Scene Change Detection and Topological Map Construction Using Omnidirectional Image Sequences

Huei-Yung Lin, Yu-Hsiang Lin, and Jia-Wei Yao
Department of Electrical Engineering
Advanced Institute of Manufacturing with High-Tech Innovation
National Chung Cheng University
168 University Rd., Min-Hsiung, Chiayi 621, Taiwan
hylin@ccu.edu.tw, iopp50@gmail.com, B9515002@mail.ntust.edu.tw

Abstract

Visual place recognition for mobile robot localization is an important problem in robotics and computer vision research. In this paper, we present a novel scene recognition technique that resembles human cognition based on the scene change observation. Our semantic scene descriptor, Extend-HCT, is based on the SURF features. We build a scene change judgement system by analyzing the parameters of Extend-HCT codewords. We design an algorithm to mark the interested regions and construct the topological map after the autonomous mobile robot passing through an unknown path. It can then be used to assist for the place recognition, localization and navigation tasks if the robot goes through the environment in the future. The experimental results for the indoor and outdoor scenes are presented.

1 Introduction

Place recognition for mobile robot localization is an active research topic in robotics. Scientists and engineers have designed many kinds of sensors in order to enable robots to explore the unknown environment, let robots perceive the outside world, and understand the relationship between their locations and obstacles. Based on the current sensor technologies, researchers can use a variety of sensors to develop key algorithms to achieve their objectives. Some important tasks are creating a map for an unknown environment, sensing the obstacle information in the environment, or asking the robot to arrive an assigned destination, etc. [13]. In the literature, most studies adopt traditional distance sensors to deal with these problems. The commonly used techniques include sonar, laser range finder, odometer, and image-based approaches, etc.

To let autonomous mobile robots have the ability to navigate and recognize the environment, the robot must be able to label the areas of the environment automatically. Scene change detection is a newly explored approach for the robotics community to accomplish this task. It has been used in the field of multimedia, e.g. for the purpose of image compression by continuous image segmentation [2]. The related issues are also investigated in mobile robotics to mark the important locations of the environment and create a topological map [3]. Some researchers installed the camera on an unmanned aerial vehicle (UAV) and used optical flow method to detect scene changes and construct the topological map [5].

For the application of visual place recognition, using the catadioptric camera with a full 360-degree field of view is one of the most suitable methods to acquire the rich information from the surrounding environment [12]. In this work, we install an omnidirectional vision system on the mobile robot, and use the recorded panoramic images for place recognition and localization. In the past few years, this problem has attracted considerable attention in the field of computer vision and robotics [6, 9]. Some feature descriptions such as SIFT [4], SURF [1], or the color information [8, 10] are used by researchers to describe and construct the visual scenes.

This work presents a scene change detection and topological map construction technique using omnidirectional image sequences. The visual information associated with the environment is obtained from a panoramic camera device mounted on a mobile robot platform. Based on the previous work using the multiple convex hull scene description [11], we develop a data analysis system to detect the dramatic scene changes in the environment. The mobile robot is then able to navigate in an unknown path while construct the topological map according to the scene change nodes recorded in the database.

In this paper, we describe the important scenes by Extended Hull Census Transform (Extended-HCT). It is used to transform the images to a series of binary codes for representation. This method is developed for scene change detection from omnidirectional images. Since the lighting change is one major issue for scene change detection, we design an algorithm which can generate a similar descriptor at the same scene with different illumination conditions in the environment. We have shown that our method is robust and can adjust to the environment in a variety of fast and large illumination changes. Consequently, the topological map can be constructed using an omnidirectional vision system, and then used to facilitate the place recognition during the mobile robot navigation. The experimental results are presented for both the indoor and outdoor scenes.

2 Extended-HCT and Change Detection

If a mobile robot needs to perform the localization and scene recognition tasks at the same time, it usually adopts the topological map to locate its current position. Thus, it is an intuitive way to solve the localization problem by combining the visual recognition system and the topological map. This paper uses the semantic descriptor named “Extended-HCT” to establish a topological map based on the hull census transform (HCT) [11]. The proposed method aims to rep-

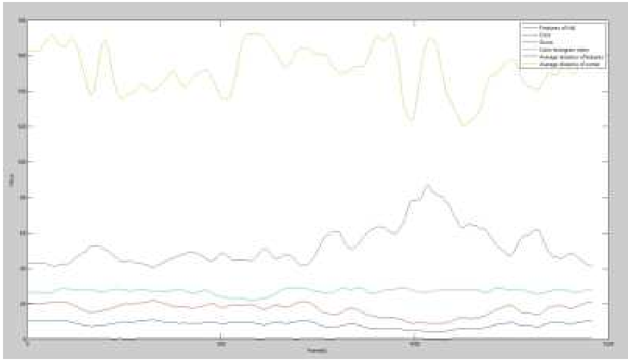


Figure 1. Scene descriptor with 6 parameters.

