

Novel experimental and software methods for image reconstruction and localization in capsule endoscopy



Authors

Anastasios Koulaouzidis¹, Dimitris K. Iakovidis², Diana E. Yung¹, Evangelos Mazomenos³, Federico Bianchi⁴, Alexandros Karagyris⁵, George Dimas², Danail Stoyanov³, Henrik Thorlacius⁶, Ervin Toth⁷, Gastone Ciuti⁴

Institutions

- 1 Centre for Liver & Digestive Disorders, The Royal Infirmary of Edinburgh, Edinburgh, UK
- 2 University of Thessaly, Department of Computer Science and Biomedical Informatics, Lamia, Greece
- 3 Centre of Medical Image Computing and Department of Computer Science, University College London, London, UK
- 4 The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy
- 5 IBM Research, Almaden California, United States
- 6 Department of Clinical Sciences, Lund University, Malmö, Sweden
- 7 Department of Gastroenterology, Skåne University Hospital, Malmö, Sweden

submitted 16.6.2017

accepted after revision 9.10.2017

Bibliography

DOI <https://doi.org/10.1055/s-0043-121882> |
Endoscopy International Open 2018; 06: E205–E210
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 2364-3722

Corresponding author

Dr Anastasios Koulaouzidis, Endoscopy Unit, The Royal Infirmary of Edinburgh, 51 Little France Crescent, Edinburgh EH16 4SA, United Kingdom
Fax: +077-89-588-408
akoulaouzidis@hotmail.com

ABSTRACT

Background and study aims Capsule endoscopy (CE) is invaluable for minimally invasive endoscopy of the gastrointestinal tract; however, several technological limitations remain including lack of reliable lesion localization. We present an approach to 3D reconstruction and localization using visual information from 2D CE images.

Patients and methods Colored thumbtacks were secured in rows to the internal wall of a LifeLike bowel model. A Pill-Cam SB3 was calibrated and navigated linearly through the lumen by a high-precision robotic arm. The motion estimation algorithm used data (light falling on the object, fraction of reflected light and surface geometry) from 2D CE images in the video sequence to achieve 3D reconstruction of the bowel model at various frames. The ORB-SLAM technique was used for 3D reconstruction and CE localization within the reconstructed model. This algorithm compared pairs of points between images for reconstruction and localization.

Results As the capsule moved through the model bowel 42 to 66 video frames were obtained per pass. Mean absolute error in the estimated distance travelled by the CE was 4.1 ± 3.9 cm. Our algorithm was able to reconstruct the cylindrical shape of the model bowel with details of the attached thumbtacks. ORB-SLAM successfully reconstructed the bowel wall from simultaneous frames of the CE video. The “track” in the reconstruction corresponded well with the linear forwards-backwards movement of the capsule through the model lumen.

