Omnidirectional vision for robot localization in urban environments

Emanuele Frontoni, Andrea Ascani, Adriano Mancini, Primo Zingaretti

Università Politecnica delle Marche
Dipartimento di Ingegneria Informatica Gestionale e dell'Automazione
Ancona – Italy
{frontoni, ascani, mancini, zinga} @diiga.univpm.it

Abstract. This paper addresses the problem of long term mobile robot localization in large urban environments. Typically, GPS is the preferred sensor for outdoor operation. However, using GPS-only localization methods leads to significant performance degradation in urban areas where tall nearby structures obstruct the view of the satellites. In our work, we use omnidirectional vision-based techniques to supplement GPS and odometry and provide accurate localization. We also present some novel Monte Carlo Localization optimizations and we introduce the concept of on line knowledge acquisition and integration presenting a framework able to perform long term robot localization in real environments. The vision system identifies prominent features in the scene and matches them with a database of geo-referenced features already known or integrated during the localization process. Results of robot localization in the old town of Fermo are presented. Results show good performance and the whole architecture behaves well also in long term experiments, showing a suitable and good system for real life robot applications.

Keywords: omnidirectional vision, robot localization, outdoor robotics, knowledge acquisition and integration.

1 Introduction

Outdoor localization of mobile robots is a difficult task for many reasons. Some range sensors like laser range finders, which play an important role in indoor localization, are not suitable for outdoor localization because of the cluttered and unstructured environment. Global Positioning System (GPS) can give valuable position information, but often the GPS satellites are occluded by buildings or trees.

Because of these problems and also for low cost robotics applications, vision has become the most widely used sensor in outdoor localization and navigation. A serious problem for vision are illumination changes, because illumination in outdoor environments is highly dependent on the weather (sunny, cloudy, ...) and on the time of day. Another problem is the perceptual aliasing: visual features may not be distinguishable enough; in a forest, every tree looks about the same.

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353

An algorithm which can partially deal with changing illumination is the Scale Invariant Feature Transform (SIFT) developed by Lowe [1]. SIFT is a feature-based method which computes descriptors for local interest points. Another similar approach, with better computational performances, is SURF (Speeding Up Robust Features) [2-3]. Today tasks as visual mapping, localization and navigation, especially in outdoor environments, take a great advantage by using the omnidirectional vision and SIFT/SURF [5-6]; common approaches range from metrical to topological techniques [7-8]; however, the choice of approach depends closely by the desired accuracy and precision in the localization process.
Here we present a framework based on omnidirectional vision, feature matching (using SIFT and SURF) and some novel Monte Carlo Localization optimizations with the scope, among others, of reducing the number of required images for faster setup operations and low memory requirements.

The vision system identifies prominent features in the scene and matches them with a database of geo-referenced features already known or integrated during the localization process. So, we also introduce the concept of on line knowledge acquisition and integration, presenting a framework able to perform robot localization in real environments with long-term changes [4]. The *enduring* and *long-term* aspects are strongly coupled; *enduring* stands for a reliable methodology for outdoor localization in urban environments, while *long-term* is the capability of system to work also in presence of changes in the environment.

All test are performed in real outdoor scenarios. Our data sets, each consisting of a large number of omnidirectional images, have been acquired over different day times both in indoor and outdoor environments. Results of robot localization in the historical centre of Fermo (Italy) are presented and discussed; in Figure 1 an example of omnidirectional views of Fermo downtown is shown.

Main novelties of the paper rely on: novel probabilistic approach specially designed for omnidirectional vision based localization, on line information integration, experiments performed in long term robot localization in real word scenarios.

The paper is organized as follows: we first introduce changes made to the generally used MCL algorithm (section 2) and the on-line information integration (section 3), then, before the conclusions, in section 4 we describe the experiments performed and the localization results obtained in the old town of Fermo, Italy.

