Encyclopedia of Information Science and Technology, Third Edition

Mehdi Khosrow-Pour
Information Resources Management Association, USA

A volume in the

Information Science REFERENCE
An Imprint of IGI Global

Managing Director:	Lindsay Johnston
Production Editor:	Jennifer Yoder & Christina Henning
Development Editor:	Austin DeMarco & Jan Travers
Acquisitions Editor:	Kayla Wolfe
Typesetter:	Mike Brehm, John Crodian, Lisandro Gonzalez, Deanna Zombro
Cover Design:	Jason Mull

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Encyclopedia of information science and technology / Mehdi Khosrow-Pour, editor.
pages cm
Includes bibliographical references and index.
ISBN 978-1-4666-5888-2 (hardcover) -- ISBN 978-1-4666-5889-9 (ebook) -- ISBN 978-1-4666-5891-2 (print & perpetual access) 1. Information science--Encyclopedias. 2. Information technology--Encyclopedias. I. Khosrow-Pour, Mehdi, 1951-
Z1006.E566 2015
020.3--dc23
2014017131

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

A Comparison of Appearance-Based Descriptors in a Visual SLAM Approach

G

L. Fernández
Miguel Hernández University, Elche (Alicante), Spain

L. Payá
Miguel Hernández University, Elche (Alicante), Spain

F. Amorós
Miguel Hernández University, Elche (Alicante), Spain

O. Reinoso
Miguel Hernández University, Elche (Alicante), Spain

INTRODUCTION

The problem of Simultaneous Localization and Mapping (SLAM) has been studied thoroughly in the past decade in the field of mobile robotics. Taking into account the information that we can find on literature, it is possible to face the SLAM problem from three different points of view: the Metric SLAM, the Topological SLAM and the hybrid Metric-Topological SLAM. When we use the metric approach, we represent the environment and we compute the robot location through geometrical information with certain accuracy. On the other hand, when we face the problem using the topological approach, the objective is to represent the environment information by means of a list of locations within a graph, maintaining connectivity relationships between them. Finally, the metric-topological approach consists of a combination of the both techniques, trying to take advantage of both methods.

Nowadays, the use of computer vision is usual when we want to build a map and localize the robot in the map, because of several advantages (they are passive sensors, have a low cost and provide us with a great amount of information). When we use a vision sensor on a SLAM problem, we can approach the problem from two points of view: using the local appearance (landmarks) or using the global appearance to extract the necessary information from the scenes. The use of local appearance implies the extraction of distinctive landmarks from the images. When we use techniques based on local appearance, we typically need more computational time to build the map and locate the robot within the map. It is due to the fact that we need to extract the distinctive landmarks from each image and find each landmark extracted, in all images that compose the map. However, it presents an advantage: the possibility of including metric information to the system. Conversely, the global appearance methods need a lower time to work (they allow us to work in real time) but they do not directly include metric information in the map.

The main objective of this work is to build and test an algorithm to solve the SLAM problem using the global appearance of omnidirectional visual information and the robot internal odometry. Taking into account the advantages and disadvantages of the methods previously listed, we have decided to use a hybrid metric-topological approach to solve the SLAM problem.

BACKGROUND

The SLAM problem is a task studied extensively in the field of mobile robotics. One of the first works we find corresponds to Moravec and Elfes (1985), where a metric map is built by means of wide-angle sonar range measurements and a probabilistic approach.

DOI: 10.4018/978-1-4666-5888-2.ch313

Subsequently laser sensors are introduced to improve the accuracy and velocity in the algorithms created. For example, Thrun (2001) presents an algorithm for the SLAM problem in which a team of robots builds a map online using laser sensors and a Monte Carlo approach. However, the use of laser sensors implies an important contribution of radiation to the environment, and the laser sensors also use mechanical systems, which introduce errors.

Later, due to the numerous advantages offered (passive sensors, low cost, large amount of information, low power consumption, etc.), the use of cameras in the field of mobile robotics became widespread. For example, Murillo et al. (2007) make use of omnidirectional images to localize a robot within a previously created map, using SURF (Speed Up Robust Features) (Bay et al., 2006). Nevertheless, in most cases, when a robot needs to perform a task it does not possess any information about the environment and therefore the robot must build a map while it is located in it (SLAM). In this line, Gil et al. (2010) present an algorithm for the visual SLAM problem in which a robot builds a map online using a stereo-camera and SURF features. The main problem presented by the use of the extraction of distinctive landmarks corresponds to the high computational cost required. As a feasible alternative, the use of topological approaches is a field of great interest in the construction of maps by means of the global appearance of visual information, due to the numerous advantages it presents in terms of simplicity and computational cost. For example, Menegatti et al. (2004) carried out a study on robot navigation using the omnidirectional visual information captured from the environment, by applying a global appearance descriptor. In this work, they carried out a task of mapping and localization, without conducting a close loop, where the authors claim not high accuracy. In this sense, Werner et al. (2009) carried out a task of topological SLAM using vision-based techniques and global appearance. They also make use of omnidirectional images, and furthermore they propose a Bayesian approach that combines the odometry information of the robot with the visual information to improve the accuracy.

