Fast Outdoor Robot Localization Using Integral Invariants

Christian Weiss[1], Andreas Masselli[1], Hashem Tamimi[2], and Andreas Zell[1]

[1] Department of Computer Science, University of Tübingen,
Sand 1, 72076 Tübingen, Germany
{c.weiss, andreas.zell}@uni-tuebingen.de, andreas.masselli@web.de,
[2] College of Administrative Science and Informatics,
Palestine Polytechnic University, P. O. Box 198, Hebron, Palestine
hashem.tamimi@uni-tuebingen.de

Abstract. Global Integral Invariant Features have shown to be useful for robot localization in indoor environments. In this paper, we present a method that uses integral invariants for outdoor environments. To make the integral invariant features more distinctive for outdoor images, we split the image into a grid of subimages and calculate integral invariants for each grid cell individually. We then concatenate the results to get the feature vector for the image. Additionally, we combine this method with a particle filter to improve the localization results. We compare our approach to a Scale Invariant Feature Transform (SIFT)-based approach on images of two outdoor areas under different illumination conditions. The results show that the SIFT approach is more exact, but the integral invariant approach is faster and allows localization in significantly less than one second.

Key words: Outdoor robot localization, Integral Invariants, SIFT

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 environment. GPS can give valuable position information, but often the GPS satellites are occluded by buildings or trees.

Because of these problems, vision has become the most widely used sensor in outdoor localization. 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 that visual features may not be distinguishable enough; in a forest, every tree looks about the same.

An algorithm which can deal with changing illumination to a certain extent is the *Scale Invariant Feature Transform* (SIFT) developed by Lowe [1]. SIFT is a feature-based method which computes descriptors for local interest points. The local features are more dependent on structure than on illumination and are very

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

distinctive. However, as the number of features per image is large (about 420 for our 320×240 images on the average), matching images is very time-consuming. Approaches that use SIFT for indoor localization are for example [2, 3]. Outdoor localization using SIFT was presented, for example, in [4]. There also exist methods that replace the gradient histogram features of the SIFT approach, for example by *Local Integral Invariants* [5]. Bradley *et al.* use a technique inspired by SIFT for outdoor localization, the so-called *Weighted Gradient Orientation Histograms* (WGOH) [6]. They subdivide the image into a grid of subimages. For each subimage, they compute an 8-bin histogram of image gradients. The subdivision makes the method robust to partial changes in the image.

Another group of vision-based localization methods are the appearance-based methods, which compute global features for images. Well-known methods for indoor localization are based on PCA [7,8] or on *Integral Invariant Features* [9,10]. These methods are more sensitive to illumination changes than the local methods, but there exist some approaches which try to make PCA-based methods illumination invariant [11]. The main advantage of these methods over local techniques is their higher speed.

Artač *et al.* combine a global technique using data tensors describing the image and a local SIFT technique for outdoor localization [12]. They first use the global technique to fastly get a set of training images which are similar to the query image. Only these database images are then used for SIFT. An overview of global and local features for mobile robot localization can be found in [13].

The outdoor localization method presented in this paper is based on integral invariants. We modified the global integral invariant approach, because the global integral invariants are not distinctive enough for our outdoor datasets. Similar to Bradley *et al.* [6], we split each image into a grid of subimages. We calculate an 8×8 histogram of integral invariants on each subimage using two relational kernels. Then we concatenate these histograms to get a global feature for each image, which we call *Weighted Grid Integral Invariant Feature* (WGII). To improve the localization, we combine the weighted grid integral invariants with a particle filter. Each image-to-image comparison is very fast, so localization is possible more than once per second. In experiments on outdoor images, we compare our approach to an accelerated SIFT approach.

The rest of the paper is organized as follows. In Section 2, we describe the global integral invariants and our modification to weighted grid integral invariants. We also shortly describe the SIFT approach we compare our method to. Section 3 explains the the particle filter and Section 4 presents results on outdoor datasets. Finally, Section 5 concludes the paper and suggests future work.

