Categorization and Segmentation of Intestinal Content Frames for Wireless Capsule Endoscopy

Santi Seguí, Michal Drozdzal, Fernando Vilariño, Carolina Malagelada, Fernando Azpiroz, Petia Radeva, and Jordi Vitrià

***Abstract*—Wireless Capsule Endoscopy (WCE) is a device that allows the direct visualization of gastrointestinal tract with minimal discomfort for the patient, but at the price of a large amount of time for screening. In order to reduce this time, several works have proposed to automatically remove all the frames showing intestinal content. These methods label frames as {*intestinal content - clear*} without discriminating between types of content (with different physiological meaning) or the portion of image covered. In addition, since the presence of intestinal content has been identified as an indicator of intestinal motility, its accurate quantification can show a potential clinical relevance. In this paper, we present a method for the robust detection and segmentation of intestinal content in WCE images, together with its further discrimination between turbid liquid and bubbles. Our proposal is based on a twofold system. First, frames presenting intestinal content are detected by a SVM classifier using color and textural information. Secondly, intestinal content frames are segmented into {*turbid*, *bubbles* and *clear*} regions. We show a detailed validation using a large dataset. Our system outperforms previous methods and, for a first time, discriminates between turbid from bubbles media.**

***Index Terms*—Wireless Capsule Endoscopy, Machine Learning, Informative Frames, Intestinal Content, Image Segmentation.**

I. Introduction

Since the appearance of the capsule endoscopy technology in 2000 [1], and due to its numerous clinical advantages, Wireless Capsule Endoscopy (WCE) has rapidly become a wide-spread clinical routine and its use has been proposed for the categorization of diverse intestinal pathologies, such as Crohn's disease [2], tract bleeding [3] and polyp search [4]. The wireless capsule device consists of an ingestible pill which contains a camera and a full electronic set which allows the radio frequency emission, of a video movie. This video, showing the whole trip of the capsule along the gastrointestinal tract, is stored into an external hard disc which is carried by the patient. The clinical protocol consists of the *a posteriori* screening of the video by a specialist in search of those features associated to intestinal pathologies.

The wireless device, named PillCam SB2, measures $11mm \times 26mm$ and weights less than 4 grams, it has a camera with 156° field of view, a battery, a wireless system and 3 optical lens. The frame rate is 2 frames per second and its image resolution is 256×256 pixels [5].

The capsule has rapidly gained recognition within the gastroenterology community thanks to its two main advantages: 1) it offers the inner visualization of the entire gastrointestinal tract and 2) it obtains the images of the gastrointestinal tract in a minimally invasive manner reducing to the patient preparation and discomfort. In contrast, standard techniques of gastrointestinal tract examination like manometry or gastroendoscopy are more invasive and produce patient discomfort or even the need of patients hospitalization.

However, procedures based on the capsule present several limitations [6]. First, the time needed by the physician to analyze the entire video: the capsule emits images at a rate of two frames per second for over 8 hours, that can result with 57.600 images for a single study. Second, the device has no therapeutic capability: it means that if any lesion that needs treatment is discovered some additional investigation must be done with standard procedures as endoscopy, radiology or surgical techniques. Finally, there is a difficulty in discerning the exact location of the visualized lesion.

Nevertheless, capsule endoscopy has undertaken a relevant boost in recent years and technological advances have been proposed both in the hardware and software areas [7] making the WCE a widely spread clinical routine [8]. This growth has been generally caused by the interest of the community in developing computer aided systems, where researchers have focused their efforts on trying to tackle the inherent drawbacks associated to the video screening stage of capsule endoscopy videos: the long time needed for visualization, the potential subjectivity of the observer due to fatigue, and finally, the presence of intestinal contents which hinders the proper visualization of the intestinal walls or lumen.

In the recent literature, we can distinguish four general research lines regarding the aim of each proposed system, namely:

1) Reduction of the needed time visualization, or adaptive control systems for video display [9], [10], [11], [12];
2) Characterisation of intestinal abnormalities such as polyps [13], [14], blood [15], [16], tumours [17], ulcers [3] or other lesions [18], [19], [20];
3) Differentiation of the diverse organs of the intestinal tract like esophagus, stomach, duodenum, jejunum-ileum and cecum [21], [22], [23].
4) Study and characterization of specific

S. Seguí, M. Drozdzal, P. Radeva and J. Vitrià are with the Departament de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, 08007 Barcelona, Spain, and also with the Computer Vision Center, Edifici O Bellaterra, 08193 Barcelona, Spain (e-mail: ssegui@cvc.uab.es; michal.drozdzal@ub.edu; petia@cvc.uab.es; jordi.vitria@ub.edu)

F. Vilariño is with the Computer Vision Center, Edifici O Bellaterra, 08193 Barcelona, Spain, and also with the Computer Science Department, Universitat Autònoma de Barcelona, Barcelona, Spain (e-mail: fernando@cvc.uab.es)

C. Malagelada and F. Azpiroz are with Digestive System Research Unit, Hospital Vall d'Hebron, 08035 Barcelona, Spain

Manuscript received June 7, 2011; revised March 6, 2012 and August 3, 2012; accepted September 24, 2012.

events/dysfunctions of intestinal motility, such as intestinal contractions or motor activity [24], [25], [26].

Additionally, the detection of intestinal content frames has been identified as an important problem in this research field for three main reasons: a) it can reduce the number of false alarm ratio for several detection methods as for example polyps, ulcers or contractions [25]; b) it can also reduce the evaluation time required by physicians for video analysis when eliminating the frames completely covered by intestinal content [27]; and c) it has been identified as an indicator of motility dysfunction [28].

A. Physiological origin of intestinal content

Previous studies with endoluminal image analysis [27], [29] have shown that intestinal contents my exhibit two appearance paradigms, namely: *bubbles* or *turbid* material. Bubble formation depends on the presence of agents that reduce surface tension, analogous to a detergent. In normal conditions this activity is due to the presence of biliopancreatic secretions, responsible of the solubilization and subsequent digestion of fat. By contrast, turbid appearance reflects the presence of chyme, that is, the meal transformed by the processes of gastric and partial intestinal digestion. In this context, the type of content depends in normal conditions of the characteristics and the time elapsed since the last meal [30].

During fasting the small bowel exhibits a cyclic activity pattern, alternating phases of quiescence with phases of intense biliopancreatic secretion into the duodenum associated with forceful propagating contractions, that pushes the content in caudad direction, clearing residues from the gut. These phases of intense motor and secretory activity occur on average every $100min$ [31]. The association of high concentration of biliopancreatic secretion with wall contractions results on a foamy appearance of contents, visually recognized by the presence of abundant bubbles. Ingestion of a meal interrupts this fasting cyclic activity pattern and induces a more homogeneous secretory and motor activity in order to digest the meal. Regardless of the characteristics and amount of food ingested, the stomach delivers into the small intestine a homogeneous liquefied chime with particles of less than $1mm$, at a steady rate adjusted to the intestinal processing capability. In fact, the small bowel controls gastric emptying and biliopancreatic secretion by a complex net of feedback mechanisms. As a consequence, postprandial intestinal content consists of a mixture of homogenized nutrients and biliopancreatic secretion in a proportion related to the types of foodstuffs in the meal. Since surfactive agents are diluted into the mixture, the appearance of chime is turbid without bubbles [30].

The presence of these types of content patterns along the small bowel reflects the relative proportion nutrients and secretions, as well as the degree of digestion, which differ at various levels of the intestine as a function of the progress of digestion. Furthermore, abnormal digestive function may affect this process and modify the pattern distribution of intestinal contents.

Despite their physiological importance and differences, these two kinds of intestinal contents have always been quantified together. In this paper we propose a system for the categorization and segmentation of frames with different classes of intestinal content.

