RAIRO-Oper. Res. 44 (2010) 365–377
DOI: 10.1051/ro/2010021

RAIRO Operations Research
www.rairo-ro.org

COMBINING ODOMETRY AND VISUAL LOOP-CLOSURE DETECTION FOR CONSISTENT TOPO-METRICAL MAPPING

S. Bazeille[1] and D. Filliat[1]

Abstract. We address the problem of simultaneous localization and mapping (SLAM) by combining visual loop-closure detection with metrical information given by a robot odometry. The proposed algorithm extends a purely appearance-based loop-closure detection method based on bags of visual words [A. Angeli, D. Filliat, S. Doncieux and J.-A. Meyer, *IEEE Transactions On Robotics, Special Issue on Visual SLAM* **24** (2008) 1027–1037], which is able to detect when the robot has returned back to a previously visited place. An efficient optimization algorithm is used to integrate odometry information and to generate a consistent topo-metrical map much more usable for global localization and path planning. The resulting algorithm which only requires a monocular camera and robot odometry data, is real-time, incremental (*i.e.* it does not require any *a priori* information on the environment), and can be easily embedded on medium platforms.

Keywords. SLAM, monocular vision, odometry, mobile robot, topo-metrical map.

Mathematics Subject Classification. 68T40, 93C85.

1. Introduction

To navigate in their environment, humans and animals use several strategies, from reactive guidance towards a visible goal to larger scale planning to reach distant goals. These last strategies require the cognitive ability to build a map

Received February 26, 2010. Accepted July 26, 2010.

[1] ENSTA ParisTech, Unité Électronique et Informatique – 32 Boulevard Victor, 75015 Paris, France. [stephane.bazeille;david.filliat]@ensta.fr

and to self-localize in it [31]. Maps-based navigation seems quite natural to humans because using a map is a very convenient way to describe an environment but it requires a lot of high level cognitive processes in order to interpret the map and to establish correspondences with the real world. However, many ethological and neurological studies showed that animals made also use of maps for navigation.

Building and using maps is based on two distinct sources of information. The first is the internal information about the movements: speed, acceleration, leg movements. The second provides external information about the environment. It may be derived from vision, odor, or touch. For the animals, the integration of these information for map building appear to take place in a part of the brain called hippocampus [12]. The navigation problem for robots is very similar and make use of the same information (*e.g.* odometry calculated from wheel rotation and perceptions taken rotation and laser-range finders or camera), which lead several author to propose navigation systems for robots inspired by neurobiological findings (*e.g.* [22]). The approach proposed in this paper is not directly inspired by biology, but has some key similarities with biological systems by using the same subjective information and being completely autonomous and incremental without requiring any information that would not be available for a human or an animal in the same scenario.

In Sections 2 and 3 we present some related work and our previous work on topological SLAM. In Section 4 we describe the new topo-metrical framework. In Section 5 we show some experimental results and we finish in Section 6 by discussing about this work and presenting some future work.

2. Related work

Over the last years, the increase in computing power pushed forward the use of visual information in robotic applications. The camera sensor is often used to replace the traditional range and bearing sensors because it provides many advantages such as smaller size, lighter weight, lower energy consumption, and above all a richer environmental information. The vision sensor is suitable for many robotic applications such as user interaction or object and place recognition [1,4], and has also been used in many SLAM solutions (*e.g.* [5,13,29]). SLAM [2] is the process of localizing a mobile robot while concurrently building a map of its environment.

The field of SLAM can be divided into topological and metrical approaches. The topological approach which models the environment as a graph of discrete locations often leads to simpler solutions. It is an abstract representation giving just relations between environment locations. It is an easy to build map, suitable for many kinds of environment and for human interactions. Its main drawback comes from lack of geometric information that only allows a global localization in previously mapped areas and local navigation with non optimal path planning. On the contrary, the metrical map is explicitly based on measures (distances, positions, lengths). The representation of the environment is geometric and clearly

