The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

3D Localization Based on Visual Odometry and Landmark Recognition Using Image Edge Points

Masahiro Tomono

Abstract— **This paper presents a method of 3D localization using image edge-points detected from binocular stereo image sequences. The proposed method calculates camera poses using visual odometry, and updates the poses by reducing the accumulated errors using landmark recognition. Landmark recognition is done based on robust and scalable image-retrieval using image edge-points with SIFT descriptors and a vocabulary tree. A randomized-ICP algorithm is employed to accurately estimate the 6-DOF camera pose from a landmark image and an edge-point based 3D map. Experiments show our edge-point based approach outperforms approaches using corners and Laplacian points.**

I. Introduction

Mobile robot localization is indispensable for most robotic tasks. Recently, highly distinctive image features [13] and efficient image-retrieval techniques [14] have been applied to mobile robot localization [22], [7], [5], [3]. Since images have a great amount of information on place and location, the image-based approach is applicable to localization on a topological map, which requires less memory and is suitable for representing a large scale environment. On the other hand, complicated robotic tasks such as obstacle avoidance and object manipulation require the accurate shape and pose (location and orientation) of the objects in the environment. A dense metric map is prerequisite to represent such precise information. The combination of image-based localization and dense metric map will be useful for enhancing robot capabilities.

This paper presents a method of 3D localization on a detailed metric map using a highly scalable image-retrieval technique. With only a stereo camera, the method estimates the robot pose based on visual odometry and 3D pose estimation using landmark images and a 3D map. The 3D map is created using a stereo SLAM scheme [20], which utilizes image edge points to represent detailed object shape. The visual odometry is implemented by revising this scheme. The bag-of-words approach is employed for efficient image retrieval, and a highly scalable image database is implemented using a vocabulary tree [14].

The contribution of this paper is to provide scalable, quantitative 3D localization by integrating a vocabulary tree and 3D pose estimation on an edge-point based map. The proposed method is applicable to non-textured environments since edge points can be extracted even from such environments. Experiments show the edge-point based approach outperforms approaches using sparse features such as corners and Laplacian points. Furthermore, edge-point based maps are more suitable than sparse features for obstacle avoidance and object manipulation since edge points can represent detailed object shape.

The procedure of our 3D localization is as follows. A 3D map is built beforehand from stereo images [20], and keyframes in the stereo images are registered as landmarks in an image database.

(1) Initial pose estimation
At the beginning of navigation, the initial pose is obtained by global localization using the image-retrieval scheme.

(2) Dead-reckoning using visual odometry
During the navigation, the visual odometry consecutively estimates robot poses from captured stereo images.

(3) 3D localization using landmark recognition
To reduce the accumulated errors in visual odometry estimates, the system detects landmarks and updates the robot pose using the landmarks and the 3D map.

By repeating step (2) and (3), the system maintains the robot pose during the navigation. If the robot gets lost for some reasons, global localization is performed as step (1). The scheme should be implemented on a Bayes filter such as Kalman filter and particle filter, but for simplicity, this paper does not consider this issue.

There are several differences between the edge-point based approach and the well-known corner-point based one since edge points are less distinctive than corner points. For example, as mentioned in Section IV, geometric consistency check is based on similarity transformation instead of epipolar geometry with RANSAC [6]. The camera pose estimation is based on a randomized ICP algorithm in connection with RANSAC-based estimation. Degeneracy check is also performed to avoid the false matches due to aperture problems in edge-based matching. The number of edge points is approximately ten-fold larger than that of corner points, but the system does not suffer from it in image retrieval due to the bag-of-words approach, which classifies the edge point features into a fixed number of visual words. Each image is represented by the weighted sum of visual words, and memory consumption and retrieval efficiency is not much affected by the number of features.

II. Related Work

SLAM using stereo vision has been studied intensively in the last decade [4], [8], [16], [10]. Most of them utilizes corner-point features and build sparse feature maps. Some

M. Tomono is with Future Robotics Technology Center, Chiba Institute of Technology, Narashino, Chiba 275-0016, Japan. tomono@furo.org

978-1-4244-6676-4/10/$25.00 ©2010 IEEE

of them provide global localization capability using SIFT descriptors [16], but the image retrieval scheme is not very scalable.

Recently, visual vocabulary [14], [17] has been applied to mobile robot localization. Global localization using visual vocabulary was introduced by [22]. The visual vocabulary approach has also been applied to topological mapping and localization [5], and very large-scale topological SLAM [3]. The vocabulary tree [14] provides highly scalable image retrieval capability (more than million images), and was utilized for topological mapping and localization [7].

