Appearance-based Localization of Mobile Robots using Local Integral Invariants

Hashem Tamimi [a], Alaa Halawani [b] Hans Burkhardt [b] and Andreas Zell [a]
[a] *Computer Science Dept., University of Tübingen, Sand 1, 72076 Tübingen, Germany.*
[b] *Institute for Computer Science, University of Freiburg, 79110 Freiburg, Germany.*

Abstract. In appearance-based localization, the robot environment is implicitly represented as a database of features derived from a set of images collected at known positions in a training phase. For localization the features of the image, observed by the robot, are compared with the features stored in the database. In this paper we propose the application of the integral invariants to the robot localization problem on a local basis. First, our approach detects a set of interest points in the image using a Difference of Gaussian (DoG)-based interest point detector. Then, it finds a set of local features based on the integral invariants around each of the interest points. These features are invariant to similarity transformation (translation, rotation, and scale). Our approach proves to lead to significant localization rates and outperforms a previous work.

Keywords. Appearance-based robot localization, integral invariants, (DoG)-based interest point detector.

1. Introduction

Vision-based robot localization demands image features with many properties. On one hand the features should exhibit invariance to scale and rotation as well as robustness against noise and changes in illumination. On the other hand they should be extracted very quickly so as not to hinder other tasks that the robot plans to perform. Both global and local features are used to solve the robot localization problem. Global features are extracted from the image as a whole, such as histograms [18], Principal Component Analysis (PCA)-based features [3], and integral invariants [17]. On the other hand, local features are computed from areas of high relevance in the image under consideration such as Scale Invariant Feature Transform (SIFT) [6], kernel PCA-based features [16], and wavelet-based features [15]. Local features are more commonly employed because they can be computed efficiently, are resistant to partial occlusion, and are relatively insensitive to changes in viewpoint. There are two considerations when using local features [5]: First, the interest points should be localized in position and scale. Interest points are positioned at local peaks in a scale-space search, and filtered to preserve only those that are likely to remain stable over transformations. Second, a signature of the interest point is built. This signature should be distinctive and invariant over transformations caused by changes in camera pose as well as illumination changes. While point localization and signature aspects of interest point algorithms are often designed together, they can be considered independently [7].

In this paper we propose the application of the integral invariants to the robot localization problem on a local basis. First, our approach detects a set of interest points in the image based on a Difference of Gaussian (DoG)-based interest point detector developed by Lowe [6]. Then, it finds a set of descriptive features based on the integral invariants around each of the interest points. These features are invariant to similarity transformation (translation, rotation, and scale). Our approach proves to lead to significant localization rates and outperforms a previous work that is described in Section 4.

2. Integral Invariants

Following is a brief description of the calculation of the rotation- and translation-invariant features based on integration. The idea of constructing invariant features is to apply a nonlinear kernel function $f(\mathbf{I})$ to a gray-valued image, $\mathbf{I}$, and to integrate the result over all possible rotations and translations (Haar integral over the Euclidean motion):

$$\mathbf{T}[f](\mathbf{I}) = \frac{1}{PMN} \sum_{n_0=0}^{M-1} \sum_{n_1=0}^{N-1} \sum_{p=0}^{P-1} f(g(n_0, n_1, p\frac{2\pi}{P})\mathbf{I}) \tag{1}$$

where $\mathbf{T}[f](\mathbf{I})$ is the invariant feature of the image, M, N are the dimensions of the image, and g is an element in the transformation group $\mathscr{G}$ (which consists here of rotations and translations). Bilinear interpolation is applied when the samples do not fall onto the image grid. The above equation suggests that invariant features are computed by applying a nonlinear function, f, on the neighborhood of each pixel in the image, then summing up all the results to get a single value representing the invariant feature. Using several different functions finally builds up a feature space. To preserve more local information we remove the summation over all translations. This results in a map $\mathbf{T}$ that has the same dimensions of $\mathbf{I}$:

$$(\mathbf{T}[f]\mathbf{I})(n_0, n_1) = \frac{1}{P} \sum_{p=0}^{P-1} f\left(g\left(n_0, n_1, p\frac{2\pi}{P}\right)\mathbf{I}\right) \tag{2}$$

