INTECH
open science | open minds

OPEN ACCESS
ARTICLE

International Journal of Advanced Robotic Systems

Efficient Topological Localization Using Global and Local Feature Matching

Regular Paper

Junqiu Wang[1,*] and Yasushi Yagi[1]

1 The Institute of Scientific and Industrial Research, Osaka University, Japan
* Corresponding author E-mail: jerywangjq@gmail.com

Received 1 Jun 2012; Accepted 24 Dec 2012

DOI: 10.5772/55630

Abstract We present an efficient vision-based global topological localization approach in which different image features are used in a coarse-to-fine matching framework. 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. The computation of OACHs and interest points is efficient due to the fact that these features are computed in an integrated process. The matching of local features is improved by using approximate nearest neighbor searching technique. 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. This work has also been compared with previous works. The comparison results show that our approach has better performance with higher correct ratio and lower computational complexity.

Keywords A Localization, Global Features, Local Features, Feature Matching

1. Introduction

Mobile robot localization is essential for robot autonomous navigation and task fulfillments [17]. Vision-based localization using natural landmarks is highly desirable for a wide range of applications. With the development and dissemination of camera-equipped mobile devices, vision-based localization can help a person to determine his position in an unfamiliar environment by using one image [14]. Vision-based localization can be implemented in a two-stage process in which global localization and local tracking are carried out consequentially [20]. We will consider global localization as a place recognition problem by matching a query image with those images in a location database that represents an environment. The image matching process adopts a coarse-to-fine framework in which both improved global and local features are employed. In this paper, we focus on improving the efficiency and reliability of the localization system.

Using image features with desirable properties is the key to an efficient and reliable location recognition system [13]. Since the viewpoints of a camera system (mounted on a mobile robot or mobile device) tend to be different in the training and testing stages, the features taken as natural landmarks should be robust against viewpoint changes. In addition, those features have to deal with partial occlusions and illumination condition changes. Global image features such as color histograms can be computed and matched easily. However, it is well known that color histograms are sensitive to illumination changes. Other global features have been investigated for localization systems and success has been achieved. Despite of these successes, those localization systems that only employ global features cannot do pose recovery because no local feature correspondence is available for the relative position computation.

Local features invariant to different viewpoint changes have proved useful in object recognition [9] and vision-based localization [7, 8]. Using a local feature detector such as the Harris-Laplace interest point detector or DoG detector, a few hundreds interest points can be detected in one image. Thus, the matching of local features of one image to many is slow especially when a large number of images are used to represent a large environment. Moreover, local feature descriptors do not have global context information. Based on the above analysis, we believe that the integration of global and local features can lead to an efficient and reliable localization system.

It has been found that image gradient orientation information is not sensitive to light conditions. Thus gradient orientation histogram is a possible choice for location representation. In this work, we extend the traditional gradient orientation histogram to Orientation Adjacency Coherence Histogram (OACH) to improve coarse localization. Traditional gradient orientation histograms do not have any spatial information that is important for reliable image matching. The proposed OACH will capture global image information as well as certain spatial information. The Harris-Laplace detector and SIFT descriptor are adopted here for local feature detection and description. The combination of global and local features makes our system efficient and reliable. In addition, the local features matching results can be used for relative position computation [21]. The computation of OACHs and Harris-Laplace interest points is efficient because they are computed in an integrated process.

According to the representation method of an environment, existing localization approaches can be classified into three categories: metric, topological, and hybrid [19]. The approach proposed in this paper belongs to the second one, in which environments are described by using adjacency graphs. The goal of a topological localization system is to determine the node of the graph that corresponds to the camera system's location.

1.1 Overview

The localization system is trained using the images captured in an environment. Global (OACHs) and local (Harris-Lapalce interest points described by the SIFT descriptor) features are computed in these images. These features are indexed into two databases: OACH database for coarse localization and SIFT database for fine localization. A coarse-to-fine strategy is adopted here to do localization using one input image. Global and local features are also computed in the input image. In coarse localization, the global feature is efficiently matched against the OACH database. A set of locations with high similarities is chosen as the input for the fine stage. Local features are matched and the final decision is made according to the number of local features successfully matched. Relative position can be recovered based on the local feature matching results.

