Identification of Intestinal Motility Events of Capsule Endoscopy Video Analysis

Panagiota Spyridonos, Fernando Vilariño, Jordi Vitrià , Petia Radeva,

Computer Vision Center, Universitat Autònoma de Barcelona
08193 Bellaterra, Spain

{panagiota, fernando, jordi, petia}@cvc.uab.es
http://www.cvc.uab.es

Abstract. Purpose: To develop a system for assisting the analysis of capsule-endoscopy (CE) video data and identifying sequences of frames related to small intestine motility. Material and Methods: Six videos were analyzed and labelled manually by an expert, indicating the events of contractions. For addressing the imbalanced recognition task of small intestinal contractions we employed an efficient two-level video analysis system. At the first level of the system, each video was processed resulting in a number of possible sequences of contractions. In the second operating part of the system, the final recognition of contractions sequences was carried out by means of a SVM classification algorithm. To encode the patterns of intestinal motility a panel of textural and morphological features of the intestine lumer were extracted. Results: The system exhibited an overall sensitivity of 73.53% in correct detecting events of contractions. The false alarm ratio (false positives over the true positives) was of the order of 59.92%.Conclusion: these results serve as a first step for developing assisting tools for computer based CE video analysis, reducing drastically the physician's time spent in image evaluation and enhancing the diagnostic potential of CE examination.

1 Introduction

Conventional endoscopic techniques for examining the small intestine (SI) are limited by its length (3.5-7.0 m) and by its complex looped configurations [1]. The current methods for imaging the SI include, primarily, barium X-rays and enteroscopy. However, the diagnostic value of radiographic means for lesions such as angiodysplasias, and neoplasms is low [2]. On the other hand, direct visual inspection by enteroscopy, is highly invasive and is associated with discomfort and occasionally complications [3].

Capsule endoscopy (CE) is a new wireless endoscopy examination of the entire SI. Moreover, CE is a technological invention designed to aid the gastroenterologist in diagnosing SI diseases with higher sensitivity. The CE system is composed of the ingestible capsule, the data recorder, and the work station supplied with the appropriate image-visualization software. The capsule acquires two images per second and during a typical 8-hour examination, the recording device of the capsule stores about 50.000 images. After examination, images are downloaded to a PC workstation [4]. An expert physician is needed to inspect visually the video and to diagnose the presence (or absence) of abnormality.

However, the visualization of the whole study (video) is a burden and time consuming procedure. In most of the cases the time it takes for a physician to review the capsule study is between one and two hours. This is quite a heavy load for the physician that renders the diagnostic task difficult and subject to variations in individual interpretation. [5]. Subsequently, it would be particularly useful for physician to have an adjunctive tool able to short the reading time of a study and to automatically recognize sequences of frames meaningful for analysis.

Digital image processing and analysis techniques offer potential solutions to endoscopic images understanding and objective interpretation. Several researches have reported that endoscopic images of the lower gastro-intestinal part, carry rich information which, if quantified in terms of textural, colour or other morphological features such as the lumen region, can allow the diagnosis of certain types of colon cancer [6-8]. However, up today no preceding work has been reported on computerized analysis of CE data.

In the present study, we introduce a CE video analysis system for detecting specific patterns related to intestine motility. The frequency and the type of contractions are of main interest and seemed to be correlated to the presence of several SI diseases [9].
The value of the proposed system is its ability to highlight special patterns of intestinal activity which might carry diagnostic information, reducing significantly the reading time of a CE study.

2 Material and Methods

2.1 Material

Clinical data (videos) where obtained by CE from six volunteers , in Digestive Diseases Dept., Hospital General “Vall D’ Hebron” in Barcelona, Spain. The endoscopic capsule used, was developed by Given Imaging Limited, Israel [10]. Measuring 11x26 mm, the capsule contains 6 light emitting diodes, a lens, a colour camera chip two batteries, a radio frequency transmitter, and an antenna. The capsule acquires two images per second at 256x256x24-bit resolution and transmits the data via radiofrequency to a recording unit located outside the body. Upon completion of the examination the data were transferred to the workstation for further visualization. Contractions were considered as dynamic events occurred in sequences of nine frames in the intestinal part between duodenum and cecum. Six videos were analyzed and labelled manually by an expert, specifying the time interval between duodenum and cecum and indicating the central frame in each sequence of contraction. . In Table I, are indicated the number of frames per video registered between duodenum and cecum and the number of findings in this interval. Typical sequences of contractions are illustrated in Fig. 1.

