Collecting outdoor datasets for benchmarking vision based robot localization

Emanuele Frontoni*, Andrea Ascani, Adriano Mancini, Primo Zingaretti

Department of Ingegneria Infromatica, Gestionale e dell'Automazione (DIIGA), Università Politecnica delle Marche, Ancona, ITALY - * frontoni@diiga.univpm.it.

Abstract—This paper aims to propose a roadmap for collecting datasets for benchmarking vision based robot localization approaches and for good experimental methodologies in this field. Also a performance metric for result comparison is proposed. In particular we present a novel way to measure performances with respect to the particular dataset used and we show the application of this method on two big datasets of omnidirectional outdoor images collected in the historical center of the town of Fermo (Italy). We believe that the robot vision field requires such benchmarking efforts to objectively measure its progress and to provide a direct quantitative measure of progress understandable to sponsors and evaluators of research as well as a guide to practitioners in the field

Index Terms—Localization, Vision, Feature Extraction, benchmarking, performance metric, data sets.

INTRODUCTION — This paper discusses the strategy for devising and using a new series of benchmarks in the robot vision field and in particular in the field of vision based robot localization. The aim is to provide a direct quantitative measure of progress understandable to sponsors and evaluators of research as well as a guide to researchers in the field. In particular a first set of benchmarks in two categories is proposed: vision based topological and metric localization. We believe that the robot vision field requires such benchmarking efforts to objectively measure its progress. There have been some previous benchmarking attempts in vision, but they dealt mainly with measuring the performance of computer architectures running vision algorithms, rather than with the performance of the vision algorithms and systems. The goal of creating challenging benchmark problems in robot vision is to pose a reasonably comprehensive set of vision problems to which proposed advances can be subjected to experimental evaluation. There would be challenging problems in various categories. e.g. indoor and outdoor scenes, recognition and grasping of manmade objects, time-varying scenes over days and seasons, and so on.

Here in particular we cope with the localization problem. The knowledge about its position allows a mobile robot to efficiently fulfil different useful tasks like, for example, office delivery. In the past, a variety of approaches for mobile robot localization has been developed. They mainly differ in the techniques used to represent the belief of the robot about its current position and according to the type of sensor information that is used for localization. In this paper we consider the problem of vision-based mobile robot topological and metric localization. Compared to proximity sensors, which are used by a variety of successful robot systems, cameras have several desirable properties. They are low-cost sensors that provide a huge amount of information and they are passive so that vision-based navigation systems do not suffer from the interferences often observed when using active sound- or light-based proximity sensors (i.e. soft surfaces or glasses). Moreover, if robots are deployed in populated environments, it makes sense to base the perceptional skills used for localization on vision like humans do. Over the past years, several vision-based localization systems have been developed. They mainly differ in the features they use to match images or for the appearance based matching (i.e. color histograms, Fourier signatures, etc.) in opposition to metric based approaches (i.e. binocular vision, calibrated omnidirectional vision, etc.). Several other

approaches differ for the probabilistic framework that uses the image similarity for position estimation. Local feature matching has become a commonly used method to compare images. For mobile robots, a reliable method for comparing images can constitute a key component for localization and loop closing tasks. 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.
Two different types of image feature extractor algorithms, SIFT and the more recent SURF, have been used to compare the images [1,3,4,8].
We yet proposed this dataset and a basic idea of performance metric in [11] and here we want to focus the attention to the way we collect datasets and to a good experimental methodology here proposed, taking into account also the kidnapping problem. Results of robot localization in the historical centre of Fermo (Italy) are presented and discussed.

