

Mapping and Localization for Mobile Robots through Environment Appearance Update

Bladimir BACCA^{a,b,1}, Joaqu  m SALV  ^a, Xavier CUF  ^a
^a *Universitat de Girona, Girona, Catalunya, Spain*
^b *Universidad del Valle, Cali, Colombia*

Abstract. The strength of appearance-based mapping models lies in their ability to represent the environment through high-level image features; and provide human-readable information. However, developing localization and mapping methods with these models could be very challenging, especially if robots must deal with long-term mapping, localization, navigation, occlusions, and dynamic environments. This paper proposes an appearance-based mapping and localization 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, these FSH register local feature stability over time through a voting scheme, and most stable features are considered for mapping and Bayesian localization. Experimental results are presented using omnidirectional images acquired through long-term acquisition considering: illumination changes (day time and seasons), 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 localization.

Keywords. Appearance-based, omnidirectional vision, localization and mapping.

Introduction

Nowadays, mobile robots are needed to interact within non-structured environments, to deal with people, moving obstacles, perceptual aliasing, weather changes, occlusions, and robot-human interaction, and to resolve mapping, localization and navigation issues as best as possible. These requirements are useful for service robots designed to conduct surveillance, inspect, deliver, clean, and explore. In addition, to localization, mapping and navigation problems, have to guarantee a high level of autonomy through long-term navigation using stable features.

Mapping and localization methods can be geometrical, topological or hybrid. Our work is focused on topological localization and mapping using appearance-based models of the environment. Topological maps are compact, consume less computer memory, can be stored in efficient data structures, and speed up the navigation process. They use graphs for environmental modeling, and the appearance of the environment can be introduced through vision sensors [1]-[2]. Omnidirectional vision is receiving special attention nowadays due its long term landmark tracking, its one-shot environment sense regardless of heading, its robustness to occlusions, and it can be fused with range data [3]. Appearance-based methods for mapping and localization

¹ Corresponding Author: University of Girona, Campus Montilivi, Building P-IV, Girona, Spain, ZIP 17071. E-mail: bladimir@eia.udg.edu, evbacca@univalle.edu.co.

have also gained increasing attention in recent years. The strength of these models lies in their ability to represent the environment through high-level image features. They use similarity measures to localize robots and to decide if new information can be added to the robot map. In a review of some remarkable studies of appearance-based mapping and localization, we found different types of sensors, such as single camera sensors [4]-[7], laser range finder (LRF) sensors (2D and 3D) [8]-[9], fusion between an LRF and a single camera sensor [10], omnidirectional cameras [11]-[16] and an omnidirectional camera combined with LRF sensors [17]. We also discovered different techniques for feature extracting/matching, and the similarity measures used. Environmental features are diverse, but our review shows that SIFT and SURF are commonly used to describe the appearance of the environment [4], [6], [13], [15], [16-17]. The similarity measure used is commonly L1 or L2, since SIFT and SURF descriptors can ensure an adequate level of local discrimination. These kinds of local features are stable and fast enough even for omnidirectional images. Other local features used are discrete cosine transforms (DCT) [5], and PCA [3], but in these cases the appearance of the environment is a bit lost. Intuitively, appearance-based models of the environment describe the environment as it is, taking advantage of its natural features. This has been done in [7] using image information content implemented through quad-tree decomposition, or depth images taken by laser scans [8-10], PHLAC features [11], and vertical lines features [12], [14].

The main contribution of our work is an appearance-based mapping and localization approach based on the Feature Stability Histogram (FSH), which is inspired by human memory model [18] to deal with dynamic environments, where temporary or definitive changes occur with occlusions and illumination changes, and long-term mapping and localization is needed. Unlike [13], we build a histogram using a voting scheme instead of a hard-wired finite state machine. This histogram stores a stability value for each feature; stable/unstable features are distinguished using a threshold, and stable features are used for mapping and Bayesian localization. We perform global and local localization tests and obtain better results compared with static environment representations.