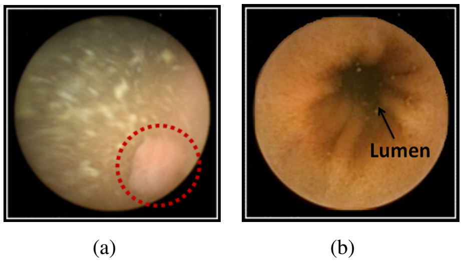


Fig. 1: Images from WCE: (a) Image partially covered by intestinal content presenting a polyp (dashed circle). (b) image where lumen (indicated by the arrow) is visualized surrounded by intestinal wall.

B. State-of-the-art method of intestinal content detection

In the recent literature, several methods have been proposed to detect intestinal content frames. In [29] Vilariño et al. presented a method for detecting bubble-like shape of intestinal juices based on Gabor filters. Another method was proposed by Vilariño et al. in [25] based on SVM classifier which uses 125-bin color histogram as a feature vector. The most recent work by Bashar et al. [27] presented a method for informative frame detection. In this method the highly contaminated by turbid fluid, fecal materials and/or residual fluids frames are named non-informative and all other frames are considered informative. The method was a two-stage cascade: in the first stage color information was used to characterize turbid, and in second stage, texture descriptor was applied to characterize bubbles images.

The main contributions introduced in this paper can be summarized as follows:

1) The development and validation of a fast automatic method for the detection of WCE frames with intestinal content. Obtained results, using a very large dataset, shows that the presented method, which uses color and textural frame information, outperforms the results obtained by other reported methods [27].
2) The development of a segmentation method for detecting bubbles and turbid media in WCE images. The proposed method is two fold: on the one hand turbid parts of the image are segmented using color information, and on the other hand, bubble regions are segmented using textural information. Finally, the output of the proposed method is an image segmentation using 3 different labels { *clear*, *turbid* and *bubbles*}.
3) The definition of a new characteristic, based on the image area covered by each kind of intestinal content, to characterize WCE videos. This information can be used as a new physiologically-based feature for automatic systems in diverse areas, such as intestinal motility, where it could add further support for motility disorder when turbid liquid secretions are present in the proximal small bowel.

4) Finally, we carry out the validation of our proposals in a large database with more than 95.000 frames, addressing in this way one of the main drawback present in previous literature: lack of statistical support of the results reported in [27], [29]. All these images were obtained from the set of 50 studies from different subjects including healthy volunteers and patients with motility disorders.

This paper is organized as follows: Section 2 describes the endoluminal scene; Section 3 presents the proposed system for detection and characterization of frames with intestinal content; Section 4 presents the experimental results; and finally, Section 5 presents a discussion and conclusions about the presented method.

II. Endoluminal Scene: frames with intestinal content

The different parts of the intestinal tract (stomach, duodenum, jejunum-ileum and cecum) presents a variety of appearances with multiple textures and colors. In addition to that, video frames can be described in three different component of the inner gut: intestinal wall, intestinal content and intestinal lumen (see Fig. 1(a) and 1(b)).

When a capsule is centered in the intestinal tube, a perspective of the lumen is obtained. However, both the free motion of the camera and the contractions that the gut undertakes produces a variety of orientations and perspectives of the scene (see Fig.2(a)). This provides a high variability in the resulting images: The contraction of the lumen is visualized with a *wrinkle pattern* that is usually centered in the middle of the images, but which can also present an offset or even lay out from the field of view of the camera. In this case, only the *intestinal wall* is shown and the lumen is lost for a sequence of frames. Additionally, lesions such ulcers, polyps, etc., are visualized in the endoluminal scene when present (see Fig. 2(b)).

Intestinal content is usually transparent and allows a nitid, clear view of the intestinal walls. However, some images are blurred by intestinal content. In this case, content is visualized as a turbid liquid secretion or as bubbles. Moreover, the intestinal content may hinder the proper visualization of the scene (see Fig. 2(c) and 2(d)). In a normal video, with standard clinical patient preparation, between 5% and 40% of video frames contain intestinal content. The degree of intestinal content in a single frame can vary from covering a small area of the image to completely occluding the intestinal wall and lumen. Intestinal content frames are presented in a high variability of colors and textures within a video. The colors and the textures are highly correlated with the ingested food by the patients. Generally, according to the visual appearance of these images, the *turbid* (*food in digestion* and *intestinal juices*) can be easily differentiated from *bubbles*.

- *Turbid* is usually presented in the frame as an homogeneous texture with a wide range of colors (see Fig.2(c)). The predominant colors presented in turbid frames varies from brown to yellow, however, sometimes can also be presented in less common colors like green or red.
- *Bubbles* are presented in the image as well-defined texture (see Fig.2(d)). This texture is characterized by several ellipsoidal blobs that can vary in the size. The most predominant colors of the bubbles are: white, yellow and green. However, sometimes bubbles are practically transparent and the only visible part of the bubble is the contour.

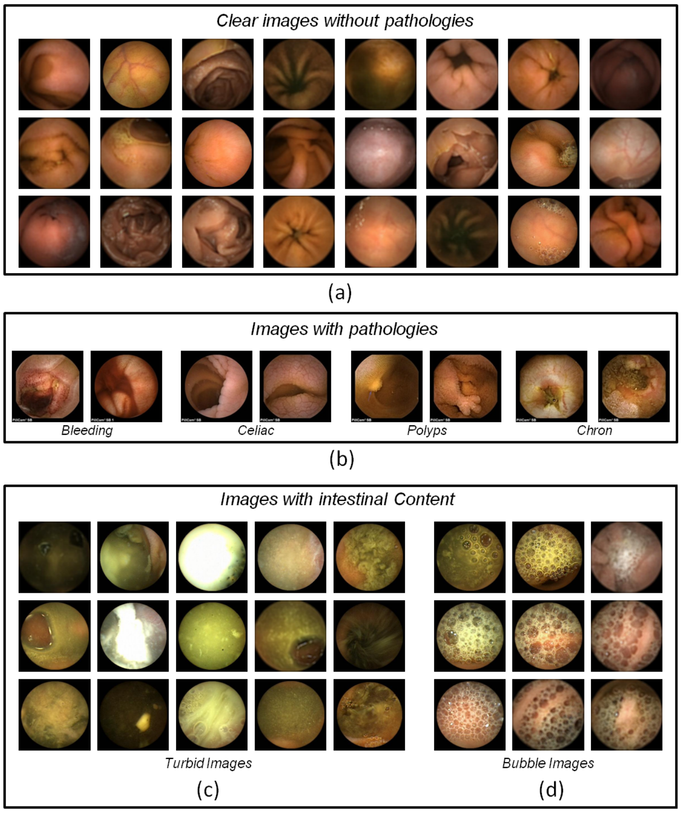


Fig. 2: Example of images from WCE videos. (a) Clear images without pathologies. (b) Images presenting some pathologies: bleeding, celiac, polyps and chron. (c) Images with turbid content. (d) Images with bubbles.

III. System for automatic detection and characterization of frames with intestinal content.

The proposed system is divided into two consecutive steps: 1) detection and 2) segmentation. The aim of the detection step is to target the segmentation of the intestinal content only in the frames where the intestinal content, reducing in this way the computation cost. The advantage of the second step is three-fold: 1) it provides information about the percentage of the image covered by intestinal content (this parameter can be used to automatically remove these sequences for visualization in order to reduce the screening time); 2) an accurate segmentation of intestinal content allows to maximize the area of intestine visualized (in Fig1(a) we can observe a frame which is mostly covered by intestinal content but showing a relevant pathology in the clean tissue) and 3) the accurate measurement of the amount of turbid and bubbles can be used as indicators associated to motility dysfunctions [28], [18]. The complete system scheme is presented in Fig. 3.

