

# Combining Odometry and Visual Loop-Closure Detection for Consistent Topo-Metrical Mapping

S. Bazeille, D. Filliat

ENSTA ParisTech, Unité Electronique et Informatique - Cognitive Robotics

32 Boulevard Victor, 75015 Paris

{stephane.bazeille, david.filliat}@ensta.fr

**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 [1] 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 in this method to generate a consistent topo-metrical map. The resulting algorithm which only requires a monocular camera and odometry data and is simple, and robust without requiring any *a priori* information on the environment.

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

## I. 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 and to self-localize in it [2]. 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 correspondence with the real world. However, many ethological and neurological studies showed that animals made also use of maps for navigation.

Building such map and using it is based on two distinct sources of information. The first is the internal information about the movements: speed, acceleration, leg movement. The second provides external information about the environment. It may be derived from vision, odor, or touch. In animals, the integration of these information for map building appear to take place in a part of the brain called hippocampus [3]. The navigation problem for robots is very similar and make use of the same information (e.g. wheel rotation and laser-range finders or camera), which lead several author to propose navigation systems for robots inspired by neurobiological findings (e.g. [4]).

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.

## II. PREVIOUS AND 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][5], and has also been used in many Simultaneous Localization And Mapping (SLAM) solutions (e.g. [6][7][8]). SLAM [9] is the process of localizing a mobile robot while concurrently building a map of its environment. In this very active research area, we are more specifically interested in topological SLAM [10] that models the environment as a graph of discrete locations.

In this paper we focus on extending the work done by Angeli [1], who has developed a real-time vision-based topological SLAM framework. The proposed method is fast and fully incremental (i.e. the system can be used without any *a priori* information about the environment), and uses appearance information from a single camera to build a topological map of the places the robot is visiting. The environment model is learned on-line, in real-time as the robot discovers its surroundings. This method presents many advantages such as its simplicity, speed, and efficiency but its main limitation is the lack of metrical information that makes the map ill-posed for robot guidance (see Fig.1). Indeed, localization is only possible in previously mapped areas and no information is stored about the guidance of the robot between places. In this article, we will present an extension of this method with the use of metrical information given by the odometry data. This additional information helps creating a more robust and efficient algorithm and make it possible to build a consistent topo-metrical map that is suitable for navigation.

Odometry is often used on robots, whether they be legged or wheeled, to estimate their position relative to a starting location. For example, rotary encoders placed on robot wheels make it possible to compute the relative movement between the current position and the previous one at any given time. 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. In our

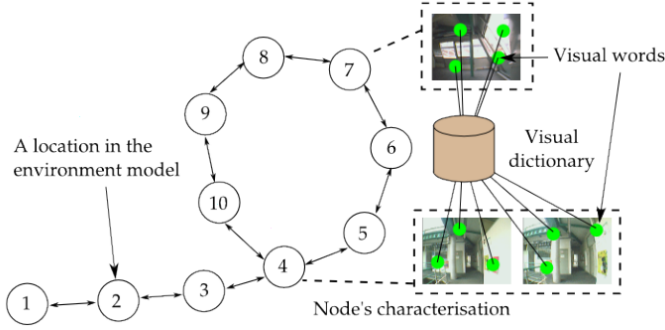


Fig. 1. Overview of a topological map of the environment used in our work.

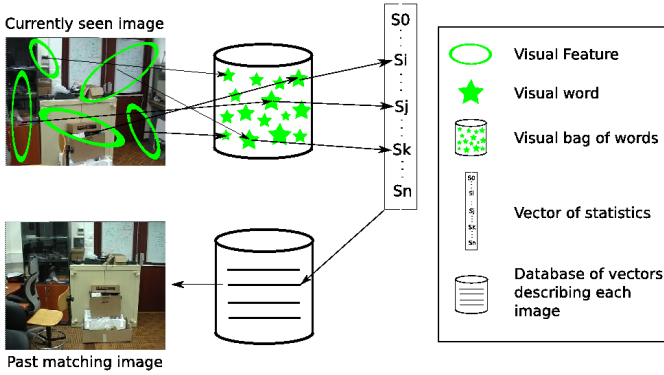


Fig. 2. Inverted index structure used to estimate the likelihood that the current image is similar with each image stored in the map.

