2008 IEEE/RSJ International Conference on Intelligent Robots and Systems
Acropolis Convention Center
Nice, France, Sept, 22-26, 2008

Lumen detection for capsule endoscopy

Xenophon Zabulis, Antonis A. Argyros and Dimitris P. Tsakiris

Abstract— **In this paper, two visual cues are proposed, to be exploited for the navigation of active endoscopic capsules within the gastrointestinal (GI) tract. These cues consist of the detection and tracking of the lumen and of an illumination highlight in capsule endoscopy (CE) images. The proposed approach aims at developing vision algorithms which are robust with respect to the challenging imaging conditions encountered in the GI tract and the great variability of the acquired images. Cases where no or more than one lumens exists, are also detected. The proposed approach extends the state-of-the-art in lumen detection, and is demonstrated for in-vivo video sequences acquired from endoscopic capsules.**

I. INTRODUCTION

Capsule endoscopy (CE) [1], [2] is a diagnostic procedure, in which a pill-sized capsule acquires images of the gastrointestinal (GI) tract, utilizing a microcamera mounted on one of its tips (see Fig. 1). The duration of this diagnostic procedure may range from 6 to 18 hours, because the capsule moves passively due to the peristalsis of the GI tract. More recently, research is focused on methods of capsule endoscopy where the capsule moves actively through the GI tract. This reduces the duration of the whole procedure, while allowing the explicit control of the camera's line-of-sight, a feature that is very valuable for diagnostic purposes [3], [4]. Example methods employed for the propulsion of the capsule include magnetically-driven motion, legged locomotion and others. Some of these principles are currently under scientific investigation, while some others have already been patented. Such methods include magnetically-driven motion of the capsule [5], [6], legged capsule locomotion [4], [7], [8], vibratory actuation [9] and others [10].

Even in the case of active capsule motion, the examination is not, yet, foreseen to last less than several hours. For this reason, methods to automate at least parts of the capsule navigation are being investigated. The objective is to relieve the capsule operator from the tedious task of continuously steering the capsule and allow him to focus on the diagnostic part of the process. Toward this goal, a fundamental reactive behaviour to be implemented by the navigation system of the capsule would be to strive to follow the lumen of the GI tract, and avoid collisions with its tissue. In this paper, the term *lumen* refers to the image region that depicts the farthest imaged piece of tissue, relative to the camera, when the capsule is - even coarsely - aligned with the GI tract. In cases of complete misalignment, only the tissue very close to the capsule is imaged; in such cases the lumen is absent.

This work was supported in part by the European Commission, through the IST project VECTOR (FP6-033970). The authors are with the Institute of Computer Science, Foundation for Research & Technology - Hellas, N. Plastira 100, Vassilika Vouton, GR-700 13, Heraklion, Crete, Greece, {zabulis,argyros,tsakiris}@ics.forth.gr The second author is also with the Computer Science Department, University of Crete, argyros@csd.uoc.gr.

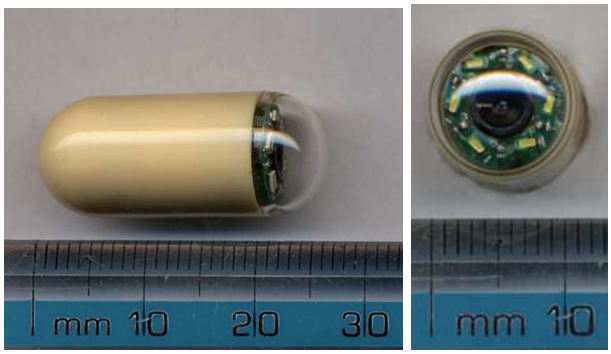

Fig. 1. A prototype endoscopic capsule from two views (left is side view and right is top view). A microcamera surrounded by LEDs is mounted at the tip of the capsule, inside the plastic dome.

CE is facilitated by LEDs mounted on the capsule and near the camera. The illumination is not ambient and, thus, tissues closer to the light sources are imaged with greater intensity than the rest. Thus, the lumen is expected to appear darker than the rest of the tissue. When the capsule is misaligned with the GI tract, the lumen is partially seen (if at all) and the visual field is dominated by tissue that is close, or in direct contact with the capsule. In this case, this tissue appears relatively brighter. The corresponding, bright image region is referred to as *highlight*.