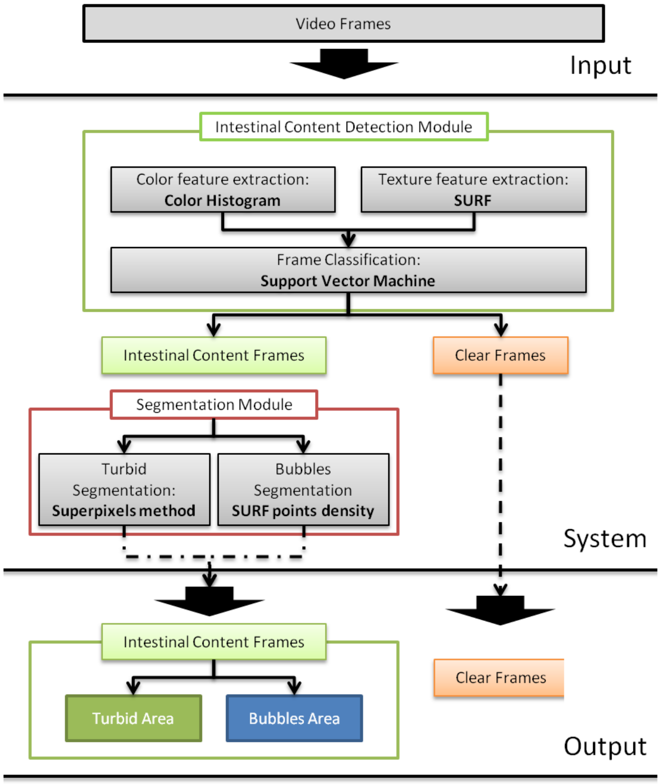


Fig. 3: System Architecture for detection and segmentation of intestinal content.

A. Image Classification

The first step of the system finds those frames with intestinal content. As explained in Section 2, there are two different types of intestinal content: turbid, which is characterized by color information, and bubbles, which are characterized by texture. In order to detect both types of intestinal content, two feature descriptors are used: color histograms and texture descriptor. These both image features are merged to learn a SVM Classifier [32].

1) Color Features: Color Features: Experts recognition of the turbid frames relies more on color information than on texture information. Color variability of intestinal content is very high and depends on: the patient, the clinical preparation, the food ingested by the patient as well as on the camera type, the light source, the reflections and the distance of the content to the capsule.

In order to reduce the image information a color quantization is perform from 16 million colors to 64 colors. Typically, color image quantization is performed dividing the original color space into smaller subregions of equal size [27]. It is well known that in WCE videos only a subset of colors is observed. For instance there are colors which practically does not appear in WCE frames like blue or violet colors. Furthermore, most of the observed colors in WCE are concentrated in small region of the RGB space. This information can be used to reduce the dimensionality of color representation with minimum loss to the 64 colors. This set of 64 selected colors defines the color map, will be referred along the article as *Intes Color Map*.

The *Intes Color Map* was created using all the frames from 80 WCE external studies (not included in our database). The three-dimensional RGB data representing all observed colors in these videos was clustered into 64 clusters using k-means technique [33]. Using this learned color map the mean quantization error is 10.02 ($std = 9.98$) while using an uniform partition of the original *rgb* color map the mean quantization error is 33.33 ($std = 19.23$). The measure to evaluate the quantization error for each pixel is the Euclidean distance between the original RGB color and the assigned color centroid from the used colormap.

2) Textural Features: Visually the bubbles are described by the presence of several circular blobs. The opacity of the bubbles is very variable and it is directly correlated with the color of intestinal liquid. However, sometimes the opacity of the bubbles is very low and its appearance is nearly transparent. These set of images are presented as a blurred image with similar colors to the intestinal wall and only characteristic that describes these frames is the contour of the circular bubbles.

In [29] and [27] the authors presented a method to detect bubble images based on Gabor filters [34] and the Gauss Laguerre function [35]. These methods achieve satisfactory results, but both methods suffer from a high computational cost. Additionally, the correct choice of the filter parameters is a critical step in these methods since bubble size can vary not only between different frames but also in single image. In this paper, we propose to use the Speeded-Up Robust Feature (SURF) detector [36]. The SURF method is a scale- and rotation-invariant interest point detector and descriptor. The method uses an integer approximation of the determinant of Hessian blob detector in order to detect points of interest. Given a point $\mathbf{x} = (x, y)$ in an image I, the Hessian matrix $H(\mathbf{x}, s)$ in $\mathbf{x}$ at scale s is defined as:

$$H(x, s) = \begin{vmatrix} L_{xx}(\mathbf{x}, s) & L_{xy}(\mathbf{x}, s) \\ L_{xy}(\mathbf{x}, s) & L_{yy}(\mathbf{x}, s) \end{vmatrix} \quad (1)$$

where $L_{xx}(\mathbf{x}, s)$ is the convolution of the Gaussian second order derivative $\frac{d^2}{d\mathbf{x}^2}G(s)$ with the image I in point $\mathbf{x}$, and similarly for $L_{xy}(\mathbf{x}, s)$ and $L_{yy}(\mathbf{x}, s)$. The determinant of the Hessian-Matrix is the blob detector response. If the response is lower than given threshold thr_surf the response is rejected and not considered as point of interest. Thus, the SURF detector can be seen as a blob detector, and SURF method can be applied to the problem of bubble detection because of the fact that one bubble can be considered as one blob. Our assumption is that the more points of interest are detected the more bubbles are in the frame. One of the advantages of SURF method in comparison to the Gabor filter is the computational cost, being 50 times faster.

3) Classification: In order to classify the frames in the classes $\{intestinal\ content,\ clear\}$, both color and textural features are merged. This is done by simply expanding the color histogram by one extra bin representing the number of points of interest. In this way a 65 bin feature vector is obtained. Afterwards each frame is classified using the Support Vector Machine (SVM) classifier [32].

The classical implementation of SVM classifier looks for the hyperplane which separates the data into two subspaces (positive and negatives samples) while maximizing margin. Originally, the algorithm proposed by V. Vapnik was a linear classifier, however it can be easily enhanced to non-linear classifier applying the kernel trick [37]. The margin which defines the classifier is defined as the distance between hyperplane and instances of positive and negative samples. Given a training set $\mathbf{X}_{1..N}$ containing N training labeled samples and coefficients $\alpha_{1..N}$ learned in training step, the decision function of SVM takes the following form:

$$y(x) = \sum_i \alpha_i K(\mathbf{X}_i, x) + b \tag{2}$$

where $K()$ is a kernel function and x the input vector. In our system, as our features are represented by histograms, we use the Histogram Intersection Kernel [38] defined as follows:

$$K_{int}(z, z') = \sum_j^m \min(z_j, z'_j) \tag{3}$$

where $z = \{z_1, .., z_m\}$ and $z' = \{z'_1, .., z'_m\}$ are the histograms with $m - 1$ bins representing color information and one bin representing number of points of interest.

B. Image Segmentation

A frame classified as *intestinal content* frame can be either completely or partially covered by turbid or bubbles. In the former case these frames are usually filtered by the system, without further processing. However, in the latter case it can be important to identify the image region not hindered by intestinal content, since it can potentially convey relevant information (polyps, ulcers, bleeding, etc.). Additionally, while the physiological meaning of bubbles and turbid is completely different the output of the proposed segmentation method is based on three labels: {*clear*, *turbid* and *bubbles*}. Regarding the visual difference between bubbles and turbid two methods are proposed.

1) Intestinal Content Segmentation based on color: In order to obtain the exact area covered by intestinal content in the image (which includes both turbid and bubbles) each pixel should be labeled as *intestinal content/clear*. However, a single pixel is not enough descriptive about the frame content. Many existing algorithms in computer vision use the pixel-grid as the underlying the image representation. However, this pixel-grid is not a natural representation of visual scenes and could contain pixels from both classes *intestinal content/clear*. In order to overcome this problem we propose to detect and classify homogeneous regions in the images. We refer homogeneous regions as a group of pixels with a perceptually consistent information (color and texture).