case, this correction will 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 estimate a globally consistent topo-metric map and correct the odometry estimate. The fusion of vision and odometry sensors will also make the system more robust, notably in case of vision system failure (e.g. sensor occlusion, strong lighting change, dark areas) where odometry alone will be able to keep estimating the robot position.

#### A. Topological global localization using Bayesian filtering

Several vision-based techniques consider the problem of topological SLAM [7][11] or topological localization [12][13]. 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 this image-to-node matching scheme (based on a similarity measure between the current image and the images of a node previously visited), two main methods can be used: maximum likelihood which only consider the current image for matching (e.g. [13]) or maximum *a posteriori* scheme which exploits the similarity of image sequences to reduce false alarms (e.g. [14]). The Bayesian filtering framework we are using uses such a maximum *a posteriori* method in order to ensure the temporal consistency of the estimation. 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 maximises the probability of loop-closure with the current image:

$$N_i = \underset{i=0, \dots, n}{\operatorname{argmax}} p(S_t = i | I_t, M) \quad (1)$$

where  $S_t = i$  is event “ $I_t$  comes from  $N_i$ ” and  $M = N_0, \dots, N_n$  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:

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

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  $p(S_t | S_{t-1} = j, M)$  taking into account the robot movement since last localization. In Angeli’s work [15], as no information about the robot movement is available, a sum of gaussian is used in order to diffuse the probability of a node to its neighbours.

Then, the similarity of the current image with images stored in the map is taken into account through the likelihood model  $p(I_t | S_t, M)$ . This model is computed using a representation of images as a set of unordered elementary visual features taken from a dictionary (i.e. the bags of visual words model [16]). 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 (see Fig. 2). In our work, two kinds of visual features are used (SIFT [17] and local colour histograms) and likelihood is computed using the number of correspondences between images through a voting scheme. An interesting point of this approach is the use of an incremental dictionary to store the visuals words, which do not need any learning stage before operating in a new environment.

In addition, a “no loop-closure” event is added as a virtual location made of the mostly seen visual words. When this last event is the more probable, the map of the environment is extended so as to memorize the new place discovered by the robot.

#### B. Adding metrical information

The main objective of this work is the integration of metrical information to the existing topological map and loop-closure detection algorithm (see Fig. 1) so as to obtain a globally consistent map with which robot guidance is possible. The most appealing solution to this problem is probably the use of visual odometry, where images coming from neighbouring nodes or image sequences taken between nodes are matched to estimate the robot displacement [8][7][18][19][20]. 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 [21]. The robot can then be directly guided to the neighbouring 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.

Following several authors [22][23][24], we have chosen the simpler solution of using the information given by odometry data. Although this solution requires a second sensor for odometry, the information provided efficiently complements the image data and remains available in situations where visual information is unavailable. Moreover, the computing power required by this solution is negligible compared to visual loop-closure detection.

Beside the capacity to guide the robot, the inclusion of metrical information in the map also opens the possibility of using a much more informative evolution model in the bayesian filter. Through a probabilistic model of odometry, the evolution model can then take into account not only the nodes topological proximity, but also their relative position. 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.

### III. SYSTEM OVERVIEW

The overall processing of the original topological SLAM framework can be defined as follow:

- 1) acquire images at 1 Hz.
- 2) extract features from the image and reject image if it is too similar with the previous one.
- 3) compute likelihood for loop-closure with each place of the map, then predict and update the probabilities of each place and multiply by likelihood to obtain the *a posteriori* probability.
- 4) check and sort the probabilities higher than a threshold (i.e. potential loop-closure detection).
- 5) verify in the descending order the potential loop-closure detection by multiple-view geometry [25].

