Intestinal Motility Assessment With Video Capsule Endoscopy: Automatic Annotation of Phasic Intestinal Contractions

Fernando Vilariño*, Panagiota Spyridonos, Fosca DeIorio, Jordi Vitrià, Fernando Azpiroz, and Petia Radeva

***Abstract*—Intestinal motility assessment with video capsule endoscopy arises as a novel and challenging clinical fieldwork. This technique is based on the analysis of the patterns of intestinal contractions shown in a video provided by an ingestible capsule with a wireless micro-camera. The manual labeling of all the motility events requires large amount of time for offline screening in search of findings with low prevalence, which turns this procedure currently unpractical. In this paper, we propose a machine learning system to automatically detect the phasic intestinal contractions in video capsule endoscopy, driving a useful but not feasible clinical routine into a feasible clinical procedure. Our proposal is based on a sequential design which involves the analysis of textural, color, and blob features together with SVM classifiers. Our approach tackles the reduction of the imbalance rate of data and allows the inclusion of domain knowledge as new stages in the cascade. We present a detailed analysis, both in a quantitative and a qualitative way, by providing several measures of performance and the assessment study of interobserver variability. Our system performs at 70% of sensitivity for individual detection, whilst obtaining equivalent patterns to those of the experts for density of contractions.**

***Index Terms*—Imbalanced data classification, intestinal motility, video capsule endoscopy.**

I. Introduction

SMALL intestine motility dysfunctions are shown to be related to certain gastrointestinal pathologies which can be manifest in a varied symptomatology [1]. Ileus, bacterial overgrowth, and the irritable bowel syndrome have been reported as major clinical disorders. The analysis of the intestinal contractions of the small bowel, in terms of number, frequency, and distribution along the intestinal tract, represents one of the methods with the highest clinical significance [2], [3]. Current techniques for assessment of small intestinal motility are multiple and complementary [2], [4], but small intestinal manometry is widely accepted as the most reliable so far. Manometry is an invasive and discomforting test based on the measure of the pressure in certain points of the gut, lacking of sensitivity over certain types of weak intestinal contractions.

In this paper, we address the study of intestinal contractions in a novel approach using wireless capsule video endoscopy (WCVE) as data source. WCVE consists of a capsule with a camera, a battery, and a set of led lamps for illumination attached to it, which is swallowed by the patient, emitting a radiofrequency signal which is received and stored in an external device. The result is a video movie which records the trip of the capsule along the intestinal tract with a rate of two frames per second, and that can be easily downloaded into a PC with the camera software installed. This technique overcomes most of the drawbacks related to manometry: it is much less invasive, since the patient simply has to swallow the pill, which will be expelled in the normal cycle through defecation. Moreover, there is no need of hospitalization nor expert support through the process and the patient can lead an ordinary life, since the attached device is recording the video movie emitted by the camera in the capsule. Once the video is downloaded into the workstation, the expert visualizes the zone of interest and labels those frames where an intestinal event is detected, obtaining the temporal pattern of intestinal contractions which is to be used as a base for the intestinal motility dysfunction assessment. However, the visualization and precise interpretation of the capsule recordings is not straightforward, but it is time consuming and stressful, since the prevalence of contractions in video is very low (1:50 frames). Visualization time can vary depending on the frame rate used for this purpose, but generally speaking, it is common that for a visualization study of the whole small intestine the expert takes about 2 h, making it not feasible as a clinical routine.

In order to deal with these drawbacks, and make the analysis of the information provided by the capsule feasible for clinical routines, we focus our efforts on the design of a system for the automatic annotation of intestinal contractions in capsule video endoscopy. We provide the physicians with one of the fundamental measurements they need for the assessment of intestinal motility, namely, the position of the intestinal contractions in capsule endoscopy videos in an automatic way. This information allows the specialists to tackle the analysis of diverse parameters related to these findings, such as their number, their typology, their distribution and frequency, etc. This information is to be contextualized within a new and general framework for motility assessment, in which other parameters, such as the study of transit time, the analysis of quiet periods, the examination of intestinal content, etc., should play their important role, as recent research studies propose [5], [6]. The extent to what

Manuscript received November 04, 2008; revised January 13, 2009. First published May 05, 2009; current version published February 03, 2010. This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, H. U. Vall d'Hebron, Barcelona, Spain, and in part by Project TIN2006-15308-C02 and Project FIS-PI061290. The technology and methods embraced by this disclosure have been filed for patent protection. *Asterisk indicates corresponding author*.

*F. Vilariño iswith the Computer Vision Center and Computer Science Department, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain.

P. Spyridonos, J. Vitrià, and P. Radeva are with the Computer Vision Center and Computer Science Department, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain.

F. DeIorio and F. Azpiroz are with H.U. Vall d'Hebron and Universitat Autònoma de , 08193 Barcelona, Spain.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2009.2020753

all these issues could be potentially utilized as a basis for further pathological diagnosis, within a wider clinical framework, constitutes a novel and open field of clinical investigation. Consequently, for this research scenario to be successfully tackled, the position of the intestinal contractions in video appears as an essential pillar, although the current procedures used for this aim, based on manual labeling, have made its deployment unfeasible so far.

Several works have been reported in the fieldwork of classical endoscopy, addressing the support of automatic systems for the diagnosis of different pathologies, such as ulcer or cancer, with applications based on digital image analysis and processing. The state-of-the art in this area includes the use of numerical descriptors obtained from histogram analysis [7], co-occurrence matrices [8], homogeneity and hue of color images and binary pattern/intensity models for texture segmentation [9], texture spectrum analysis [10], [11], and wavelet features [12] among others. From all these works one can conclude that intensity, color and texture are relevant visual cues when processing endoscopic videos. In addition, all of these approaches consist of frame-based analyses which fit the nature of the applications for which they are designed.

In our opinion, one of the fundamental novelties of our contribution is to propose an integrated system that implements a sequence-based approach instead of a frame-based approach to analyze dynamic events such as intestinal contractions within an adequate framework, extract heterogeneous visual features and classify them in a robust way. Our proposal is based on a machine learning system which automatically learns and classifies contractions from a capsule video source, providing the expert with the portion of the video which is highly likely to contain the intestinal contractions. This yields a considerable reduction in visualization time, and the consequent reduction of stress, since most of the sequences to be analyzed are real contractions. One of the main advantages of our system is related to its ability to dynamically adapt itself to the different patterns of intestinal activity associated with intestinal contractions in a robust way. Furthermore, our implementation appears to be flexible and easily extensible, since the modular design of our approach will potentially allow the expert to include domain knowledge into the system by means of the addition of new modular stages.

The rest of the paper is organized as follows. In Section II, we develop the analysis and explanation of WCVE images, the different visual appearance of the different types of intestinal contractions and the difficulties inherent to their detection. In Section III, we describe the feature extraction and the classification procedure. Section IV presents our experimental results. Finally, we devote the last sections to the discussion of our system, and the exposition of our proposals for future research on intestinal motility with video capsule endoscopy.

II. Intestinal Contractions in Video Capsule Endoscopy

A. Basic Concepts on Gastrointestinal Motility

Muscle layers of the gut wall and their innervation are organized to provide the motor functions along the intestinal tract [1].

As a result of muscular stimulation, a contractile activity and tone are produced, and intestinal contractions are generated. From a physiological point of view, the different patterns of contractions can be gathered into two main categories, namely, *phasic* and *tonic*. The former are characterized by a sudden closing of the intestinal lumen, followed by a posterior opening, while the latter corresponds to a sequence in which the intestinal lumen remains completely closed for an undetermined span in time.

Both the type and the spatial frequency of intestinal contractions depend on the region of the gastrointestinal tract (stomach, small intestine, or colon), and the temporal patterns they present are different during fasting (before the ingestion of nutrients) and postprandial stages (after the ingestion of nutrients). A typical number around 400–900 phasic contractions can be found in a 4-h-long study, distributed in periods of different activity. The number and length of tonic contractions within the same period can present a high variability, ranging from a set of few long contractions to several tens of shorter ones. In this work, we restricted our field of research to the study of small intestinal motility assessment by means of the analysis of phasic contractions, in an attempt to provide a first approach to the global problem.

B. Intestinal Contractions Sequences With Capsule Endoscopy

Video capsule endoscopy images show a circular field of view, in which the intestinal wall and the intestinal lumen are shown. The phasic contraction is observed as a closing movement of the lumen which is spanned over a few frames. Fig. 1 shows a mosaic where the frames of a video have been deployed in a sequential way and different intestinal contractions have been outlined in a green rectangle. The maximal frequency of phasic contractions is known to be between 11 and 12 events per minute, spanning 4–5 s in average for the whole open-close-open cycle [1], [2], and the frame acquisition rate of cameras is typically set on 2 frames per second [13]. Thus, we adopted the convention of bounding the span of a phasic contraction in a sequence of nine frames. In the rest of the paper, we refer to a contraction sequence as a nine frames sequence, where the central frame is set to be the frame that will be labeled as a detected contraction. The intensity with which the intestinal walls concentrically contract is not the same for all the contractions, and sometimes the closing of the lumen is not complete. If the lumen is fully closed during a contractile activity, this kind of event is referred to as an *occlusive* contraction; in case the lumen closing is not total, the intestinal contraction is referred to as *nonocclusive*. Nonocclusive contractions are hard to detect by classical manometry, since the intestinal walls do not exert sufficient pressure to be detected. In video capsule endoscopy this kind of contractions is clearly shown, though. Fig. 1 renders out two clear examples of occlusive and nonocclusive contractions labeled as (a) and (b), respectively.