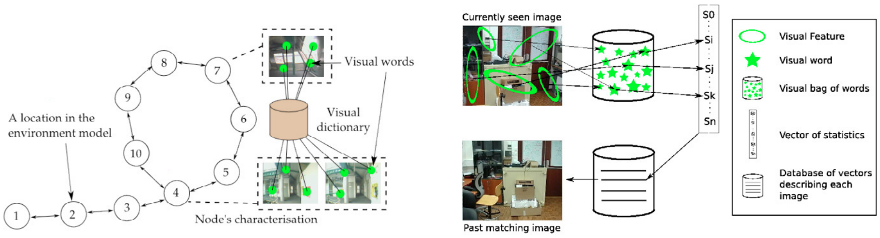

FIGURE 1. (a) An example of topological map. (b) Inverted index structure used to estimate the likelihood between the current image and images stored in the map.

corresponds to the real world. The localization can be done continuously and planned navigation is easier. The major problem of this kind of map is due to the required geometric consistency between position and perception which makes the map hard to build. To overcome those problem, number of approaches have attempted to use a combination of the two representations. For example, metrical maps can be embedded into graphs of higher level to enhance scalability [9] or other graph-based solutions can be used to infer a precise metrical position for the robot, while still allowing for large scale mapping [18]. In our previous work, we have developed [1] a real-time vision-based topological SLAM which presents many advantages such as its incremental, real-time and easy embedded, but the topological map created is not suitable for robot guidance (Fig. 1) because localization is only possible in previously mapped areas and no information is stored about the guidance of the robot between places. To enable global localization and navigation we decide to extend our topological map to a topo-metrical maps adding the metrical information given by the odometry data. The integration of metrical information to the existing topological map and loop-closure detection algorithm can be done in several ways. The most appealing solution to this problem is probably the use of visual odometry, where images coming from neighboring nodes or image sequences taken between nodes are matched to estimate the robot displacement [13,17,24,28,29]. Instead of estimating node positions, another solution is to use visual servoing, also known as vision-based robot control which uses feedback information extracted from a vision sensor to control the motion of a robot [6]. The robot can then be directly guided to the neighboring nodes without explicitly computing their relative positions. The advantage of these solutions is to use only vision, but they require a lot of processing and are not robust in absence of visual information, in dark areas for example. Like several authors [8,11,27], we have chosen the simpler solution of using the information given by robot odometry. Odometry is often provided by robots, whether they be legged or wheeled, to estimate their position relative to a starting location. The main drawback of odometry is the continuous growth of error in the position estimate due to the integration of noisy measurements over time. As a consequence, efficiently using odometry information requires complementary information to enable a correction of this cumulative drift errors. This correction can be obtained through the position constraint given by the visual loop-closure detection when the robot has returned at the position of a previous passing. These constraints, integrated through the application of a relaxation algorithm, will make it possible to build a topo-metrical map globally consistent. Different relaxation methods exist to deal with this problem [7,14,16]. To solve the particular graph-based formulation of SLAM problem in which the poses of the robots are modeled by nodes in a graph, and constraints between poses resulting from observations or from odometry are encoded in the edges between the nodes, recent solutions are more efficient [15,25].

3. Previous work: topological SLAM using Bayesian filtering

Several vision-based techniques consider the problem of topological SLAM [13, 26] or topological localization [3,32], the main idea is to seek for the past images that look similar to the current one and consider they come from close viewpoint. To solve the image-to-node matching problem (based on a similarity measure between the current image and the images of a node previously visited), we decided to use a maximum *a posteriori* scheme which exploits the similarity of image sequences to reduce false alarms and ensure the temporal consistency of the estimation (*e.g.* [21]). A complete description of the approach is given in [1], but a short overview is provided here for clarity.