Applying a set of different f's will result in a set of maps. A global multi-dimensional feature histogram is then constructed from the elements of these maps. The choice of the non-linear kernel function f can vary. For example, invariant features can be computed by applying the monomial kernel, which has the form:

$$f(\mathbf{I}) = \left(\prod_{p=0}^{P-1} \mathbf{I}\left(\mathrm{x_p}, \mathrm{y_p}\right)\right)^{\frac{1}{P}} \tag{3}$$

One disadvantage of this type of kernels is that it is sensitive to illumination changes. The work in [10] defines another kind of kernels that are robust to illumination changes. These kernels are called relational kernel functions and have the form:

$$f(\mathbf{I}) = rel\left(\mathbf{I}(\mathrm{x}_1, \mathrm{y}_1) - \mathbf{I}(\mathrm{x}_2, \mathrm{y}_2)\right) \tag{4}$$

with the ramp function

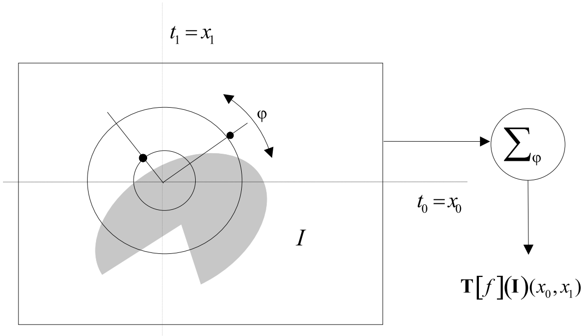

Figure 1. Calculation of $f = rel\,(\mathbf{I}(0,3) - \mathbf{I}(4,0))$, by applying Equation 2, involves applying the function *rel* to the difference between the grey scale value of the pixels that lie on the circumference of circle of radius 3 and pixels that lie on the circumference of another circle of radius 4 (taking into consideration a phase shift of $\frac{\pi}{2}$ between the corresponding points) and taking the average of the result [17].

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

centered at the origin and $0 < \varepsilon < 1$ is chosen by experiment. Global integral invariants have been successfully used for many applications such as content-based image retrieval [12], texture-classification [9], object recognition [8], and robot localization [17]. Figure 1 illustrates how these features are calculated. Please refer to [12] for detailed theory.

3. DoG-based Point Detector

The interest points, which are used in our work, were first proposed as a part of the Scale Invariant Feature Transform (SIFT) developed by Lowe [6]. These features have been widely used in the robot localization field [11] [14]. The advantage of this detector is its stability under similarity transformations, illumination changes and presence of noise.

The interest points are found as scale-space extrema located in the Difference of Gaussians (DoG) function, $D(x,y,\sigma)$, which can be computed from the difference of two nearby scaled images separated by a multiplicative factor k:

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * \mathbf{I}(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma) \end{aligned} \tag{6}$$

where $L(x,y,\sigma)$ defines the scale space of an image, built by convolving the image $\mathbf{I}(x,y)$ with the Gaussian kernel $G(x,y,\sigma)$. Points in the DoG function, which are local extrema in their own scale and one scale above and below are extracted as interest points. The interest points are then filtered for more stable matches, and more accurately localized to scale and subpixel image location using methods described in [2].

4. Using Global Integral Invariants For Robot Localization

In [17], integral invariants are used to extract global features for solving the robot localization problem by applying Equation 2 to each pixel (n_0, n_1) in the image $\mathbf{I}$. The calculation of the matrix $\mathbf{T}$ involves finding an invariant value around each pixel in the image which is time consuming. Instead of this, Monte-Carlo approximation is used to estimate the overall calculation [13]. This approximation involves applying the nonlinear kernel functions to a set of randomly chosen locations and directions rather than to all locations and directions.

