Applied Mechanics and Materials Vols. 284-287 (2013) pp 2142-2146

doi:10.4028/www.scientific.net/AMM.284-287.2142

Robot Simultaneous Localization and Mapping Using Speeded-Up Robust Features

Yin-Tien Wang[1,a], Chen-Tung Chi[2,b] and Ying-Chieh Feng[1,c]

[1]Dept. of Mechanical and Electro-Mechanical Eng., Tamkang U., New Taipei City 251, Taiwan

[2]Dept. of Mechanical Eng., Taipei Chengshih U. of Science and Technology, Taipei 112, Taiwan

[a]ytwang@mail.tku.edu.tw, [b]cdchi@tpcu.edu.tw, [c]696372076@s96.tku.edu.tw

Keywords: Robot Mapping, Speeded-Up Robust Features (SURF), EKF-SLAM

Abstract. An algorithm for robot mapping is proposed in this paper using the method of speeded-up robust features (SURF). Since SURFs are scale- and orientation-invariant features, they have higher repeatability than that of the features obtained by other detection methods. Even in the cases of using moving camera, the SURF method can robustly extract the features from image sequences. Therefore, SURFs are suitable to be utilized as the map features in visual simultaneous localization and mapping (SLAM). In this article, the procedures of detection and matching of the SURF method are modified to improve the image processing speed and feature recognition rate. The sparse representation of SURF is also utilized to describe the environmental map in SLAM tasks. The purpose is to reduce the computation complexity in state estimation using extended Kalman filter (EKF). The EKF SLAM with SURF-based map is developed and implemented on a binocular vision system. The integrated system has been successfully validated to fulfill the basic capabilities of SLAM system.

Introduction

To build a persistent map with robust image features (landmarks) is an important step for implementing the visual simultaneous localization and mapping (SLAM). The robust image features must have the properties of high repeatability and unique description to be successfully detected at each time step. Furthermore, the map is sparse but persistent for real-time implementation and being able to represent the characteristic of the environment. The research in this paper focuses on the development of an algorithm to build a persistent map for robot visual SLAM based on spare representation of scale-invariant features. The spare representation has the advantage of reducing the computational cost for information update in SLAM tasks using extended Kalman filter (EKF).

Harris corner detector [1] is the most popular method for image feature detection in visual SLAM task [2,3]. This method detects image corners or point positions by investigating the eigenvalues of the second moment matrix. The Harris corner detector is a simple algorithm and is easy to implement for robot SLAM task. However, it is difficult to track the corner features robustly when the camera is moving. When the distance and angle of camera viewpoint is changed, the scale and orientation of the corner feature will be changed and then result in the failure of image feature tracking. Therefore, more efforts are needed to recover the scale and orientation of image features [2]. On the other hand, the detection method of scale- and orientation-invariant features [4,5] can automatically resolve above-mentioned problems and provide a robust method for feature representation in SLAM task [6]. However, the shortcoming of the scale- and orientation-invariant method is computational inefficient. In order to improve the computational performance, Bay et al. [7] utilized the concepts of box filter and integral image [8] for the detection of scale- and orientation- invariant features, called speeded-up robust features (SURF). In the literature, it was suggested that SURF is superior to other methods for detection and representation of image point features [9]. However, the computational cost of SURF method in visual SLAM was not evaluated. Many researchers have applied SURF in robot localization and mapping tasks. Murillo et al. [10] utilized SURF to replace the SIFT method for image feature detection and recognition. The results showed that the efficiency for robot visual localization has been improved.

In this paper, we propose a novel algorithm for map building based on modified SURF feature detection and recognition procedures. A binocular vision is utilized as the sensing device to implement the visual SLAM tasks. Meanwhile, two experiments on a real system are carried out in static SLAM scenes to validate the proposed algorithm.

Speeded-Up Robust Features

Lindeberg proposed the concept of automatic scale selection to overcome the disadvantage of Harris corner [4]. He established a Hessian matrix whose elements are the convolution of the image and Laplacian of Gaussians (LoG). Then the feature points can be detected by investigating the determinant of the Hessian matrix. Lindeberg's method has the advantages of stability and high repeatability. Lowe [5] replaced the LoG by Difference of Gaussians (DoG). Bay et al. [7] utilized the box filter to approximate LoG and the determinant of the Hessian matrix. The box filter was further combined with the method of integral image [8] to reduce the image processing time. They coined this method of feature detection as SURF [7].