The inclusion of odometry information to complement visual sensor and to obtain a topo-metrical mapping required four main modifications (see Fig. 3):

- in step 1, images are now acquired when the robot has moved for a given distance or turned of a given angle and relative odometry since last processed image is recorded.
- in step 3, the gaussian transition model is replaced by an odometry based transition model.
- in step 5, the acceptance test is modified to constrain loop-closure to be detected only for very close locations.
- a new step 6 is added to memorize the relative position between nodes on each link of the graph and to apply a relaxation algorithm each time a loop-closure is detected to correct cumulative odometry drift.

#### A. Mapping and graph relaxation

The topological map is constituted of a set of nodes associated with an image and linked by edges. We integrated

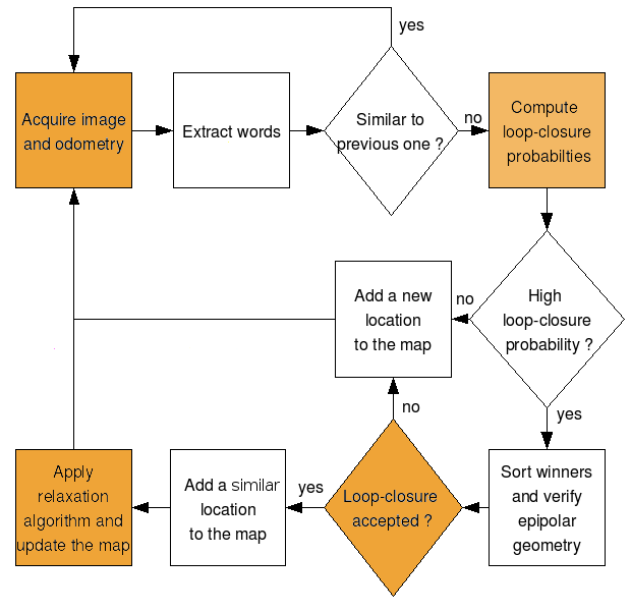


Fig. 3. Processing diagram of the topo-metrical map construction. The orange boxes indicate the steps modified or added after inclusion of odometry information.

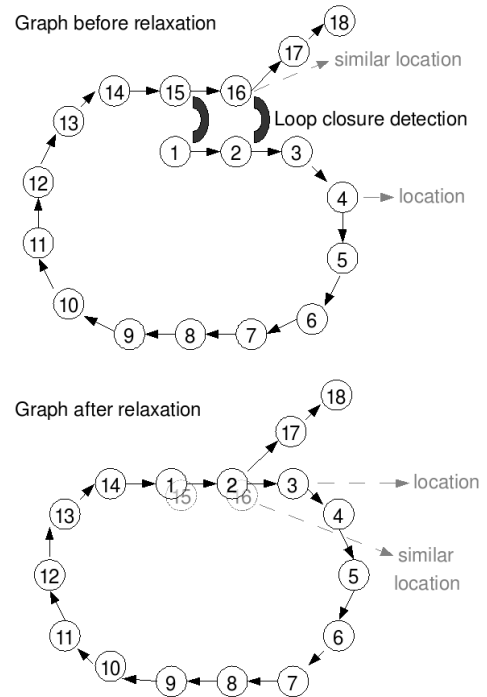


Fig. 4. Illustration of the graph relaxation process used to correct the map.

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  $(x, y, \theta)$ , where  $x$  and  $y$  are the 2D position coordinates and  $\theta$  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  $(d, \alpha, \phi)$ , where  $d$  and  $\alpha$  are the polar coordinates of the second node in the coordinate space of the first, and  $\phi$  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. In this case, the robot is assumed to have returned exactly at the position of a previous passing by constraining the two nodes to have the same position. This hypothesis is justified by the acceptance policy of loop-closure (step 5 of the algorithm) that has been modified to only accept loop-closure with very close views thereby allowing only small variations between the corresponding positions and orientations. This acceptance policy require 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.

