

Efficient Topological Localization Using Orientation Adjacency Coherence Histograms

Junqiu Wang and Hongbin Zha
National Laboratory on Machine Perception
Peking University
Beijing 100871, China
jerywang@public3.bta.net.cn

Roberto Cipolla
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
cipolla@eng.cam.ac.uk

Abstract

This paper describes an efficient vision-based global topological localization approach that uses a coarse-to-fine strategy. Orientation Adjacency Coherence Histogram (OACH), a novel image feature, is proposed to improve the coarse localization. The coarse localization results are taken as inputs for the fine localization which is carried out by matching Harris-Laplace interest points characterized by the SIFT descriptor. Computation of OACHs and interest points is efficient due to the fact that these features are computed in an integrated process. We have implemented and tested the localization system in real environments. The experimental results demonstrate that our approach is efficient and reliable in both indoor and outdoor environments.

1. Introduction and Related Work

Vision-based topological localization can be considered as a place recognition problem by matching a query image with representative images in a location database [3, 6, 8, 10]. Image features with desirable properties lead to an efficient and reliable localization system. Since the viewpoints of a robot tend to be different in the training and testing stages, the features used as landmarks should be robust against viewpoint changes. In addition, the features have to deal with partial occlusions and illumination condition changes.

Local features invariant to different kinds of viewpoint changes have been used in existing localization algorithms. Kosecka and Yang characterized scale-invariant key points by the SIFT descriptor in their localization system [3]. Harris-Laplace interest points described by the SIFT descriptor have been used as natural landmarks for localization [7]. Column matching has also been employed for navigation [1]. These systems are based on local feature match-

ing that is slow, especially when a lot of images are used to represent a large environment.

Wang et al. proposed a coarse-to-fine global localization strategy in which the *Vector Space Model* (VSM) is employed to accelerate the localization process [8]. Both the coarse and the fine localization are carried out based on local features. The performance of this algorithm is not satisfying in outdoor environments [9]. In addition, the coarse localization is not efficient enough. In this work, we extend traditional gradient orientation histogram to Orientation Adjacency Coherence Histogram (OACH) to improve coarse localization. Traditional gradient orientation histograms do not have any spatial information. The proposed OACH captures global image properties as well as some spatial information. The computation of OACHs is efficient because they are computed with Harris-Laplace interest points in an integrated process.

Color histogram is an alternative that could be used in coarse localization. However, localization using color histograms is sensitive to illumination changes [6]. Zhang and Kosecka proposed a hierarchical localization system, in which color histograms are computed only on buildings [11]. Their system is limited to building recognition. In contrast, our system has good performance in indoor and outdoor environments.

2. Feature Detection

2.1. OACH

OACH is an extension based on traditional gradient orientation histograms. Different from the SIFT descriptor, OACH captures global image properties while retaining some spatial information.

2.1.1. Orientation Histogram

The image derivatives in the u and v directions, L_u and L_v , are computed separately. The computation is imple-

mented by convolution with the differential of Gaussian kernel of standard deviation σ_D :

$$L_u(\mathbf{x}, \sigma_D) = I(\mathbf{x}) * G_u(\mathbf{x}, \sigma_D), \quad (1)$$

$$L_v(\mathbf{x}, \sigma_D) = I(\mathbf{x}) * G_v(\mathbf{x}, \sigma_D), \quad (2)$$

The orientation of the pixel is

$$\theta(\mathbf{x}) = \arctan\left(\frac{L_v}{L_u}\right) \quad (3)$$

The gradient orientation is quantized into m bins (o_1, \dots, o_m). There are three requirements for the setting of the m : (a) gradient orientation histograms should be robust against image rotations; (b) these histograms should use a limited amount of memory; (c) the orientation histogram should be sufficiently discriminative. The m is experimentally set to 8 based on the above requirements.