This article is organized as follows: section 1 presents a description of our method, section 2 the robot localization method, section 3 experiments and results, and section 4 our conclusions.

1. Appearance-Based Mapping and Feature Stability Histograms

Topological maps are a suitable environment representation since they can be used to move on to a semantic environment representation, they are compact and hold high level environmental information. Our topological map is composed of several nodes; each one stores one or more omnidirectional views. The map is defined as follows:

- A node is denoted by n_i , $i \in \{1, \dots, N\}$ where N is the number of nodes in the map.
- A node is composed of a set of SIFT [21] descriptors extracted from views, which are denoted by $Dn(i,j)$ where i is the node index, and $j \in \{1, \dots, K\}$, K is the number of feature descriptors stored within a node.
- A node stores its own FSH denoted $fsh(i,t)$, $i \in \{1, \dots, N\}$ at time t , where t denotes the number of times the FSH has been updated; also, the FSH evolution through time is stored and denoted $rfsh(i,p)$, $i \in \{1, \dots, N\}$, $p \in \{1, \dots, t\}$.

- Edges between nodes define neighboring positions and store a set of corresponding features extracted from a two-view geometry process and denoted Ed_r , $r \in \{1, \dots, R\}$, where R is the number of edges between nodes. This edge can be defined as: $Ed_r = match(Dn_{i-1}, Dn_i)$, where Dn_{i-1} and Dn_i are the previous and actual set of SIFT descriptors at each $(i-1)$ -th and i -th node, and $match()$ denotes the matching process described in subsequent sections.
- Estimated camera motion is also stored. It is constrained to a planar motion denoted as $m_i = [x_i, y_i, \theta_i]^T$, where x_i and y_i can be recovered up scale, but θ_i is estimated using a modified version of [22].

In this work we used a calibrated omnidirectional camera to extract full size SIFT feature descriptors from the region of interest of the image. A major challenge in appearance-based approaches, when a new image is acquired, is to decide when to add a new node to a topological map. We have defined a visual similarity between two sets of descriptors in Eq. (1).

$$S_{v,i} = \frac{match(v_{desc}, Dn_i)}{\sqrt{match(v_{desc}, Dn_i) * Dn_R}} \quad (1)$$

Where v_{desc} is the new image descriptors set, Dn_i is the i -th node descriptor set, $match(v_{desc}, Dn_i)$ is the number of corresponding features between the new image and the node features, and Dn_R is the number of total features in the current node. We defined a geometrical average in the denominator of Eq. (1) to prevent high values influence of Dn_R . The matching process occurred in two parts. First, given an $(i-1)$ -th image and an actual image, tentative matched features were extracted using a nearest neighbor method as described in [21]; second, an epipolar geometry estimation was performed with these matches using RANSAC, where p and q are mirror points of the first and second view; a look-up table was created to speed-up the process of lifting the image points on the mirror's sphere-equivalent model. The essential matrix E was estimated such that for all correspondences: $q^T E p = 0$. An essential matrix was calculated for every nine random correspondences. A correspondence is regarded as an inlier if in the second image the point v satisfies $v^T C v = 0$, where C is defined in Eq. (2) according to [23].

$$C = \begin{bmatrix} n_x^2(1 - \xi^2) - n_z^2\xi^2 & n_x n_y(1 - \xi^2) & n_x n_z \\ n_x n_y(1 - \xi^2) & n_y^2(1 - \xi^2) - n_z^2\xi^2 & n_y n_z \\ n_x n_z & n_y n_z & n_z^2 \end{bmatrix} \quad (2)$$

Where ξ is the mirror parameter, and n_x , n_y , and n_z form the normal vector to the plane between the feature and the central projection points of the two views, which is calculated by $[n_x, n_y, n_z] = Eu$. The E-matrix with the maximum of inliers was chosen.

