MOBILE ROBOT LOCALIZATION IN INDOOR ENVIRONMENT USING SCALE-INVARIANT VISUAL LANDMARKS

Soon Young Park, Suk Chan Jung, Young Sub Song, Hang Joon Kim

Department of Computer Engineering, Kyungpook National Univ., Republic of Korea
{sypark, scjung, yssong, hjkim}@ailab.knu.ac.kr

ABSTRACT

This paper presents a three-dimensional mobile robot localization system using visual landmarks. We use the scale-invariant feature points as visual landmarks. This feature is independent of camera view point, thus it is proper to use the landmark. The keypoints detected by SIFT are invariant to scale change and rotations, thus we use the keypoint as a landmark. Since the image coordinates for the landmarks are projected depending on the camera pose, the camera pose is determined using the relation between the two-dimensional image coordinates and three-dimensional world coordinates for the landmarks. The camera pose is considered the same as the robot pose, as the camera taking the images is fixed to the robot. The inclusion of falsely detected landmarks has an adverse effect on the accuracy of the robot localization. Therefore, the proposed method estimates the robot pose, while eliminating any falsely detected landmarks. To evaluate the proposed method, experiments are performed using a mobile blimp robot in an indoor environment. The results confirm that the proposed method can estimate the robot pose with a good accuracy.

Index Terms— robot, localization, visual landmark, SIFT

1. INTRODUCTION

Mobile robot localization has recently become an active area of study, and involves estimating the relative position and orientation of a robot in a particular environment. Although various types of sensor can be used for localization, including sonar, lasers, and cameras [1][2][3], the improved computational capabilities of processors have facilitated the use of more vision-based approaches, among which the landmark based method is simple and robust for accurate localization [4][5][6].

Accordingly, this paper presents a vision-based robot localization system that uses a single camera. In this paper, we consider the robot localization as a robot pose estimation problem. To estimate the robot pose, scale-invariant feature points are used as visual landmarks. The keypoints detected by the SIFT algorithm are invariant to scale changes, rotation, affine transformations, and illumination changes [7], making them independent of the camera viewpoint, and suitable landmarks for robot localization. The proposed method then estimates the robot pose using the relation between the image coordinates and the corresponding world coordinates. Since this relation depends on the pose of the camera, the camera pose is estimated using landmarks. When the camera is fixed to the robot, the camera pose is considered as the robot pose.

An overview of the proposed system is presented in Fig.1. It is assumed that the world coordinates of the landmarks are known. Thus, before running the system, a landmark database is built. When the robot then travels the environment, images are periodically captured by a camera, and landmarks are detected in the input images. The landmarks are detected by matching the keypoints in the image with the landmarks in the landmark database. When matches are found, the relation between the image coordinates and the corresponding world coordinates is calculated. Then the camera pose can be determined by applying a camera calibration method to several matched points [8]. The RANSAC algorithm is used to remove the effect of falsely matched data [9][10]. To evaluate the described robot localization method, the pose of a mobile blimp robot is estimated in an indoor environment. Since the blimp can travel in all directions without any constraint, the height of the blimp also needs to be estimated. As a result, the robot pose is estimated using only the visual information.

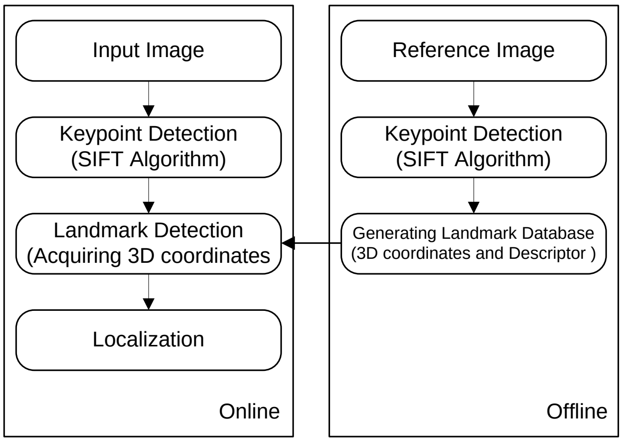
Input Image
Keypoint Detection (SIFT Algorithm)
Landmark Detection (Acquiring 3D coordinates
Localization
Online
Reference Image
Keypoint Detection (SIFT Algorithm)
Generating Landmark Database (3D coordinates and Descriptor)
Offline

Fig.1. System overview.