If the camera is focusing the lumen during the whole contraction, as pictured in Fig. 2(a), the contraction pattern appears clearly. However, these visual patterns present a high variability, which is strictly related to 1) the movement of the device along the gut and 2) the presence of intestinal content

1) Camera Movement: Since the capsule is freely moving into the gut, multiple changes in direction (namely, focusing

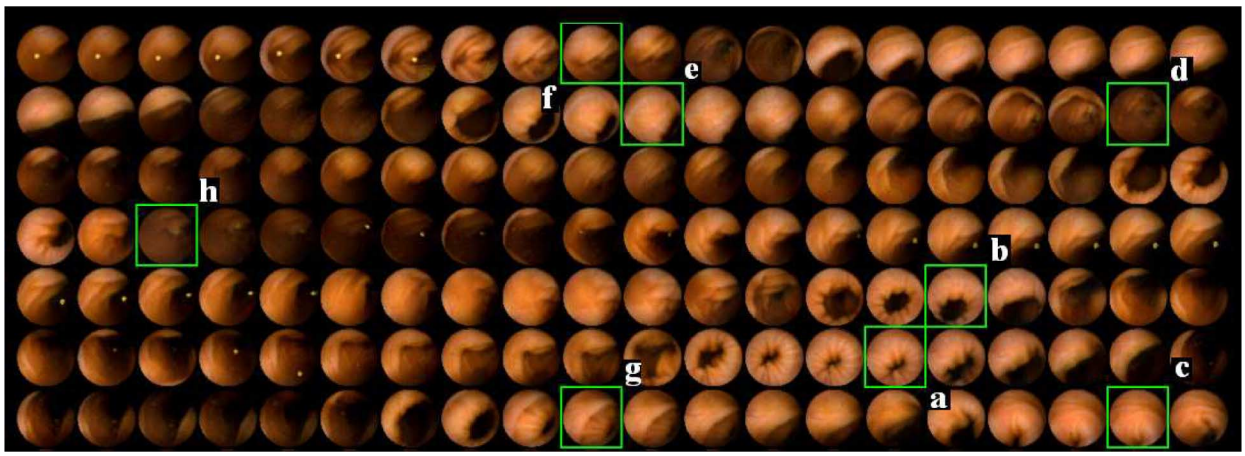


Fig. 1. An example of 140 consecutive frames, corresponding with 70 s, from a small intestine capsule endoscopy video study. The image shows a paradigmatic visual example of the dynamics involved in intestinal motility. The green rectangles surround different contraction frames labeled by the experts. In the occlusive and nonocclusive contractions labeled as (a) and (b), respectively, the camera focused the lumen during the whole contraction. In (c)–(h), the lumen was partially or totally missed in different parts of the sequence due to the free movement of the camera within the gut.

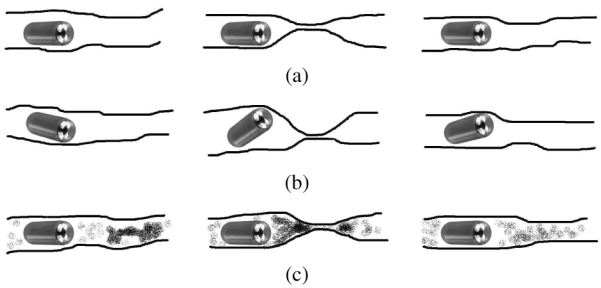


Fig. 2. Graphical representation in three steps (before, at the time of, and after the contraction event). (a) The paradigm of a complete phasic contraction. (b) The camera pointing towards the intestinal wall. (c) The presence of turbid liquid hindering the visualization.

the intestinal lumen or the lateral intestinal wall) and orientation (i.e., facing the proximal or distal parts of the tract) are performed. As a result, the camera is not always focusing the central part of the lumen, see Fig. 2(b) for a graphical representation. This yields incomplete contractions when the central frame shows the intestinal lumen but it is not centered in the image, lateral contractions when the first or the last part of the sequence is missed, but the central frame is present, and out-of-plane contractions when the central frame of the contraction is completely out of plane and the contraction event can only be deduced by the remaining part of the sequence, Fig. 1 shows different examples of these sequences, marked as (c)–(h).

2) Turbid Liquid: The good visibility of the intestinal lumen and wall is usually hindered by the presence of intestinal juices mixed up with remains of food, see Fig. 2(c). This is visualized as a semi-opaque turbid liquid in a wide range of colors from brown to yellow. In addition to this, the turbid liquid may be accompanied by the presence of bubbles and other artifacts related to the flux of the different liquids into the gut. As a result, the fluid is interposed between the camera and the intestinal contraction event, obstructing its right visualization. Fig. 3 shows two example sequences containing turbid liquid.

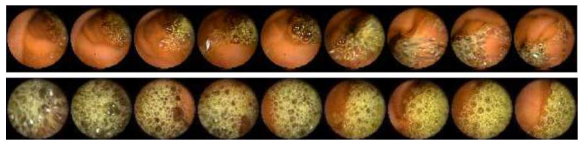


Fig. 3. Two sequences of intestinal contractions with presence of turbid liquid, which hinders partially (top) and completely (bottom) the correct visualization of the event.

III. Cascade System for the Detection of Intestinal Contractions in Capsule Endoscopy

Our proposal for the automatic detection of phasic contractions in capsule endoscopy video is deployed in a sequentially modular way, namely, a cascade, which is graphically explained in Fig. 4. Each part of the cascade receives as an input the output of the previous stage. The main input consists of the video frames, and the main output consists of the frames suggested as contractions. The rejected frames are distributed among three different stages. A first stage detecting dynamic patterns related to intestinal contractions, where most of the noncontraction frames are filtered. A second stage, removing nonvalid frames due to occlusions or a wrong orientation of the camera, as described in Section II-B. A final classification stage based on a support vector machine classifier (SVM) [14], where the final output defines the suggested contractions. The learning steps of each stage of the cascade involve a set of parameters P for tuning the classification performance. The turbid frames step and the final classification step consist of two SVMs trained with a data set which has been labeled from previous studies. Grey level images were used throughout, unless otherwise stated, by eliminating the hue and saturation information while retaining the luminance [15].

A. Stage 1: Detecting Dynamic Patterns

The aim of the first stage is to prefilter all the video frames according to the visual pattern of phasic contractions described

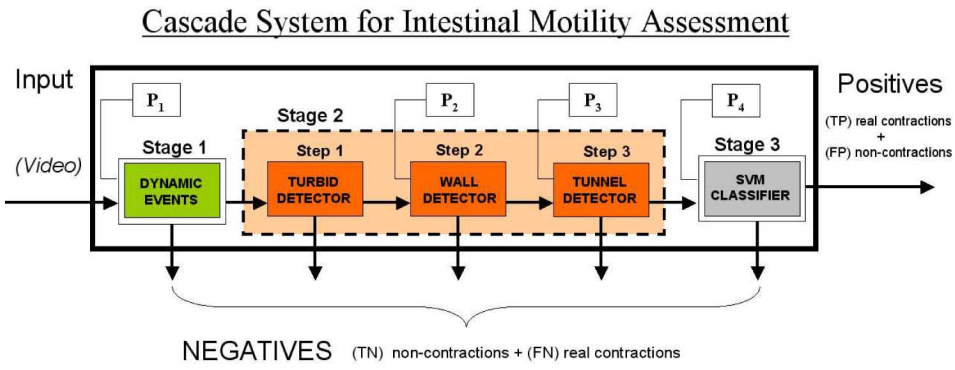


Fig. 4. Cascade system for intestinal motility assessment. The input is the video study and the output are the intestinal contraction frames suggested by the system. Each stage rejects sequences of noncontractions. The global performance can be tuned by the set of parameters P.

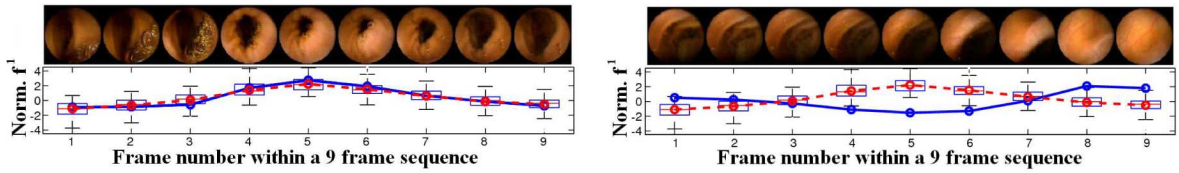


Fig. 5. Pattern of f^1 (solid blue line) for (a) one contraction and (b) a random sequence. The dashed red line corresponds to the averaged pattern of all the labeled intestinal contractions and the box plots define their lower quartile, median, and upper quartile (the whiskers show the extent of the rest of the data).

in Section II-B. This is implemented by means of the normalized intensity $f^1(n)$, defined in

$$f^1(n) = I_n - \frac{\sum_{i=-4}^{4} I_{n+i}}{9}. \tag{1}$$

For each frame n, we take into account the four previous and the four following frames. For each one of these nine frames, we calculate the overall intensity, $I_{n+i}, i = -4, \ldots, 4$, as the sum of the intensity values of its pixels. The final value of $f^1(n)$ represents a normalized intensity of the central frame within the nine frames sequence. Should the central frame n be darker than its neighbors, the difference in $f^1(n)$ would tend to be negative, and vice-versa. For the specific visual pattern of phasic contractions, the presence of an open lumen in the previous and following frames makes the central frame of a sequence of an intestinal contraction have a higher value of intensity than its neighbors. Thus, $f^1(n)$ is designed in order to present a symmetric concave pattern when the central frame of a nine frames sequence corresponds with an intestinal contraction, and a smooth pattern with a different shape otherwise. A plot of $f^1(n)$ for (a) one contraction sequence and (b) one arbitrary sequence of nine frames is pictured in Fig. 5. A discriminant function g^1 is defined in terms of f^1 and the threshold P_1 as

$$g^1(n) = \begin{cases} 1, & \text{if } f^1(n) > P_1 \\ -1, & \text{otherwise} \end{cases} \tag{2}$$

rejecting all the frames for which $g^1(n) = -1$.