In order to improve the robustness of feature representation and reduce the computational cost, we modify the SURF detection method in two aspects [11]: first, we utilize the box filters only in 2 low-level octaves for feature detection to reduce the computation cost. Second, we set up a threshold value ($D_{\text{threshold}}$) as the lowest limit for the determinant of Hessian matrix in order to control the number of detected features in an image. The value of $D_{\text{threshold}}$ can be obtained by an online procedure. Our purpose is that, even in the dull background or environment with fewer features, there will be enough number of features detected.

After the features are detected from the image, the description vector is utilized to represent the characteristics of features. For the orientation of a feature, Bay et al. [7] computed the Haar wavelet responses in the x and y direction of the feature area. The orientation of a feature is defined at the direction with the largest sum of the Haar wavelet responses. Furthermore, a high-dimensional description vector is utilized to describe the uniqueness of the feature. First, choose a square area with the center located in the feature point and its direction along the direction of the feature. Second, divide the square area into 4x4 sub-areas. There is a 4D descriptor vector for each sub-area, expressed as v=(Σdx, Σdy, Σ|dx|, Σ|dy|), where Σdx, Σdy, Σ|dx|, and Σ|dy| are the sums of the Harr wavelet responses and their absolute values in x and y direction which is defined by the coordinate along the orientation of the feature. This will result in total dimensions of 4x4x4=64 for the description vector.

The Nearest-Neighbor (NN) search method [12] is the most popular method for matching high-dimensional description vectors. The Euclidean distance d between the arbitrary point p in P set and the query point q is usually defined by the norm l_2. The distance between two descriptors would be treated as the basis of feature matching. Meanwhile, we use d_{match} as the threshold of the Euclidean distance and as a judgment whether the matching between two descriptors is successful. When the Euclidean distance d is less than d_{match}, the matching is successful,

$$d = \|p-q\|_2 = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} < d_{\text{match}}. \tag{1}$$

One example is implemented to validate the modified SURF algorithm for image feature detection and matching. An image with 320×240 pixels is captured by a CMOS webcam. The modified SURF algorithm is applied to detect image features on the image. We use box filters only in 2 low-level octaves. The detecting results using different $D_{\text{threshold}}$ values are depicted in Fig. 1. The figure shows that the number of detected features is varied from 837 to 41 when the value of $D_{\text{threshold}}$ is changed from 0 to 10000. However, the computation time is almost the same for different value of $D_{\text{threshold}}$.

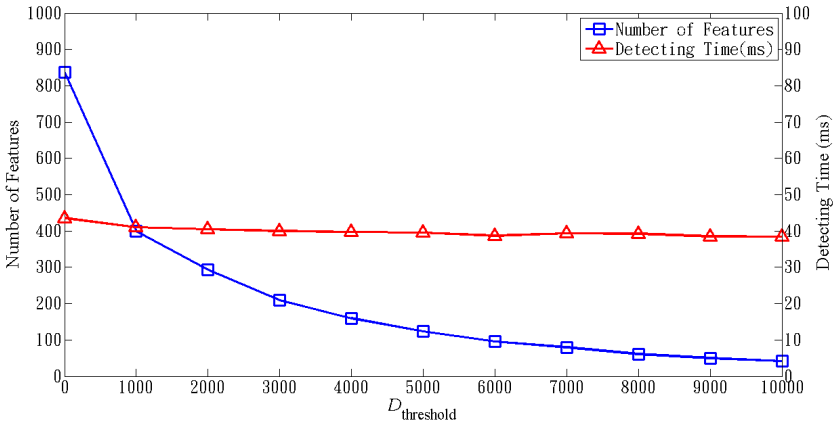

Fig. 1. Example using different $D_{threshold}$ values

Robot SLAM

In robot SLAM tasks, the states of robot and landmarks in the environment are estimated based on the information of measurements. State estimation is treated as a target tracking problem in this article. The state sequence of a system at time step k can be expressed as

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, w_{k-1}). \tag{2}$$

where $\mathbf{x}_k$ is the state vector; $\mathbf{u}_k$ is the input; w_k is the process noise. The objective of the tracking problem is to recursively estimate the state $\mathbf{x}_k$ of the target according to the measurement $\mathbf{z}_k$ at k,

$$\mathbf{z}_k = [\mathbf{z}_{1k}^T \ \mathbf{z}_{2k}^T \cdots \mathbf{z}_{mk}^T]^T; \quad \mathbf{z}_{ik} = \begin{bmatrix} I_{ix} \\ I_{iy} \end{bmatrix}, i = 1, 2, \cdots, m \tag{3}$$

where m is the number of measured image features at time k; (I_{ix}, I_{iy}) are the pixel coordinates of ith feature in the image plane. In the recursive state estimation algorithm, a hand-held binocular vision is utilized as the only sensing device for the measurement. The image depth of extracted features can be obtained using the concept of stereo vision [13]. We treat this camera as a free-moving robot system with unknown inputs. The states of the system can be estimated by solving the target tracking problem using EKF estimator [2,3].

