Aerial Robot Simultaneous Localization and Mapping with a Monocular Vision

Yin-Tien Wang*, Ting-Wei Chen, Jr-Syu Yang

Department of Mechanical and Electro-Mechanical Engineering, Tamkang University, New Taipei City 251, Taiwan

*Corresponding Author: ytwang@mail.tku.edu.tw

Abstract

The objective of this study is developing a visual simultaneous localization and mapping (vSLAM) system using a monocular vision. The vSLAM system could provide the aerial robot the capabilities of autonomous navigation in global positioning system (GPS) denied environments. When the aerial robot navigates in a GPS-denied environment, the vSLAM system collects the measurement information for robot state estimation and environmental mapping. Considering the carrying capacity of the aerial robot, a single camera is used in this study, and the image is transmitted to PC-based controller for image processing. The extended Kalman filter is used as the state estimator to recursively predict and update the states of the aerial robot and the environment landmarks.

The map scale determination problem of monocular vision is solved by using a one-dimensional ultrasonic distance sensor. Meanwhile, the image depth is represented by using the method of inverse depth parameterization. The robot vSLAM system integrates the sensor inputs, image processing, and state estimation. The resultant system is used to perform the tasks of simultaneous localization and mapping for aerial robots.

Keywords: Visual simultaneous localization and mapping (vSLAM), image depth parameterization, data association, map management.

1. Introduction

When an autonomous robot is navigating in the environment, it relies on sensors to recognize the outside world and then estimate the state of the robot itself. Many kinds of sensors can be used including the laser range finder, global positioning system (GPS), and vision sensor. The vision sensor has a reasonable cost and is generally used as a robot's sensing device, especially in a GPS-denied environment. Only single camera is used in this study, as shown in Fig. 1, because of the carrying capacity of the aerial robot. The monocular vision sensor captures only two-dimensional images and lacks for the depth information of environmental objects. Therefore, the spatial coordinates of a new landmark could not be determined and the map scale of the environment cannot be estimated initially. Many researchers have developed landmark initialization procedures either in time-delayed method or un-delayed method for the monocular vision[1,2]. The un-delayed method will be employed in this research. When an image feature is selected, the spatial coordinate of the image feature is represented using the method of inverse depth parameterization[2]. Meanwhile, an ultrasonic sensing system is developed in this paper to provide one-dimensional distance measurement and solve the map scale determination problem of monocular vision.

The image features detected from the vision sensor can be used to represent the landmarks in the environment and build an environmental map. In this study, a detection method based on the scale-invariant feature was employed[3]. This method, called speeded-up robust features (SURFs), significantly reduces the calculation time and robustly detects the features from monocular RGB images[4]. Meanwhile, the extended Kalman filter (EKF) is used to recursively predict and estimate the robot state as well as the environmental landmarks[5].

The remaining sections of the paper are organized as follows. Section 2 describes the structure of the aerial robot SLAM. Visual measurement and the method of SURF-based mapping are presented in Sections 3 and 4, respectively. Section 5 depicts the experimental applications and results.
Finally, concluding remarks are given in the last section.

2. Aerial Robot SLAM

In this study, a monocular vision system is used as the only measuring device in the state estimation algorithm for robot SLAM tasks. The camera is carried by the aerial robot and treated as a free-moving system with unknown inputs. System states are estimated using the EKF estimator to perform the SLAM tasks. The state vector of a system at time step $k$ is

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1})$$

where $x_k$ is the state vector, $u_k$ is the input, and $w_k$ is the process noise. The state vector contains the states of the robot and landmarks,

$$x = [x_c^T, M^T]^T = [x_c^T, m_1^T, m_2^T, \ldots, m_n^T]^T$$

where $x_c = [r^T, \dot{r}^T, \omega^T]^T$ denotes the robot position and velocity in world frame, and $m_i$ represents the $i$th landmark in the environment map $M$. The objective of the robot SLAM tasks is to estimate the state $x_k$ of the target recursively according to the measurement $z_k$ at $k$,

$$z_k = g(x_k, v_k)$$

where $v_k$ is the measurement noise. Since the sensor frame is set at the center of the camera, the coordinates of $i$th observed image feature in the world frame (Fig. 2) is

$$m_i = r + h_i^W = r + R h_i^C$$

where $r$ is the position vector of the sensor frame; $R$ is the rotational matrix from the world frame to the sensor frame; $h_i^W$ and $h_i^C$ are the ray vectors of $i$th image feature in the world and sensor frames, respectively. Because of the lack of one-dimensional range information in monocular vision, how to initialize the image features as new landmarks becomes an important topic.

3. Monocular Visual Measurement

3.1 Inverse Depth Parameterization

The un-delayed method is used in this research to initialize a new landmark of the environment map. When an image feature is selected, the spatial coordinates of the image feature are calculated by employing the method of inverse depth parameterization. Assume that there are $n$ image features with position vectors, $m_i$, $i = 1, \ldots, n$, which are described by the 6-dimensional (6D) state vectors

$$\hat{m}_i = [\hat{r}^W_i, \hat{\psi}^W_i, \hat{\eta}^W_i, \hat{\theta}^W_i, \hat{\phi}^W_i, \hat{\rho}^W_i]^T$$

where $\hat{\rho}^W_i$ indicates the estimated state of the camera when the feature was observed, as shown in Fig. 2; $\hat{\rho}_i$ is the estimated image depth of the feature; $\hat{\phi}^W_i$ and $\hat{\psi}^W_i$ are the longitude and latitude angles of the spherical coordinate system which locates at the camera center. To compute the longitude and latitude angles of a normalized vector $\hat{\eta}_i^W$ in the direction of the ray vector is constructed by using the perspective project method:

$$\hat{\eta}_i^W = R(\hat{\phi}_i^W) \begin{bmatrix} I_{ix} - u_0 \\ I_{iy} - v_0 \\ 1 \end{bmatrix}^T$$