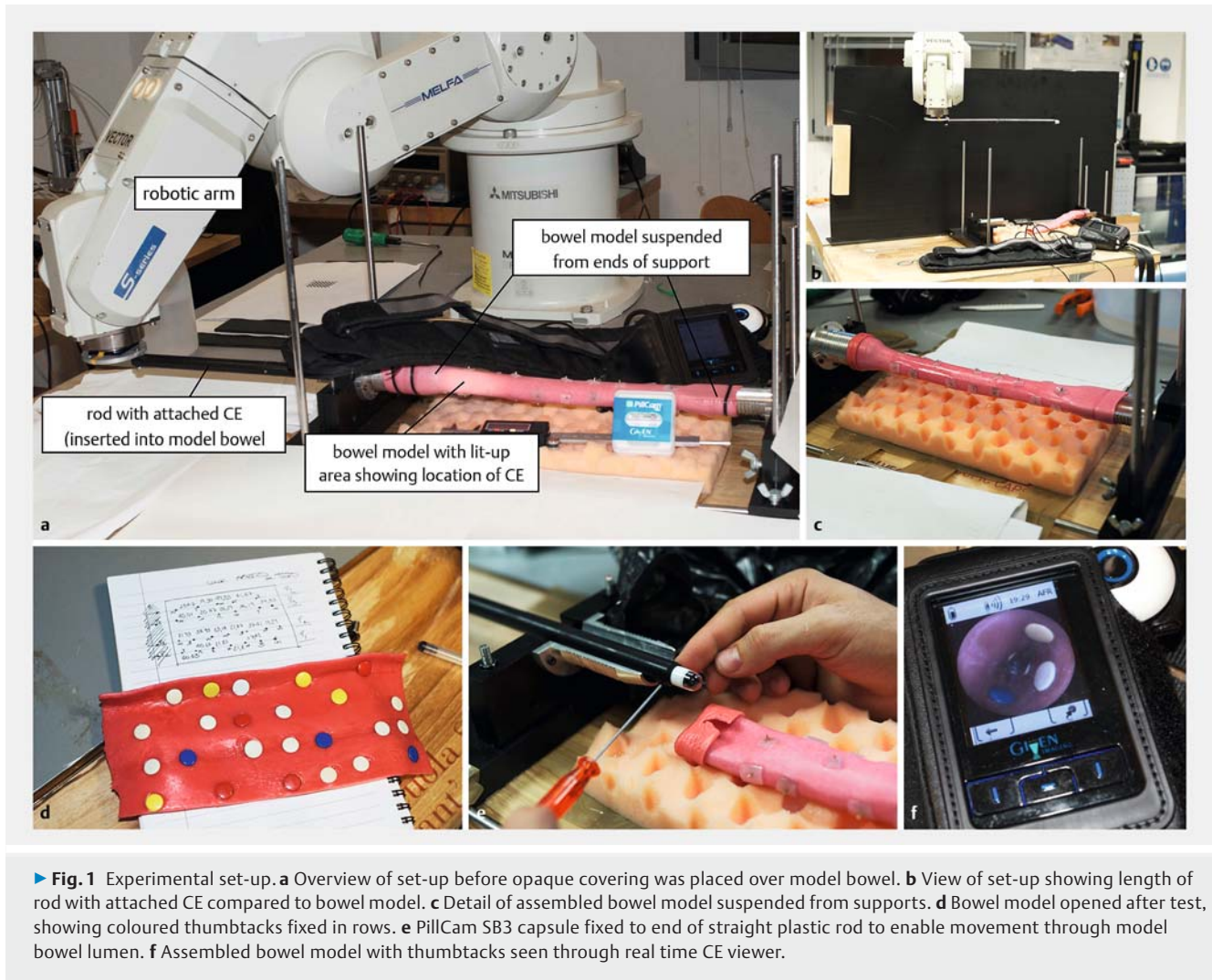
Conclusion The reconstruction methods, detailed above, were able to achieve good quality reconstruction of the bowel model and localization of the capsule trajectory using information from the CE video and images alone.

Introduction

Capsule endoscopy (CE) is a key technology for minimally invasive small-bowel investigation, with good sensitivity for major pathology. Nevertheless, CE continues to face several technological limitations including lack of reliable lesion localization capability [1] and the 2-dimensional (2D) nature of CE images

which hampers lesion characterisation [2,3]. Consequently, it is difficult to determine the precise location of lesions detected within the body. This information is vital to establish prognosis and for treatment planning, e.g. deciding the appropriate route for device-assisted enteroscopy.

