyframe to be added to the map. Our loop closure approach makes use of a discrete Bayes filter. This filter is updated with every image irrespective of whether the image has been discarded or not. If a loop closure is not found, the current image is considered as a new keyframe and is added to the map as a new node. Otherwise, we create a link between the current location of the robot and the loop closure candidate and, then, a map refinement process is performed. The topological robot position within the map is updated accordingly. In order to avoid false loop closure detections between the current image and its neighbours in the sequence, new keyframes are not inserted directly as loop closure hypotheses in the filter. They are instead stored in a temporarily cache list and pushed into the filter once a certain number of images have been considered. Our approach is outlined in Fig. 1. The image description and matching process and the loop closure detection algorithm are detailed in the following sections.

A. Image Description and Matching

As commented above, in our approach, each image is described using the SIFT [10] algorithm. The loop closure detection algorithm, as we will see shortly, needs to match efficiently the features of the current image with features of all previously considered keyframes, in order to determine whether it is a revisited place. Therefore, a method for

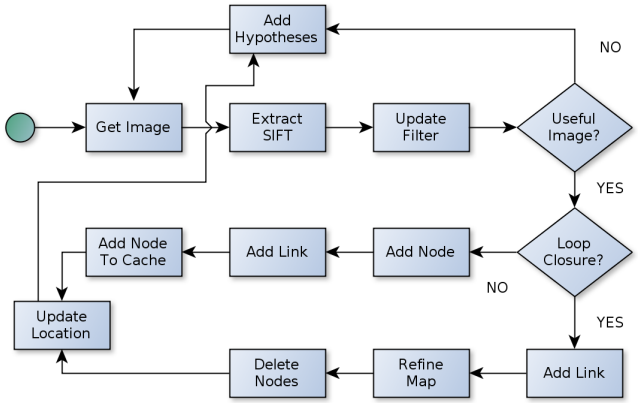

Fig. 1. Overall algorithm diagram. See text for details.

an efficient nearest neighbour search is needed in order to match these high-dimensional descriptors. Tree structures have been widely used to this end, since they reduce the search complexity from linear to logarithmic. To the same purpose, we maintain a set of randomized kd-trees containing all the SIFT descriptors of previously detected keyframes. An inverted index structure, which maps each feature to the keyframe where it was found, is also created. Given a query descriptor, these structures allow us to obtain, traversing the tree just once, the top K nearest keypoints among all keyframes in an efficient way.

B. Probabilistic Loop Closure Detection

A discrete Bayes filter is used to detect loop closure candidates. This filter estimates the probability that the current image closes a loop with previously seen locations, allowing us to deal with noisy measurements and uncertainty in the robot location and helping us to discard false recognitions. The Bayesian framework is also used for ensuring temporal coherency between consecutive predictions, integrating past estimations over time.

Given the current image I_t at time t, we denote z_t as the set of SIFT descriptors extracted from this image. These are the observations in our filter. We also denote L_i^t as the event that image I_t closes a loop with image I_i, where $i < t$. Using these definitions, we want to detect the image of the map I_c whose index satisfies:

$$c = \underset{i=0,\dots,t-p}{\arg\max} \left\{ P\left(L_i^t | z_{0:t}\right) \right\}, \tag{1}$$

where $P\left(L_i^t | z_{0:t}\right)$ is the full posterior probability at time t given all previous observations up to time t. As in [3], the most recent p images are not included as hypotheses in the computation of the posterior since I_t is expected to be very similar to its neighbours and then false loop closure detections will be found. This parameter p delays the publication of hypotheses and needs to be set according to the frame rate or the velocity of the camera. The posterior can be derived as:

$$P\left(L_i^t | z_{0:t}\right) = \eta P\left(z_t | L_i^t\right) \sum_{j=0}^{t-p} P\left(L_i^t | L_j^{t-1}\right) P\left(L_j^{t-1} | z_{0:t-1}\right), \tag{2}$$

where η represents the normalizing factor, $P\left(z_t | L_i^t\right)$ is the observation likelihood, $P\left(L_j^{t-1} | z_{0:t-1}\right)$ is the posterior distribution computed at the previous time instant and $P\left(L_i^t | L_j^{t-1}\right)$ is the transition model. See [9] for a detailed decomposition of the posterior.

