

Image-Based Monte-Carlo Localisation without a Map

Emanuele Menegatti*, Mauro Zoccarato, Enrico Pagello†, Hiroshi Ishiguro ‡

Intelligent Autonomous Systems Laboratory
Department of Information Engineering
The University of Padua, Italy

† *also with* Institute of Biomedical Engineering
of the National Research Council (ISIB-CNR) Padua , Italy

‡Department of Adaptive Machine Systems
Osaka University
Suita, Osaka, 565-0871 Japan

*emg@dei.unipd.it

Abstract. In this paper, we propose a way to fuse the image-based localisation approach with the Monte-Carlo localisation approach. The method we propose does not suffer of the major limitation of the two separated methods: the need of a metric map of the environment for the Monte-Carlo localisation and the failure of the image-based approach in environments with spatial periodicity (perceptual aliasing). The approach we developed exploits the properties of the Fourier Transform of the omnidirectional images and uses the similarity between the images to weights the *beliefs* about the robot position. Successful experiments in large indoor environment are presented in which we do not used a priory information on the metrical map of the environment.

1 Introduction

In mobile robotics the localisation problem is a fundamental problem. A robot in order to effectively fulfill an assigned task, must be able to move from point to point and to be able to know its position in the environment. Several solutions have been proposed to this problem. In this work we focused on two successful solutions developed in the past years: the image-based localisation and the Monte-Carlo localisation. Both of these techniques are well known and well known are their limits, i.e. the frame of assumptions under which they are effective. In this work we propose a way to fuse these two techniques trying to overcome, with this fusion, their intrinsic limitations.

The image-based localisation approach [1, 5, 13, 15, 19, 9] is based on the idea that most of the time, the robot does not need a metrical localisation and a topological localisation is enough. We can choose some points in the environment, called **reference locations**. Every time the robot makes a new measurement (i.e. it grabs a new image), it can infer at which of these points it is closer (topological localisation). The robot infers the reference location corresponding to its actual

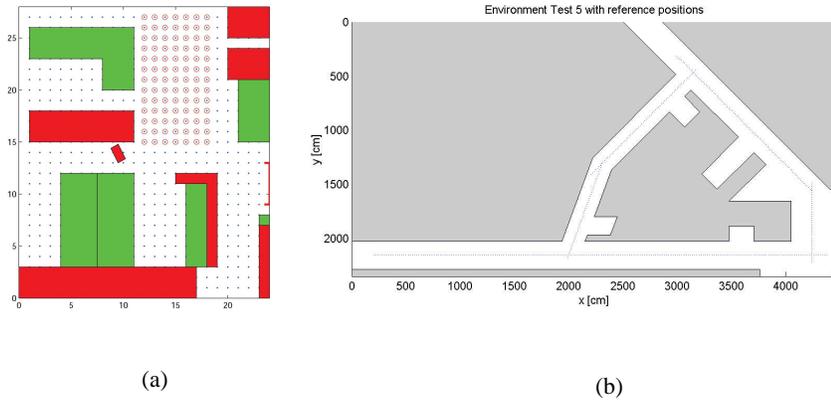


Fig. 1. (a) The sketch of the test environment in a laboratory cluttered with many pieces of furniture. The dimensions of the test environment are about $5m \times 2m$. (b) The sketch of the environment where most of the tests were carried out. This is a large environment of about $50m \times 25m$.

position by comparing the current image it grabbed, called **input image**, with the **reference images** previously grabbed at the reference points. The flaw in this approach is the fact the localisation is purely reactive: to a sensorial input, immediately corresponds a localisation. The image-based localisation approach does not work if more than one reference location corresponds to a input image. In this case the robot cannot choose between the two possible solutions. This happens in environments with highly periodical structures where the appearance of the environment is the same at different locations, like in a maze or in a hospital or in an office building.

The Monte-Carlo localisation approach was developed using range finder sensors like sonars or laser range finders. It is based on the idea of scanning the environment at a certain location and matching this scan to a geometrical map [10]. The possible location of the robot is one of the locations from which the robot would have experienced that scan. The belief about the robot position is managed with a Bayesian approach, as detailed in Section 3. This probabilistic approach allows to manage multi-modal probability distribution of the believed robot position and to fuse the robot's odometry with the range observations. The limit of this approach is that it requires to know a detailed geometrical map of the environment in order to associate a probable robot location to a certain scan.