We present an approach to localization using visual information derived from CE images, without additional external or in-



► **Fig. 1** Experimental set-up. **a** Overview of set-up before opaque covering was placed over model bowel. **b** View of set-up showing length of rod with attached CE compared to bowel model. **c** Detail of assembled bowel model suspended from supports. **d** Bowel model opened after test, showing coloured thumbtacks fixed in rows. **e** PillCam SB3 capsule fixed to end of straight plastic rod to enable movement through model bowel lumen. **f** Assembled bowel model with thumbtacks seen through real time CE viewer.

ternal sensory devices. Earlier such approaches include topographic video segmentation, i. e. division of video frames into a number of consecutive segments corresponding to different parts of the gastrointestinal (GI) tract [4]. Later approaches were based on motion estimation to localise the CE with respect to anatomical landmarks [1]. We propose a CE localization system based on landmark or feature extraction and tracking in consecutive video frames [5]. This system implements visual odometry to provide estimations of relative movement of the CE during its passage through the GI tract [6]; this information can also be used to achieve 3-dimensional (3D) reconstruction of the bowel lumen.

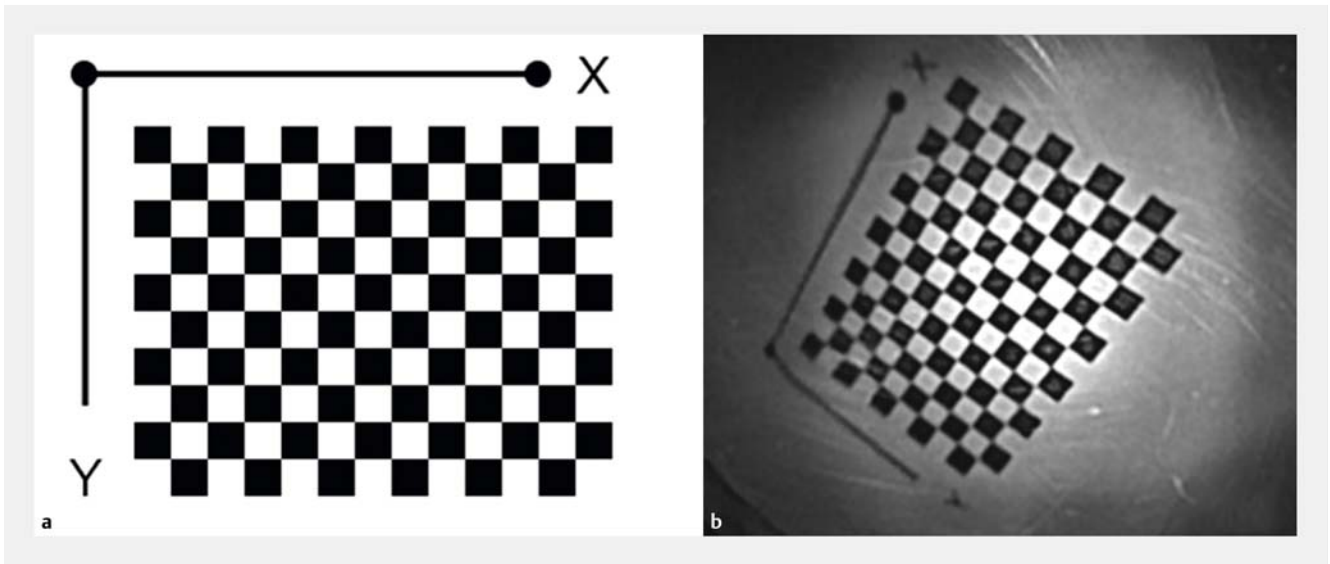
Patients and methods

Experimental procedure

The experiment was performed in a controlled setting using a commercially available capsule fixed to a robotic arm which was used to move the capsule through an in vitro bowel phantom. The setup modules are detailed below (► **Fig. 1**):

- High-precision robotic arm (RV-3SB robot, Mitsubishi, Tokyo, Japan): able to move the capsule forwards and backwards through the bowel phantom at programmed velocities.
- Straight plastic rod attached to the robotic arm, with the capsule fixed to one end; the rod was longer than the total length of the model to allow the capsule to traverse the entire lumen. The capsule was aligned to the center of the lumen.
- Pillcam SB3 (Medtronic, Minneapolis, USA) capsule with camera resolution 320×320 pixels, variable frame rate of 2 to 6 frames per second (fps), and 156° field of view.
- 30-cm lifelike bowel model (LifeLike Biotissue Inc, Ontario, Canada); the model was fixed and suspended in a custom-made support. The internal diameter was about 23 mm, consistent with that of adult humans.