The optical axis of the omnidirectional camera was vertically aligned to and mounted on a mobile robot constrained to horizontal plane movements. Once the correspondence problem was solved, the planar motion was estimated and denoted as $m_i = [x_i, y_i, \theta_i]^T$, where x_i and y_i can be recovered up scale, but θ_i was estimated using a modified version of [22] described here:

- Extract the corresponding image key-points denoted u_{kp} and v_{kp} .

- Compute the orientation error for all corresponding features as defined in Eq. (3).

$$\theta_{err} = \tan^{-1} \frac{v_{kp,y}}{v_{kp,x}} - \tan^{-1} \frac{u_{kp,y}}{u_{kp,x}} \quad (3)$$

Where, $v_{kp,y}$, $v_{kp,x}$, $u_{kp,y}$ and $u_{kp,x}$ are the key points vector components of the actual and previous image. The corresponding features have to satisfy another constraint, they have to be equally distributed along the previous omnidirectional image, and only the features that satisfy this constraint were selected.

- Using these features, a histogram H was computed using the orientation error vector of Eq. (3) with a resolution of 1° between -180° and 180° . The orientation was estimated using: $\hat{\theta} = \text{argmax}(H)$.

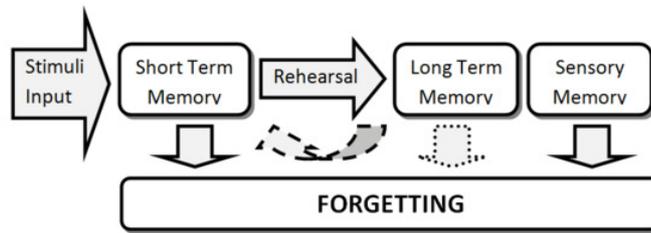


Figure 1. Atkinson and Shiffrin model of human memory.

Robot mapping and localization in real environments with temporary occlusions, pedestrians, illumination changes, among others leads to changes in the robot environment model. In appearance-based approaches these inconveniences are more pronounced. A solution inspired by nature can separate stable features from unstable ones, and use only the stable features to focus mapping and localization. According to [18] and [20] this process occurs naturally in animals with motor skills. In the specific case of human beings, [18] proposes a memory model that is valid for all human memories, feelings and experiences. This approach is shown in Figure 1. 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 a lifetime. Another component is the sensory model, which experimentally demonstrates the capability of sensory organs to discriminate information for subsequent processing. Note that the forgetting stage is always present. This model has been applied to robot mapping approaches [13] and mobile robot control architectures [19]. Our approach does not consider sensory memory. It modifies the human memory model, proposed for use in mobile robotics, as follows: first, in the very beginning only one image is stored per node, and there is no difference between LTM and STM features (all features are assumed to be LTM); second, 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; and third, the forgetting stage for the LTM is modified to allow not forgetting at all.

According the human memory model described above, a feature descriptor can be defined as an LTM if it has a high value in the FSH, otherwise it is considered an STM

feature. If an LTM feature is temporarily occluded, it will suffer a relative decrease in its FSH value. However, if the occlusion remains, the LTM feature will become an STM feature, but with a chance of becoming an LTM feature again if the occlusion disappears in the future. STM features are commonly produced by pedestrians, illumination changes, shadows and temporary occlusions, and for this reason they are not considered for map building and localization; indeed, only LTM features are used for these purposes. 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 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 tested that a threshold of 0.6 shown a good commitment between the inconveniences described above and scalability in large environments.

Through our approach we were able to automatically construct a topological map from a set of omnidirectional images taken at regularly spaced intervals. Each image is a node in the map. Later, the map was updated eight times with other omnidirectional images obtained under different environmental conditions. The map building algorithm works as follows: after image acquisition and feature extraction, a high similarity check is done to prevent robot stand-by images; then, the matching process begins and a similarity threshold is used to determine if a current image belongs to a node or not. If so, the FSH and its register are updated over time, if not, a new node is created and the matching features with the last node are kept for motion estimation.