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353

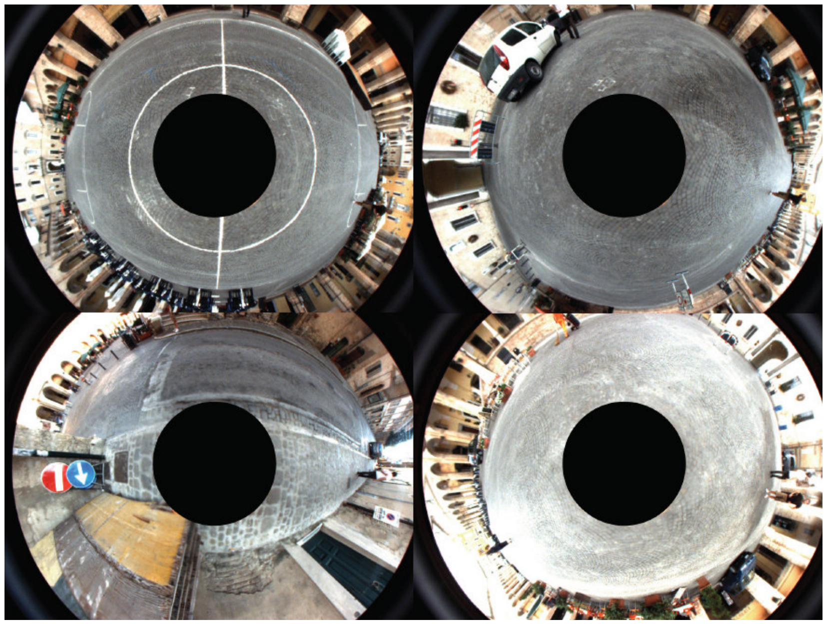

Figure 1 Example of omnidirectional views

2 Changes to basic MCL algorithm

All the proposed modifications described below are designed for omnidirectional vision. We do not take into account angular rotation and the resampling is always done working only in X and Y coordinates. The first step of the original MCL algorithm [9-10] was sorting particles in a random way on the whole map, even in areas far from dataset images, which represent robot knowledge. So, we chose to assign the same number of particles to each dataset image, and to put them inside a square of side 2δ centered on image location, where δ is one of the algorithm parameters.
The square described before covers an area equal to:

$$\frac{map\_area}{number\_image\_dataset}$$

Where *map_area* is the total area of the map (measured in square meters) and *number_image_dataset* is the number of image contained in the dataset (a sort of

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353

knowledge density of the environment).

A second change is related to the *particle weight* assignment. When the motion model is applied to every particle, each of them is attributed to a cluster of ray 2δ; then, a weight can be computed for each particle as:

$$w_j = similarity \frac{number\_particle\_cluster}{total\_number\_particles}$$
$$j = 1,...,total\_number\_particles$$

where *similarity* is the similarity value for the dataset image closest to the particle, *number_particle_cluster* is the number of particles of the cluster and *total_number_particles* is the total number of particles.
Resulting from this new weight assignment method, we changed robot pose estimate, using the weighted (by particle-weight) mean of particle-position in the best cluster. In particular, robot pose x_s is obtained from the evaluation of the following expression:

$$x_s = \sum_{i=1}^{k} x_s^i w_i, \quad k = number\_particles\_cluster$$

Where w_i is the weight of each particle.
Finally the *re-sampling* function was modified so that a number z of particles equal to:

$$z_j = w_j * total\_number\_particles$$

is put only around particles with highest weight (resulting from the sensor model that update every particle weight).

We also add other particles to the best one as:

$$z_k = z_k + 0.4 * diff$$

Where k is the particle with best weight and *diff* is the difference between *total_number_particles* and the total number of particles distributed over other particles: sum(z_j). The remaining percentage is placed randomly and uniformly, as in the first step of the MCL algorithm, trying to resolve *kidnapping* robot problem.

Starting from the previous improvements, we added an option that allows to place initial particles near or far from dataset images, depending on a priori knowledge of the robot path. So, during initial positioning of particles, the value of δ is computed as described in the first paragraph of this section with *far* option, and it is smaller with

near one (the half of the *far* option).
Functions that implement motion model application and particles-images assignment have not been changed.
Functions concerning particle weight assignment and re-sampling, on the contrary, were substituted with a single one described here following. Also this modification works well with omnidirectional vision but not with directional one due to the fact that is difficult to associate each particle to a certain image considering also the orientation.
A weight equal to its normalized similarity value is attributed to every dataset image; then, the particles linked to each image are counted and their centre of mass is computed. Now, if the number of particles is different from zero and the weight is greater than a certain threshold (computed as the 25% of the biggest weight), we replace around the previous center of mass a new number of particles as:

$$nPart_j = \alpha * w_j * total\_number\_particles + \beta * npart\_associat_j,$$

$$j=1,\dots, number\_images\_dataset$$

where α and β are parameters experimentally evaluated and *npart_associat* is the number of previously associated particles. The window centered in the centre of mass has the same form described at the beginning of the paragraph or the initial distribution with the "near" option. Finally we find the best image in terms of:

$$w_j * npart\_associat_j$$

and we add a number of particles to the related cluster as in the first variant (0.4*diff). Robot pose is finally estimated after this function using a robust mean shown in the first variant, but cluster weight w_j is defined as:

if "near"

