Diagnostic System for Intestinal Motility Disfunctions Using Video Capsule Endoscopy

Santi Seguí[1], Laura Igual[1], Fernando Vilariño[1], Petia Radeva[1,2], Carolina Malagelada[3], Fernando Azpiroz[3], and Jordi Vitrià[1,2]

[1] Computer Vision Center, Universitat Autònoma de Barcelona, Bellaterra, Spain
[2] Departament de Matemàtica Aplicada i Anàlisis, Universitat de Barcelona, Barcelona, Spain
[3] Digestive System Research Unit, University Hospital Vall d'Hebron; Ciberehd; and Department of Medicine, Autonomous University of Barcelona, Barcelona, Spain

Abstract. Wireless Video Capsule Endoscopy is a clinical technique consisting of the analysis of images from the intestine which are provided by an ingestible device with a camera attached to it. In this paper we propose an automatic system to diagnose severe intestinal motility disfunctions using the video endoscopy data. The system is based on the application of computer vision techniques within a machine learning framework in order to obtain the characterization of diverse motility events from video sequences. We present experimental results that demonstrate the effectiveness of the proposed system and compare them with the ground-truth provided by the gastroenterologists.

1 Introduction

Medical image analysis is an important research topic in the field of computer vision. Nowadays, images are crucial in many clinical procedures such as laser surgery, endoscopy and ultra sound scan, only to cite a few [1]. Moreover, in the last years a considerable effort has been done in the development of new image acquisition techniques increasing image accuracy and getting more image modalities with different views of the human body. Miniaturization of hardware is used to design sophisticated techniques allowing the analysis of parts of organs of the human body unattainable before. Wireless Video Capsule Endoscopy (WVCE) [2] is a novel technique using an ingestible device which allows to access to the entire bowel without surgery. This device consists of a small capsule of $11mm$ diameter and $25mm$ length, equipped with four illuminating leds, a camera, a battery and a wireless system (see Figure 1). The capsule is swallowed by the patient and travels along the intestinal tract. The captured images are transmitted by the radio frequency communication channel to a data recorder providing a video with two frames per second and 256×256 pixels of resolution. The analysis of intestinal motility activity is an important source of information for gastroenterologists to assess the presence of certain intestinal disfunctions. So far, motility assessment has been mainly performed by using invasive techniques, such as intestinal manometry. In this specific scenario, WVCE represents a much

A. Gasteratos, M. Vincze, and J.K. Tsotsos (Eds.): ICVS 2008, LNCS 5008, pp. 251–260, 2008.

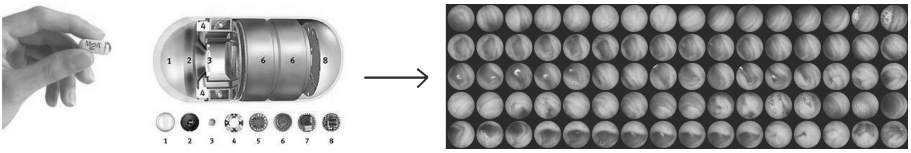

Fig. 1. Left: Wireless Video Capsule Endoscopy device: 1) Dome. 2) Lens holder. 3) Lens. 4) LEDs. 5) CMOS image sensor. 6) Battery. 7) ASIC transmitter. 8) Micro-antenna. Right: Example of capsule endoscopy video frames.

less invasive alternative, but it has the main drawback of the high amount of time needed for the off-line analysis by the expert, which is typically set around $4-6$ hours. This makes the visual analysis of WVCE unfeasible as a clinical routine. Novel approaches using video endoscopy data have been developed and successfully applied to detect some intestinal affections such as cancer, ulcers and bowel Crohn's disease [3,4,5,6,7,8]. However, up to our knowledge, automatic analysis based on WCVE has not yet been used to deal with intestinal motility diseases.