In this work we present a comparison of three global appearance descriptors in a process of robotic mapping, using a hybrid topological/metric SLAM.

APPEARANCE-BASED SLAM

In the approach we present in this work, we have decided to fuse the metric and the topological approach, with the goal of getting the advantages of each method. The topological method allows us to build a global compact representation of the environment and the metrical approach uses the information provided from the topological method to detect loop closures, so that it is possible to correct the possible errors in the position of the robot.

Constructing a Topological Map

Since we have decided to work with the global appearance of the images, we need to represent the global information that each image has, building a specific descriptor. The descriptor should retain the information in a compact and efficient manner, it must be computed quickly and it must be robust against changes in the environmental lighting conditions. In this work we make use of and compare three different global image descriptors: Fourier Signature (FS) (Payá et al., 2009), Gist-Gabor (Friedman, 1979) and Histogram of Oriented Gradient (HOG) (Amorós et al., 2010).

With the objective of representing the environment, we have used a graph, so that, each node of the graph represents an area of the environment with similar visual appearance (containing one or more images), and each edge indicates the connectivity relationships between the nodes (Romero et al., 2010). Taking into account that each node can contain one or more images, we have decided to compute the most representative image of each node. With this objective we use the following equations to compute the similarity between every two images in the node $S(i,j)$:

$$S(i,j) = {}^{1}\!/_{D(i,j)}$$
$$D(i,j) = \sqrt{\sum_m \sum_n \left(Des_i(m,n) - Des_j(m,n)\right)^2}, \tag{1}$$

where Des_i represents the descriptor of the panoramic image I_i, and m and n are the number of components the descriptor has in rows and columns, respectively. Then, the closer the images are, the smaller the distance D is and the greater the similarity S is. Once we have computed the similarity between every two images in the node, we use the following equations to compute the most representative image:

$$R = \arg\max_{i \in P}(\min_{j \in P, i \neq j}(S(i,j)))$$
$$N_R = \max_{i \in P}(\min_{j \in P, i \neq j}(S(i,j))), \tag{2}$$

where P represents the set of nodes in the graph, R is the image that best represents the node and N_R is its minimum similarity factor.

Then, to build the topological map we use the following algorithm:

1. Whenever the robot captures a new image I_K, it computes the descriptor Des_K and compares it with the descriptor of the image that represents the current node R_c, Des_C. If the similarity $S(K, R_c)$ is over certain threshold ($Th_{\min}$), it is added to the current node, and the node representative is re-calculated.
2. If the computed similarity $S(K, R_c)$ does not exceed the threshold $Th_{\min}$, we compare Des_K with the node representative of all the neighbor nodes to the current one. The node with the higher similarity is chosen, $S(K, R_{nei})$, so that if the new similarity exceeds the threshold, the current image is added to that adjacent node, and the node representative is re-calculated.
3. In the case that the similarity $S(K, R_{nei})$ does not exceed the threshold, the current image descriptor is compared with the representative descriptor of the rest of the nodes. If the similarity exceeds the threshold in any case, it is added to that node, and the node representative is re-calculated.
4. If no match is found, a new node is added to the system, so an edge between the new node and the previous one is added too (in the following, they are neighbor nodes). The new node includes the current image, which is also the node representative until a new image arrives, when this representative will be updated.

Constructing a Metric Map

We compute the metric map using a visual SLAM algorithm based in Payá et al. (2010). We combine a Monte-Carlo localization (MCL) algorithm with a landmark estimation process, so that, the robot decides when its current location is included as a metric map landmark. The main improvements over previous work mainly lie in the combination of a mapping process and a localization process simultaneously, combining a topological method with a metric method.

In a Monte-Carlo localization problem, the aim is the estimation of the robot's pose $x_{t+1} = (x, y, \theta)$ at time $t+1$ using a set of measurements $z_{1:t+1} = \{z_1, z_2, \ldots, z_{t+1}\}$ from the environment and the movements $u_{1:t+1} = \{u_1, u_2, \ldots, u_{t+1}\}$ of the robot (Fox et al., 1999). The probability density function $p(x_{t+1} \mid z_{1:t+1}, u_{1:t+1})$ is represented by a set of M random samples $x_{t+1} = x_{t+1}^i, i = 1 \ldots M$ extracted from it (particles), where each particle can be understood as a hypothesis of the true state of the robot $x_{t+1}^i = (x^i, y^i, \theta^i)$. The weight of each particle determines the importance of the particle and the set of particles defines a discrete probability function that approximates the continuous belief. The original MCL algorithm consists of three main phases: *Prediction phase*, in which a set of particles $\bar{\chi}_{t+1}$ is generated based on the set of particles χ_t and a control signal u_{t+1}; *Update phase*, in which the weight w_{t+1}^i of each particle in the set $\bar{\chi}_{t+1}$ is computed using the observation z_{t+1}; and *Resampling phase,* in which the resulting set χ_{t+1} is computed by resampling with replacement from the set $\bar{\chi}_{t+1}$, where the probability of resampling each particle is proportional to its importance weight w_{t+1}^i, in accordance with the literature on the SIR algorithm (Sampling Importance Resampling) (Smith et al., 1992). Finally, the set χ_{t+1} represents the distribution $p(x_{t+1} \mid z_{1:t+1}, u_{1:t+1})$.