The detection of the lumen and highlight can provide information that can be used to reorient the camera and align it with the GI tract. When the lumen is imaged, the capsule navigation system would rotate the capsule, so that the lumen is centered in the image. In cases where the lumen is invisible or partially imaged, the navigation system would rotate the capsule, so that the highlight exits the image, and the entire lumen appears centered in it. This work focuses on the robust detection of the lumen based on data from the capsule's camera. The design of the control system that actively orients the capsule depends on the specific propulsion method employed for the active capsule, and is beyond the scope of the present work.

An important component of the capsule navigation process is the detection of situations where the above process may be prone to error, in which case the operator of the capsule is alerted to resume control. Thus, cases where the lumen is absent, as well as cases where two lumens appear in the acquired images, are identified. To our knowledge, there are two conditions under which two lumens may occur in the

acquired image. First, when this is an anatomical fact. While there are no "junctions" in the GI tract anatomy, concavities give rise to a second lumen that is indistinguishable from the correct one (a well-known example is the appendix). Second, there are cases where the capsule faces the crease of a tissue fold or the inner part of a colon "turn".

This work also focuses on the particularly great variability of lumen appearance in CE. In contrast to conventional endoscopy, in CE the colon is not distended by gas and, thus, may appear concave in the image. Also, in conventional endoscopy the probe is actively aligned and not in contact with the GI tract, to reduce the risk of tissue damage. Due to these conditions, the lumen typically appears as the largest and darkest blob in the image and exhibits a convex contour. Additionally, wiring through the (conventional) endoscopical probe facilitates high-resolution imaging and high frame date, whereas in CE the wireless transmission of images and power-consumption constraints impose limited frame rate and image resolution. As an example, the widely known *PillCam* endoscopic capsule developed by *Given Imaging* [11], acquires 256×256 pixel images, at a frame rate of 2 Hz.

The remainder of this paper is organized as follows. In Sec. II related work is reviewed. In Sec. III, the proposed approach is described and in Sec. IV experiments with the proposed approach are presented. Finally, conclusions and directions for future research are provided in Sec. V.

II. RELATED WORK

Most of the relevant work in lumen detection regards conventional endoscopy, the detection of the lumen aims at the precise control of the endoscopic probe, so that it does not collide with the tissue. Thus, most of the reviewed works assume that the lumen is present in the image [12]. The case of detecting whether the lumen is really imaged has not been thoroughly investigated and, thus, only couple of marginally-relevant works are reviewed, at the end of this section. Also, to our knowledge, the use of highlights as a cue to CE navigation, has not appeared in the literature.

A starting point for several approaches to lumen detection is to focus interest on a coarse and dark image region, assuming that it corresponds to the lumen, and then refine the shape of this region. In this context, the consistent appearance of the lumen as the largest dark blob in endoscopic images [13], gave rise to image-thresholding approaches. To avoid the empirical selection of the threshold value, methods for its automatic adjustment were proposed in [14] and [15]. These approaches exhibit weak performance at cases of low contrast, where the difference between the lumen and the rest of the image pixels is not striking.

The thresholded image is combined in [16] with gradient information capitalizing on the characteristic pattern of intensity, which gradually increases outwards the lumen boundary. To detect this pattern the *iris* filter [17] is utilized, in order to evaluate the degree of convergence of the gradient vectors towards some image region. The size of the filtering kernel is set to approximate the image size of the candidate. The result of threshold and filtering is prone to local maxima of intensity, due to mucosal reflections, shadows, or noise. To better approximate the lumen boundary, a region-growing process was employed in [18], [14], [15]. In [19], color and texture were combined in the lumen's region-extraction process utilizing an image segmentation technique [20]. The method in [21], detects the lumen in the frequencies domain, based on template-matching in the Fourier transform of the image. The prototype is obtained through training, with images in which the lumen prominently appears. In [22], fuzzy logic was utilized to extract the lumen.

The methods in [23], [24], [18], [14], [15] tackle local maxima with a coarse-to-fine approach that is based on image pyramids and quad-trees. The coarse-to-fine treatment of the data facilitates the acceleration of the process. Also, for time-efficiency, the methods of [16] and [15] were hardware-optimized in [25] and [26], respectively.

To our knowledge and as stated in [19], *"... deciding whether an image contains the distant colon lumen or not, has not been investigated in the literature"*. The work in [19] implicitly detects the occurence of the lumen in the image, by estimating the ratio of lumen and tissue pixels (utilized as an indicator of image informativeness). The work in [27] detects lumen contraction, i.e. the abrupt change of lumen diameter, by detecting a temporal peak of the mean image intensity. However, this method cannot detect a non-contracting lumen.