1) Transition Model: Before updating the filter using the current observation, the loop closure probability at time t is predicted from $P\left(L_j^{t-1} | z_{0:t-1}\right)$ according to an evolution model. The probability of loop closure with an image I_j at time $t-1$ is diffused over its neighbours following a discretized Gaussian-like function centered at j. In more detail, 90% of the total probability is distributed among j and exactly four of its neighbours. The remaining 10% is shared uniformly across the rest of loop closure hypotheses according to $\frac{0.1}{\max\{0, t-p-5\}+1}$. This implies that there is always a small probability of jumping between hypotheses far away in time, improving the sensitivity of the filter when the robot revisits old places.

2) Observation Model: Once the prediction step is performed, the current observation needs to be included in the filter. We have to compute the most likely locations given the current image I_t and its keypoint descriptors z_t, but we want to avoid comparing I_t with each previous keyframe, since this is not tractable. To this end, we use the structures described in section II-A.

For each hypothesis i in the filter, a score $s\left(z_t, z_i\right)$ is computed. Initially, these scores are set to 0 for all frames from 0 to $t-p$. For each descriptor in z_t, the K closest descriptors among the previous keyframe images are retrieved; next, each of them, denoted by n, adds a weight w_n to the score of the image where it appears. This value is normalized using the total distance of the K candidates retrieved:

$$w_n = 1 - \frac{d_n}{\sum_{k \in K} d_k}, \forall n = 1, \dots, K, \tag{3}$$

where d is the Euclidean distance between the considered query descriptor in z_t and the nearest neighbour descriptor found in the tree structure. This value is accumulated onto a score according to:

$$s\left(z_t, z_{j(n)}\right) = s\left(z_t, z_{j(n)}\right) + w_n, \forall n = 1, \dots, K, \tag{4}$$

being $j(n)$ the index of the image where the candidate descriptor n was extracted. The computation of the scores is finished when all descriptors in z_t have been processed. Then, the likelihood function is calculated according to the following rule [3]:

$$P\left(z_t | L_i^t\right) = \begin{cases} \frac{s(z_t, z_i) - s_\sigma}{s_\mu} & \text{if } s\left(z_t, z_i\right) \geq s_\mu + s_\sigma \\ 1 & \text{otherwise} \end{cases}, \tag{5}$$

being respectively s_μ and s_σ the mean and the standard deviation of the set of scores. After incorporating the observation

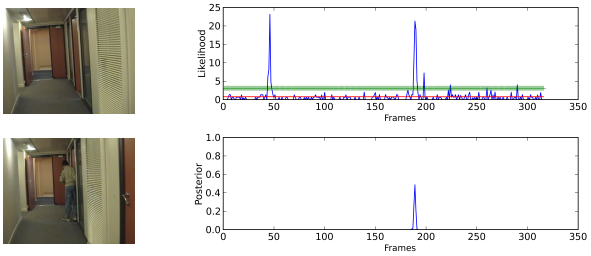

Fig. 2. Example of loop closure detection visiting several times the same place and with changes in the environment in the Lip6Indoor dataset. Image 331 (Top, Left) closes a loop with image 189 (Bottom, Left) and image 48 (not shown). As can be seen in (Top, Right), the current likelihood presents two strong peaks despite a person in the current image occludes part of the same. Peaks correspond to loop candidates. After the normalization step, the posterior (Bottom, Right), shows a single peak in the last candidate frame. Red and green lines show respectively s_μ and $s_\mu + s_\sigma$ values.

to our filter, the full posterior is normalized in order to obtain a probability distribution.

3) Selection of a Loop Closure Candidate: In order to select a final candidate, we do not search for high peaks in the posterior distribution, because loop closure probabilities are usually diffused between neighbouring images. Instead, for each location in the filter, we sum the probabilities along the same neighbourhood as defined in section II-B.1. The image I_j with the highest sum of probabilities in its neighbourhood is selected as a loop closure candidate. If this probability is below a threshold T_{loop}, the loop is not accepted. Otherwise, an epipolarity analysis between I_t and I_j is performed in order to validate the candidate. Matchings that do not fulfill the epipolar constraint are discarded by means of RANSAC. If the number of surviving matchings is above a threshold T_{ep}, the loop closure hypothesis is accepted; otherwise, it is definitely rejected. Finally, we define another threshold T_{hyp} to ensure a minimum number of hypotheses in the filter, so that loop closure candidates are meaningful.