The vocabulary tree approach is also employed for localization on metric maps. Visual odometry with global localization using a vocabulary tree was developed by [23] with two stereo cameras and an IMU. SLAM with global localization using a vocabulary tree was developed by [11] with a stereo camera. These systems utilize corner points, and may be hard to work in non-textured environments. As mentioned in Section V, we found edge points provides better results than corners. Also, the corner-based approach generates sparse maps, and cannot represent detailed shape of the objects in the environment.

Image edge points have been utilized for stereo SLAM and object recognition in our previous work [19], [20]. The edge-based stereo SLAM [20] is utilized to build 3D maps in this paper, but the vocabulary tree is newly integrated with the 3D mapping scheme for landmark. The framework of edge-based recognition and pose estimation in this paper is based on the previous work [19], but accuracy in 3D pose estimation is improved by randomized ICP and scalability in image retrieval is enhanced with the vocabulary tree.

III. Mapping

A. 3D Mapping Based on Edge Point ICP

A 3D map is built based on the method proposed by [20]. We briefly review the method.

1) Stereo Reconstruction: The method utilizes image edge points detected by the Canny detector [2]. Note that edge points can be obtained from not only long segments but also fine textures. We refer to a pair of left and right images as *stereo frame* (*frame*, for short). Intra-frame reconstruction (i.e., between the left and right images) is performed based on the epipolar geometry in parallel stereo. We search the matching pair of edge points between left and right images along the scanline since epipolar lines are horizontal for parallel binocular stereo cameras. The matching criterion is the normalized correlation of a small window around the edge point. Also, the orientation of the image gradient at the edge point is optionally used to reduce outliers. Multiple candidate matches are inevitably obtained especially when the edge direction is nearly parallel with the epipolar line. We employ the DP matching approach [15] to address this problem. The 3D edge point $P_c = (X, Y, Z)^T$ is calculated from point $(x_l, y_l)^T$ on the left image and point $(x_r, y_r)^T$ on the right image based on the parallel stereo formula.

2) Camera Motion Estimation and Map Building: The camera motion from time $t-1$ to t is estimated by matching the edge points in frame I_{t-1} and those in frame I_t. Our method employs 3D-2D matching, in which the 3D points reconstructed from I_{t-1} are matched with the 2D points detected in I_t. The registration is performed using a variant of ICP algorithm [1] on the image plane. Let r_t be the camera pose at t, P^i_{t-1} be the i-th 3D edge point reconstructed at $t-1$, and p^i_{t-1} be the projected point of P^i_{t-1} onto image I_t. Let q^i_t be the image edge point at t which corresponds to p^i_{t-1}. A cost function F is defined as follows.

$$F(r_t) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} d(q^i_t, p^i_{t-1}) \qquad (1)$$

Here, $d(q^i_t, p^i_{t-1})$ is the perpendicular distance between p^i_{t-1} and the edge segment on which q^i_t lies.

Camera pose r_t and edge point correspondences are searched by minimizing $F(r_t)$ using the ICP. The initial value of r_t is set to r_{t-1}, and the initial correspondence q^i_t of p^i_{t-1} is set to the edge point which is the closest to p^i_{t-1} in terms of Euclidean distance. By repeating the minimization of $F(r_t)$ and edge point matching, the optimal r_t and edge point correspondences are obtained. In this process, outliers are coped with by a robust cost function [9].

Based on the obtained camera pose, a 3D map is built by transforming the intra-frame 3D points from the camera coordinate system to the world coordinate system.

B. Landmark Image Database

Landmark images are sampled from the images used in 3D mapping and are stored in a landmark image database in order to retrieve them at the localization phase. Landmark data is a tuple (I_n, E_n, r_n), where I_n is the n-th landmark image, E_n is the set of edge points detected in I_n, and r_n is the camera pose from which I_n was captured. r_n is estimated in the 3D mapping process. Each edge point in E_n has a SIFT descriptor [13].

We detect edge points using the Canny detector [2] with multiple scales. Then, the scale-space analysis is performed in order to make edge points invariant to scale change [12]. Note the analysis is not performed in the stereo SLAM because it is time consuming. Using the scale-invariant edge points, edge-point matching is performed robustly even when the object size in the input image is different from that in the landmark image. Each edge point has a quadruple (x, y, θ, s), that is, location, orientation, and scale.

The vocabulary tree [14] is employed to build an image database. The vocabulary tree is a tree structure for storing visual words (quantized descriptor vectors) efficiently based on hierarchical k-means clustering, and can provide highly scalable image-retrieval. Each node in the tree corresponds to a visual word and has an inverted file to store the identifiers of the images that contain the features corresponding to the visual word.