Section 2 explains the method used to detect the landmarks and construct the database, section 3 describes how the robot pose is estimated based on the relation between the image coordinates and the world coordinates. Experimental results using the proposed method are presented in section 4, and some conclusions are given in section 5.

2. VISUAL LANDMARKS

The proposed system for the mobile robot localization uses visual landmarks that must be recognizable features in the image. We use feature points with world coordinates that are known to the robot as landmarks, and they are stored in a database. The robot pose is then determined from its position relative to these landmarks. When a robot travels an environment, the images change continuously. Therefore, in the case of a vision-based localization method, the landmarks must be invariant under rotation and scale changes. For this purpose, we use the keypoints which are detected by the SIFT algorithm as the landmarks.

This section then describes the methods used to detect the keypoints using the SIFT algorithm, along with the structure of the landmark database.

2.1. Keypoint detection using SIFT

The SIFT-based keypoint extraction procedure is composed of the following four steps [7]:

1) Scale space extrema detection: the candidate keypoints are detected by finding the maxima and minima pixels in the images based on the DoG (Difference of Gaussian).

2) Keypoint localization: the unstable keypoints are rejected, and the remaining keypoints are assigned a location and scale.

3) Orientation assignment: each keypoints is assigned one or more orientations based on the local image gradient.

4) Keypoint descriptor generation: each keypoint is assigned a 128-dimensional descriptor.

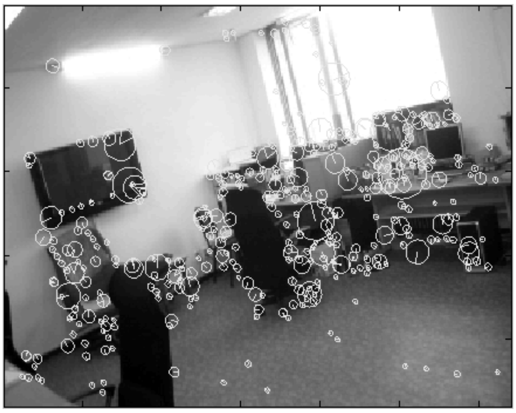

Fig.2. Image with SIFT keypoints marked.

After above steps, the detected keypoints are assigned locations, scales, orientations, and SIFT descriptors.

Fig. 2 shows the detected keypoints in an input image when using the SIFT algorithm described above, where the center of each circle represents the coordinates of the detected keypoint, the radius of the circle denotes the extrema scale, and the line in the circle indicates the orientation of the keypoint.

2.2. Landmark database

When localizing a mobile robot, it is assumed that the world coordinates for each landmark are known. Thus, a landmark database needs to be constructed before starting the localization procedure. For this purpose, we construct a landmark database as follows. First, several images are captured of the given environment within which the robot travels. Keypoints are then detected within these images using the SIFT algorithm, and some selected as the visual landmarks. The landmark database entries are as follows:

$$L_i = \left[M_i,\ SIFT_{128}\right] \tag{1}$$

where *i* is the index for each landmark. M_i represents the three-dimensional world coordinates for the *i* th landmark, which are measured manually, and $SIFT_{128}$ is the 128-dimensional descriptor. This vector is used when we searching for landmarks that match the keypoints in an input image. Fig. 3 shows two-dimensional projection of the world coordinates for all the landmarks in the constructed database.

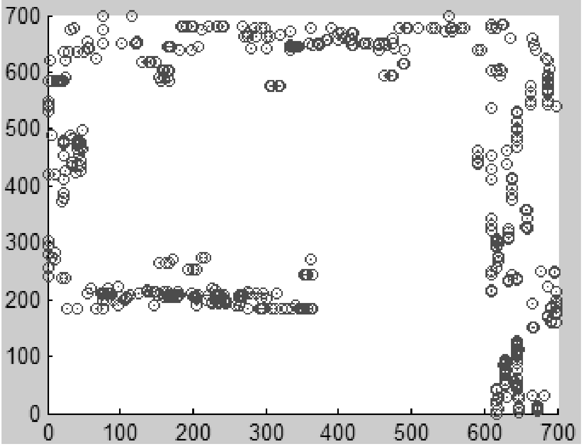

Fig.3. Landmarks in database.