In this work, we combine the image-based localisation approach with the Monte-Carlo Localisation approach. The advantage is that we do not need a geometrical map of the environment, because the probable position of the robot is calculated using the similarity between the current input image and the reference images. The Monte-Carlo localisation approach allows us to manage the case in which the current image is similar to more of one reference image (i.e. multi modal probability distribution for the robot position) and to overcome the limitations of the reactive localisation. We will see in Section 3, how the Monte-Carlo localisation

sation keeps track of the history of the robot states by updating the past beliefs about the robot position based on the actual motion performed by the robot. Other authors proposed a similar fusion. Wolf, Burgard and Burkhardt proposed to use a perspective camera and an image retrieval system for the image-based localisation part. In our opinion, the disadvantage of his method is that it needs a large number of reference images and it needs a geometrical map of the environment [23]. Kröse et al. used an omnidirectional camera and an image matching method based on Principal Component Analysis to reduce memory required to store the reference images. For managing the probability distribution of the robot position, he used a Markov Localisation with an occupancy grid. They focus on the description of the environment rather than on the localisation process. They do not use a model of the motion of the robot and every time the robot takes a new step all positions in the grid have the same probability to contain the robot. The approach we propose in this paper has been extensively tested in real world indoors environments. We performed preliminary test in the small laboratory of Fig. 1 (a) and more challenging tests in the corridors of Fig. 1 (b). Both are at Floor 7 of Wakayama University, Japan.

2 Image-based Localisation

The main problem tackled by researchers in the image based localisation approach is how to store and to compare a large number of reference images. In fact, if a perspective camera is used, the appearance of the environment from a reference location changes with the direction of gaze. Nevertheless, there are several successful works where perspective cameras have been used. A solution could be to constrain the movements of the robot in order to have the perspective camera gazing always in the same direction [5]. An alternative solution could be to acquire a large number of images with different gaze directions, but to extract from the perspective images compact distinctive features which reduce the amount of memory required to store the images [23].

The best solution in order to reduce the number of reference images is to use omnidirectional cameras. For every reference location we will have only one omnidirectional image, containing the appearance of the environment in all possible gazing directions. The visual memory constructed with the collection of reference images is usually pre-processed to further reduce the memory requirements. The most common approach is to extract a set of eigenimages from the set of reference images and to describe every image as linear combination of these eigenimages. Unfortunately, this approach does not lead to a rotational invariance of the images. In other words, if the image grabbed by the robot is rotated with respect to the corresponding reference image the two images are not longer recognised as similar images. So, authors using this approach need to constrain the heading of the sensor as in [15, 20] or to further preprocess the images in order to obtain the rotational invariance as in [1, 14, 9].

In this work, we decided to reduce the number of images required to describe the environment by mounting an omnidirectional camera on the mobile robot [12, 16, 17]. In order to efficiently store and process the omnidirectional images we exploited the properties of the Fourier transform of panoramic images as detailed in a previous work [13]. Every time an omnidirectional image is grabbed, it is transformed in a compact representation. The calculation of the compact representation is performed as follow (Fig. 2). The omnidirectional image is unwrapped

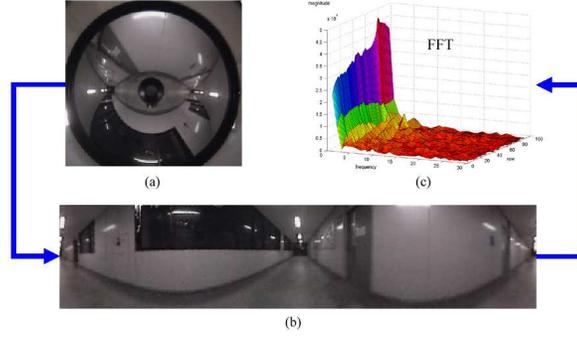


Fig. 2. The process of generation of the Fourier Signature. The omnidirectional image (top left) is unwrapped in a panoramic cylinder (bottom) and the Fourier coefficient of every row of the panoramic cylinder are calculated (top right).

creating a **panoramic cylinder**. The panoramic cylinder is expanded row by row into Fourier series. Only the magnitude and the phase Fourier coefficients associated to the lowest frequency components are stored. The result of this procedure is a compact representation of the original omnidirectional image that we called **Fourier signature**. The Fourier signature drastically reduces the amount of memory required to store a view at a reference location and the computational power required to match a current view against the visual memory. In addition, the Fourier signature is the natural representation to implement a rotational invariance, because the magnitude coefficients of the Fourier components are invariants to rotation of the sensor heading. For more details on this procedure, please refer to [13].