This method searches for the node N_i of the map that is the more similar to the current image I_t, in other words, it searches for the node N_i that maximizes the probability of loop-closure with the current image:

$$N_i = argmax_{i=0,\ldots,n} p(S_t = i | I_t, M) \tag{1}$$

where $S_t = i$ is event "I_t comes from N_i" and $M = N_0, ..., N_n$ is the map of the environment. Bayes rule, marginalization and Markov assumption [1] lead to the incremental computation of the *a posteriori* probability as follow:

$$p(S_t|I_t, M) = \eta . \underbrace{p(I_t|S_t, M)}_{\text{likelihood model}} . \sum_{j=0}^{n} \underbrace{\underbrace{p(S_t|S_{t-1} = j, M)}_{\text{transition model}} \underbrace{p(S_{t-1} = j|I_{t-1}, M)}_{a\ priori\text{ probability}}}_{\text{prediction}} . \tag{2}$$

In this equation, the prediction is computed using the *a priori* probability (*i.e.* the probability at the previous time step) multiplied by an evolution model diffusing the probability of a node to its neighbors to take into account the robot motion since last localization. Then, the result of this computation called prediction is multiplied by the likelihood (number of correspondences between images through a voting scheme) to obtain the *a posteriori* probability. The likelihood model is computed using a representation of images as a set of unordered elementary visual features (SIFT [20] and local color histograms) taken from a dictionary (*i.e.* the bags of visual words model [10]). An inverted index makes it possible to very efficiently compute this likelihood in time linear with the number of visual words of the current image (Fig. 1b).

To discard outliers we select the best loop-closure candidates after voting with a multiple view geometry stage like [19,32]. In details we sort the probabilities higher than a threshold and we verify in the descending order the potential loop-closure detection by multiple-view geometry [23] (we try to find the fundamental matrix between the two loop-closure images).

The presented topological SLAM method is simple real-time and fully incremental (*i.e.* the environment model is learned on-line as the robot discovers its surroundings) and uses only appearance information from a single camera to build a topological map of the places the robot is visiting. But it suffers from the lack of metrical information that makes the map ill-posed for robot guidance.

4. System overview

To add the metrical information in the map and make the graph more usable for global localization and navigation, we have chosen the simpler solution of using the information given by odometry data. Because we dispose of such an information we also change the Bayesian framework including a much more informative evolution model to make the prediction more accurate and so make the loop-closure detection reliable. An overview of the new algorithm is detailed in Figure 2. Although this solution requires a second sensor for odometry, the information provided also efficiently complements the image data in situations where visual information is unusable or unavailable (*e.g.* sensor occlusion, strong lighting change, dark areas), giving an estimating of the robot position. The inclusion of metric information in our previous algorithm to obtain a consistent topo-metrical mapping required four main modifications:

- images previously acquired at 1 Hz are now acquired with the relative odometry when the robot has moved enough from its previous position (images do not need to be processed when the robot do not move);
- the Gaussian transition model is replaced by an odometry based transition model in the Bayesian filter;
- the geometry validation stage is modified to constrain loop-closure to be detected only for very close locations;
- the relative position between nodes is recorded on each link of the graph and we apply a relaxation algorithm each time a loop-closure is detected to correct cumulative odometry drift.

4.1. Adding coherent metrical information

The topological map is a graph constituted of a set of nodes associated with an image and linked by edges. We integrated metrical information in two forms in order to produce a topo-metrical map. First, each node is associated with an absolute position in the map (x, y, θ), where x and y are the 2D position