Global features achieve robustness mainly because of their histogram nature. On the other hand, local features, extracted from areas of high relevance in the image under consideration, are more robust in situations where the objects in images are scaled or presented in different views [4]. Such situations are often encountered by the robot during its navigation. In the next sections we modify the global approach by applying the integral invariants locally around a set of interest points.

5. Extracting Local Integral Invariants

Unlike the existing approach, explained in Section 2, the features that we propose are not globally extracted; they are extracted only around a set of interest points. Our approach can be described in the following steps:

1. **Interest point detection:**
   The first stage is to apply the DoG-based detector to the image in order to identify potential interest points. The location and scale of each candidate point are determined and the interest points are selected based on measures of stability described in [6].
2. **Invariant features initial construction:** For each interest point located at (n_0, n_1) we determine the set of all points which lie on the circumference of a circle of radius r_1. We use bilinear interpolation for sub-pixel calculation. Another set of points that lie on a circumference of a circle of radius r_2 are determined in the same manner. Both circles have their origin at (n_0, n_1). To make the features invariant to scale changes, the radii are adapted linearly to the local scale of each interest point. This way the patch that is used for feature extraction always covers the same details of the image independent of the scale.
3. **Nonlinear kernel application:** A non-linear kernel function is applied to the values of the points of the two circles. Each point located at (x_0, y_0) on the circumference of the first circle is tackled with another point located at (x_1, y_1) on the circumference of the second circle, taking into consideration a phase shift θ between the corresponding points. This step is repeated together with step 2 for a set of V kernel functions f_i, $i = 1, 2, \cdots, V$. The kernels differ from each other by changing r_1, r_2 and θ. Finally we apply Equation 7 for each interest point located at (n_0, n_1).

$$F_i(n_0, n_1) = \frac{1}{P} \sum_{p=0}^{P-1} f_i \left(g \left(n_0, n_1, p \frac{2\pi}{P} \right) \mathbf{I} \right), \; i = 1, 2, \cdots, V. \tag{7}$$

   We end up with a V-dimensional feature vector, F, for each single interest point.

(a) Step (1) (b) Step (2) (c) Step (3) (d) Step (4)

(e) Step (5) (f) Step (6) (g) Step (7) (h) Step (8)

Figure 2. Eight different image samples that belong to the same reference location.

6. Experimental Results

In this section we present the experimental results of our local integral invariants compared with the global integral invariants reviewed in Section 4.

6.1. Setting up the Database

To simulate the robot localization we use a set of 264 gray scale images taken at 33 different reference locations. Each has a resolution of 320×240. In each reference location we apply the following scenario capturing an image after each step: (1) The robot stops. (2) It translates 50 cm to the right. (3) The pan-tilt unit rotates 20 degrees to the left. (4) The robot moves 50 cm ahead. (5) The pan-tilt unit rotates 40 degrees to the right. (6) The pan-tilt unit rotates 20 degrees to the left. (7) The pan-tilt unit moves five degrees up. (8) The pan-tilt unit moves 10 degrees down. Figure 2 includes eight sample images of one reference location.

The database is divided into two equal parts. 132 images are selected for training and 132 for testing. This partitioning is repeated 50 times with different combinations for training and testing images. The average localization rate is computed for each combination. We assume that optimal localization results are obtained when each input image from one of the reference locations matches another image that belongs the same reference location.

6.2. Global Integral Invariants

In order to compare our work with the work of [17] which involves calculating the global features, a set of 40×10^3 random samples is used for the Monte-Carlo approximation, which is also suggested in [13] for best performance. For each sample we apply a set of three different kernels. Both monomial and texture kernel functions are investigated for best localization accuracy using the above images. For each image a single

D-dimensional histogram is build with $D = 3$. Each dimension contains 8 bins which has experimentally led to best results. The histograms are compared using the $l^2 - Norm$ measure.