2 Image Feature Extraction

This section first describes how global integral invariant features are calculated for an image and how we modified the feature extraction procedure. Additionally, we shortly describe the SIFT approach that we use for comparison.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

2.1 Global Integral Invariant Features

Global integral invariant features are features which are invariant to euclidean motion, i.e. rotation and translation, and to some extent robust to illumination changes. The key idea is to apply all possible translations and rotations to the image and to calculate the features by averaging over all the transformed versions of the image. The approach compares images using their global features, i.e. features representing the whole image. These features can be a single number, or a histogram of local features evaluated at each pixel. By an appropriate choice of the kernel function, integral invariants can also be made robust to local transformations, motion of individual objects and object deformation [9].

Let $\mathbf{I} = \{\mathbf{I}(x_0, x_1), 0 \leq x_0 < N_0, 0 \leq x_1 < N_1\}$ be a grayscale image of size $N_0 \times N_1$. $\mathbf{I}(i, j)$ represents the intensity at the pixel coordinate (i, j). This image is transformed by elements g of a transformation group G, where G is the group of euclidean motions. Thus an image $\mathbf{I}$ is transformed according to

$$(g\mathbf{I})(i, j) = \mathbf{I}(k, l), \tag{1}$$

where

$$\begin{pmatrix} k \\ l \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} i \\ j \end{pmatrix} - \begin{pmatrix} t_0 \\ t_1 \end{pmatrix}. \tag{2}$$

ϕ specifies the rotation angle, t_0 and t_1 specify the translation.

To obtain features $F(\mathbf{I})$ which are invariant to euclidean motion of the images, one must integrate over the transformation group G, given a kernel function $f(\mathbf{I})$:

$$F(\mathbf{I}) = \frac{1}{RN_0N_1} \sum_{t_0=0}^{N_0-1} \sum_{t_1=0}^{N_1-1} \sum_{r=0}^{R-1} f\left(g\left(t_0, t_1, \phi = 2\pi\frac{r}{R}\right)\mathbf{I}\right). \tag{3}$$

Eq. 3 calculates a single number as feature for an image. A more distictive feature for an image is a histogram of local features. The rotations are performed at each pixel of the image individually and the histogram is formed using the values for each pixel.

There are different possible choices for the kernel function f. Well known functions are the *monomial* and *relational* kernel. Both involve a local neighborhood of the pixel in the calculation of the feature. For our purposes, we found the relational kernel to work better, because it seems to be more robust to illumination changes. For two pixel coordinates $p_1 = (x_1, y_1)$ and $p_2 = (x_2, y_2)$, the relational kernel is calculated by

$$f(\mathbf{I}) = rel\left(\mathbf{I}(x_1, y_1) - \mathbf{I}(x_2, y_2)\right) \tag{4}$$

with the ramp function

$$rel(\gamma) = \begin{cases} 1 & \text{if } \gamma < -\epsilon, \\ \frac{\epsilon-\gamma}{2\epsilon} & \text{if } -\epsilon \leq \gamma \leq \epsilon, \\ 0 & \text{if } \epsilon < \gamma. \end{cases} \tag{5}$$

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

Example choices for p_1 and p_2 are $p_1 = (2, 0)$ and $p_2 = (0, 4)$. This means that values lying on two circles with radius 2 and 4 around a pixel and a phase shift of 90° are used for the calculation of the kernel. It is also possible to use more than one kernel and to form a multidimensional histogram as feature for an image.

To calculate the similarity between a query image $\mathbf{Q}$ and a database image $\mathbf{D}$, we compare their feature histograms $\mathbf{q}$ and $\mathbf{d}$ using normalized histogram intersection

$$\bigcap_{\mathbf{norm}} (\mathbf{q}, \mathbf{d}) = \frac{\sum_{k\in\{0,1,\ldots,m-1\}} \min(q_k, d_k)}{\sum_{k\in\{0,1,\ldots,m-1\}} q_k}, \tag{6}$$

where m is the number of histogram bins.

2.2 Weighted Grid Integral Invariant Features

In our experiments, we found that ordinary global integral invariant features are not distictive enough for outdoor localization, even when using multidimensional histograms formed using three kernels. Thus, we adopted the idea of Bradley *et al.* [6], who split the image into a grid of subimages and calculate gradient histograms for each subimage. Additionally, they use a weighting such that pixels near the center of a subimage get a higher weight than pixels near the borders of subimages, because the pixels near the borders are more likely to fall into another subimage under image translations or rotations.