2. Localization

A topological map gives sparse locations in the environment because the training image set does not cover all possible positions. Hence, given an image, the localization algorithm is able to find the node where the robot is likely to be, and this node is related to a real world position in the environment. Robot localization involves two main localization problems: global and local. The former is a concern when no a priori information on the location is available. In contrast, if the robot knows the initial pose, its mission is to track subsequent poses under the assumption that closer locations are more likely than distant ones. We propose an appearance-based mapping and probabilistic localization approach to deal with both problems. SIFT features are commonly used as local features, but perceptual aliasing in the environment can confuse robots. However, the proposed mapping approach based on the human memory model described above and using a Bayesian filtering technique for robot localization can reduce the location ambiguity and assign a probability value at each time instant.

Our state is defined as $x \in \{n_1, \dots, n_N\}$, a node in the topological map, and $z_v = v_{desc}$ is the observation at time t and composed by the current SIFT descriptors. Given a collection of LTM features $Z = \{Dn_1, \dots, Dn_N\}$, the goal is to find the node location x_t where the image was taken. The Bayesian filter recursively calculates the posterior state distribution $p(x_t | z_{1:t})$. Applying the Bayes' rule Eq. (4) can be defined.

$$p(x_t | z_{1:t}) = \frac{p(z_t | x_t, z_{t-1}, \dots, z_0) p(x_t | z_{t-1}, \dots, z_0)}{p(z_t | z_{t-1}, \dots, z_0)} \quad (4)$$

Where, the denominator can be replaced by a normalization factor. Bayesian filtering methods assume that the dynamics of the system is Markovian, which means

that future locations do not depend on past locations, Eq. (4) can be expressed as Eq. (5) allowing recursive position estimations since $p(x_{t-1} | z_{1:t-1})$ is the last estimation.

$$p(x_t | z_{1:t}) = \alpha p(z_t | x_t) \sum_{x_{t-1} \in \{n_1, \dots, n_N\}} p(x_t | x_{t-1}) p(x_{t-1} | z_{1:t-1}) \quad (5)$$

Eq. (5) has two unknown distributions: $p(x_t | x_{t-1})$ and $p(z_t | x_t)$. The first is called the motion model, expresses the probability transition between two locations in the map. To define it, we enforce the temporary coherence of the position estimation and assume transitions between closer places are more likely than transitions between more distant ones. We model this as a Gaussian distribution centered at x_t and expressed in Eq. (6).

$$p(x_t | x_{t-1}) = \gamma e^{-\frac{\|x_t - x_{t-1}\|}{\sigma_x^2}} \quad (6)$$

Where, γ is a normalization constant, $\|x_t - x_{t-1}\|$ is the distance between the two nodes in the map, and σ_x^2 is the variance of the distances on the map. The second unknown in Eq. (5) is $p(z_t | x_t)$, or the sensor model. In our case this distribution is related to the visual similarity between the current view z_v at x_t and the observations stored in the topological map $Z = \{Dn_1, \dots, Dn_N\}$, such that there is an expected and high probability of finding a maximum of $Z(x)$ at x_t rather than at other nodes of the map. As a result of perceptual aliasing, the sensor model can have more than one maximum value. To overcome this inconvenience and avoid discarding other possible hypotheses, 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 $sim(z_v, z(x_t))$ and this value minus σ_z (the variance of the similarity measure) as denoted in Eq. (7). Then, the sensor model assumed is shown in Eq. (8).

$$l = \max_Z \{sim(z_t, z(x_t)) - \sigma_z\} \quad (7)$$

$$p(z_t | x_t) = \delta_{sum} \sum_l w_l e^{-\frac{sim(z_t, z(x_t, l))}{\sigma_l^2}} \quad (8)$$

Where, δ_{sum} is a normalization factor, w_l is the mixture weight which equals the selected peak value of Eq. (7), $sim(z_t, z(x_t))$ is the similarity measure defined by Eq. (1), and σ_l^2 is the variance of this measure.

