Vision and Inertial-Based Image Mapping for Capsule Endoscopy

Yasmeen Abu-Kheil*, Gastone Ciuti†, Marco Mura†, Jorge Dias‡, Paolo Dario† and Lakmal Seneviratne *
* Khalifa University of Science, Technology and Research, Abu-Dhabi, U.A.E
†The BioRobotics Institute, Scuola Superiore Sant’Anna, Pisa, Italy
‡Biomedical Engineering, University of Technology, Pisa, Italy
Emails: yasmeen.abu-kheil@kustar.ac.ae, jorge.dias@kustar.ac.ae and lakmal.seneviratne@kustar.ac.ae

Abstract—Capsule endoscopy is a non-invasive procedure for gastrointestinal diagnosis. It does not require sedation and it is comfortable and well tolerated by patient. However, the problem with such procedure is that a huge number of images is collected, which require time to investigate and diagnose; furthermore, the capsule movement is not controlled leading, in some cases, to inaccurate diagnosis. In this context, a mapping of the lumen is required to guarantee a higher reliability of the inspection, enabling the medical doctor to evaluate all the parts of the lumen for a better diagnosis. In this paper, we propose a method for mapping images from a capsule-based endoscope: the technique uses visual and inertial-based data fusion to obtain a 3D map of the lumen from 2D capsule images, also paving the way for the implementation of a path planning and autonomous locomotion and inspection.

Keywords—Capsule endoscopy; mapping; vision-inertial fusion.

I. INTRODUCTION

Standard techniques for endoscopy procedures adopt flexible endoscopes, which are introduced into the oral or rectal orifices. Even if such technique enables reliable diagnosis, it is considered traumatic and frequently poorly tolerated by patients [1]. Capsule endoscopy (CE) allows medical doctors to examine internal organs, i.e. digestive districts, without the need of sedation. CE also offers a noninvasive and painless investigation of the gastrointestinal (GI) tract [2]. However, one of the limitation of CE is the amount of time required by physicians to inspect and diagnose a large amount of images, usually more than 50,000, in one examination; CE examination takes around 90 minutes to review such a large amount of images [3]. Another limitation of the CE is that images may only visualize a limited portion of the GI tract due to the irregularity and unpredictability of the capsule motion [3]. Therefore, there is a need for GI tract map generation to help and support endoscopists in visualizing and perceiving such images to provide an accurate diagnosis in a faster time.

Various methodologies were proposed in the literature to develop 3D maps and surface reconstruction of the GI tract. Traditional methods for vision-based 3D reconstruction require the use of stereo cameras that are not applicable to capsule endoscopy due to size limitations of the capsule, power consumption and transmission rate [4] - [5]. Another approach is monocular 3D reconstruction techniques that are based on static target structure and a fast frame rate able to produce enough features for structure recovery [6] - [7]. In [3], Yichen et al. proposed to reconstruct the inner surfaces of the GI tract; they used a scale invariant feature transform (SIFT) to extract image features between two image frames, applied epipolar geometry to calculate the 3D spatial point location and finally generate meshes by Delaunay Triangulation and texture mapping. In [8], Sun et al. also proposed to reconstruct 3D capsule surface by detecting features and tracking them between one frame to another. The limitation of their approach, as reported by the authors, is that their experimental results were only based on two image sequences. Other researchers proposed to use Shape From Shading (SFS) technique to obtain a 3D structure of the GI tract inner surface by only using a monocular camera. The limitations of SFS to recover GI tract inner surface are the need to estimate and recover the surface albedo prior to the endoscopic procedure [4], [9] and its inability to provide enough depth accuracy for diagnosis.

In this paper, image and 3D motion flow based algorithms [10] are used to extract useful image features and information for image mapping and surface reconstruction. Then, a set of measured optical flow features between two selected images were fused with motion parameters obtained from Inertial Measurements Unit (IMU) to generate a map of the GI tract. This approach was previously proposed by Lobo et al. [10] and recently used for navigation and control of aerial devices [11].

The paper is organized as follows: in section II, some basic fundamentals and system models were described; in section III, the proposed system architecture is described while the experimental approach to verify the proposed system is presented in section IV. Finally, conclusions are given in section V.