In our case, we first compute the integral invariants for each pixel of the image using two kernels. We then split the image into a 4×4 grid of subimages. On each subimage, we calculate a weighted two-dimensional histogram of integral invariant features. We weight the integral invariants by a 2D Gaussian with mean at the center of the subimage and standard deviations of 0.25 times the width and the height of the subimage, respectively. After that, we concatenate the resulting histograms to get the final feature vector for the image.

To calculate the histograms, we use two relational kernels. The first kernel uses the pixel coordinates $p_1 = (6, 0)$ and $p_2 = (0, 9)$, the second kernel uses $p_1 = (10, 0)$ and $p_2 = (0, 20)$. The parameter ϵ is 0.098 for both kernels, and the number of rotations is 10. For each subimage, we compute an 8×8 histogram. As an image has 16 subimages, the final feature vector for each image is of size 1024×1. We chose these values of the parameters because experimentally, they lead to the best results.

2.3 SIFT

For comparison to our method, we use a localization approach based on the *Scale Invariant Feature Transform* (SIFT) [1]. In this approach, the most similar training image to a test image is the one which contains the highest number of features that can be matched to the features of the test image.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

To speed up the SIFT-based localization, we reduce the number of features of each image. The idea is to delete "noisy" SIFT-features, which are likely not to appear in more than one image. To reduce the number of features of the training images, we match each training image to the two neighboring training images. We only keep the features that can be matched to a feature of at least one of the two neighboring images. In the localization phase, the only neighbor of the current test image is the test image that was taken directly before the current image. Thus for each test image, we only keep the features that can be matched to a feature of the previous image.

Depending on the dataset, this technique reduces the number of features per image to about 20% to 50% of the original number of features. The reduction is more significant for images showing vegetation like grass, bushes and trees than for images mainly showing artificial objects like buildings, cars and roads. Due to the smaller number of features, matching images is accelerated by a mean factor of about 5 without loss of accuracy.

3 Combination with the Particle Filter

For localization, we combine the weighted grid integral invariants with a particle filter [14]. To get a good comparison to the SIFT approach, we also use a particle filter for the SIFT localization. Thus, if not stated differently, all following explanations hold for both the WGIIs and SIFT.

Particle filters represent the belief $Bel(x)$ of the robot about its position by a set of m particles. Each particle consists of a pose (x, y) together with a non-negative *importance factor*, which determines the weight of each particle. The estimated pose of the robot is given by the weighted mean of the particles. For global robot localization, the inital belief is approximated by particles which are randomly distributed over the robot's universe. All importance factors are set to $\frac{1}{m}$. The particles are updated for each test image using 3 steps:

1. Draw m random particles $x_{t-1}^{(i)}$ from $Bel(x_{t-1})$ according to the importance factors p_{t-1} at time $t-1$.
2. Update the sample $x_{t-1}^{(i)}$ to sample $x_t^{(j)}$ according to an action u_{t-1}. As we do not use a motion model, for example from odometry, we randomly update the particle according to a gaussian distribution with a standard deviation of 4 m. Additionally, we move each sample in the direction to the nearest training image. The distance we move the particle is 0.2 times the distance of the particle to the nearest training image.
3. Weight the sample $x_t^{(j)}$ by the importance factor $p(y_t|x_t^{(j)})$, i.e. the likelihood of the sample $x_t^{(j)}$ given the measurement y_t. In our method, we first search the nearest training image to each particle. The test and the training image are then matched using WGIIs or SIFT, and the score of the match becomes the new weight of the particle. We additionally multiply the new weight with a factor that decreases with the distance of the particle to its nearest training image. In the case of integral invariants, we potentiate the new weight by 20,

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

because the differences between weights are all low (but still distinctive at this low level). This way the difference between the weights becomes clearer.

After the third step, we normalize the importance factors and calculate the estimated position. Before repeating the three steps for the next test image, we delete the worst 5% of the particles. For these particles, we randomly insert new ones with importance factors $\frac{1}{m}$. After that, we renormalize all importance factors. The random insertion of new particles ensures that the robot can fastly recover its position if the position was lost.

To speed up the calculation of the weights, we save for each particle the matching result between the test image and the nearest training image to the particle. If another particle has the same nearest training image, we do not have to recalculate the match. In the case of SIFT, this method speeds up the estimation of a new position by a factor of about 5. For the weighted grid integral invariants, we only get a slight speedup.