Due to the cumulative noise of odometry, the map is not coherent after a loop-closure detection (i.e. the position of similar nodes are different) and it is necessary to correct the position of each node. To do so, we apply a relaxation algorithm to estimate the position of nodes that best satisfied the loop-closure constraints (see Fig. 4). The relaxation algorithm we chose is the Tree-based network Optimizer (TORO) [26], because of its speed and its high efficiency. TORO is an extension of Olson’s algorithm [27] which introduced a tree-based parametrization for the nodes in the graph. This technique solves the problem of learning maximum likelihood maps for mobile robots. It is based on a graph-formulation of the simultaneous localization and mapping problem and applies a gradient descent-based optimization scheme. The method is very fast and is called each time a new loop-closure is found to estimate the consistent node configuration which maximally satisfy the odometry constraints between nodes.

It must be noted that false positive loop-closure detections have to be avoided when we apply the relaxation because this would result in a map totally incoherent with the real environment. The Bayesian visual loop-closure detection we use is robust enough to make this assumption and is parametrized to avoid false positives, at the cost of missing some true positive.

### B. 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 neighbouring locations in the graph. The weight was defined as a sum of gaussian centered on the current location (see Fig. 5, Top). The limitation of this model is that diffusion is done in all directions without preference, because it only assume that the neighbouring images in time are close together, without any information about the real robot movement.

Because a reliable metrical information is now available, we integrate odometry in the evolution model to predict more precisely the evolution of the probability. Thus, starting from a given node, we distribute the probability to each neighbouring location in the map depending on the deviation of these nodes relative positions with the robot displacement since the last update  $d_u, \alpha_u, \phi_u$  measured by odometry (see Fig. 5, Bottom).

We used the standard motion model for robot odometry [28],

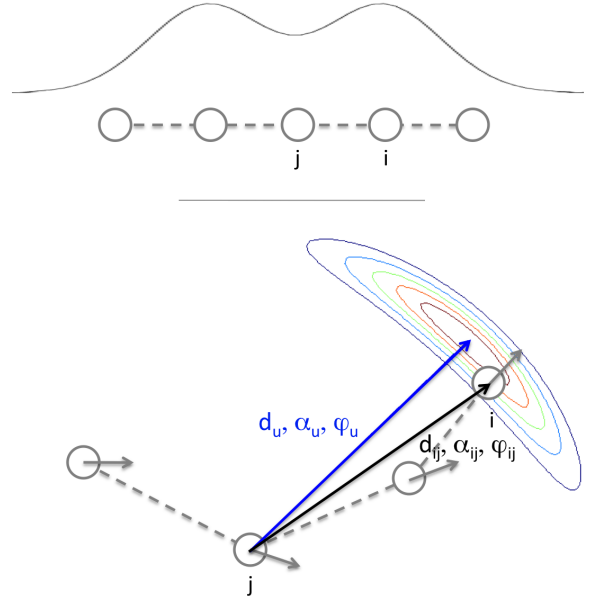


Fig. 5. Illustration of the modification of the evolution model. *Top* : The original model only take the graph connectivity into account when propagating probability from node  $j$  to node  $i$ . *Bottom* : Including odometry, the new evolution model is more precise and preferentially propagate probability from node  $j$  to the nodes  $i$  that correspond to a movement coherent with the odometry.

assuming gaussian noise on the robot displacement measured in polar coordinates:

$$p(d, \alpha, \phi | d_u, \alpha_u, \phi_u) =$$

$$G_{\mu_d, \sigma_d}(d - d_u) G_{\mu_\alpha, \sigma_\alpha}(\alpha - \alpha_u) G_{\mu_\phi, \sigma_\phi}(\phi - \phi_u)$$

where  $d, \alpha$  gives the odometry displacement in polar coordinates in the frame of the previous robot position and  $\phi$  is the variation of robot direction during movement.  $G_{\mu, \sigma}(X)$  is the gaussian distribution of mean  $\mu$  and variance  $\sigma^2$ .