The rest of this paper is structured as follows. In the next section, previous work is introduced; Section 3 addresses global feature (OACH) and local feature (Harris-Laplace interest point) detection methods; Section 4 describes coarse localization based on OACHs, while Section 5 presents fine localization in which local feature matching is used. Experimental results, performance evaluation and conclusions are given in the final two sections.

2. Previous Work

Vision-based localization has been investigated for more than 30 years. DeSouza and Kak [3] gave a comprehensive survey of vision-based methods for indoor and outdoor localization and navigation. Refer their paper for an overview.

In a vision-based topological localization system, a location in an environment can be represented by different kinds of image features. Global image features are employed in many works because they can be computed easily. Ulrich and Nourbakhsh [19] presented an efficient topological localization system using color histograms as representation of environments. Torralba et al. [18] suggested a context-based place recognition system. In their system, each location node is represented by a Gaussian mixture model derived from the integration of responses of bank of wavelet filters. They model the spatial relationships between locations by a Hidden Markov Model, which improves the system's performance. Yagi et al. [23] described a route navigation method using a dual active contour model in omni-directional images to represent the environment. Their system works well in indoor environments. However, it is difficult to apply their method in a large environment. Katsura et al. [6] developed an outdoor localization system based on segmentation of the images. Their system can obtain the location by matching areas of trees, sky, and buildings. However, occlusions bring trouble to the matching. Those systems solely based on global features cannot recover the relative position for the camera with respect to the scene (except [23], in which relative pose can be recovered thanks to the property of omni-directional cameras). Wang et al. [30] proposed a topological localization method using local feature matching directly.

In recent years, many local feature detectors are proposed to conquer partial occlusion and viewpoint change problems. Local feature descriptors were also invented to encode appearance in the neighborhood of interest points. Lowe [9] presented the scale invariant feature detector that searches for local scale space maxima of Difference-of-Gaussian (DoG). At the same time, he introduced a novel local feature descriptor using local gradient orientation information. Se et al. [16] adopted the SIFT detector and descriptor in their localization system as 3D landmarks. Trinocular camera system is employed to estimate the 3D position of landmarks and their region of confidence. They update a 3D map by selecting reliable SIFT features by using a Kalman filter. Goedeme et al. [4] tried to detect and match line segments using color-based descriptors. They have shown that those line segments are invariant to small viewpoint changes. Harris-Laplace interest point detector, another scale invariant feature detector, was used in [20] to determine the location of a camera system. Although it seems that affine or perspective invariant features are better choices for a localization system, they have not been used in localization because the computations of those features usually are very expensive.

Noticing that the matching of local features is a challenge especially for localization in large environments, Wang et al. [21, 29] proposed coarse-to-fine global localization strategy in which the Location Vector Space Model (LVSM) was proposed to accelerate the localization process. 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. In addition, the coarse localization is not efficient enough. Beis and Lowe [1] presented Best Bin First(BBF) algorithm based on the K-D tree for fast approximate nearest neighbor searching. The technique has been used in a panoramic image producing system [11], in which less than 50 images are processed. In contrast, the proposed approach can deal with a much larger environment. Zhang and Kosecka [24] proposed a hierarchical localization system in which color histograms are computed only on buildings for the coarse localization. Their system is limited to building recognition. In their work, it is assumed that there is only one building in each image, which is not always true in practice. The well-known illumination sensitivity of color histograms is another drawback of their work.

Valgren and Lilienthal [25] introduced spectral clustering techniques into their localization system. The experimental results show that their system works despite of the high computational cost. Tipaldi et al. [26] developed a system to create approximate maps in dynamic environments. Ni et al. [28] proposed epitomic representation of environments. These works tried to assemble local representations into location representation. However, global features are not combined with local interest points.

The proposed approach to vision-based global localization is also related to efforts in wearable computing. Davison [2] presents a real-time 3D SLAM system using reliable features. In his work, a covariance matrix is kept for local feature updating. His system gives good performance in small environments. However, if the environment is large and cannot be covered by a few hundred features, the feature covariance updating is computationally prohibitive.

