Visual Self-Localization for Small Mobile Robots with Weighted Gradient Orientation Histograms

Marius Hofmeister, Maria Liebsch and Andreas Zell

Department of Computer Science, University of Tübingen, Germany
E-mail: {marius.hofmeister, andreas.zell}@uni-tuebingen.de

Abstract: Research on small mobile robots is challenging due to the low computational power and limited sensing of the robots. In this paper, we present a method that enables these types of robots to localize themselves visually in indoor environments. Our approach uses a compass to cope with the restricted visual content that a low-resolution image can provide. Therefore, in the localization phase the robot orients itself towards a given direction and uses global image features to determine its position. Also, the robot's rotation impreciseness is included in the way the mapping is done. By real-world experiments we show that our method works despite of the restricted processing capabilities and the low resolution of the images.

1. INTRODUCTION

Small mobile robots in the dimension of around 15 cm have shown to be useful in many cases. A variety of different designs has been developed: Mondada et al. [1993] presented the well-known *Khepera* robot, which has been used in various scenarios, e.g., to simulate biologically inspired behaviour in a group of robots (Martinoli and Mondada [1995]). Further notable are for instance the *e-puck*, which is intended to be used for educational purposes (Cianci et al. [2007]), and the *Swarm-Bot* for the self-organization and self-assembling of multiple robots (Mondada et al. [2003]). In mobile sensor networks, inexpensive small robots, which have the ability to autonomously change their position, will be able to play an important role in the future.

In this paper, we use one of 13 *c't-bots* [1] of our lab to realize the experiments. Since our future plans are to build up a swarm of robots, we had to decide in favor of small and inexpensive robots. This implies the challenging part of the work: we have to cope with poor odometry, highly inaccurate motion and restricted computational power. As a result, our method works also on robots that are further miniaturized, provided that they possess a similar camera resolution and comparable processing capabilities.

Self-localization is a fundamental problem for mobile robots and has been studied by researchers now for many years. A variety of sensors has been examined to solve this task, ranging from laser range finders to GPS sensors. In this paper we consider the problem of vision-based self-localization since cameras are cheap and flexible sensors. The existing approaches to visual positioning often differ in the features they use to match images. On the one hand, there are local image features, like the well-known *Scale-Invariant Feature Transform* (SIFT, by Lowe [2004]), which describe only patches around points of interest in an image. As the number of interest points in an image is usually large, extracting features and matching them is very computation-intensive and therefore not feasible in our context. Examples of vision-based localization with local image features were presented by Barfoot [2005], Se et al. [2001], Wolf et al. [2005] and e.g., Tamimi and Zell [2005].

On the other hand, there are global image features, which describe an image as one single fixed-length feature vector. Usually, global features are more sensitive to illumination changes than local features. However, the global methods are faster, because matching two images by comparing two vectors is very efficient. Common global features for robot localization are color histograms (Ulrich and Nourbakhsh [2000]) or Global Integral Invariants (Weiss et al. [2007a]). Bradley et al. [2005] further presented Weighted Gradient Orientation Histograms (WGOH). Tamimi [2006] gave an overview of local and global techniques for vision-based robot localization.

One problem arises with the use of global image features in combination with a camera that is not omnidirectional: they are sensitive to translations. This comes from the fact that the whole image information is included in the feature vector. This means that, if the robot is not standing at the same position and is not exactly oriented towards the same direction, the content of an image changes considerably and thus also the feature vector. Furthermore, the problem gets even worse when using low-resolution images indoor, where the distance between the camera and the objects is smaller than in an outdoor environment. This makes the content of an image change a lot easier.

In this paper, we propose a method that enables small robots to localize themselves with global image features at a reasonable accuracy. The reason for using global image features arises from the low computational power that small robots often only have, making it difficult to compute and match local features.