Table I. Number of frames and findings per video

	Number of frames for analysis	Number of Findings
Video_1	29424	716
Video_2	28783	487
Video_3	27796	524
Video_4	38865	718
Video_5	17599	347
Video_6	27156	911

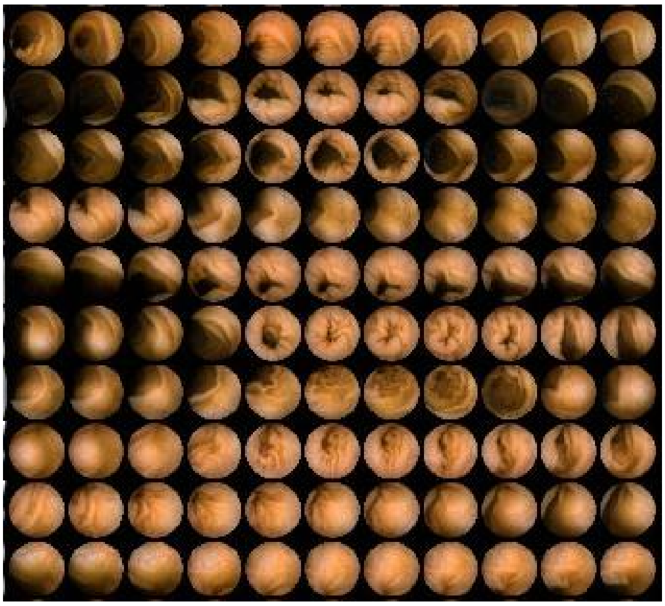

Figure1. Examples of sequences of intestinal contractions

2.2 Methods

The automatic recognition of contractions in a CE video is a highly skewed classification problem on the order of 50 to 1(Table I), which hardly can be tackled with a conventional direct classification process. Class imbalance is a well known issue for several real pattern recognition applications and has been addressed mainly by assigning distinct costs to training samples [11], by re-sampling the original dataset [12] or using cascade classifiers [13]. In this work, we addressed the imbalanced recognition task of small intestinal contractions by means of an efficient two-level video analysis process. At the first level of the system, each video was processed resulting in a number of possible sequences of contractions, under the hypothesis that a contractions might be described as a rapid closing and opening of the intestinal lumen and subsequently could be characterized by a sharp variation of the grey-level intensity.

The feature used to capture the intensity variation in a sequence was the locally normalized mean intensity of the image IN given by:

$$I_N = I - \bar{I} \qquad (1)$$

where, I is the mean grey-level intensity of the frame, and $\bar{I}$ is the averaged intensity estimated every 9-frame sequence:

$$\bar{I} = \frac{\sum_{i=1}^{9} I_i}{9} \qquad (2)$$

Following this estimation, the intervals of sequential frames with positive IN were extracted from the whole series of the study. The frame with the maximum value IN in each interval was considered as the central frame of a possible sequence of contraction. An example is illustrated in Fig.2 and Fig.3.