We first create a generic vocabulary tree from edge points detected from generic images (2M points from 200 images in the experiment). This generic tree is a skeleton, in which no images are stored yet. Then, we store landmark images to the vocabulary tree by adding the image index $(I_j, f_{j,k})$ to the inverted files. Here, I_j is the identifier of the j-th image, and $f_{j,k}$ is the frequency of a visual word v_k in I_j. Creating a generic vocabulary tree is time consuming, but storing a new image to the tree is very efficient since it just updates the inverted files.

IV. Localization

A. Visual Odometry

Our visual odometry is basically the same with the stereo SLAM above. Only the difference is that no global maps are generated by the visual odometry to reduce memory consumption. The purpose of the visual odometry is camera motion estimation, and local maps are necessary for it. Camera motion estimation is performed using a local map which is created by integrating 3D points from multiple frames since 3D points reconstructed from one stereo frame can have large errors.

B. Landmark recognition by Image Retrieval

Some images in the image sequence for the visual odometry are used for landmark recognition. We refer to these images as query images. Our landmark-based localization can be performed using monocular images, and either left or right image in a stereo frame is utilized. First, Edge points are detected from a query image I_n in the manner mentioned in III-B. Then, the landmark images matched with I_n are retrieved from the vocabulary tree based on the TF-IDF (Term Frequency-Inverse Document Frequency) scoring scheme [17], [14]. The images with top M scores are employed as candidates ($M = 10$ in the experiment).

C. Pruning Bad Images

The vocabulary tree provides candidate images very efficiently from a large scale image database, but the candidates include false positives. We remove them to select a good image for 3D pose estimation.

1) Geometric Consistency: Each retrieved image contains many false edge-point matches. The false matches can be removed based on the geometric consistency that the query image and the landmark image must have similar layouts. The epipolar geometry with RANSAC for outlier rejection is widely used to check the consistency between two images. However, this method is not suitable for edge points since edge points are less distinctive than corner points and many false matches can be obtained. The epipolar geometry check needs at least five point pairs and the RANSAC will not be efficient for such less distinctive edge points. Therefore, we employ the consistency check based on 2D similarity transformation [21]. We employ a voting scheme to find the similarity transformation parameters. In our scheme, each edge point has location (x, y), orientation θ, and scale s. Thus, we can calculate a set of similarity transformation parameters from one point pair between the two images. Then, we cast a vote for each point pair in the parameter space (translation, rotation, and scale ratio). The maximum in the voting space is employed as the best similarity transformation between the two images. This method is efficient since it is based on the voting by one point pair.

We define a score $S_1 = N_m/N_{all}$, where N_m is the number of edge points which meets the geometric consistency, and N_{all} is the number of all edge points in the query image. If S_1 is below a threshold, the image is removed from the candidate list.

2) Degeneracy Check: If the geometric layout in the image is too simple, pose estimation could suffer from degeneracy and no correct estimates will be obtained. Although there are many types of degeneracies, this paper considers the most typical case in 2D space. If the matched edge points lie on a line or parallel lines, a degeneracy occurs. In indoor environments, this sometimes occurs when the camera moves closely to a wall or large planar object with few textures. Note that sufficient corner-points cannot be detected in this case.

This type of degeneracy can be detected easily by examining the distribution of edge point orientations. We make the histogram of edge point orientations and find the peak of the histogram. We define a score $S_2 = N_d/N_m$, where N_d is the number of edge points having the same orientation as the peak orientation in the histogram and N_m was defined above. If S_2 exceeds a threshold, the image is removed from the candidate list.

The 2D degeneracy check eliminates many false matches, but other types of degeneracies in 3D-2D matching can theoretically occur. The detection of such degeneracies is future work.

D. Refinement of Edge Point Matching

After the best landmark image is selected based on S_1 and S_2, the edge point correspondences for the best image are refined. The edge points in the best image are transformed to the query image coordinate frame according to the similarity transformation parameters obtained in Section IV-C.1. Then, the candidate edge points in the query image that can be matched with an edge point e_j in the landmark image are searched within a small region around e_j using SIFT descriptors. This process obtains dense edge-point matches, which are useful for the ICP-based pose estimation to be mentioned below.

E. 3D Pose Estimation by Randomized ICP

To estimate the camera pose, we find edge point correspondences between the 3D map and the query image. This is done based on the edge point correspondences between the 3D map and the stereo image (landmark image) obtained by the stereo SLAM, and also based on the edge point correspondences between the landmark image and the query image obtained in the previous section. Then, we calculate the camera pose relative to the 3D map by minimizing the average reprojection errors of the 3D edge points onto the query image.