[1] http://www.ct-bot.de

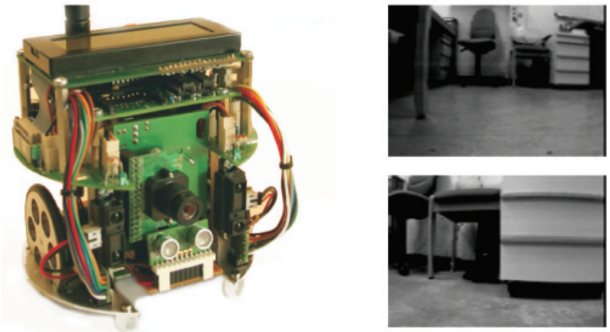


Fig. 1. c't-bot and example images by the on-board camera.

This paper is organized as follows. In the next Section we describe our method including a description of the robot we used and the employed type of image features. In Section 3, we present our real-world results and in Section 4, we finally draw a conclusion and present ideas for future work.

2. APPROACH

2.1 Overview

Our method works as follows: in the mapping phase, images have been grabbed on reference points in different specified directions. For localization, the robot rotates towards on of these known directions. It grabs an image, computes the feature vector and matches it to all feature vectors that have been grabbed in the same direction before. The vector with the highest similarity is assumed to represent best the current position of the robot. This method is feasible despite of the poor odometry and inaccuracy of the robot's drive because of the use of a compass module which provides reliable information on the robot's orientation. Due to magnetic deflections of furnishings etc. the orientation which was indicated by the compass was not always true, but repeatable. Thus it could be seen as a function of the position.

2.2 Robot

The *c't-bots*[1], which were developed by the German computer magazine *c't*, have a diameter of 12 cm and are 19 cm high (see Fig. 1). Their sensor system consists of two Sharp GP2D12 infrared distance sensors, an optical mouse sensor for motion estimation and infrared reflex light barriers for the detection of steps and lines. The wheel encoders provide a resolution of 60 counts per wheel. A Devantech SRF08 ultrasonic sensor is used for more accurate distance measurements than the infrared sensors can provide.

Additionally, our robot is equipped with a *POB-Eye* color camera. This camera has an image processing module which allows to compute all the image processing directly on it and to send the extracted image features via I2C bus to the robot. The camera provides a resolution of 120×88 pixels and possesses an ARM7TDMI processor with 60 MHz and 64 KB RAM. The angle of view of the used 2.5 mm lens is 64 ° horizontally and 41 ° vertically. Furthermore, the robot has a WLAN interface to send data to a PC. To store the image features, our robot contains a low-cost SD card. The "heart" of the robot is an ATmega644 microprocessor with 64 KB flash program memory, 16 MHz clock frequency and 4 KB SRAM. For localization, we also use a Devantech CMPS03 compass with a declared accuracy of 3-4 °.

Our algorithm runs on the robot itself. The WLAN interface is only used for debugging and monitoring purposes.

2.3 Image Feature Extraction

The Weighted Gradient Orientation Histograms (WGOH) were presented by Bradley et al. [2005], originally for outdoor environments. WGOH were inspired by SIFT features (Lowe [2004]) and are similar to features presented by Kosecka and Li [2004].

Bradley et al. first split the image into a 4×4 grid of subimages. On each subimage, they calculate an 8-bin histogram of gradient orientations, weighted by the magnitude of the gradient at each point and by the distance to the center of the subimage (see Fig. 2). In our implementation of WGOH, we use a 2D Gaussian for weighting, where the mean corresponds to the center of the subimage and the standard deviation corresponds to half the width and the height of the subimage, respectively (Weiss et al. [2007b]). We took these parameters from SIFT, where a Gaussian with half the width of the descriptor window is used for weighting. The 16 histograms are concatenated to a 1×128 feature vector, which is normalized subsequently. To reduce the dependency on particular regions or some strong gradients, the elements of the feature vector are limited to 0.2, and the feature vector is renormalized.

We decided to use global image features, in which a single feature is extracted from the whole image, instead of local features. This decision is based on the fact that our computational power is restricted. We chose WGOH features, because they are fast to compute and also have the advantage to be robust to illumination changes.