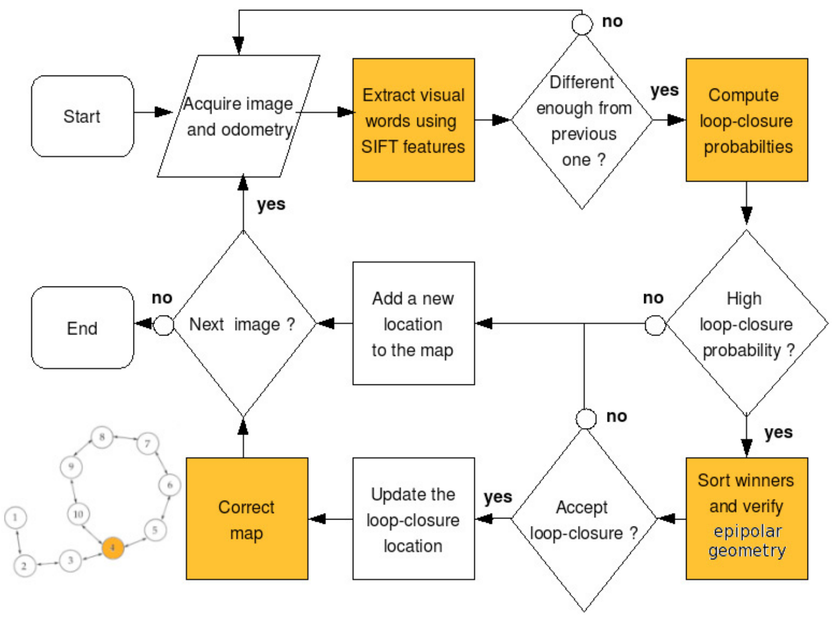

FIGURE 2. Processing diagram of the topo-metrical map construction.

coordinates and θ an angle representing the direction of the robot when the image was taken. Secondly, the edges are associated with a relative position between two nodes defined by (d, α, ϕ), where d and α are the polar coordinates of the second node in the coordinate space of the first, and ϕ is the difference angle between the two nodes direction.

During the localization and mapping process, each time a new image is acquired, a new location is created. When a loop-closure is detected this location is added as a similar location to the existing loop-closing node and we apply a relaxation algorithm to estimate the position of nodes that best satisfied the loop-closure constraints (Fig. 3). The relaxation algorithm we choose is the Tree-based network optimizer (TORO) [15], because of its speed and its high efficiency. TORO is an extension of Olson's algorithm [25] which introduced a tree-based parametrization for the nodes in the graph. It is based on a graph-formulation of the SLAM problem and applies a gradient descent-based optimization schemeto estimate the consistent node configuration which maximally satisfies the odometry constraints between nodes.

4.2. Including odometry in the evolution model

In the original framework, the evolution model used to obtain the prediction given the *a priori* probability applied a diffusion of the probability over the neighboring locations in the graph. The weight was defined as a sum of Gaussian

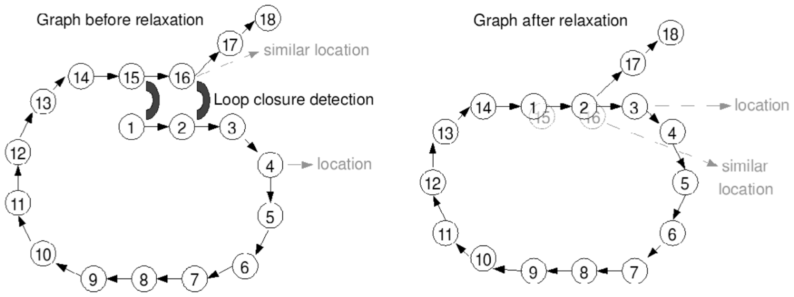

FIGURE 3. Illustration of the graph relaxation process.

centered on the current location (Fig. 5, top). The limitation of this model is that diffusion is done in all directions without preference, because it only assumes that the neighboring images in time are close together, without any information about the real robot movement.

Because a reliable metrical information is now available, we can integrate odometry in the evolution model to predict more precisely the evolution of the probability. Thus, the evolution model can then takes into account not only the nodes topological proximity, but also their relative position. To do so, starting from a given node, we distribute the probability to each neighboring location in the map depending on the deviation of these nodes relative positions with the robot displacement since the last update d_u, α_u, ϕ_u measured by odometry (Fig. 5, bottom). We used the standard motion model for robot odometry [30], assuming gaussian noise on the robot displacement measured in polar coordinates:

$$p(d, \alpha, \phi | d_u, \alpha_u, \phi_u) = G_{\mu_d,\sigma_d}(d - d_u) G_{\mu_\theta,\sigma_\theta}(\alpha - \alpha_u) G_{\mu_\phi,\sigma_\phi}(\phi - \phi_u) \tag{3}$$

where d, α gives the odometry displacement in polar coordinates in the frame of the previous robot position and ϕ is the variation of robot direction during movement. $G_{\mu,\sigma}(X)$ is the gaussian distribution of mean μ and variance σ^2. Using this model, the evolution model becomes:

$$p(S_i | S_j, u_t, M) = G_{\mu_d,\sigma_d}(d_{ij} - d_u) G_{\mu_\theta,\sigma_\theta}(\theta_{ij} - \theta_u) G_{\mu_\phi,\sigma_\phi}(\phi_{ij} - \phi_u) \tag{4}$$

where $u_t = d_u, \theta_u, \phi_u$ gives the odometry displacement and $d_{ij}, \theta_{ij}, \phi_{ij}$ is the relative position between nodes i and j. The substitution makes the prediction of the *a posteriori* probability more precise, improving robustness and responsiveness of the algorithm. This will make it possible to enhance the reactivity of loop-closure detection, which required several consecutive effective loop-closure before detection in the original approach.

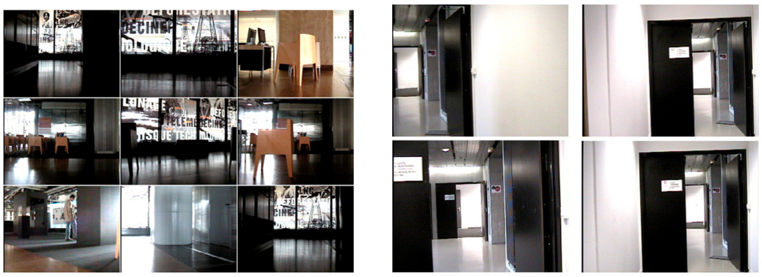

FIGURE 4. (a) Samples of the images used in our experiment (Museum). (b) Examples of the loop-closure validation images (on two couples of images previously accepted by the epipolar geometry the first one is now rejected with the new validation stage) (Office sequence).

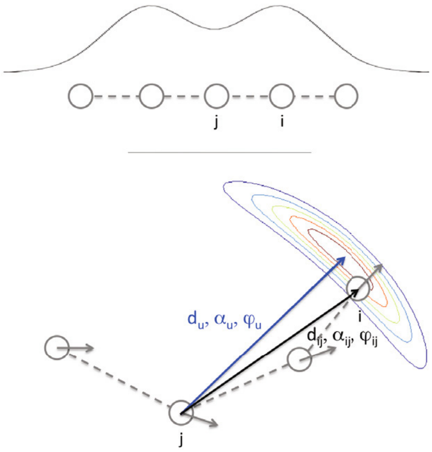

FIGURE 5. Illustration of the modification of the evolution model. *Top*: the original model only takes the graph connectivity into account when propagating probability from node j to node i. *Bottom*: including odometry, the new evolution model is more precise and preferentially propagates probability from node j to the nodes i that corresponds to a movement coherent with the odometry.

4.3. Validating loop-closures

When a loop-closure is detected by the Bayesian filter, the robot is assumed to have returned exactly at the position of a previous passing. By constraining the two nodes of the graph to have the same position, we correct the cumulative noise of odometry but we make the map incoherent if the loop-closure is not valid (false

