ARTICLE IN PRESS

Computers in Biology and Medicine ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Comparative assessment of feature extraction methods for visual odometry in wireless capsule endoscopy

Evaggelos Spyrou [a,b], Dimitris K. Iakovidis [a], Stavros Niafas [a], Anastasios Koulaouzidis [c]

[a] Dept. of Computer Engineering, Technological Educational Institute of Central Greece, 3rd km Old National Road Lamia-Athens, 35100 Lamia, Greece
[b] National Centre for Scientific Research – Demokritos, Institute of Informatics and Telecommunications, Computational Intelligence Laboratory (CIL), 60037, Patr. Grigoriou and Neapoleos, Agia Paraskevi, Athens, Greece
[c] The Royal Infirmary of Edinburgh, Endoscopy Unit, 51 Little France Crescent, Old Dalkeith Road, Edinburgh EH16 4SA, UK

ARTICLE INFO

Article history:
Received 17 January 2015
Accepted 16 May 2015

Keywords:
Wireless capsule endoscopy
Localization
Visual odometry
Feature extraction
Algorithm
Small bowel
Gastrointestinal tract

ABSTRACT

Wireless capsule endoscopy (WCE) enables the non-invasive examination of the gastrointestinal (GI) tract by a swallowable device equipped with a miniature camera. Accurate localization of the capsule in the GI tract enables accurate localization of abnormalities for medical interventions such as biopsy and polyp resection; therefore, the optimization of the localization outcome is important. Current approaches to endoscopic capsule localization are mainly based on external sensors and transit time estimations. Recently, we demonstrated the feasibility of capsule localization based—entirely—on visual features, without the use of external sensors. This technique relies on a motion estimation algorithm that enables measurements of the distance and the rotation of the capsule from the acquired video frames. Towards the determination of an optimal visual feature extraction technique for capsule motion estimation, an extensive comparative assessment of several state-of-the-art techniques, using a publicly available dataset, is presented. The results show that the minimization of the localization error is possible at the cost of computational efficiency. A localization error of approximately one order of magnitude higher than the minimal one can be considered as compromise for the use of current computationally efficient feature extraction techniques.

1. Introduction

Wireless capsule endoscopy (WCE) is a non-invasive procedure for the visual inspection of the gastrointestinal (GI) tract [1]. It is performed with a miniaturized, swallowable endoscope packed into a capsule of the size of a large vitamin pill. It is equipped with one or more color cameras and during its journey to the anus it wirelessly transmits color images to an external receiver using radiofrequency (RF) signals [2].

The knowledge of the exact location of the capsule endoscope within the GI tract is important for the localization of abnormalities. Since WCE is currently used only for screening purposes, any medical intervention for biopsy or treatment is performed at a secondary phase. For example, if a lesion, such as a polyp, is detected in the small bowel by WCE, a biopsy and/or lesion resection can be performed at a second phase with a more invasive technique, such as surgery or device-assisted enteroscopy [3]. Therefore, the accurate estimation of the location of the lesion by WCE increases the yield and outcome of subsequent interventions. In the case of an open surgical procedure this can result in minimal loss of intestine tissue during the operation, whereas in the case of double-balloon enteroscopy it can determine the insertion route (antegrade or retrograde) of the enteroscope as well as the insertion depth so as to minimize the patients' discomfort [4,5].

In clinical practice, the location of an abnormality can be approximated by the estimation of the expected transit time of the capsule with respect to anatomic landmarks, and decision-making is based on a devised time index [6]. However, this localization approach can be very inaccurate, especially if the capsule cannot visualize the cecum [7,8].

In addition to the transit time estimation, commercial WCE platforms provide the localization of the capsule in a 2-dimensional (2D) projection of the body with respect to the umbilicus, e.g., the abdominal quadrant where the capsule is located. This is accomplished by a wearable RF sensor array, which receives the RF signals transmitted by the capsule while it wirelessly transmits the images. The estimation of the position of the capsule is based on the triangulation principle, with average position errors ranging between 3.7 and 11.4 cm [9,10]. A more recent approach enables localization of the capsule in the 3-dimensional (3D) human body space with an average localization error of 13.3 cm^3 [11].

RF signals are influenced by tissue densities, juxtaposition of different organs, and other anatomic considerations which can affect capsule localization accuracy [6]. Considering that magnetic waves

http://dx.doi.org/10.1016/j.compbiomed.2015.05.013
0010-4825/

are less sensitive in the presence of human tissue, various magnetic localization techniques have been proposed [10]. The most promising ones involve magnetic sensor arrays capable of localizing capsules equipped with a permanent magnet in the 2D or 3D body space. The average position errors reported from ex-vivo experiments range between 1.8 and 10 mm and the average orientation errors between 2° and 3° [9]. Other localization techniques of capsule endoscopes are also possible, e.g., using x-ray and magnetic resonance imaging; however, they require expensive equipment and could result in adverse health effects [10].

A drawback of both the RF and magnetic localization techniques is that they do not directly provide information related to the distance the capsule travels within the GI tract. Having the 2D or 3D body coordinates of the capsule endoscope it is difficult to determine its location with respect to an anatomic landmark, such as the pylorus, needed in the case of a surgical intervention after WCE. To cope with this issue a concept capsule endoscope design with three wheels has been proposed [12,13]. The wheels are mounted on expandable and retractable legs designed to keep contact with the mucosal surface. As the capsule moves within the GI tract the wheels spin and measure the distance traveled by conventional (wheel) odometry.