III. EXPERIMENTAL RESULTS

Our approach has been assessed using five datasets conforming more than 3000 images. These datasets were obtained from indoor and outdoor environments, under different environmental conditions. Fig. 2 shows the suitability of the Bayes framework in a loop closure detection situation. In this case, the camera visited twice the same place. When it returns to this place again, two high peaks corresponding to the previous visits can be observed in the likelihood, representing possible loop candidates for the current image. After the prediction, update and normalization steps, the posterior presents only one single peak at the second candidate image, i.e. the filter ensures temporal coherency between predictions. This figure also shows an example of situation where a loop is detected despite there is a person in the image who was not in the previous visit, what proves the ability of the filter for detecting loops when the appearance of the environment changes. Our approach accepts the loop closure since the epipolar constraint between the two images is satisfied.

In order to obtain global performance measures, each dataset was provided with a ground truth, which indicates, for each image in the sequence, which images can be considered as a loop closure with it. The assessment against this ground truth has been performed counting for each sequence the number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), where positive is meant for detection of loop closure. Then, the two following metrics were computed:

- *Precision*. Ratio between real loop closures and total amount of loop closures detected $\left(\frac{TP}{TP+FP}\right)$.
- *Recall*. Ratio between real loop closures and total amount of loop closures existing in the sequence $\left(\frac{TP}{TP+FN}\right)$.

The results for each sequence are shown in Table I. As can be seen, no false positives resulted in any case. This is essential, since false positives can induce errors in mapping and localization tasks. As a consequence, the classifier always reaches 100% in precision for all datasets. The best recall rates for 100% precision are shown in the table. A high rate of correct detections were obtained from all experiments. False negatives are due to, on the one hand, the sensitivity of the filter. In effect, when an old place is revisited, the likelihood associated to that hypothesis needs to be higher than the other likelihood values during several consecutive images in order to increase the posterior for this hypothesis. This introduces a delay in the loop closure detection, which derives in false negatives. This sensitivity can be tuned by modifying the transition model of the filter, although a higher sensitivity can introduce loop detection errors, i.e. false positives. On the other hand, false negatives can also be due to camera rotations. When the camera is turning around a corner, it is difficult to find and match features in the images, which prevents the hypothesis from satisfying the epipolar constraint and leads to the loop closure hypothesis to be rejected, despite the posterior for this image is higher than T_{loop}. However, in spite of the difficulties of the UIBIndoor dataset, our approach is able to succeed, as can be seen in Table I.

The path followed by the camera in one of the datasets is shown in Fig 3. Whenever the camera explores new places, no loop closures are found. When a place is revisited, the algorithm starts to find loop closures. Several images are usually needed until closing the loop, due to the filter inertia. These images correspond to the false negatives found.

We also validate our framework for mapping and localization. To this end, the loop closure detection algorithm was adapted to be used with the detected keyframes. An example of these maps is shown in Fig. 4. The main zones of this map were labelled with letters to simplify the identification of each part in the topological structure, since the resulting topological map do not preserve the shape.

IV. CONCLUSIONS

A complete appearance-based mapping and localization framework based on local invariant features is presented here.

TABLE I

RESULTS FOR THE FIVE DATASETS. ^{a}AVERAGE FOR ALL SEQUENCES.

Dataset	#Imgs	Size	TP	TN	FP	FN	Pr	Re
Lip6Indoor	388	240×192	191	151	0	31	100	86
Lip6Outdoor	1063	240×192	551	435	0	52	100	91
UIBSmallLoop	388	300×240	194	172	0	2	100	99
UIBLargeLoop	997	300×240	439	491	0	47	100	90
UIBIndoor	384	300×240	157	177	0	30	100	84
	3220		1532	1426	0	162	100^a	90^a

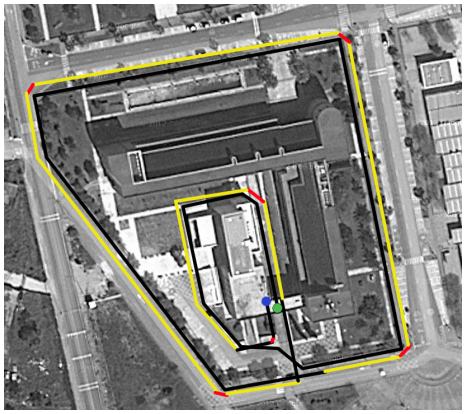

Fig. 3. Path followed by the camera during the UIBLargeLoop experiment. Green and blue points indicate respectively the beginning and the end of the sequence; the black lines show no loop closure detections (highest posterior probability is under T_{loop}), the red lines show rejected hypoteses (no epipolar geometry is satisfied) and the yellow lines represent loop closure detections (highest probability is above T_{loop} and the epipolar constraint is satisfied). Notice that the camera passes through the same place in successive loops, but the lines are drawn in parallel for visualization purposes.