In this paper we manage this challenging open field of research by proposing an automatic system for the diagnosis of severe intestinal motility disfunctions using WVCE. Our proposal is based on the application of computer vision techniques within a machine learning framework in order to obtain the characterization of diverse motility events from video sequences. In particular, this study is focused on the automatic diagnosis of severe diseases by the analysis of video data from the portion of intestine comprised between the post-duodenum and cecum. In these sequences, the lumen, the gut wall and the intestinal contents can be distinguished. The lumen is the cavity where digested food goes through and can be recognized as a dark area, generally not centered in the image due to the free movement of the capsule within the gut. The intestinal wall is the visible part of the intestine and presents a range of colors spanning from orange to brown. The intestinal contents consist of remains of food in digestion and intestinal juices. The intestinal contractions are the result of muscular stimulation produced by the enteric nervous system. They appear as a closing of the intestinal lumen in a concentric way followed by an opening. Open lumen sequences are those sequences of frames in which the lumen appears static and opened for a long period of time when the intestine is relaxed, and there are not contractive movement. Sometimes, the camera has a null apparent motion and the visualized frames show the same image continuously. These frames belong to periods of repose in intestinal activity. In Figure 2 we display several examples of video frames. In the last years, machine learning has been applied to multiple medical image modalities in order to support clinical decisions. Support Vector Machines (SVM) [9] have become one of the state-of-the-art classification techniques due to their good performance and generalization power. Lately, different authors have shown the probabilistic version of SVM, namely Relevance Vector Machines (RVM) [10], to be useful in problems where an explicit outcome in

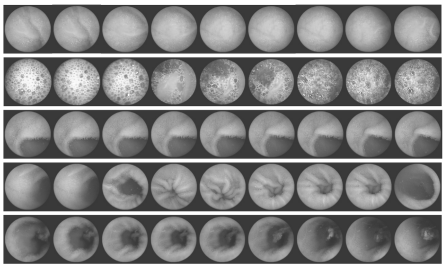

Fig. 2. Some examples of intestinal video images. From top to bottom: static sequence, turbid frames, tunnel sequence, occlusive contraction, non-occlusive contraction.

terms of the reliability of the classification result is important. Our system uses both SVM and RVM as essential parts of a global framework, in which multiple features are associated to the diverse motility events mentioned above.

This paper is organized as follows: Section 2 describes in detail the system architecture. In section 3 we present the experimental results and in Section 4 we expose the conclusions of this work.

2 System For Automatic Detection of Small Intestine Pathologies

Our system is developed in a modular way as is shown in the system architecture diagram displayed in Figure 3. It has two main blocks, namely: 1) the automatic capsule endoscopy video analysis, and 2) the characterization of pathologies. The first block is divided in three different modules which extract information of the video frames. These modules are referred to: *Intestinal Contents Analysis*, *Analysis of Idleness in Contractile Activity*, and, finally *Analysis of Manifest Contractile Activity*. In the second block, we use the extracted information in order to get new features with clinical relevance which describe the video in a higher level of interpretation. Then, we use them to learn and classify videos as belonging to patients and non-patients. Through the following paragraphs, we explain in more detail the system blocks.

2.1 Block 1: Automatic Capsule Endoscopy Video Analysis

1. Intestinal Contents Analysis Module: This module is in charge of finding all frames where intestinal contents are visualized. The intestinal contents can be present as intestinal juices, as well as food in digestion. The presence of the intestinal contents are characterized in terms of color, which usually ranges from brown to yellow, but mainly centered around green. Since the variability in color of the intestinal contents is very high and patient-specific, we propose a semi-automatic method for the detection of frames showing intestinal contents. Our approach consists of creating a non-supervised clustering with a Self-Organizing

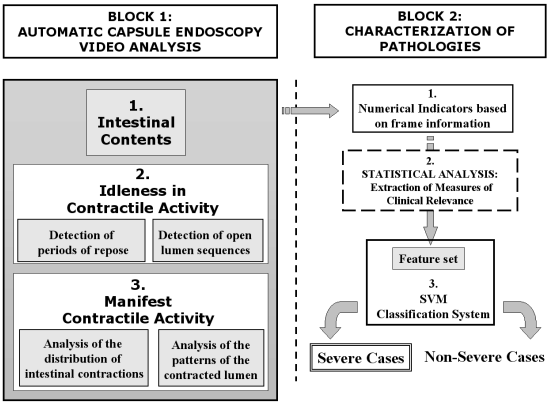

Fig. 3. System Architecture for automatic capsule endoscopy video analysis and pathology characterization

Map (SOM) method [11] with all video frames. To manage this, we first, convert each video image into the CIE-Lab [12] color space, then we compute the mean value of the components a and b of this color space for all video frames, and we clusterize the data using the SOM method. This process returns a set of cells which contains frames with similar color feature vector. Then, we manually select those cells of the SOM where intestinal contents can be appreciated (see Figure 4). We refer to these frames as *turbid frames*.