B. Stage 2: Rejection of Turbid, Wall and Tunnel Frames

The aim of stage two is to reject the turbid frames and those frames where the camera is focusing on the intestinal wall. In addition to this, those frames where the lumen appears static for a long sequence of time are rejected as well, since these frames do not carry motility information.

1) Turbid Frames: The presence of turbid liquid is characterized by color, which is usually in a range from brown to yellow, mainly centered around green. For each frame, a color quantization is performed in the following way: each RGB component of the image is quantized into five bins in a linear way, spanning all the range of the color component. This yields a 125-bin histogram (5^3), which is used as a feature vector. In order to train the turbid classifier, a data set of 2000 turbid frames from seven specific studies was randomly chosen among those video regions that the experts labeled as "region showing intestinal content." The nonturbid frames were randomly chosen among the remaining sequences using undersampling, i.e., taking a random sample of nonturbid equal to the number of turbid. The seven selected studies represented paradigmatic examples of intestinal content appearance regarding the experts' opinion.

The SVM has two main generalized parameters to be set: the kernel type and the kernel parameter. We used a radial basis function kernel and a $\gamma = 0.1$. Equation (3) shows the mathematical representation of the radial basis function kernel

$$K_{\text{rbf}}(x, x_i) = \exp \frac{-|x - x_i|^2}{2\sigma^2}, \quad \gamma = 1/(2\sigma^2). \tag{3}$$

The choice of the kernel and the γ parameter was obtained in an empirical way with an exhaustive analysis, using as a reference for validation the visual assessment of the experts. The SVM classifies all the video frames into turbid and nonturbid. In order to incorporate the dynamic characteristics of the intestinal contractions as performed in the first stage, we adopted as a final criterion the rejection of those frames with more than four neighbors labeled as turbid frames within the nine frames sequence (the number of four frames was strictly based on the

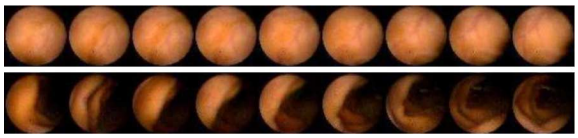


Fig. 6. Paradigmatic sequences of wall (top), and tunnel frames (bottom). This type of sequences lack information regarding contractions, and the system detects and rejects them as system negatives in the second stage.

experts' assessment), letting the remaining frames pass to the next step. By automatizing the intestinal contents detection in this way, the specific intervention of the expert, which was suggested as a plausible alternative in previous contributions [16], can be avoided.

2) Wall and Tunnel Frames: Wall and tunnel frames are those due to the stable orientation of the camera towards the intestinal wall and lumen respectively, for a span of time where no motility activity is present. Fig. 6 shows two examples of wall (top) and tunnel sequences (bottom). Both wall and tunnel frames were described by means of the sum of the area of the lumen throughout the sequence of nine frames.

In order to estimate the area of the lumen in each frame, a Laplacian of Gaussian filter (LoG) was applied [15]. The LoG filter is a second order symmetric filter with a tuning parameter σ which plays the role of a scale parameter. The output of the LoG is high when a dark spot is found, providing a higher response the closer the diameter of the spot is fitting the span of the Gaussian defined by σ, and the higher the contrast is between the dark spot and its bounds. The value of σ was fixed to $\sigma_{\text{lum}} = 3$, the minimum size of the lumen in the central frame of a contraction sequence (this was straightforward to obtain after testing different values of several contraction sequences a quarter of the original size). The whole procedure is graphically deployed in Fig. 7 (this figure is shown as appearing in [17]). For each sequence of nine frames, the LoG filter is applied (second row). Following, a greater-than-zero threshold is performed to the filter output, which provides a binary image with one or more connected components or blobs (third row). In case that only one blob is obtained, its area $f^2(n)$ is taken as the lumen area from

$$\begin{aligned} L_n &= \text{LoG}_{\sigma_{\text{lum}}}(I_n) \\ \text{Blob}_n &= F(x,y) \\ &= \begin{cases} 1, & L_n(x,y) > 0 \\ 0, & \text{otherwise} \end{cases} \\ f^2(n) &= \sum_{\text{pixels}} \text{Blob}_n. \end{aligned} \tag{4}$$

In case that several blobs are obtained, the one with the highest global response of the filter (i.e., presumed to be the one with the highest contrast and best fitting in size) is selected based on the function

$$f^3(n) = \sum_{\text{pixels}} (\text{Blob}_n \star L_n) \tag{5}$$

where $\star$ represents the element-by-element product of the two image matrices. The last row in Fig. 7 shows an example of the lumen segmentations obtained with this procedure. The subsequent characterization of wall and tunnel frames is straightforward: the system classifies a frame as a wall frame if the sum of the lumen area throughout the nine frames sequence is less than a certain threshold, while the same frame is classified as a tunnel frame if the sum of the lumen area throughout the nine frames sequence is greater than certain threshold. These two values form the system tuning parameters P_2 and P_3.

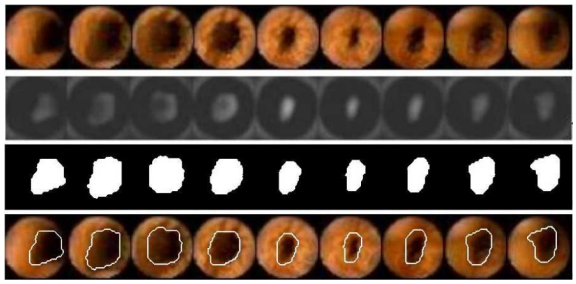


Fig. 7. Original image, LoG filter response, binary blob and final lumen segmentation for the nine frames of an intestinal contraction sequence.

C. Stage 3: Final Classifier

The last stage of our approach consists of a SVM classifier, which receives as an input the output of the second stage of the cascade, with an imbalance ratio which has been typically reduced from 1:50 to 1:5 frames. The output of the SVM consists of frames suggested to the specialist as the candidates for intestinal contractions. The choice of the SVM is underpinned by its robust mathematical background, being one of the most widely used classification techniques, with a remarkable success in multiple and diverse applications through the recent years [18]. A radial basis function kernel was used with a $\gamma = 0.01$. The γ parameter controls the operation point of the SVM, and corresponds to the fourth tuning parameter of the system, P_4.

In order to characterize the intestinal contractions, a set of 33 features were computed. These features included the following.

1) The three previously described functions f^1, f^2, and f^3.
2) ix textural features obtained from the co-occurrence matrix [19] of the original gray-level image after gray level normalization: energy, entropy, inverse differential moment, shade, inertia, and promenance [7], [20].
3) Eighteen features obtained from the histogram of the Rotation Invariant Uniform Local Binary Units operator (LBPriu2), applied in a circular symmetric neighborhood using $P = 16$ different orientations and a radius $R = 2$ [21]. The results reported in the bibliography for textured patterns point out these values as a fine compromise between the number of features obtained (18) and the performance in textural discrimination.
4) Finally, six statistical features: standard deviation, skewness, and kurtosis were calculated on the normalized gray level image and the local binary pattern vector. The use of these textural features, in addition to the normalized intensity, blob area and blob contrast, is underpinned by the fact that during the contraction, the folds and wrinkles of the intestinal wall tend to show certain patterns which endow these images with a typical texture carrying discriminative power from a visual point of view. In this sense, textural

information can also be useful for the explicit characterization of certain types of turbid liquid, such as the bubbles shown in Fig. 3, which might be tackled by means of texture analysis using Gabor filters [22]

A feature vector was constructed taking into account the previous and following four frames, so that a final $33 \times 9 = 297$ dimensional feature vector was assigned to each frame. In order to address the high dimensionality of this feature space, a sequential forward feature selection method was applied [23].

IV. Results

Our experimental tests were performed using 10 capsule studies obtained from 10 different volunteers. These volunteers were aged between 22 and 33, presented no evidence of gastrointestinal pathologies and were asked to abstain from eating and drinking for 12 h prior to the start of the studies, which were conducted at the Digestive Diseases Department of the General Hospital de la Vall D'Hebron in Barcelona, Spain. The endoscopic capsules used were developed by Given Imaging, Ltd.[1] The capsules dimensions were 11×26 mm, contained six light emitting diodes, a lens, a color camera chip, two batteries with a mean life of about 6 h, a radio-frequency transmitter, and an antenna. The capsule acquisition rate was two frames per second with a resolution of $256 \times 256 \times$ 24-bit. For each study, one expert visualized the whole video and labeled all the frames showing intestinal contractions between the first post-duodenal and the first cecum images. These findings were used as the gold standard for testing our system. The parameter vector P was set to the initial value $P^0 = \{P_1 = 0, P_2 = 50, P_3 = 650, P_4 = 0.01\}$ using an exhaustive search in the following way. For P_1, we looked for value which let 95% of contractions pass to the second stage. This value guarantees that most of the contractions pass to the second stage, while a substantial reduction in the number of frames to be analyzed is achieved. The 95% threshold was chosen after several tests and the visual analysis of the filtered sequences. For P_2 and P_3, we created a first qualitative validation step by means of visual mosaics, in which all the video frames were sequentially represented and those frames selected as wall (or tunnel) frames were surrounded by a color square, in the same way as in Fig. 1. The fine tuning of the P_2 and P_3 values was performed by means of the manual counting of the true positives and false positives of those mosaics with the best qualitative results. Finally, the selected P_2 and P_3 values corresponded to the parameter values showing the best results in terms of precision, which in both cases were around 92%.