The setup was covered with an opaque plastic box to minimize external illumination, similar to in vivo conditions. The real-time viewer used to show the images captured by the Pillcam SB3 capsule. Colored thumbtacks (diameter 0.95 mm) were se-



► **Fig. 2** Checkerboard pattern used for initial CE calibration. **a** The CE calibration pattern as printed. **b** The CE calibration pattern as viewed from the CE.

cured in four rows along the lumen and the appearance and location of each marker from the rim of the model were carefully documented. Normal gut peristalsis was not simulated at this stage to ensure accurate measurements of distances and therefore the reproducibility of results in this preliminary experiment.

Calibration and estimation of 2D trajectory

Camera calibration is a fundamental process for determining the unknown intrinsic parameters of a camera, such as its focal length. It is used by the 3D reconstruction software to produce estimates of camera position in real-world units (meters). Calibration is usually performed only once, during system development, for a given camera model. Following activation of the PillCam SB3, calibration was performed before beginning the experiment, to correct for lens distortion and calculate the unspecified intrinsic parameters of the camera including focal length. The set-up used images of a chessboard with 3-mm squares arranged in a 10×13 configuration, ► **Fig. 2**. The capsule was mounted on the plastic rod and robotically navigated into the model lumen. It was moved forwards and backwards in a straight line through the length of the model. Several passes were made at different constant velocities of 0.5 to 8 mm/second.

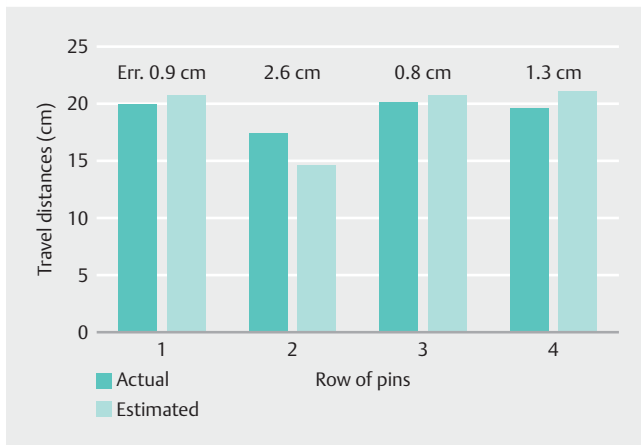
Calibration was performed using Kannala and Brandt's method, best suited for the calibration of conventional, wide-angle and fish-eye lenses [7]. The motion estimation algorithm detects corresponding points of interest (POI) in consecutive video frames; represented by the drawing pins lining the bowel wall. Relative distances between the POI and camera lens were used to estimate actual distances travelled by the capsule. The mean absolute error (MAE) of localization was used to quantify accuracy, calculated as the mean of the absolute difference between estimated and actual travel distances of the capsule.