Using this model, the evolution model becomes:

$$p(S_i | S_j, u_t, M) =$$

$$G_{\mu_d, \sigma_d}(d_{ij} - d_u) G_{\mu_\theta, \sigma_\theta}(\theta_{ij} - \theta_u) G_{\mu_\phi, \sigma_\phi}(\phi_{ij} - \phi_u)$$

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.

## IV. 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 7 shows image samples taken from this run. The path of the experiments as measured by odometry is shown on the top of Fig. 6. As a landmark we stop the run precisely on the path previously taken (and with

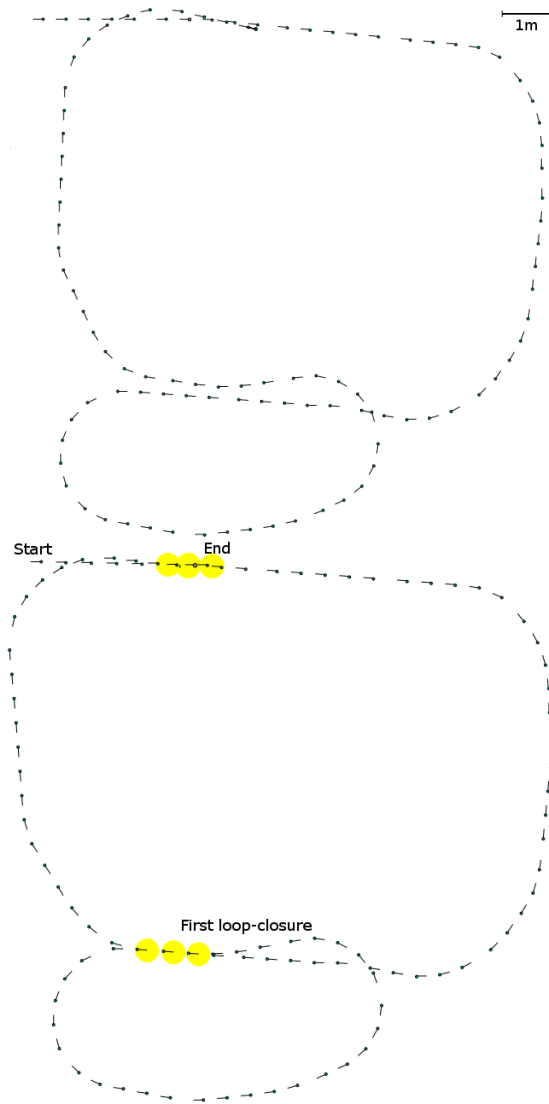


Fig. 6. Top: Topo-metrical map generated using only the relative odometry given by the robot, without loop-closure detection. Bottom : Topo-metrical map generated with the presented framework using relaxation after each loop-closure detection. Run length  $\approx 40\text{m}$ .

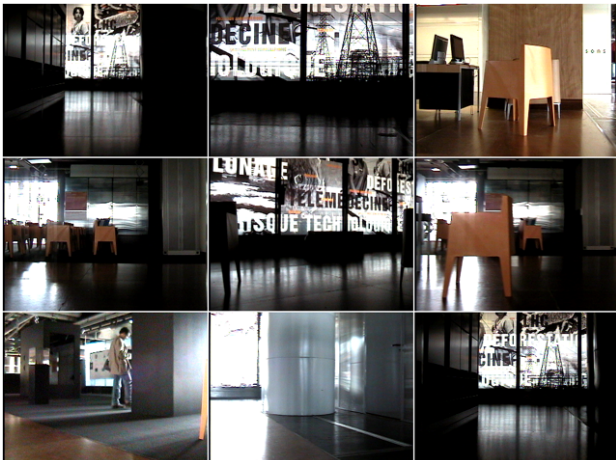


Fig. 7. Samples of the image sequence used in our experiment.

the same direction). The images and the odometry relative information were taken each time the robot move at least 50 cm or turn of at least 30 degrees. On this example, the odometry drift is quite low, but we can see that the loop-closures are correctly detected and that the result of the relaxation provides correct results (Fig. 6, Bottom, 8).