We followed the leave-one-out strategy for the experimental design: one video was separated for testing while the nine remaining videos were used for training the SVM classifiers using undersampling, i.e., taking a random sampling of noncontractions equal to the number of contractions. Feature vectors of 54 elements were made up for each sample. These vectors consisted of nine features obtained from the feature selection procedure pointed out in Section III-C, which consisted of f_1, f_2, f_3, entropy, inertia, and inverse differential moment, applied to the sequence of nine frames defined by each sample.

[1]http://www.givenimaging.com

A. Positive and Negative Detection Rates

In order to accomplish a detailed performance analysis of our approach, we provide the study of each separate stage in the cascade. Tables I, II, and III show the performance results of stages 1, 2, and 3, respectively. For each stage, certain number of frames arrive at the input (column **Frames**), containing a number of intestinal contractions labeled by the expert (column **Findings**). The quotient *Noncontraction frames/Number of findings* represents the imbalance rate at the input of the stage (column **Imbalance Rate**). The output columns consist of the number and the percentage of frames and findings passing to the next stage, and the resulting imbalance rate. In addition to this, the rate of missing findings, i.e., findings which were wrongly filtered as noncontractions, and the rate of noncontraction frames, i.e., those noncontractions passing to the next stage of the cascade, is provided.

- **Stage 1:** Table I shows that the overall number of frames at the output of stage one is about 11% the input, i.e., the system rejects 89% of the frames in this stage. But despite this high reduction in the number of frames, almost every finding was kept (97%), i.e., just around 3% of the findings were wrongly rejected as noncontractions. At the output of stage one, the imbalance ratio was reduced about 10 times, from 60.3 to 5.9.
- **Stage 2:** At the output of this stage about 28% of the frames were rejected, keeping 96% of the findings provided by stage one. The overall imbalance rate was reduced to 3.6. In addition to this, the sum of the loss of findings, taking into account both stage one and stage two, set the rate of detected contractions at the output of stage two about 93%, as can be observed in the column **%Findings in video** in Table II. As in the previous stage, the reduction of the imbalance rate is substantial, while the loss in contractions appears to be reasonable (only 7% of all existing contractions in video).
- **Stage 3:** The output of stage three is at the same time the output of the system. Thus, we can analyze the output of stage three both in terms of stage performance and global performance. The stage performance is pictured in Table III, while the global performance analysis is deployed in Table IV and will be the object of analysis in the next paragraphs. Table III shows that the SVM classifier yields a reduction of about 71% in the number of frames at the output, keeping 75% of the contractions provided by stage two. Moreover, the imbalance rate of the final data set is reduced to 0.7.

Finally, a global system assessment must provide: how many of the existing contractions our system is able to detect (*sensitivity*), how many of the existing noncontractions our system is able to reject (*specificity*), and which the ratio between real contractions detected and the total number of sequences at the output of the system is (*precision*). In addition to the former, a false alarm ratio (FAR) between the false contractions at the output of the system and the existing contractions in the video provides the expert with useful information. A rigorous definition of these quantities in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) can be stated as follows.

TABLE I
PERFORMANCE ANALYSIS FOR THE FIRST STAGE OF THE CASCADE

	INPUT			OUTPUT							
Study	Frames	Findings	Imb. Rate	Frames (%)		Findings (%)		Imb. Rate	Missed Findings (%)		Non-Cont. Frames (%)	
Video 1	29444	747	38.4	3192	10.84%	720	96.39%	3.4	27	3.61%	2472	77.44%
Video 2	28803	529	53.4	3027	10.51%	502	94.90%	5.0	27	5.10%	2525	83.42%
Video 3	27816	575	47.4	3185	11.45%	561	97.57%	4.7	14	2.43%	2624	82.39%
Video 4	38885	733	52.0	4025	10.35%	717	97.82%	4.6	16	2.18%	3308	82.19%
Video 5	17619	356	48.5	1849	10.49%	349	98.03%	4.3	7	1.97%	1500	81.12%
Video 6	27360	476	56.5	2943	10.76%	459	96.43%	5.4	17	3.57%	2484	84.40%
Video 7	27176	918	28.6	2903	10.68%	890	96.95%	2.3	28	3.05%	2013	69.34%
Video 8	12620	150	83.1	1366	10.82%	143	95.33%	8.6	7	4.67%	1223	89.53%
Video 9	25994	206	125.2	2953	11.36%	198	96.12%	13.9	8	3.88%	2755	93.29%
Video 10	27967	397	69.4	2948	10.54%	385	96.98%	6.7	12	3.02%	2563	86.94%
Avg:			60.3		10.78%		96.65%	5.9		3.35%		83.01%

TABLE II
PERFORMANCE ANALYSIS FOR THE SECOND STAGE OF THE CASCADE

	OUTPUT								
Study	Frames (%)		Findings (%)		Imb. Rate	Missed Findings (%)		Non-Cont. Frames (%)		% Findings in Video
Video 1	2774	86.90%	697	96.81%	3.0	23	3.19%	2077	74.87%	93.31%
Video 2	2346	77.50%	474	94.42%	3.9	28	5.58%	1872	79.80%	89.60%
Video 3	2623	82.35%	548	97.68%	3.8	13	2.32%	2075	79.11%	95.30%
Video 4	3170	78.76%	673	93.86%	3.7	44	6.14%	2497	78.77%	91.81%
Video 5	1740	94.10%	341	97.71%	4.1	8	2.29%	1399	80.40%	95.79%
Video 6	2288	77.74%	453	98.69%	4.1	6	1.31%	1835	80.20%	95.17%
Video 7	2692	92.73%	869	97.64%	2.1	21	2.36%	1823	67.72%	94.66%
Video 8	804	58.86%	134	93.71%	5.0	9	6.29%	670	83.33%	89.33%
Video 9	678	22.96%	184	92.93%	2.7	14	7.07%	494	72.86%	89.32%
Video 10	1538	52.17%	363	94.29%	3.2	22	5.71%	1175	76.40%	91.44%
Avg:		72.41%		95.77%	3.6		4.23%		77.35%	92.57%

TABLE III
PERFORMANCE ANALYSIS FOR THE THIRD STAGE OF THE CASCADE

	OUTPUT								
Study	Frames (%)		Findings (%)		Imb. Rate	Missed Findings (%)		Non-Cont. Frames (%)		% Findings in Video
Video 1	904	32.59%	595	85.37%	0.5	102	14.63%	309	34.18%	79.65%
Video 2	607	25.87%	343	72.36%	0.8	131	27.64%	264	43.49%	64.84%
Video 3	646	24.63%	405	73.91%	0.6	143	26.09%	241	37.31%	70.43%
Video 4	981	30.95%	547	81.28%	0.8	126	18.72%	434	44.24%	74.62%
Video 5	433	24.89%	266	78.01%	0.6	75	21.99%	167	38.57%	74.72%
Video 6	768	33.57%	339	74.83%	1.3	114	25.17%	429	55.86%	71.22%
Video 7	835	31.02%	603	69.39%	0.4	266	30.61%	232	27.78%	65.69%
Video 8	189	23.51%	111	82.84%	0.7	23	17.16%	78	41.27%	74.00%
Video 9	228	33.63%	130	70.65%	0.9	54	29.35%	98	42.11%	63.11%
Video 10	363	23.60%	248	68.32%	0.5	115	31.68%	115	31.68%	62.47%
Avg:		28.42%		75.70%	0.7		24.30%		39.65%	70.08%

Table IV summarizes the performance results of the cascade system. Our approach achieves an overall sensitivity of 70.08%, peaking 80% for the study referred to as *Video 1*—the high overall specificity value of 99.59% is typical of imbalanced problems. However, FAR and precision carry insightful information about what the output is like. The resulting precision value of 60.26% tells us that 6 out of 10 frames in the output correspond to true findings. FAR is similar, but in terms of noise (the bigger the FAR, the larger the number of false positives), and normalized by the number of existing contractions. For different videos providing an output with a fixed precision, those with the highest number of findings in video will have lower FAR. In this sense, a FAR value of one tells us that we have obtained as many false positives as existing contractions in video. FAR and precision values are usually related, and Table IV shows that the peaks of performance for both measures are found in the same two studies (*Video 6* and *Video 7*, outlined in bold type).