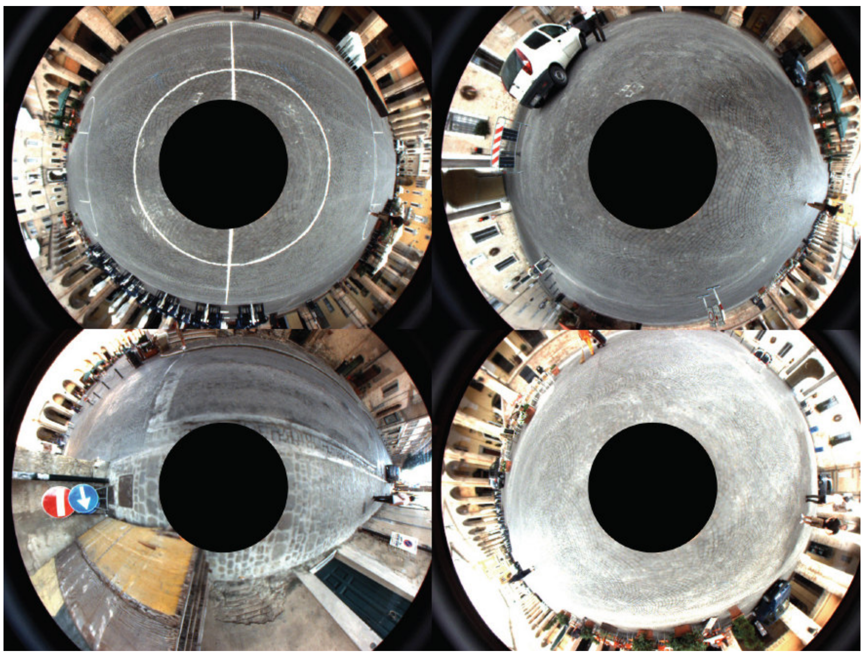

Figure 1 Example of omnidirectional images used for test

COLLECTING DATASETS IN OUTDOOR ENVIRONMENTS – The robot used for dataset collecting is equipped with an omnidirectional camera and a GPS. We register images together with the odometric position of the robot and the GPS position, used as ground truth.
Several different runs are performed and some of them have artificially created very dynamic scenes with occlusions, done by some students walking around the robot.
Figure 2 show the whole are of 12000 square meters explored by the robot and the lines are the routes done by the robot for dataset collecting and for test sets recording.
For every dataset we need to have the image acquired by the robot, its topologic label, its GPS posizion, the odometric position of the robot and some other general features of the whole dataset, like for example lighting conditions, dynamic conditions, camera parameters.

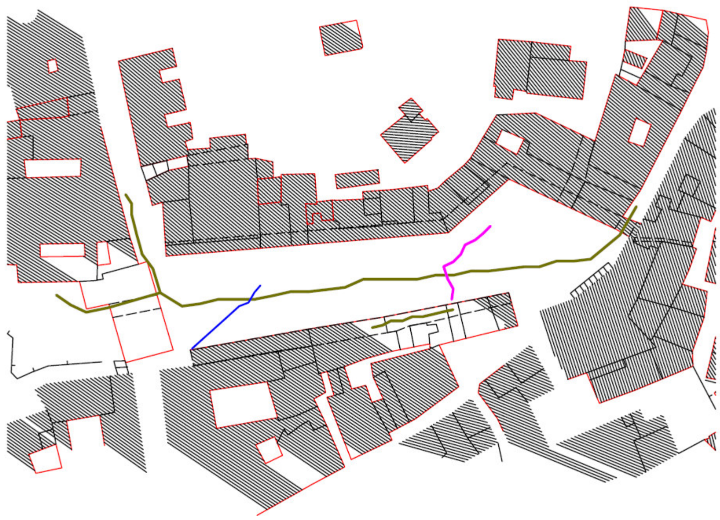

Figure 2 The 12000 square meters area of the historical center of Fermo used for dataset collecting; green line represents the dataset and blue and red one represents the paths.

PERFORMANCE METRIC EVALUATION - All the experiments should be performed in different datasets provided that they are collected using the following criteria and published on line to permit comparison. Every dataset must provide a reference image database collected in different environment conditions (in particular light or weather conditions for indoor or outdoor images). Every dataset must provide at least five different test sets collected in the same environment, but in different positions with respect to the data set of reference images. If possible, test sets should be dynamic and cover the case of presence of occlusion. The performance evaluator takes into account different aspects of the vision based localization problem:
- the size of the dataset S (considering both the extension of the area covered by the robot S1 and the number of images collected in the reference image database NS);
- the aliasing of the dataset A (a measure of the self-similarity of the environment and of the refence dataset);
- the percentage of correct localization, in the case of topological localization, or the average localization error, both represented by L, in the case of metric localization (average of at least ten trials, measured in percentage or meters respectively);
- the resolution R used in the image processing module, measured using the total number of pixels of the image (lowest resolutions are considered better in robotics due to the purpose of needing low cost commercial robotics systems); The evaluator E results in the formula: $E = A + L / S + R$

An extensive evaluation of this metric with respect to different approach to vision based robot localization can be found in [11].

BENCHMARKING PROCEDURE - Here we present results obtained using the collected dataset and measured using the metric presented above.

The purpose is not only to present good localization results, but also to present a good way of performance discussion and comparison.

Localization performances are evaluated using Monte Carlo Localization (MCL): the estimated robot pose is the result of the probabilistic particle filter applied using a vision similarity based sensor model and a classic motion model.

A comparison, using the proposed evaluator, among two different methods to image similarity measure using SIFT and SURF is also presented.

We tested our work changing number of particles in input, and the combination Path-Dataset. Other simulations were useful to calibrate other algorithm parameters.