3D reconstruction

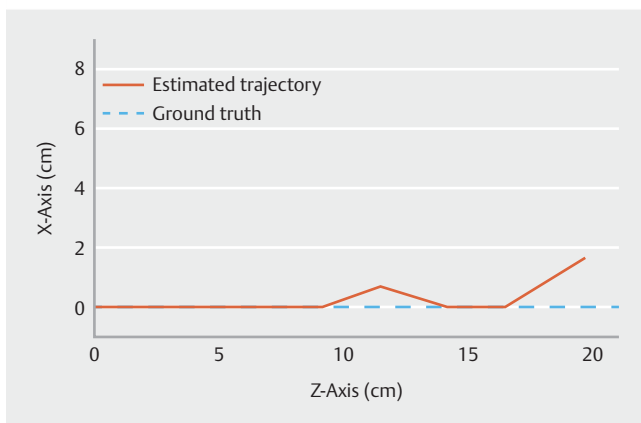
The 2D images obtained from the capsule were then processed to achieve 3D reconstruction of the bowel model. A modified Shape-from-Shading (SfS) technique was used to reconstruct a 3D surface from 2D images. SfS refers to a computer vision technique that recovers 3D shape and depth information from 2D digital images by investigating the variation of illumination across the image. The major assumption that this technique is based on is that the amount of reflected light is dependent on the orientation (shape) of the scene that is imaged. The majority of SfS approaches assume a light source either coinciding with the optical center or infinitely far away from the scene. However, these conditions are unrealistic for endoscopic recordings. Despite the small distance between camera and light source, the observed tissue is also very close to the camera and images are therefore affected by small illumination changes. To overcome this limitation, the method used approximates the position of the light source at the tip of the endoscope and uses the position directly in the algorithm. Given the small size and the density of the circular LED array of the capsule, its overall illumination can be considered equivalent to that of such a single light source following an approximately uniform illumination aggregation model [8]. Traditional SfS can recover depth up to an unknown scale factor, using the albedo of the imaged surface [9]. Albedo is a physical measure of reflectance or brightness of a surface. For a given surface, albedo is defined as the ratio of the reflected irradiance to the incident irradiance and it is dimensionless. Irradiance is a physical measure defined as the radiant flux (power) received by a surface per unit area. Furthermore, in our technique, because we consider the camera and light source as separate entities, we can model the SfS problem such that the unknown albedo is parameterized and calculated, thus providing a more accurate metric estimation of depth [10].

► **Table 1** Best results for travel distance estimation obtained using Kannala and Brandt's method.

Row of pins	Travel distances (cm)		
	Actual	Estimated	Absolute error
1	19.8	20.7	0.9
2	17.4	14.8	2.6
3	19.9	20.7	0.8
4	19.6	20.9	1.3



► **Fig. 3** Best results in travel distance estimation after calibration per row of thumbtacks. The error between the actual and the estimated travel distance is presented on top of the respective bars.



► **Fig. 4** Graph showing estimated vs actual CE trajectory.

The ORB-SLAM (Oriented FAST and Rotated BRIEF – Simultaneous Localization and Mapping) is used to estimate the pose (location and orientation) of a camera by finding matching points in image sequences as in videos [11]. From these matching points and the known calibration parameters of the camera, an estimation of the camera's pose as well as a sparse 3D reconstruction (mapping) of the environment can be extracted. Using a sequence of images from the CE video, the entire trajectory ("tracks") of the CE can be estimated. In ORB-SLAM the matching points in consecutive images are extracted using a

specific type of customized image features called ORB. ORB features include Features from Accelerated Segment Test (FAST), used for detection of points of interest within the image [12] and Binary Robust Independent Elementary Features (BRIEF) [13], used for the representation of image content at the points of interest. These features offer the advantage of fast calculation, facilitating the real-time operation of SLAM, as well as being invariant to viewpoint rotation and scale changes.

Results

Calibration and travel distance estimation

Seventeen video frames of the checkerboard (► Fig. 2) were used for calibration. As the capsule was navigated through the model bowel, the number of video frames per movement ranged from 42 to 66, due to the variable frame rate of the capsule. Overall, the MAE in the estimated distance travelled by the capsule was 4.1 ± 3.9 cm, for a camera focal length of 1.16 mm. Minimum error achieved was 1.4 ± 0.8 cm, and the respective results per row of thumbtacks are illustrated in ► Table 1 and ► Fig. 3.

The 2D reconstruction of the capsule's trajectory through the model bowel is shown in ► Fig. 4. The solid red line represents the estimated capsule movement, in comparison to the actual path shown by the straight broken line.