3. Experiments and Results

We implemented our approach in a mobile robot equipped with an omnidirectional vision setup composed of a *RemoteReality* parabolic mirror and a *Sony FCB-IX47AP* color camera. Additionally, the robot was controlled by an embedded computer at 900 MHz. To evaluate our approach, in the map building phase we first guide the mobile robot through an indoor environment, and this first image set was used to build the early topological map and to estimate the camera motion. Figure 2 shows the environment representation and the node locations obtained in the topological map

building phase. The map shows the first floor of the Computer Engineering Department at the University of Girona where there normally are a lot of pedestrians, different passages that are quite similar to each other, and big windows (nodes 12 to 28, 63 to 66 and 70 to 80) that allow big changes in illumination due to weather conditions.

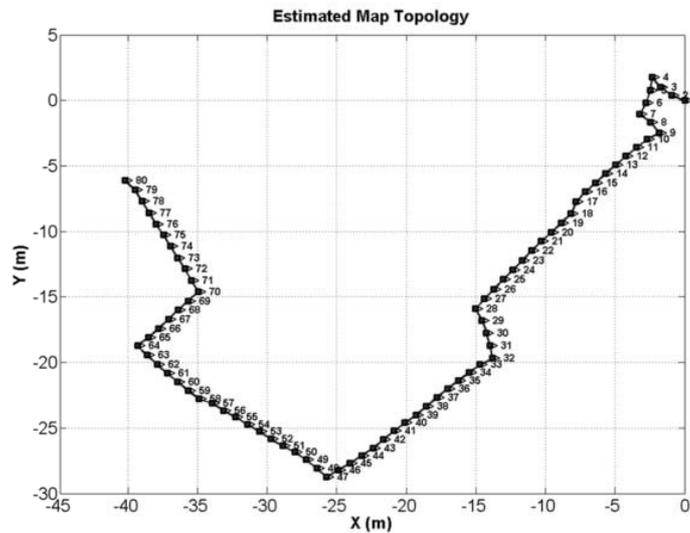


Figure 2. Camera motion estimation and topological map representation.

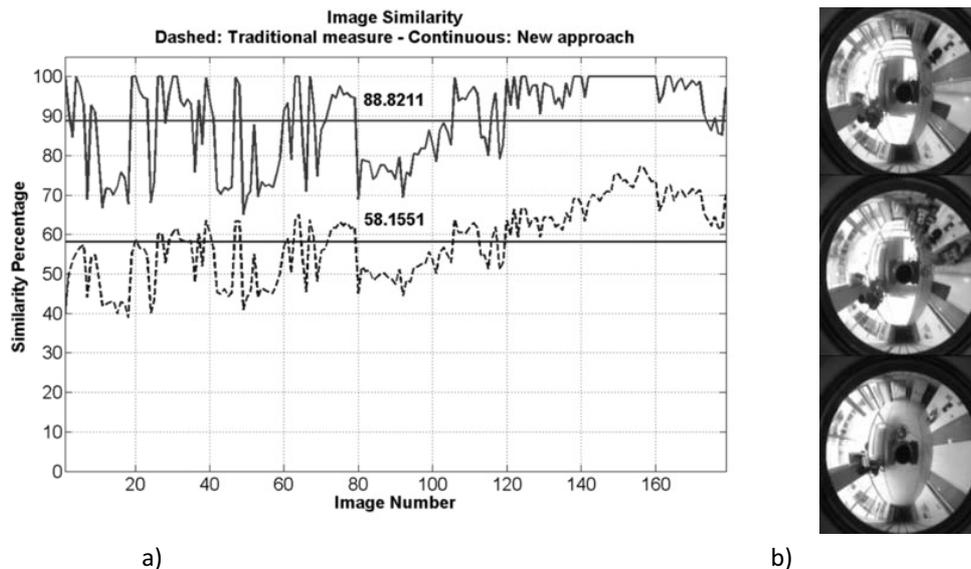


Figure 3. a) Similarity measure of node 20 to the main office of the Master's program and a big window. LTM-based similarity mean of 88.82% vs. 58.15%. b) Omnidirectional images example.

