Advanced Science and Technology Letters
Vol.85 (Information Technology and Computer Science 2015), pp.46-50
http://dx.doi.org/10.14257/astl.2015.85.11

Robot localization method based on visual features and their geometric relationship

Sangyun Lee[1] , Changkyung Eem[2], and Hyunki Hong[3]

[1] Department of Imaging Science and Arts, GSAIM, Chung-Ang University, Seoul, Korea
[2] College of ICT Engineering, Chung-Ang University, Seoul, Korea
[3] School of Integrative Engineering, Chung-Ang University, Seoul, Korea

Abstract. This paper presents a novel method to recognize current location of a mobile robot from input stereo images with visual and geometric features. We extract structural planes from 3D depth data by using SLIC (Simple Linear Iterative Clustering) based superpixel and RANSAC algorithm. The experimental results show that the proposed method using visual features and their geometric relationship provides better performance in robot localization.

Keywords: Mobile Robot localization, visual feature, structural plane

1 Introduction

Vision based robot localization and navigation have been extensively studied in robotics and computer vision [1][2]. Localization to estimate a precise robot position is a basic requirement for robotic application. This capacity in complex environments relies on a map which can be either given to the robot, or learned while the robot discovers its surroundings.

The retrieval methods that categorizes the bag of words based images or instantiates the vocabulary tree scheme have been used widely in scene recognition from an image sequence [3][4]. Most localization and map-learning systems employ a visual word, carrying any kind of appearance information in feature space such as color patch or KLT feature. However, human beings recognize their locations by using visual features as well as the geometric positions. This paper introduces a novel localization method based on visual and geometric features. The proposed method is applicable in both a stereo camera and a kinect sensor for 3D data acquisition in robot navigation.

At first, we detect visual features and describe their neighbor intensity distribution by using BRISK (Binary Robust Invariant Scalable Keypoints)[5]. BRISK descriptor of the feature is employed to determine efficiently whether the features of the request are matched with the key points of the model or not. Then, we appply SLIC (Simple Linear Iterative Clustering) based superpixel method [6] to an input depth map and then 3D plane patch is extracted. RANSAC (Random Sample Consensus) is used to compute surface normal information of the segmented plane patch.

ISSN: 2287-1233 ASTL

We assume that the surrounding environment where the robot drives is constructed generally with many structural planes such as walls and ceilings. In man-made indoor environment, the image patch is a small region with appearance features and the scene is represented with a collection of patches on the structural plane. In this paper, plane patches are represented with graph-based data structure and we combine these patches to determine the structural plane[7]. Geometric features are generated from the structural plane information, and visual features belonging to each plane are selected as scene index words for database in a learning stage. Visual and geometric features are extracted from a query image and then FLANN (Fast library for performing Approximate nearest Neighbor)[8] search is proceeded in a search stage. Once the image candidates for database matched to the input image are generated, we examine features on the structural planes extracted from each candidate to improve the localization performances.

2 Proposed Method

By applying RANSAC approach to 3D depth map of the stereo images, we can obtain the planes where many 3D points are distributed. RANSAC has been popular in regression problem with samples contaminated with outliers, but reconstruction performance of RANSAC is much highly affected by the complicated scene elements and the camera viewpoint. By applying SLIC-based superpixel to 3D depth map directly, we can segment 3D plane patches precisely. Then RANSAC algorithm is employed to each 3D patch to derive a reliable plane equation, excluding unwanted effects of outliers. Every patch is represented as a graph data structure, in which the vertex denotes a normal vector of each plane and the edge does the angle between adjacent plane patches.

(a) (b) (c)

Fig. 1. (a) Input image, (b) segmented superpixel patches in 3D depth map, (c) segmented structural planes.

Because 3D patch segmentation of SLIC-based superpixel has high computational load, the proposed method make sub-sampling 3D data of input depth map for computation efficiency. A distance value between centroids of the neighboring two planes and an angle difference between two planes are computed. Then we obtain a sum of relative weighted distance and angle difference. Figure 1 shows the structural planes obtained in hierarchical merging process with 3D plane patches, which are represented with graph-based data structure. An input scene is segmented totally into

50 plane patches and 10 structural planes are obtained (Fig. 1. (c)). Here a structural plane is represented with a unique color.

In this paper, visual features on 3D structure planes and their geometric relationship are used as index words for scene location recognition. Then the best-bin-first method in FLANN about the query image is employed to search the database. The proposed method considers only the case that the number of feature points of the query image matched to database image is larger than the threshold. Here it experimentally set as 10. In a search stage, we compute firstly the ratio of the total number of the features detected from each candidate to that of the matched feature points. Then the geometric structure relationship between the keypoints is computed. For further details, we obtain geometric relationship that is measured with the cross product of a normal vector of the structural plane of the matched candidate and that of the query image. The candidate with the minimum cost value is determined as final matched scene.