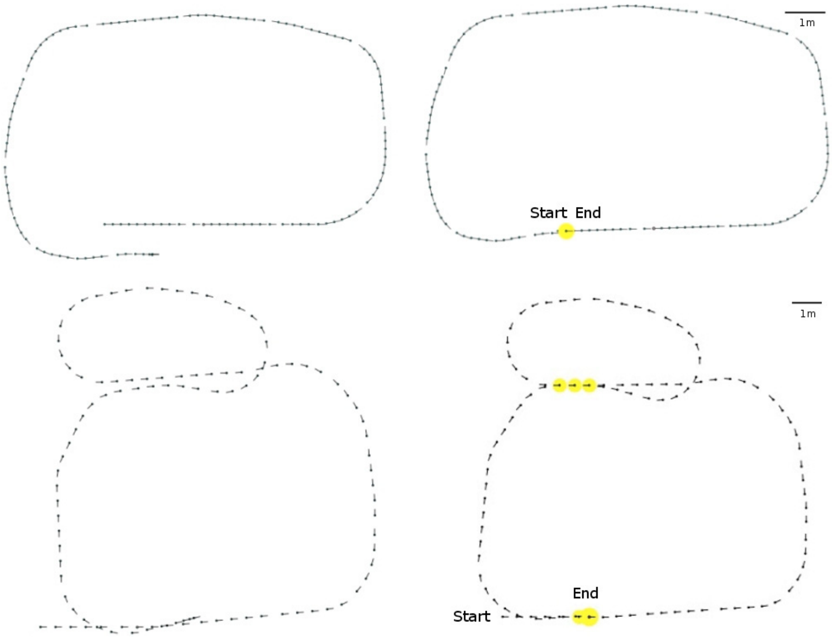

FIGURE 6. Two topo-metrical maps generated though our algorithm (Museum sequences). On the left the raw odometry, on the right the corrected topo-metrical map.

alarm) or not accurate (important translation or rotation between the two images). That is why the acceptance policy of loop-closure has been modified to only accept loop-closure with very close views thereby allowing only small variations between the corresponding positions and orientations (Fig. 4, right). This acceptance policy requires that 90% of the SIFT points matched between the two images validate the epipolar geometry constraints, and additionally, that the total displacement of these points in the image space is below a threshold.

5. Experimental results

To demonstrate the quality of the approach we have used data acquired with a Pioneer 3 DX mobile robot. The robot was guided to do some loops in an indoor environment showing strong perceptual aliasing conditions (several distinct places looks similar). Figure 4 shows image samples taken from the run (Museum and Office). As a landmark we stop the run precisely on the path previously taken (and with the same direction). The images and the odometry relative information were taken each time the robot moves at least 50 cm or turns of at least 30 degrees (Figs. 6, right, 7, top). We can clearly see in Figure 7 the work of the relaxation to correct the odometry drift.

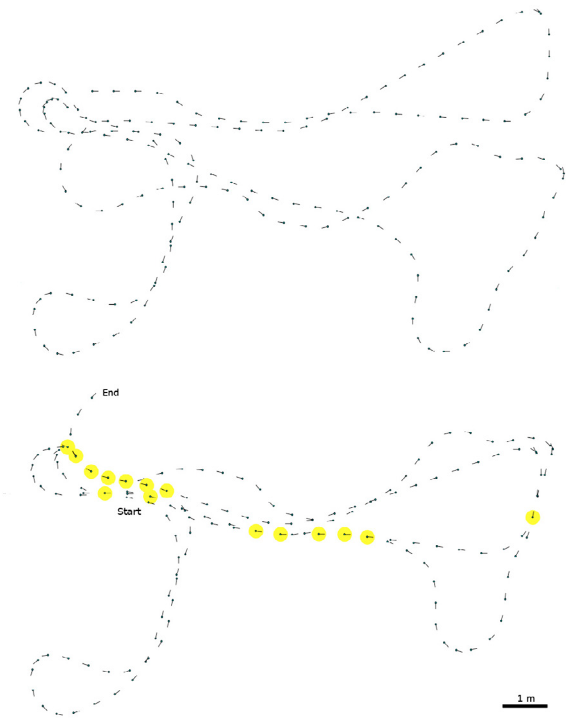

FIGURE 7. Example of topo-metrical map generated (Office sequence).

Besides the build of consistent topo-metrical map, the use of odometry in the evolution model improves the responsiveness and the robustness of the algorithm: during the experiment, only two consecutive similar frames are now required before effective loop-closure detection, instead of three or four with the original model and successive loop-closure are always detected when taking a path that has already been taken (Fig. 8). Multiple loop-closure detection on the same node while the robot is moving and loop-closure detection from distant places which make the map not consistent with the environment are also discarded, thanks to the odometry consideration and the use of drastic loop-closure acceptance conditions. The new image acquisition policy enforces a more regular sampling of positions in the environment, independent of the robot velocity and also reduces the computational burden of the algorithm when the robot is not moving.