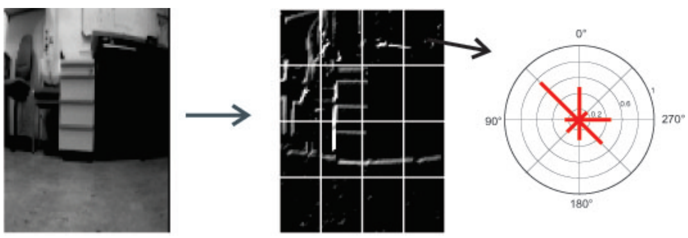


Fig. 2. WGOH computation steps.

2.4 Image Matching

To calculate the similarity between two images Q and D, we compare their feature histograms q and d using normalized histogram intersection

$$\bigcap_{norm}(q,d) = \frac{\sum_{k\in\{0,1,\ldots,m-1\}} \min(q_k, d_k)}{\sum_{k\in\{0,1,\ldots,m-1\}} q_k}, \qquad (1)$$

where m is the number of histogram bins. The position of the image with the highest similarity is then chosen to be the current position of the robot.

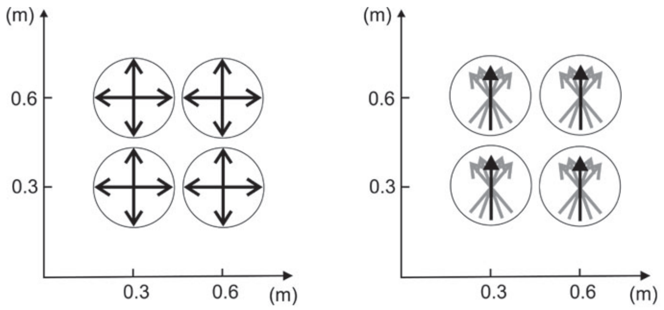


Fig. 3. Mapping directions of experiments 1 and 2.

3. EXPERIMENTAL RESULTS

Our experiments were held in an office environment. Since the robot did not have the ability to determine its ground truth position through GPS or other accurate sensors like laser scanners, we manually took measurements every 0.3 m in an area of 3 m × 1.8 m. We defined *correct matches* as the ones where the test image was matched successfully to one of the four nearest training images, similarly to a match within 0.43 m. A more detailed information about the preciseness of our method gives the mean localization error in Fig.4 and Fig.5.

Experiment 1 In the first experiment, we took 4 images at each position in the directions north, west, east, and south (see Fig. 3). This corresponds to a total of 240 images at 60 positions as training data. Then, we put the robot at arbitrary positions and oriented it by hand exactly towards one of the above mentioned directions. In this way, we recorded 33 images as test data.

Then, we matched the images of the test data against all images of the training data. As a result, we got correct matches in 29 out of 33 images, which corresponds to a localization rate of 88 %. As mentioned before, in the first part of this experiment we put the robot by hand towards the specified directions. However, since our aim was to let the robot move and localize itself autonomously, we also had to consider the properties of its motion model, especially its ability to rotate towards a specified direction. Thus we teleoperated the robot approximately 0.3 m through the area and, if it was not oriented in a range of $\pm 20\,^\circ$ to one of the directions north, west, east or south, we made it rotate autonomously towards one of these directions, allowing an orientation error of $\pm 20\,^\circ$. We determined the orientation by means of the compass. Then, we grabbed a test image for localization and rotated the robot back to continue its path.

Because of the imprecise rotations of our robot we got correct matches in only 6 of 27 images, which corresponds to a localization rate of 22 %. It turned out that, if the robot was exactly oriented towards the directions in which the training data had been recorded, the localization rate was high. But, if not precisely oriented to one of the mentioned directions, the robot was not able to localize itself properly. Fig. 4 gives an overview of the results.