4 Experimental Results

In our experiments, we use images collected by our RWI ATRV-JR outdoor robot. We took one 320×240 pixel grayscale image per second with the left camera of the Videre Design SVS stereo camera system mounted on top of the robot. As we used a constant velocity of about 0.6 m/s, the positions of subsequent images are about 0.6 meters away from each other. The robot is also equipped with a differential GPS (DGPS) 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. As we know that we moved the robot on a smooth trajectory, we corrected some wrong GPS values manually. As we also always used a constant velocity, we closed gaps by linearly interpolating between the positions before and after the gap.

In our experiments, we used two different datasets. Dataset 1 consists of six rounds around a big building. Each of the rounds is about 260 m long and contains about 400 images. We collected three of the rounds under sunny conditions. However, there are some short sections (about 5 to 10 seconds long) during which the sun was covered. Six weeks later, we collected the other three rounds on a cloudy day. The images of dataset 1 contain many artifical objects like buildings, streets and cars. Additionally, there are some dynamic objects like cars and people passing by. We also traversed a parking lot, where on the two days different cars were parked.

We acquired the images of dataset 2 in a different area mostly containing vegetation like grass, bushes and trees. We again collected different rounds under varying illumination conditions. We recorded two rounds in the early afternoon, in which the sun was shining brightly. In the evening, we collected the third and fourth round. It was cloudy and starting to get dark. There are also some dynamic objects in the images of the evening rounds, namely people playing

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

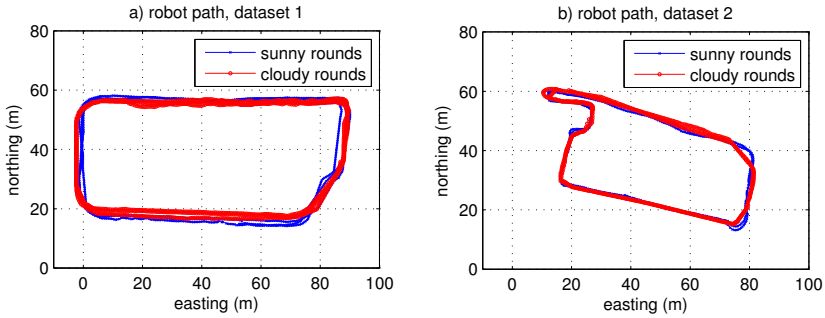

Fig. 1. GPS ground truth data. a) Six rounds around a big building (dataset 1). b) Four rounds on a meadow (dataset 2).

soccer and moving one of the goals around. Each round of dataset 2 is about 220 m long and consists of about 350 images. Fig. 1 shows the GPS ground truth data of dataset 1 and 2. Fig. 2 shows example images of dataset 1 and 2 under different illumination conditions.

In our experiment, we calculated the error of all possible training/test combinations of rounds. For each round of test images, we also repeated the experiment n times, where n is the number of test images. For each of these experiments, we used a different image as starting image for the localization. Then we calculated the mean error of all experiments that are similar, e.g. all experiments in which we used the sunny images of dataset 1 for training and the cloudy images of dataset 1 for testing. In all experiments, we used $m = 300$ particles.

Figure 3 a) and b) as well as Tab. 1 show the results of the experiments in which the training and test images have similar illumination. Figure 3 shows a mean curve for the two results of each dataset. In all experiments, the error decreases rapidly. The mean error for the WGIIs (2.75 m) is about 1.36 times higher than the mean error for the SIFT approach (2.02 m).

Table 1. Mean localization errors using the particle filter (in meters)

dataset	training images	test images	SIFT	WGII
dataset 1	sunny	sunny	2.15	3.53
	cloudy	cloudy	2.06	2.18
dataset 2	sunny	sunny	1.78	1.48
	cloudy	cloudy	2.10	3.79
dataset 1	sunny	cloudy	3.27	5.97
	cloudy	sunny	2.52	6.10
dataset 2	sunny	cloudy	2.88	3.82
	cloudy	sunny	2.74	3.66