In our algorithm, as we want to build a map of the environment while the position of the robot is computed (SLAM), we begin the experiment without a map of the environment. Taking into account this consideration, the initial set of samples is represented by a set of samples drawn from a narrow Gaussian centered at the initial point $(x = 0, y = 0)$. At the start of the experiment, the first map landmark (l_0) corresponds with the first pose of the robot x_0 and when the robot moves and captures a new image, the set of particles of the initial pose also moves adding some error to the movement of each particle, taking into account the movements $u_{1:t+1} = \{u_1, u_2, \ldots, u_{t+1}\}$ of the robot. Moreover, the topological algorithm computes the representative pose of the node every time a new image is added, and it will appear on the metric map as the set of particles that represents the metric localization of the node. This way, each representative pose of the topological map has one metric map landmark (l_i) associated. So, although we compute each set of particles that represents each movement of the robot, we only add to the metric map the set of particles that represents the node. On the other hand, our MCL algorithm differs with respect to the traditional MCL algorithm in that we do not carry out an update and resampling process in each movement of the robot but only when the topological algorithm detects a loop closure, even though we perform a prediction phase at each movement of the robot.

When the topological algorithm detects a loop closure, an update and resampling process is activated. To compute the weight of each particle w^i_{t+1} we use as input both, the metric information (odometry) and the visual information (global image descriptor) through the following equation:

$$w^i_{t+1} = \exp\{-v_i \Sigma_l^{-1} v^T{}_i\} \exp\{-h_j \Sigma_d^{-1} h^T{}_j\}, \tag{3}$$

where v_i represents the difference between the position of the landmark l_j and the position (x_i, y_i) of each particle i $(v_i = (l^j_x, l^j_y) - (x^j, y^j))$, Σ_l is a diagonal matrix $\Sigma_l = diag(\sigma_l^2, \sigma_l^2)$ where σ_l^2 has been chosen in order to minimize the error in the localization of the robot. On the other hand, $h = | d_j - d_t |$ is the difference between the appearance descriptor associated to the current observed image and the descriptor associated to the landmark l_j. Once the weight of each particle has been computed, the resulting set χ_{t+1} is computed by resampling with replacement from the set $\overline{\chi}_{t+1}$, where the probability of resampling each particle is proportional to its importance weight w^i_{t+1}.

