Robust object tracking for underwater robots by integrating stereo vision, inertial and magnetic sensors

Norimitsu Sakagami  
School of Marine Science and Technology  
Tokai University  
Shizuoka, Japan 424-0902  
Email: sakagami@scc.u-tokai.ac.jp

Song K. Choi  
Department of Mechanical Engineering  
University of Hawaii at Manoa  
Honolulu, HI, USA 96822  
Email: schoi@hawaii.edu

Abstract—As described in this paper, we propose a robust tracking system with a stereo vision, IMU, and magnetic sensor for underwater robots to estimate the position of a stationary object even when the object is out of the field of view (FOV) of a camera or occlusion occurs. During maneuvering of underwater robots, they are subject to external disturbances. Their cameras are always moving. Therefore, objects in an image move out of the FOV of a camera unexpectedly.

As described in this paper, we integrate an IMU and a magnetic sensor with a stereo vision sensor to estimate an object’s position robustly. In the process of the estimation, Kalman filters are used to estimate the sensor system orientation. First, we conduct numerical analyses to investigate the estimation accuracy of the proposed method against sensor noise. Then, a preliminary experiment is conducted to demonstrate the robustness of the proposed method against occlusion and moving out of the FOV. As a first step, we examine the tracking performance with regard to rotational motion of the sensor system.

I. INTRODUCTION

Sensor fusion is applied in various research fields such as augmented reality (AR), human measurements, and mobile robots. Integrating multi-sensor data based on the benefits of each sensor can achieve more wide range, accurate, rapid, and robust measurement than measurement by a single sensor. In AR research, sensor fusion based on vision sensing is used to track objects correctly in a real image sequence and to overlay virtual images on it (e.g., [1],[2],[3]). In the area of human measurement, vision-based sensor fusion is used to measure human body motion[4]. Sensor fusion for mobile robots with multi-sensors (e.g., inertial measurement unit (IMU), magnetic compass, vision sensor) is apparent in many reports of the relevant literature[5][6].

Inertial sensors and visual sensors are mutually complementary and mutually compatible. Actually, inertial sensors can quickly respond to dynamic motions, and can measure widely various motions. Nevertheless, they are suited to short-term sensing because their sensor data include inherent noise and bias drift. In contrast, a vision sensor is generally preferred for long-term sensing but its measurement range is limited in a camera’s FOV. Vision sensing is not robust against occlusion or blurs. In some reports of the relevant literature[3][7][8], fusing inertial sensors with vision sensors improves the robustness and accuracy of object tracking based on their benefits.

As presented herein, we describe the integration of stereo vision, IMU, and magnetic sensors for application to underwater robotic systems. Free-floating underwater robots are subject to external disturbances. Therefore a target object is not always visible in the FOV of their cameras. To extend the vision sensor capabilities, we use an IMU and magnetic sensors for underwater object tracking. Kalman filters are used to integrate these measurement data to provide continuous target positioning even when a target is out of the FOV.

The developed tracking system is planned for implementation on an underwater robot, for use in assisting 1) automatic maneuvering control and 2) the human support interface.

The contributions of this paper are three-fold: first, we quantitatively investigate the estimation accuracy of the proposed method when an object is out of the FOV of the camera system and occlusion occurs. Second, we develop a sensor system composed of stereo vision, and inertial and magnetic sensors. Then we conduct a preliminary experiment in a laboratory. As the initial step in developing the sensor system, we examine the tracking performance with regard to the rotational motion of the system. Finally, we assess the possibility of tracking underwater archaeological objects using a feature detection algorithm.

The outline of this paper is the following. Section 2 describes related works on fusion of IMU and vision sensing, and our approach. Section 3 presents discussion of the proposed integration method. In Section 4, we describe numerical analysis undertaken to investigate the estimation accuracy against sensor noise. We also demonstrate the effectiveness of the developed system in a preliminary experiment in Section 5. In Section 6, we discuss the possibility of underwater object tracking.

II. RELATED WORK

Sensor fusion is divisible roughly into two approaches: sensor fusion with a Kalman filter and without a Kalman
When an underwater robot translates and/or rotates, the position estimation using a detection algorithm for the tracking of underwater archaeological objects. Kalman filters are used for robust visual tracking. Another approach [8] is to use the inertial sensor only when the vision system is unreliable.