3. Feature Computation

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

3.1 Orientation Histogram

The image derivatives in the *u* and *v* directions, $\mathrm{L_u}$ and $\mathrm{L_v}$ are computed separately. The computation is implemented by convolution with the differential of Gaussian kernel of standard deviation σ_D

$$\mathrm{L_u}(\mathrm{x},\sigma_\mathrm{D}) = \mathrm{I}(\mathrm{x}) * \mathrm{G_u}(\mathrm{x},\sigma_\mathrm{D}) , \tag{1}$$

$$\mathrm{L_v}(\mathrm{x},\sigma_\mathrm{D}) = \mathrm{I}(\mathrm{x}) * \mathrm{G_v}(\mathrm{x},\sigma_\mathrm{D}) , \tag{2}$$

where *I* is the input image; and $\mathrm{G_u}$ and $\mathrm{G_v}$ are the Gaussian filters, respectively. We calculate the orientation of the pixel $\mathrm{I}(\mathrm{x})$ by

$$\alpha(\mathrm{x}) = \arctan(\frac{\mathrm{L_v}(\mathrm{x})}{\mathrm{L_u}(\mathrm{x})}) , \tag{3}$$

The gradient orientation is quantized into *m* bins $(\theta_1, \theta_2, ..., \ \theta_\mathrm{m})$. 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 set experimentally to 8 based on the above requirements.

For a pixel $\mathbf{x} = (u, v)$, let $\|\alpha(\mathrm{x})\|$ denote its quantized gradient orientation, the traditional orientation histogram *h* of image *I* is defined for $\mathrm{i} \in \{1, 2, ..., \mathrm{m}\}$. Given a pixel in *I*, $\mathrm{h}_{\mathrm{OH}}(\|\alpha(\mathrm{x})\| = \theta_{\mathrm{i}})$ indicates the probability that the gradient orientation of a pixel is θ_{i} .

3.2 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 **x** at a city block distance 1 from the center pixel is θ_{j} when the orientation of the center pixel is θ_{i} :

$$\mathrm{h}_{\mathrm{OAH}}(\theta_{\mathrm{j}}, \theta_{\mathrm{i}}, \mathrm{x}) = \mathrm{p}(\|\alpha(\mathrm{x}_{\mathrm{N}})\| = \theta_{\mathrm{j}}, | \, \|\alpha(\mathrm{x})\| = \theta_{\mathrm{i}}). \qquad (4)$$

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

3.3 Orientation Adjacency Coherence Histogram

The OAH is further extended to OACH based on the Harris detector. In [5], an image is decomposed into three kinds of regions: smooth areas, edges and corners. Base on this decomposition we calculate two OAHs 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:

$$\mathrm{C}(\mathrm{x}, \sigma_{\mathrm{I}}, \sigma_{\mathrm{D}}) = {\sigma_{\mathrm{D}}}^2 \mathrm{G}(\mathrm{x}, \sigma_{\mathrm{I}}) * \begin{pmatrix} \mathrm{L}_{\mathrm{u}}^2 & \mathrm{L}_{\mathrm{u}}\mathrm{L}_{\mathrm{v}} \\ \mathrm{L}_{\mathrm{u}}\mathrm{L}_{\mathrm{v}} & \mathrm{L}_{\mathrm{v}}^2 \end{pmatrix}, \qquad (5)$$

where e σ_{D} is the derivation scale, and σ_{I} the integration scale; G the Gaussian filter.

Edges and interest points are computed based on

$$\det(\mathrm{C}) - \mathrm{k} \cdot \mathrm{trace}^2(\mathrm{C}) < \mathrm{T}_{\mathrm{E}} , \qquad (6)$$

And

$$\det(\mathrm{C}) - \mathrm{k} \cdot \mathrm{trace}^2(\mathrm{C}) > \mathrm{T}_{\mathrm{C}} , \qquad (7)$$

where *k* is the coefficient of the Harris function, T_{E} is the threshold of the Harris function ($\mathrm{T}_{\mathrm{E}} < 0$); T_{C} is the threshold for corner regions ($\mathrm{T}_{\mathrm{C}} > 0$). We detect edges using Equation (6); and detect corners using Equation (7). The edge detection is carried out at the first scale.

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

3.4 Harris-Laplace Interest Point Detection and Characterization

Harris interest points are found at several scales based on the Harris function and Harris-Laplace interest points are detected by searching for the maximum of the Laplacian function [21]. To accelerate the local feature computation, the Harris-Laplace interest point detection in this work follows the method proposed in [21], in which Harris interest points are detected at four scales with the initial scale 1.2 and the maximum scale is 2.07. Although the detector can deal with smaller scale changes than the original implementation, we can detect reliable local features and the scale change is enough for a feature matching in a localization system. Moreover, the time for feature detection is significantly reduced.