6.3. *Local Integral Invariants*

We use the following parameters when implementing our descriptive features: For each interest point we set $V = 12$ which gives us a 12-dimensional feature vector that is generated using a set of either relational kernel functions or monomial kernel functions. Best results were obtained with $\varepsilon = 0.098$ in Equation 5.

When evaluating the localization approach we first compare each individual feature vector from the image in the query with all the other feature vectors, extracted from the training set of images, using the $l^2 - Norm$ measure. Correspondences between the feature vectors are found based on the method described in [1] which leads to robust matching. Then we apply a voting mechanism to find the corresponding image to the one in query. The voting is basically performed by finding the image that has the maximum number of matches. Figure 3 gives an example of the correspondences found between two different images using the proposed approach.

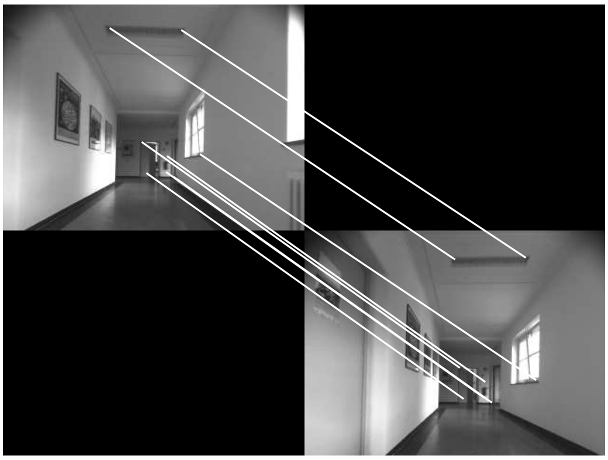

Figure 3. Matching two different images using the proposed approach.

6.4. *Results*

Figure 4 demonstrates the localization rate of the proposed approach and the existing one. It can be seen that the local integral invariants-based approach performers better than the global integral invariants-based approach using any of the two kernel types but gives best results using the relational kernel function. The global integral invariants-based approach has better results when using the monomial kernel function. One way to test if the results of the proposed approach are statistically *significant* or not, is to calculate the confidence interval for each experiment, and to check if the confidence interval of the mean values of the proposed approach does not overlap the confidence interval of the mean values of the

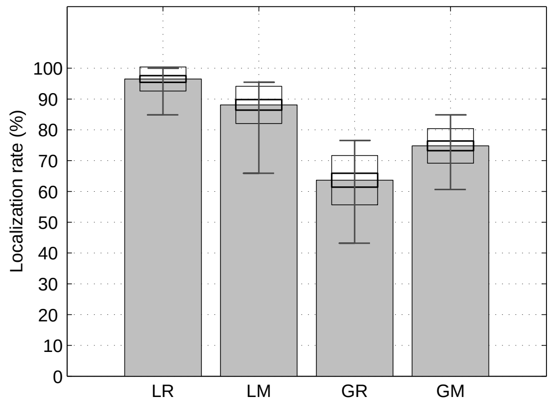

Figure 4. The localization rate of the local integral invariants-based approach using relational kernel functions (LR) and monomial kernel functions (LM), compared with the existing global integral invariants-based approach using relational kernel functions (GR) and monomial kernel functions (GM).

existing approach. The 95%-confidence interval test is used here, which means that the real mean value of each experiment lies in this interval with the propability of 95%. On the top of each bar in Figure 4, the following information are represented from outside to inside: The maximum and minimum values of the data, the standard deviation and the 95%−confidence interval. The confidence interval in the bar (LR) does not overlap any confidence interval of the other bars. This means that the results of this local integral invariants-based approach are significantly better than the other results.

The average localization time of the global approach and the local approach are 0.42 and 0.86 seconds respectively using a 3GHz Pentium 4. The additional computation time in the case of the local approach is due to the additional complexity in the feature extraction stages and the computation overhead during the comparison of the local features.