3D reconstruction

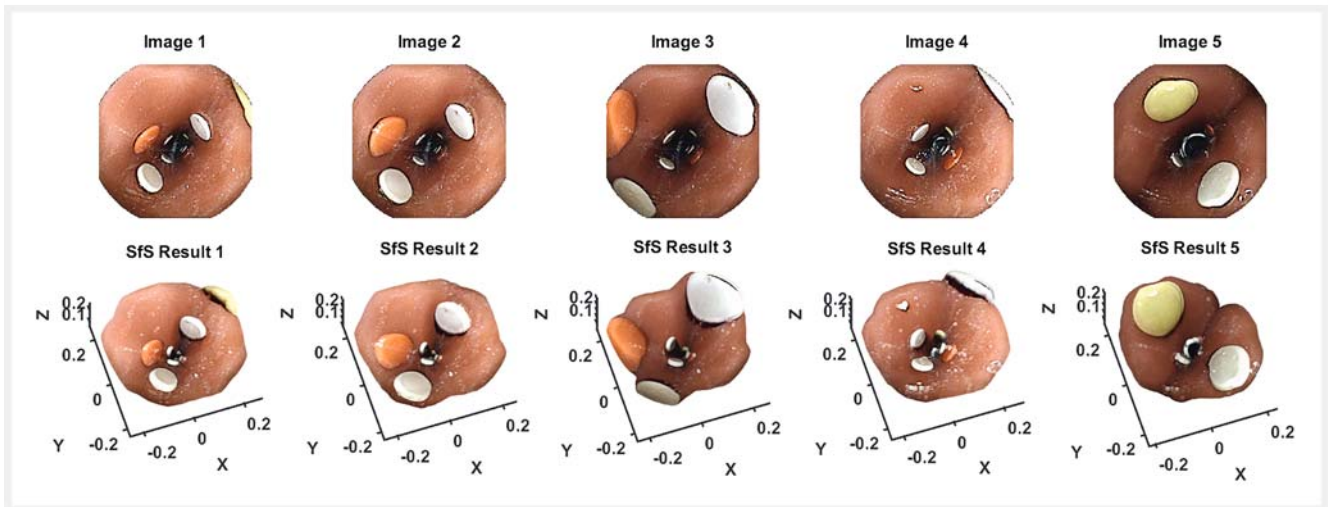
Both 3D reconstruction methods detailed above were able to achieve a good, but not optimal, reconstruction of the bowel model using information from the CE video alone.

Using the modified SfS technique, the cylindrical shape of the model bowel, with details of the tissue and attached thumbtacks, was successfully reconstructed. Examples of reconstructed bowel lumen, with corresponding original images, are shown in ► Fig. 5.