In order to match interest points, the SIFT descriptor [9] is used here to characterize these points. The SIFT descriptor consists of orientation, scale, coordinates and a 128D vector.

4. Coarse Localization

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

4.1 Topological Environnent Model

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, G_{L} is set to 1200 and G_{H} is set to 15000.

4.2 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 L1 distance, L2 distance, Jeffrey divergence, χ2 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 and those (X^l) in the database is defined as the sum of two Jeffrey divergences between two OAHs computed on edges and corners:

$$d(Q,X^l)=\sum_{k=1}^{m^2}(q_k^E\log\frac{q_k^E}{r_k^E}+x_k^{E,l}\log\frac{x_k^{E,l}}{r_k^E})$$

$$+\sum_{k=1}^{m^2}(q_k^C\log\frac{q_k^C}{r_k^C}+x_k^{C,l}\log\frac{x_k^{C,l}}{r_k^C}), \qquad (8)$$

where $r_k^E=\frac{q_k^E+x_k^{E,l}}{2}$ and $r_k^C=\frac{q_k^C+x_k^{C,l}}{2}$. *E* denotes edges and C denotes corners.

4.3 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 i^{th} place should be included in the candidate set for the fine localization, a confidence measure c_i is defined as:

$$c_i=\frac{d_m}{d_i}, \qquad (9)$$

where d_m is the minimum matching distance of all locations in the database, d_i is the distance of the location ranked at the i^{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 set to 0.5 in our experiments.

5. Fine Localization

5.1 Feature Selection

These local features detected by Harris-Laplace detector and described by the SIFT descriptor in the training images have different discriminative ability for the localization. Certain local features with similar descriptors occur in most of the training images. They have little contribution to the matching of images. At the same time, those local features appearing in only one or two training images usually are not reliable. We try to select those local features that are appropriate for the matching.

Wang et al. [21] proposed a feature selection method by using the Zipf's law. First, the feature descriptors are clustered by using k-means algorithm. The cluster centers are taken as the terms of a visual vocabulary. The terms with very high frequency and very low frequency are deleted from the visual vocabulary. According to the Zipf's law, the terms whose products of the frequency and the rank are approximately a constant are kept in the vocabulary. Wang et al. discard those terms that are not discriminative enough.

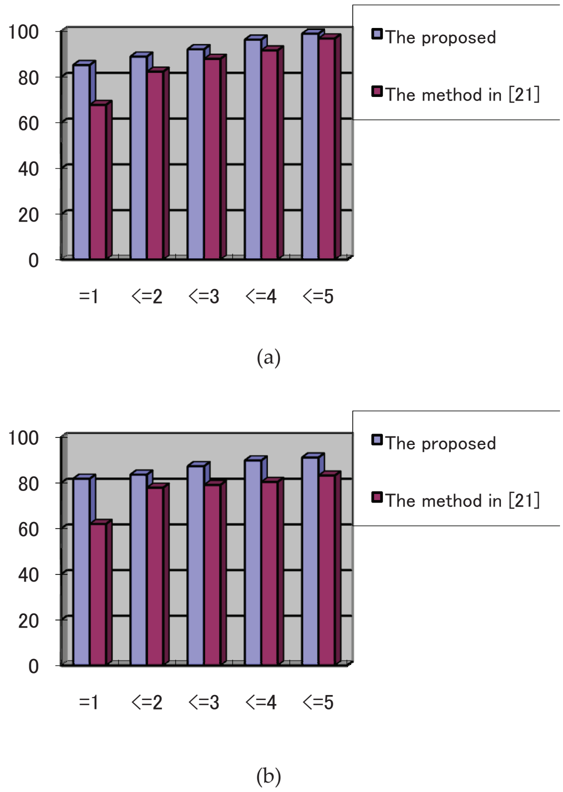
In this work, a visual vocabulary is built and the Zipf's law is applied to the vocabulary. However, we do not delete those terms with low discriminative ability (we will call the set of these terms as stop list) to identify good features from the features in the training images. During the training stage, we assign a term to one feature according to the distance between the features to each term. If the term belongs to the stop list, we will discard this feature. Otherwise the feature will be added to the location database. The feature selection method is effective because those features with low discriminative ability are discarded and an environment can be represented by a smaller feature set. The vocabulary used in this work consists of 1024 terms in which 689 terms are put into the stop list. Although the number of the terms in the stop list is large, the remaining local features after the feature selection are enough for the localization. The matching of local features will be computationally less expensive.