We defined a **dissimilarity function** on the Fourier signatures. The higher this value, the more two images are dissimilar. The dissimilarity function is defined as the $L1$ norm of two Fourier signatures:

$$Dis(O_i, O_j) = \sum_{y=0}^{l-1} \sum_{k=0}^{m-1} |F_{iy}(k) - F_{jy}(k)| \quad (1)$$

where k indicates the frequency and y is the index of the row of the panoramic cylinder.

For convenience we defined also a **similarity function** obtained rescaling the dissimilarity values between 0 and 1000.

$$Sim(O_i, O_j) = 1000 - 1000 \frac{Dis(O_i, O_j) - Min_{i,j}\{Dis(O_i, O_j)\}}{Max_{i,j}\{Dis(O_i, O_j)\} - Min_{i,j}\{Dis(O_i, O_j)\}} \quad (2)$$

A robot localisation technique to be used in a populated environment should be able to cope with moving objects, like people, that may cause temporary occlusions in the image. Experiments we conducted, but that we cannot report because of lack of space show the Fourier Signature is robust to occlusion.

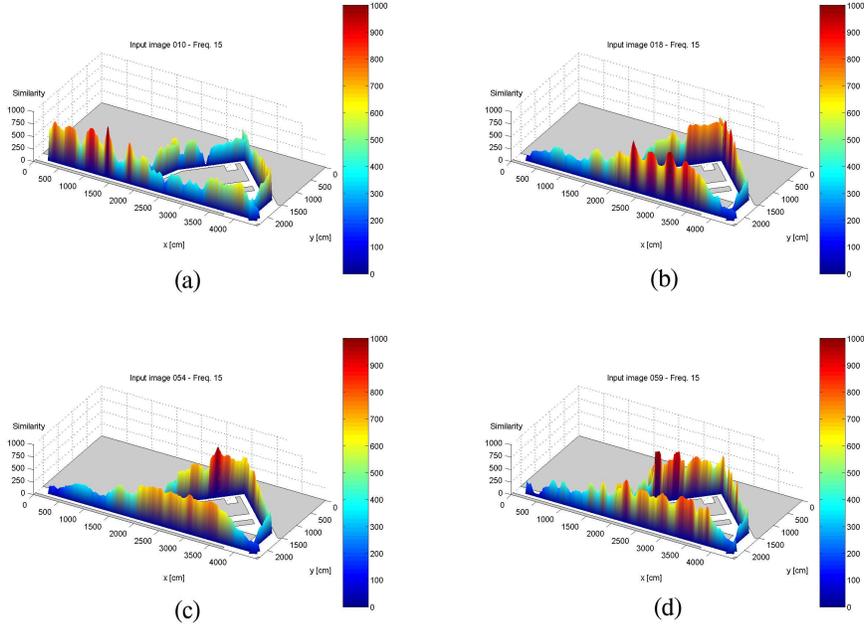


Fig. 3. Some examples of perceptual aliasing in our test environment. The colour scale indicate the amount of similarity of the different images: red the heighest similarity, blue the lowest similarity. (a) Input Image 10 matches not only the correct reference image at position 1380 on the x-axis, but also reference images at 320, 620 and 1060. The same happens in plots (b), (c) and (d).

As we stated in the introduction, the image-based localisation approach has a flaw: i.e. it fails in environments where two different locations look the same. This is the trick used to get people lost in a labyrinth, i.e. all the junctions of the labyrinth looks the same and one cannot distinguish one junction from another. This situation is called “perceptual aliasing”. Nonone of the presented works on image-based localisation can cope with such a situation, except [23]. More information than the simple image similarity is needed to be able to localise the robot. The labyrinth is an extreme example, but also in our everyday-life indoor environment perceptual aliasing is present. In Fig. 3, we present the calculation of the similarity of a current image with respect to all reference images in the environment. The current image is assessed as similar not only to the corresponding image, but is similar also to other reference images at different locations. In order to provide robust image-based localisation, despite the perceptual aliasing we exploited the Monte-Carlo Localisation approach to keep track of the belief about the robot position.