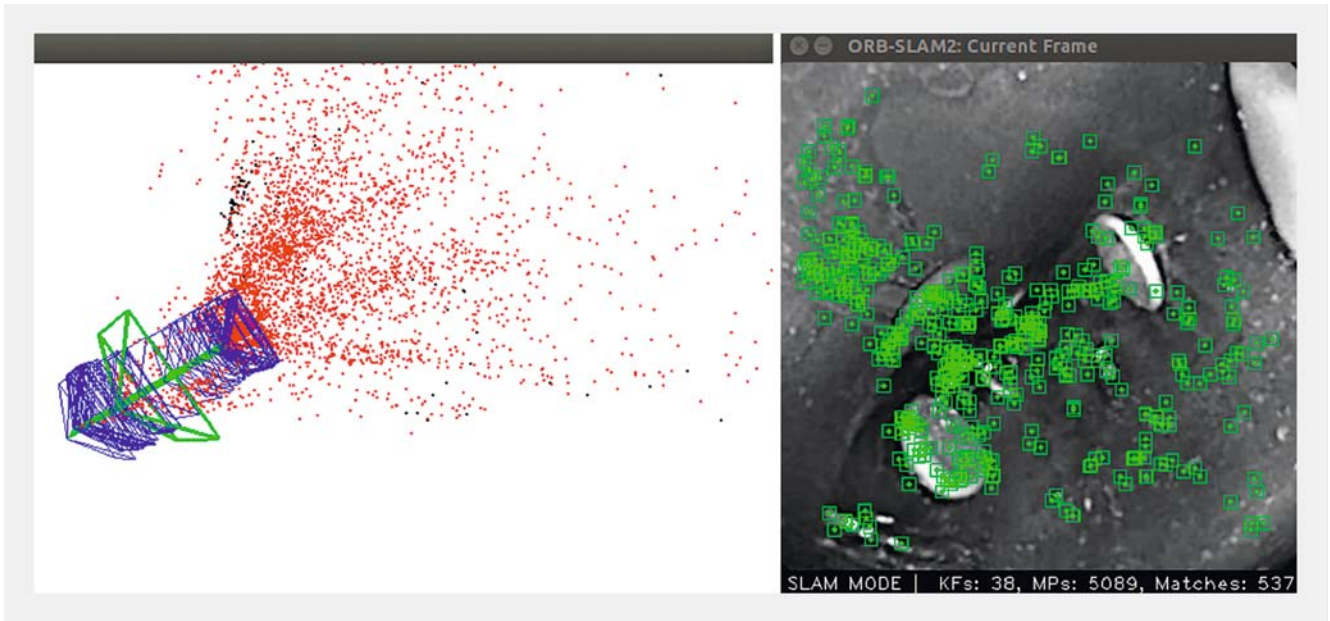
The ORB-SLAM method of 3D reconstruction produced good localization of the capsule within the reconstructed model. Results using this method are shown in ► Fig. 6. The blue triangles, corresponding to the outline of the reconstructed bowel wall from each frame of the video, are positioned in a straight line, with the overall "track" denoted by the green line passing through the triangles. This corresponds with the linear forwards-backwards movement of the capsule in the straight bowel model used.

Discussion

CE technology has progressed significantly since its introduction to routine clinical practice; however, the interpretation of a CE examination in order to reach a diagnosis remains heavily reliant on human readers [14]. Furthermore, the long reading times required also diminish its clinical efficiency. Therefore, further technological developments should aim to reduce CE reading times and minimize variability in CE reading. An ideal way to do so is to develop methods for computer-assisted and eventually automated diagnosis.



► **Fig. 5** Reconstruction results using the modified SfS technique. Selected frames from the CE video are shown above, with the corresponding reconstructions below.



► **Fig. 6** Results obtained using the ORB-SLAM algorithm. The location and post of the CE camera is estimated for each frame (current track in green rectangle; previous tracks in blue rectangles). The green line denotes the overall CE trajectory. The sparse 3D reconstruction is illustrated as a point cloud.

A significant limitation of CE is the lack of accurate localization. Current approaches to capsule and hence lesion localization include: transit time estimation from anatomical landmarks, localization in 2D or 3D space with respect to external sensors and radiofrequency triangulation, active magnetic localization, magnetic resonance, ultrasound and positron emission imaging-based approaches [4, 15, 16]. Our method provides comparable performance to methods based on external sensor arrays, without their use. Furthermore, because CE is a wireless minimally invasive system, information is mainly obtained as videos and images. 3D information could facilitate more detailed diagnostic evaluation of lesions seen [17]. Due

to the difficulty in accessing the human small-bowel, more invasive investigations or procedures such as deep enteroscopy should be optimally planned.

Typically, in CE, monocular vision provides the only information for 3D reconstruction. Therefore, our modified SfS method uses assumptions more applicable to CE images, obtained in the confined environment of the bowel lumen, and where manual focus is impossible due to the passive nature of capsule propulsion. To determine depth, this method estimates the albedo (whiteness coefficient, or measure of reflection) by using specular highlights and the corresponding surface normals of the reconstructed surface [10].

Our setup has inherent limitations due to currently available technology. First, the intrinsic parameters of the PillCam SB3 are unknown; therefore, vital information such as the focal length of the lens had to be estimated via calibration. Secondly, we assumed that the capsule moved at constant velocity following the centre of the bowel lumen. Finally, the model bowel was linear, immobile and had an elliptical cross-section throughout; furthermore, there was no luminal content. These do not entirely reflect actual human bowel structure and function, nor the usual clinical conditions under which CE operates.