III. PROPOSED METHOD

Due to the interplay of the light shed by LEDs and the 3D shape of the colon, it is reasonable to assume that both the lumen and the highlight correspond to size-dominant local extrema of image intensity. Specifically, the lumen corresponds to the most distant part of the tissue and therefore appears as the darkest image area. Similarly, the highlight is expected to be the brightest image area.

The input to the proposed method is a circular image of radius R. Unless stated otherwise, the algorithm operates on a monochromatic image with intensities in the range $[0, 1]$. When color is considered, the image is initially transformed to the HSV color space and the *hue* component is used. The output consists of:

- One or two candidate highlight regions.
- Zero (minimum), one (typically), or two (maximum) lumen regions.

The representation of each region includes its pixels, its contour, and a *representative* point (defined below).

The proposed method detects the lumen and the highlight by a coarse-to-fine version of the Mean Shift algorithm (MS) (Sec. III-A). The MS algorithm runs several times with different seed points and the detected extrema are spatially clustered. For each cluster, a representative is kept which is then grown to a size-dominant bright or dark image region, based on a region-growing technique (Sec. III-B). The resulting regions are then further evaluated as to whether they are likely to correspond to the pursued lumen(s) and highlight (Sec. III-C). Finally, the detected regions are tracked over time (Sec.III-E). In addition, the case that the

lumen is not present in the image is investigated (Sec.III-D). An overview of this method is illustrated in Fig. 2.

A. Localization of size-dominant local extrema of intensity

The localization of size-dominant local extrema of intensity is performed with a variant of the MS algorithm [28]. This algorithm is a mode detector for data samples drawn from a probability density function (PDF). MS iteratively shifts a weighting kernel $\mathcal{K}$, by centering it to its centroid, which is computed by weighting the locations of the kernel's points with their underlying PDF values. In this work, the samples are the image pixels, the PDF is the image intensity function, and $\mathcal{K}$ is uniform and circular; at a given position, its center is denoted as $\vec{c_o}$ and its centroid as $\vec{c_c}$. The method iterates between (a) computing $\vec{c_c}$ and (b) shifting $\mathcal{K}$ to be centered at $\vec{c_c}$. The algorithm terminates either when $\mathcal{K}$ convergences to a stable location (i.e. when $|\vec{c_o}-\vec{c_c}|$ becomes less than some threshold τ) or when a maximum number of iterations is reached. The initial value of $\vec{c_o}$ is an input to the algorithm and called a *seed* point. The size of $\mathcal{K}$ determines the scale of the detected mode. A small kernel can be trapped in extrema of small spatial extent. In contrast, a large kernel will converge to coarse-scale modes, but with less accuracy caused by the intense blurring due to the large $\mathcal{K}$.

For the detection of size-dominant extrema, an iterative, coarse-to-fine version of MS is proposed. The proposed algorithm starts with a kernel $\mathcal{K}$ of radius ρ_0, which is progressively reduced. In each subsequent iteration $i \geq 1$, the basic MS step is performed and the resulting location forms the seed of the next iteration. The kernel radius and shift threshold are reduced as $\rho_i = \rho_0\beta^i$ and $\tau_i = \tau_0\beta^i$, respectively, where $\beta \in (0,1)$. A large kernel ensures robustness against extrema of minor size, while reducing its size increases the extrema localization accuracy. The result of this process is a point referred to as a Mean Shift Result (MSR) (see Fig. 2b). Being initialization-dependent, the MS algorithm does not provide a global result. Thus, it is multiply executed for seed points that are log-polarly distributed around the image center $\vec{o} = (0,0)$ (see Fig. 2c).