6. Conclusion and future work

We have introduced in this paper a system that is able to build a topo-metrical map in real time while a robot is discovering an unknown environment. The developed framework is an extension of our previous work on real time visual

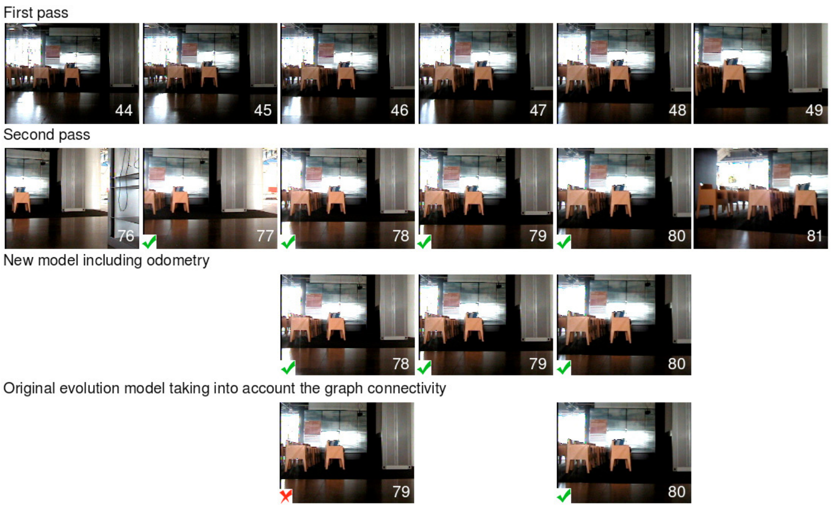

FIGURE 8. Comparison between the previous and the novel evolution model on the first loop-closure of the sequence (Museum). We can consider on this 1.5 m of the run four loop-closure locations. The loop-closure detected using the odometry evolution model corresponds to the ground truth, but those using the Gaussian transition model are visually correct but are inconsistent with the robot trajectory as they present a gap between detections (image 47 is not matched).

loop-closure detection [1] to which we added metrical information given by robot odometry. It builds topo-metrical map instead of the existing topological map and replaced the evolution model of the Bayesian filter with a new odometry-based model.

The extended algorithm, which only requires a monocular camera and odometry data, is more robust, more responsive and still does not require any *a priori* information on the environment. It is a simple solution, which works in real-time and which can be easily embedded on medium platforms. The resulting map is geometrically consistent and is usable for robot guidance.

Our future work will be to optimize visual processing to reduce computational cost and to implement this framework on mobile toy robots using remote processing methods. Using remote processing will notably requires to embed odometry processing and guidance on the platform while performing image processing and relaxation on remote servers in an asynchronous process.

References

[1] A. Angeli, D. Filliat, S. Doncieux and J.-A. Meyer, A fast and incremental method for loop-closure detection using bags of visual words, *IEEE Transactions On Robotics, Special Issue on Visual SLAM* **24** (2008) 1027–1037.

[2] T. Bailey and H. Durrant-Whyte, Simultaneous localisation and mapping (slam): Part ii. *IEEE Robot. Autom. Mag.* **13** (2006) 108–117.

[3] O. Booij, B. Terwijn, Z. Zivkovic and B. Kröse, Navigation using an appearance based topological map, in *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2007).

[4] M. Cummins and P. Newman, Fab-map: Probabilistic localization and mapping in the space of appearance. *Int. J. Robot. Res.* **27** (2008) 647–665.

[5] A.J. Davison, I.D. Reid, N.D. Molton and O. Stasse, Monoslam: Real-time single camera slam. *IEEE Trans. Pattern Anal. Mach. Intell.* **29** (2007) 1052–1067.

[6] A. Diosi, A. Remazeilles, S. Segvic and F. Chaumette, Outdoor visual path following experiments, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'07* (2007).

[7] T. Duckett, S. Marsland and J. Shapiro, Fast, on-line learning of globally consistent maps. *Autonomous Robots* **12** (2002) 287–300.

[8] T. Duckett, S. Marsland and J. Shapiro, Learning globally consistent maps by relaxation, in *Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA)* (2000), pp. 3841–3846.