Therefore, from Fig. 2, the longitude and latitude angles of the spherical coordinate system can be obtained as

$$\hat{\phi}_i^W = \tan^{-1} \left( \frac{\hat{\eta}_{iy}^W}{\hat{\eta}_{ix}^W} \right)$$

$$\hat{\psi}_i^W = \tan^{-1} \left( \frac{\sqrt{\hat{\eta}_{ix}^W \hat{\eta}_{iy}^W}}{\hat{\eta}_{iz}^W} \right)$$

Fig. 1. Quadrotor aerial robot (left) and monocular vision sensor (right).

Fig. 2. Coordinate of monocular vision sensor.
When image features are selected to be new landmarks, the inverse depth parameterization vector in eq. (5) is assigned to be new augmented states in the EKF-based SLAM.

3.2 Determination of Map Scale

For determining the map scale in monocular SLAM problem, we developed a one-dimensional distance detector based on the ultrasound technology. The distance detector consists of an ultrasound sensor chip (HC-SR04), a radio frequency transmitter (3Dr Telemetry), and a microchip (Arduino Nano). When the aerial robot is taking off, the ultrasound sensor is designed to measure the distance from the ground. The SLAM task begins to work if the height of the quadrotor is 1.5 m from the ground. At the beginning of the SLAM task, some SURF features obtained from the first image are chosen as the map landmarks and their states are initialized according to eq. (4). In the equation, the depth information of these SURF features is obtained from the ultrasound sensor. With these initial SURF features, the map scale is also calculated. After the map scale is obtained, the ultrasound sensor is turned off and the newly added landmarks are initialized by using the 6D state vector in eq. (5).

4. SURF-Based Mapping

Robot visual mapping requires a robust method of representing visual landmarks detected in an image. In this study, we used the SURF method to detect and represent visual landmarks for robot mapping during SLAM tasks. The SURF method developed by Bay et al. uses a box filter instead of a difference of Gaussians to approximate the determinant of the Hessian matrix\(^{(4)}\). The box filter is further combined with the integral image method to reduce the image processing time\(^{(8)}\). After the features are detected from the image, the description vector is computed to represent feature characteristics. A high-dimensional description vector is used to describe the uniqueness of the feature. For matching high-dimensional description vectors, the most popular method is the nearest-neighbor search method\(^{(9)}\). The criterion for matching two image features is usually to determine the smallest Euclidian distance between their descriptors.

For implementing the navigating tasks, the monocular vision is integrated with the free-moving motion model, the measurement model, and the SURF detection algorithm to form a SLAM system. Once the images are captured by the camera, image features are detected by using the SURF method. The system performs data association of the map landmarks and the image features using the proposed matching criterion. A map management system is also designed to coordinate the newly added features and the “bad” features in the system. New features are chosen as landmarks and added to the map when the robot explores an unknown environment. The state variables of all new landmarks are augmented in the state vector in eq. (1). However, features that are not continuously detected during the task are considered as “bad” features and are deleted from the state vector.

5. Experimental Results

The SLAM experiment has been implemented on a quadrotor aerial robot to validate the proposed algorithms. In this experiment, the monocular vision is carried by the quadrotor to follow a forward-backward trajectory of 4 m length, as shown in Fig. 3. The camera lens always faces downward when the aerial robot is flying along the forward-backward trajectory 2.5 times. The SLAM system starts when the quadrotor's height is 1.5 m which is measured by the ultrasound sensor. Fourteen SURF features from the first image are chosen as map landmarks and their state vector is initialized according to eq. (4) in which the image depth information is obtained from the ultrasound sensor. After that new landmarks are added to the map constantly and their state vectors are initialized using eq. (5). As the monocular vision follows the forward-backward trajectory, the SLAM system concurrently builds the environment map and estimates the robot pose. Figure 4 shows the 3000th RGB image frame obtained in the experiments. At this frame, 13 landmarks are detected and the map size is increased to 170. The top-view (xy-plane) and side-view (yz-plane) plot of the environmental map is depicted in Figs. 5 and 6, respectively. The map size is increased to be 140 landmarks at the end of the first forward-backward trip, be about 160 landmarks at the end of the second trip, and be 176 landmarks at the end.
of the experiment. The average sampling frequency is about 15Hz and the lowest frequency is about 7Hz. During the SLAM task, the average pose deviation is less than 10 cm, and the highest peak is about 20 cm.

6. Conclusions

We developed an algorithm for aerial robot simultaneous localization and mapping using a monocular vision. In this study, we solved the problem of determining the map scale as well as initializing new landmarks by utilizing an ultrasound range detector. For the aerial robot SLAM system, the map scale was determined from the pixel coordinates of image features and the distance information provided by an ultrasonic sensing system. For each observed landmark, the state was initialized by one 6D vector using inverse depth parameterization method. The experiments were carried out to validate the performance of the vector aerial robot SLAM systems.

Acknowledgment

This work was partially supported by the Ministry of Science and Technology, Taiwan, under grant nos. MOST103-2221-E-032-041, MOST103-2632-E-032-001-MY3, and MOST104-2221-E-032-020.

References