Regarding sensor fusion using Kalman filters, in [10], the authors proposed online multi-rate Extended Kalman Filter (EKF) to fuse vision and inertial sensor data to estimate the absolute position and velocity of vehicles in a monocular SLAM framework for autonomous robots. In [11], two sensor fusion methods based on EKF were developed to estimate the pose of the monocular camera with the IMU sensor. One report of the literature [3] describes a robust tracking system that combines the IMU sensor and camera for Mobile Augmented Reality (MAR) applications. A relative position estimator for AUVs [6] was developed by fusing a monocular camera with an accelerometer and a rate gyroscope based on EKF.

As described in this paper, we developed a robust tracking system with a stereo vision, IMU, and magnetic sensor. This tracking system is expected to be useful for automatic maneuvering control of underwater robots and observation support for human operators. The developed system estimates the position of an object and presents the object position, even when an object is out of the FOV of the camera system and occlusion occurs. Kalman filters are used for position estimation. Moreover, we discuss the possibility of using a detection algorithm for the tracking of underwater archaeological objects.

### III. PROPOSED APPROACH

This section presents the sensor integration of an IMU with magnetic and vision sensors.

#### A. Coordinate system

As presented in Figure 1, we define four coordinate frames: the robot-fixed frame $\Sigma_r$, the camera-fixed frame $\Sigma_c$, the IMU-fixed frame $\Sigma_s$, and the image frame $\Sigma_{image}$. The camera and IMU sensors are fixed in an underwater robot. The vision sensor is used to measure the object position. The IMU and magnetic sensors are used to calculate the robot orientation.

![Fig. 1. Relation among four coordinate frames.](image)

#### B. Position estimation

We assume that an object is fixed in an inertial reference frame and that the camera system mounted on a robot measures the depth $z_c$ using a conventional depth estimation method based on a stereo matching approach. The relation between the object position $p_c \in \mathbb{R}^{3\times1}$ in the camera-fixed frame and the object position $p_i \in \mathbb{R}^{2\times1}$ on the image plane is given as presented below. [12]

$$ p_c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = K^{-1} \begin{bmatrix} sx_i \\ sy_i \\ s \end{bmatrix} = K^{-1} \begin{bmatrix} sp_i \\ s \end{bmatrix} \quad (1) $$

Therein, $K \in \mathbb{R}^{3\times3}$ is a camera intrinsic matrix; $s$ is the homogeneous scaling factor.

When an underwater robot translates and/or rotates, the position of the object $p_s \in \mathbb{R}^{3\times1}$ in the IMU-fixed frame can be given as shown below.

$$ \begin{bmatrix} p_s \\ 1 \end{bmatrix} = \begin{bmatrix} R_s & t_s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_{cs} & t_{cs} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_c \\ 1 \end{bmatrix} = T_s T_{cs} \begin{bmatrix} p_c \\ 1 \end{bmatrix} \quad (2) $$

where $R_s \in \mathbb{R}^{3\times3}$ and $t_s \in \mathbb{R}^{3\times1}$ respectively denote the rotation matrix and translation vector updated using the IMU sensor and magnetic sensors. $R_{cs} \in \mathbb{R}^{3\times3}$ and $t_{cs} \in \mathbb{R}^{3\times1}$ respectively represent the constant rotation matrix and translation vector between the camera-fixed frame and the IMU-fixed frame.

The inverse matrices of the above matrices are useful to estimate the position $p'_r$ on the image after translation and/or rotation of the robot. The estimated position can be described as shown below.

$$ p'_r = K \begin{bmatrix} I[0] & T_{cs}^{-1} \end{bmatrix} \begin{bmatrix} p_s \\ 1 \end{bmatrix} \quad (3) $$

Therein, $I \in \mathbb{R}^{3\times3}$ is a unit matrix.