2. Analysis of Idleness in Contractile Activity: In this module we analyze the degree of quietness of the intestine from two different point of views: the movement of the camera and the movement of the intestine.

Detection of Periods of Repose: analyzes the quietness of the intestine from the point of view of the camera motion. We get three indicators for each image. (1) The *static degree* for each frame is computed by using the Earth Mover's Distance (EMD) method [13]. This method computes the distance between two distributions [13], which are represented by signatures. The signatures are sets of weighted features which capture the color distributions. The result is based on the minimal cost to transform a distribution into another. To compute the signatures of a color space, we first reduce the original RBG color space of the images into an RGB color space of 64 colors, resulting in a 64-bin histogram. For the computation of the EMD between two consecutive images, we use the Euclidian distance in this RGB color space. The output of the EMD method at the frame i represents the static degree, denoted SD_i. (2) Then, we define the *static label* of the frame i, denoted SL_i, as follows:

$$SL_i = \begin{cases} 1, \text{ if } \dfrac{1}{2n} \displaystyle\sum_{i-n<j<i+n} SD_j < \text{thr}_{\text{emd}} \\ 0, \text{ otherwise,} \end{cases}$$

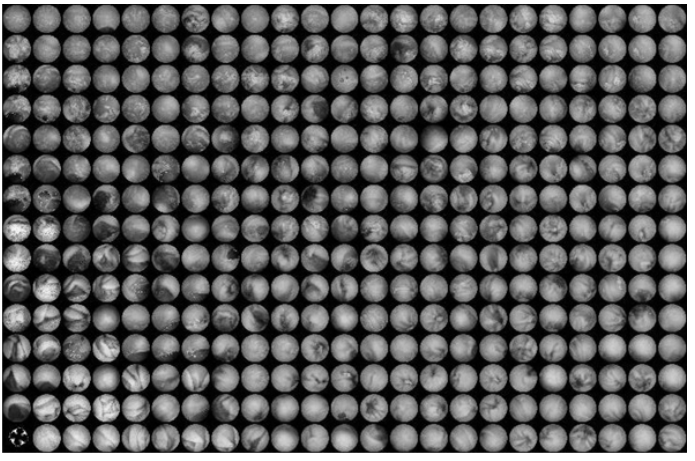

Fig. 4. Example of the cell prototypes of a SOM

where the threshold thr_{emd} and the value n were set to 0.01 and 40 respectively, after an empirical search. (3) Finally, in order to characterize static parts of the video, we consider the binary vector, $v = (SL_1, \ldots, SL_N)$ associated to a video sequence of size N and we perform a morphological closing to it with a structural element of size s (this value was fixed to 10 after several tests) to avoid small errors. A sequence with an associated vector v is defined as a *static sequences* if $\sum_{i=1,\ldots,N} SL_i \geq L$, where L was set to 60 experimentally. *Detection of open lumen sequences:* analyzes the quietness from the point of view of the intestinal lumen motion. Open lumen sequences are important cues to be studied, since they provide useful information about the degree of relaxation of the intestine. These sequences are described in terms of the lumen area along a sequence of nine frames. Nine is the number of frames usually involved in an contraction, since two is the rate of frames per second of the video and that the open-close-open cycle of a contraction takes between four to five seconds. In order to estimate the area of the lumen, denoted L_A, a Laplacian of Gaussian filter was applied (LoG)[14]. The LoG filter is a second order symmetric filter with a tunning parameter σ which plays the role of a scale parameter. The value of LoG is high when a dark hole is found in the image. The value of σ was fixed to 3, the minimum size of the lumen computed empirically. The system uses the following definition of the *tunnel label*:

$$TL_i = \begin{cases} 1, \text{ if } \dfrac{1}{9} \displaystyle\sum_{i-4<j<i+4} L_{Aj} < \text{thr}_{\text{lumen}} \\ 0, \text{ otherwise.} \end{cases}$$