5.2 Local Feature Matching

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 [20]. The computational complexity of similarity search on very large data sets is high especially the descriptors are in high dimensional spaces. It is necessary to accelerate the fine localization since the SIFT descriptor is a 128D vector. We find a set of candidate feature matches using an approximate nearest neighbor algorithm. A k-d tree is built based on the selected features.

To further improve the efficiently of local feature matching, we index local features according to the orientation information of SIFT descriptors. In the computation of SIFT descriptors, a consistent orientation is assigned to each descriptor based on local gradient orientations in the neighborhood of a interest point. The descriptor is represented relative to this orientation and therefore achieves rotation invariance. Full invariance to rotation is not necessary for local feature matching in a localization system because the rotation of a camera system rarely exceeds 45 degrees. The SIFT descriptors computed from the representative images are indexed into four bins. The indexing is a ”soft” assigning process, in which those descriptors with orientations near the boundary is assigned to both neighboring bins.

If the fine localization is ambiguous, the Epipolar geometry constraints can be used here to verify the matching results. Following the approach in [21], we estimate fundamental matrix by using the RANSAC algorithm. If a match is accepted by the fundamental matrix, it is an inlier. Otherwise it is an outlier. Only inliers are counted for the location determination. This approach is effective in the fine localization.



(a)

(b)

Figure 1. (a). The performance comparison tested on the first indoor image sequence. The figures show the percentage of correctly retrieved images ranked in top 5 locations; (b). The performance comparison tested on the second indoor image sequence. The figures show the percentage of correctly retrieved images ranked in top 5 locations.

6. Experiments

We test our localization method on an indoor and an outdoor image dataset. The performance of our method is compared with the one proposed in [23]. Since our OACH is a global feature that is used in the first stage of our localization system, we compare our OACH with another widely used global feature – GIST. GIST is defined as the meaningful information that an observer can identify from a glimpse at a scene. Our OACH can be used in image retrieval task without any difficulty. The performance of the OACH is compared with the GIST in two datasets.

We test our approach and other methods on 4 different datasets. The first dataset covers an indoor environment. The images taken on different days include illumination and viewpoint variations. The second dataset is an outdoor dataset consisting of many images captured in a campus. The third dataset is the ZuBuD dataset, which consists of 1005 the images of 201 buildings. There 115 query images that were taken in different days. The last dataset is the Holidays dataset, which is the most difficult one.

In all the experiments, 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.

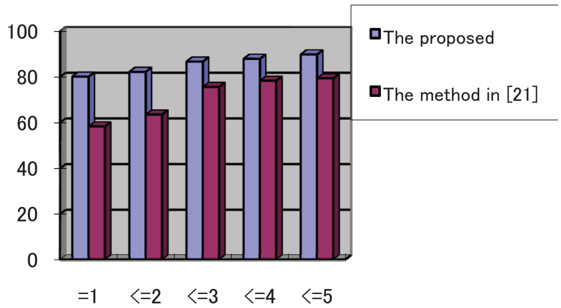


Figure 2. The performance comparison tested on the outdoor image sequence. The figures show the percentage of correctly retrieved images ranked in top 5 locations.

6.1 Indoor Experiments

We use the dataset which has been tested in [21]. The dataset was captured in the ground floor of a building.

The environment model consists of 127 locations. The representative images are extracted from the training image sequence. These representative images are indexed into two databases: OACH database and local feature database. 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).

To compare the performance of our system with that of related work, the computational power of the computer used in this work is same to the one used in [21]. In addition, localization is conducted using the same training images and the same testing images by using the approach in [21]. The comparison criteria here are the percentage (correct ratio) of test images that are correctly localized from the location database. Two comparison tests are conducted using the test image sequences: Figure 1(a) shows the experimental results on Sequence-I and Figure 1(b) shows the results on Sequence-II. Based on Equation (9), the coarse localization determines the candidate locations for the fine localization. The proposed approach provides better performance in these tests than that in [21]. In Figure 4(a), the correct location is found in the fine localization stage although the illumination condition of the test image is different from that of the image in the database.