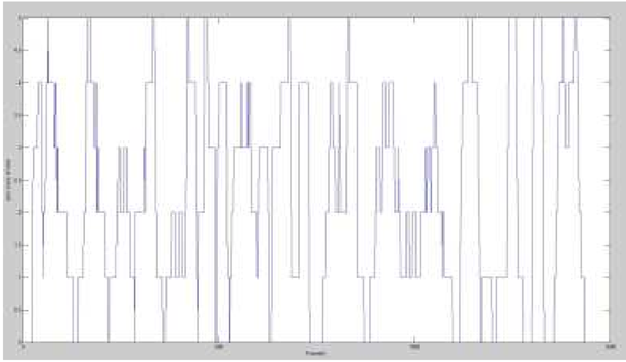


Figure 2. Zero-crossing.

resent the visual places, its features are quite similar to the descriptors used in the HCT. It can adapt to the fast and drastic illumination change, and is ideal for the images captured from a catadioptric vision system. In addition, Extended-HCT includes the color information of the environment and the structure information of the convex hull feature points. Thus, it is more suitable to describe the scene of visual recognition compared to the original HCT.

Hull Census Transform is an important step in the preprocessing of Extended-HCT. The convex hull of a set of points in the Euclidean plane is the smallest convex set that contains the set of points. When the set is a non-empty finite subset of the plane, a convex polygon is formed by connecting the elements in the subset. The important features in the image are extracted and processed with the hull census transform. The descriptor can represent the image using more structured information without losing the original characteristics of the features [11].

For the Extended-HCT, the visual codewords are adapted to the fast moving scenes, especially when exploring the unknown environment. The most critical breakthrough in our approach is that the feature coding vectors are derived based on the relationship between the feature points. It is able to tolerate the lighting changes without adjusting the parameter settings. We use six kinds of parameters to describe the images, including the relationship between the feature vectors, the structured relation among the feature points in the image, and the color histogram indexing information associated with the environment.

The detailed descriptions of the parameters are given as follows:

Features of Hull: The SURF features are first extracted from the omnidirectional images. This parameter represents the number of feature points for every layer of the convex hulls after performing the HCT.

Cost: Each layer of the HCT codes is represented by a binary string. The cost is given by the differences value of the HCT codes of two consecutive image frames.

Score: The score value is associated with the decimal format in its ranking number of bits converted from the HCT binary codes. It is defined by the union of the norm of HCT code length.

Color Histogram Index: The color information of the environment is represented by the green, red, blue and gray channels. This parameter is used for counting the associated color histogram indexing [10].

Structural Relation: Two parameters represent the structural relation among the features. For each layer of convex hull in an image, one is the centroid of the feature point locations and the other is associated with the total distance between any two feature point locations.

The continuous omnidirectional image sequence is used to perform the scene change detection. Since the consecutive images captured at the nearby locations have very similar feature characteristics, their Extended-HCT codes will also be highly correlated. Strictly speaking, the parameters of the Extended-HCT codes as described previously will change gradually for the continuously captured images. With this property, we can then analyze the information provided by the Extended-HCT codes to derive the scene change of the environment.

In the implementation, the SURF features are extracted for Extended-HCT codes. The SURF descriptor has the advantages that it is invariant to the rotation, scaling, and brightness changes. Furthermore, it also remains the high stability under viewpoint changes and the affine transformation. Thus, the detected locations of the SURF feature extraction are very close for two similar images.

To detect the scene change locations, a series of filtering process is carried out on the diagrams of the six kinds of descriptors. As illustrated in Figure 1, there exist close relations and high dependencies among the parameters, such as some peaks and valleys appeared at the nearby locations. Since they are the results from the Extended-HCT, we can analyze the common peaks and valleys to derive the images with significant scene changes. Due to the lack of equations for parameter descriptions, it is not possible to identify the peak and valley locations by differentiation. Thus, the convolution mask

$$G_m = [-1, 0, 1]$$

is used to detect the gradient changes of the parametric functions. By checking the zero-crossings, some positions with extreme values can be obtained.