Conclusion

In conclusion, based on our experimental set-up, we present methods for both 2D and 3D localization of a capsule using visual information alone. Such methods are feasible and have potential to be of clinical use. However, there remains a significant margin of error, indicating that much further work is required to refine these processes.

Acknowledgements

The work described in this paper was partially supported by the European Commission within the framework of the endoscopic versatile robotic guidance, diagnosis and therapy of magnetic-driven soft-tethered endoluminal robots Project-H2020-ICT-24-2015 (EU Project-G.A. number: 688592). The authors thank all the collaborators of the EU project.

Competing interests

None

References

- [1] Iakovidis DK, Koulaouzidis A. Software for enhanced video capsule endoscopy: challenges for essential progress. *Nat Rev Gastroenterol Hepatol* 2015; 12: 172–186
- [2] Grove PM. The Psychophysics of binocular vision. In: *3D-TV System with Depth-Image-Based Rendering*. New York, NY: Springer New York; 2013: 347–373
- [3] Van Rijn JC, Reitsma JB, Stoker J et al. Polyp miss rate determined by tandem colonoscopy: A systematic review. *Am J Gastroenterol* 2006; 101: 343–350
- [4] Baptista V, Marya N, Singh A et al. Continuing challenges in the diagnosis and management of obscure gastrointestinal bleeding. *World J Gastrointest Pathophysiol* 2014; 5: 523–533
- [5] Spyrou E, Iakovidis DK. Video-based Measurements for Wireless Capsule Endoscope Tracking. *Meas Sci Technol* 2014; 25: 14
- [6] Spyrou E, Iakovidis DK, Niafas S et al. Comparative assessment of feature extraction methods for visual odometry in wireless capsule endoscopy. *Comput Biol Med.* 2015; 65: 297–307
- [7] Kannala J, Brandt SS. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE Trans Pattern Anal Mach Intell* 2006; 28: 1335–1340
- [8] Cool S, Pieters JG, Mertens KC et al. Development of a high irradiance LED configuration for small field of view motion estimation of fertilizer particles. *Sensors* 2015; 15: 28627–28645
- [9] Ciuti G, Visentini-Scarzanella M, Dore A et al. Intra-operative monocular 3D reconstruction for image-guided navigation in active locomotion capsule endoscopy. In *Biomedical Robotics And Biomechanics (Biorob)*, 2012 4th IEEE Ras & Embs International Conference. 2012: 768–774
- [10] Visentini-Scarzanella M, Stoyanov D, Yang G-Z. Metric depth recovery from monocular images using Shape-from-Shading and specularities. In: *2012 19th IEEE International Conference on Image Processing*. 2012: 25–28
- [11] Mur-Artal R, Montiel JM, Tardos JD. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics* 2015; 31: 1147–1163
- [12] Rosten E, Drummond T. Machine learning for high-speed corner detection. *Computer Vision–ECCV 2006* 2006: 430–443
- [13] Calonder M, Lepetit V, Strecha C et al. Brief: Binary robust independent elementary features. *Computer Vision–ECCV 2010* 2010: 778–792
- [14] Lo SK, Mehdizadeh S, Ross A et al. How Should We Do Capsule Reading? *Tech Gastrointest Endosc* 2006; 8: 146–148
- [15] Keuchel M, Kurniawan N, Baltes P et al. Quantitative measurements in capsule endoscopy. *Comput Biol Med* 2015; 65: 333–347
- [16] Than TD, Alici G, Zhou H et al. A review of localization systems for robotic endoscopic capsules. *IEEE Trans Biomed Eng* 2012; 59: 2387–2399
- [17] Sakata S, Grove PM, Stevenson ARL et al. The impact of three-dimensional imaging on polyp detection during colonoscopy: a proof of concept study. *Gut* 2016; 65: 730–731