Experiment 2 This led to our second experiment. First, we measured the rotation standard deviation of the robot's motion model for a rotation of $180\,^\circ$ and came to a value of $\sigma_{rot} = 13.38\,^\circ$. To overcome these imprecise rotations, we mapped five images at each reference point in the room: one in the direction north and the others with a rotation difference of $10\,^\circ$ (see Fig. 3). Altogether, we grabbed 300 images at 60 measuring points as training data. Then, we acquired test data from two different rounds in the mapped area (see Fig. 6). Therefore, we made the robot move approximately 0.3 m in an arbitrary direction. Then, if the robot was not oriented in the range of $\pm 20\,^\circ$ towards north, we rotated it autonomously to grab an image and after this rotated it back to continue its path. Note that we teleoperated the robot to simulate autonomous movements and did not move the robot by hand. We even used the compass for this experiment.

	matches	loc. rate	MLE±SD
Oriented exactly	29/33	88 %	0.21±0.14
Oriented imprecisely	6/27	22 %	0.84±0.61

Fig. 4. Results of experiment 1. Shown are correct matches, localization rates, mean localization errors (MLE) and standard deviations (SD) in meters.

The test data of round 1 was grabbed directly after the mapping process, while the test data of round 2 was grabbed in a slightly changed environment to test the robustness of the method.

In round 1 we got 24 correctly matched images out of 27 images. This corresponds to a localization rate of 89 %. In round 2 we got 21 correctly matched images out of 29 images, which corresponds to a localization rate of 72 %. This is a lower rate as in round 1, but if we take a look at the results in detail, we discover that only 3 test images are matched totally wrong while the rest (90 %), even though not correctly matched to one of the four nearest training images, have a maximum localization error of 0.45 m. See also the mean localization error in Fig. 5.

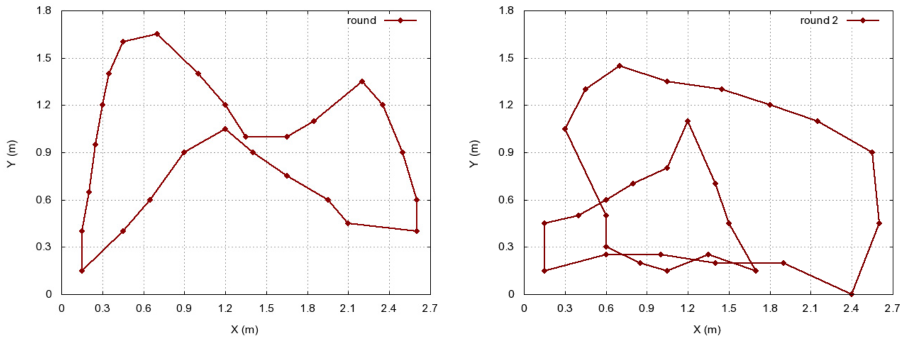
Generally, we can conclude that, if the area is mapped with respect to the rotation impreciseness of the robot, an autonomous localization is possible.

The computation of a WGOH feature vector from one image took 0.87 s on the robot camera. The comparison of one feature vector with 100 vectors which had been saved on the SD card took 6.9 s, with 200 vectors 11.0 s and with 300 vectors 15.0 s. A particle filter could help to speed up this matching process by reducing the number of images which have to be matched.

	matches	loc. rate	MLE±SD
Round 1	24/27	89 %	0.21±0.16
Round 2	21/29	72 %	0.32±0.21

Fig. 5. Results of experiment 2. Shown are correct matches, localization rates, mean localization errors (MLE) and standard deviations (SD) in meters.

Repeatability Test To test whether the results of our localization method are statistically significant, we performed cross-validation. Therefore, we divided the 300 training images of *experiment 2* into five folds. Each fold consisted of 60 images which were now declared as test data. Remember that the images had been mapped as



(a) (b)

Fig. 6. True robot paths of experiment 1 (a) and experiment 2 (b).

shown in Fig. 3. All images were taken from different positions such that no position appeared more than once in each test data. The orientation of the images was chosen randomly. In this way, we got five independent random datasets as samples, each consisting of 60 test images and the 240 remaining images, which served now as training data. Then we matched the test data against the training data. The mean localization errors of the five datasets are shown in Fig.7. These results are significant with 95 % confidence interval from 0.29 m to 0.33 m.