For a pixel $p = (u, v)$, let $O(p)$ denote its gradient orientation, the traditional orientation histogram h of I is defined for $i \in \{1, 2, \dots, m\}$. Given a pixel in I , $h_{OH}(o_i)$ gives the probability that the gradient orientation of a pixel is o_i .

2.1.2. Orientation Adjacency Histogram

In an Orientation Adjacency Histogram (OAH), the distributions of gradient orientations of a pixel's 4-neighbor is counted. It captures spatial correlation between orientations of adjacent pixels. The OAH gives the probability that a pixel at a city block distance 1 from the center pixel is o_j when the orientation of the center pixel is o_i :

$$h_{OAH}(o_j, o_i) = p(O_N = o_j | O_C = o_i). \quad (4)$$

The orientations of the neighborhood of the center pixel are accumulated and then normalized by the number of the center pixels of orientation o_i . The OAH requires m^2 space for its storage (64 in this work).

2.1.3. Orientation Adjacency Coherence Histogram

The OAH is further extended to OACH based on the Harris detector. In [2], an image is decomposed into three kinds of regions: smooth areas, edges and corners. Two OAHs are computed respectively in the edge and corner regions and stacked together to form an OACH. Since more spatial information is included, the discriminative ability of OACH is better than that of OAH.

The basic idea of the Harris detector is to use the autocorrelation function in order to determine locations where the signal changes in one or two directions. A matrix related to the auto-correlation function is computed:

$$C(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 G(\mathbf{x}, \sigma_I) * \begin{pmatrix} L_u^2 & L_u L_v \\ L_u L_v & L_v^2 \end{pmatrix} \quad (5)$$

where σ_D is the derivation scale, σ_I the integration scale, and G the Gaussian.

Edges and interest points are computed based on:

$$\det(C) - \alpha \cdot \text{trace}^2(C) < T_E, \quad (6)$$

and

$$\det(C) - \alpha \cdot \text{trace}^2(C) > T_C. \quad (7)$$

Edges are computed based on Eq.(6), where α is the coefficient of the Harris function and T_E is the threshold of the Harris function ($T_E < 0$). The edge detection is carried out at the first scale. Corner regions are detected by using Eq.(7), T_C is the threshold for corner regions ($T_C > 0$).

Orientations of the pixels in an image are accumulated and put into two OAHs according to the pixel classification results.

2.2. Harris-Laplace Interest Points Detection and Characterization

Harris interest points that are scale invariant can be detected at different scales based on the Harris function [5]. In order to match interest points, the SIFT descriptor [4] is use here to characterize these points.

3. Coarse Localization

During the training stage, many images are captured in an environment. Representative images are extracted from these images. These images are employed to represent the environment.

3.1. Topological Environment Models

The topological structure of the environment is extracted based on the images taken during the training stage. The structure extraction needs a definition of similarity measure in the image appearance space. In this work, the Jeffrey divergence distance between the OACHs of two images is used to measure the similarity. If the distance between two adjacent images is below a threshold G_L , they are considered as belonging to one distinct place and are clustered based on the similarities between OACHs. The distances are accumulated along the path. A new location is created if the accumulated distance is greater than another threshold G_H . In our experiments, the G_L is set to 1200 and G_H is set to 15000.

3.2. Coarse localization

To determine the location of the input image captured for localization, we have to measure the similarity between the OACH of the input image and those of each image in the location database for coarse localization. Similarity measures

for histograms can be broadly partitioned into bin-by-bin and cross-bin measures. We have investigated different bin-by-bin measures such as L_1 distance, L_2 distance, Jeffrey divergence, χ^2 distance and histogram intersection. According to our experiments, the Jeffrey divergence provides the best matching results, followed by histogram intersection.

The Jeffrey divergence between the OACH (Q) of an input image captured at an unknown location ($Q = \{q_k | k = 1, 2, \dots, 2m\}$) and those ($E^l = \{e_k^l | k = 1, 2, \dots, 2m\}$) in the database is defined as

$$d(Q, E_j^l) = \sum_k (q_k \log \frac{q_k}{r_k} + e_k^l \log \frac{e_k^l}{r_k}), \quad (8)$$

where $r_k = \frac{(q_k + e_k^l)}{2}$.

