Appearance-based mapping and localisation using feature stability histograms

B. Bacca, J. Salvi, J. Batlle and X. Cufí

Proposed is an appearance-based mapping and localisation method based on the human memory model, which is used to build a feature stability histogram (FSH) at each node in the robot topological map. FSH registers local feature stability over time through a voting scheme, and most stable features are considered for mapping and Bayesian localisation. Experimental results are presented using omnidirectional images acquired through long-term acquisition considering: illumination changes, occlusions, random removal of features and perceptual aliasing. This method is able to adapt the internal node's representation through time to achieve global and local robot localisation.

Introduction: Nowadays, mobile robots interact within non-structured environments to deal with moving obstacles, perceptual aliasing, weather changes, occlusions, and to resolve mapping, localisation and navigation issues as best as possible. Mapping and localisation methods can be geometrical, topological or hybrid. Our work is focused on topological localisation and mapping using appearance-based models of the environment. The strength of the actual appearance-based models lies in their ability to represent the environment through high-level image features, using similarity measures to localise robots and to decide if new information can be added to the robot map. These methods are highly affected by changes in the environment appearance owing to weather conditions, occlusions by pedestrians, daytime and perceptual aliasing, which leads to spending a relatively high amount of computation time trying to localise a robot based in similarity measures. This inconvenience is increased when the robot is in operation for a long time. This problem has not been addressed directly in appearance-based approaches [1]. Many solutions have been proposed based on stable enough features descriptors such as SIFT and SURF [1], which use Manhattan or Euclidean similarity measures to build a map and Bayesian probabilistic framework for localisation in outdoor environments, but they do not report long-term mapping experiments or how to manage occlusions and lost features by environmental changes. Intuitively, appearance-based models describe the environment as it is, gathering its natural landmarks. This has been done in [2] using quad-tree decomposition, vertical line features [3], or invariant column segments [4], and textured planes extracted from catadioptric images [5]. But they report experiments conducted in static environments, without updating its appearance. In this Letter, we present an appearance-based mapping and localisation approach based on the feature stability histogram (FSH), which is inspired by the human memory model [6] to deal with dynamic environments. Unlike [7], we have built a histogram using a voting scheme instead of a hard-wired finite state machine. We report global and local localisation tests and obtain better results compared with static environment representations.

Appearance-based mapping: Our topological map is composed of several nodes; each one stores a set of SIFT [8] descriptors extracted from views. It stores its own FSH and the FSH evolution through time. An edge stores a list of corresponding features between two incident nodes and the estimated camera motion. We have defined a visual similarity between two sets of descriptors as the ratio of the number of corresponding features between the new image and the node features, and its geometrical average with the number of total features in the current node. The matching process is based on RANSAC robust epipolar geometry estimation.

Feature stability histograms: Appearance-based mapping and localisation in dynamic environments leads to changes in the robot environment model. A solution inspired by nature can separate stable features from unstable ones, and use only the stable features to focus mapping and localisation. In human beings, [6] proposes a memory model such that stimuli input enters the short-term memory (STM) which retains information long enough to use it. If the information in the STM is rehearsed or reinforced in some way it becomes part of the long-term memory (LTM) and can be used for a lifetime. Our approach modifies this model to be used in mobile robotics, as follows: first, at the beginning only one image is stored per node, and there is no difference between LTM and STM features; secondly, the LTM and STM features are distinguished thanks to a voting scheme, which is stored in the FSH, and the rehearsal method consists of associating a counter to each node feature in the FSH, incrementing it each time the robot re-observes a feature in this node, then having the robot use the FSH to measure how persistent a feature is by just observing its actual value. According to how the FSH is built, a good way to distinguish LTM features is to select a threshold such that FSH values greater than the threshold are considered to be LTM features. But a fixed threshold could drastically reduce the number of LTM features, or increase them considerably, reducing feature representativeness in both cases. We experimentally found that a threshold of 0.7 (0.3 is not appropriate for scalability) has shown a good commitment between successful localisation results and scalability in large environments, as shown in Fig. 1. This test was performed using 50 random image sequences for each change of 0.1 in the threshold. Successful localisation in Fig. 1 means that the estimated position was ± 1.5 nodes from the ground true (distance between nodes is 1 m).