3 Experimental Results and Conclusion

The computational equipment includes a PC with Intel Core (TM) i7 3.08GHz and 8GB RAM with Nvidia GTX780 graphics card. Stereo images are captured by a Bumblebee 3 from Point Grey Inc. at a rate of 15 frames per second (fps). There are 200 indoor scene images composed of color and 3D depth map in the robot navigation environment that are learned and matched to tell where the robot is located (Fig. 2).

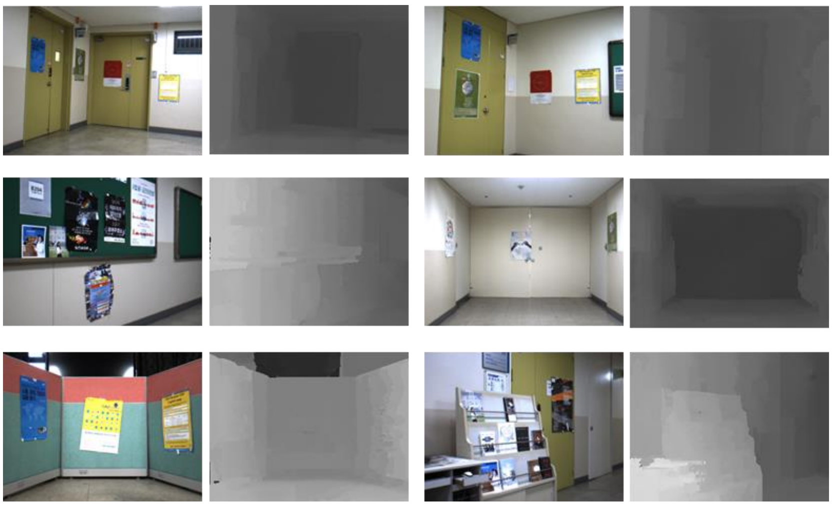

Fig. 2. Database image (color and 3D depth map).

In input scenes (Fig. 2), we compare recognition performances of the previous method [4] and those of the proposed system as Fig. 8. The performance of any classification system is subject to a tradeoff between the rate of misdetections (false negatives) and the rate of false detections (false positives). In general, a lower threshold yields higher detection rates and higher false positive rates. A receiver operator characteristic (ROC) curve indicates the error rate for misdetection (or the true positive rate) against the false alarm rate. Figures 3 demonstrates the performances of two methods according to the final matching scores.

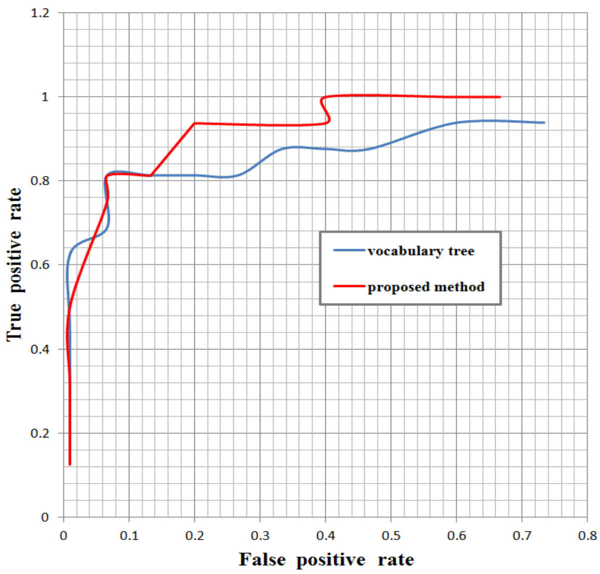

Fig. 3. ROC curves of two methods.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2013R1A1A2008953).

References

1. Cummins, M., and Newman, P.: FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. Journal of Robotics Research, 27 (6), pp. 647--665 (2008).
2. Zikovic, Z., Bakker, B., Krose, B.: Hierarchical Map Building Using Visual Landmarks and Geometric Constraints. In: IEEE Intelligent Robots and Systems, pp. 2480—2485 (2005).
3. Filliat. D.: A Visual Bag of Words Method for Interactive Qualitative Localization and Mapping. In: 2007 IEEE International Conference on Robotics and Automation, pp. 3921--3926 (2007)
4. Nister, D., Stewenius, H.: Scalable Recognition with a Vocabulary Tree. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 2, pp. 2161--2168 (2006)
5. Leutenegger, S., Chli, M., Siegwart, R. Y.: BRISK: Binary Robust Invariant Scalable Keypoints. In: International Conference on Computer Vision, pp. 2548-2555 (2011)

6. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels Compared to State-of-the-art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34 (11), pp. 2274-2282 (2012).
7. Wang, T., Chen, L., Chen, Q.: A Graph-based Plane Segmentation Approach for Noisy Point Clouds. In: Control and Decision Conference, pp. 3770--3775 (2013)
8. Muja, M., Lowe, D. G.: Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In: Int'l. Conf. on Computer Vision Theory and Applications, pp. 331-340 (2009)