In Figure fig:Tunnel we display some tunnel frames. Then, any sequence has a binary vector of this form $u = (TL_1, \ldots, TL_N)$ associated and it is called

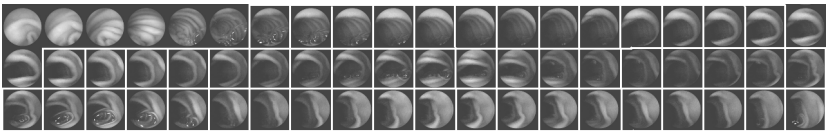

Fig. 5. Example of video frames. Frames with tunnel label are squared in white.

an *open lumen sequence* if after apply a morphological closing to u it verifies: $\sum_{i=1,\ldots,N} TL_i \geq M$, where M is the minimum length (experimentally set to 60).

3. Analysis of Manifest Contractile Activity: This module analyzes the contractile activity of the intestine, and extracts the star-wise visual pattern that is present when the intestine is contracting.

Patterns of manifest contractile activity. In this module, we analyze the intestinal contractile activity based on the method proposed by Vilariño et al. [15], which is focused onto the detection of intestinal contractions. Each step of the system filters a different subset of video frames.

- **Step 1: Intestinal Contents Filter.** The aim of this step is to reject the frames where we cannot appreciate correctly whether there is contractile activity. This happens when the intestinal contents hinder the right visualization of the lumen. In particular, this step rejects the frames labelled as (*turbid frames*) in the module 1 *Intestinal Contents Analysis*.
- **Step 2: Periods of Repose Filter.** We focus on rejecting all those frames where the camera is apparently stopped. We explain this effect by the absence of the contractile activity. The frames previously labelled as static frames in the module 2 *Analysis of Idleness in Contractile Activity* are rejected here.
- **Step 3: SVM Classifier.** This last module performs the classification, which is based on SVM with a radial basis function kernel (the σ value was fixed by using cross-validation method). The feature vectors are built up by using the same 54 features proposed in the literature [15], which have shown to provide optimal results for the discrimination of contractile activity. The output of this last step gives us the distribution of the intestinal contractions.

In this module we also extract the *occlusive degree* of the frames with contractile activity. The area of the lumen is the principal feature to characterize this property, since generally, in non-occlusive contractions the lumen area in the central frame is bigger than in occlusive contractions. We compute the lumen area for each frame and its four previous and following neighbors. The resulting vector is used as feature vector for the classification. The result of the RVM Classifier is the occlusion degree of each frame.

Analysis of the contracted lumen. The star-wise pattern is an omnipresent characteristic of the contracted lumen. This pattern is characterized by strong edges of the folded intestinal wall, distributed in an approximately radial way around the intestinal lumen. In order to distinguish this pattern, an accurate wrinkle detector is essential. We follow the strategy proposed by Spyridonos et al.

[16]. First, the skeleton of the wrinkle pattern is extracted and the center of the intestinal lumen is detected. Finally, a set of descriptors are estimated taking into account the near radial organization of the wrinkle skeleton around the center of the intestinal lumen. At the end of this process 14 features are introduced and processed by a RVM Classifier.

2.2 Block 2: Feature Extraction and Classification

Feature Extraction. Once the important information at frame level is extracted we have to define a set of features at video level to characterize intestinal patients in motility disease. We used 19 features, which were selected by the physicians, provided their clinical relevance. Let us define all of them:

1. Related to *open lumen sequences feature*, we extract the percentage of open lumen sequences in the valid parts (valid parts are all frames except those where the intestinal contents do not allow a good visibility), the mean of the static values, computed by EMD, at all open lumen sequences frames and the mean length of the open lumen sequences.
2. From the *intestinal contents feature*, we extract the percentage of the video that are turbid, the percentage of turbid frames that are static, and the mean of static degree of all turbid frames.
3. From *static feature*, we extract the percentage of frames that are static from the set of valid frames and from all video frames, the mean of static degree from all video frames and the mean length of static sequences.
4. Related to *contractile activity*, we extract the number of detected contractions per minute in valid parts, as well as in the whole video, the percentage of contractions in which wrinkles are presented and the percentage of contractions that are non-occlusive.
5. Finally, related to *contracted lumen*, we extract the percentage of frames where the star-wise pattern can be observed, we get this feature considering all video frames, and only considering valid video frames. We also get the percentage of frames that has a low level of wrinkle presence, considering a low value 0.1, and the percentage of frames that have a high level of wrinkle presence, considering as all the frames that have a wrinkle level higher than 0.9. Additionally we get the mean length of wrinkle sequences.

