An Algorithm for a pose estimation of a robot using Scale-Invariant Feature Transform

이재광*, 허욱열**, 김학일***
(Jae-Kwang Lee, Uk-Youl Huh, Hak-Il Kim)

Abstract – This paper describes an approach to estimate a robot pose with an image. The algorithm of pose estimation with an image can be broken down into three stages : extracting scale-invariant features, matching these features and calculating affine invariant. In the first step, the robot mounted mono camera captures environment image. Then feature extraction is executed in a captured image. These extracted features are recorded in a database. In the matching stage, a Random Sample Consensus(RANSAC) method is employed to match these features. After matching these features, the robot pose is estimated with positions of features by calculating affine invariant. This algorithm is implemented and demonstrated by Matlab program.

Key Words : Scale-Invariant Feature Transform ; Pose Estimation ; RANSAC ; Affine Invariant

1. Introduction

In recent years, an autonomous mobile robot serves people in various areas such as guiding a museum for tourists, helping patients in a hospital, catering food, keeping house safely, etc. A core problem for the robot to perform these tasks autonomously in dynamic environment, is the determination of the position and orientation of the robot. This self localization for a robot is performed by finding the robot pose supported by the most features matched in an image. Therefore, this paper deals with a pose estimation algorithm with a vision sensor.

A CCD camera is a popular choice for mobile robot sensing because it is not inherently dependent on environmental geometry like ranging devices[1]. However to estimate the robot pose with a vision sensor is too sensitive for illumination changes, image rotation, translation and occlusion. To overcome these problems Scale-Invariant Feature Transform(SIFT)[2] is employed in this paper. This paper describes a robot pose estimation with SIFT method through matching features.

The outline of the rest of the paper is as follows : Section 2 introduces how the scale-invariant features extracted from an image. Section 3 describes the matching procedures for the pose estimation and the following section shows simulation results. In section 5, concluding remarks and future works are provided.

저자 소개
* 李 宰 光 : 仁荷大學校 電氣工學科 碩士課程
** 許 旭 烈 : 仁荷大學校 電氣工學科 敎授・工博
***金 學 一 : 仁荷大學校 情報通信工學部 敎授・工博

2. Scale-Invariant Feature Extraction

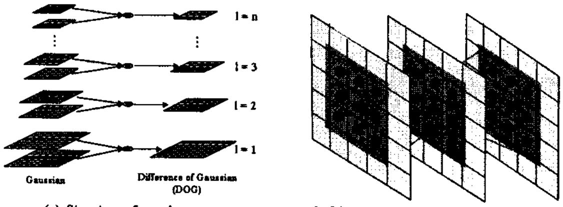
The features should invariant to image translation, rotation, illumination changes and occlusion to be used for pose estimation. SIFT[3] approach is proper method for pose estimation due to its invariant character. SIFT is detailed in the following subsections.

2.1 Gaussian Image Pyramid

Pyramid representations have the advantage that the minimum number of samples are used to represent the image at each scale, which greatly speeds up computation in comparison with a fixed resolution scheme. A scale space sampling of the image and its Gaussian can be efficiently constructed by a series of convolution, subtraction and sub-sampling steps[4], as shown in figure 1-(a). The sub-sampling and Gaussian standard deviation must be chosen such that each layer of the pyramid represents a 'correct sampling' of the image at some scale. In practice a Gaussian kernel with standard deviation 1.5 pixels is used and a sub-sampling of 1.5:1 in the Gaussian Pyramid.

2.2 Find Maxima and Minima

The SIFT feature locations are efficiently detected by identifying maxima and minima of a difference of Gaussian (DoG) function in scale space. The maxima and minima are extracted by comparing three adjacent level of DoG.



(a) Structure of a scale space (b) Maxima and minima of the DoG

Fig. 1 Image Pyramid and Maxima and Minima Computation

Figure 1-(b) is a pictorial representation of nearest neighbors in a current image and adjacent images for a pixel under consideration. The extracted maxima and minima can be considered as SIFT features.