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

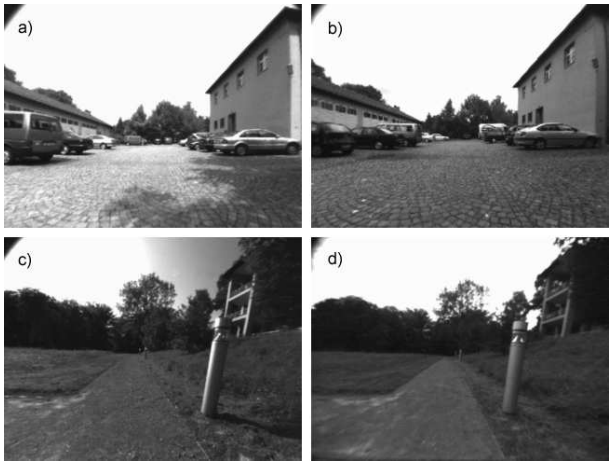

Fig. 2. Example images of dataset 1 and 2. a) dataset 1, sunny. b) dataset 1, cloudy. c) dataset 2, sunny. d) dataset 2, cloudy.

Figure 3 c) and d) as well as Tab. 1 present the results of the experiments in which the training and test images were taken under different illumination conditions. In average, the error using the weighted grid integral invariants is about 1.78 times higher than for images with similar illumination. Using SIFT, the error rises by a factor of 1.41. Thus, if we use SIFT as a reference for robustness against illumination changes, the weighted grid integral invariants also seem to be reasonably robust to illumination changes.

The experiments show that the SIFT approach generates more exact position estimates than the WGII approach. But even though we used a particle filter and the two techniques described in Sections 2.3 and 3 to accelerate the SIFT localization, it is still slow. On our robot, which is equipped with a 1.8 GHz Pentium M Processor and 1 GB of RAM, the SIFT feature extraction for one test image takes about 0.821 s, the SIFT match to the preceeding image needed for reducing the number of SIFT features takes about 0.104 s and one particle filter step takes about 0.771 s. This sums up to a total of 1.696 s for one test image (Fig. 4).

In contrast, our WGII approach runs in less than one second. On our robot, feature extraction for one test image needs 0.545 s. One particle filter step needs only 0.106 s, due to the much faster calculations of image similarities. In total, localization for one image only takes 0.651 s.

Thus, we can see that there is a trade-off between accuracy and computation time. The errors using the grid integral invariants approach are between about 1.4 and 2 times larger than using the SIFT approach. On the other hand, localization using the WGIIs is about 2.6 times faster than using SIFT.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

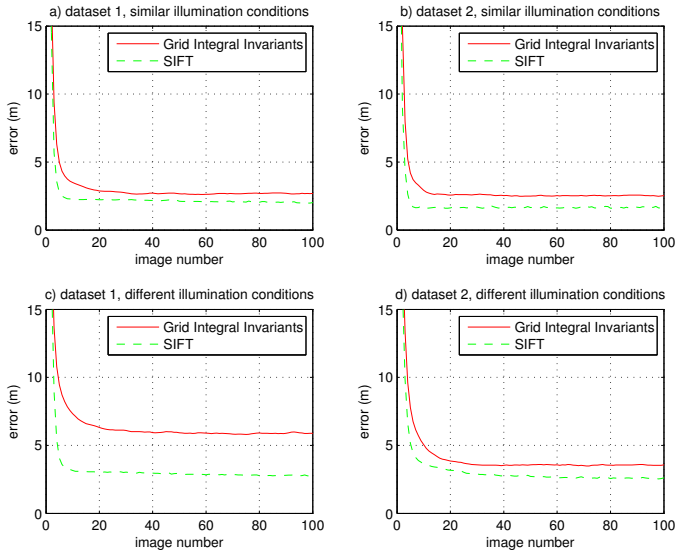

Fig. 3. Mean errors for particle filter experiments up to image 100. There is no significant change for the following images. The mean initial errors are about 36 m for dataset 1 and about 26 m for dataset 2.

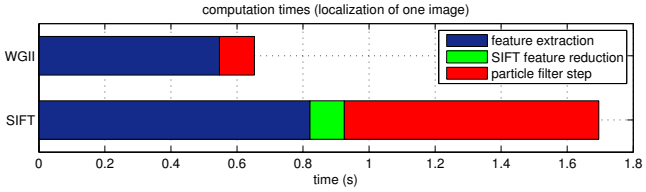

Fig. 4. Computation times for localization of one test image.