Classification. The last stage of our approach consists of applying the SVM classifier to detect intestinal pathology from input capsule video. The output of the SVM consists of video data suggested by the specialist as the candidates for patients on motility diseases.

Semi-supervised learning. Medical field of work usually deals with imbalanced problems. It is common that the acquisition of labelled data is costly and not always possible, however, it is also usual to dispose of another data set of unlabelled examples. Standard classifiers only use labelled data, in contrast, the semi-supervised learning classifiers address the imbalanced problems by tacking profit

of the available unlabelled data set. In our case, by using unlabelled data we pretend to increase the number of patient examples. We propose the algorithm displayed in Table 1. The basic idea is as follows: to create an initial training set with labelled data, and in each new iteration to add a new patient example from the unlabelled data set -this new example has to be the one classified as patient with the highest score-. This iterative loop will stop when no example within the unlabelled data set is classified as patient. Finally, all examples that were not classified as patient will be included in the healthy subjects set.

Table 1. Semi-supervised algorithm for imbalanced problem

1. Pick your favorite classification method.
2. Train a classifier f from (X_s, Y_s).
3. Use f to classify all unlabelled items $x \in X_u$.
4. Pick x^* such that $f(x^*) \in C_1$ (C_1: Patient class) and having the highest confidence, add $(x^*, f(x^*))$ to the labelled data.
5. Repeat until all samples are classified as C_2 (C_2: Non-Patient class).
6. Add the remaining $(x, f(x))$ to the labelled data.

3 Experimental Results

For our experiments, we considered a set 86 of videos obtained using the WVEC device developed and provided by Given Imaging, Ltd., Israel [17]. The capsule endoscopy interventions were conducted at Digestive Diseases Department of General Hospital "Vall d'Hebron" in Barcelona, Spain. The video data set is divided in three groups: control set of volunteers showing no apparent disease symptoms, patient set with pathological test manometry, and another set of subject with symptoms, but without pathological test manometry, and thus, without any diagnostics. Volunteers were randomly selected from a big pool of subjects without any apparent symptom. We considered these cases as *healthy* subjects without performing a manometry test, because the probability of them of being patient is too small. The final testing set was formed by 86 videos: 50 from healthy volunteers, 19 from patients with pathological manometry test and 17 from subjects with symptoms but without pathological manometry test. The first two groups, healthy volunteers and patients with pathological manometry test, were considered as *labelled data*. However, the data set composed by patients without pathological manometry test was considered as *unlabelled data*.

From a clinical point of view, the gastroenterologists are interested in assessing: a) *sensitivity*, b) *specificity*, c) system *precision*, and finally, d) the *False Alarm Ratio*. All these features are described in terms of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). Table 2 summarizes these definitions.

Table 2. Validation Measures

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

Classification results. A first test was performed with a data set formed only by the labelled data set. The first row of Table 3 shows the obtained results performing leave-one-out validation method [18] over this data set. As we can observe we get a 4.35% of error, 84.21% of sensitivity and 100.00% of specificity and precision. Three of the patients were considered as healthy subject and all of the healthy subjects were correctly classified.

The previous test presents an important drawback: the small size of the data set. In order to overcome this problem we perform a second test considering a new data set including the labelled and unlabelled data set after applying the algorithm proposed in Table 1. We perform a leave-one-out validation method with the new data set. Note that the obtained classification error is computed only by using labelled data, and the unlabelled data is only used in order to create the training data set. In the second row of Table 3 we display the obtained results to this test. The error decreased from 4.35% to 1.16% and the sensitivity increased from 84.21% to 94.74% without affecting to the specificity, precision and FAR.