Recently we proposed a visual odometry approach to capsule endoscope localization that requires neither external sensors nor wheels to measure the distance traveled by the capsule from an anatomic landmark [14]. It is based on the estimation of the motion of the capsule by tracking visual features extracted from consecutive video frames. In this paper, we perform an extensive comparative study to determine the optimal feature extraction scheme that minimizes the displacement and orientation error for capsule endoscope localization. Implementation and time performance issues of the feature extraction methods are addressed and novel results indicating up to one order of magnitude lower error rates are obtained in the majority of the test cases.

The rest of this paper consists of four sections: Section 2 reviews the state-of-the-art capsule endoscope localization approaches that have been based solely on visual features. Section 3 describes the methods compared in this study, and Section 4 presents the results obtained. A summary of conclusions is provided in the last section.

2. Related work

The concept of visual localization of capsule endoscopes appeared in 2008, with techniques based on topographic video segmentation, i.e., the segmentation of the WCE video into a number of consecutive segments that correspond to different parts of the GI tract [15,16]. These techniques were based on the supervised classification of the video frames using color, texture and motion features. This way the different parts of the GI tract, or the transitions between them, such as the esopahgogastric junction, the pylorus and the ileocecal valve, can be automatically recognized. Since then, several variations of these techniques have been proposed [17,18]. Recently the feasibility of using an unsupervised learning technique to solve the localization problem has been demonstrated. This technique is based on Scale Invariant Feature Transform (SIFT) for the extraction of local image features, and on a probabilistic latent semantic analysis model for data clustering [19]. Such topographic video segmentation techniques can be used to infer in which part of the GI tract a capsule is located, and not the exact location of the capsule within each part. The latter can be estimated on the basis of the transit time [15] and/or by application of a more accurate localization technique within the part of the GI tract that is of interest.

Such a localization technique can be based on capsule motion estimation. A motion estimation algorithm relies on the analysis of consecutive video frames to infer how much the capsule has moved, in terms of relative displacement and orientation. This is usually performed in three steps [20]. The first step is feature extraction from each frame. Contemporary feature extraction approaches [21,22] involve both the automatic detection of salient key-points within frames and the estimation of numeric descriptors from the neighborhood of each key-point. In the second step, for each key-point of two consecutive frames its similarity with the key-points of the next frame within a fixed distance is estimated. In the third step the geometric transformation, e.g., scaling and rotation of each frame with respect to its previous one is estimated, and the displacement and orientation of the traveling capsule endoscope can be determined.

In the context of WCE three approaches of this kind have been proposed for the estimation only of the orientation of the capsule endoscope, as a means to assist sensor-based localization [23–25]. Two of them [23,24] were based on the Scale Invariant Feature Transform (SIFT) algorithm [21] and [25] was based on Speeded-Up Robust Features (SURF) [22]. Due to the lack of ground truth data, e.g., measurements of the actual rotation of the capsule within the GI tract, these algorithms were validated with simulation data, produced by manual rotation of real WCE images at various degrees. In [23] rotation estimation was based on the Lucas–Kanade–Tomasi (KLT) feature tracker [26], and in [24] it was based on Singular Value Decomposition (SVD) of a correlation matrix between point sets. Both of these approaches were able to estimate the orientation of the capsule with reasonable errors for rotations of up to 20–25°. In [25] the Random Sample Consensus (RANSAC) algorithm [27] was used for the estimation of the geometric transformation between consecutive frames, resulting in reasonable orientation errors for rotations of up to 35° [28].

Simulation data including not only rotation but also scaling (to simulate the forward or backward displacement of the capsule) were used in our recent works for capsule localization by visual odometry [14,28]. In these works we showed the feasibility of motion estimation to determine both the orientation and the displacement of a capsule endoscope. They were based on an extended SURF-RANSAC scheme capable of estimating not only the rotation but also the scaling of the video frames with a reasonable error for scaling factors between 0.2 and 5.

An emulation platform [29] was considered as an alternative to the simulation approach used for the validation of the localization of the capsule endoscope. This was created by bending, twisting and painting a 1.5 m plastic tube with a diameter of 3 cm, similar to the small intestine. A wired camera was used instead of a capsule. Motion estimation was based on an extension of the SIFT-SVD scheme [24] for the estimation of both the orientation and rotation of the capsule endoscope. The emulation platform enabled absolute distance measurements in cm, and showed that the Mean Square Error (MSE) reached up to 4 cm for the total length of the tube.

A preliminary ex-vivo experiment with animal intestine (phantom model) was performed in [30], where we have compared the performance of the SURF and Maximally Stable Extremal Regions (MSER) algorithms for feature extraction with respect to forward or backward motion estimation. The results indicated an advantage of the SURF algorithm over MSER, with an average accuracy of 81.5%.

In another study [31] a motion estimation algorithm that follows a different strategy for visual localization of capsule endoscopes was proposed. It uses the Fibonacci searching technique to find the maximum mutual information between two frames and to estimate the corresponding transformation, from which the displacement and the rotation of the capsule can be derived. Its experimental evaluation was based on the publicly available

Please cite this article as: E. Spyrou, et al., Comparative assessment of feature extraction methods for visual odometry in wireless capsule endoscopy, Comput. Biol. Med. (2015), http://dx.doi.org/10.1016/j.compbiomed.2015.05.013