[9] E. Eade and T. Drummond, Monocular slam as a graph of coalesced observations, in *Proc. of the Int. Conf. on Computer Vision* (2007).

[10] D. Filliat, A visual bag of words method for interactive qualitative localization and mapping, in *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2007).

[11] D. Filliat and J.A. Meyer, Global localization and topological map learning for robot navigation, in *Proc. of the 7th Int. Conf. on Simulation of Adaptive Behavior (SAB02), From Animals to Animats 7* (2002).

[12] D. Filliat and J.-A. Meyer, Map-based navigation in mobile robots – I. A review of localisation strategies. *J. Cogn. Systems Res.* **4** (2003) 243–282.

[13] F. Fraundorfer, C. Engels and D. Nistér, Topological mapping, localization and navigation using image collections, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2007).

[14] U. Frese, P. Larsson and T. Duckett, A multilevel relaxation algorithm for simultaneous localization and mapping. *IEEE Trans. Robot. Autom.* **21** (2005) 196–207.

[15] G. Grisetti, C. Stachniss, S. Grzonka and W. Burgard, A tree parameterization for efficiently computing maximum likelihood maps using gradient descent, in *Proc. of Robotics: Science and Systems, Atlanta, GA, USA* (2007).

[16] A. Howard, M.J. Mataric and G. Sukhatme, Relaxation on a mesh: a formalism for generalized localization, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (2001), pp. 1055–1060.

[17] K. Konolige, J. Bowman, J.D. Chen, P. Mihelich, M. Calonder, V. Lepetit and P. Fua, View-based maps, in *Proc. of Robotics: Science and Systems, Seattle, USA* (2009).

[18] K. Konolige and M. Agrawal, Frameslam: From bundle adjustment to real-time visual mapping. *IEEE Trans. Robot.* **24** (2008) 1066–1077.

[19] J. Kosecká, F. Li and X. Yang, Global localization and relative positioning based on scale-invariant keypoints. *Robotics and Autonomous Systems* **52** (2005) 209–228.

[20] D.G. Lowe, Distinctive image feature from scale-invariant keypoint. *Int. J. Comp. Vis.* **60** (2004) 91–110.

[21] E. Menegatti, M. Zoccarato, E. Pagello and H. Ishiguro, Image-based monte-carlo localisation with omnidirectional images. *Robot. Auton. Syst.* **48** (2004) 17–30.

[22] M. Milford and G. Wyeth, Hippocampal models for simultaneous localisation and mapping on an autonomous robot, in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA 2004)* (2003).

[23] D. Nistér, An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (2004) 756–777.

[24] D. Nistér, O. Naroditsky and J. Bergen, Visual odometry for ground vehicle applications. *J. Field Robot.* **23** (2006).

[25] E. Olson, J. Leonard and S. Teller, Fast iterative alignment of pose graphs with poor initial estimates, in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2006)* (2006), pp. 2262–2269.

[26] J.M. Porta and B.J.A. Kranse, Appearance-based concurrent map building and localization. *Robot. Auton. Syst.* **54** (2006) 159–164.

[27] P. Rybski, F. Zacharias, J. Lett, O. Masoud, M. Gini and N. Papanikolopoulos, Using visual features to build topological maps of indoor environments, in *Proc. of the IEEE Int. Conf. on Robotics and Automation* (2003).

[28] G. Sibley, C.r Mei, I. Reid and P. Newman, Adaptive relative bundle adjustment, in *Robotics Science and Systems* (*RSS*), Seattle, USA (2009).

[29] B. Steder, G. Grisetti, S. Grzonka, C. Stachniss, A. Rottmann and W. Burgard, Learning maps in 3d using attitude and noisy vision sensors, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent RObots and Systems* (2007).

[30] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents).* The MIT Press (2005).

[31] E.C. Tolman, Cognitive maps in rats and men. *Psychol. Rev.* **55** (1948) 189–208.

[32] J. Wang, H. Zha and R. Cipolla, Coarse-to-fine vision-based localization by indexing scale-invariant features. *IEEE Trans. Syst. Man Cybern.* **36** (2006) 413–422.