Our second step was 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. To do so, we conducted a static experiment in which the illumination changed due to weather conditions, curious pedestrians cause a lot of temporary occlusions and the furniture was manipulated. Figure 3a shows the similarity percentage vs. the image number at node 20, which is close to the Master's program main office and to a big window ensuring real world conditions. The 180 images were acquired over five days. Figure 3a shows both similarity measures: the dashed curve was obtained without updating the node map; whereas the continuous curve was made using our approach. As can be seen, occlusions due to pedestrians, changes in furniture,

and appearance change due to weather conditions cause low similarity measures when the representation of the environment at the corresponding node is not updated accordingly. But, in the case of an LTM-based similarity measure this effect is reduced because most LTM features remain present and a good representation of the environment is maintained.

The third step of our experiment is topological localization. For this experiment eight map updates were performed during the day, at night, and in summer, winter and spring. A database of 720 images of 640x480 pixels each was obtained. After that, two additional new image sets were taken at completely different base line distances (1m and 2.5m approx.), and different positions and orientations compared with the image map updates. With these two sets, global and local localizations were tested using the Bayesian filtering approach described above. For each image in these sets the real node in the map was stored, to extract the position error between this value and the estimated location node. Global and local localization results were extracted from four tests: the first one used the original set of images; in the second 15% of noise was added to the current image; and in the third and fourth ones additional artificial occlusion was added by randomly removing 25% and 50% respectively, of the current image features. To evaluate the localization performance in each set, 100 random initial positions were generated for each test, and then 11 consecutive images were presented to each map update. We evaluated the global localization performance of our approach using the first image of the 11 consecutive images; in this we ensured that not previous knowledge about the location was available. These experiments were performed for both maps: the map without an updated appearance (not LTM-based) and the one with an updated appearance.

Table 1. Correct global localization results.

Test	Long base line set				Short base line set			
	Not LTM-based		LTM-based		Not LTM-based		LTM-based	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Original set	74.28%	0.71	89.11%	0.59	43.66%	0.59	72.3%	0.57
Noise added	72.26%	0.62	78.22%	0.59	38.61%	0.6	64.4%	0.54
Noise + 25% occlusion	58.44%	0.6	78.2%	0.64	34.65%	0.55	63.4%	0.64
Noise + 50% occlusion	56.42%	0.67	62.4%	0.61	29.7%	0.57	56.44%	0.65

Table 2. Correct local localization results.

Test	Long base line set				Short base line set			
	Not LTM-based		LTM-based		Not LTM-based		LTM-based	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Original set	43%	0.64	70.8%	0.78	39.1%	0.7	72%	0.9
Noise added	43%	0.89	64%	0.827	41%	1.03	67.6%	0.97
Noise + 25% occlusion	40.2%	0.92	61.4%	0.84	31.68%	1.05	62%	1
Noise + 50% occlusion	32.34%	0.93	55.4%	0.88	28.3%	0.96	55.1%	1.05