The proposed approach provides much better results than the method described in [21]. Our global features is useful in this environment.

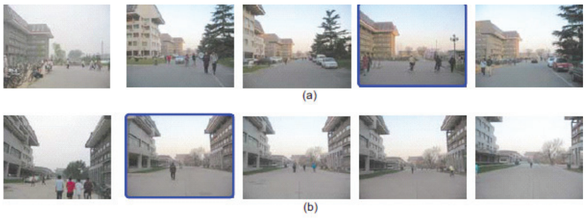


Figure 3 The coarse localization examples in the outdoor environment. The first image in each row is the input image for localization. The coarse localization results are shown following the input image. The correct representative images are framed by the blue rectangles.

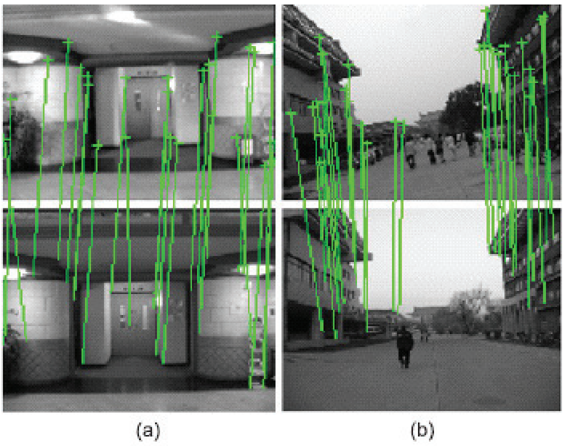


Figure 4. Based on the coarse localization, the localization system finds the correct location by the local feature based refinement. The images in this figure show local feature matching results between the input images and the representative images found in the database. The input images are shown in the first row, and the representative images found by our system are shown in the second row.

6.2 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) 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 Figure 3(a), the color distribution in the input image is different from the images from the training stage due to the different illumination conditions. Thus coarse localization by using color information is not reliable. We find the correct location by matching the OACHs during the coarse localization stage. We find the correct location in the fine localization by using local feature matching. In Figure 3(b), the location is correctly found although there are many people in the test image. The local feature matching results are shown in Figure 4.

We compare the proposed approach with the one in [21]. The overall performance of the proposed approach is shown in Fig. 1. The proposed approach demonstrates better performance than that in [21].

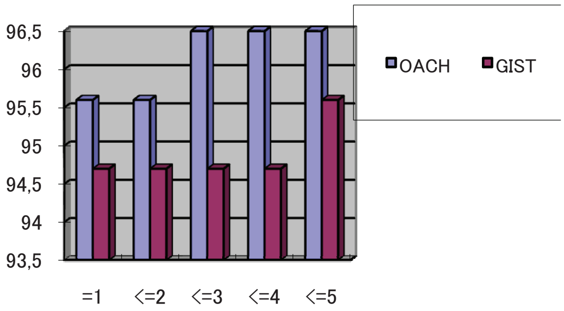


Figure 5. The performance comparison tested on the Zurich building dataset. The figures show the percentage of correctly retrieved images ranked in top 5 locations. We compare the performance of our global feature OACH and the GIST feature [13] on this dataset.

6.3 Experiments on the ZuBuD Building Dataset

Our OACH is a global feature that is used in the first stage of our localization system. This dataset contains 201 buildings, each of which is represented by 5 images. Therefore, the dataset has 1005 images in the training set. The query set includes 115 images that were taken under different dates and weather conditions. We test our global feature OACH and the GIST feature in [13]. The GIST has been widely used as a global feature description of scenes. We compare the performance of our OACH and the GIST feature on this building dataset. The GIST feature is a good at describing an image with low dimensionality. It has been widely applied in many works. The performance comparison is illustrated in Figure 5. The proposed OACH outperforms the GIST feature a little. Although the GIST has similar performance as our OACH feature, the detection process in our method is an integrated process. We can apply our local features smoothly after the OACH filtering process.