$w_k = number\_particle\_cluster * w_j * e^{-distance(j)}$

else

$w_k = number\_particle\_cluster * wt_j$

where w_j is the weight of the particle nearest to the centre of mass and *distance(j)* is the distance between image and associated particle. Position error is evaluated as the Euclidean distance between robot pose estimate and the real one.

3 On line information integration

Several vision based techniques are integrated in our framework, which can be synthesized as follows. During the learning phase a visual tour in the environment is performed to collect reference images (i.e., snapshots of the environment,

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353

omnidirectional images), stored in database as feature vectors; the visual tour supplies also information about the presence of obstacles in the environment, which can be obtained using a set of sonars to scan the area near the robot. Besides, a stochastic position evaluator, which will be used to describe the belief distribution about the actual position of the robot in the environment, is initialized. At the beginning of the localization phase the robot does not know its position (kidnapped robot). The robot first compares the appearance of the image acquired from the unknown position with images in the database. Then the system selects the most similar image from the database and starts moving, according to a planning strategy given by a high level control layer.

Usually this kind of approach need a precise and almost complete exploration of the environment, making the application of this approach in everyday life very difficult, most of all in huge outdoor city like environments.

One of the purposes of this paper is to propose a way of using this approach also using very few images, thanks to omnidirectional vision and on line information integration.

Figure 2 - On line knowledge integration: red circles represent the previous knowledge (omnidirectional images with relative position in the environment); blu line is the robot trajectory; blu dots are MCL particles that, in case of good accumulation and low image density, bring to the knowledge update.

The general idea, depicted in Figure 2, is to cover unexplored areas of the environment with images acquired on line, according to precise rules based on the correct robot localization at the acquisition time. These rules are based on the concentration of MCL particles around a certain position; in this condition the robot is considered as well localized and the image is added to the reference image database with the actual robot position estimated by the MCL process. The main drawback of

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353

this method could be the increase of memory occupation during information integration for long term localization processes. To solve this limitation we introduce the concept of image density that is given be the number of known images, contained in the reference image database, per square meter. This number gives a threshold for the update procedure assuring that the number of new images added to the environment knowledge is limited.

4 Experiments and results

In our experiments, we use images collected by our outdoor mobile robot ActiveMedia P3-AT. We took one 1034x789 pixel grayscale image every almost 5 meters (not covering all the test area) using a Basler Scout A1300c camera with an omni directional mirror mounted on top of the robot. The robot is also equipped with a differential GPS (DGPS with EGNOS corrections) system, which we used to get ground truth data for the position of each image. Under ideal conditions, the accuracy of the DGPS is below 0.5 m. However, due to occlusion by trees and buildings, the GPS path sometimes significantly deviated from the real position or contained gaps. From our knowledge that we moved the robot on a smooth trajectory, we corrected some wrong GPS values manually. This was used for ground truth registration.

Several different runs were performed and some of them exhibit very dynamic scenes with occlusions, deriving from some students walking around the robot. Figure 3 shows the whole are of 12000 square meters explored by the robot and the lines are the routes followed by the robot for dataset collecting and for test set recording. For every dataset we need the image acquired by the robot, its topologic label, its GPS position, the odometric position of the robot; other general features of the whole dataset, like, for example, lighting conditions, dynamic conditions, camera parameters are also considered.