5 Conclusion

We presented a new method for outdoor mobile robot localization based on integral invariant features, which we call weighted grid integral invariants (WGII). We split the images into a 4×4 grid of subimages and calculate a multidimensional 8×8 histogram of integral invariant features for each subimage individually. The feature vector for the whole image is the concatenation of the histograms of the subimages. We also combined the WGIIs with a particle filter.

Experiments on image datasets of two different areas and under varying illumination show that localization using our approach is possible with mean errors between about 2 and 6 meters, depending on the dataset. A comparison to a SIFT-based approach showed that the SIFT approach is more exact. But in contrast to the SIFT approach, localization using our approach is possible in less than one second.

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de

In future work, we will examine combinations of the two approaches. In most situations, in which the robot is relatively sure about its position, the fast weighted grid integral invariant approach will give good position estimations. If the robot is not sure about its position, the more exact but slower SIFT approach can be used to give more reliable position estimates.

References

1. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. Intl. Journal of Computer Vision, vol. 60, no. 2 (2004) 91–110
2. Se, S., Lowe, D., Little, J.: Local and Global Localization for Mobile Robots Using Visual Landmarks. In Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS 2001), Maui, Hawaii (2001) 414–420
3. Tamimi, H., Zell, A.: Global Robot Localization using Iterative Scale Invariant Feature Transform. In 36th Intl. Symposium on Robotics (ISR 2005), Tokyo, Japan (2005)
4. Barfoot, T. D.: Online Visual Motion Estimation using FastSLAM with SIFT Features. In Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS 2005), Edmonton, Canada (2005) 3077–3082
5. Tamimi, H., Halawani A., Burkhardt H., Zell, A.: Appearance-based Localization of Mobile Robots using Local Integral Invariants. In Proc. of the 9th Intl. Conf. on Intelligent Autonomous Systems (IAS-9), Tokyo, Japan (2006) 181–188
6. Bradley, D. M., Patel, R., Vandapel, N., Thayer, S. M.: Real-Time Image-Based Topological Localization in Large Outdoor Environments. In Proc. of the IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS 2005), Edmonton, Canada (2005) 3062–3069
7. Jogan, M., Leonardis, A.: Robust Localization using an Omnidirectional Appearance-based Subspace Model of Environment. Robotics and Autonomous Systems, vol. 45, no. 1 (2003) 51–72
8. Jogan, M., Artač, M., Skocaj, D., Leonardis, A.: A Framework for Robust and Incremental Self-Localization. In Proc. of the 3rd Intl. Conf. on Computer Vision Systems (ICVS 2003), Graz, Austria (2003) 460–469
9. Siggelkow, S.: Feature Histograms for Content-Based Image Retrieval. Ph.D. Dissertation, Institute for Computer Science, University of Freiburg, Germany (2002)
10. Wolf, J., Burgard, W., Burkhardt, H.: Robust Vision-based Localization by Combining an Image Retrieval System with Monte Carlo Localization. IEEE Transactions on Robotics, vol. 21, no. 2 (2005) 208–216
11. Jogan, M., Leonardis, A., Wildenauer, H., Bischof, H.: Mobile Robot Localization under Varying Illumination. In Proc. of the Intl. Conf. on Pattern Recognition (ICPR 02), Quebec, Canada (2002) 741–744
12. Artač, M., Leonardis, A.: Outdoor Mobile Robot Localisation using Global and Local Features. In Proc. of the 9th Computer Vision Winter Workshop (CVWW), (2004) 175–184
13. Tamimi, H.: Vision-based Features for Mobile Robot Localization. Ph.D. Dissertation, Department of Computer Science, University of Tübingen, Germany (2006)
14. Thrun, S., Fox, D., Burgard, W., Dellaert, F.: Robust Monte Carlo Localization for Mobile Robots. Artificial Intelligence, vol. 128, no. 1-2 (2000) 99–141

Proceedings of the 5th International Conference on Computer Vision Systems (ICVS 2007)
Published in 2007 by Applied Computer Science Group, Bielefeld University, Germany, ISBN 978-3-00-020933-8
This document and other contributions archived and available at: http://biecoll.ub.uni-bielefeld.de