It is clear that the number of scene changes depends on the criteria for image content evaluation. Moreover,

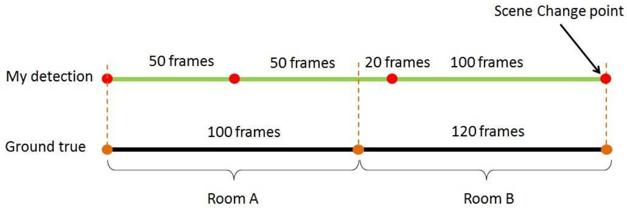


Figure 3. Evaluation for system accuracy.

the peaks and valleys will not appear at exactly the same locations in general. To make the zero-crossing on the parametric descriptors for scene change detection feasible, we employ a sliding window to search the extrema of the descriptors within a fixed time interval. First, the image frame is marked if there exists a zero-crossing for any parameters in the sliding window. They are then superimposed to generate a single representation as illustrated in Figure 2. Finally, the scene change frame is selected for the values greater than a preset threshold.

3 Experimental Results

In the experiments, we demonstrate the scene change detection and topological map construction for the indoor and outdoor scenes. To evaluate the performance of the proposed system, the detected scene changes are compared to the manually marked groundtruth. The scene change frames are used to partition the groundtruth to several regions. In each region, a positive frame is defined as those belong to the largest number of consecutive frames. The system accuracy is then given by the percentage of positive frames in terms of total frames as illustrated in Figure 3.

For the indoor scenes, the COLD dataset [7] is used for performance evaluation. Three different robot platforms with two heterogeneous cameras (catadioptric and perspective ones) are used to gather the image data under varying conditions and times in different environments. These videos are also captured under human motions and different weather conditions (e.g. cloudy and night). The dataset is useful for testing the visual place recognition algorithms because the three different environments have similar rooms such as print rooms, one-person offices, etc.

Figures 4 and 5 show the topological maps constructed from the COLD dataset images captured from the sequence "Freiburg Part A". The connections between two colored curves indicate the scene change locations. Although the scene change locations do not totally agree with the human observation, they are fairly consistent for these two experiments. The system accuracy is shown in Figure 6, which is better than the results reported in the literature [7].

We have also compared the experimental results of the scene change detection using Extended-HCT and HCT. Figures 7(a) and 7(a) show the results of the COLD-Freiburg Part A night 1 sequence for Extended-HCT and HCT, respectively. The topological maps are constructed using the same color segment to represent the same scene and the connection point between different color segments to indicate the scene change location. In the proposed method, the images are transformed to 29 parameters by Extended-HCT (5 layers of convex hull and 6 kinds of parameters). On the other hand, HCT uses only 6 parameters (3 layers of convex hull and 2 kinds of parameters). Thus, in this work we get much richer information than what is used in HCT, including many structural relationships of feature points and the color information of the environment. If the extraction of feature points are too few to obtain enough convex hulls, then the proposed technique has better tolerance for the environment change than HCT, like the two rooms on the lower left corner in Figure 7. The performance comparison for the proposed method and HCT is shown in Figure 8. A significant improvement can be seen for the night scene sequence.

For the outdoor scene experiments, the omnidirectional camera is mounted on top of an SUV and moves around the university campus. The location of the SUV is recorded by the GPS to generate a metric topological map. Similar to the indoor experiments, 6 kinds of parameters are used for Extended-HCT and scene change detection. Figure 9 shows the generated topological map overlaid on a Google map where the red

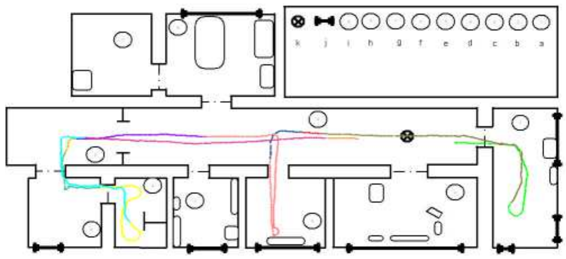


Figure 4. COLD-Freiburg Part A seq1 sunny2 (accuracy: 71.04%, positive frames: 1075, total frames: 1517).

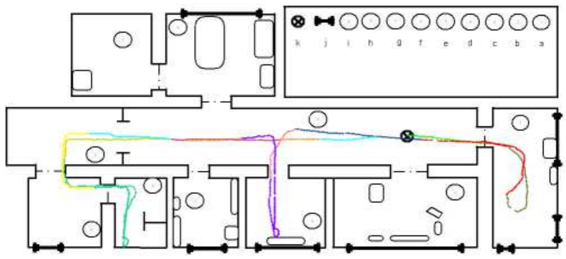


Figure 5. COLD-Freiburg Part A seq1 cloudy1 (accuracy: 67.16%, positive frames: 980, total frames: 1459).