When a kernel is partially outside the image, inaccurate results are obtained, because unknown pixel values have to be considered. To ensure robust behavior, $\mathcal{K}$ is prevented from crossing these boundaries. To retain K's convergence towards the extremum, its motion is adjusted to follow the boundary until convergence. The adjustment accounts for the circular shape of the CE image and proceeds as follows: Let $\mathcal{K}$ be centered at $\vec{c_o}$, its centroid $\vec{c_c}$, and $\vec{v} = \vec{c_c} - \vec{c_o}$. If centering $\mathcal{K}$ at $\vec{c_c}$ would cause $\mathcal{K}$ to cross the image boundary ($|\vec{c_o} - \vec{c_c}| > R - \rho_i$), then $\vec{v}$ is decomposed into two orthogonal vectors. These are $\vec{v_r}$ and $\vec{v_q}$, with $\vec{v_r}$ being parallel to $\vec{t} = \vec{c} - \vec{o}$. The shift of $\mathcal{K}$ is then performed in two steps. In the first, $\mathcal{K}$ is shifted in the direction of $\vec{v_r}$ for a distance $R - |\vec{c_o} - \vec{o}| - \rho_i$; this guarantees that $\mathcal{K}$ is exactly tangent to the image boundary. In the second, $\mathcal{K}$ is shifted in the direction and distance defined by $\vec{v_q}$. As a result, $\mathcal{K}$ "slides" tangentially to the image boundary.

B. Region extraction

1) Clustering MSRs: The multiple MSRs of the previous step occur closely clustered at the spatial vicinities of size-dominant intensity extrema (see Fig. 2d). In contrast, when such extrema do not exist in the image, the MSRs are scattered across the image with great variance (see Fig. 3). Based on this observation, a criterion is evaluated to detect if the lumen is really present in the image or not. Details are presented in Sec. III-D. If the lumen is not present in the image, the remainder of this step deals only with highlight detection.

The input MSRs are clustered based on spatial proximity, through a generalized connected components labeling (CCL) algorithm. Two MSRs are grouped together if their Euclidean distance is less than δ. The CCL algorithm computes the transitive closure of this relation and partitions the set of MSRs into a number of equivalence classes, each being a cluster of "connected" MSRs. For each such cluster, the MSR with the highest (for highlight) or lowest (for lumen) mean intensity at its neighborhood is selected as its *representative*.

2) Extracting coarse representations of image regions: The representatives extracted in the previous step are used as input seed points to a Recursive Region Extraction (RRE) process. This is a region-growing process, whose goal is to extend the cluster representatives into their corresponding image regions and which operates as follows. Initially, each region contains a seed point $\vec{p}$, which is the representative of a cluster of MSRs. Immediate neighbors of pixels already assigned to the region are added to it, if their intensity differs less than a threshold κ to the mean intensity in the neighborhood of $\vec{p}$. This process is an iterative one. Each iteration j results in an image region $\mathcal{R}_j$. Moreover, in each iteration, κ is modulated as $\kappa_j = \kappa_0 + j\kappa_0/\gamma$, where γ is an input constant. The process is terminated if (a) the region does not grow any more, or if (b) by growing the region by one more iteration $(j+1)$ any of the following would occur:

- The area of $\mathcal{R}_j$ increases by an order of magnitude.
- The compactness[1] of $\mathcal{R}_j$ decreases by an order of magnitude.

3) Refining image regions: The regions extracted in the previous step are refined utilizing an active contour (or "snake") [29] (see Fig. 2e). The contour energy minimized by the snake contains terms depending on color, gradient and contour smoothness.

C. Region evaluation

In this step, the extracted bright and dark regions are evaluated, based on their likelihood to correspond to highlights and lumens, respectively. For bright regions, only the best scoring region is selected. Dark regions are sorted with respect to a measure of this likelihood, since there might be more than one lumen appearing in the image. For a dark region $\mathcal{D}$ the metric is:

$$S = I^2 \cdot C \cdot A, \tag{1}$$

[1]The compactness of an image region is measured as the ratio of the region's area to that of a circle with the same perimeter.