3 Monte-Carlo Localisation

To manage the uncertainty about robot position caused by the perceptual aliasing, we adopted a well-known approach based on Monte Carlo Methods. This approach has been successfully used by many researchers [21, 6, 3, 22, 23] and is referred as Monte Carlo Localisation.

This is a probabilistic methods based on Bayesian Filtering (Markov Localisation in robotics) [2, 23, 8]. It calculates the probability density of robot position (the **belief**) and recursively propagates this probability density using motion and perception information. In our implementation motion data come from the odometry sensors and perception data come from the omnidirectional vision system. The belief about robot position is updated every time the robot grabs a new image (i.e. a new observation of the world).

In robotics, many technique has been applied to approximate a continuous density for the beliefs. Two of the most used techniques were Kalman Filter and Grid-based Methods, but these methods presented many disadvantages [7] and new techniques have been proposed like the Particle Filter approach.¹

The Particle Filter approach represents the belief about the robot position with a set of discrete points in the configuration space of the robot. These points are called **samples**. To update the belief over time, these samples are updated. To every sample is associated a weight indicating the probability that the robot is occupying that position. The samples with the lowest weights are deleted and the samples with highest weights survive and generate new samples.

This approach is efficient and can solve the well know problems of Position tracking and Global Localisation [8]. Nevertheless, Particle Filters alone cannot solve the problem of the Kidnapped robot [8]. In [8] was proposed to continuously generate a subset of samples uniformly distributed in the environment at random positions . This approach relies on the hope that some samples of this subset will be generated by chance close to the position where the robot has been moved. In this case, these samples will have height weights in the next measurements and will survive and cluster the other samples close to them. The drawback of this approach is that it takes several iterations for the samples to cluster around the correct sample and sometimes it takes some iteration to generate a sample close to the position where the robot was kidnapped. In the next section, we will present our solution to speed up the position recovery.

3.1 The proposed kidnapping strategy

In order to understand the new kidnapping strategy we propose, we need to briefly review the theory behind the Monte-Carlo Localisation approach. In the Bayesian Filtering problem one has to calculate the probability density of the robot position $Bel(s_t) = p(s_t|O_t, a_t)$ over the time. What is known is: the **prior probability density** $Bel(s_0)$, which describes the initial robot uncertainty about its position; the **prediction model** $p(s_t|s_{t-1}, a_{t-1})$ that applies motion data a_{t-1} to actual state s_{t-1} obtaining a new state s_t ; the **observation model** $p(O_t|s_t)$ that represents the probability of making observation O_t from state s_t .

¹ For a complete review and evaluation of localisation method refer to the two enlightening works of Gutmann [10, 11].

Using the Bayes Formula and Markov assumption about the state space the equation to calculate the belief is

$$Bel(s_t) = \eta p(O_t | s_t) \int p(s_t | s_{t-1}, a_{t-1}) Bel(s_{t-1}) ds_{t-1} \quad (3)$$

where η is a normalisation factor. In Monte-Carlo Localisation the probability density of the robot position (i.e. the belief) is approximated with a set of weighted samples $\{s_t^i, \omega_t^i\}_{i=1 \dots N}$. These samples are updated recursively with a procedure called **sampling-importance-resampling** composed of following steps: For each samples s_{t-1}^i in previous set, a new samples \hat{s}_t^i is drawn from $p(s | s_{t-1}^i, a_{t-1})$ and s_{t-1}^i is replaced with \hat{s}_t^i ; Given the new observation O_t , each sample \hat{s}_t^i is weighted proportionally to $P(O_t | \hat{s}_t^i)$; The samples are replaced proportionally to the weights [4], i.e. if a sample has weight 0.7, the sample is replaced with 7 identical sample.