The homogeneous regions are obtained using the ”superpixel” method proposed by Ren and Malik in [39]. ”Superpixels” are obtained using the Normalized Cuts method (NCuts) [40]. NCuts is a the classical region segmentation method which uses spectral clustering to exploit pairwise brightness, color and texture affinities between pixels. Rather than focusing on local features and their consistencies in the images, the aim of NCuts consist of extracting the global impression of an image. The number of ”superpixels” depends on the entrance parameter and can be set using cross-validation.

In the proposed system the only frames being superpixelized are the frames previously detected (step 1) as intestinal content frames. In order to classify each superpixel as *intestinal content/clear* region the linear SVM classifier is used. In order to classify each ”superpixel” we used the mean intensity of the pixels (for each channel R,G and B) inside the ”superpixel” region as a feature descriptor.

Algorithm 1 Algorithm for intestinal content segmentation

```
Input: image $I$ and number of regions $N$
  Compute $N$ regions $R_N$ using NCuts method.
  for $i = 1$ to $N$ do
    Compute feature vector $f_i = [f_i^r, f_i^g, f_i^b]$ as the mean values of each rgb channel of pixels in $R_i$
    Classify $R_i$ using Linear SVM classifier and feature vector $f_i$
    Based on the classification result assign boolean value $\{intestinal\ content, clear\}$ to all pixels in $R_i$
  end for
Output: Binary image representing segmented regions $\{intestinal\ content, clear\}$.
```

2) Bubble Frame Segmentation: Bubble image area is estimated by analyzing the spatial distribution of the interest points of SURF method. SURF method detects blobs which are correlated with the number of bubbles in the image, and hence the segmentation of bubble area can be done by analyzing the density of interest points. The area of the image with high point density is considered to be a bubble region, otherwise it is considered to be a clear region. The density is estimated using a kernel density method. Let s be a location in the image I and $p_{1..n}$ are the locations of the interest points detected by SURF. The estimation of the intensity using the kernel method is given by:

$$\hat{f}_k(s) = \frac{1}{\sigma_k(s)} \sum_{i=1}^{n} \frac{1}{h^2} k\left(\frac{s - p_i}{h}\right) \tag{4}$$

where $\sigma_k(s)$ is the correction for edge effects for location s, k is the kernel and h is the bandwidth. We use the quadratic kernel proposed by Bailey and Gatrell in [41]:

$$k(\mathbf{u}) = \frac{3}{\pi}(1 - \mathbf{u}^t\mathbf{u})^2 \qquad \mathbf{u}^T\mathbf{u} \leq 1 \tag{5}$$

When this kernel function is substituted into Equation 4 and $\sigma_k(s)$ if fixed to 1 the following estimate function of the intensity is obtained:

$$\hat{f}_k(s) = \sum_{d_i \leq h} \frac{3}{\pi h^2} \left(1 - \frac{d_i^2}{h^2}\right)^2 \tag{6}$$

where d_i is the distance between location s and interest points locations p_i. Finally, given the bubble density image, the bubble area is defined as the region where the density value is higher than a threshold thr_b.

C. Final labeling

According to the output from two proposed segmentation methods the system output is defined as:

- *Bubbles*: All image pixels that belong to the bubble area estimated by the bubble segmentation method.
- *Turbid*: The set of image pixels in the area estimated by intestinal content segmentation based on color and not considered as bubbles.
- *Clear*: All other image pixels (not bubbles and not turbid).

IV. Experimental Results

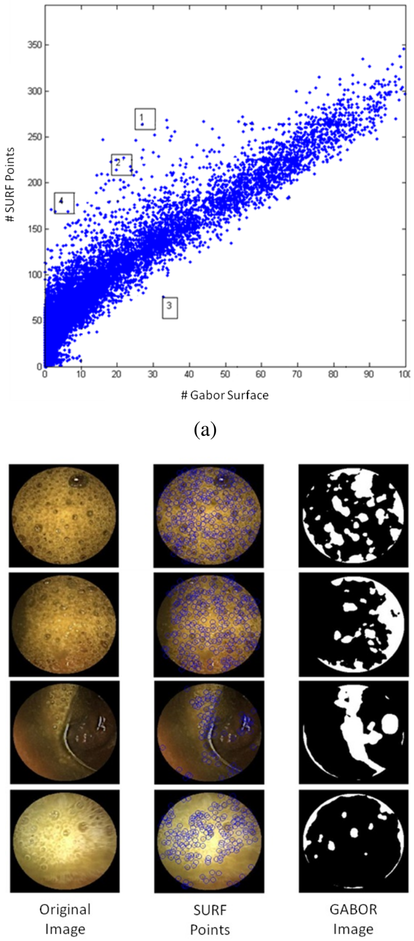
In this Section, we present the experimental results of the proposed system for automatic characterization of intestinal content frames. First, we describe the data set and the evaluation procedure, and then, we show the qualitative and quantitative results of all parts of the proposed system. In detail we will present the validation of:

- SURF detector;
- Intestinal content detector;
- Intestinal content segmentation.

A. Database

The data set was obtained using the SB2 capsule endoscopy camera developed by Given Imaging, Ltd., Israel [42]. All cases were conducted in the same conditions at Digestive Diseases Department, Hospital General ”Vall d’Hebron” in Barcelona, Spain [28].

For the experimental setup a set of 50 studies from different subjects has been used. For every video the duodenum and cecum entrance was marked by medical experts, and the video was analyzed only inside those thresholds. A random set of frames from each video was selected and then labeled as intestinal content, clear by a medical expert. The number of frames per video is a number between 1000 and 2000 and depends on video length. These frames represents between 5% to 10% of the video frames from the duodenum until the the cecum. Table I shows the list of videos used in the experiments, indicating the number of frames from each class. As it can be observed, there is a high variability in terms of percentage of intestinal content in videos: there are some videos which practically does not present intestinal content (video 36 and 37) and there are some which intestinal content is present in more than 80% of the frames of the video (video 29).



(a)

(b)

Fig. 4: This figure shows the correlation between Gabor and SURF methods, both applied to the detection and segmentation of bubble frames. Figure a) shows the correlation graph ($r = 0.95$) between Gabor surface and Surf points in bubble detection problem. Each point in the graph represents one single frame. Ordinate axis represents the $\%$ of frame surface covered by bubbles following the Gabor method. Abscise axis represents the number of SURF points detected in that frame. With numbers 1, 2, 3 and 4 some outliers have been marked. The outliers and the output of SURF and Gabor methods for these frames are shown in Figure b).

B. SURF detector validation

In this experiment we compare the obtained results using the SURF method with those obtained by the proposed method in [29], which estimates the bubble area using Gabor filters. Threshold thr_surf has been fixed by cross-validation to 65.000.

In Fig. 4 we present a scatter plot showing the correlation graph between the output of both methods: the number of interest points detected by SURF method and the surface area estimated by Gabor filters. Pearson correlation coefficient r is used in order to evaluate the output of both methods. The obtained value ($r = 0.95$) indicates that the methods are highly correlated. As it can be seen, there are only some samples which present a significant difference between methods. In the same figure we show four images (marked with blue square) where the methods present a low correlation. As it can be seen, the qualitative analysis of these outliers show that the proposed method performs better than the Gabor filter for the case of blurred bubbles, and with an extremely low computational cost.

TABLE I: Database: List of 50 used videos indicating the number of clear and turbid frames used from each video.