Table 1 and 2 show the correct global and local localization results. Correct localization means that the estimated position was ± 1.5 nodes from the real one. All location estimates were selected in a *winner-takes-all* way. As can be seen, our approach shows better results if the map appearance is updated. In addition, during the result extraction phase we observed that position uncertainty decreased and the initial belief broadness got shrugged as the updates were incorporated into the topological map. Table 2 shows the correct local localization, where our approach has better results than a map without an appearance update. Table 2 shows small difference between the two data sets. In the local localization context this can be explained by the long base line between images, greater changes in environment appearance, and higher position uncertainty, which leads our sensor model to keep track of more hypotheses than the short base line set. Throughout the experiments, the sum of Gaussians assumed for the sensor model allowed our approach to recover from bad global location estimations, and then progressively obtain a good localization update as the image sequence continued. In this case, it was assumed that the robot continues on its path to collect more evidence about its locations, which is not far from reality since active map building and localization algorithms often use this technique [24]. This is advantageous in indoor environments like office corridors with high perceptual aliasing. Finally, our results might be improved by changing the environment description itself. SIFT features are often used in standard pin-hole cameras, and they cope well when used in omnidirectional images, but a better way to increase feature representation might be to use another feature close to the environment appearance [25].

4. Conclusions and Future Works

We have proposed an innovative feature management approach for topological mapping and localization and appearance-based environment representation. Our approach is based on a modified human memory model, and implements concepts such as LTM and STM to distinguish stable from non-stable features. These concepts were applied to topological mapping and localization using FSH, which store at each node a statistic about what features have been observed repeatedly. STM and LTM features are distinguished using a threshold, and LTM features only used for robot mapping and localization. Using the voting schema implemented through the FSH our method can deal with temporary occlusions caused by dynamic environments, and illumination changes. It was tested in static and dynamic experiments. The former included images acquired over a long period of time to show that our approach and the image similarity measure offers better results than a static description of the environment. The latter used a topological map that was updated up 8 times, and a Bayesian-based localization approach for global and local localization experiments. These experiments were conducted using two data sets with highly different positions, orientations, base lines and environmental conditions from those stored in the topological map. We used a threshold method to distinguish LTM from STM features, but an automatic classification of these features could be obtained using pattern recognition techniques such as *k-means*. Finally, SIFT features behave well in our experiments, but they do not have enough representative information to be used as global features. So another improvement will be to use features closer to the appearance of the environment.

Acknowledgements

This work has been partially funded by the Commission of Science and Technology of Spain (CICYT) through the coordinated project DPI-2007-66796-C03-02, the LASPAU-COLCIENCIAS grant 136-2008, the University of Valle contract 644-19-04-95, and the consolidated research group's grant SGR2005-01008.