6.4 Experiments on the Holidays Dataset

The Holidays dataset is composed of a set of holiday images and other images taken under different viewpoints and illumination conditions. A large variety of scene types can be found in this dataset. The images of a distinct scene or object are collected in one group. There are 500 image groups in the dataset. The query image set is composed of the first images of each image group.

The Holidays dataset is more difficult than the ZuBuD dataset. The viewpoint variations are large in many image groups. The illumination conditions in many images of the same group are very different. We compare the performance of our method and the GIST descriptor on this dataset. Our method gives a little performance gain. It is better than the GIST in the first, third, fourth, and fifth places. However, our correct ratio ranked in the second place is 0.2% lower than the GIST.

Although our global descriptor only has slight performance gain than the GIST, it has an important advantage: the feature detection process is done in an integrated way. Therefore, we do not need to detect local features for the fine localization.

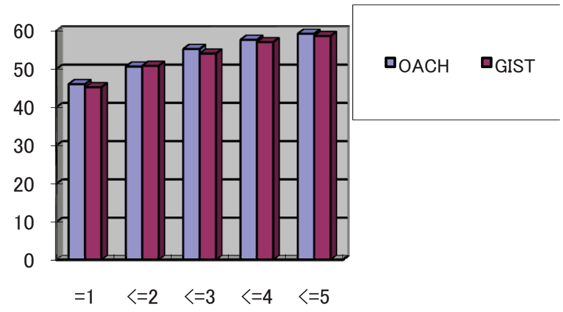


Figure 6. The performance comparison tested on the Holidays dataset. The figure shows the percentage of correctly retrieved images ranked in top 5 locations. We compare the performance of our global feature OACH and the GIST feature [13] on this dataset.

6.5 Time Complexity

Computation of OACHs: The time complexity of the OACH is $O(n_1n_2)$ (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. However, its performance was significantly improved even with small number of bins. The small number of bins is important for the robustness against rotations. It takes about 0.02 seconds to compute an OACH in a 640 × 480 image.

Time for Localization: In [21], 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 1 shows the time comparison between the proposed approach and the approaches in [21] and in [8]. It is clear that the proposed approach is more efficient.

Methods/Stage	Term	Coarse	Fine	Total
The Method [24]				0.4
The method [21]	0.018	0.002	0.024	0.044
Our Approach	0	0.001	0.02	0.02

Table 1. The comparison of the average times used in the localization (seconds).

7. Conclusions and Future Work

We present an efficient and reliable vision-based topological localization approach by combining global and local features. The merit of the features used in this work is that they are detected simultaneously. The novel global feature, OACH, is successfully used in coarse localization and the performance has been improved. OACH can deal with viewpoint and illumination changes that bring difficulties to topological localization. It is also possible to use OACH in other applications such as content-based image retrieval.

8. References

[1] J. S. Beis and D. G. Lowe,"Shape indexing using approximate nearest-neighbor search in high-dimensional spaces," In IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1000-1006, 1997.

[2] A. J. Davison. "Real-Time Simultaneous Localisation and Mapping with a Single Camera", in Proc. of Int. Conf. on Computer Vision, pp. 1403-1410, 2003.

[3] G. N. DeSouza and A. C. Kak. "Vision for Mobile Robot Navigation: A Survey", IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(2), pp. 237-267, 2002.

[4] T. Goedeme, T. Tuytelaars, L. Van Gool. "Fast wide baseline matching for visual navigation", In IEEE Conf. on Computer Vision and Pattern Recognition, pp. 24- 29, 2004.

[5] C. Harris and M.J. Stephens. "A combined corner and edge detector", In Proc. of Alvey Vision Conference, pp. 147-152, 1988.

[6] H. Katsura, J. Miura, M. Hild, and Y. Shirai. A view-based outdoor navigation using object recognition robust to changes of weather and seasons. In Proc. of IEEE Intl. Conf. on Intelligent Robots and Systems, pp. 2974-2979. 2003.

[7] J. Kosecka and Fayin Li. "Vision based topological Markov localization", in Proc. of IEEE Intl. Conf. on Intelligent Robotics and Automation, pp. 1481-1486, 2004.

[8] J. Kosecka, F. Li, X. Yang: Global localization and relative positioning based on scale-invariant keypoints. Robotics and Autonomous Systems 52(1): 27-38 (2005).

[9] D. G. Lowe, ”Object recognition from local scale-invariant features”, in Proc. Int. Conf. Computer Vision, pp. 1150-1157, 1999.