3. Matching and Pose Estimation

Pose estimation is implemented by employing RANSAC[5] matching method and affine invariant.

3.1 RANSAC Matching

Given two sets of SIFT features, we would like to find the coordinate frame translation and rotation that will result in the most matches between SIFT features in the first image and SIFT features in the second image. Since a mobile robot is limited to planar motion, there are only 3 parameters (2 for translation and 1 for rotation) for this alignment.

This can be formulated as a hypothesis testing problem, where multiple alignment hypotheses are considered and the best one corresponds to the alignment which can match the most features from the first image to the second image.

RANSAC has been used in many applications for model fitting, hypothesis testing and outlier removal. RANSAC is employed to test the alignment hypotheses and find the inlier features.

3.2 Affine Invariant

The motivation for choosing this class of transformation is that a small planar surface patch when viewed from a varying viewpoint undergoes an affine distortion[6].

We intend to match scale-space features - features with a spatial 2D image location and an associated scale. If we hypothesize a match between such features in two images, then assuming an affine model. There are 3 unknown transformation parameters for the local image distortion : Translation, skew and rotation.

This information is illustrated graphically in figure 2. The translation (2 parameters) is determined by the spatial position of the features and the scale change(1 parameter) is determined by the relative scale of the features.

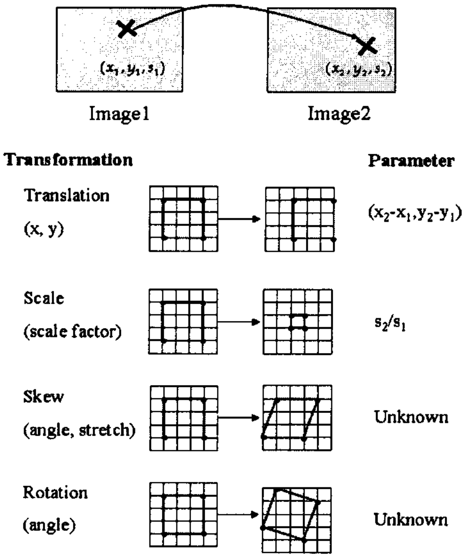
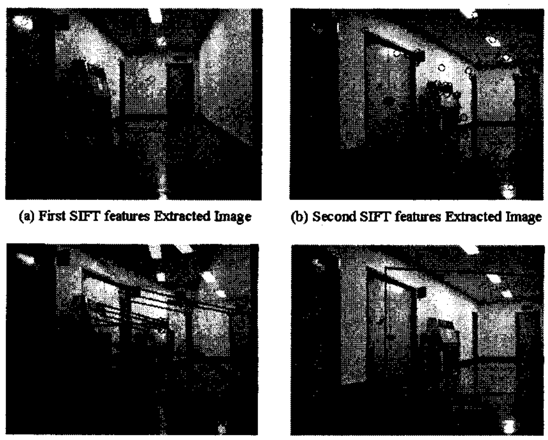


Fig. 2 Hypothesized Correspondence

4. Simulation Results

We have tested the proposed algorithm in various environment and conditions. One of the results is shown in figure 3. Figure 3 is pictures of corridor in front of our lab with about 40% translation between two images. Figure 3-(a), (b) shows the two input images. SIFT features extracted are marked in each image. Figure 3-(c) is a mixed picture of two input images and matched features are shown. The accuracy of matched features is about 92%. Figure 3-(d) means a estimated pose of the robot with marked red line compared to the first image. It is relative pose of the first image to the second image.