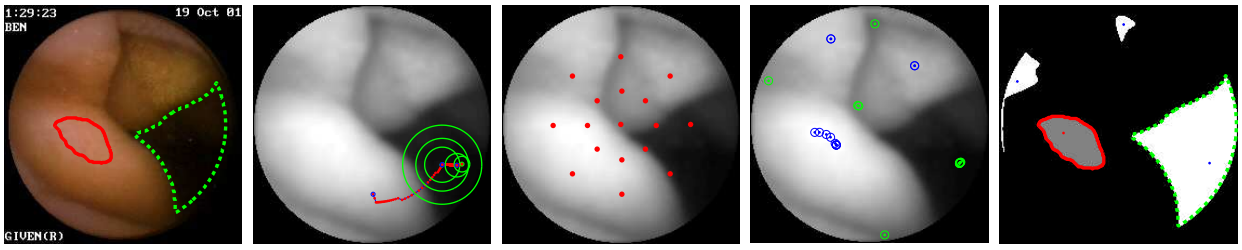

Fig. 2. Overview of the steps of the proposed method. Left to right: (a) Original image with superimposed result; solid line outlines highlight and dashed line the lumen. (b) Behavior of the proposed variant of the MS algorithm. Figure shows the trajectory of $\mathcal{K}$ and the progressive reduction of its radius. Notice that before converging to the intensity extremum, the trajectory of the kernel is parallel to the boundary of the image. (c) Initialization seeds for multiple runs of the MS algorithm. (d) The resulting MSRs for the dark (bright dots) and bright (dark dots) areas. (e) Extracted regions and their boundary approximation; region representatives indicated with dots.

where A is $\mathcal{D}$'s area, C is the ratio of $\mathcal{D}$'s area over its perimeter, $I = 1 + (1 - m_I)$ and m_I the mean intensity of its representative point. In the above equation, the I factor is set to be the most influential, because it represents the "darkness" at the neighborhood of the representative. For bright regions, the metric is simplified to the value of the mean intensity of the region. Due to the multiplicative nature of the above metric, runner up solutions exhibit a large metric difference to the winning one. A second lumen is considered only if the score of the second-winning region is in the same order of magnitude as the first.

D. Absence of lumen

Judging whether the lumen appears or not in an image depends on the evaluation of four criteria. Fig. 3 shows a typical case of an image in which the lumen is not present. The central part of the image is bright due to LED illumination and, consequently, the image periphery is relatively dark. As a result, the MSRs corresponding to the dark areas (dMSRs) are scattered in the periphery of the image. Thus, an image is considered to be lumen-less if all of the following hold:

- The majority ($> 90\%$) of locations of the dMSRs occur at the periphery of the image.
- The spatial variance of dMSRs locations is high.
- Clustering of dMSRs yields more than one cluster.
- The RRE region extracted for each of the above clusters exhibits low compactness.

E. Tracking and target selection

The detected regions are tracked across time by associating regions across frames. This association is based on region overlap and is relatively simple because, in a certain frame, the detected regions are usually few, sparsely dispersed and non-occluding each other.

The tracking process is utilized in target selection. There are cases at which multiple size-dominant dark blobs occur in an image, without any of them really corresponding to the lumen. Such cases are due to scene complexity (e.g. colon folds, turns, etc.) and shadows. Such artifacts occur only for a few frames in the temporal sampling of our experiments. The tracking history is, therefore, utilized to reject lumen region hypotheses whose temporal duration is short. In practice, a certain region should be considered as a valid lumen hypothesis only if it has been tracked for more than 5 frames.

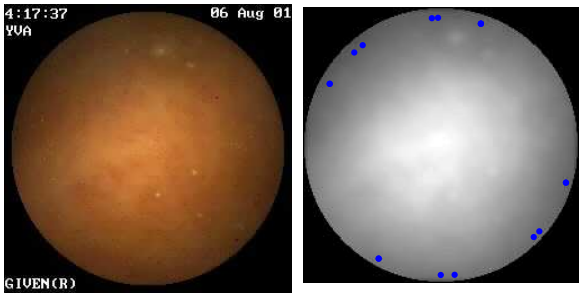

Fig. 3. Absence of lumen. When the lumen is not visible, the MSRs are scattered in the periphery of the image.

Finally, there are cases where multiple lumens do occur and which are tracked by the proposed approach. Additional visual cues or other type of evidence is required to purposefully select the best candidate in context-dependent and task-driven application scenarios.

IV. EXPERIMENTS

In this section, the effectiveness of the proposed approach is compared to existing approaches to lumen detection and evaluated in particular situations that can be encountered in CE. Images were acquired by a *Given Imaging* endoscopic capsule with nominal size 256×256 pixels and, effective $R = 124$ pixels. The frame rate was $2\,fps$. The values for the parameters of the proposed method are: $\rho_0 = 2R/3$, $\tau_0 = 1$ pixel, $\beta = 2/3$, $\kappa = 7$, $\gamma = 4$, and the number of MS iterations in step 1 of the method was 5. Fig. 2 and Fig. 3, were briefly discussed above to illustrate the steps of the proposed method. Fig. 2 shows a conventional case of lumen detection, which is treated adequately by several of the existing methods. Fig. 3 shows a typical case where the lumen is not imaged.