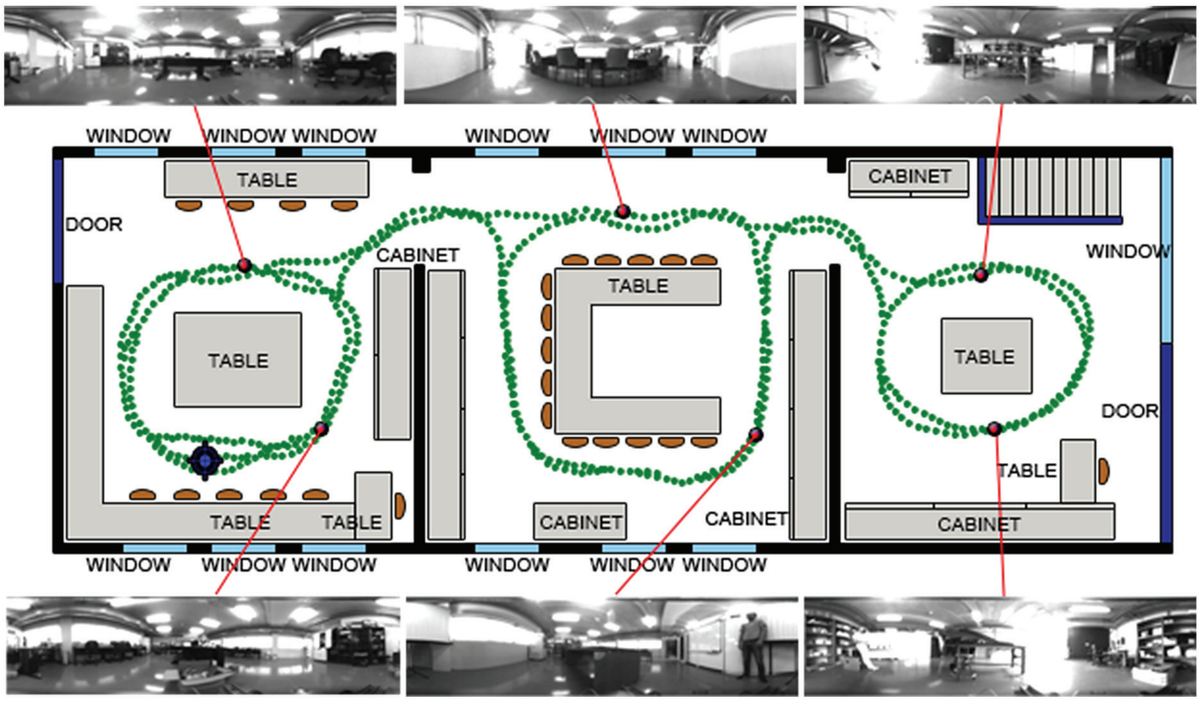
Results and Discussion

With the objective of evaluating our SLAM algorithm, we performed a realistic experiment with a mobile robot in an indoor environment. We captured a data set with all the data we need and we carried out several sets of experiments using a desktop PC (Figure 1). The robot was manned to travel a specified route into the environment, so that several loop closures occur. The ground truth is computed using the method presented in Stachniss et al. (2004) from the data collected with a laser.

Based on previous work (Payá et al., 2010) and to evaluate our method, we have decided to use the Procrustes analysis (Seber, 1984), where we get a measure of how accurate is the layout of the landmark after the SLAM process, comparing to the real layout. As a result of this process we obtain a parameter $\mu \in [0,1]$ (shape difference), where μ is a measure of the shape correspondence between the sets of points *A* and *B*, so that the lower is μ, the more similar are A and B.

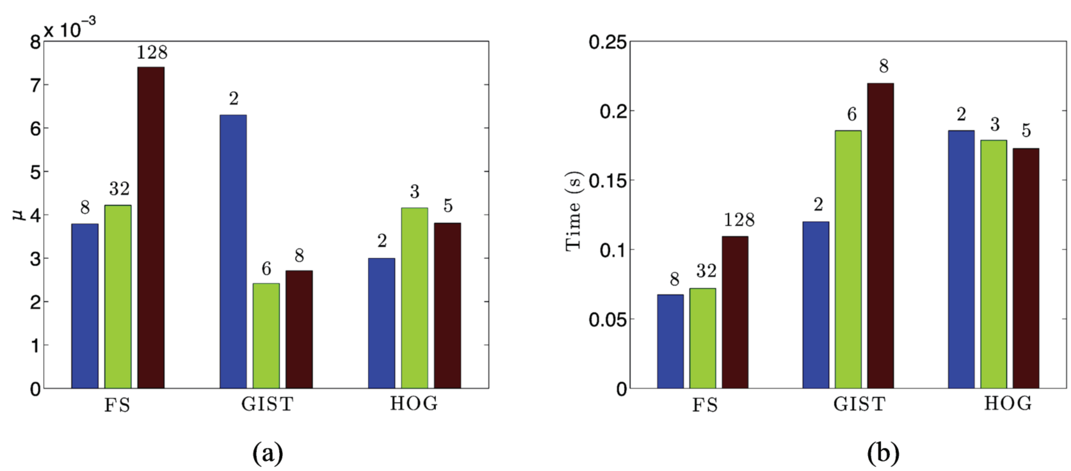
In our experiments, we have compared the performance of our hybrid algorithm when we use three different appearance-based descriptors (FS, GIST and HOG). We use the robot odometry as the input of the prediction phase in Monte-Carlo algorithm. In the simulation we have tested the influence of the number of components of each descriptor, as well as the influence of the number of particles used in the Monte-Carlo algorithm. We have used the μ factor to evaluate the accuracy of the resulting map (metric), and the step time needed by the algorithm to evaluate the feasibility of the algorithm to work in real time. In Figure 2 we can see the μ factor and the step time needed by the algorithm t, depending on the characteristics of the descriptor used. We have performed several series of experiments using the three descriptors (separately) and a number of particles equal to 200. When we build the descriptor of each image, we have to decide the number of components that we want to store. In the case of Fourier Signature (FS) we have to decide the number of components of each row of the signature that we store. In the case of GIST descriptor, we can change the size of the mask used to calcu-

Figure 1. Bird's eye view of the environment where the experiments have been carried out. We can see the path followed by the robot to get the necessary data to test the performance of our algorithm (green dots) and some examples of the images captured along this path.



late the descriptor. This will change the size of the descriptor. At last, in the case of HOG descriptor, it is possible to modify the size of the descriptor by means of changing the size of the window used to compute the descriptor. It can be seen that for FS and GIST, the accuracy depends of the components used to compute the descriptor greater extent than in the case of HOG. However, in general, we can say that we get a good

Figure 2. (a) Shape difference μ versus number of components using the FS (left), GIST (center), and HOG (right); (b) processing step time to build the map versus number of components using the FS (left), GIST (center), and HOG (right)



performance in accuracy for the three methods, even in the worst case. The μ factor is lower than 0.0075 in all the cases. We must take into account that the μ factor we get by using only the internal odometry without our algorithm is around 0.1421. With respect to the time t, FS clearly improves the other two descriptors. Anyway, with all three descriptors it is possible to work in real time since the times obtained are under 0.25 sec per iteration.