3.3. Candidates Selection for Fine Localization

The coarse localization results are taken as inputs for the fine localization. To determine whether a representative image ranked at l th place should be included in the candidate set for the fine localization, a confidence measure c_l is defined:

$$c_l = \frac{d_m}{d_l} \quad (9)$$

where d_m is the minimum matching distance of all locations in the database, d_l is the distance of the location ranked at l th place. The ambiguity values range between 0 and 1. The location ranked at the first place has confidence value 1, and thus it is always considered as a possible location for the fine localization. The confidence value is high if a candidate location has similar distance that of the location ranked at the first place. A location is put into the candidate set if its confidence value is greater than a threshold c_h . The threshold is experimentally set to 0.5.

4. Fine Localization

The fine localization is to find the correct location in a small set that has been determined by the coarse localization results. Fine localization is carried out based on local feature matching [8].

5. Experiments

For experimental verification, we have implemented and tested the proposed approach. All of these tests are conducted on a 1.4GHz laptop with 128M memory. The size of the images is 640×480 .

We do not have any assumption of the position of the robot before the coarse-to-fine localization is carried out. The purpose of our system is to find the correct location based on one image.

Table 1. Correct ratio of the localization tests.

Images	= 1	≤ 2	≤ 3	≤ 4	≤ 5
Test-A	84.4%	90.1%	92.8%	96.5%	98.8%
Test-C	68.2%	84.3%	89.8%	95%	98.3%
Test-B	82.5%	86.3%	89.5%	91.2%	92.6%
Test-D	62%	79.9%	82.5%	84.2%	88.7%
Test-E	80.6%	83.4%	86.2%	87.3%	89.2%
Test-F	58.5%	64.7%	77.5%	80.1%	81.8%

5.1. Indoor Experiments

The representative images are indexed into two databases. Two image sequences are captured for testing of our approach. The first one (Sequence-I) contains 315 images captured roughly along the path of the first exploration by a camcorder. The second one (Sequence-II) contains 482 images captured along a path deviating from the first exploration (about 0.5 meters from the first exploration paths).

Two tests are conducted using the test image sequences. Test-A tests Sequence-I and Test-B tests Sequence-II. Based on Eq.(9), the coarse localization determines the candidate locations for the fine localization.

The proposed approach provides good performance in these tests. In Fig. 1(a), the correct location is found although the illumination condition of the test image is different from that of the image in the database. In Fig. 1, the input images are shown in the first row, and the representative images found by our system are shown in the second row. The local feature matching results (interest point correspondences) are also illustrated.

To compare the performance of our system with that of related work, we have implemented and test the approach proposed in [9]. Localization is conducted using the same training images and the same testing images by using the approach in [9]: Test-C tests Sequence-I and Test-D tests Sequence-II. The correct ratios of the coarse localization are shown in Table 1. A localization result is taken as correct when the following conditions are met: a) the correct one is found in the fine localization stage; b) there are more than 8 correct matches, the minimum number needed for relative position recovery. The correct ratio in Table 1 is defined as the percentage of correctly retrieved images ranked before the i^{th} place.

5.2. Outdoor Experiments

We carried out the outdoor experiments on a campus. We captured 320 images at different locations (roughly 2 meters between the positions where two adjacent images are taken)

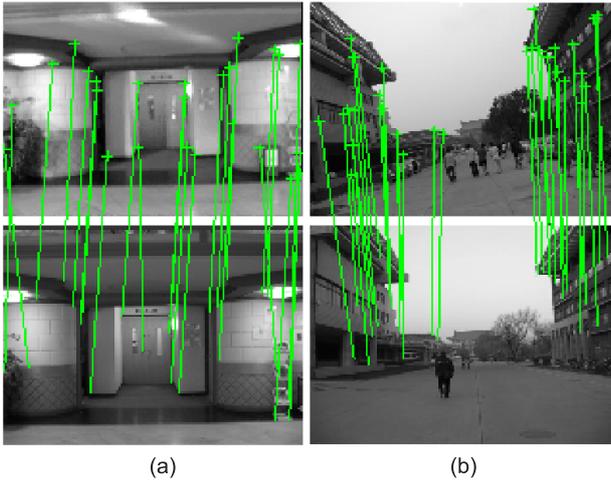


Figure 1. Indoor and outdoor localization examples.