When considering underwater robot control and manipulation missions[13], one must estimate the object position $p_r$ in the robot-fixed frame $\Sigma_r$ defined as

$$ \begin{bmatrix} p_r \\ 1 \end{bmatrix} = \begin{bmatrix} R_{sr} & t_{sr} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_s \\ 1 \end{bmatrix} \quad (4) $$

#### C. Inertial and magnetic sensors

The IMU and magnetic sensors, as well as vision systems are commonly used for underwater robots. As explained in
Figure 2, in this paper, we use IMU (which combines an accelerometer and a gyroscope) and magnetic measurement data to calculate the rotation $R_s$. We estimate the position of a stationary object when a vision system does not work. Therefore, accurate measurement of the IMU, magnetic sensor is extremely important. In general, accelerometer output is adversely affected by noise; moreover, gyroscopes are subject to drift. To overcome such shortcomings, a Kalman filter [14] is used for accurate estimation of the robot orientation.

To shorten the processing time of a micro-controller used in experiments in Section V, we assume that the state equation for roll angle is linear and given as

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k,$$

where $k$ is the time step, $\hat{x}_k \in \mathbb{R}^{2 \times 1}$ is the estimated state (estimated roll angle $\theta_r$ and gyroscope bias $\dot{\theta}_r$), $A \in \mathbb{R}^{2 \times 2}$ is the state transition matrix, $B \in \mathbb{R}^{2 \times 1}$ is the control matrix, and $u_k$ is the control variable (gyroscope measurement $\dot{\theta}_r$).

The measurement update is given as

$$\hat{x}_k = \hat{\hat{x}}_k + \hat{K}_k (z_k - H\hat{x}_k),$$

where $H \in \mathbb{R}^{1 \times 2}$ is the measurement matrix, $\hat{K}_k \in \mathbb{R}^{2 \times 1}$ is the Kalman gain, and $z_k$ is measurement variable (roll angle $\dot{\theta}_r$) from the accelerometer. We also use the same state equation and the measurement update equation for pitch angle estimation.

We integrate the magnetic sensor measurement with the gyroscope measurement for quick response. We also use the same equations, but the angle value from the accelerometer above is replaced by the angle value from the magnetic sensor to estimate the yaw angle.

\[ D. Vision sensor \]

The stereovision sensor can detect an object in the FOV and estimate its three-dimensional position using conventional image processing algorithms. As presented in Figure 2, the translation vector $t_s$ is updated using the stereo vision sensor when an object is on an image. However, the vector can not be updated when an object is out of the FOV. Therefore, we assume that vector $t_s$ is constant when an object is out of the FOV for this study. However, this assumption is somehow reasonable because the object displacement caused by translation during maneuvering of an underwater robot can be ignored compared with the displacement caused by rotation.

Underwater vision sensing entails one more difficulty. It is difficult to detect underwater objects because they are sometimes visually imperceptible. For underwater vision sensing, we have been testing the SURF algorithm[15], which is a fast and robust algorithm in general to detect key points of an image. However, our code is not sufficiently fast to detect underwater objects for real time operations at this moment. We must develop a faster algorithm to detect underwater objects. We will describe the possibility of detection for underwater objects in the Discussion section of the paper. Therefore, in this paper, we apply a simple template matching on the selected ROI of an image to detect an artificial object for a preliminary experiment.

\[ IV. NUMERICAL ANALYSIS OF ACCURACY \]

In this section, we describe numerical examination of the effects of sensor noise on the estimation accuracy.

Using Eqs. (2) and (3), one can estimate an object position $p'_i$. In practice, the matrices in these equations can be considered intuitively to include some noise because their components are calculated using sensor measurements. Therefore, the sensor noise effects on the estimation accuracy are numerically investigated before developing a hardware system.

We assume that an object is in the center of an image, and that the rotational matrix in Eq. (2) is obtained from the measurements with random noise of the sensor system (accelerometer, gyroscope, and magnetometer). When the sensor system inclines, the object position in the image can be estimated using Eqs. (2) and (3). However, roll $\theta_r$, pitch $\theta_p$, and yaw $\theta_y$ angles in the rotational matrix respectively entail random noise $\Delta\theta_r$, $\Delta\theta_p$, and $\Delta\theta_y$. The random noise was generated by the Mersenne Twister algorithm[16]. In this analysis, we consider two amplitudes of noise, for which the ranges are $[-1.0$ and $1.0$ or $[-2.0$ and $2.0$.

Figure 3 presents the analysis result in the image frame $\Sigma_{image}$. Black dots show true positions when the sensor measurements contain no noise. The gray and light gray dots
respectively denote the estimation results with random noise between -1.0 and 1.0, and between -2.0 and 2.0.