First, the proposed method is compared with other existing approaches. Region growing techniques (e.g. [14]) perform adequately in clearly defined lumen occurrences, but are

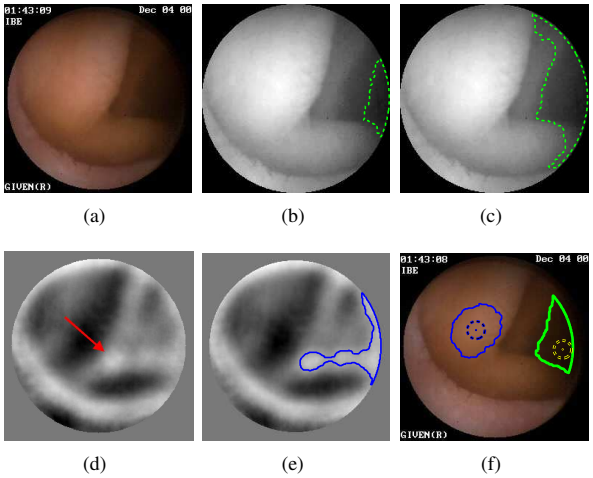

Fig. 4. Comparing approaches to lumen detection: (a) original image, (b) region growing with low tolerance, (c) region growing with high tolerance, (d) iris filter response, (e) combination of region growing and iris response, (f) proposed approach.

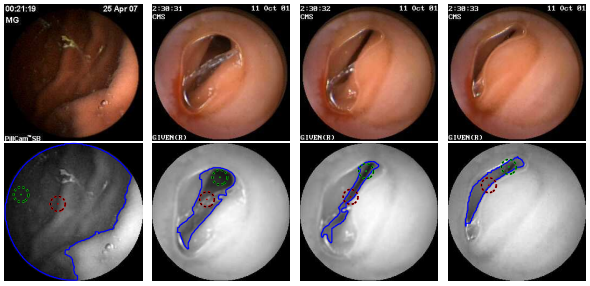

Fig. 5. Comparison of region representatives. Red targets plot the centroid (mass center) and green targets indicate the darkest lumen region.

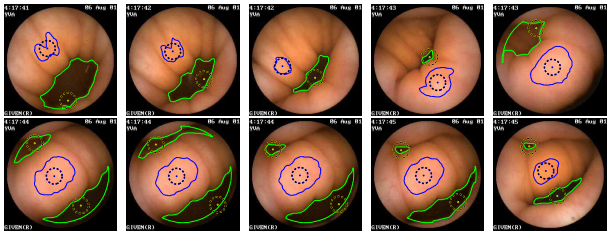

Fig. 6. Processing an image sequence where a second lumen appears. A single lumen appears in the initial frames of the above sequence (top row). In the bottom row, a second lumen gradually appears and is tracked. Dashed lines outline the detected lumens and a solid line outlines the highlight. Region representative points are marked with dashed circles.

hindered by the presence of shadows (see Fig. 4). Due to the shadow, there is no ideal threshold value that correctly segments the lumen without over- or under-estimating its area (see Fig. 4b and c). The convergence of gradient towards the inner of the candidate lumen area can be combined to refine this result, as in [14]. However, in the case of Fig. 4 the response of the iris filter (shown in Fig. 4d) does not disambiguate the boundary, because, due to the shadow, the gradient in the region is weak. In fact in this image, the dominant filter response (shown by the arrow) is located on the well-focused tissue fold, and on not the visible part of the lumen, which occurs near the image border. Combining the two responses leads to the less accurate contour of Fig. 4e. In the proposed approach, the consideration of contour compactness in the energy of the active contour restricts the overestimation of the lumen. In addition, tracking multiple candidate lumens, as proposed, facilitates, not only the better description of the imaged scene (cases of two lumens), but also the compensation against short failures of lumen detection (see also Fig. 7 below).

Fig. 5 corroborates that it is more accurate to select the region of lower intensity rather than the center of mass or the contour, when tracking the representative of the lumen. In the left column, the obliqueness by which the lumen is imaged sets the center of mass approximately in the center of the image, whereas the darkest region indicates more accurately the orientation of the capsule relative to the axis of the GI tract. The remaining columns are a sequence of frames, which were acquired shortly apart (1 sec each). They show how the collapse of tissue tangent to the capsule dome can partially occlude the GI tract in front of the capsule. As seen in the first frame, the "pathway" through the GI tract is on the top part of the lumen. In the remaining frames, most of the lumen is collapsed and the centroid of its image region does not point to the direction that the capsule is to be oriented.