Video	#clear/IC frames	Video	#clear/IC frames	Video	#clear/IC frames	Video	#clear/IC frames	Video	#clear/IC frames
Video1	1327 / 672	Video11	993 / 1006	Video21	615 / 1225	Video31	1857 / 143	Video41	1788 / 212
Video2	1451 / 549	Video12	992 / 353	Video22	1815 / 140	Video32	1766 / 234	Video42	988 / 1012
Video3	1205 / 795	Video13	1369 / 470	Video23	200 / 1453	Video33	1892 / 108	Video43	1769 /232
Video4	1008 / 992	Video14	1184 / 499	Video24	457 / 1535	Video34	1921 / 79	Video44	1886 / 114
Video5	1135 / 865	Video15	434 / 1389	Video25	1383 / 298	Video35	1517 / 483	Video45	1406 / 594
Video6	1530 / 434	Video16	502 / 1375	Video26	1904 / 85	Video36	1993 / 7	Video46	1307 / 693
Video7	1346 / 453	Video17	1418 / 192	Video27	1390 / 527	Video37	1998 / 2	Video47	1041 / 959
Video8	1203 / 797	Video18	750 / 738	Video28	709 / 847	Video38	1631 / 369	Video48	1120 / 880
Video9	1337 / 613	Video19	1556 / 302	Video29	223 / 1763	Video39	1974 / 26	Video49	1743 / 257
Video10	624 / 1261	Video20	1288 / 476	Video30	828 / 867	Video40	1762 / 238	Video50	1714 / 286

C. *Intestinal Content Classification*

In this Section we evaluate the proposed system for detecting frames with intestinal content. In order to assess the method a leave one video out validation method is used. The parameters used to evaluate a classifier are:

- Accuracy (A) = $\frac{TP+TN}{TP+FP+TN+FN}$
- Sensitivity (S) = $\frac{TP}{TP+FN}$
- Specificity (K) = $\frac{TN}{TN+FP}$
- Precision (P) = $\frac{TP}{TP+FP}$

where TP = true positive, TN = true negative, FP = false positive and FN = false negative. The frames with intestinal content are considered the positive samples and clear frames the negative cases.

We compare the results of our system with two previously proposed methods by Bashar et al. in [27]: 1) SVM classifier using *Color Moment Features* and, 2) SVM classifier using *HSV 64 bin color Histogram* features. Additionally, we test a simplified version of our proposed system that uses only 64 bin color histogram (without texture information). The results are presented in Table II where the mean value and the standard deviation of different methods are presented. We can see that the proposed method that uses color and textural information achieves the best result outperforming others methods in all measurements (accuracy, sensitivity, specificity and precision). The box plots of accuracy are presented in the Fig. 5, where it can be seen that the proposed method have the smallest variance.

TABLE II: Accuracy of intestinal content detection methods

	Accuracy	Sensitivity	Specificity	Precision
Color Moments	83.6 ± 10.9%	54.4 ± 24.1%	92.3 ± 10.5%	73.4 ± 26.9%
HSV 64bin	89.9 ± 7.8%	73.7 ± 22.6%	93.1 ± 9.5%	83.8 ± 21.8%
IntesColorMap 64bin	91.2 ± 6.9%	78.3 ± 17.7%	92.8 ± 8.1%	82.5 ± 18.4%
IntesColorMap 64bin + Bubbles	91.6 ± 6.6%	80.1 ± 16.7%	93.1 ±7.9%	83.0 ± 18.2%

D. *Study of reliability*

Classifier is reliable when the training data used in the model construction process well represents the data that are to come in test process. In order to ensure that our approach is accurate and consistent with respect to new videos an analysis of the training set is done. In this analysis the 50 (labeled as $\{$*intestinal content*, *clear*$\}$ frames) WCE videos are used (for details see Table I).

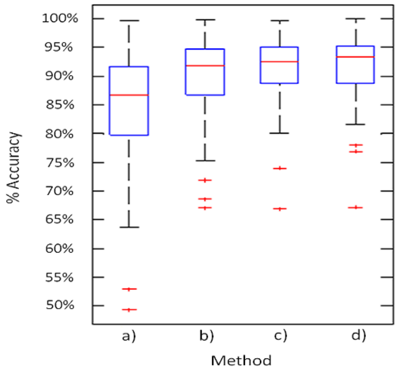


Fig. 5: Box plots of intestinal content classification results using different sets of features: a) Color Moment Features b) HSV histogram with 64 bins c) *Intes Color Map* histogram with 64 bins and d) *Intes Color Map* histogram with 64 bins + bubbles. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

In the experiments presented in this section two questions are tackled. In the first one, we would like to answer the question of the turbid variability between different subjects. In the second one, we would like to determine the minimum number of videos that builds a reliable classifier.

The results on turbid variability between different subjects are presented on Fig. 6 and Fig. 7. Fig. 6 represents the results of training the classifiers with one video and testing the classifier accuracy with remaining 49 videos. The experiment is repeated 50 times (once for each video in training set). As it can be seen, the variability of the accuracy is quite high. For instance there are some videos (#10, #21 and #38) that, when used for training the classifier, give good and generalizable models for remaining 49 videos. For these videos the median accuracy is high and the variance is small. These videos contain high variability of both intestinal content and clear frames that well represents the images in remaining videos. On the other hand, some models obtained by using videos #36, #37 or #39 are not able to generalize the results in the remaining set of videos, the median accuracy is low and the variance is high. There are two possible justifications to this observation: 1) those videos are very homogenous and they do not offer enough information to learn a good classifier, or 2) these videos contain information (e.g. strange turbid color) that is not seen in other videos (they can be outlier videos). Those videos are interesting because they can provide new information about data distribution that might be helpful in discriminative model construction. Second

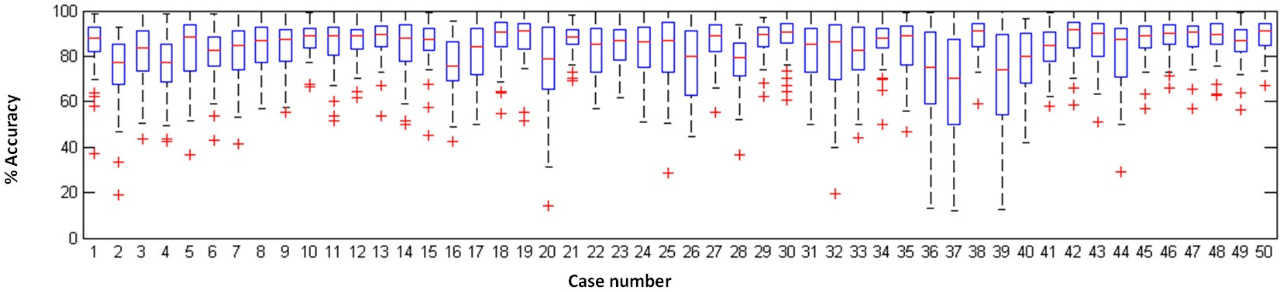


Fig. 6: Each boxplot represents the accuracy obtained by testing a classifier learned by the video of x-axis with all other 49 videos in our dataset.

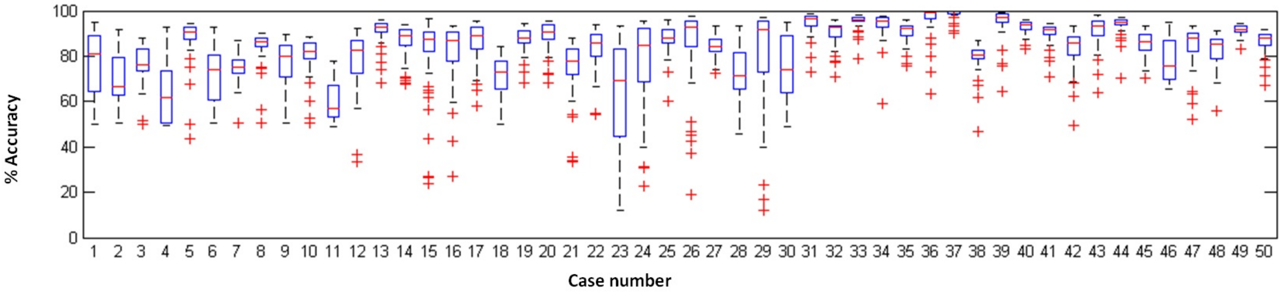


Fig. 7: Each boxplot represents the accuracy obtained by testing the video represented by x-axis using using 49 different classifiers which are learned using one different video in each case.

experiment, designed to study the variability between different subjects, shows how good a video is represented by the set of videos (see Fig. 7). Here each video were classified using 49 classifiers trained with the data from remaining videos, the experiment was repeated 50 times once for each video. As it can be seen, there are some videos (#33 and #37) that are well classified by all trained models, they have high median accuracy and small variance. These are homogenous videos with frequently appearing images in other videos. On the other hand, there are some videos (#23 and #29) where the majority of the learned classifiers have problem to correctly classify them. These videos contain frames with color and texture that are not frequently observed in WCE video.