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353

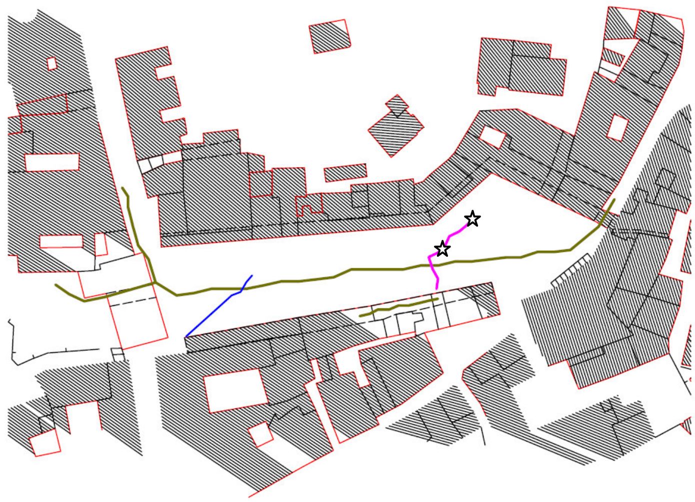

Figure 3 The area of 12000 square meters of the historical centre of Fermo used for dataset collecting; green line represents the dataset and blue and red ones represent the paths.

The modified MCL algorithm together with on line information integration were tested with the same path and dataset images. The localization error in this outdoor environment, expressed in meters, and compared with old MCL algorithm, decreases for the new approach from an average error of 5 meters to an average error around 1 meter. Table 1 shows detailed results for the localization error in terms of the mean error and its standard deviation. As it can be seen, the difference between 500 particles and 300 particles is not high. Besides, as regard standard deviation, respective results suggest a good stability of the algorithm independently of the particle number. The same stability is visible with respect to the feature matching algorithm (SIFT and SURF) showing that the influence of the feature matching module is less important than the proposed MCL improvements.

The average error of 1 meter is an extremely good result if we consider the number of images used as references; if compared with classic MCL approaches we obtained a decrease of error of more then the 50%.

Environment	**#particles**	**Algorithm**	$E[err]$	σ_{err}
Outdoor	500	SIFT	1,2 m	0,2 m
Outdoor	300	SIFT	1,1 m	0,2 m
Outdoor	500	SURF	1,0 m	0,2 m
Outdoor	300	SURF	1,1 m	0.3 m

Table 1 Numerical results of outdoor localization error simulation for different feature matching algorithms and number of particles.

Unlike the old one, the new MCL algorithm decreases particles number at every step thanks to its new re-sampling function previously explained. Figure 4 shows the particle number trend during outdoor simulation described above, with an input value of 500.

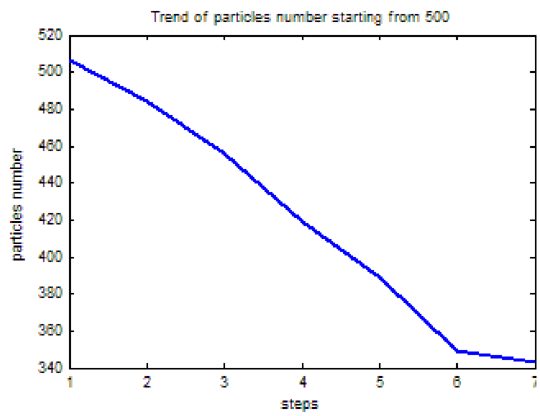

Figure 1 Particle number trend in outdoor simulation

Good results were obtained also for *kidnapping* cases (that in populated outdoor environments can be, for example, associated to unexpected odometric errors due to floor imperfection or steps, or due to collisions with people passing by or simply people moving the robot). To simulate them, we firstly linked path 2 (5 steps) with path 1 (7 steps); in this way, we could sample the algorithm in a more critical condition. In fact, we introduced an odometry error between the two steps linking the different paths, so we could check how the algorithm responds to this type of common mistake. Figure 4 shows the result of one simulation with an odometry error equal to ¾ of the real displacement, between step 5 and 6.

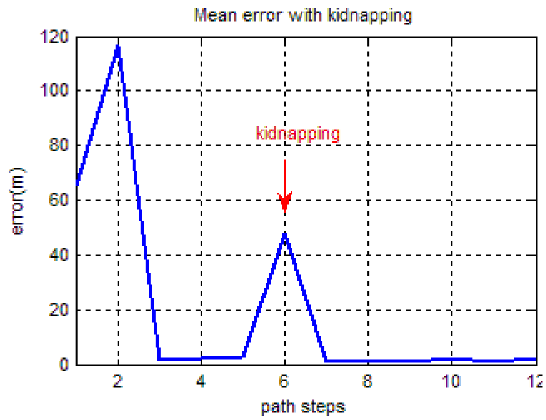

Figure 5 Results of mean error with a kidnapping situation

The result is very interesting: robot finds its position after kidnapping with only one step, when it is able to localize itself with an error of about 1-2 meters.