Map Management

The proposed SLAM is implemented on the free-moving vision system by integrating the motion and sensor models, as well as the extraction of SURF. A flowchart for the developed SLAM system is depicted in Fig. 2. The images are captured by binocular camera and features are extracted using the modified SURF method. In the flowchart, a map managerial tactic is designed to manage the newly extracted features and the bad features in the system. The properties of the newly extracted features are investigated and the moving objects will be discriminated from the stationary objects by using a detection algorithm. All the stationary landmarks are included in the state vector. On the other hand, those features which are not continuously detected at each time step will be treated as bad features and erased from the state vector.

Experimental Results

In this section, the experiment of robot SLAM is carried out on the real system to validate the performance of the proposed algorithm. The camera is carried by a person to circle around a bookshelf (1.5m×2m floor dim.) in our laboratory. The resultant map and the camera pose estimation are plotted

in Fig. 3. In this figure, the estimated states of the camera and landmarks are illustrated in a 3D map plot. The camera is carried to move from first image frame and circles around the bookshelf three times. The SLAM system also starts up from first image frame and captures image features with unknown positions. These features will be initialized and stored as landmarks in the map. The SLAM system builds the environment map and estimates the camera pose concurrently, when the camera is carried to circle around the bookshelf. The 32nd, 645th, and 1484th frame of color images in Fig. 3 belong to first camera circulation around the bookshelf. The results show that the SLAM task is successfully implemented. In each color image, the (blue) circular marks indicate the landmarks extracted from the captured image with an unknown image depth, while the (red) square marks represent the landmarks with a known and stable image depth. The estimated states of the camera and landmarks are illustrated in 3D map in the center of Fig. 3. The red (dark) ellipses represent the uncertainty of the landmarks which have known image depths and the green (light) ellipses denote the uncertainty of the landmarks which have unknown image depths. Meanwhile, the trajectory of the estimated camera pose is plotted as solid lines in 3D map. The 2298th frame in Fig. 3 is the end of the first camera circulation. At this location, the camera comes to the place it had visited before and the trajectory loop is closing. Some old landmarks, for example landmark No. 41 in this image, are captured again and the covariance of the state vector is reduced gradually. In the color images of 2739th and 3039th frames, which belong to second camera circulation, the landmark numbers under 450 are successfully refound by the SLAM system. The 3331st and 4209th in Fig. 3 are the end of the second and third camera circulation, respectively. The second and third times of loop-closure are expected at these two locations. In these frames, old landmarks are visited again and the covariance of the state vector is reduced further.

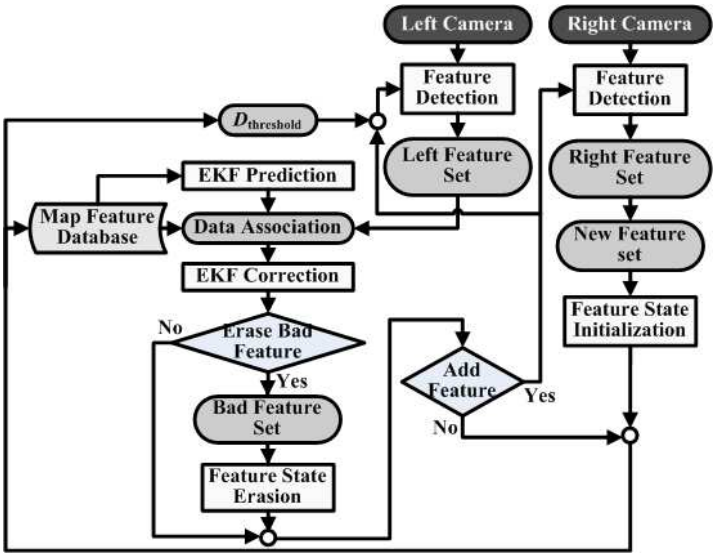

Fig. 2. Flowchart for the robot SLAM

Conclusions

In this research, we developed an algorithm for building a persistent map to improve the robustness of robot visual SLAM system. The SURF method is utilized to provide a robust detection of image features and a stable description of the features. Using the modified SURF algorithm, the experimental work has been carried out on a real system. The results showed that the binocular SLAM system with the proposed algorithm has the capability to support robot systems simultaneously navigating and mapping in the environment.

Acknowledgments: This paper was partially supported by the National Science Council in Taiwan under grant no. NSC100-2221- E- 032-008 and NSC101-2632-E-032-001-MY3.