In Fig. 6, an ambiguous case of lumen detection is shown. In the first frames of the sequence, a single lumen is detected and localized. In subsequent frames, the lumen size decreases, its image intensity increases, and a second lumen appears. Whichever lumen the capsule should follow, there are frames where the assumption that the lumen is the largest and darkest blob in the image does not hold. By selecting the extremum as the representative, the potential target direction of the capsule is guaranteed to point within the lumen. In the video accompanying this paper [30], a number of experiments in indicative CE image sequences are shown.

Finally, Fig. 7 shows a case where a spurious second lumen briefly occurs in the image. The spurious lumen is removed from consideration, as it occurs for only a few frames.

Processing a video frame requires about $3\,sec$ on a conventional Pentium PC that runs an unoptimized MATLAB implementation of the proposed method.

Fig. 7. Rejecting candidate image regions that occur for short time periods. The candidate region outlined with a dashed line is rejected, because it occurs only for a few frames. Lumen and highlight outlined with green and blue solid lines, respectively; representatives are marked with circles.

V. CONCLUSIONS

In this paper, two visual cues to endoscopic capsule navigation are proposed. These cues are based on the detection and tracking of the lumen and of an illumination highlight in CE images. The lumen is considered as a target direction for navigation, whereas the intense presence of a highlight signifies the misalignment of the capsule relative to the colon. For example, in the top row of Fig. 6, the highlight is (approximately) always in the opposite semi-circle on which lumen resides. In the second row, where two lumens appear, the highlight occurs between them. In both cases, rotating the camera so that the highlight exits the visual field, the lumen would become more centered in this field. The proposed approach identifies cases where the lumen is not imaged, to prevent from feeding the capsule navigation process with spurious information. It also identifies the case where two lumens are imaged.

This work extends the state-of-the-art in lumen detection by coping with the large image variability and difficult imaging conditions encountered in CE, which are significantly different than those of conventional endoscopy. However, the results might also be useful to conventional endoscopy, as the microrobotic control of the corresponding probe is also a current research topic, aiming at the simplification of probe control and the prevention of tissue damage. In this context, the location of the intensity minimum within the contour of the lumen is a more useful cue than its centroid.

Two main follow up studies can be identified. The first is to validate (in simulation and in-vitro) the adequacy of the proposed visual cues in supporting capsule navigation. The second is to semantically analyze the combined motion patterns of lumen(s) and highlight in order to infer information about the orientation of the capsule relative to the GI tract. Another direction for future work is the integration of the proposed method with existing methods which detect CE video frames that are not useful for navigation, due to severe clutter or illumination artifacts. Such cases occur when the visual field is dominated by bubbles formed by intestinal juices [31], or when the information value of the acquired image is low [19], due to low contrast or lack of imaged tissue.

REFERENCES

[1] F. Nebeker. Golden accomplishments in biomedical engineering. *IEEE Engineering in Medicine and Biology Mag.*, 21:17–47, 2001.

[2] D. Panescu. An imaging pill for gastrointestinal endoscopy. *IEEE Engineering in Medicine and Biology Mag.*, 24(4):12–14, 2005.

[3] I. Kassim, L. Phee, W. S. Ng, F. Gong, P. Dario, and C. A. Mosse. Locomotion techniques for robotic colonoscopy. *IEEE Engineering in Medicine and Biology Mag.*, 25(3):49–56, 2006.

[4] A. Menciassi, M. Quirini, and P. Dario. Microrobotics for future gastrointestinal endoscopy. *Minimally Invasive Therapy*, 16(2):91–100, 2007.

[5] B. Kim, S. Lee, J. Park, and J. Park. Design and fabrication of a locomotive mechanism for capsule-type endoscopes using shape memory alloys (SMAs). *IEEE/ASME Trans. on Mechatronics*, 10(1):77– 86, 2005.

[6] A. Uehara and K. Hoshina. Capsule endoscope NORIKA system. *Minimally Invasive Therapy and Allied Technologies*, 12(5):227–234, 2003.

[7] C. Stefanini, A. Menciassi, and P. Dario. Modeling and experiments on a legged microrobot locomoting in a tubular, compliant and slippery environment. *Int. J. of Robotics Research*, 25(5-6):551–560, 2006.