In the field of information integration we have some preliminary results showing that the integration allows to add only a few images in the localization process, but sufficient to cover unexplored area of the environment and to allow correct localization also during robot trajectory not covered, initially, by reference images. In the experiments reported before we added no more than 1 to 3 new image every ten steps and we saw that the new image helped to cover more uniformly the test area. In fig. 3 the pink route was performed far away from the explored part of the environment, but the localization was still well obtained and two new images (depicted by white stars in the figure) where added to the reference image database. Even if preliminary results must be extended and deeply investigated the proposed method seems to be reliable for real life application and long term localization in real dynamic environments. The vision based localization of the future will start with a man taking some pictures around in the environment with his mobile phone equipped with GPS; this will be the initial robot knowledge and then, after a fast installation process based on very few reference images, the robot will work in the environment basing the localization process only on low cost omnidirectional vision.

5 CONCLUSIONS AND FUTURE WORKS

In this paper we presented some novel Monte Carlo Localization optimizations especially designed for omnidirectional vision based locaziation and we also introduce the concept of on line knowledge acquisition and integration presenting a framework able to perform long term robot localization in real environments.

In summary we can assert that the two variants (MCL modifications and on line knowledge integration) have new features, with respect to the old algorithm, that introduce improvements in term of localization accuracy, precision and robustness. Besides, these ones are two new approaches to probabilistic robotics which can be delved into and improved so as they can give higher performances, with the integration of other sensors information too.

Current and future works are going in the direction of finding best and more robust vision based algorithm for feature matching. In particular we also address the issues of appearance-based topological and metric localization by introducing a novel group matching approach to select less but more robust features to match the current robot view with reference images [11]. Feature group matching is based on the consideration that feature descriptors together with spatial relations are more robust than classical approaches.

References

1. S. Se, D. Lowe, and J. Little, "Vision-based mobile robot localization and mapping using scale-invariant features," in Proceedings of the IEEE International Conference on Robotics and Automation (ICRA) 2001, Seoul, Korea, May 2001, pp. 2051–2058.
2. H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in Ninth European Conference on Computer Vision, 2006.
3. H. Bay, B. Fasel, and L. V. Gool, "Interactive museum guide: Fast and robust recognition of museum objects," in Proc. Int. Workshop on Mobile Vision, 2006.
4. C. Valgren and A. Lilienthal, "Sift, surf and seasons: Long-term outdoor localization using local features," in Proc. of 3rd European Conference on Mobile Robots, Freiburg, Germany, 2007.
5. B. Krose, O. Booij, and Z. Zivkovic, "A geometrically constrained image similarity measure for visual mapping, localization and navigation" in Proc. of 3rd European Conference on Mobile Robots, Freiburg, Germany, 2007.
6. E. Frontoni, A. Mancini, and P. Zingaretti, "Vision based approach for active selection of robots localization action," in Proc. of the 15th Mediterranean Conference on Control & Automation, Athens, Greece, 2007.
7. H. Andreasson and T. Duckett, "Topological localization for mobile robots using omni-directional vision and local features," in 5th IFAC Symposium on Intelligent Autonomous Vehicles (IAV), Lisbon, 2004.
8. C. Valgren, T. Duckett, and A. Lilienthal, "Incremental spectral clustering and its application to topological mapping," in Proc. IEEE Int. Conf. on Robotics and Automation, 2007, p. 42834288.
9. F. Dellaert, W. B. D. Fox, , and S. Thrun, "Monte carlo localization for mobile robots," in Proc. 1999 IEEE Intl. Conf. on Robotics and Automation (ICRA), 1999.
10. E. Frontoni, A. Ascani, A. Mancini, P. Zingaretti, "Performance metric for vision based robot localization", in Proc. of Robotics: Science and Systems Conference, Zurich, Switzerland, 2008.
11. A. Ascani, E. Frontoni, A. Mancini, and P. Zingaretti, "Feature group matching for appearance-based localization" in Proc. of the International Conference on **I**ntelligent Robots and Systems, Nice, France, 2008.

Workshop Proceedings of SIMPAR 2008
Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS
Venice(Italy) 2008 November,3-4
ISBN 978-88-95872-01-8
pp. 343-353