7. Conclusion

In this paper we have proposed an appearance-based robot localization approach based on local integral invariants. The local features have a compact size but are capable of matching images with high accuracy. In comparison with using global features, the local features show better localization results with a moderate computational overhead. Using local integral invariants with relational kernel functions leads to significant localization rate than the monomial kernel functions.

Acknowledgment

The first and second author would like to acknowledge the financial support by the German Academic Exchange Service (DAAD) of their PhD. scholarship at the Universities of Tübingen and Freiburg in Germany. The authors would also like to thank Christian Weiss for his assistance during the experimental part of this work.

References

[1] P. Biber and W. Straßer. Solving the Correspondence Problem by Finding Unique Features. In *16th International Conference on Vision Interface*, 2003.

[2] M. Brown and D. Lowe. Invariant Features from Interest Point Groups. In *British Machine Vision Conference, BMVC*, pages 656–665, Cardiff, Wales, September 2002.

[3] H. Bischof H, A. Leonardis A, and D. Skocaj. A Robust PCA Algorithm for Building Representations from Panoramic Images. In *Proceedings 7th European Conference on Computer Vision (ECCV)*, pages 761–775, Copenhagen, Denmark, 2002.

[4] A. Halawani and H. Burkhardt. Image retrieval by local evaluation of nonlinear kernel functions around salient points. In *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, volume 2, pages 955–960, Cambridge, United Kingdom, August 2004.

[5] Y. Ke and R. Sukthankar. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. In *CVPR (2)*, pages 506–513, 2004.

[6] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[7] K. Mikolajczyk and C. Schmid. A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630, 2005.

[8] O. Ronneberger, H. Burkhardt, and E. Schultz. General-purpose Object Recognition in 3D Volume Data Sets using Gray-Scale Invariants - Classification of Airborne Pollen-Grains Recorded with a Confocal Laser Scanning Microscope. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 2, Quebec, Canada, 2002.

[9] M. Schael. Invariant Grey Scale Features for Texture Analysis Based on Group Averaging with Relational Kernel Functions. Technical Report 1/01, Albert-Ludwigs-Universität, Freiburg, Institut für Informatik, January 2001.

[10] M. Schael. Texture Defect Detection Using Invariant Textural Features. *Lecture Notes in Computer Science*, 2191:17–24, 2001.

[11] 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)*, pages 2051–2058, Seoul, Korea, May 2001.

[12] S. Siggelkow. *Feature Histograms for Content-Based Image Retrieval*. PhD thesis, Albert-Ludwigs-Universität Freiburg, Fakultät für Angewandte Wissenschaften, Germany, December 2002.

[13] S. Siggelkow and M. Schael. Fast Estimation of Invariant Features. In W. Förstner, J. M. Buhmann, A. Faber, and P. Faber, editors, *Mustererkennung, DAGM*, pages 181–188, Bonn, Germany, September 1999.

[14] H. Tamimi, H. Andreasson, A. Treptow, T. Duckett, and A. Zell. Localization of Mobile Robots with Omnidirectional Vision using Particle Filter and Iterative SIFT. In *Proceedings of the 2005 European Conference on Mobile Robots (ECMR05)*, Ancona, Italy, 2005.

[15] H. Tamimi and A. Zell. Vision-based Global Localization of a Mobile Robot using Wavelet Features. In Informatik aktuell, editor, *Autonome Mobile Systeme (AMS), 18. Fachgespräch, Karlsruhe, 4. - 5. December*, pages 32–41, Karlsruhe, Germany, 2003.

[16] H. Tamimi and A. Zell. Vision based Localization of Mobile Robots using Kernel approaches. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, pages 1896–1901, 2004.

[17] 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.

[18] C. Zhou, Y. Wei, and T. Tan. Mobile Robot Self-Localization Based on Global Visual Appearance Features. In *Proc. IEEE Int. Conf. Robotics & Automation (ICRA)*, pages 1271–1276, 2003.