Table V shows the specific detection rate of the intestinal contractions corresponding to the complete occlusive and complete nonocclusive patterns characterized in Section II-B, i.e., those sequences in which the camera is pointing to the center of the lumen for the nine frames, showing either an occlusive or a nonocclusive contraction. These sequences represent the clearest patterns of contractions and their detection rates, which should be very high, provide relevant information regarding the robustness of the system, as well as potential indicators for a further clinical assessments. The comparison between the de-

TABLE IV
GLOBAL SYSTEM PERFORMANCE

Study	Sensitivity		Specificity		FAR		Precision	
Video 1	595/747	**79.65%**	29135/29444	98.95%	309/747	41.37%	595/904	65.82%
Video 2	343/529	64.84%	28539/28803	99.01%	264/529	49.90%	343/607	56.51%
Video 3	405/575	70.44%	27575/27816	99.13%	241/575	41.91%	405/646	62.69%
Video 4	547/733	74.65%	38451/38885	98.88%	434/733	59.21%	547/981	55.76%
Video 5	266/356	74.72%	17452/17619	99.05%	167/356	46.91%	266/433	61.43%
Video 6	339/476	71.22%	26931/27360	**98.43%**	429/476	**90.13%**	339/768	**44.14%**
Video 7	603/918	65.69%	26944/27176	99.15%	232/918	**25.27%**	603/835	**72.22%**
Video 8	111/150	74.00%	12542/12620	99.38%	78/150	52.00%	111/189	58.73%
Video 9	130/206	63.11%	25888/25994	99.59%	106/206	51.45%	130/228	57.02%
Video 10	248/397	**62.46%**	27852/27967	**99.59%**	115/397	28.96%	248/363	68.32%
Avg:		**70.08(±5.81)%**		**99.12(±0.38)%**		**48.71(±17.92)%**		**60.26(±7.84)%**

TABLE V
DETECTION RATE FOR PATTERNS OF COMPLETE OCCLUSIVE + COMPLETE NONOCCLUSIVE CONTRACTIONS

	Occlusive				Non-occlusive				Occlusive + Non-occlusive	
	Proportion		Sensitivity		Proportion		Sensitivity		Sensitivity	
Video 1	125/747	16.7%	119/125	95.2%	85/747	11.4%	76/85	89.4%	195/210	92.8%
Video 2	38/529	7.1%	36/38	94.7%	34/529	6.4%	27/34	79.4%	63/72	87.5%
Video 3	59/575	10.2%	56/59	94.9%	23/575	4.0%	17/23	73.9%	73/82	89.0%
Video 4	72/733	9.8%	66/72	91.7%	176/733	24.0%	140/176	79.5%	206/248	83.0%
Video 5	32/356	8.9%	29/32	90.6%	42/356	11.8%	40/42	95.2%	69/74	93.2%
Video 6	42/476	8.8%	37/42	88.1%	44/476	9.2%	35/44	79.5%	72/86	83.7%
Video 7	178/918	19.3%	129/178	72.5%	61/918	6.7%	49/61	80.3%	178/239	74.5%
Video 8	12/150	8.0%	8/12	66.7%	12/150	8.0%	11/12	91.7%	19/24	79.1%
Video 9	22/206	10.7%	19/22	86.4%	15/206	7.3%	13/15	86.7%	32/37	86.5%
Video 10	45/397	11.3%	39/45	86.7%	28/397	7.0%	26/28	92.9%	65/73	89.0%
Avg:			**86.7(±9.7)%**				**84.8(±7.2)%**		**85.8(±5.9)%**	

Sensitivity	Specificity	FAR	Precision
$\frac{TP}{TP+FN}$	$\frac{TN}{TN+FP}$	$\frac{FP}{TP+FN}$	$\frac{TP}{TP+FP}$

tection rate for this type of contractions and the specific case of occlusive + nonocclusive set is plotted in Fig. 8. A parallel behavior between the detection rate of occlusive and nonocclusive patterns and the generic contractions can be observed, except for the case of *Video 8*, where a slight decrease in sensitivity is observed for the occlusive case (notice the low number of occlusive contractions found in this video, which may likely produce an impact in the obtained value). An overall sensitivity of 85.8% was achieved for the occlusive+nonocclusive set, peaking at 93.2% for *Video 5*.

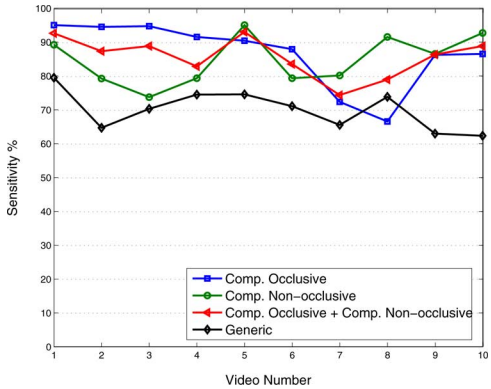


Fig. 8. High detection rates (up to 95%) are obtained for complete occlusive and complete nonocclusive contractions. Detection rates for the generic case are plotted in black for comparison purposes.

B. Assessment of the Density of Contractions

The main aim of this study is focused on the assessment of the ability of our system to describe the pattern of the density of intestinal contractions for each video. In order to assess this, we must answer two main questions: 1) which is the divergence shown in the labeling of the same video by different specialists, and 2) whether our system performs a labeling which is similar to that provided by the specialists. The former is linked to the interobserver variability, while the latter is linked to the divergence of the labeling of our system in terms of interobserver variability.

The first step is the assessment of the variability of the labeling between-experts, both in terms of absolute labeling and in terms of density of contractions. For this aim, two different specialists labeled all the videos. For the absolute labeling assessment, we calculated the % of labels of Expert 1 that were present within the Expert 2 labels set, and vice versa. Thus, for each video we had two measures of coincidence of labels; in the ideal case, both experts would provide 100% of coincident labels. The results plotted in Fig. 9 show a high agreement between experts in terms of absolute labeling. Fig. 9(a) plots the histogram of studies regarding the coincidence of labels, and Fig. 9(b) renders the box plot showing the distribution of coincident labeling in quartiles, with half of the studies over 90% of coincident labeling, three out of four videos over 83%, and only two outliers in 20 measures showing coincidence values around 65%. Regarding the interobserver assessment in terms of density of contractions, we performed the following analysis. For each video, a histogram of the intestinal contractions was created, grouping all the labeled contractions in bins spanning 3 min (360 video

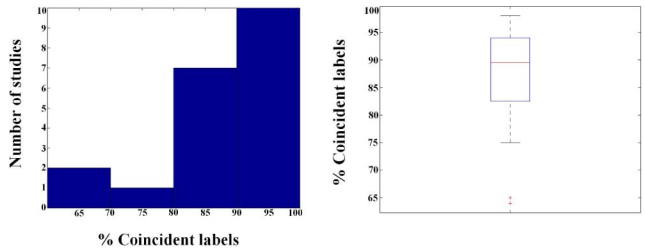


Fig. 9. Interobserver agreement in terms of coincident labels. (a) Histogram of studies regarding coincident labels. (b) Box plot showing median value around 90% of coincident labels and three out of four studies over 83% of agreement.

frames). The first postdoudenal and first cecal images, namely, the region of analysis, were fixed by the first specialist. A Kolmogorov–Smirnov hypothesis test [24] was performed. The null hypothesis for this test is that X_1 and X_2 have the same continuous distribution, where X_1 and X_2 are the labeling patterns provided by the experts 1 and 2, respectively. The alternative hypothesis is that they have different continuous distributions. We reject the null hypothesis if the test is significant at the 95% level. The test resulted negative for all the 10 videos, providing a mean p-value of $0.83(\pm 0.21)$. This result yields the conclusion that both specialists obtain similar patterns in terms of density of labels.

In order to answer the second question, we performed the same hypothesis test, including the system output. The null hypothesis for this test is that X_1 and X_S have the same continuous distribution, where X_1 is the pattern provided by the expert 1 and X_S is the pattern provided by the system. We repeated the test by substituting X_1 by X_2, and calculated the p-value as the mean of both tests, only if the test resulted negative for both the experts. This test resulted negative for 9 out of 10 videos at the 95% level, showing a mean p-value of $0.21(\pm 0.23)$. Fig. 10 shows the bar charts of the histograms for specialist 1, system output, and the cumulative density function (cdf), on which the Kolmogorov–Smirnov test is based, for (a) a negative test (*Video 3*) and (b) a positive test (*Video 5*). It can be seen that the only video rejected by the test, which showed a p-value of 0.0004, presents a systematic over-detection of contractions, which is distributed in a similar amount along the video. From this result we can conclude that the patterns obtained by the experts and the patterns obtained by our approach are similar in terms of density of labels, observing a unique case of divergence for Video 5, which presents a systematic increase in the labeling counts. One likely reason for this divergence can be founded in the presence of an unusual high number of long sustained contractions for this video, whose analysis is out of the scope of this paper since the paradigm of sustained contractions is completely new, and a rigorous clinical validation should be needed before any further conclusion might be drawn.