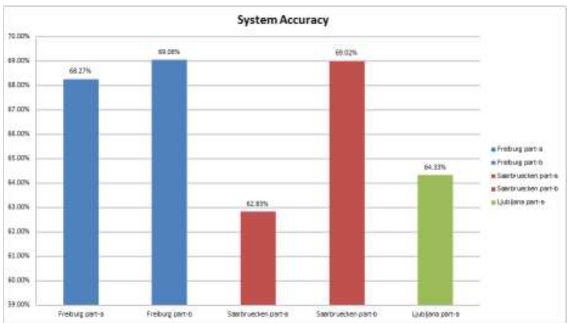
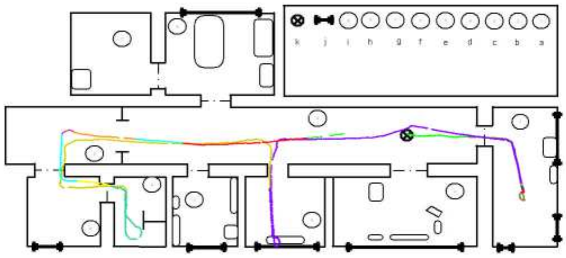
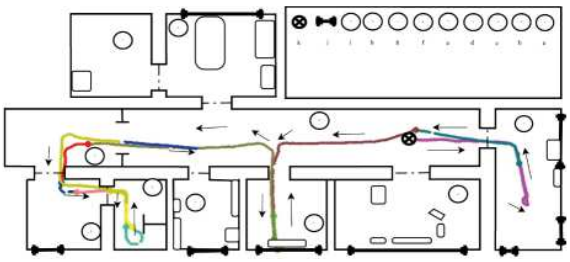


Figure 6. System accuracy.

Figure 9. Topological map of the outdoor scene.



(a) COLD-Freiburg Part A night1 Accuracy: 73.30% (the experiment result in this paper).



(b) COLD-Freiburg Part A night1 Accuracy: 64.45% (the experiment result using only HCT)

Figure 7. The result of comparison analysis in night path1.

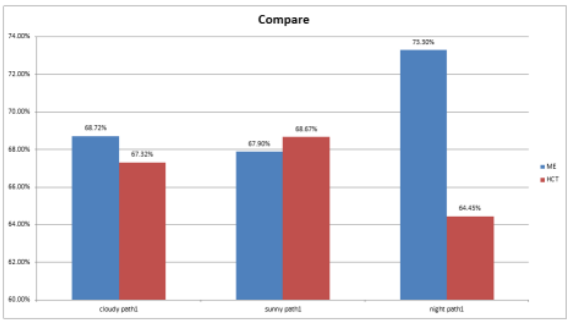


Figure 8. Accuracy comparison.

lines are the navigation path of the SUV. The markers on the map indicate the scene change locations derived from the omnidirectional images. Due to the NMEA and GPS receiver errors, we can see some parts of the red lines are not on the road correctly. The problem is out of scope of this study and will not be further investigated.

4 Conclusion

This paper presents a new technique for scene change detection and topological map construction. It is an approach closely related to the human perception for the scene change detection. We built a scene change detection model based on Extend-HCT codewords, and analyzed the generated change of the parameter of codewords corresponding to environment. The experiments are carried for both the indoor and outdoor scenes.

References

[1] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.

[2] U. Gargi, R. Kasturi, and S. H. Strayer. Performance characterization of video-shot-change detection methods. *IEEE Trans. Cir. and Sys. for Video Technol.*, 10(1):1–13, Feb. 2000.

[3] M. Liu, D. Scaramuzza, C. Pradalier, R. Siegwart, and Q. Chen. Scene recognition with omnidirectional vision for topological map using lightweight adaptive descriptors. In *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, IROS'09, pages 116–121, 2009.

[4] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.

[5] N. Nourani-Vatani and C. Pradalier. Scene change detection for vision-based topological mapping and localization. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 3792 –3797, oct. 2010.

[6] A. Pronobis and B. Caputo. Confidence-based cue integration for visual place recognition. In *IROS*, pages 2394–2401. IEEE, 2007.

[7] A. Pronobis and B. Caputo. Cold: The cosy localization database. *Int. J. Rob. Res.*, 28(5):588–594, May 2009.

[8] M. J. Swain and D. H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7(1):11–32, Nov. 1991.

[9] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin. Context-based vision system for place and object recognition. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 273–, 2003.

[10] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference on*, volume 2, pages 1023 –1029 vol.2, 2000.

[11] M.-L. Wang and H.-Y. Lin. A hull census transform for scene change detection and recognition towards topological map building. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 548 –553, oct. 2010.

[12] Y. Yagi. Omnidirectional sensing and its applications. *IEICE Trans, on Information and Systems*, E82-D(3):568–579, 1999.

[13] J.-H. Zhou and H.-Y. Lin. A self-localization and path planning technique for mobile robot navigation. In *Intelligent Control and Automation (WCICA), 2011 9th World Congress on*, pages 694–699, June.