Outlier rejection is a crucial issue in this process. One solution is a RANSAC-based approach [19]. However, RANSAC sometimes suffers from large matching errors due to variation in random sampling especially when using less distinctive edge points. The ICP algorithm can provide more accurate pose estimation. However, the ICP used in the stereo SLAM is not suitable for localization since the camera pose of the query image can be distant from that of the landmark image. If we simply use the ICP, it will easily fall into local minima. Note that, in the stereo SLAM, the camera poses of two consecutive frames is very close and the ICP rarely falls into local minima. To address this problem, this paper proposes a randomized ICP approach, where randomized initial values are fed to the ICP to avoid falling into local minima. The procedure is as follows.

(1) Selection of initial pose
An initial camera pose r_t is obtained using the RANSAC-based approach [19].

(2) Creation of initial values to the ICP
Multiple initial values $\{b_k | b_k = r_t + d_k \quad (k = 1, 2, ...)\}$ are created, where d_k is a displacement sampled from a random distribution in 3D pose space.

(3) ICP execution
For each b_k, the ICP algorithm mentioned in Section III-A is performed.

(4) Selection of the best candidate
The candidate with the best matching score is selected from the results of step (3). The score here is defined as $S_3 = N_f / N_{all}$, where N_f is the number of matched edge points in the ICP.

In this procedure, if a good initial pose is not obtained by RANSAC in step (1), good candidates may not be obtained in step (3). To increase success rate, the procedure is repeated at most L times until finding a candidate pose whose score exceeds a threshold ($L = 10$ in implementation).

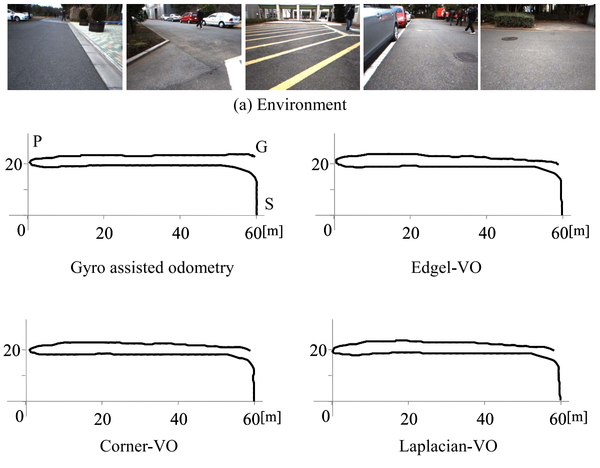
V. Experiments

We conducted experiments using Point Grey Research's binocular camera Bumblebee2. The baseline distance is 120 [mm]. The image size was reduced to 320×240 pixels. 3D localization was made using only the images captured by the left camera.

For comparison, we conducted the same experiments using other features: corners and Laplacian points. Corners were detected in the same way as the initial point selection in the KLT tracker [18]. Laplacian points were detected by finding local maxima of the trace of the Hessian matrix. The orientation of a corner or Laplacian point is necessary to compute rotation-invariant SIFT descriptors, and it was obtained by finding the mode in the histogram of gradients in the neighborhood of the feature point.

A. Visual Odometry

We evaluated visual odometry for each feature type in two environments. We refer to each visual odometry as edgel-VO, corner-VO and Laplacian-VO, respectively.


(a) Environment

Gyro assisted odometry

Edgel-VO

Corner-VO

Laplacian-VO

(b) Trajectories

Fig. 1. VO Trajectories in a textured environment.

TABLE I
VO errors at points P and G in Fig.1

	Edgel-VO	Corner-VO	Laplacian-VO
P	1.2[m]	2.0[m]	1.4[m]
G	3.3[m]	3.0[m]	3.6[m]

(1) Experiment in a textured environment

Images were captured with a stereo camera mounted on a mobile robot in an outdoor environment shown in Fig. 1(a). A number of edge points, corner points, and Laplacian points were detected from the road surface and vegetation under moderate illumination conditions on a cloudy day.