along the paths around the buildings. The total distance is about 800 meters. Among them, 124 representative images are extracted to represent the outdoor environment and 12 locations are created. The test set consists of 215 images, which are different from the training images. These images are taken randomly along the path within 2 meters deviation from the first exploration route under different weather conditions and in different seasons. Our system demonstrates good performance in the outdoor environment. In Fig. 1(b), the location is correctly found although there are many people in the test image. Test-E is the test of outdoor localization using the proposed approach. For comparison, Test-F is conducted based on the approach in [8]. The overall performance of the proposed approach is shown in Table 1. The proposed approach demonstrates better performance than that in [8].

5.3. Time complexity

Computation of OACHs: The time complexity of the OACH is $O(n_1 n_2)$ (n_1 is the width of the images and n_2 the height). Since the gradient orientations of 4-neighbor pixels have to be counted, the complexity of the proposed method is slightly higher than that of the traditional orientation histograms. It takes about 0.1 seconds to compute an OACH in a 640×480 image.

Time for Localization: In [8], the computation in the localization includes term assignment, coarse localization from the LVSM and fine localization from the database. There is no term assignment in our approach. Table 2 shows the time comparison between the proposed approach and the approach in [8]. It is clear that the proposed approach is more efficient.

Table 2. Comparison of the average times used in the localization (seconds)

	Term	Coarse	Fine	Total
Approach in [8]	0.09	0.013	0.12	0.245
Our approach		0.005	0.10	0.105

6. Conclusions and Future Work

We present an efficient and reliable vision-based topological localization approach. The OACH is successfully used in coarse localization and the performance is improved. OACH can deal with viewpoint and illumination changes that bring difficulties for topological localization. Our localization system provides better performance than that in closely related work.

References

- [1] T. Goedeme, T. Tuytelaars, and L. V. Gool. Fast wide baseline matching for visual navigation. In *Proc. CVPR'04*, pages I-24 – I-29, 2004.
- [2] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. of Alvey Vision Conference*, pages 147–152, 1988.
- [3] J. Košecká and F. Li. Vision based topological markov localization. In *Proc. ICRA'04*, pages 1481–1486, 2004.
- [4] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. ICCV'99*, pages 1150–1157, 1999.
- [5] K. Mikolajczyk and C. Schmid. Indexing based on scale-invariant features. In *Proc. ICCV'01*, pages 525–531, 2001.
- [6] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. ICRA'00*, pages 1023–1029, 2000.
- [7] J. Wang, R. Cipolla, and H. Zha. Image-based localization and pose recovery using scale invariant features. In *Proc. of IEEE Int'l. Conf. on Robotics and Biomimetics*, pages 711–715, 2004.
- [8] J. Wang, R. Cipolla, and H. Zha. Vision-based global localization using a visual vocabulary. In *Proc. ICRA'05*, pages 4241–4246, 2005.
- [9] J. Wang, H. Zha, and R. Cipolla. Coarse-to-fine vision-based localization by indexing scale-invariant features. *IEEE Trans. on Systems, Man and Cybernetics*, 36(2):413 – 422, 2006.
- [10] J. Wolf, W. Burgard, and H. Burkhardt. Robust vision-based localization by combining an image retrieval system with monte carlo localization. *IEEE Trans. on Robotics*, 21(2):208–216, 2005.
- [11] W. Zhang and J. Košecká. Localization based on building recognition. In *IEEE Workshop on Applications for Visually Impaired in CVPR*, 2005.