The use of odometry in the evolution model improves the responsiveness 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. /refexperiment). 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.

## V. 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 Angeli's work on real time visual loop-closure detection [1] to which we added metrical information given by robot odometry to build a topo-metrical map instead of the existing topological map and replaced the evolution model of the Bayesian filter with a new odometry-based model.

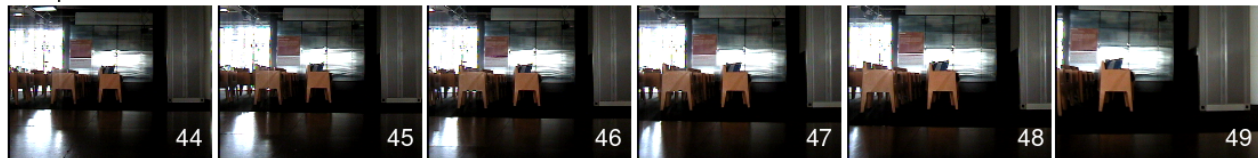
The algorithm, which only requires a monocular camera and odometry data, is more robust, more responsive and still does not require any *a priori* information on the environment. It is a simple solution, which works in real-time and which can be easily embedded on medium platforms. The resulting map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

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

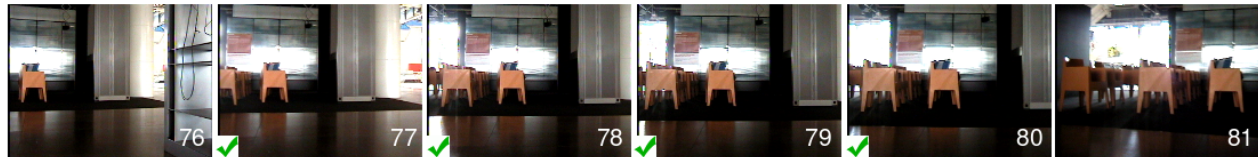
## REFERENCES

- [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*, vol. 24, no. 5, pp. 1027–1037, October 2008.
- [2] E. C. Tolman, "Cognitive maps in rats and men," *Psychological Review*, vol. 55, pp. 189–208, 1948.
- [3] D. Filliat and J.-A. Meyer, "Map-based navigation in mobile robots - I. a review of localisation strategies," *Journal of Cognitive Systems Research*, vol. 4, no. 4, pp. 243–282, 2003.
- [4] M. Milford and G. Wyeth, "Hippocampal models for simultaneous localisation and mapping on an autonomous robot," in *IEEE International Conference on Robotics & Automation (ICRA 2004)*, 2003.
- [5] M. Cummins and P. Newman, "Fab-map: Probabilistic localization and mapping in the space of appearance," *The International Journal of Robotics Research*, vol. 27, pp. 647–665, 2008.

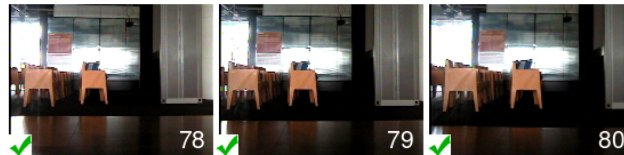
### First pass



### Second pass



### New model including odometry



### Original evolution model taking into account the graph connectivity



Fig. 8. Comparison between the previous and the novel evolution model on the first loop-closure of the sequence. We can consider on this 1.5m of the run four loop-closure location. The loop-closure detected by the novel algorithm are : 46-78, 47-79, 48-80 which correspond to the ground truth. The loop-closure detected by the previous model are 46-79 and 48-80 which are visually correct but are inconsistent with the robot trajectory as they present a gap between detections (image 47 is not matched).