[10] H.P. Luhn. The automatic creation of literature, IBM Journal of Research and Development. Vol.2, No.2, pp. 159-165, 1958.

[11] M. Brown and D. G. Lowe, ”Recognising panoramas,” in Proc. of Int. Conf. on Computer Vision, pp. 1218-25, 2003.

[12] K. Mikoljczyk and C. Schmid, ”Indexing based on scale-invariant features”, in Proc. of Int. Conf. on Computer Vision, pp. 525-531, 2003.

[13] A. Oliva, and A. Torralba. ”Building the Gist of a Scene: The Role of Global Image Features in Recognition” Visual Perception, Progress in Brain Research, vol 155. pp. 23-36, 2006.

[14] D. Robertson and R. Cipolla. An image-based system for urban navigation. In Proc. of British Machine Vision Conference, 2004.

[15] Y. Rubner, C. Tomasi, and L. J. Guibas. ”The Earth Mover’s Distance as a metric for image retrieval”, Int. Journal of Computer Vision, 40(2), pp. 99-121, 2000.

[16] S. Se, D. Lowe and J. Little, Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks, Int. Journal of Robotics Research, Vol. 21, No. 8, August 2002, pp. 735-758.

[17] S. Thrun. Learning metric-topological maps for indoor mobile robot navigation, Artificial Intelligence, Vol: 99, pp. 21-71, 1999.

[18] A. Torralba, K. P. Murphy, W. T. Freeman and M. Rubin. ”Contextbased vision system for place and object recognition”, in Proc. of Int. Conf. on Computer Vision, pp. 273-280, 2003.

[19] I. Ulrich and I. Nourbakhsh. ”Appearance-based place recognition for topological localization”, in Proc. Int. Conf. Robotics and Automation, pp. 1023-1029, 2000.

[20] J. Wang, R. Cipolla, H. Zha,. ”Vision-based global localization using a visual vocabulary”, Proc. of Int. Conf. on Robotics and Automation, pp. 4241 - 4246, 2005.

[21] J. Wang, H. Zha, R. Cipolla. Coarse-to-fine vision-based localization by indexing scale invariant features, IEEE Trans. on System, Man, Cybernetics, Part B, Vol. 36, Issue 2, pp: 413-422, 2006.

[22] J. Wolf, W. Burgard, and H. Burkhardt, ”Robust vision-based localization by combining an image retrieval system with Monte Carlo localization”, in IEEE Trans. on Robotics, Vol.21, No.2 pp. 208-216, April, 2005.

[23] Y. Yagi, K. Imai, K. Tsuji, M. Yachida, ”Iconic Memory-Based Omnidirectional Route Panorama Navigation”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.27, no.1, pp.78-87, January, 2005.

[24] W. Zhang and J. Kosecka. ”Localization based on Building Recognition”, in IEEE Workshop on Applications for Visually Impaired in CVPR, 2005.

[25] C. Valgren and A. J. Lilienthal. ”Incremental Spectral Clustering and Seasons: Appearance-Based Localization in Outdoor Environments”, Proc. of Int. Conf. on Robotics and Automation, pp. 1856 - 1861, 2008.

[26] G. D. Tipaldi, D. Meyer-Delius, M. Beinhofer and W. Burgard. “Lifelong Localization and Dynamic Map Estimation in Changing Environments”, RSS Workshop on Robots in Clutter, 2012.

[27] A. Ramisa, A. Tapus, D. Aldavert, R. Toledo and R. Mantaras. “Robust vision-based robot localization using combinations of local feature region detectors”, Autonomous Robots Journal, vol. 27, no. 4, pp. 373-385, 2009.

[28] K. Ni, A. Kannan, A. Criminisi, J. M. Winn. “Epitomic Location Recognition”, IEEE Trans. Pattern Analysis Machine Intelligence, vol. 31, no. 12, pp. 2158-2167, 2009.

[29] J. Wang, H. Zha, R. Cipolla. "Efficient Topological Localization Using Orientation Adjacency Coherence Histograms”, ICPR (2) 2006: 271-274.

[30] J. Wang, R. Cipolla, H. Zha. "Image-base localization using scale invariant features", in Proc. of IEEE Int. Conf. on Robotics and Biomimetics 2004.