To tackle the kidnapped robot problem the standard approach is to replace, after the prediction step, a certain amount of samples with others samples randomly drawn in the entire environment. This technique is robust, but in general require many steps to re-localise the robot, especially in large environments. We propose a new solution: instead of randomly drawing the sample, the new samples should be generated around the reference images that best match the current input image. This assures that the newly generated samples are concentrated around possible locations only. This approach is made possible by the technique we use to match the input image with the reference images. In fact, in the calculation of the similarity, Eq. 2, if we stop the summation over the frequencies at low values of k , we obtain a broader localisation, as detailed in [18]. In other words, if we use only the Fourier components corresponding to the lowest spatial frequency, the current input image does not match only the corresponding reference image, but also other similar reference images. Usually, these reference images are close to the correct reference image, i.e. the new samples concentrate in a region around the correct reference image, but sometimes these similar reference images can be one far away from the other (e.g. in case of perceptual aliasing), i.e. the new samples take into account alternative possibility for the actual location of the robot. In both cases, our system recover very quickly the correct position after the kidnapping. To be noted that our technique not only re-localise quickly the robot after a kidnap but also speed-up the global localisation process.

We tested our system in a large indoor environment with a long corridor and a loop (a path about 100 m long): the robot faced many false hypothesis, but the obtained results are good. Particularly we can exploit hierarchical localisation to consider a wide number of hypothesis: that is we can use a lower number of frequencies to generate the new samples with the strategy. Remember that if we use m frequencies in weight process automatically we have all similarity data for $m' \leq m$. Using a lower number of frequencies we obtain a broad localisation that is a wide number of hypothesis that after will be controlled by MCL.

4 Experiments

We tested our system in the two environment of Fig. 1 (a) and (b). The first environment is a small cluttered laboratory and we use this as a first testbed for our system. We obtained encouraging results in Global Localisation, Position

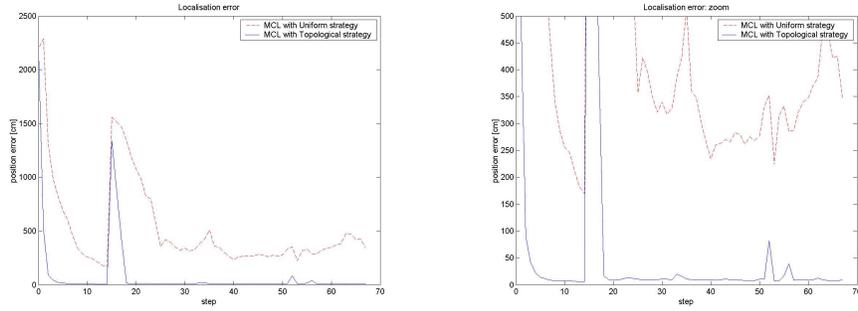


Fig. 4. Localisation error during the path of 68 step of Fig. 5(a). The Topological strategy we proposed outperforms the uniform strategy in robustness and speed in recovering the correct position.

Tracking and Kidnapped Robot problem. We presented these tests in [18]. To confirm the good performances of our system we carried out several experiments in the environment of Fig. 1 (b). This is a very challenging environment. It is a corridor 90 m long with a highly periodical structure and with many locations which look similar, i.e. the same lighting, the same junctions, etc. This means that image-based localisation fails because it could not decide between the various topological localisations as earlier described in Fig. 3. In this environment the kidnapping robot problem is particularly challenging, due to this high perceptual aliasing. The **uniform strategy** proposed by Burgard et al. give non-optimal results in this environment (cnf. with [18]). In Fig. 4 can be noted the *uniform strategy* requires many steps to localise the robot from scratch during the Global localisation. In the same way, it requires many steps to re-localise the robot after a kidnapping. This is because, in such a large environment, the probability to generate a sample close to the new robot position is low, so it takes many steps to cluster the samples around correct position. This lead to a wide error during the re-localisation process. As it could be seen in Fig. 4 our strategy is more effective. With the mixed technique proposed in this paper, it is possible to speed-up robot global localisation and also to quickly relocalise the robot after the kidnapping. In the experiments of Fig. 7, the robustness of our technique with respect to perceptual aliasing is tested, because the kidnapping is made in a position in the longer corridor with high perceptual aliasing as you could see in Fig. 3 (b). So the re-localisation process have to trial several hypothesis after kidnapping and require 5-6 step to assess the new position.