(a) First SIFT features Extracted Image (b) Second SIFT features Extracted Image

(c) Features Matched with RANSAC (d) Estimated Pose of the Robot

Fig. 3 Simulation Result for the process of the algorithm (Corridor)

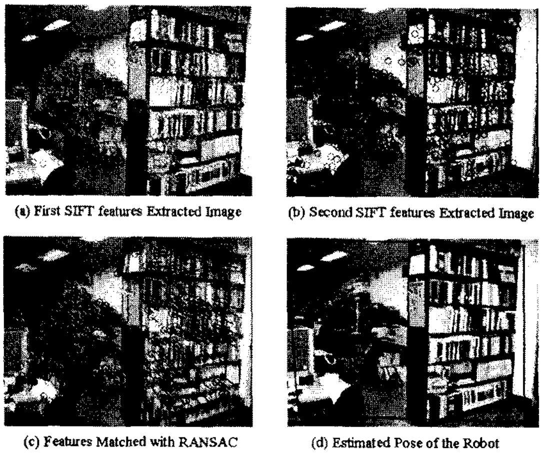


Fig.4 Simulation Result for the process of the algorithm (Inside of our lab.)

Figure 4 shows the same results with Figure 3. Figure 4 has more SIFT features and difference between the first image and the second image is smaller. It has small translation and skew about 10%. The accuracy is about 95%.

Figure 5 shows the results of accuracy for feature points matching compared the matched features computed for Harris corner detector[7]. We can see that SIFT obtained better results in various conditions. In addition, the accuracy of SIFT is not sensitive to its conditions but Harris does. In the figure, percent indicates the rate of difference for two images.

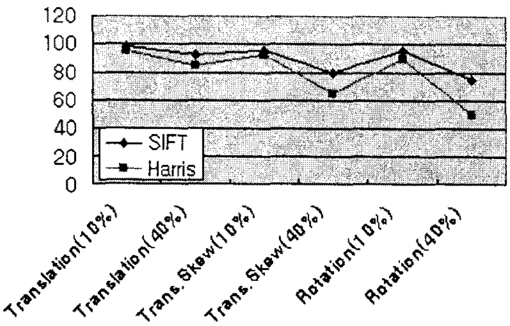


Fig. 5 The accuracy of two feature detectors (SIFT and Harris)

5. Conclusions and Future Works

In this paper, we described briefly a vision-based robot pose estimation algorithm based on the SIFT features. SIFT method developed by Lowe for image feature generation has characteristics of invariant to image translation, scaling, rotation and illumination changes. RANSAC matching method and affine invariant is used to estimate robot pose. We evaluate the performance of SIFT and previous approach to feature detection, Harris corner detector, one of the widely used and Harris corner detector is sensitive to the scale of an image. Therefore is not suitable for pose estimation and localization that can be matched from a range of robot position.

We are currently looking into concurrent map building and localization based on SIFT features using mono vision and encoder information of two motors.

References

[1] J.A. Perez, J.A. Castellanos, J.M.M. Montiel, J. Neira, and J.D. Tardos, "Continuous Mobile Robot Localization : Vision vs Laser," IEEE Int. Conf. on Robotics and Automation, pp. 2917~2923, 1999

[2] D.G. Lowe, "Object Recognition from Local Scale-invariant Features," Proc. of the 7th Int. Conf. on Computer Vision, pp. 1150~1157, 1999

[3] S. Se, D.G. Lowe, and J. Little, "Vision-based Mobile Robot Localization and Mapping using Scale-Invariant Features," Proc. of the IEEE Int. conf. on Robotics and Automation, pp. 2051~2058, 2001

[4] D.G. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," Int. Journal of Computer Vision, Vol. 60, No. 2, pp. 91~110, 2004

[5] M.A. Fischler and R.C. Bolles, "Random Sample Consensus : a Paradigm for Model Fitting with Application to Image Analysis and Automated Catrography," Commun. Assoc. Comp. Vol. 24, pp. 381~395, 1981

[6] E.J. Pauwels, T. Moons, L. VanGool, P. Kempenaers, and A. Oosterlinck, "Recognition of Planar Shapes under affine distortion," Int. Journal of Computer Vision, Vol. 14, No. 1, pp. 49~65, 1995

[7] C.G. Harris and M. Stephens, "A Combined Corner and Edge Detector," Proc. 4th Alvey Vision Conf., pp. 147~151, 1988