3. ROBOT LOCALIZATION

This paper considers the mobile robot localization as determining the pose of a mobile robot. A camera fixed to the robot captures periodic images while the robot travels the given environment. The robot pose is then estimated using visual landmarks in the input images. As the landmarks are projected in the input images according to the camera pose, the robot pose can be determined using the relation between the image coordinates and corresponding world coordinates for the landmarks using the camera calibration theory [8]. Therefore, the proposed method includes a procedure for landmark detection in the input images. Plus, any falsely detected landmarks, which can drastically increase the pose error, are removed using the RANSAC algorithm [9][10].

This then section describes the procedure used to detect landmarks in the input image, followed by the estimation of the robot pose.

3.1. Landmark detection

The proposed robot localization method involves obtaining the two-dimensional coordinates of the landmarks in the input image. The first step is to detect landmarks in the input image. The landmark detection is performed by matching between every detected keypoint in the input image with all the landmarks in the database. If matches are found, the corresponding two-dimensional image coordinates are obtained for that landmark. The criterion for a match is based on a similarity test between the SIFT descriptor for a detected keypoint and the SIFT descriptor for a landmark in the database. The criterion of a match is whether the following relation is satisfied:

$$d(k, L_a) < d(k, L_b) \times Th_{sim} \tag{2}$$

In the above equation, k is the identifier for the keypoint detected in the input image, L_a is the identifier for the database landmark with the least distance between the descriptor vector for the keypoint and that for the database landmark, L_b is the identifier for the database landmark with the second least distance, and Th_{sim} is a predefined threshold value, which was 0.8 in this study according to Lowe [7]. Meanwhile, *d(P, Q)* is the distance between the two 128-dimensional vectors identified by P and Q. If the vector components of the two 128-dimensional SIFT descriptors are as follows:

$$\begin{aligned} P &= (p_1, p_2, \ldots, p_{128}) \\ Q &= (q_1, q_2, \ldots, q_{128}) \end{aligned} \tag{3}$$

the Euclidean distance between the two vectors is then

$$d(P, Q) = \sqrt{\sum_{i=1}^{128} (p_i - q_i)^2} \tag{4}$$

3.2. Robot pose estimation

The robot pose is then estimated using the matched landmarks. The camera pose is considered the same as the robot pose, as the camera that captures the images is fixed to the robot. A robot pose p is a pair of a position vector t and an orientation vector θ.

$$\begin{gathered} p = (t,\ \theta) \\ t = [t_x\ t_y\ t_z]^T, \theta = [\theta_x, \theta_y, \theta_z]^T\ . \end{gathered} \tag{5}$$

The camera pose is determined based on the relation between the image coordinates and world coordinates of the landmarks. The relation is calculated using a camera projection function that transforms the 3D world coordinates into 2D image coordinates. Here, the pinhole camera model is used as the projection function [8]. Pinhole camera model uses homogeneous coordinates. Using the pinhole camera model, the camera projection function f is expressed as

$$\begin{gathered} f(p, M_i) = \begin{bmatrix} u/w \\ v/w \end{bmatrix} \\ where, \begin{bmatrix} u \\ v \\ w \end{bmatrix} = K \times [R \mid T] \times \begin{bmatrix} M_i \\ 1 \end{bmatrix} \end{gathered} \tag{6}$$

, where M_i represents the world coordinates for the ith landmark in the database. K represents the internal parameter, and R and T are the external parameters. As such, the external parameters transform the world coordinates into camera coordinates, then K projects the camera coordinates into the image coordinates. The camera pose is determined using these parameters. K is a 3×3 matrix composed of the internal parameters, including the camera focal length and center coordinates. The internal parameters can be calculated using the camera and the images captured by the camera [11]. Meanwhile, the camera pose p is calculated using the external parameters, R and T. R is a 3×3 rotation matrix that can be expressed as

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

and T is a 3×1 translation vector. After obtaining R and T, the camera pose p is calculated as follows. The camera position is defined as the point where the camera coordinates are $[0, 0, 0]^T$. Therefore the camera position represents the world coordinates that satisfy the following equation

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \left[R \mid T \right] \begin{bmatrix} t_x \\ t_y \\ t_z \\ 1 \end{bmatrix} \tag{8}$$

The camera orientation is then calculated using the calculated rotation matrix.