References

- [1] Bailey, T., Durrant-Whyte, H.F., "Simultaneous Localisation and Mapping (SLAM): Part II - State of the Art". *Robotics and Automation Magazine*, pp. 10, 2006.
- [2] Durrant-Whyte, H.; Bailey, T. "Simultaneous Localization and Mapping (SLAM): Part I The Essential Algorithms". *Robotics and Automation Magazine* 13: 99–110, 2006.
- [3] Gaspar, J., Winters, N., Grossmann, E and Santos-Victor, J., "Toward Robot Perception through Omnidirectional Vision", *Studies in Computational Intelligence* 70, 223–270. Berlin: Springer, 2003.
- [4] Angeli, A.; Doncieux, S.; Meyer, J.-A.; Filliat, D., "Incremental vision-based topological SLAM," *Intell. Robots and Sys., IROS 2008. IEEE/RSJ Int. Conference on*, vol., no., pp.1031-1036, 22-26 Sept. 2008
- [5] Porta, J.M.; Krose, B.J.A., "Appearance-based concurrent map building and localization using a multi-hypotheses tracker," *Intell. Robots and Sys., Proc. IEEE/RSJ Int. Conf. on*, vol.4, pp. 3424-3429. 2004
- [6] Šegvić, S., Remazeilles, A., Diosi, A., and Chaumette, F. 2009. A mapping and localization framework for scalable appearance-based navigation. *Comput. Vis. Image Underst.* 113, 2 (Feb. 2009), 172-187.
- [7] Suján, V. A., Meggiolaro, M. A., and Belo, F. A. 2006. Information based indoor environment robotic exploration and modeling using 2-D images and graphs. *Auton. Robots* 21, 1 (Aug. 2006), 15-28.
- [8] Nieto, J., Bailey, T., and Nebot, E. Recursive scan-matching SLAM. *Robot. Auton. Syst.* 55, 1, pp. 39-49, 2007.
- [9] Nüchter, A. and Hertzberg, J. 2008. Towards semantic maps for mobile robots. *Robot. Auton. Syst.* 56, 11 (Nov. 2008), 915-926.
- [10] Newman, P.; Cole, D.; Ho, K., "Outdoor SLAM using visual appearance and laser ranging," *Robot. and Auto., ICRA 2006. Proc. IEEE International Conference on*, vol., no., pp.1180-1187, 15-19 May 2006
- [11] F. Linaker, M. Ishikawa, Real-time appearance-based Monte Carlo localization, *Robotics and Autonomous Systems*, Volume 54, Issue 3, 31 March 2006, Pages 205-220, ISSN 0921-8890.
- [12] Goedemé, T., Nuttin, M., Tuytelaars, T., and Van Gool, L. 2007. Omnidirectional Vision Based Topological Navigation. *Int. J. Comput. Vision* 74, 3 (Sep. 2007), 219-236.
- [13] Dayoub, F.; Duckett, T., "An adaptive appearance-based map for long-term topological localization of mobile robots," *Intell. Robots and Sys., IEEE/RSJ Int. Conf. on*, vol., no., pp.3364-3369, 22-26, 2008.
- [14] Murillo, A. C., Sagüés, C., Guerrero, J. J., Goedemé, T., Tuytelaars, T., and Van Gool, L. 2007. From omnidirectional images to hierarchical localization. *Robot. Auton. Syst.* 55, 5 (May. 2007), 372-382.
- [15] Zivkovic, Z., Booij, O., and Kröse, B. From images to rooms. *Robot. Auton. Syst.* 55, 5 2007, 411-418.
- [16] Andreasson, H.; Duckett, T.; Lilienthal, A.J., "A Minimalistic Approach to Appearance-Based Visual SLAM," *Robotics, IEEE Transactions on*, vol.24, no.5, pp.991-1001, Oct. 2008
- [17] Scaramuzza, D., Fraundorfer, F., Pollefeys, M. and Siegwart, R. Closing the Loop in Appearance-Guided Structure-from-Motion for Omnidirectional Cameras. *Proc. of the Eighth Workshop on Omnidirectional Vision (OMNIVIS'08)*, Marseille, France, October, 2008.
- [18] R. Atkinson and R. Shiffrin. "Human memory: A proposed system and its control processes". In K.W. Spence & J.T. Spence (Eds.), *The Psychology of Learning and Motivation*, 2: pp. 89–195, 1968.
- [19] R. Barber, "Desarrollo de una arquitectura para robots móviles autónomos : aplicación a un sistema de navegación topológica", Ph.D. dissertation, Dept. Elect. Eng., U. Carlos III, Madrid, Spain, 2000.
- [20] R. Llinás., "I of the Vortex: From Neurons to Self". Cambridge, MA: MIT Press, ch. 6-8, 2001.
- [21] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2) (2004) 91–110.
- [22] Scaramuzza, D., Fraundorfer, F., and Siegwart, R. Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. In *Proc. 2009 IEEE Int. Conf. on Rob. Aut.*, May 12-17, 2009.
- [23] Barreto, J. P. and Araujo, H. Geometric Properties of Central Catadioptric Line Images and Their Application in Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 8 2005, 1327-1333.
- [24] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, Cambridge, MA, 2005.
- [25] B. Bacca, J. Salvi, and X. Cufi, Appearance-Based SLAM for Mobile Robots. In *Proc. of Conf. on Art. intell. Res. and Dev., Frontiers in Art. Intell. and App.*, vol. 202. IOS Press, Amsterdam, 55-64, 2009.