[8] M. Karagozler, E. Cheung, J. Kwon, and M. Sitti. Miniature endoscopic capsule robot using biomimetic micro-patterned adhesives. In *IEEE Biomedical Robotics and Biomechatronics*, pages 105–111, 2006.

[9] X. Zabulis, M. Sfakiotakis, and D. P. Tsakiris. Effects of vibratory actuation on endoscopic capsule vision. In *30th Intl. Conf. of IEEE Engineering in Medicine and Biology Society (EMBC08)*, Vancouver, Canada, 20-24 August, 2008. (to appear).

[10] B. Kim, S. Lee, J. Park, and J. Park. Inchworm-like microrobot for capsule endoscope. In *IEEE/RAS-EMBS Robotics and Biomimetics*, pages 458–463, 2004.

[11] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain. Wireless capsule endoscopy. *Nature*, 405:417, 2000.

[12] S. J. Phee, W. S. Ng, I. M. Chen, F. Seow-Choen, and B. L. Davies. Automation of colonoscopy - part ii: visual-control aspects. *IEEE Engineering in Medicine and Biology*, 17(3):81–88, 1998.

[13] L. Sucar and D. Gillies. Knowledge-based assistant for colonoscopy. In *IEA/AIE*, volume 2, pages 665–672, 1990.

[14] K. Asari. A fast and accurate segmentation technique for the extraction of gastrointestinal lumen from endoscopic images. *Medical Engineering and Physics*, 22:89–96, 2000.

[15] S. Kumar, K. Asari, and D. Radhakrishnan. A new technique for the segmentation of lumen from endoscopicimages by differential region growing. In *Midwest Symposium on Circuits and Systems*, volume 1, pages 414–417, Las Cruces, NM, USA, 1999.

[16] H. Tian, T. Srikanthan, and V. K. Asari. A recursive Otsu-Iris filter technique for high-speed detection of lumen region from endoscopic images. In *Applied Imagery Pattern Recog.*, pages 182–186, 2001.

[17] H. Kobatake and S. Hashimoto. Convergence index filter for vector fields. *IEEE Tran. on Image Processing*, 8:1029–1038, 1999.

[18] S. Kumar, K. Asari, and D. Radhakrishnan. Online extraction of lumen region and boundary from endoscopicimages using a quad structure. In *Image Proc. and its Applications*, volume 2, pages 818–822, 1999.

[19] S. Hwang, J. Oh, J. Lee, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. de Groen. Automatic measurement of quality metrics for colonoscopy videos. In *ACM Conf. on Multimedia*, pages 912–921, 2005.

[20] Y. Deng and B. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(8):800–810, 2001.

[21] C. Kwoh and D. Gillies. Using Fourier information for the detection of the lumen in endoscope images. In *IEEE TENCON*, pages 981–985, Piscataway, NJ, USA, 1995.

[22] G. Tan, S. Knshnan, and K. Chan. Endoscopic image processing with fuzzy reasoning for identification of intestinal lumen contour and abnormalities. In *Medical Physics and Biomedical Engineering*, pages 61–66, 1994.

[23] G. Khan and D. Gillies. Detecting insertion direction for an automatic endoscope. In *MEDINFO*, pages 1455–1459, 1989.

[24] G. Khan and D. Gillies. *Computer Vision and Image Processing*, chapter Parallel-hierarchical image partitioning and region extraction, pages 123–140. Academic Press, San Diego, 1992.

[25] H. Tian, T. Srikanthan, and V. Asari. Hardware efficient technique for rapid lumen segmentation from endoscopic images. In *JCIS*, pages 716–719, 2002.

[26] K. Asari, T. Srikanthan, S. Kumar, and D. Radhakrishnan. A pipelined architecture for image segmentation by adaptive progressive thresholding. *Microprocessors and Microsystems*, 23(8):493–499, 1999.

[27] P. Spyridonos, F. Vilarino, J. Vitria, and P. Radeva. Identification of intestinal motility events of capsule endoscopy video analysis. In *Advanced Concepts for Intelligent Vision Systems*, pages 20–23, Antwerp, Belgium, 2005.

[28] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[29] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. on Image Processing*, 7(3):359–369, March 1998.

[30] http://www.ics.forth.gr/∼zabulis/iros08.html.

[31] F. Vilarino, P. Spyridonos, O. Pujol, J. Vitria, and P. Radeva. Automatic detection of intestinal juices in wireless capsule video endoscopy. In *Int. Conf. on Pattern Recognition*, pages 719–721, 2006.