In Fig. 6 you can see the snapshot of the initial belief at the beginning of the global localisation process. As the system has no information about the robot position, the samples are generated uniformly in the area of the environment. Notice that we don't control if the samples are in an allowed area or not, Fig. 1 (b). This is not necessary. If a sample is inside an object it will be far away from any reference image. Its weight will be zero and at the next step it will be removed by the *sampling-importance-resembling*. In this way we don't use any map of the

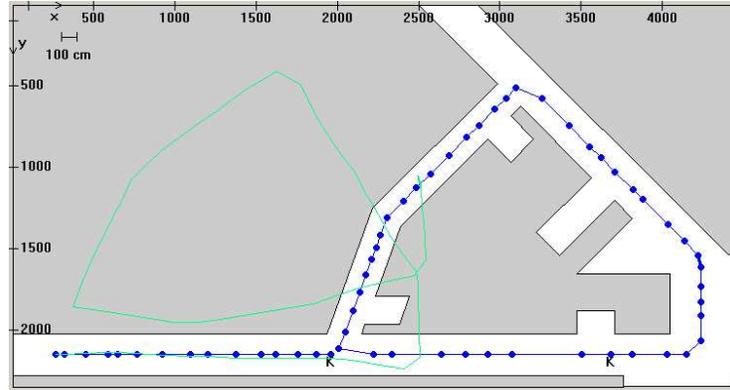


Fig. 5. The path we used to test our system in environment of Fig. 1. The blue dotted line is the real path with indicated the positions where the robot grabbed the images. The green line is the path reported by odometry.

environment. In Fig. 7 you can see some snapshot of the tests with the uniform strategy (left series of images) and with the proposed topological strategy (right series). As you can see our system localise robot quickly after about 5 steps and after this it continues as a position tracking with clustered samples around true position. When the robot is kidnapped our system consider the hypothesis based to vision information and quickly cluster the samples around the true position re-localising the robot in about 4 steps. Instead using uniform strategy it need about 10 steps to re-localise robot. Moreover after re-localisation our system continue as a position tracking whereas with the uniform strategy the samples are generated randomly in the environment and this make less accurate the estimation.

5 Conclusions

In this paper we propose a way to fuse the image-based localisation approach based on an omnidirectional sensor with the Monte-Carlo localisation approach. Using an omnidirectional camera we can reduce the amount of memory required to construct a visual memory for the robot because we can reduce the number of reference images with respect to a perspective camera. Using Fourier signature to have a compact representation of an image, we obtain useful properties: a natural rotational invariance, a robustness to occlusion and a hierarchical localisation. To improve our image-based approach in the case of environment with high perceptual aliasing, we fused it with sample-based Monte-Carlo localisation. Particularly we proposed an improvement of this technique that exploits the vision information to generate some of the samples. Using hierarchical localisation we can concentrate samples around the true position while considering others hypothesis. We tested this approach in a challenging environment with high perceptual aliasing and we demonstrate that our system speed-up the global localisation task, and quickly relocalise after a kidnapping. Moreover, our system

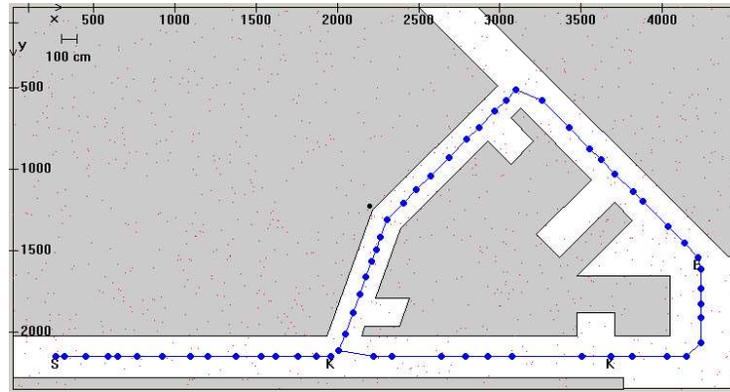


Fig. 6. Here you can see the initialisation of the belief in a global localisation task. The samples are generated uniformly in the area without any knowledge of the environment map. Black dot is the current mean estimated position.

does not need to know the geometrical map of the environment. We compared our technique with the standard uniform approach and showed the advantages of topological strategy. At the moment of writing we are testing our system in a large outdoor environment, particularly to understand the influence of luminosity pattern in our image-based system.