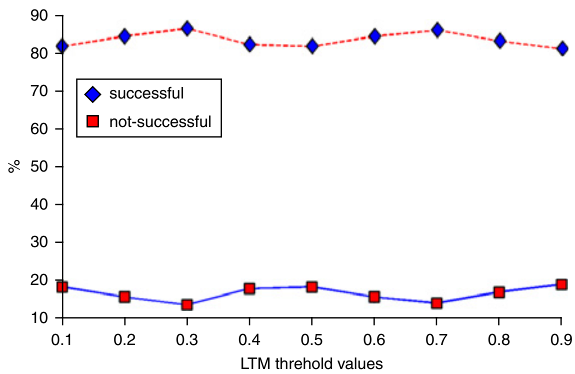

Fig. 1 *Successful and not-successful localisation against LTM threshold*

Localisation: We propose an appearance-based mapping and probabilistic localisation approach to deal with local and global robot localisation. SIFT features are commonly used as local features, and perceptual aliasing in the environment can confuse robots. However, the proposed mapping approach based on the human memory model and using a Bayesian filtering technique for robot localisation can reduce the location ambiguity and assign a probability value at each time instant. Our state is defined as $\boldsymbol{x} \in \{n_1, \ldots, n_N\}$, a node in the topological map. The Bayesian filter recursively calculates the posterior state distribution $p(x_t \mid z_{1:t})$:

$$p(\boldsymbol{x}_t|z_{1:t}) = \alpha p(z_t|\boldsymbol{x}_t)\Sigma_{\boldsymbol{x}_{t-1}\in\{n_1,\ldots,n_N\}} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|z_{1:t-1}) \quad (1)$$

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \gamma e^{-(\|x_t - x_{t-1}\|/\sigma_x^2)} \quad (2)$$

$$p(z_t|\boldsymbol{x}_t) = \delta_{sum}\Sigma_l w_l e^{-(sim(z_{t,1}, z(x_{t,l}))/\sigma_l^2} \quad (3)$$

where $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ is called the motion model, γ is normalisation constant, $\|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|$ is the distance between the two nodes in the map, and σ_x^2 is the variance of the distances on the map; the motion model enforces the temporary coherence of the position estimation and assumes that transitions between closer places are more likely than transitions between more distant ones. $p(z_t \mid x_t)$ is the sensor model, δ_{sum} is a normalisation factor, w_l is the mixture weight, $\text{sim}(z_t, z(x_t))$ is the similarity measure and σ_l^2, is the variance of this measure; the sensor model is related to the visual similarity between the current view z_v at x_t and the observations stored in the map. As a result of perceptual aliasing, the sensor model can have more than one maximum value. To overcome this inconvenience the sensor model can be defined as a sum of Gaussians, corresponding to the number of peak values between the maximum of the similarity measure and the variance of the similarity measure.

Results: We tested our approach in a mobile robot equipped with an omnidirectional vision setup composed of a parabolic mirror of diameter of 74 mm and a Sony colour camera of 640×480 pixels. We first guide the mobile robot through an indoor environment (49×28 m) to build the early topological map and to estimate the camera motion, ensuring a lot of pedestrians, different passages that are quite similar to each other, and changes in illumination due to weather conditions. Next, we conducted a static experiment where 180 images were acquired over five days to evaluate how our approach works by updating the map information in one node of the topological map, and detecting the most stable features to compute the image similarity. The left side of Fig. 2 shows both similarity measures: the dashed curve was obtained without updating the node map; whereas the continuous curve was made using our approach; the mean similarity value along the test is also shown. As can be seen, dynamic changes in the environment (right side of Fig. 2) cause low similarity measures when its representation is not updated accordingly as our approach does.