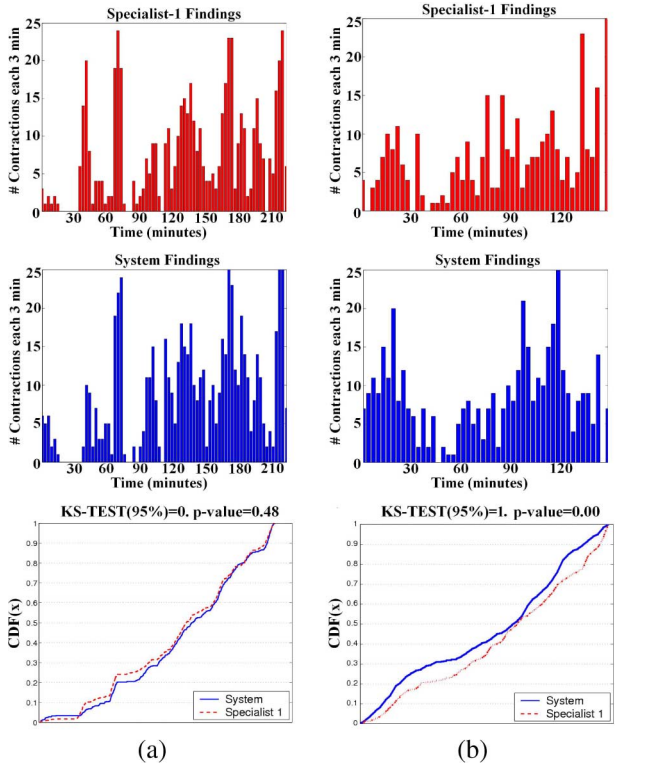


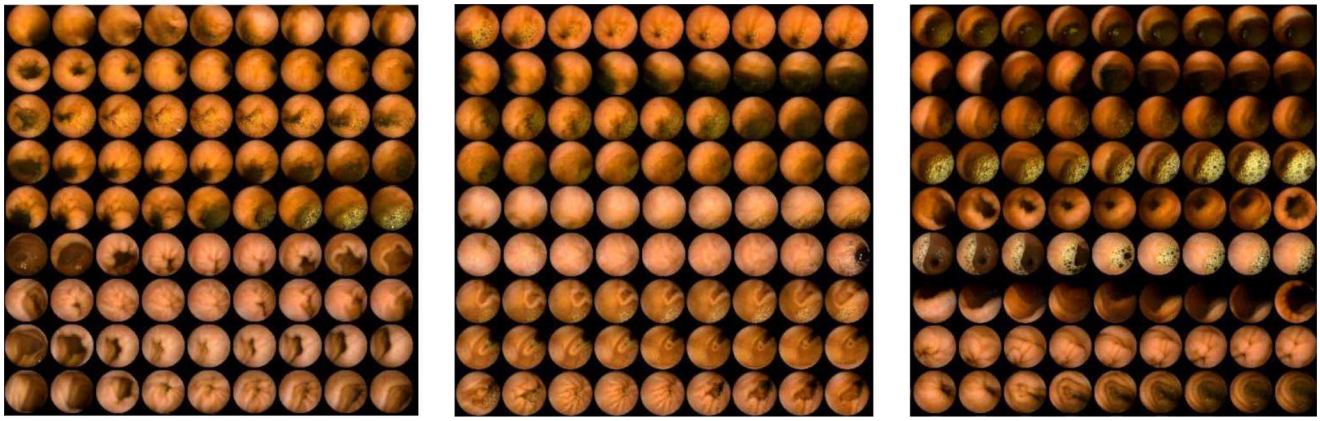
Fig. 10. Human and system labelling histograms of intestinal contractions for (a) Video 3 and (b) Video 5. Each bin contains the number of contractions each 3 min. The last row shows the cdf and the KS-test result.

V. Discussion

A. Cascade Classification System

Many authors have applied diverse strategies in order to tackle the impact that the imbalance ratio has in the performance of classification, involving stratified sampling, cost-sensitive approaches, different implementations of decision trees and bagging, and the use of several metrics for performance measurement, mainly [25]–[29]. In our strategy, each stage is tuned to prune as many noncontraction frames as possible, trying to minimize the loss of true positives, and achieving in this way an effective reduction in the imbalance ratio of the data. The last stage of the cascade, consisting of the SVM trained by means of under-sampling, is to face a classification problem with an imbalance ratio about 1:5, in contrast with the 1:50 at the input of the system. This reduction in the imbalance ratio is shown to be an effective way of tackling the problem of classification in this kind of scenario. In addition to this, one more important feature must be outlined: the modular shape of the system lets the expert identify new targets in the video analysis procedure, providing the chance to easily include them as new filter stages, and adding domain knowledge to the system in a natural and flexible way.

One of the main considerations taken into account for the selection of the SVM classifier was its sensitivity to the skewed distribution of the data sets. The SVM takes into account samples which are close to the decision boundaries, namely, the support vectors, and it tends to be unaffected by samples lying far away. Additionally to the former, stratified sampling techniques have been proved to be efficient in the improvement of performance of several classifiers, including SVMs [30]. In our approach, several methods of sampling were tested, and under-sampling yielded the highest reliability (a detailed analysis and discussion about the design of these experiments can be



(a) (b) (c)

Fig. 11. Some example sequences provided by the system. (a) Correctly detected contractions. (b) Nondetected contractions (false negatives). (c) Sequences which had not been labeled by the experts, but detected as contractions (false positives).

found in the [31] and [32]). Regarding the computation time, the feature extraction and training procedure were accomplished in about 3 h, while the final classification was performed in less than 1 min using a standard PC (2.4 GHz processor).

B. Qualitative Analysis of the Classification Results

Fig. 11 shows a set of paradigmatic examples for (a) detected contractions (true positives), (b) not detected contractions (false negatives), and (c) sequences which had not been previously labeled by the experts, but which our system classified as contractions (false positives). The detected contractions [see Fig. 11(a)] basically correspond to the paradigm of phasic contractions described in Section II-B. It must be noticed that the presence of turbid liquid in some frames does not result in a rejection of this sequence, because only the manifest turbid sequences are rejected.

The missed contractions [see Fig. 11(b)] share some common features. Firstly, the open lumen is not always present at the beginning and the end of the sequence. It must be noticed as well that the motion impression that the expert perceives during the video visualization is not present in the deployed sequence of frames. In this sense, we performed some tests which consisted of showing the experts a set of paradigmatic sequences containing doubtful contractions both by visualizing them in the video at a visualization rate of two frames per second, and showing the same sequences deploying the nine frames as in Fig. 11. We found that the experts usually labeled a higher number of contractions during the video visualization than looking at the deployed sequence. This fact drives us to think that the motility characterization should be performed in more subtle detail, in order to detect the apparently slight changes in some sequences shown in Fig. 11(b), but which actually seem to be clear for the expert during the visualization process. Regarding those contractions missed by the rejection of turbid frames, we must highlight that although the percentage of frames showing intestinal juices in video widely ranges from 15% to 45%, about 2%–15% of missed contractions were estimated to be due to this effect.

Finally, the false positives analysis supplies very interesting results. On the one hand, our system shows its ability to detect real contractions which the experts did not get to label. An example of these sequences is rendered in the fifth row of Fig. 11(c). A rough study over the false positives of the ten analyzed videos showed that about 10% of the false positives consisted of this kind of sequences. On the other hand, the sequences shown in Fig. 11(c) display the inherent difficulty related to the high variability of patterns present at the output of the system: the lateral movement of the camera while focusing the lumen which can be confused with the pattern of its contraction, the differences in illumination creating shadows which can be confused with the lumen, the multiple patterns of wrinkles which can provide a high response to the lumen detector, and the residual presence of patterns of turbid liquid, share the main responsibility in the false positives. We suggest that many of these problems may be tackled by a deeper study about the textural information provided by the lumen, both in the relaxed stage and the contraction activity.

Regarding the assessment of the absolute labelling, we would like to highlight a final comment. Providing that the length of a phasic contraction is expected to be less or equal to nine frames, a finding can be assumed to be detected if the system label is set within a window of $+/-2$ frames, which is the minimal jitter physiologically supported. This is the strategy employed for the results of absolute labelling shown in Section IV. The use of only one expert as a ground truth is justified since the interobserver variability study provides statistical evidence of the equivalence of the labelling of different experts. Nevertheless, we admit that the use of an averaged result from multiple experts would grant an asymptotical outcome, provided that those studies are available. For the sake of clarity, and without any statistical weight, the obtained results can be compared with those drawn from a uniform distribution from which a number of findings is randomly sampled. This number of findings can be obtained from a normal distribution with the mean and variance empirically obtained from the experts' findings in the different studies. In this case, our experiments show a ground under 20% sensitivity, 20% precision and 80% FAR, which would be the result associated to a random labelling under the hypothesis of

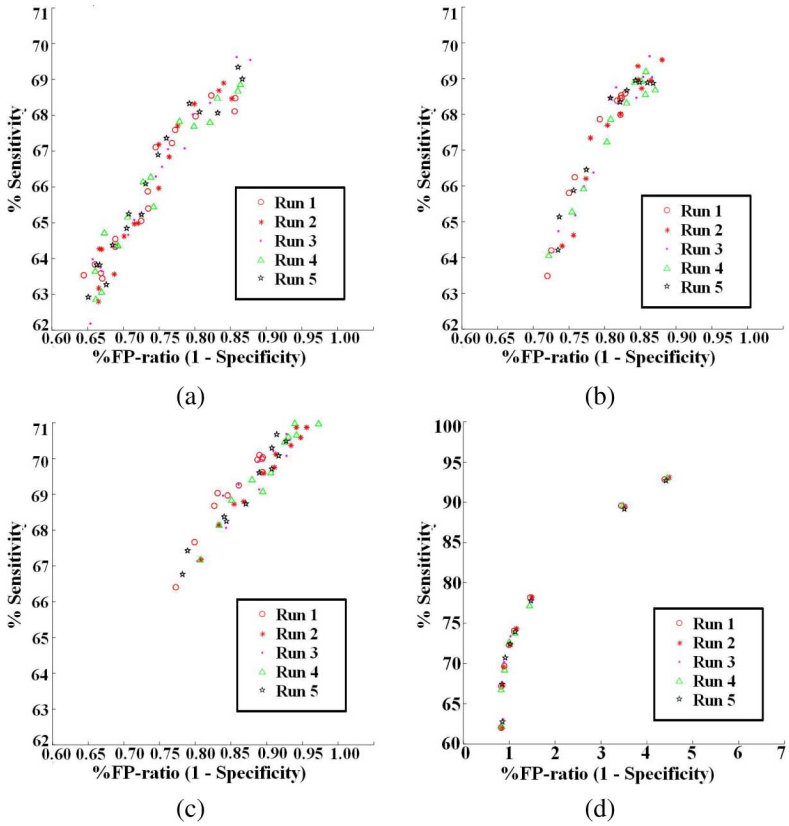


Fig. 12. Operation points from the ROC curve segments using the forward parameter selection procedure for (a) P_1, (b) P_2, (c) P_3, and (d) P_4. Each symbol represents each of the different five runs. The different points of each symbol represent the different performance pairs of sensitivity versus FP-ratio.

a uniform distribution of findings. In addition to the former, the proposed strategy allows the thorough numerical analysis provided in Tables I–V, which conveys specific information about each step in the system.