On the other hand, Figure 3 shows the μ factor and the step time needed by the algorithm t, versus the number of particles in the Monte Carlo algorithm for each descriptor. It can be seen that for the three descriptors as the number of particles increases, the shape factor decreases but it reaches a minimum value from which the decrease is hardly noticeable (μ equal to 0.0020 using FS and 2000 particles, μ equal to 0.0036 using GIST and 2000 particles, and μ equal to 0.0033 using HOG and 2000 particles). Although the μ factor is lower when the FS is used, the results obtained when using the three descriptors are very similar in terms of the μ factor. With respect to the time, we can see that when the number of particles increases, so does the time required for each step of the process for the three descriptors. Nevertheless, the time is lower in the case of FS, and better in HOG than GIST, although the time grows in a similar manner using the three descriptors. It is due to the fact that the time depends mainly of the Monte-Carlo algorithm. It is necessary to emphasize that to obtain the graphs shown in Figures 2 and 3, we have performed a number of simulations equal to 500 for each case in each descriptor and for each number of samples, respectively, where the starting point for each simulation has been changed.

Finally, Figure 4 shows an example of some intermediate steps during a simulation experiment using the FS, 32 components per row and a number of particles equal to 200. As expected, the dispersion in the particles of each landmark grows when the robot moves from the initial position until the robot detects a loop closure, when the dispersion of the samples decreases because at this moment the algorithm carries out a resampling of these samples. It can be observed that the map obtained by means of our algorithm is considerably more accurate than the map obtained through the internal odometry.

FUTURE RESEARCH DIRECTIONS

We are working now in the improvement of our hybrid algorithm to get a better accuracy including new measures in the sampling process when we detect a loop closure. In this sense, we try to add the information of other images that are included in the node where we detect the loop closure. Moreover, it can be interesting to incorporate a hypothesis tree during the topological

Figure 3. (a) Shape difference μ versus number of samples using the FS (blue line), GIST (red line), and HOG (black line); (b) processing step time to build the map versus number of samples using the FS (blue line), GIST (red line), and HOG (black line)