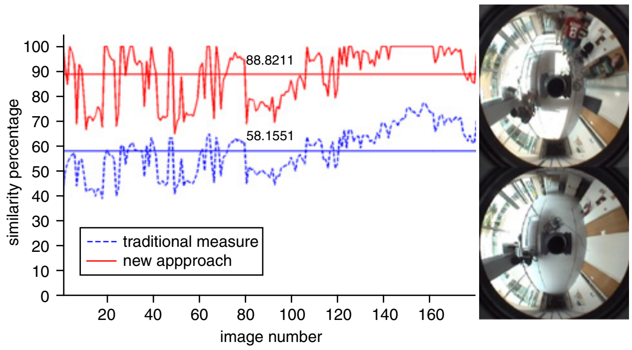

Fig. 2 *Similarity measures of static experiment and typical images*

The third step is topological localisation. For this experiment eight map updates were performed during the day, at night, and in summer, winter and spring. A database of 720 images was obtained. One additional image set was taken at completely different positions and orientations compared with the image map updates. For this set the real node in the map was stored for error analysis purposes. Global and local localisation results were extracted from four tests: the first one used the original set of images; in the second, Gaussian noise ($\mu = 0$, $\sigma = 0.15$) was added to the current image; and in the third and fourth tests additional artificial occlusion was added by randomly removing 25 and 50%, respectively, of the current image features. To evaluate the localisation performance 100 random image sequences were generated for each test. The global localisation performance was evaluated using the first image of the sequence, ensuring that no previous knowledge about the location was available. Table 1 summarises our results, where successful localisation means that the estimated position was ± 1.5 nodes from the ground true. It shows our approach (LTM-based) has better results when the map appearance is updated in comparison with traditional approaches where the map appearance is not updated. In addition, a successful localisation against image noise test was performed using our approach; the image noise was varied from 2.5 to 25%, resulting in 67.3% ± 4.2 and 32.6% $\pm$ 4.2 for successful and not successful robot localisation, respectively.

Table 1: Topological localisation results

Successful global localisation				
Test	Not LTM-based		LTM-based	
	Success	STD	Success	STD
Original set	44%	0.6	72%	0.6
Noise added	39%	0.6	64%	0.5
Noise + 25% occlusion	35%	0.5	63%	0.7
Noise + 50% occlusion	30%	0.6	56%	0.7
Successful local localisation				
Original set	39%	0.7	72%	0.9
Noise added	41%	1	68%	1
Noise + 25% occlusion	32%	1.1	62%	1
Noise + 50% occlusion	28%	1	55%	1

Conclusions: We have proposed an innovative feature management approach for topological mapping and localisation and appearance-based environment representation, which is based on a modified human memory model, and implements concepts such as LTM and STM to distinguish stable from non-stable features. Using the voting schema implemented through the FSH our method can deal with temporary occlusions and changes in illumination caused by dynamic environments.

Acknowledgments: The authors acknowledge the Commission of Science and Technology of Spain, project DPI-2007-66796-C03-02, the University of Valle contract 644-190495, the LASPAUCOLCIENCIAS grant 136-2008, and the research group's grant SGR2005-01008.

11 June 2010

doi: 10.1049/el.2010.1599

One or more of the Figures in this Letter are available in colour online.

B. Bacca and J. Batlle (*University of Girona, Girona 17071, Spain*)

E-mail: bladimir@eia.udg.edu

J. Salvi and X. Cufí (*VICOROB and the Department of Computer Engineering, University of Girona, Girona 17071, Spain*)

B. Bacca: Also with Universidad del Valle, Columbia

References

1 Šegvić, S., Remazeilles, A., Diosi, A., and Chaumette, F.: 'A mapping and localization framework for scalable appearance-based navigation', *Comput. Vis. Image Underst.*, 2009, **113**, (2), pp. 172–187

2 Sujan, V.A., Meggiolaro, M.A., and Belo, F.A.: 'Information based indoor environment robotic exploration and modeling using 2-D images and graphs', *Auton. Robots*, 2006, **21**, (1), pp. 15–28

3 Scaramuzza, D., Martinelli, A., and Siegwart, R.A.: 'Robust descriptor for tracking vertical lines in omnidirectional images and its use in mobile robotics', *Int. J. Robot. Res.*, 2009, **28**, (2), pp. 149–171

4 Goedemé, T., Nuttin, M., Tuytelaars, T., and Van Gool, L.: 'Omnidirectional vision based topological navigation', *Int. J. Comput. Vis.*, 2007, **74**, (3), pp. 219–236

5 Orghidan, R., Mouaddib, E., Salvi, J., and Serrano, J.J.: 'Catadioptric single shot rangefinder for textured map building in robot navigation', *IET Comput. Vis.*, 2007, **1**, (2), pp. 43–53

6 Atkinson, R., Shiffrin, R., and Spence, K.W.: 'Human memory: a proposed system and its control processes', *in* K.W. Spence and J.T. Spence (Eds): 'The psychology of learning and motivation' (Academic Press, 1968), Vol. 2, pp. 89–195

7 Dayoub, F., and Duckett, T.: 'An adaptive appearance-based map for long-term topological localisation of mobile robots'. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, Nice, France, 2008, pp. 3364–3369

8 Lowe, D.G.: 'Distinctive image features from scale-invariant keypoints', *Int. J. Comput. Vis.*, 2004, **60**, (2), pp. 91–110