References

1. H. Aihara, N. Iwasa, N. Yokoya, and H. Takemura. Memory-based self-localisation using omnidirectional images. In A. K. Jain, S. Venkatesh, and B. C. Lovell, editors, *Proc. of the 14th International Conference on Pattern Recognition*, volume vol. I, pages 1799–1803, 1998.
2. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, Feb. 2002.
3. W. Burgard, D. Fox, M. Moors, R. Simmons, and S. Thrun. Collaborative multi-robot exploration. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2000.
4. J. Carpenter, P. Clifford, and P. Fearnhead. An improved particle filter for non-linear problems. *IEEE Proc. Radar, Sonar, Navigation*, vol.146, 1999.
5. R. Cassinis, D. Duina, S. Inelli, and A. Rizzi. Unsupervised matching of visual landmarks for robotic homing using fourier-mellin transform. *Robotics and Autonomous Systems*, 40(2-3), August 2002.
6. F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte Carlo Localization for Mobile Robots. In *IEEE International Conference on Robotics and Automation (ICRA99)*, May 1999.
7. D. Fox. KLD-Sampling: Adaptive Particle Filters. In *Advances in Neural Information Processing Systems 14 (NIPS)*. MIT Press, 2001.

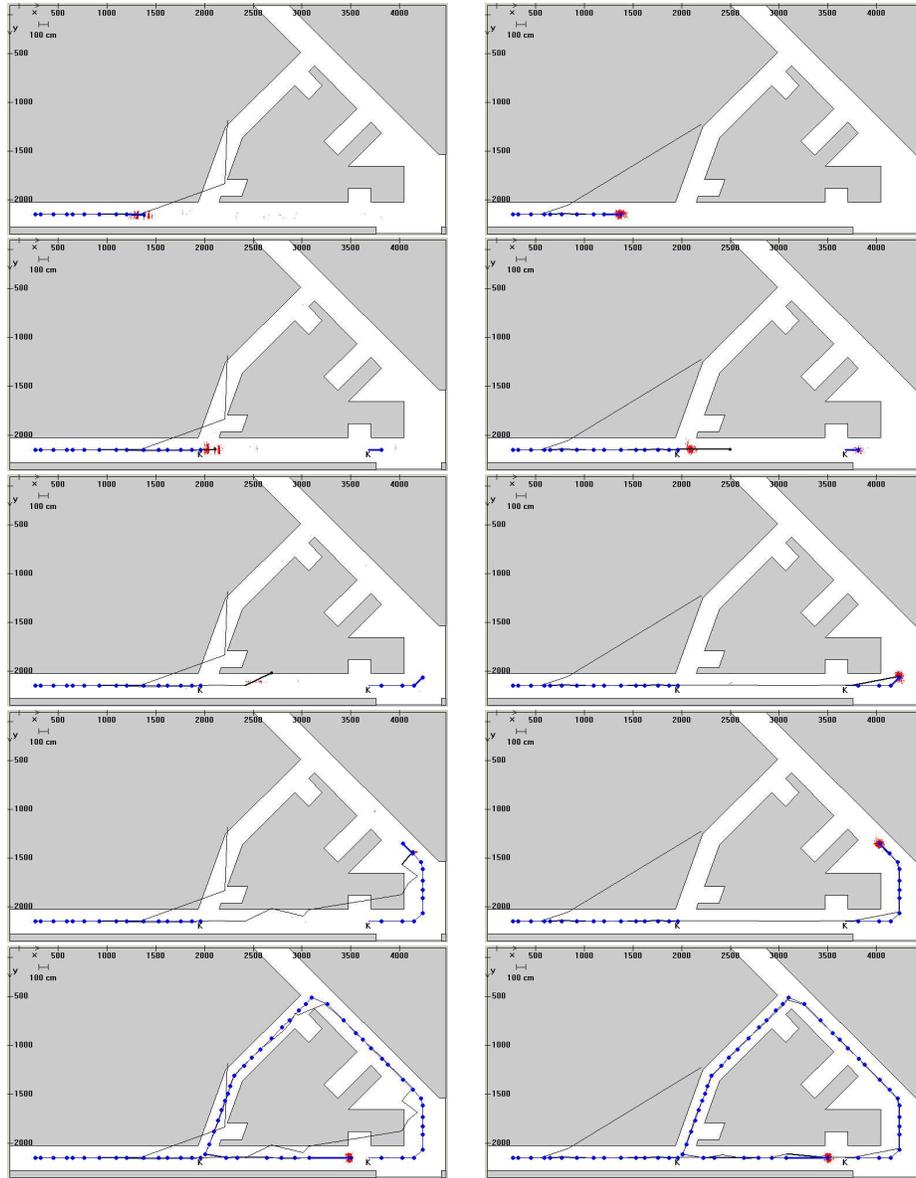


Fig. 7. Here you can see some snapshots of our tests. Uniform strategy (left series of images) is compared with our Topological strategy (right series). The blue line is the actual path of the robot and the black curve is the estimated path of the robot. You can see that with our strategy is possible to speed-up robot global localisation and also to quickly relocalise the robot after the kidnapping.

8. D. Fox, W. Burgard, F. Dellaert, and S. Thrun. Monte Carlo Localization: Efficient Position Estimation for Mobile Robots. In *Proceedings of National Conference on Artificial Intelligence (AAAI)*, 1999.
9. J. Gaspar, N. Winters, and J. Santos-Victor. Vision-based navigation and environmental representations with an omnidirectional camera. *IEEE Transaction on Robotics and Automation*, Vol 16(number 6), December 2000.
10. J.-S. Gutmann, W. Burgard, D. Fox, and K. Konolige. An Experimental Comparison of Localization Methods. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, October 1998.
11. J.-S. Gutmann and D. Fox. An Experimental Comparison of Localization Methods Continued. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), da pubblicare*, 2002.
12. H. Ishiguro. Development of low-cost compact omnidirectional vision sensors. In R. Benosman and S. Kang, editors, *Panoramic Vision*, chapter 3, pages pp. 23–38. Springer, 2001.