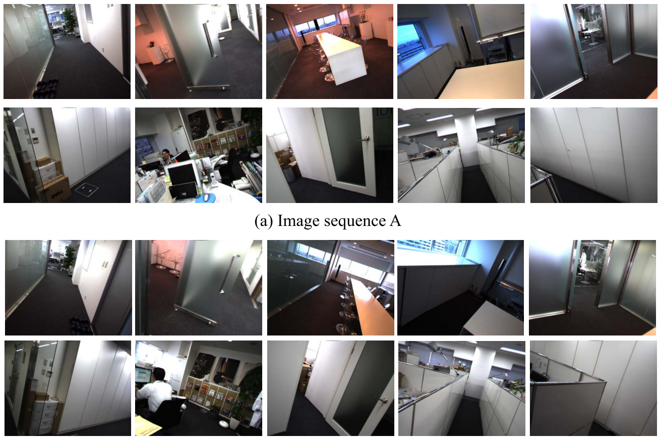
Fig. 1(b) shows the trajectories of each visual odometry and the gyro-assisted odometry (GO) of the mobile robot. The accuracy of GO is 1% in distance. Table I describes the errors of visual odometry measurements compared with GO's measurements at points P and G. There are no large differences between the trajectories due to plenty of textures and smooth motion by the robot.

(2) Experiment in an environment having non-textured areas

Two image sequences were captured manually with a stereo camera in an indoor environment which has non-textured areas. Fig. 2 shows the two image sequences, which we refer to as A and B. Image sequences A and B followed almost same paths, but image sequence A was taken one week after image sequence B.

Each VO was evaluated using image sequence A. Fig. 3 depicts the trajectory of each VO. There is no ground truth for this manual camera motion. Thus, as a reference, the top-left figure shows a 3D map and camera trajectory generated by our stereo SLAM from image sequence B. The 3D map is well-aligned and can be used as a reference to evaluate the trajectories qualitatively. Note the path of image sequence A is shorter than that of B.

As shown in the figures, edgel-VO generated a good trajectory. On the other hand, corner-VO and Laplacian-VO generated a distorted trajectory. The trajectories tend to



(a) Image sequence A

(b) Image sequence B

Fig. 2. Image sequences captured in the same environment under different conditions.

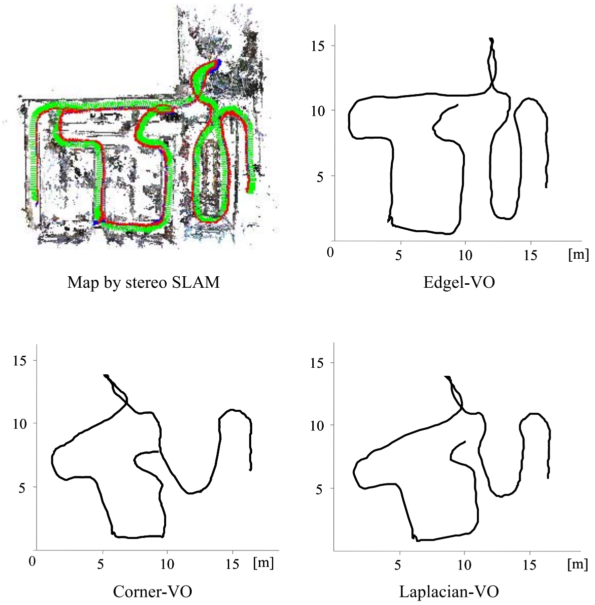
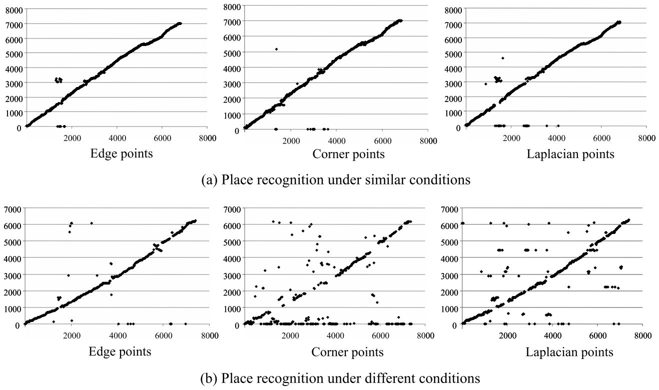


Fig. 3. Trajectories estimated by visual odometry.



(a) Place recognition under similar conditions

(b) Place recognition under different conditions

Fig. 4. Confusion matrices of image retrieval.

TABLE II

RECOGNITION RATES (#SUCCESS WITH TOP SCORE/#ALL IMAGES)

	Edge points	Corner points	Laplacian points
Similar condition	0.97	0.96	0.94
Different condition	0.93	0.51	0.78

largely distort when the camera turns in non-textured areas.

B. Place Recognition

Two experiments of place recognition were conducted in the indoor environment mentioned in Section V-A(2). Similarly to edge points, image databases were built for corners and Laplacians. For each feature type, scale space analysis was done and a vocabulary tree was created.