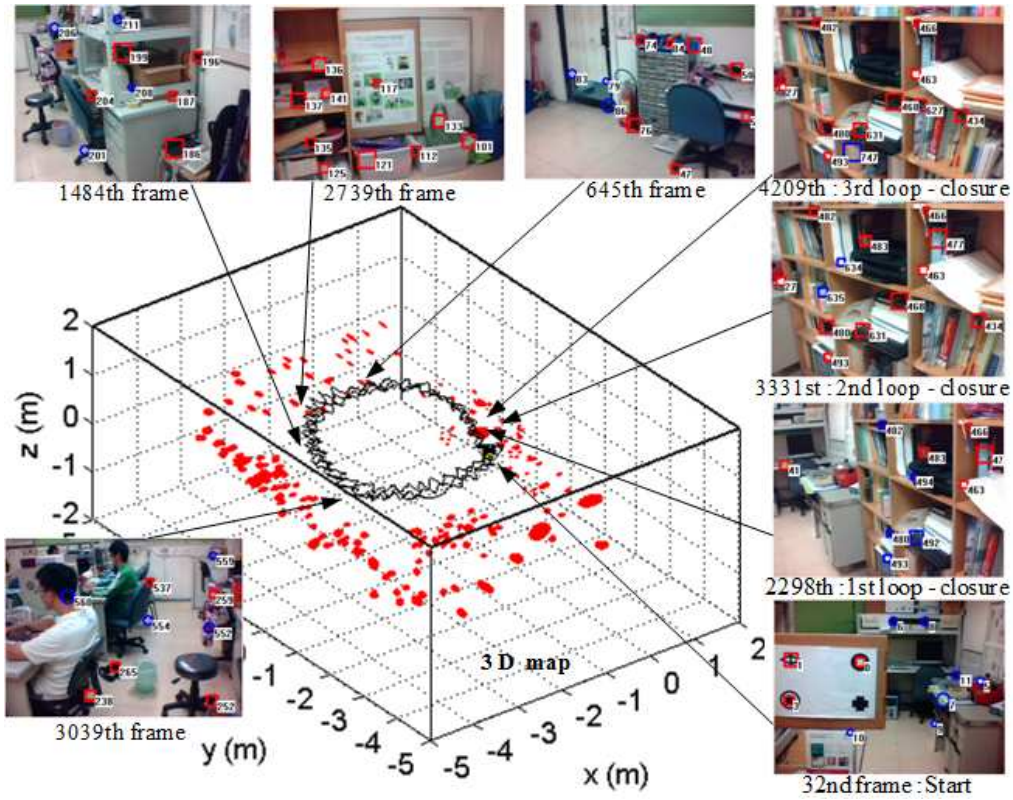

Fig. 3. Robot SLAM results

References

[1] C. Harris and M. Stephens: A combined corner and edge detector, Proceedings of the 4th Alvey Vision Conference, (1988) August 31- September 2, Univ. of Manchester, UK.

[2] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse: IEEE T. Pattern Anal. 29, 1052 (2007).

[3] L.M. Paz, P. Pinies, J.D. Tardos, and J. Neira: IEEE T. Robot. 24, 946 (2008).

[4] T. Lindeberg: Int. J. Comput. Vision, 30, 79 (1998).

[5] D.G. Lowe: Int. J. Comput. Vision, 60, 91 (2004).

[6] N. Karlsson, E.D. Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M.E. Munich: Proceedings of IEEE International Conference on Robotics and Automation (2005) April 18-22, Barcelona, Spain.

[7] H. Bay, T. Tuytelaars, and L. Van Gool: Proceedings of The ninth European Conference on Computer Vision, (2006) May 7-13, Graz, Austria.

[8] P.A. Viola and M.J. Jones: Rapid object detection using a boosted cascade of simple features, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2001) December 8-14, Kauai, HI, USA.

[9] A. Gil, O.M. Mozos, M. Ballesta and O. Reinoso: Mach. Vision Appl. 21, 905 (2010).

[10] A.C. Murillo, J.J. Guerrero and C. Sagues: SURF features for efficient robot localization with omnidirectional images, Proceedings of the IEEE International Conference on Robotics and Automation, (2007) April 10-14, Rome, Italy.

[11] Y.C. Feng: Sparse and persistent map for robot visual SLAM based on scale- and orientation-invariant features, Master thesis, Department of Mechanical and Electro-Mechanical Engineering, Tamkang University, New Taipei City, Taiwan (2011).

[12] G. Shakhnarovich, T. Darrell and P. Indyk: *Nearest-neighbor methods in learning and vision*, The MIT Press, MA, USA (2005).

[13] Y.T. Wang, C.T. Chi, and S.K. Hung: Adv. Sci. Lett. 8 (2012).