13. H. Ishiguro and S. Tsuji. Image-based memory of environment. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-96)*, pages pp. 634–639, 1996.
14. M. Jogan and A. Leonardis. Robust localization using panoramic view-based recognition. In *Proc. of the 15th Int. Conference on Pattern Recognition (ICPR00)*, vol. 4, pages 136–139. IEEE Computer Society, Sept. 2000.
15. B. Kröse, N. Vlassis, R. Bunschoten, and Y. Motomura. A probabilistic model for appearance-based robot localization. *Image and Vision Computing*, vol. 19(6):pp. 381–391, April 2001.
16. E. Menegatti, F. Nori, E. Pagello, C. Pellizzari, and D. Spagnoli. Designing an omnidirectional vision system for a goalkeeper robot. In A. Birk, S. Coradeschi, and S. Tadokoro, editors, *RoboCup-2001: Robot Soccer World Cup V*, pages pp. 78–87. Springer, 2002.
17. E. Menegatti and E. Pagello. Omnidirectional mirror design based on the vision task. In *Proceeding of 1st Nat. Conf. "Sistemi Autonomi Intelligenti e Robotica Avanzata*, pp.151–159, Frascati - Roma, Oct. 2002.
18. E. Menegatti, M. Zoccarato, E. Pagello, and H. Ishiguro. Hierarchical image-based localisation for mobile robots with monte-carlo localisation. In *Proc. of European Conf. on Mobile Robots (ECMR'03)*, (to appear), Sept. 2003.
19. T. Pajdla and V. Hlaváč. Zero phase representation of panoramic images for image based localization. In F. Solina and A. Leonardis, editors, *8-th International Conference on Computer Analysis of Images and Patterns*, number 1689 in Lecture Notes in Computer Science, pages 550–557, Tržaška 25, Ljubljana, Slovenia, September 1999. Springer Verlag.
20. A. Rizzi and R. Cassinis. A robot self-localization system based on omnidirectional color images. *Robotics and Autonomous Systems*, 34: 2001.
21. S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Cremers, F. D. Fox, D. Haehnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz. Probabilistic algorithms and the interactive museum tour-guide robot minerva. In *Intern. Journal of Robotics Research*, Vol. 19, pp. 972–999, Nov. 2000.
22. S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust Monte Carlo Localization for Mobile Robots. *Artificial Intelligence Journal*, 128(1-2), 2001.
23. J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization for mobile robots using an image retrieval system based on invariant features. In *Proc. of the IEEE Intern. Conf. on Robotics & Automation (ICRA)*, 2002.