Dataset	MLE
1	0.3010
2	0.3197
3	0.3186
4	0.3381
5	0.2936

Fig. 7. Mean localization errors (MLE) in meters of the different datasets.

4. CONCLUSION AND FUTURE WORK

In this paper we presented a method for small mobile robots to localize themselves autonomously in indoor environments. We used global image features and low-resolution images. A prerequisite is that the rotation impreciseness of the robot is included into the way the mapping of the environment is done. Our real-world experiments verify that the proposed solution works at a good accuracy.

Clearly, the drawback of our method is the relatively high number of images one has to map in a comparatively small area. But we also have to keep in mind the small size of the robots and their limitations.

Our next steps will be to evaluate and compare different image features for the use on small mobile robots. We will also take into account the robustness to illumination and translation changes.

ACKNOWLEDGEMENTS

The first author would like to acknowledge the financial support by the Friedrich-Ebert-Stiftung (FES) of his Ph.D. scholarship at the University of Tübingen.

REFERENCES

T. D. Barfoot. Online visual motion estimation using Fast-SLAM with SIFT features. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 579–585, Edmonton, Canada, 2005.

D. M. Bradley, R. Patel, N. Vandapel, and S. M. Thayer. Real-time image-based topological localization in large outdoor environments. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3670–3677, Edmonton, Canada, 2005.

C. Cianci, X. Raemy, J. Pugh, and A. Martinoli. Communication in a swarm of miniature robots: The e-puck as an educational tool for swarm robotics. In *Proceedings of Simulation of Adaptive Behavior (SAB), Swarm Robotics Workshop*, pages 103–115, 2007.

J. Kosecka and F. Li. Vision based topological markov localization. In *Proceedings of International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1481–1486, New Orleans, LA, USA, 2004.

D. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 60, pages 91–110, 2004.

A. Martinoli and F. Mondada. Collective and cooperative group behaviours: Biologically inspired experiments in robotics. In *Proceedings of the Fourth International Symposium on Experimental Robotics (ISER)*, pages 3–10, Stanford, USA, June 1995.

F. Mondada, E. Franzi, and P. Ienne. Mobile robot miniaturisation: a tool for investigation in control algorithms. In *Proceedings of the 3rd International Symposium on Experimental Robotics*, pages 501–513, Kyoto, Japan, October 1993.

F. Mondada, A. Guignard, M. Bonani, D. Bar, M. Lauria, and D. Floreano. SWARM-BOT: From concept to implementation. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 2, pages 1626–1631, 2003.

S. Se, D. Lowe, and J. Little. Local and global localization for mobile robots using visual landmarks. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 1, pages 414–420, 2001.

H. Tamimi. *Vision-based Features for Mobile Robot Localization*. PhD thesis, University of Tübingen. Tübingen, Germany, 2006.

H. Tamimi and A. Zell. Global robot localization using iterative scale invariant feature transform. In *Proceedings of the 36th International Symposium on Robotics (ISR)*, Tokyo, Japan, 2005.

I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1023–1029, San Francisco, CA, 2000.

C. Weiss, A. Masselli, H. Tamimi, and A. Zell. Fast outdoor robot localization using integral invariants. In *Proceedings of the 5th International Conference on Computer Vision Systems (ICVS)*, Bielefeld, Germany, March 2007a.

C. Weiss, A. Masselli, and A. Zell. Fast vision-based localization for outdoor robots using a combination of global image features. In *Proceedings of the 6th Symposium on Intelligent Autonomous Vehicles (IAV)*, Toulouse, France, September 2007b.

J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization by combining an image retrieval system with monte carlo localization. *IEEE Transactions on Robotics*, 21(2):208–216, 2005.