A disparity occurs between the projected image coordinates and the image coordinates of a landmark, when the world coordinates of a landmark are projected onto the image plane using equation (6). This disparity is defined using the energy E. The energy is defined as the average distance in image pixel units between the coordinates for the landmarks in the image plane and the 2D projected coordinates for the matching world coordinates with a pose parameter p.

$$E(p) = \sum_{i=1}^{n} \frac{\left\| m_i - f(p, M_i) \right\|^2}{n} \tag{9}$$
$$where\, M_i \in M, m_i \in m, n = |M| = |m|$$

, where m_i represents the two-dimensional image coordinates for the landmark in the image plane. The camera pose p is then determined based on minimizing the energy E.

$$\hat{p} = \arg\min_{p} E(p). \tag{10}$$

When the number of matched pairs is high (the more landmarks are detected in the image), the accuracy of the robot localization is also higher. However, the existence of falsely matched pairs has a drastic impact on the accuracy of the camera pose estimation. The matched pair can contain the falsely matched pair. Therefore, the RANSAC algorithm is applied to remove the effect of falsely matched pairs [9][10]. Thus, the following procedure is used to estimate the camera pose p from the matched pairs:

1) Select k (which was 3 in this paper) random pairs.
2) Calculate the hypothesis (tentative pose $\hat{p}$).
3) For each matched pair that is not selected in 1), check whether it is an inlier using the following criterion:
$$D(p, L_i) = \begin{cases} inlier \ \ if \ \left\| m_i - f(p, M_i) \right\| < Th_{dist} \\ outlier \ \ otherwise \end{cases}$$
4) When the ratio of the number of inliers is larger than a predefined threshold value Th_{rate}, stop the iterations. Otherwise repeat from step 1).

4. EXPERIMENTAL RESULTS

Experiments were performed using a mobile robot in an indoor environment. The image data were processed using a Pentium 2.8 GHz PC with a Windows XP operating system. The software was implemented using Visual C++ 6.0. The image size taken by the camera was QVGA(320 × 240).

The indoor environment where the robot could travel was a 7m × 7m area, and the predetermined database included 781 landmarks generated from 6 images. In this experiment, we use a blimp as the mobile robot, which is shown in Fig. 4. The blimp had three propellers, which enabled it to travel without constraint in all directions, including up and down. As shown Fig. 4, a camera is fixed to the blimp, and the images are taken by the camera.

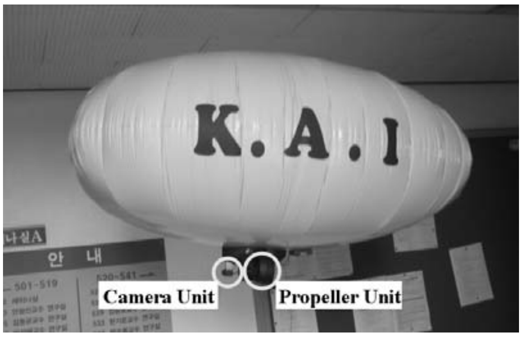

Fig.4. Mobile blimp robot.

Estimating the robot pose using the proposed system requires the determination of Th_{dist} which is the threshold value for whether a matched keypoint is an inlier or not, and Th_{rate} which is the rate of the number of inliers to the number of all matched pairs. To determine these parameters, we manually classified the matched pairs into the inliers and the outliers for sample images. Then we calculated the energy of the landmarks in case of inliers and outliers and the ratio of inliers and outliers to all pairs. Table 1 shows the energy of inliers and outliers, the rate of them. In table 1, the ratio of the number of inliers was 36.75% and the mean energy of the inliers was 4.71. Thus, based on experimental results, Th_{dist} was set at 10 and Th_{rate} set at 30%.

Table 1. Experimental results for a parameter determination.