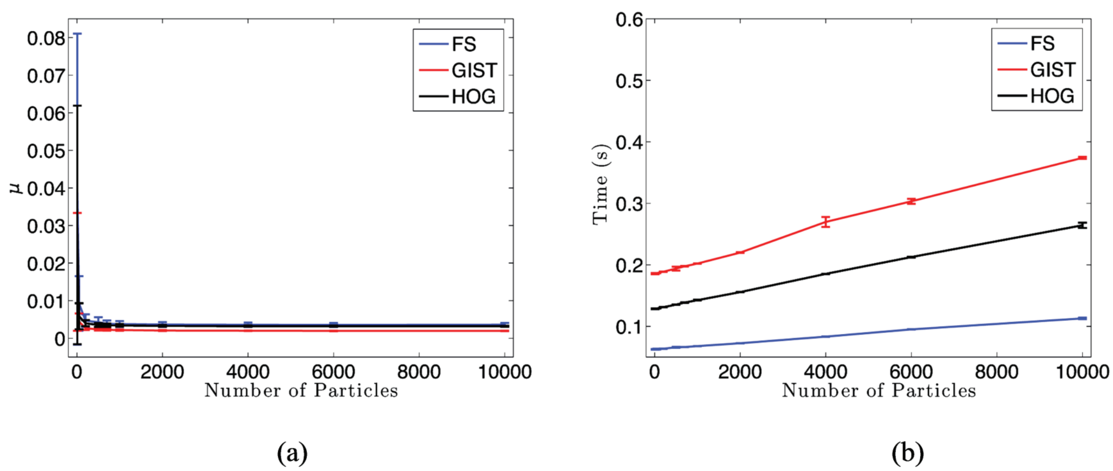
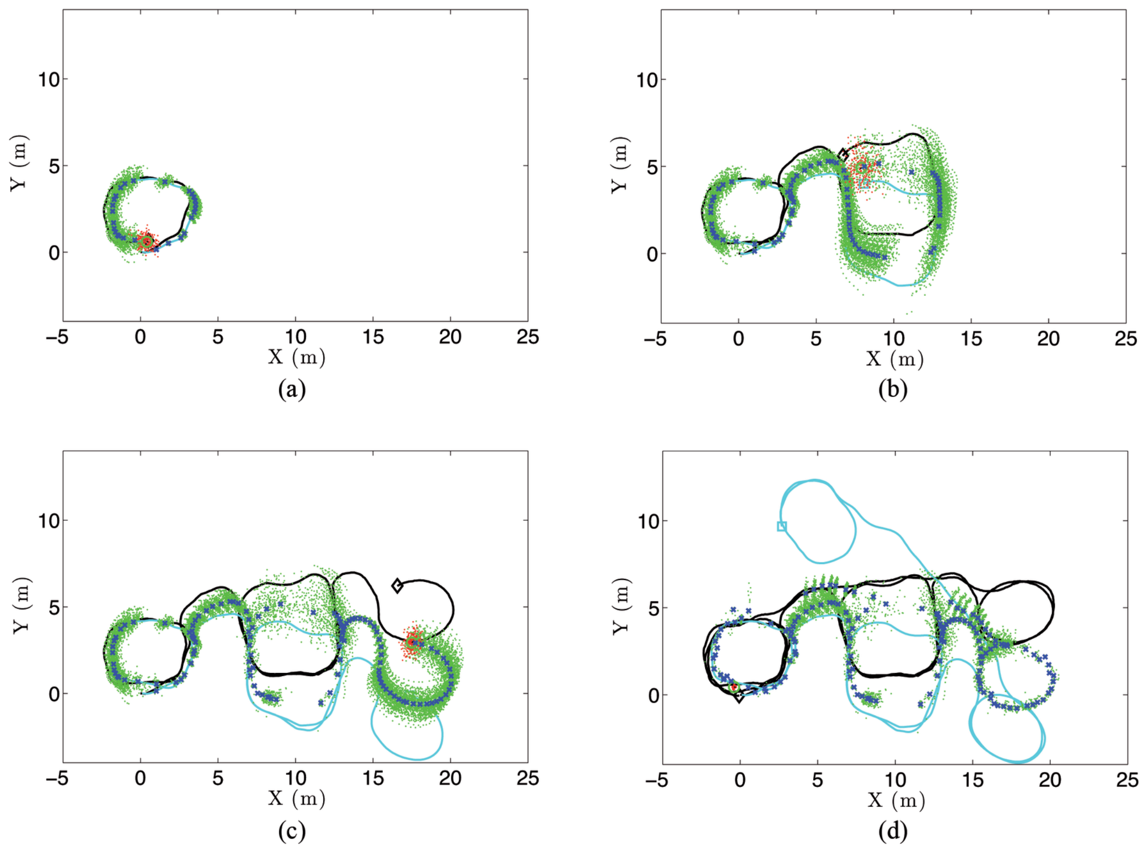


Figure 4. Example of some intermediate steps during a simulation experiment using the FS. The particles that represent the actual position of the robot are plotted as red points and their position is represented as a green circle. The particles that represent the position of the landmarks are plotted as green points and their position is represented as a blue cross. The internal odometry is plotted as a blue line where the blue squares represent the current position of the robot computed using the internal odometry. Finally, the ground truth is represented with a black line where the black square represent the position of the robot according the ground truth.



map building, so that it can be possible to maintain multiple hypotheses with the aim of reducing the number of incorrect associations when we close a loop. On the other hand, we will also try to test our algorithm in a large and outdoors dynamic environment.

CONCLUSION

In this article we have presented and evaluated a hybrid metric-topological SLAM algorithm to compute a map of the environment. We show how it is possible to get a good performance using three different appearance-based descriptors. The main contributions of our work include: 1) the development of a method to detect a loop closure by means of a topological graph, 2) the development of an algorithm to build a metric map using the robot internal odometry and the information provided by the topological graph, and 3) a evaluation of our method using three different appearance-based descriptors (FS, GIST, HOG).

We have carried out a realistic experiment in a typical laboratory environment under realistic lighting conditions. We demonstrate that we get a correct map using the three descriptors, even if the parameters chosen are not the most optimal. Notwithstanding, we show that a minimum of μ factor can be reached using a minimum number of particles, but it is necessary to reach a balance between accuracy and computational cost. We also show an example of a good performance simulation using the descriptor that gets the best results, FS. On the other hand we believe the approach we propose is useful when we work in large indoor environments, which are, in general, the most common environments where robots have to operate. Also, we demonstrate that our method is generalizable to other different kinds of descriptors.

REFERENCES

Amorós, F., Payá, L., Reinoso, O., Fernández, L., & Marín, J. M. (2010). Visual map building and localization with an appearance-based approach. In *Proceeding of the 7th International Conference on Informatics in Control, Automation and Robotics*, Madeira, Portugal (pp. 423-426).

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. In *Proceeding of the ninth European Conference on Computer Vision*, Graz, Austria (pp. 404-417).