Usually, the larger training set the better classification results. However it is important to remember that the data acquisition and labeling have an associated cost limiting the size of the training data set. To evaluate the influence of the training set size on the classifier accuracy a set of 10 test videos was randomly selected from a pool of 50 videos. Then these videos were classified using different size training data sets (from 1 to 40 videos). At each iteration of the test one video was added to the training set. The results are presented on Fig. 8, where a learning curve for each video is presented. We can observe that when training set contains 30 or more videos the classifier accuracy stabilizes. Moreover, we see that for some videos a high accuracy is achieved using small size training set. These results show an asymptotically convergent learning curve which appears to assess the validity of the size of our dataset.

E. Segmentation

In this section the intestinal content segmentation is evaluated. Note that our segmentation tasks are performed by using two methods: 1) intestinal content segmentation based on color information of the regions, and 2) bubble regions segmentation based on textural information. The turbid area can be defined as the difference between intestinal content region and bubble region. Finally, the output of the proposed method is an image segmentation using 3 different labels $\{$*clear*, *turbid* and *bubbles*$\}$.

In order to evaluate the methods a set of 350 images from the big dataset of Intestinal Content Frames (see Table I) was tested. These images were manually selected by an expert with the purpose of selecting a set of frames with high variability in terms of content percentage, texture and color. However in case of frames segmentation, the manual annotation was done by three different experts using $\{$*clear*, *turbid* and *bubbles*$\}$ labels. To measure the performance of the segmentation algorithm we use the % of the image correctly labeled based on the ground-truth provided by the experts. The manual segmentation of intestinal content is a complex task. This complexity sometimes leads to an ambiguity between annotations from same/different experts. The uncertainty of the experts arises from the variability of intestinal content. Frequently the limits between the intestinal content and intestinal wall or lumen are questionable while a clear contour is not preserved. Moreover, the variability is higher in case of semi-transparent intestinal content.

In order to evaluate intra-user variability the overlapping area between the annotations of the three experts was cal-

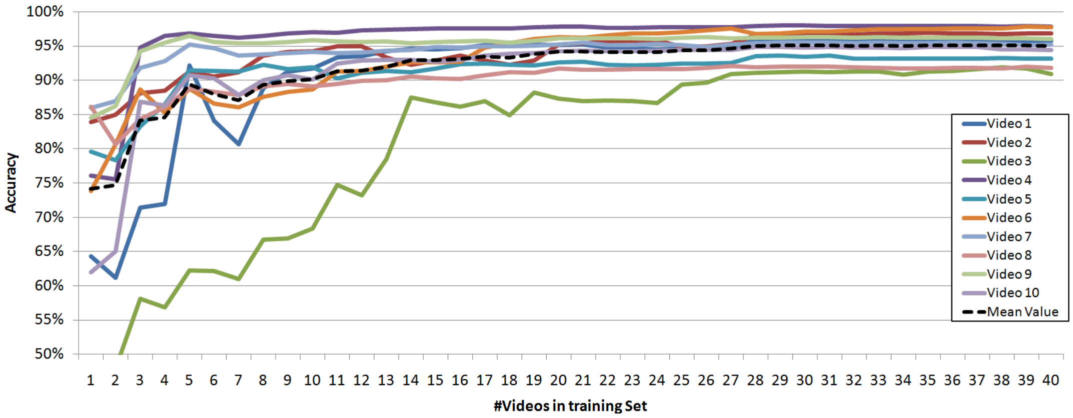


Fig. 8: System accuracy for 10 videos from different subjects. X-axis represents the number of videos in the training data set.

culated. The results are presented in Table III where the mean intra-user variability is presented. As it can be observed, the user variability on bubbles regions is low, presenting an overlap regions higher that 99% between the experts. However, turbid regions presents a user-variability between annotations of 10%.

TABLE III: Intestinal Content Segmentation: User-variability

	Overlap Area		
	Expert 1&2	Expert 1&3	Expert 2&3
Turbid	91.86%	89.71%	92.63%
Bubbles	99.04%	99.41%	98.30%
Intestinal Content (Bubbles + Turbid)	91.22%	89.39%	91.99%

In the second experiment we evaluated both segmentation methods. In order to evaluate the methods we present qualitative and quantitative results. The first evaluated method is the turbid segmentation method. As it was commented in the methodology, this method has two steps: 1) divide the image in N ”superpixels” regions using NCut Method, and 2) the regions are classified using a Linear SVM classifier. The N parameter which defines the number of regions was set to 60 and the regularization constant parameter C of SVM classifier was set to 1 after cross-validation. In order to evaluate the system all tests was performed using the leave-one-image-out validation method.

In Fig. 9, qualitative results for the intestinal content segmentation method based on color are presented using 9 different images containing bubbles and turbid. The second column in the mosaic shows the ”superpixel” regions, and the third column shows the regions classified as intestinal content. As it can be observed, both turbid and bubble regions are classified as intestinal content regions in most of the images. The bubble segmentation method has two parameters which were fixed by cross-validation: the kernel bandwidth $h = 50$ and the threshold $thr_b = 0.1$. The qualitative results of the method are presented in Fig. 10. For each evaluated sample the original image, the output of the density estimation method and the final binary output are presented. As it can be seen, only the image regions containing bubbles are detected.

In Fig. 11 we present the overall qualitative results. A set of 20 random images from the test set is shown. For each image we present the segmentation output with the associated labels $\{$*clear*, *turbid* and *bubble*$\}$. The labels are represented with the following colors $\{$*black*, *gray* and *white*$\}$. In this mosaic we can observe that using both segmentation systems we are able to differentiate the physiological meaning of the intestinal content.

Finally, in Table IV the quantitative results are presented. This table shows the overlap area between the annotations of the 3 experts and output area from each method (*turbid*, *bubbles* and *intestinal content* segmentation). We can appreciate that the bubble segmentation method outperforms the result on the turbid segmentation. This result is supported by the results obtained in the intra-user variability experiment.

TABLE IV: Segmentation results

	Overlap Area		
	Expert 1	Expert 2	Expert 3
Turbid	78.04%	81.60%	79.16%
Bubbles	92.43%	92.25%	92.60%
Intestinal Content (Turbid + Bubbles)	81.71%	85.28%	82.88%

V. Conclusion

In this paper, we propose and evaluate an automatic system for categorization and segmentation of intestinal content frames for WCE. The three main contributions of this paper are: 1) development and validation of an automatic system for intestinal content detector; 2) development and validation of a segmentation method for detection bubbles and turbid media in WCE images; and 3) definition of a new image feature of WCE: area covered by each kind of intestinal content.

The presented method is divided in two steps. In the first step the frames with intestinal content are detected using a color and textural feature and a Linear SVM classifier. In the second step of the system, those detected frames are segmented obtaining the image regions of bubbles and turbid media.