	Inliers	Outliers
Energy	4.71	90.89
Rate	36.75%	63.25%

To demonstrate the effectiveness of the proposed mobile robot localization method, the estimated robot pose was compared with the real robot pose. Then, the real camera pose was manually measured. The experimental results then showed that the mean distance error for the robot position was 26cm, whereas the orientation error was about 7.5°. Fig. 5 shows the several estimated poses with the corresponding real robot poses. In Fig. 5, the center of the circle represents the position of the robot, while the line through the circle represents the orientation of the robot.

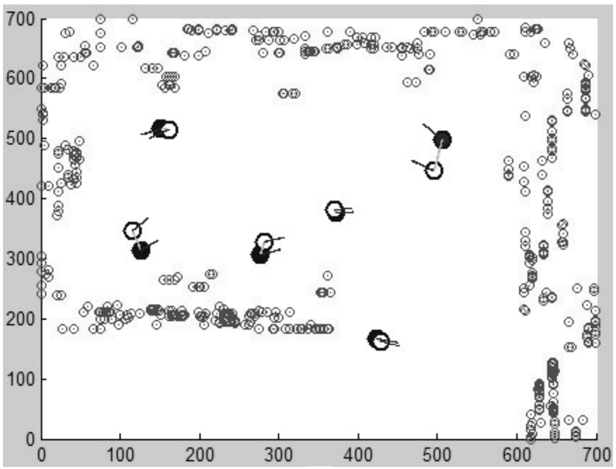

Fig.5. Estimated robot poses: ● is the real robot poses, ○ is the estimated robot poses using the proposed method.

5. CONCLUSION

In this paper, we have presented a robot localization system using scale-invariant visual landmarks. The landmarks are feature points with world coordinates that are known to the robot, and stored in a landmark database before running the system. When using the system, a camera fixed to the robot captures the periodic images. And the keypoints in the input images are detected using the SIFT algorithm. The detected keypoints are then matched to the landmarks in the landmark database. The pose of the robot is estimated using the relation between the two-dimensional image coordinates and the three-dimensional landmark coordinates. The RANSAC algorithm was applied to eliminate any falsely matched keypoints. Experimental results demonstrate the proposed method works well for estimating the three-dimensional position and orientation of a robot.

6. REFERENCES

[1] O.Wijk, H.I.Christensen, “Localization and navigation of a mobile robot using natural point landmarks extracted from sonar data,” *Robotics and Autonomous Systems*, vol.31, issues 1-2, pp. 31-42, 2000.

[2] S.Zhang, L.Xie and M.D.Adams, “Feature extraction for outdoor mobile robot navigation based on a modified Gauss–Newton optimization approach,” *Robotics and Autonomous Systems*, vol.54, issue 4, pp. 277-287, 2006.

[3] P.Skrzypczynski, “Spatial Uncertainty Management for Simultaneous Localization and Mapping,” *Robotics and Automation, 2007 IEEE International Conference on*, pp. 4050-4055, 2007.

[4] M. Betke, L.Gurvits, “Mobile Robot Localization Using Landmarks,” *IEEE Transaction Robotics and Automation*, vol.13, No.2, pp. 251-263, 1997.

[5] R. Sim, G. Dudek, “Mobile robot localization from learned landmarks,” *In Proceedings of the International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1060-1065, 1998.

[6] S. Thrun, “Finding landmarks for mobile robot navigation,” *In Proceedings of the International Conference on Robotics and Automation*, vol. 2, pp. 958-963, 1998.

[7] D. Lowe, “Distinctive Image Feature from Scales-Invariant Keypoints,” *International Journal of Computer Vision*, pp. 91-110, 2004.

[8] R. Cipolla, P. J. Giblin,“Visual Motion of Curves and Surface,” *Cambridge University Press*, 2001.

[9] M. A. Fischler and R.C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". *Comm. of the ACM 24*: 381-395, 1981.

[10] D. A. Forsyth and J. Ponce, Computer Vision, a modern approach. *Prentice Hall* (2003). ISBN 0-13-085198-1.

[11] R. Hartley and A. Zisserman, “Multiple View Geometry in Computer Vision” *2nd edition, Cambridge University Press*. 2003.