Fox, D., Burgard, W., & Thrun, S. (1999). Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, *11*, 391–427.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for Gist. *Journal of Experimental Psychology. General*, *108*, 316–355. doi:10.1037/0096-3445.108.3.316 PMID:528908

Gil, A., Reinoso, O., Ballesta, M., & Juliá, M. (2010). Multirobot visual SLAM using a Rao-Blackwellized particle filter. *Robotics and Autonomous Systems*, *58*(1), 68–80. doi:10.1016/j.robot.2009.07.026

Menegatti, E., Zoccarato, M., Pagello, E., & Ishiguro, H. (2004). Image-based memory for robot navigation using properties of omnidirectional images. *Robotics and Autonomous Systems*, *47*(4), 251–267. doi:10.1016/j.robot.2004.03.014

Moravec, H. P., & Elfes, A. E. (1985). High resolution maps from wide angle sonar. In *Proceeding of the IEEE International Conference on Robotics & Automation*, St. Louis, MO, USA (pp. 116-121).

Murillo, A. C., Guerrero, J. J., & Sagüés, C. (2007). SURF features for efficient robot localization with omnidirectional images. In *Proceeding of the IEEE International Conference on Robotics & Automation*, San Diego, CA, USA (pp. 3901-3907).

Payá, L., Fernández, L., Gil, A., & Reinoso, O. (2010). Map building and Monte Carlo localization using global appearance of omnidirectional images. *Sensors (Basel, Switzerland)*, *10*(12), 11468–11497. doi:10.3390/s101211468 PMID:22163538

Payá, L., Fernández, L., Reinoso, O., Gil, A., & Úbeda, D. (2009). Appearance-based dense maps creation: Comparison of compression techniques with panoramic images. In *Proceeding of the 6th International Conference on Informatics in Control, Automation and Robotics*, Milan, Italy (pp. 250-255).

Romero, A., & Cazorla, M. (2010). Topological slam using omnidirectional images: Merging feature detectors and graph-matching. In J. Blanc-Talon, D. Bone, W. Philips, D. Popescu & P. Scheunders (Eds.), *Advanced Concepts for Intelligent Vision Systems: Proceedings of the 12th International Conference (ACIVS 2010)*. Sydney, Australia (LNCS 6474, pp. 464–475).

Seber, G. A. F. (1984). *Multivariate observations*. New York, NY: Wiley Interscience. doi:10.1002/9780470316641

Smith, A. F. M., & Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, *46*(2), 84–88.

Thrun, S. (2001). A probabilistic online mapping algorithm for teams of mobile robots. *The International Journal of Robotics Research*, *20*(5), 335–363. doi:10.1177/02783640122067435

Werner, F., Maire, F., & Sitte, J. (2009). Topological SLAM using fast vision techniques. In J.-H. Kim, S. S. Ge, P. Vadakkepat, N. Jesse, A. Al Manum, S. Puthusserypady K,…, D. Ahlgren (Eds.), *Advances in Robotics: Proceedings of the FIRA RoboWorld Congress 2009*. Incheon, Korea (LNCS 5744, pp. 187-196).

G

ADDITIONAL READING

Amorós, F., Payá, L., Reinoso, O., & Fernández, L. (2012). Map building and localization using global-appearance descriptors applied to panoramic images. *Journal of Computer and Information Technology*, *2*(1), 55–71.

Andreasson, H., & Duckett, T. (2004). Topological localization for mobile robots using omni-directional vision and local features. In *Proceeding of the 5th Symposium of Intelligent Autonomous Vehicles*. Lisbon, Portugal.

Angeli, A., Doncieux, S., Meyer, J. A., & Filliat, D. (2009). Visual topological SLAM and global localization. In *Proceedings of the International Conference on Robotics and Automation*, Kobe (pp. 4300-4305).

Cummins, M., & Newman, P. (2008). Fab-map: Probabilistic localization and mapping in the space of appearance. *Journal of Robotics Research*, *27*(6), 647–665. doi:10.1177/0278364908090961

Davison, A. J. (2003). Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the International Conference on Computer Vision*, Sendai, Japan (pp. 1403-1410).

Dellaert, F., Burgard, W., Fox, D., & Thrun, S. (1999). Using the CONDENSATION algorithm for robust, vision-based mobile robot localization. In *Proceeding of the IEEE Computer Society Conference on Computer VIsion and Pattern Recognition*, Fort Collins, CO, USA.

Elfes, A. (2002). Using occupancy grids for mobile robot perception and navigation. *Journal of Computer*, *22*(6), 46–57. doi:10.1109/2.30720

Fernández, L., Gil, A., Payá, L., & Reinoso, O. (2010). An evaluation of weighting methods for appearance-based Monte Carlo localization using omnidirectional images. In *Proceedings of the IEEE International Conference on Robotics and Automation (Workshop on Omnidirectional Robot Vision)*, Anchorage, Alaska, USA (pp. 13-18).