II. BACKGROUND

A. Notations

- \( \{B\} \): body frame (inertial)
- \( \{C\} \): camera frame
- \( \hat{P}^C(X, Y, Z) \): 3D image point
- \( \hat{p}^C(x, y) \): 2D image point
- \( (f_x, f_y) \): focal distances
- \( Z \): depth relative to camera frame \( \{C\} \)
- \( \vec{v}^B = (v_x, v_y, v_z) \): linear velocity vector
- \( \vec{ω}^B = (ω_x, ω_y, ω_z) \): angular rate vector
- \( \vec{a}^B = (a_x, a_y, a_z) \): linear acceleration vector
- \( \vec{p}^C = (x, y) \): optical flow
- \( \hat{P}^T = (x_T, y_T) \): translational optical flow

\[ ⃗a = ⃗ω \cdot ⃗v \cdot \frac{1}{T} \]


- **Map**: the generated map between image sequences

### B. Camera Model

Pinhole projection camera model is a typical model that is used to describe the mapping of a 3D model point \(P=(X, Y, Z, 1)\) in a homogeneous coordinates onto the image point \(p=(x,y,1)\), as described in (1):

\[
\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}
\]

### C. Optical Flow Model

The optical flow can be computed by differentiating (1) and after geometrical transformations, the optical flow can be expressed using (2):

\[
\dot{\vec{p}}_C = \frac{1}{Z} \begin{bmatrix} f_x & 0 & x \\ 0 & f_y & y \end{bmatrix} \vec{v}_B + \begin{bmatrix} \frac{xu}{f_v} - \frac{x^2}{f_v} \frac{y f_x}{f_v} \\ \frac{yu}{f_v} - \frac{y^2}{f_v} \frac{x f_y}{f_v} \end{bmatrix} \vec{\omega}_B
\]

As shown in (2), the optical flow is divided into two main parts: i) a translational part that depends on body linear velocities and image depth, and ii) a rotational part that depends on angular rates. Therefore, the optical flow measurements and the angular rate data were used to estimate the translational optical flow, which in turn can give the depth and map sequence.

### III. METHODOLOGY

#### A. General Overview and 3D Mapping Architecture

The proposed technique aims at using visual and inertial data fusion [10] to obtain a 3D map of the GI tract from 2D capsule images. This will provide: i) a better representation of the GI tract for the physician, ii) reduction of the diagnostic time and iii) improvement of the pathological recognition rate. The general capsule locomotion platform and mapping architecture is shown in Fig. 1. The overall system will consist of a six degrees of freedom robotic arm equipped with an external permanent magnet (EPM), a capsule device that includes an internal permanent magnet (IPM) that interacts with the external magnet as well as mapping algorithm. The movement of the capsule is guaranteed by the magnetic field interaction between the EPM, attached to the robotic arm, and the IPM, integrated inside the capsule [12]. The mapping algorithm consists of three main modules: i) the vision module, ii) the inertial module and iii) the fusion module. The detailed description of each module is summarized in sections III-B, III-C and III-D. Then, the mapping algorithm outputs (i.e., acquired images, linear accelerations and angular velocities) will be sent to a remote computer where they are fused and processed. The mapping algorithm computes the optical flow features from the acquired image sequence and then uses a non-linear Kalman filter algorithms to estimate the translational optical flow from the measured angular velocities and the computed optical flow [11]. The translational optical flow can provide the depth and map sequence; processed information will be also fed back to the robotic arm for control aspects.

#### B. Vision Module

The vision module consists of a camera that is equipped with a light source embedded into the capsule prototype. The camera intrinsic and extrinsic parameters were estimated with a standard checkerboard using Matlab Camera Calibration Toolbox [4]. The main function of the vision module is to approximate the capsule motion parameters based on motion changes extracted from the images by computing the optical flow between two consecutive images. The implemented algorithm used to calculate the optical flow was computed from Pyramidal Lucas-Kanade optical flow using Shi-Tomasi [13] - [14] features. The limitations of optical flow computations are the ability to extract useful features during the optical flow computations in real-time and the ambiguity arises from the similarity between the translational and rotational flow fields. Fig. 2 shows the vision module within our prototype capsule.

#### C. Inertial Module

The capsule prototype is also equipped with an IMU system, which measures the orientation and acceleration of the capsule (see Fig. 3). The inertial module consists of 9-axis motion sensing system, a 3-axis digital gyroscope and a logic unit. The advantage of inertial navigation is represented by the robustness to external disturbances. To estimate the linear velocities from the inertial sensors, the linear accelerations are

![Fig. 1: Overall system architecture](image1)

![Fig. 2: Vision module](image2)
The measured IMU linear acceleration and angular velocities are modeled in (3) and (4), respectively.