(1) Place recognition under the similar conditions

Two image sequences were captured manually in only textured areas of the environment. Note that these images were different from those in Section V-A(2). The second image sequence were captured just after the first image sequence, and they are very similar. An image database was created for each feature type from 357 images regularly sampled from the first image sequence (7140 images). 685 query images were regularly sampled from the second image sequence (6845 images)

Fig. 4(a) shows the confusion matrices created from image retrieval results. The horizontal axis is query image ID, and the vertical axis is database image ID which was retrieved with the top score. Since the two image sequences followed the same path, correct recognitions are on the diagonal in the graph. Image ID is -1 when no retrieval is obtained. Table II shows the recognition rates, which is the ratio of the number of successfully retrieved images with the top score to the total number of the query images. In this experiment, all the feature types provided good results since two image sequences were very similar and plenty of feature points were obtained from the textured areas.

(2) Place recognition under different conditions

This experiment was conducted using the images in Fig. 2. An image database was created for each feature type from 315 images regularly sampled from image sequence B (6300 images). 740 query images were regularly sampled from image sequence A (7400 images).

There are many differences in image sequences A and B. Image sequence A was taken in the evening and image sequence B was taken in the daytime. Therefore, their illumination conditions are different. Also, the camera was slanted in the opposite direction. Objects including people, chairs, a whiteboard and many boxes moved during the one week. Furthermore, due to manual manipulation, the camera viewpoints are slightly different in A and B, and some areas are not overlapped.

Fig. 4(b) shows the confusion matrices created from image retrieval results. Table II shows the recognition rates. The edge-point based recognition has the best recognition rate. False recognitions in the edge-point based recognition were caused by degeneracies mentioned in Section IV-C.2 and non-overlapped areas between the two image sequences. In addition to these factors, false recognitions in the corner-based recognition were caused by poverty of reliable feature points in non-textured areas.

C. Global 3D Localization

Experiments of global 3D localization were done using 3D maps and the images retrieved from the image database with the top score. A 3D map was created for each feature type using our stereo SLAM. No predictions based on motion models were used and localization were performed independently for all the query images.

The experiments were conducted using the images in Fig. 2. As in the previous sections, 3D maps and image databases were created from image sequence B. 740 query images were regularly sampled from image sequence A. Remind that there are many differences in the two image sequences as mentioned in Section V-B(2).

Fig. 5 shows the estimated camera poses and 3D maps as a reference. Localization is done with respect to a map, and the result is good if the estimated camera trajectory is close to the reference map. For edge points, pose estimation was done successfully at a high rate and a trajectory similar to the 3D map was obtained. On the other hand, for corners, pose estimation failed at many points and dispersion is large.

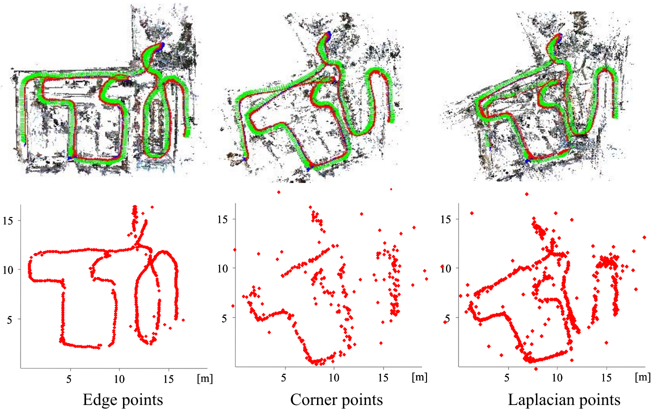
Fig. 6 shows examples of 3D localization using each feature type. As shown in the figure, there are plenty of matched points for edge points, which provides very stable localization. In the case of corners, there are only a small number of matched points, which can result in localization failure due to small conditional changes. Laplacian-based one lies between the two extremes.

Fig. 7 shows examples of edge-point based localization, in which corner-based localization failed. In general, the corner-based approach tends to fail in noisy and non-textured areas. Localization failures were mostly caused by failures in place recognition mentioned in Section V-B. Even if place recognition is successful, pose estimation often fails for corners. This is because sufficient point correspondences are not obtained due to sparseness of features.

D. Integration of Visual Odometry and Global Localization

An experiment integrating visual odometry and global localization was done using edge points under the same condition as Section V-C. Edgel-VO calculated the camera pose for every frame in image sequence A (7400 images), and global 3D localization was performed once for every 100 frames, in which the visual odometry estimate was replaced with the pose estimated by global localization. In the case the pose by global localization was largely different from the previous pose, it was discarded as outlier and the visual odometry estimate was used. In current implementation, no probabilistic data fusion was conducted.

Fig. 8 (a) shows the comparison of the trajectory by visual odometry only and the trajectory by visual odometry and global localization. The discrepancy between them at point G is 0.87[m].