We want to give a roadmap of all significant results in a global localization and traking experiment, consisting in:
- a measure of the localization error between the estimated position and the ground trout, expressed in meters, reporting the mean error and its standard deviation evaluated from the third step of each simulation until the end of the experiments;
- the number of particle used in the global localization and in the tracking phase; a graph of the trend of the particle number if necessary;
- the trend of a kidnapping experiment.
All results must be the summary of several trials and be statistical significant.
We do not want the computational time is reported in this phase, but a particular benchmark can be presented for this evaluation.
Using this ideas we present our results as follow. The number of particles used were set to 500 or 300, and the algorithm was repeated five times for each simulation. In Fig.3 we report also the trend on the particles number during the localization process.
The main objective of this paper remains the possibility to evaluate a benchmark quality value for each of the proposed approach. Table 1 summarizes the performances about some of the above described experiments according to the proposed evaluator. It shows the mean error and its standard deviation evaluated from the third step of each simulation.

Environment	Nparticles	Vision Algorithm	μ	σ	E
Outdoor Fermo	500	SIFT	4,19 m	0,16 m	0.3063
Outdoor Fermo	500	SURF	4,95 m	0,65 m	0.4003
Outdoor Fermo	300	SIFT	5,13 m	0,24 m	0.4163
Outdoor Fermo	300	SURF	6,90 m	0,49 m	0.4603

Table 1 Numerical results of outdoor simulation, comparing SIFT and SURF and using the performance metric E

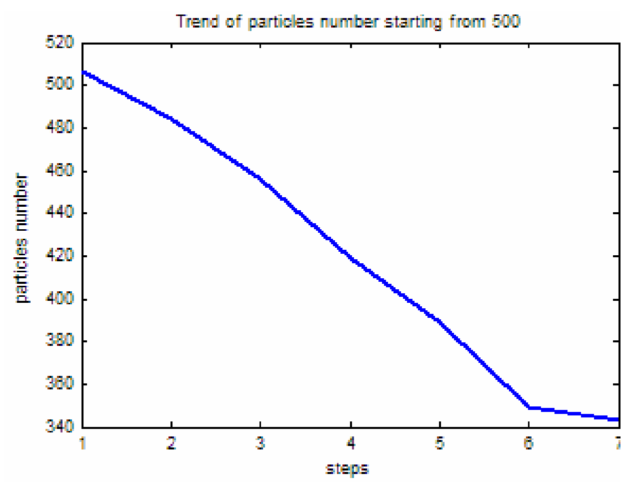

Figure 3 Trend of particles number in outdoor simulation

A good result was obtained for a *kidnapping* happening. To simulate it, we firstly linked path 2 (5 steps) with path 1(7 steps), reported in fig.2; 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.

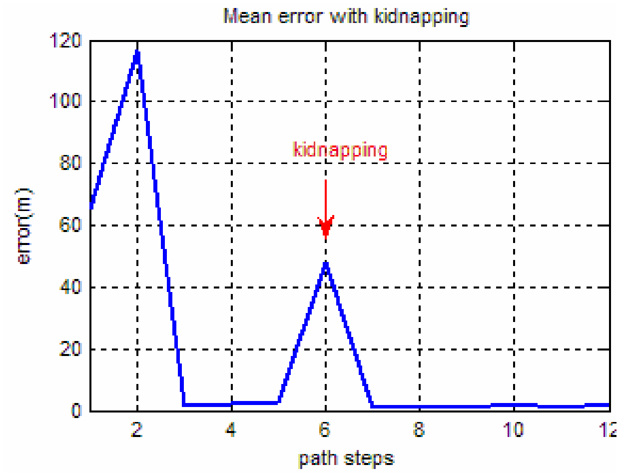

Figure 4 Results of mean error with a kidnapping simulation

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.

CONCLUSIONS - We have taken the initial steps in developing a new set of machine vision benchmarks in the areas of robot localization. The purpose of this paper is not only to present good localization results, but also to present a good way of performance discussion and comparison.
Next steps involve more careful delineation of the experimental protocol, selection of other specific problems, and the gathering of imagery and other test data. We welcome comments on this new benchmarking effort and in particular on the proposed performance metric measure. We also welcome datasets gathered in this way and made publicly available on the web.

SHORT BIOGRAPHY

[1] D. Lowe, “Distinctive image features from scale-invariant keypoints,”International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110,2004.

[2] Y. Ke and R. Sukthankar, “Pca-sift: A more distinctive representation for local image descriptors,” IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 506–513, 2004.

[3] H. Bay, T. Tuytelaars, and L. V. Gool, “Surf: Speeded up robust features,” in Ninth European Conference on Computer Vision, 2006.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] E. Frontoni, A. Ascani, A. Mancini and P. Zingaretti, Performance metric for vision based robot localization, Robotics Science and Systems 2008, Workshop on Good Experimental Methodologies, ETH Zurich, June 2008.