$$\vec{a}_m = C(\vec{a} - g) + \vec{b}_a + \vec{n}_a \quad (3)$$

$$\vec{\omega}_m = \vec{\omega} + \vec{b}_w + \vec{n}_w \quad (4)$$

where:
- $C$: 3x3 matrix representing the accelerometer axis misalignment
- $\vec{a}$: actual linear acceleration vector
- $g$: gravity vector
- $\vec{b}_a$: accelerometer bias
- $\vec{n}_a$: accelerometer Gaussian measurement noise with zero mean and co-variance matrix $\sum Q_a$
- $\vec{b}_w$: rotational velocities bias
- $\vec{n}_w$: rotational velocities Gaussian measurement noise with zero mean and co-variance matrix $\sum Q_\omega$

**D. Fusion Module**

A non-linear Kalman Filter (KF) algorithm is used to estimate the translational optical flow from the angular velocities and the computed optical flow. Non-linear KF is a recursive algorithm that estimate the translational optical flow based on minimum mean square error estimator [15]. Fig. 4 shows the block diagram of the fusion module. In our implemented Kalman filter:

- The state vector is:

$$S = \begin{bmatrix} \vec{P}_T, \vec{\omega} \end{bmatrix}^T$$

where:

$$\vec{P}_T = \frac{1}{Z} \begin{bmatrix} f_x & 0 & x \\ 0 & f_y & y \end{bmatrix} \bar{v}^B$$

- The measurement vector is:

$$Y = \begin{bmatrix} \vec{P}, \vec{\omega} \end{bmatrix}^T$$

where: $Y_k = HS_k + n_k$. The matrix H can be deduced from (2). For a detailed implementation of Kalman filter equations, you can refer to [11].

**IV. EXPERIMENT AND EVALUATION**

**A. Experimental Setup**

In order to check the feasibility of the proposed approach, a prototype of the capsule system was equipped with:

- iNEMO-M1 STMicroelectronics [16]: an IMU sensor to measure the body linear acceleration and angular velocities. The sensor also has a microcontroller that executes an extended Kalman filter algorithm to retrieve information about the board orientation.

- The vision module consists of a 500 x 582 charge-coupled device camera with 120 degrees field of view (provided by Karl Storz GmbH, Tuttingen, Germany) and four high-efficiency white light-emitting diodes (NESW007BT; Nichia Corp., Tokushima, Japan).

The capsule prototype is then attached to a robotic arm RV-6SL whose movement controls the capsule motions direction and speed. Then, a plastic colon simulator (M40; Kyoto Kagaku Co., LTD, Japan) is placed in front the capsule to allow data collection. The camera, IMU and the robotic arm were also connected to a computer that saves all required sensor readings to be processed. The implemented algorithm has a low computational time compatible with an active locomotion capsule endoscopic procedure. Fig. 5 shows the experimental setup. All of the above mentioned components are compatible with a capsule endoscope system for colonoscopic inspections in terms of dimension and power supply. The experimental setup and systems represent only a functional test bench with the aim to test and evaluate the accuracy and reliability of the methodology; then the entire components will be moved in a capsule-shape engineered device.
B. Experimental Scenarios

We applied the proposed algorithm to two different sets of motion: translational and rotational motions of the capsule, obtained with the robotic arm, as shown in Fig. 6 (a) and Fig. 6 (b), respectively. The translational motion includes forward and background movements along the optical axis of the camera. Additionally, the rotational motion includes rolling and pitching movements. The speed of the capsule were varied from 3 mm/s to 6 mm/s with steps of 1 mm/s during the translational motion. Similarly, during the rolling and pitching movement, the capsule angular speed were approximately set to 5 degrees/s.

![Fig. 6: Experimental scenarios: (a) translational motion and (b) rotational motion.](image)

C. Experimental Results

Table I shows preliminary qualitative results of the generated maps for different motion types: translational and rotational motions. The figures shown in Table I present a consistent, even if qualitative, estimation of the colon simulator structure.

<table>
<thead>
<tr>
<th>Motion Type</th>
<th>Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translational</td>
<td><img src="image" alt="Map" /></td>
</tr>
<tr>
<td>Rotational</td>
<td><img src="image" alt="Map" /></td>
</tr>
</tbody>
</table>

V. CONCLUSION

The feasibility of integrating optical flow and inertial data to obtain a qualitative analysis estimation of colon structure has been preliminary verified. Integrating both vision and inertial data will not only make motion estimation more robust to mistracked image features but also will remove the discontinuities in estimated motion. The presented results are preliminary results that will be verified with a ground truth reference to quantify the accuracy of the approach. Next step is to exploit the mapping data of the lumen towards the implementation of planning and assisted navigation strategies for a magnetically-propelled capsule endoscope.

REFERENCES