Edge points Corner points Laplacian points

Fig. 5. Trajectories of consecutive global 3D localization. Top: 3D maps generated from image sequence B. Bottom: Localization using image sequence A.

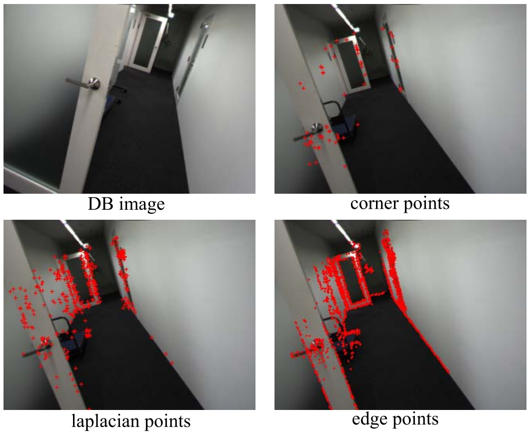

DB image corner points
laplacian points edge points

Fig. 6. Comparison of 3D localization. The overlaid red dots represent the 3D map reprojected onto the image according to the estimated pose.

Fig. 8 (a) shows the trajectory overlaid on the reference map. The integration of visual odometry and global localization provides a trajectory well matched with the map by compensating each other. Although a smooth trajectory was obtained without probabilistic schemes in this experiment, the accuracy is expected to improve by probabilistic data fusion. This is future work.

E. Discussion

The experiments show that the proposed method works well in indoor environments. The edge-point based approach provides better recognition rates and accuracy than other features especially in non-textured environments. Corners are detected sparsely in non-textured areas and it makes place recognition and pose estimation unstable as seen in Fig. 5. One practical solution to these problems is to use multiple stereo cameras for wider field of view and an IMU for robust motion model [23]. Our approach can provide good results using only one stereo camera although performance will be

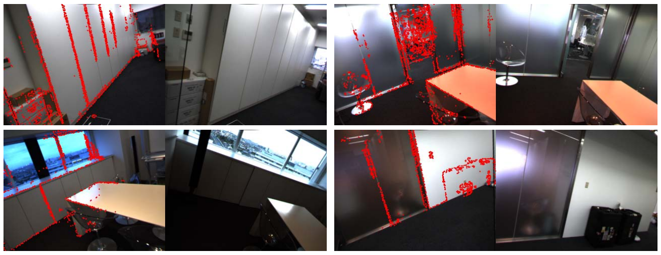
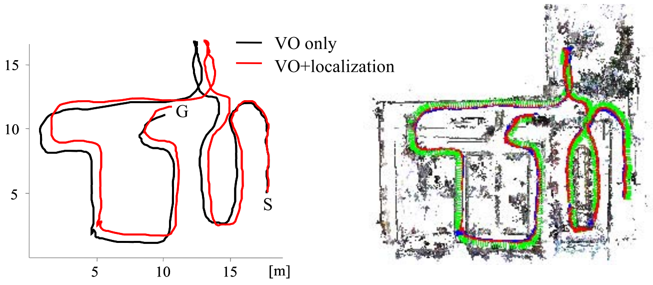


Fig. 7. Examples of 3D localization done by edge points. The red dots represent the 3D map reprojected onto the image. Corners failed localization using these images.



(a) Camera trajectory

(b) Trajectory on a map

Fig. 8. Trajectories by visual odometry and global localization.

enhanced by such additional devices.

Table III depicts the computation time. Our method is more time-consuming since the number of points is larger. This is noticeable in visual odometry. In localization, however, there are no large differences. This is because scale space analysis, which is the most time-consuming part, takes similar computation time for each feature type. Our system is implemented in Java and runs on top of Core2 Duo 3.06GHz using only a single core. In our experience, C++ implementation is three to five faster than Java implementation. This will make the system much faster.

Our method works well in both textured and non-textured environments especially in indoors. In our experiences, map building and visual odometry work in outdoor environments under various illumination conditions [21]. However, place recognition and 3D localization work successfully only in well-textured regions under moderate lighting conditions. Fig. 9 shows examples. Place recognition easily fails under poor illumination conditions in outdoors because feature detection is affected by drastic illumination changes. Solution to this problem is future work.

VI. Conclusions

This paper has presented a method of 3D localization with a stereo camera based on visual odometry and landmark recognition. The method detects landmarks using a scalable landmark-image database based on a vocabulary tree, and estimates 3D camera poses accurately using a randomized ICP algorithm. Future work includes the introduction of Bayes filter for data fusion and the incorporation of the proposed method into a navigation system.