The evaluation of the proposed system, using a large data set, shows that the presented method for detecting intestinal content frames outperforms the results of the state-of-the-art methods. Moreover, we observe that, regarding the intestinal content variability in terms of color and texture, a large data set is needed to ensure the generalization of the method, and in this sense, our experiments confirm the statistical robustness

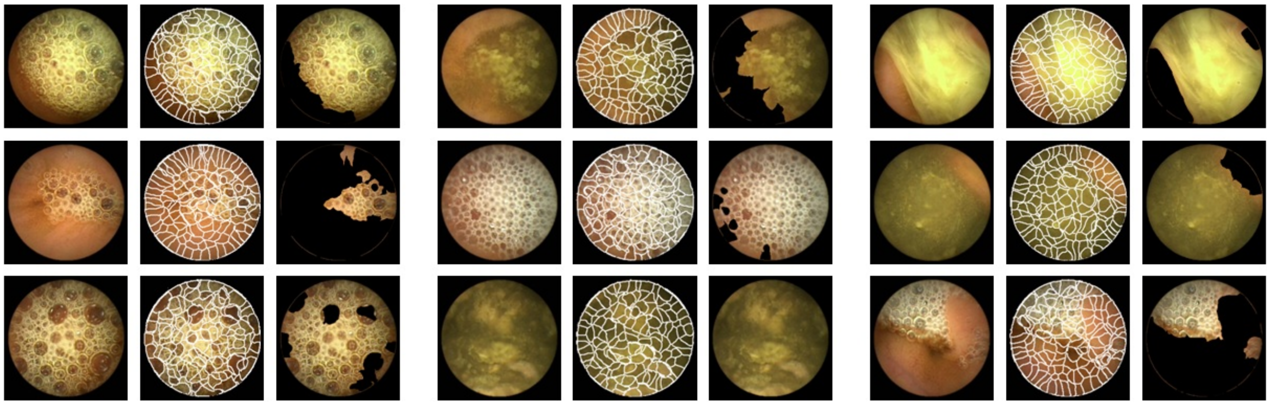


Fig. 9: Qualitative results obtained from 9 random images by using the proposed intestinal content segmentation method.

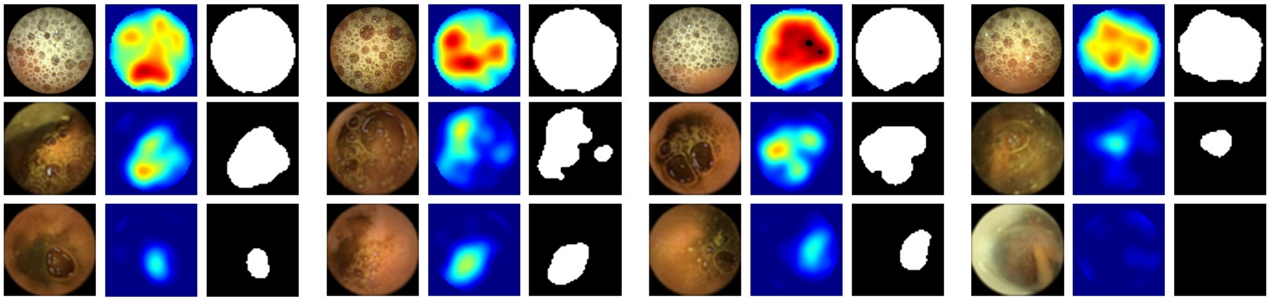


Fig. 10: Qualitative results obtained from 12 random images by using the proposed bubble segmentation method.

of the presented outcomes. Finally, qualitative and quantitative results of segmentation method present good performance when discriminating intestinal content in bubbles and turbid.

As a future work, an analysis of the presence and dynamic distribution of different kinds of intestinal content (turbid and bubbles) along the intestinal tract will be studied. Regarding physiological meaning of different kinds of intestinal content, the analysis of the proposed feature could be useful for the evaluation of disorders of intestinal motility.

Acknowledgment

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, as well as by the MICINN Grants TIN2009-14404-C02 and CONSOLIDER-INGENIO 2010 (CSD2007-00018).

References

[1] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, “Wireless capsule endoscopy,” *Nature*, vol. 405, pp. 4–7, 2000.

[2] Z. Fireman, E. Mahanja, E. Broide, M. Shapiro, L. Fich, A. Sternberg, Y. Kopelman, and E. Scapa, “Diagnosing small bowel crohns disease with wireless capsule endoscopy,” *Gut*, vol. 52, pp. 390–392, 2003.

[3] B. Li and M. Q. H. Meng, “Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments,” *Comput. Biol. Med.*, vol. 39, pp. 141–147, February 2009.

[4] K. Schulmann, S. Hollerbach, K. Kraus, J. Willert, T. Vogel, G. Moslein, C. Pox, M. Reiser, A. Reinacher-Schick, and W. Schmiege, “Feasibility and diagnostic utility of video capsule endoscopy for the detection of small bowel polyps in patients with hereditary polyposis syndromes,” *The American Journal of Gastroenterology*, vol. 100, no. 1, p. 27, 2005.

[5] Y. Metzger, S. Adler, A. Shitrit, B. Koslowsky, and I. Bjarnason, “Comparison of a new pillcam sb2 video capsule versus the standard pillcam sb for detection of small bowel disease,” *Reports in Medical Imaging*, vol. 2, pp. 7–11, 2009.

[6] M. Mackiewicz, *Capsule Endoscopy - State of the Technology and Computer Vision Tools After the First Decade, New Techniques in Gastrointestinal Endoscopy*. Oliviu Pascu and Andrada Seicean (Ed.), 2011, vol. 1.

[7] A. Moglia, A. Menciassi, and P. Dario, “Recent patents on wireless capsule endoscopy,” *Recent Patents on Biomedical Engineering*, vol. 1, pp. 24–33, 2008.

[8] T. Yamada, D. Alpers, A. Kalloo, D. Powell, and C. Owyang, *Principles of Clinical Gastroenterology*. John Wiley & Sons, 2011. [Online]. Available: http://books.google.es/books?id=MYDoT1PT5N4C

[9] V. Hai, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa, and Y. Yagi, “Adaptive control of video display for diagnostic assistance by analysis of capsule endoscopic images,” in *Proceedings of the 18th International Conference on Pattern Recognition - Volume 03*, ser. ICPR ’06, 2006, pp. 980–983.

[10] Y. Yagi, H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, and T. Arakawa, “A diagnosis support system for capsule endoscopy,” *Inflammopharmacology*, vol. 5, no. 2, pp. 78–83, 2007.

[11] H. Vu, R. Sagawa, Y. Yagi, T. Echigo, M. Shiba, K. Higuchi, T. Arakawa, and K. Yagi, “Evaluating the control of the adaptive display rate for video capsule endoscopy diagnosis,” in *Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics*, 2009, pp. 74–79.

[12] P. M. Szczypinski, R. D. Sriram, P. V. Sriram, and D. N. Reddy, “A model of deformable rings for interpretation of wireless capsule endoscopic videos,” *Medical Image Analysis*, vol. 13, no. 2, pp. 312 – 324, 2009, includes Special Section on Functional Imaging and Modelling of the Heart.

[13] A. Karargyris and N. Bourbakis, “Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos,” *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 10, pp. 2777 –2786, 10 2011.

[14] S. Hwang and M. E. Celebi, “Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature,” in *Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, march 2010, pp. 678 –681.

[15] Y. S. Jung, Y. H. Kim, D. H. Lee, and J. H. Kim, “Active blood detection in a high resolution capsule endoscopy using color spectrum transformation,” in *Proceedings of International Conference on BioMedical Engineering and Informatics*, 2008, pp. 859–862.