C. Validation of the System Operation Point

Providing that the parameter set $P^0 = \{P_1^0, P_2^0, P_3^0, P_4^0\}$ was obtained in an exhaustive empirical search, we must assess that P^0 does correspond with an optimal operation point, in terms of system performance. In order to assess this issue, we proceeded with a forward-propagation algorithm for parameter selection, which is deployed in detail in Appendix I. Both the fast forward algorithm and the performance criteria chosen are justified by the following reasons. On the one hand, it must be taken into account that when we vary one parameter, we must rerun all the system for each of the nine videos used for training, and we must apply a new leave-one-out strategy for each of them, training their classifiers using the eight remaining videos. On the other hand, a performance criterion function based on the global classification error does not appear to be a reliable metric in this context. In order to tackle the issue of performance assessment in imbalanced problems, several authors have proposed different solutions, including the use of the g-metric, the F-metric, and others [30]. Among the clinical community, the use of a trade-off between sensitivity and some other measure is widely extended. For our case, we requested our experts to provide us with the reference of the performance threshold which should be used in the trade-off function, arriving at a final compromise around sensitivity $= 70\%$.

Figs. 12 and 13 represent a zoom into the region of the receiver operating characteristic (ROC) curve [33] in which our system is operating. Regarding FP-ratio, the system performance ranges between the intervals [0.6%, 0.9%] for Fig. 12(a) and (b) and Fig. 13(a), and between [0%, 5%] for Fig. 12(b). Regarding sensitivity, the system performance ranges between the intervals [62.0%, 70.0%] for Fig. 12(a) and (b) and Fig. 13(a) and between [60.0%, 95.0%] for Fig. 13(b). We must recall that one of the characteristics of imbalanced problems is related to the high values in specificity that they present, typically about 99%, which provides the consequent output of FP-ratio to be around the intervals [0.6%, 0.9%] for Fig. 12(a) and (b) and Fig. 13(a). Fig. 12 plots the points of the ROC curve segments corresponding to the different operation points provided by the different values of the parameter vector P after five runs. Each run is represented with a different symbol and color. Each graph (a), (b), (c), and (d) corresponds with one parameter in P (P_1, P_2, P_3, and P_4). Fig. 13 shows the points of the same ROC curve segments clustered by the same parameter. In these plots, each operation point is centered in the mean value of sensitivity and FP-ratio after the five runs, and the length of the ellipses axes is proportional to its standard deviation. The trade-off between sensitivity and specificity is kept for each run: A lower FP-ratio implies a lower specificity, and it is also accompanied by a lower value of sensitivity. Furthermore, our system appears

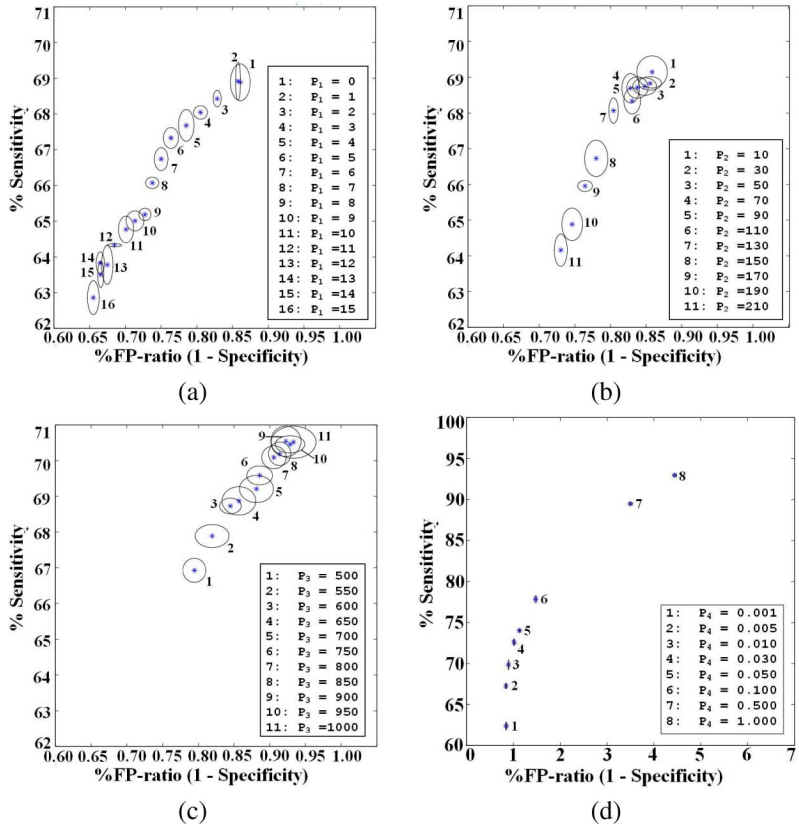


Fig. 13. Operation points from the ROC curve segments using the forward parameter selection procedure for (a) P_1, (b) P_2, (c) P_3, and (d) P_4, grouped by parameter value. The mean of each ellipse represents the mean of the performance pair obtained for that parameter after five runs. The axes of the ellipses are proportional to the resulting mean variance.

TABLE VI
PERFORMANCE OPERATION POINT FOR THE DIFFERENT PARAMETERS

Parameter	Sensitivity(std)		FAR(std)	
P^0	70.88	(0.51)	46.96	(0.79)
$P^{Best_{i,i=1:5}}$	71.35	(1.10)	47.96	(1.58)
P^{Best}	71.68	(0.44)	48.72	(0.54)

to be stable, in the sense that for several runs, the resulting operation point is confined in the ellipses drawn in the plots rendered in Fig. 13(a)–(c), showing no hysterical responses. We can observe the monotonically growing curves for the different parameter values, and the global displacement of the curve segment from the 60% to the 70% of sensitivity. Parameter P_4 presents the widest range of variability, being consistent with the role of γ, which controls the margins which directly affect the support vectors used for classification.

The final performance of the system was calculated in two different ways: 1) averaging the performance point of the five runs of the validation procedure tuned with $P^{\mathrm{Best}_{i,i=1:5}}$, and 2) averaging five runs of the system tuned in P^{Best}. Table VI shows these results in comparison with the performance of the system tuned to P^0, previously exposed. The final outcome confirms our hypothesis over P^0, since the confidence intervals of the performance values for the heuristically obtained parameters and those provided by the forward-propagation algorithm overlap both for sensitivity and FAR, assessing the equivalence of P^0 and P^{Best} in terms of performance.

A final remark must be mentioned regarding the suitability of including σ_{lum} value in the previous analysis. The σ_{lum} represents the scale value which best fits the lumen size. If this value were smaller, the LoG detector would tend to highlight folds, wrinkles, and/or intestinal content as false detections. Conversely, should this value be larger, the LoG detector would not detect the lumen hole at all at the maximum contraction frame. The case of P_1, P_2, P_3, and P_4 is different. The modification of each of these thresholds produces a different output at each level of the cascade, which will finally have a direct impact in the deal between false positives and true positives. The final SVM classifier will learn depending on the sequences arriving at the final stage, and the characteristics of these sets are to be different depending on the P values. In other words, while the σ_{lum} represents the optimal scale value for lumen segmentation, the P values represent degrees of freedom within the system to be tuned.

VI. CONCLUSION

This work addressed the problem of the automatic detection of intestinal contractions in capsule video endoscopy, a novel and highly challenging issue in medical imaging. The main novelty of our contribution is based on tackling the assessment of intestinal motility with a machine learning approach, which joins both classical image processing techniques and the use of diverse strategies for facing the low prevalence of contractions in

video. The main outcome is that we turned a useful but not feasible clinical routine, such as the manual labeling of intestinal contractions in video endoscopy studies, into a feasible clinical routine by means of their automatic detection, obtaining reasonable performance results, and providing the specialists with the main source of information needed for the further development of this novel field of clinical research.

We showed the design of the system in terms of sequential stages to be helpful from a two-fold perspective. On the one hand, by using this modular perspective, domain knowledge can be easily added to the system by means of the inclusion of new sequential stages to the cascade. On the other hand, we showed the rejection of negatives in a sequential way to be a useful strategy for dealing with this type of skewed data. We evaluated the accuracy of our approach, assessing by means of the interobserver variability tests that the patterns obtained by both our system and the experts are similar.

VII. Future Work

A special consideration must be taken about the enormous amount of information provided by this type of images. In this sense, the further analysis of dynamics involved in intestinal motility represents one of the most interesting perspectives of this work from the point of view of the physician. We argue that using dynamic information obtained through the optical flow analysis of each sequence can provide a promising line of research, which could help in the typification of the intestinal contractions and the discrimination of certain pathologies. The analysis of the speed of the intestinal contractions, their length in time (including the study of sustained contractions) and their degree of occlusion appear as new paradigms which deserve deeper and separate consideration. The extent to which diverse information (obtained from the automatic detection of sequences containing intestinal juices, quiet regions of the video, parameters related to transit time, relative and/or absolute position of the pill within the gut, etc.) can be joined together in order to provide an initial framework for clinical diagnosis constitutes challenging fieldwork for future contributions, in which promising results have been presented recently [34]–[36].

Other lines of research our group is currently involved in comprise finding textural descriptors for the inclusion of information associated to the wrinkle pattern by means of both the analysis of characteristic radial patterns [37], [38], and other features based on eigenmotion [39]. Finally, we propose to handle the skewed distributions of the data sets by both automatically identifying wide regions of the feature space associated with noncontractions and improving the classifier performance by means of ensemble methodologies [40].