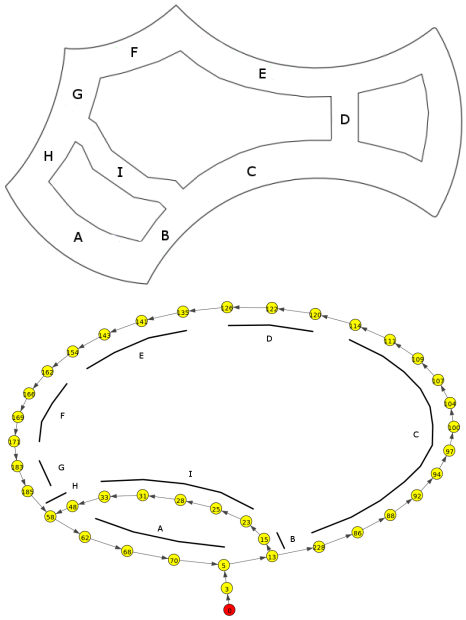

Fig. 4. (Top) Reference map for the Lip6Indoor dataset. (Bottom) Topological map generated using our approach. Each part of the map is identified with a letter in both maps. The red node identifies the beginning of the sequence. Maps locations are visited in the following order: A-B-I-H-A-B-C-D-E-F-G-H-A-B-C-D-E-F-G-H-A-B-I.

When a new useful image is acquired, a discrete Bayes filter is used to select a loop closure candidate and decide whether this frame is a loop closure or a new node to be added to the map. This probabilistic filter presents a novel observation model based on an efficient matching scheme between the current image and the features of the current nodes in the map, using an index based on a set of randomized kd-trees. As a result, a topological map of the environment is obtained, which represents the scenario of the robot as a graph.

In order to validate our solution, results from an extensive set of experiments, using datasets from different environments, have been reported. These results are very promising, showing that our mapping and localization approach can be employed for generating topological maps of the environment that, if they are provided with odometry information, can also be used for navigating in the current scenario in an efficient way.

Referring to future work, we intend to explore: (a) the use of other kinds of image descriptors based on local invariant features, such as binary descriptors, since they can improve our approach in computational terms; (b) the execution of the Bayes filter in a Graphics Processing Unit (GPU) to further speed up the loop closure detection; and (c) the use of the full algorithm for mapping larger environments, since unused descriptors in the tree structures should be purged to maintain a reasonable response time of the loop closure detection algorithm.

REFERENCES

[1] F. Fraundorfer, C. Engels, and D. Nister, "Topological Mapping, Localization and Navigation Using Image Collections," in *IROS*, 2007, pp. 3872–3877.

[2] M. Cummins and P. Newman, "Probabilistic Appearance Based Navigation and Loop Closing," in *ICRA*, 2007, pp. 2042–2048.

[3] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "A Fast and Incremental Method for Loop-Closure Detection Using Bags of Visual Words," *Trans. Rob.*, vol. 24, no. 5, pp. 1027–1037, 2008.

[4] A. Oliva and A. Torralba, "Modeling the Shape of the Scene : A Holistic Representation of the Spatial Envelope," *Int. J. Comp. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[5] G. Singh and J. Kosecka, "Visual Loop Closing using Gist Descriptors in Manhattan World," in *Workshop on Omnidirectional Robot Vision*, 2010.

[6] Y. Liu and H. Zhang, "Visual Loop Closure Detection with a Compact Image Descriptor," in *IROS*, 2012, pp. 1051–1056.

[7] N. Sunderhauf and P. Protzel, "BRIEF-Gist - Closing the Loop by Simple Means," in *IROS*, 2011, pp. 1234–1241.

[8] D. Bradley, R. Patel, N. Vandapel, and S. Thayer, "Real-Time Image-Based Topological Localization in Large Outdoor Environments," in *IROS*, 2005, pp. 3670–3677.

[9] E. Garcia-Fidalgo and A. Ortiz, "Vision-Based Topological Mapping and Localization by means of Local Invariant Features and Map Refinement," *Robotica*, 2014.

[10] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comp. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.