[16] G. Pan, G. Yan, X. Qiu, and J. Cui, “Bleeding detection in wireless

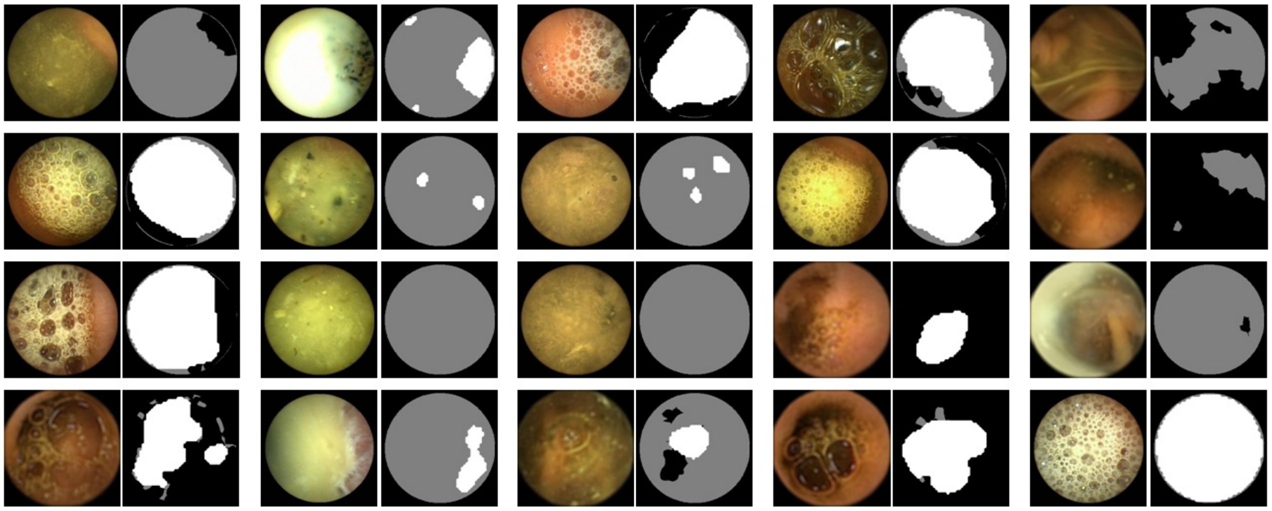


Fig. 11: Results of Intestinal content segmentation from 20 random images. Black areas mean clear region, white means region with bubbles and grey areas means turbid regions.

capsule endoscopy based on probabilistic neural network," *Journal of Medical Systems*, vol. 35, pp. 1477–1484, 2011.

[17] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, A. Member, and M. Tzivras, "Computer-aided tumor detection in endoscopic video using color wavelet features," *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, pp. 141–152, 2003.

[18] C. Malagelada, F. De lorio, S. Seguí, S. Mendez, M. Drozdzal, J. Vitrià, P. Radeva, J. Santos, A. Accarino, J. R. Malagelada, and F. Azpiroz, "Functional gut disorders or disordered gut function? small bowel dysmotility evidenced by an original technique," *Neurogastroenterology & Motility*, vol. 24, no. 3, pp. 223–e105, 2012.

[19] E. J. Ciaccio, C. A. Tennyson, S. K. Lewis, S. Krishnareddy, G. Bhagat, and P. H. R. Green, "Distinguishing patients with celiac disease by quantitative analysis of videocapsule endoscopy images," *Comput. Methods Prog. Biomed.*, vol. 100, pp. 39–48, October 2010.

[20] R. Kumar, Q. Zhao, S. Seshamani, G. Mullin, G. Hager, and T. Dassopoulos, "Assessment of crohn's disease lesions in wireless capsule endoscopy images," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 355 –362, feb. 2012.

[21] L. Igual, J. Vitrià, F. Vilariño, S. Seguí, C. Malagelada, F. Azpiroz, and P. Radeva, "Automatic discrimination of duodenum in wireless capsule video endoscopy," *IFMBE Proceedings*, vol. 22, pp. 1536–1539, 2008.

[22] J. P. S. Cunha, M. Coimbra, P. Campos, and J. M. Soares, "Automated topographic segmentation and transit time estimation in endoscopic capsule exams," *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 19–27, 2008.

[23] J. Lee, J.-H. Oh, S. K. Shah, X. Yuan, and S. J. Tang, "Automatic classification of digestive organs in wireless capsule endoscopy videos," in *Proceedings of the 2007 ACM symposium on Applied computing*, 2007, pp. 1041–1045.

[24] H. Vu, T. Echigo, R. Sagawa, K. Yagi, M. Shiba, K. Higuchi, T. Arakawa, and Y. Yagi, "Detection of contractions in adaptive transit time of the small bowel from wireless capsule endoscopy videos," *Comput. Biol. Med.*, vol. 39, pp. 16–26, January 2009.

[25] F. Vilariño, P. Spyridonos, F. De Iorio, J. Vitrià, F. Azpiroz, and P. Radeva, "Intestinal motility assessment with video capsule endoscopy: automatic annotation of phasic intestinal contractions." *IEEE Trans Med Imaging*, vol. 29, no. 2, pp. 246–59, 2010.

[26] S. Seguí, L. Igual, F. Vilariño, P. Radeva, C. Malagelada, F. Azpiroz, and J. Vitrià, "Diagnostic system for intestinal motility disfunctions using video capsule endoscopy," in *ICVS*, 2008, pp. 251–260.

[27] M. Bashar, T. Kitasaka, Y. Suenaga, Y. Mekada, and K. Mori, "Automatic detection of informative frames from wireless capsule endoscopy images," *Medical Image Analysis*, vol. 14, no. 3, pp. 449–470, 2010.

[28] C. Malagelada, F. D. Iorio, F. Azpiroz, A. Accarino, S. Seguí, P. Radeva, and J.-R. Malagelada, "New insight into intestinal motor function via noninvasive endoluminal image analysis," *Gastroenterology*, vol. 135, no. 4, pp. 1155 – 1162, 2008.

[29] F. Vilariño, P. Spyridonos, O. Pujol, J. Vitrià, and P. Radeva, "Automatic detection of intestinal juices in wireless capsule video endoscopy," in *18th Inter. Conf. on Pattern Recognition (ICPR)*, 2006, pp. 20–24.

[30] K. Barrett, L. Johnson, F. Ghishan, J. Merchant, and H. Said, *Physiology of the Gastrointestinal Tract*, ser. Physiology of the Gastrointestinal Tract. Elsevier Science, 2006, no. v. 2. [Online]. Available: http://books.google.es/books?id=j6Z5tQAACAAJ

[31] J.-R. Malagelada and F. Azpiroz, *Determinants of gastric emptying and transit in the small intestine*. John Wiley & Sons, Inc., 2010.

[32] V. N. Vapnik, "An overview of statistical learning theory," *Neural Networks, IEEE Transactions on*, pp. 988–999, 1999.

[33] S. Lloyd, "Least squares quantization in pcm," *Information Theory, IEEE Transactions on*, vol. 28, no. 2, pp. 129 – 137, Mar. 1982.

[34] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 891–906, 1991.

[35] P. Campisi and G. Scarano, "A multiresolution approach for texture synthesis using the circular harmonic functions," *Image Processing, IEEE Transactions on*, vol. 11, no. 1, pp. 37 –51, Jan. 2002.

[36] H. Bay, A. Ess, T. Tuytelaars, and L. van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346–359, June 2008.

[37] A. Aizerman, E. M. Braverman, and L. I. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and Remote Control*, vol. 25, pp. 821–837, 1964.

[38] S. Maji, A. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conf. on*, June 2008, pp. 1 –8.

[39] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 10 2003, pp. 10 –17 vol.1.

[40] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888 –905, Aug. 2000.

[41] T. C. Bailey and A. C. Gatrell, *Interactive Spatial Data Analysis*. London: Longman Scientific and Technical, 1995.

[42] (2001) Given imaging, ltd. [Online]. Available: http://www.givenimaging.com/