- [6] A. Davison, I. Reid, N. Molton, and O. Stasse, "Monoslam: Real-time single camera slam," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, June 2007.
- [7] F. Fraundorfer, C. Engels, and D. Nistér, "Topological mapping, localization and navigation using image collections," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.
- [8] B. Steder, G. Grisetti, S. Grzonka, C. Stachniss, A. Rottmann, and W. Burgard, "Learning maps in 3d using attitude and noisy vision sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007.
- [9] T. Bailey and H. Durrant-Whyte, "Simultaneous localisation and mapping (slam): Part ii," *IEEE Robotics and Automation Magazine*, vol. 13, no. 3, pp. 108–117, 2006.
- [10] B. Kuipers and Y.-T. Byun, "A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations," *Journal of Robotics and Autonomous Systems*, vol. 8, pp. 47–63, 1991.
- [11] J. Porta and B. Krse, "Appearance-based concurrent map building and localization," *Robotics and Autonomous Systems*, vol. 54, pp. 159–164, 2006.
- [12] J. Wang, H. Zha, and R. Cipolla, "Coarse-to-fine vision-based localization by indexing scale-invariant features," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 36, no. 2, pp. 413–422, April 2006.
- [13] O. Booij, B. Terwijn, Z. Zivkovic, and B. Kröse, "Navigation using an appearance based topological map," in *IEEE International Conference on Robotics and Automation*, 2007.
- [14] E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro, "Image-based monte-carlo localisation with omnidirectional images," *Robotics and Autonomous Systems*, vol. 48, no. 1, pp. 17–30, 2004.
- [15] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer, "Real-time visual loop-closure detection," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2008, pp. 1842–1847.
- [16] D. Filliat, "A visual bag of words method for interactive qualitative localization and mapping," in *IEEE International Conference on Robotics and Automation*, 2007.
- [17] D. Lowe, "Distinctive image feature from scale-invariant keypoint," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [18] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. –, 2006.
- [19] K. Konolige, J. Bowman, J. D. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-based maps," in *Proceedings of Robotics: Science and Systems*, Seattle, USA, June 2009.
- [20] G. Sibley, C. Mei, I. Reid, and P. Newman, "Adaptive relative bundle adjustment," in *Robotics Science and Systems (RSS)*, Seattle, USA, June 2009.
- [21] A. Diosi, A. Remazeilles, S. Segvic, and F. Chaumette, "Outdoor visual path following experiments," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'07*, 2007.
- [22] T. Duckett, S. Marsland, and J. Shapiro, "Learning globally consistent maps by relaxation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2000, pp. 3841–3846.
- [23] D. Filliat and J. A. Meyer, "Global localization and topological map learning for robot navigation," in *From Animals to Animats 7. The Seventh International Conference on simulation of adaptive behavior (SAB02)*, 2002.
- [24] P. Rybski, F. Zacharias, J. Lett, O. Masoud, M. Gini, and N. Papanikolopoulos, "Using visual features to build topological maps of indoor environments," in *IEEE International Conference on Robotics and Automation*, 2003.
- [25] D. Nistér, "An efficient solution to the five-point relative pose problem," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 756–777, 2004.
- [26] G. Grisetti, C. Stachniss, S. Grzonka, and W. Burgard, "A tree parameterization for efficiently computing maximum likelihood maps using gradient descent," in *Proceedings of Robotics: Science and Systems*, Atlanta, GA, USA, June 2007.
- [27] E. Olson, J. Leonard, and S. Teller, "Fast iterative alignment of pose graphs with poor initial estimates," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, 2006, pp. 2262–2269.
- [28] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.