Although it is possible to estimate object positions using Eqs. (2) and (3) from this analysis, as an object moves away from the image center, the estimation accuracy worsens because of the presence of noise on sensor measurements. Therefore, the Kalman filters are important for estimation with high certainty.

V. PRELIMINARY EXPERIMENTS

Our preliminary experiment tested the performance of the developed system in the laboratory.

A. Experimental setup

The developed sensor comprises an accelerometer-gyroscope package sensor (ITG3200/ADXL345), a magnetometer (LSM303D), and USB cameras (UCAM-C0220FE; Elecom Co. Ltd.) as portrayed in Figure 4. Image processing was performed using a laptop computer with a 1.9 GHz CPU (Celeron; Intel Corp.) and 4 GB RAM. An image template matching was used for detecting a marker (a target object).

B. Experiment: estimation error

In the same manner as the analysis in Section IV, we experimentally estimated the object position to investigate the estimation accuracy. The object was in the center of an image, and the rotational matrix in Eq. (2) was obtained from the sensor measurements. When the sensor system inclined, the object position in the image was estimated using Eqs. (2) and (3). The initial distance between the camera and the object was 780 mm.

Figure 5 and the upper part of Table I show the experimentally obtained results in the image frame. The yellow plots present the proposed method results with actual sensor noise. The red plots denote the results obtained using the conventional template matching. Table I shows the RMS estimation errors in pixels and in millimeters between the estimated and the actual positions.

C. Experiment: tracking results

The developed system was rotated about the pitch and yaw axes with time to demonstrate its performance. Figures 6 and 7 show the experimentally obtained results. The yellow plots present the proposed object tracking. The red plots show the results obtained using the conventional template matching, which worked only in the FOV.

Figure 6 presents the tracking result. Even when the object was out of the FOV of the camera, the sensor system continued tracking it (yellow plots). Figure 7 shows the result obtained when occlusion occurred. The obstacle (human hand) was always included in the images during the experiment. The
proposed method continued estimation of the object position when the object was hidden behind the obstacle.

From these experimentally obtained results, we verified that the proposed sensor system tracks an object when it is in and out of the FOV of the camera and when occlusion occurs.

VI. Discussion

We assumed that the integrated sensor detected a target object using image processing in this paper. Therefore, in the experiment, we used an artificial object to be easily detected using image processing. Nevertheless, it is difficult in practice to recognize underwater objects automatically.

Many object recognition algorithms have been proposed to date. Some algorithms[17][15][18] are robust, fast, and widely used for field environments. We have been testing the SURF algorithm[15], a robust feature detection algorithm, to detect underwater archaeological objects[19]. Figures 8 and 9 show matching performance using the SURF algorithm (in OpenCV 2.3) and Visual Studio 2010 (Microsoft Corp.) for video sequences captured by our ROV at the Yarabuki archaeological site[20] at Ishigaki Island, Japan. The template image was the first image of the video sequences. The SURF algorithm found out key points in the images and the center of the key points was treated as the center of the objects. The speed of the algorithm was not sufficiently fast for real time operations at this moment. It is necessary to develop a faster algorithm to detect underwater objects as future work.

VII. Conclusion

This paper presented a robust tracking system with vision, IMU, and magnetic sensors for underwater robots to avoid missing a stationary object. We integrated these sensors to estimate the object position even when occlusion occurs and when the object moves out of the FOV. During the process of the estimation, the Kalman filters output the orientation of the sensor system. We investigated the estimation accuracy numerically and conducted preliminary experiments in the laboratory. The experimentally obtained results show that the proposed system provided good tracking even when an object was beyond the FOV of the camera.

As described in this paper, the translation of the underwater robot was approximated as a constant value when the object was out of the FOV of the camera. However, image processing based on image features around an object is expected to be useful to estimate the translation vector (travel of the underwater robot). Additionally, we must develop a faster detection algorithm for application to underwater objects.

ACKNOWLEDGMENT

The authors would like to thank Tokai University which provided financial support (overseas research visit program) for the research activity.

REFERENCES

Fig. 8. Image detection for underwater archaeological objects (anchor).

Fig. 9. Image detection for underwater archaeological objects (vases).