Table 3. Classification Results

	Error	Sensitivity	Specificity	Precision	FAR
supervised learning	4.35%	84.21%	100.00%	100.00%	0.00%
semi-supervised learning	1.16%	94.74%	100.00%	100.00%	0.00%

4 Conclusions

In this paper we proposed an innovative automatic system for the diagnostic of intestinal motility disfunctions based on computer vision and machine learning techniques. This diagnostic process uses the WVCE data and is a promising alternative to the currently used invasive techniques, as the manometry. The proposed system, firstly, extract the important visual features which characterize phenomenons present in WVCE, secondly, define and analyze clinical patterns to characterize the pathologies and classify the video. This work entails a deep study of video frames and the definition of the most important physiological aspects leading to a new technical and medical terminology. Another important contribution is the application of the semi-supervised algorithm for overcoming the problem of the small quantity of patient examples. The obtained results show that this system represents a promising method for patient diagnostic in intestinal motility disfunctions. The principal strengths of this system test are the low level of invasion and that the monitorization of the clinical process by gastroenterologists is not necessary.

Acknowledgements

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, H. U. Vall d'Hebron, Barcelona, Spain, as well as the projects TIN2006-15308-C02 and FIS-PI061290.

References

1. Sonka, M., Fitzpatrick, J.M.: Handbook of Medical Imaging. SPIE Press (2000)
2. Iddan, G., Meron, G., et al.: Wireless capsule endoscopy. Nature 405, 417 (2000)
3. Tjoa, M.P., Krishnan, S.M.: Feature extraction for the analysis of colon status from the endoscopic images. Biomedical Engineering OnLine 2, 3–17 (2003)
4. Karkanis, S.A., Iakovidis, D.K., et al.: Computer aided tumor detection in endoscopic video using color wavelet features. IEEE Transactions on Information Technology in Biomedicine 7, 141–152 (2003)
5. Magoulas, G., Plagianakos, V., et al.: Neural network-based colonoscopic diagnosis using online learning and differential evolution. Applied Soft Computing 4, 369–379 (2004)
6. Zheng, M.M., Krishnan, S.M., Tjoa, P.: A fusion-based clinical support for disease diagnosis from endoscopic images. Computers in Biology and Medicine 35(3), 259–274 (2005)
7. Kodogiannis, V.S., Chowdrey, H.S.: Multi-network classification scheme for computer-aided diagnosis in clinical endoscopy. In: Proceedings of the International Conference on Medical Signal Processing (MEDISP), pp. 262–267 (2004)
8. Boulougoura, M., Wadge, V., et al.: Intelligent systems for computer-assisted clinical endoscopic image analysis. In: Proceedings of the 2nd IASTED Conference on Biomedical Engineering Innsbruck, pp. 405–408 (2005)
9. Vapnik, V.N.: An overview of statistical learning theory. IEEE Transactions on Neural Networks, 988–999 (1999)
10. Tipping, M.: The relevance vector machine. In: Advances in Neural Information Processing Systems, San Mateo, CA, Morgan Kaufmann, San Francisco (2000)
11. Kohonen, T.: Self-Organizing Maps. Springer, Heidelberg (1995)
12. CIE: Colorimetry - part 4: Cie 1976 l*a*b* colour spaces. Cie draft standard ds 014-4.2/e:2006 (2006)
13. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision 40(2), 99–121 (2000)
14. Russ, J.C.: The Image Processing Handbook. CRC Press, Boca Raton (1999)
15. Vilariño, F., Spyridonos, P., Vitrià, J., de Iorio, F., Azpiroz, F., Radeva, P.: Intestinal motility assessment with video capsule endoscopy: Automatic annotation of intestinal contractions. IEEE Trans. on Medical Imaging (under revision) (2006)
16. Spyridonos, P., Vilariño, F., Vitrià, J., Radeva, P.: Anisotropic feature extraction from endoluminal images for detection of intestinal contractions. LNCS (in press, 2006)
17. Given Imaging, L. (2007), `http://www.givenimaging.com`
18. Stone, M.: Cross-validatory choice and assessment of statistical predictions (with discussion). Journal of the Royal Statistical Society B 36, 111–147 (1974)