Fernández, L., Payá, L., Valiente, D., Gil, A., & Reinoso, O. (2012). Monte Carlo localization using the global appearance of omnidirectional images. Algorithm optimization to large indoor environments. In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics*, Rome, Italy (pp. 439-442).

Fox, D., Burgard, W., Dellaert, F., & Thrun, S. (2000). Monte Carlo localization: Efficient position estimation for mobile robots. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, Orlando, Florida, USA (pp. 343-349).

Friedrich, H., Dederscheck, D., Krajsek, K., & Mester, R. (2007). View-based robot localization using spherical harmonics: Concept and first experimental Results. In F. A. Hamprecht, C. Schnörr & B. Jähne (Eds.), *Pattern Recognition: Proceedings of the 29th DAGM Symposium*. Heidelberg, Germany (LNCS 4713, pp. 21-31).

Gil, A., Ballesta, M., Juliá, M., Reinoso, O., & Ubeda, D. (2012). Cooperative simultaneous localisation and mapping using independent Rao-Blackwellised filters. *Journal of Computer Vision, IET*, *6*(5), 407–414. doi:10.1049/iet-cvi.2011.0108

Gonzalez, R. C., & Woods, R. E. (1992). *Digital image processing* (3rd ed.). Addison-Wesley.

Kröse, B., Bunschoten, R., Hagen, S. T., Terwijn, B., & Vlassis, N. (2004). Household robots look and learn: Environment modelling and localization from an omnidirectional vision system. *Journal of IEEE Robotics and Automation Magazine*, *11*(4), 45–52. doi:10.1109/MRA.2004.1371608

Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (2002). FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceeding of the Eighteenth International conference on Artificial Intelligence*, Menlo Park, CA, USA (pp. 593-598).

Nilsson, N. J. (1982). *Principles of artificial intelligence*. Berlin, New York: Springer Publisher. doi:10.1007/978-3-662-09438-9

Oliva, A. (2005). Gist of the Scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *The encyclopedia of neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier. doi:10.1016/B978-012375731-9/50045-8

Payá, L., Reinoso, O., Amorós, F., Fernández, L., & Gil, A. (2011). Probabilistic map building, localization and navigation of a team of mobile robots. Application to route following. *Journal of Multi-Robot Systems: Trends and Development*, *1*, 191–210.

Payá, L., Reinoso, O., Gil, A., & Sogorb, J. (2008). Multi-robot route following using omnidirectional vision and appearance-based representation of the environment. *Journal of Lecture Notes in Computer Science*, *5271*, 680–687. doi:10.1007/978-3-540-87656-4_84

Payá, L., Reinoso, O., Vicente, M. A., Gil, A., & Pedrero, J. M. (2007). Subspace reduction for appearance-based navigation of a mobile robot. In *Proceeding of the International Conference on Image Analysis and Processing*, Modena, Italy (pp. 123-128).

Shi, C., Wang, Y., & Yang, J. (2010). Online topological map building and qualitative localization in large-scale environment. *Journal of Robotics and Autonomous Systems*, *58*, 488–496. doi:10.1016/j.robot.2010.01.009

Thrun, S., Fox, D., Burgard, W., & Dellaert, F. (2000). Robust Monte Carlo localization for mobile robots. *Journal of Advanced Robotics*, *19*(10), 1059–1080.

Valgren, C., Lilienthal, A. J., & Duckett, T. (2006). Incremental topological mapping using omnidirectional vision. In *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China (pp. 3441-3447).

Valiente, D., Gil, A., Amorós, F., & Reinoso, O. (2013). SLAM of view-based maps using SGD. In *Proceeding of the 10th International Conference on Informatics in Control, Automation and Robotics*, Reykjavik, Iceland (pp. 329-336).

Wolf, J., Burgard, W., & Burkhardt, H. (2005). Robust vision-based localization by combining an image retrieval system with Monte Carlo localization. *Journal of IEEE Transactions on Robotics*, *21*(2), 217–229.

KEY TERMS AND DEFINITIONS

Appearance Descriptor: It is a descriptor of an image that represents the global information of the same without extracting landmarks.

Gist: It is the meaning of a scene, or in other words, the spatial envelope of the scene.

Localization: It is the estimation of the position of an autonomous agent in a given map.

Mapping: It is the creation of an internal representation of any given environment.

Metrical Map: It is a representation of the environment through geometrical information with certain accuracy.

Mobile Robot: It is an autonomous vehicle that is capable of movement in any given environment.

Omnidirectional Vision: It is a vision system that is capable of capturing all the information surrounding the system with a single image (360º).

Probabilistic Localization: It is a localization task, where the information of all previous robot locations is used to estimate its current location.

SLAM: It is the process of building a map of an environment while simultaneously the localization of the agent that compute the map is estimated.

Topological Map: It is a representation of the environment by means of a list of locations within a graph with connectivity relationships between them.