TABLE III

Computation time in Java implementation [sec/frame]

	Edge points	Corner points	Laplacian points
Visual odometry	0.31	0.12	0.13
Global localization	2.17	1.84	2.02

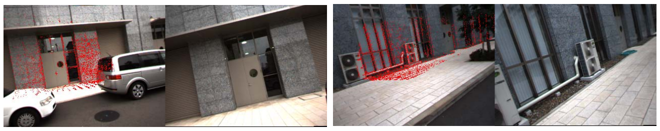


Fig. 9. 3D localization in outdoor environments. The red dots represent the 3D map reprojected onto the image.

References

[1] P. J. Besl and N. D. Mckay: “A Method of Registration of 3-D Shapes,” *IEEE Trans. on PAMI*, Vol. 14, No. 2, pp. 239-256, 1992.

[2] J. Canny: A Computational Approach to Edge Detection, *IEEE Trans. PAMI*, Vol. 8, No. 6, pp. 679-698 (1986).

[3] M. Cummins and P. Newman, “Highly Scalable Appearance-Only SLAM - FAB-MAP 2.0,” *Proc. of RSS2009*, 2009.

[4] P. Elinas, R. Sim, and J. J. Little: “σ $SLAM$: Stereo Vision SLAM Using the Rao-Blackwellised Particle Filter and a Novel Mixture Proposal Distribution,” *Proc. of ICRA2006*, pp. 1564–1570, 2006.

[5] D. Filliat: “A visual bag of words method for interactive qualitative localization and mapping,” *Proc. of ICRA2007*, 2007.

[6] M. Fischler and R. Bolles: “Random Sample Consensus: a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography”, *Communications ACM*, 24:381-395, 1981.

[7] F. Fraundorfer, C. Engels, and D. Nistér: “Topological mapping, localization and navigation using image collections,” *Proc. of IROS2007*, 2007.

[8] M. A. Garcia and A. Solanas: “3D Simultaneous Localization and Modeling from Stereo Vision,” *Proc. of ICRA2004*, 2004.

[9] R. Hartley and A. Zisserman: “Multiple View Geometry in Computer Vision,” Cambridge University Press, 2000.

[10] D.C. Herath, K.R.S. Kodagoda and G. Dissanayake:”Stereo Vision Based SLAM: Issues and Solutions,” Vision Systems, G. Obinata and A. Dutta (Eds), Advanced Robotic Systems, pp. 565-582, 2007.

[11] K. Konolige, J. Bowman, JD Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua: “View-Based Maps,”, *Proc. of RSS2009*, 2009.

[12] T. Lindberg: “Feature Detection with Automatic Scale Selection,” *Int. J. of Computer Vision*, 30(2), pp. 79-116 (1998).

[13] D. G. Lowe: “Object Recognition from Local Scale-Invariant Features,” *Proc. of ICCV*, 1999.

[14] D. Nistér, and H. Stewénius, “Scalable Recognition with a Vocabulary Tree,” *Proc. of CVPR2006*, 2006.

[15] Y. Ohta and T. Kanade: Stereo by intra- and inter- scanline search using dynamic programming,” *Trans. on IEEE PAMI*, Vol. 7, No. 2, pp. 139–154, 1985.

[16] S. Se, D. Lowe, and J. Little: “Local and Global Localization for Mobile Robots using Visual Landmarks,” *Proc. of IROS2001*, 2001.

[17] J. Sivic and A. Zisserman: “Video Google: A text retrieval approach to object matching in videos,” *Proc. of ICCV2003*, 2003.

[18] J. Shi and C. Tomasi: “Good Features to Track,” *Proc. of CVPR’94*, pp. 593-600, 1994.

[19] M. Tomono: “3-D Object Map Building Using Dense Object Models with SIFT-based Recognition Features,” *Proc. of IROS2006*, 2006.

[20] M. Tomono: “Robust 3D SLAM with a Stereo Camera Based on an Edge-Point ICP Algorithm,” *Proc. of ICRA2009*, pp. 4306–4311, 2009.

[21] M. Tomono: “Detailed 3D Mapping Based on Image Edge-Point ICP and Recovery from Registration Failure,” *Proc. of IROS2009*, to appear.

[22] J. Wang, R. Cipolla, and H. Zha: “Vision-based Global Localization Using a Visual Vocabulary,” *Proc. of ICRA2005*, 2005.

[23] Z. Zhu, T. Oskiper, S Samarasekera, R. Kumar, and H. S. Sawhney, “Ten-fold Improvement in Visual Odometry Using Landmark Matching,” *Proc. of ICCV2007*, 2007.