Appendix I
Fast-Forward Parameter Selection Algorithm

The procedure used essentially matches the following highlights: We reset all the parameters to P^0, and established a range of possible values for each of them: 16 values for P_1 within the interval [0:15], 11 values for P_2 within the interval [10:210], 11 values for P_3 within the interval [500:1000], and the eight empirically selected values for P_4 [0.001, 0.005, 0.010, 0.030, 0.050, 0.100, 0.500, 1.000]. The choice of these intervals was performed based on the minimum and maximum values for each stage. The interval of the last parameter γ was carefully selected based on the observation of a substantial variation of the classifier performance. After the initialization step, the system was evaluated for all the possible values of P_1 within the defined range, and the best operation point (P_1^{Best}) was selected according to the performance criteria defined in Algorithm 1. The value of P_1^0 was substituted by P_1^{Best}, repeating the same procedure for the rest of the parameters in a sequential way (P_2, P_3, and P_4). The whole procedure was repeated five runs and the final set P^{Best} was obtained by averaging $P^{\mathrm{Best}_{i,i=1:5}}$. A trade-off was stated from the physicians' perspective of a working point around a 70% of sensitivity. This trade-off was completed by the use of the formula shown in the performance criterion, which searches for a valance between sensitivity and FAR. This formula allows a further tuning by means of the a and b coefficients, which play the role of weighting both measures.

Algorithm 1

BEGIN
SET ranges for parameters:
$R_1 \rightarrow [0:15]$ {16 values}
$R_2 \rightarrow [10:210]$ {11 values}
$R_3 \rightarrow [500:1000]$ {11 values}
$R_4 \rightarrow [0.001, 0.005, 0.010, 0.030, 0.050, 0.100, 0.500, 1.000]$ {8 values}
for $i = 1$ to 5 **do**
  Set parameters: $P = P^0 = \{P_1^0 = 0, P_2^0 = 50, P_3^0 = 650, P_4^0 = 0.01\}$ {initialization}
for $j = 1$ to 4 **do**
  Calculate the system performance substituting P_j^0 with each value of R_j.
  Apply the ***Performance Criterion*** to obtain P_j^{Best}.
  Substitute P_j^0 with P_j^{Best} in P.
end for
$P^{\mathrm{Best}_i} = \{P_1^{\mathrm{Best}}, P_2^{\mathrm{Best}}, P_3^{\mathrm{Best}}, P_4^{\mathrm{Best}}\}$
end for
$P^{\mathrm{Best}} = \mathrm{avg}(P^{\mathrm{Best}_i})$
END

Performance Criterion:

For all the performance pairs (Sensitivity, FAR) obtained for each parameter:

if For all the pairs, Sensitivity ≥ 70 **then**

We chose the parameter that achieves the higher sensitivity.

else

We select the two parameters with a closest value to 70 (higher or lower)

We choose the parameter which minimizes the error function:

$\text{sensitivity} * (a * \text{sensitivity}^2 + b * \text{FAR}^2)$, using $a, b = 1$

end if

REFERENCES

[1] J. Kellow *et al.*, "Principles of applied neurogastroenterology: Physiology/motility-sensation," *Gut*, vol. 45, pp. II17–II24, 1999.

[2] E. M. Quigley, "Gastric and small intestinal motility in health and disease," *Gastroenterol. Clin. North Amer.*, vol. 25, no. 1, pp. 113–145, 1996.

[3] E. M. Quigley, "Disturbances in small bowel motility," *Baillieres Best Practice Res. Clin. Gastroenterol.*, vol. 13, no. 3, pp. 385–395, 1999.

[4] M. B. Hansen, "Small intestinal manometry," *Physiological Res.*, vol. 51, pp. 541–556, 2002.

[5] F. DeIorio *et al.*, "New insight into intestinal motor activity: Correlation of endoluminal image analysis and displacement," *Gastroenterology*, vol. 128, no. 4, 2005.

[6] F. DeIorio *et al.*, "In search for new parameters of intestinal motor activity in humans," *Gastroenterology*, vol. 130, no. A7, pp. 43–, 2006.

[7] M. P. Tjoa and S. M. Krishnan, "Feature extraction for the analysis of colon status from the endoscopic images," *Biomed. Eng. Online*, vol. 2, pp. 3–17, 2003.

[8] G. Magoulas *et al.*, "Neural network-based colonoscopic diagnosis using online learning and differential evolution," *Appl. Soft Comput.*, vol. 4, pp. 369–379, 2004.

[9] M. M. Zheng, S. M. Krishnan, and P. Tjoa, "A fusion-based clinical support for disease diagnosis from endoscopic images," *Comput. Biol. Med.*, vol. 35, no. 3, pp. 259–274, 2005.

[10] V. S. Kodogiannis and H. S. Chowdrey, "Multi-network classification scheme for computer-aided diagnosis in clinical endoscopy," in *Proc. ICMSP (MEDISP)*, 2004, pp. 262–267.

[11] M. Boulougoura *et al.*, "Intelligent systems for computer-assisted clinical endoscopic image analysis," in *Proc. IASTED CBEI*, 2005, pp. 405–408.

[12] S. A. Karkanis *et al.*, "Computer aided tumor detection in endoscopic video using color wavelet features," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 3, pp. 141–152, Sep. 2003.

[13] G. Iddan *et al.*, "Wireless capsule endoscopy," *Nature*, vol. 405, p. 417, 2000.

[14] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[15] J. C. Russ, *The Image Processing Handbook*. Boca Raton, FL: CRC Press, 1994.

[16] F. Vilariño *et al.*, "A machine learning framework using SOMs: Applications in the intestinal motility assessment," in *Progress in Pattern Recognition, Image Analysis and Applications*. New York: Springer, 2006, vol. 4225, Lecture Notes Computer Science, pp. 188–197.

[17] P. Spyridonos *et al.*, "Identification of intestinal motility events of capsule endoscopy video analysis," in *Advanced Concepts for Intelligent Vision Systems*. New York: Springer, 2005, vol. 3708, Lecture Notes Computer Science, pp. 531–537.

[18] I. Guyon *et al.*, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, no. 1–3, pp. 389–422, 2002.

[19] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. New York: Elsevier, 2003.

[20] P. Ohanian and R. Dubes, "Performance evaluation for four classes of textural features," *Pattern Recognit.*, vol. 25, no. 8, pp. 819–833, 1992.

[21] T. Ojala *et al.*, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[22] F. Vilariño *et al.*, "Automatic detection of intestinal juices in wireless capsule video endoscopy," in *Proc. ICPR*, 2006, vol. 4, pp. 719–722.

[23] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Proc. ECML*, 1998, no. 1398, pp. 137–142.

[24] S. Siegel, *Nonparametric Statistics for the Behavioural Science*. New York: McGraw-Hill, 1956.

[25] N. V. Chawla, "Data duplication: An imbalanced problem?," in *Workshop on Learning From Imbalanced Datasets II, ICML*, 2003.

[26] T. Fawcett and F. J. Provost, "Adaptive fraud detection," *Data Mining Knowledge Discovery*, vol. 1, no. 3, pp. 291–316, 1997.

[27] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Mach. Learn.*, vol. 30, no. 2–3, pp. 195–215, 1998.

[28] P. Domingos, "Metacost: A general method for making classifiers cost-sensitive," in *Proc. ACM SIGKDD*, 1999, pp. 155–164.

[29] M. C. Monard and G. E. Batista, "Learning with skewed class distribution," in *Adv. Logic, Artif. Intell. Robot.*, 2002, pp. 173–180.

[30] R. Akbani *et al.*, "Applying support vector machines to imbalanced datasets," in *Proc. ECML*, 2004, pp. 39–50.

[31] F. Vilariño, "A machine learning approach for intestinal motility assessment with capsule endoscopy" Ph.D. dissertation, Dept. Comput. Sci., Univ. Autònoma de Barcelona, Barcelona, Spain, 2006 [Online]. Available: http://www.cvc.uab.es/~fernando/vilarino_thesis.pdf

[32] F. Vilariño *et al.*, "Experiments with SVM and stratified sampling with an imbalanced problem: Detection of intestinal contractions," in *Pattern Recognition and Image Analysis*. New York: Springer, 2005, vol. 3687, Lecture Notes Computer Science, pp. 789–791.

[33] C. Metz, "Basic principles of ROC analysis.," in *Seminars Nucl. Med.*, 1978, vol. VII, pp. 283–298.

[34] S. Segui *et al.*, "Diagnostic system for intestinal motility disfunctions using video capsule endoscopy," in *Computer Vision Systems*. New York: Springer, 2008, vol. 5008, Lecture Notes Computer Science, pp. 251–260.

[35] M. Bahsar *et al.*, "Detecting informative frames from wireless capsule endoscopic videos using color and tecture features," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2008*. New York: Springer, 2008, vol. 5242, Lecture Notes Computer Science, pp. 603–610.

[36] J. Cunha *et al.*, "Automated topographic segmentation and transit time estimation in endoscopic capsule exams," *IEEE Trans. Med. Imag.*, vol. 27, no. 1, pp. 19–27, Jul. 2008.

[37] F. Vilariño *et al.*, "Linear radial patterns characterization for automatic detection of tonic intestinal contractions," in *Progress in Pattern Recognition, Image Analysis and Applications*. New York: Springer, 2006, vol. 4225, Lecture Notes Computer Science, pp. 178–187.

[38] P. Spyridonos *et al.*, "Anisotropic feature extraction from endoluminal images for detection of intestinal contractions," in *Proc. MICCAI*, 2006, vol. 2, pp. 161–168.

[39] L. Igual *et al.*, "Eigenmotion-based detection of intestinal contractions," in *Computer Analysis of Images and Patterns*. New York: Springer, 2007, vol. 4673, Lecture Notes Computer Science, pp. 293–300.

[40] F. Vilariño *et al.*, "ROC curves and video analysis optimization in intestinal capsule endoscopy," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 875–881, 2006.