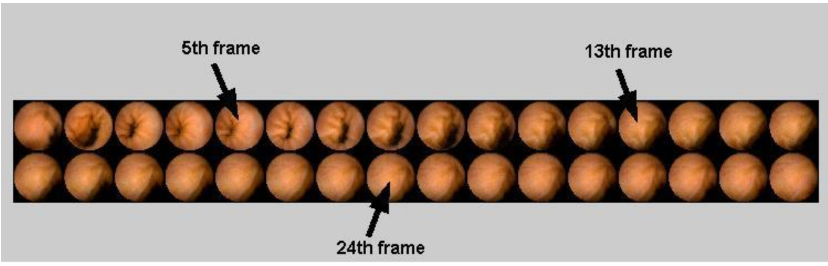

Figure 2. Example of 32 sequential video frames; the indicated frames were selected during the first level of analysis as central frames of candidate sequences of contractions

In the second operating part of the system, the final recognition of contractions sequences was carried out by means of a two class SVM classification algorithm [14]. To encode the patterns of intestinal motility a panel of textural and morphological features were extracted. Textural descriptors comprised features from first and second order statistics [15] and from Rotation Invariant Uniform Local Binary Units operator (LBPriu2) applied in a circular symmetric neighbourhood P of radius R (P=16, R=2) [16]. Morphological features of the intestinal lumen comprised measurements of blob area, blob shape (solidity), blob sharpness and blob deepness.

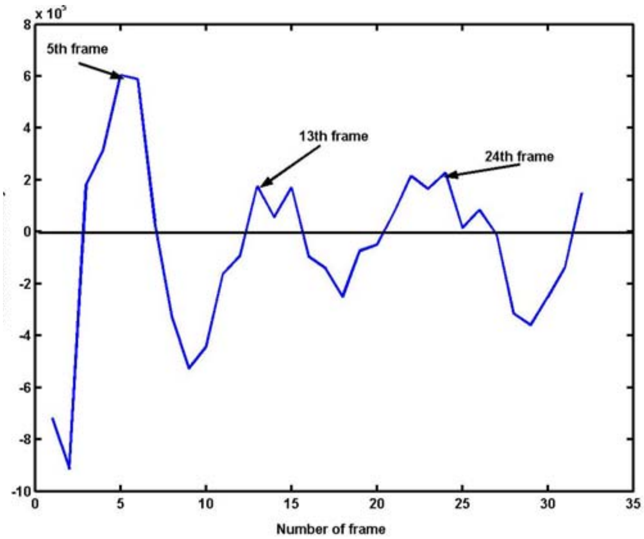

Figure 3. Graphical plot of normalized intensity for the 32 video frames; from each interval of sequential frames with positive $I_{N,}$ the frame with maximum value of I_N is pre-selected as the central frame of a possible contraction sequence.

To estimate the lumen area (blob), frames were processed by a Laplacian of Gaussian filter [17]. This filter has a high response at valleys that are dark regions surrounded of brighter regions. In our case, the region of interest in a frame was the lumen area. Dark areas were extracted by applying a greater-than-zero-threashold in the Laplacian image. The resulting binary image was superimposed to the Laplacian image. In the new image the blob sharpness was estimated by summing the pixel values of Laplacian image in the extracted objects. The object with the greater sum was selected as the blob area (Fig.4). Blob deepness is the minimum of the Laplacian valley in the blob area.

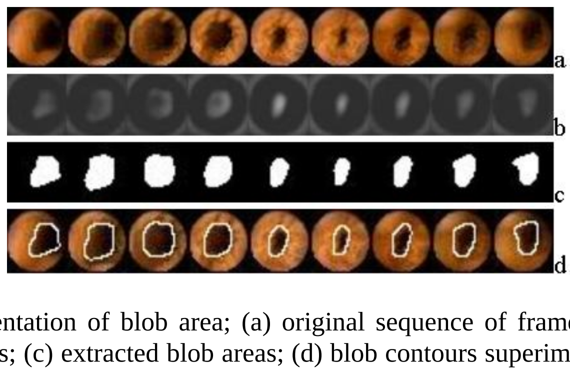

Figure 4.Segmentation of blob area; (a) original sequence of frames; (b) estimated laplacian images; (c) extracted blob areas; (d) blob contours superimposed to original images

Following the feature extraction, each sequence was represented by a 37x9 dimensional feature vector. For feature reduction, a sequential forward selection method was used based on the performance of the system [18]; starting from an empty feature set, sequentially feature x + is added that results in better classification performance when combined with the features Yk that have already been selected.

To evaluate the performance of the classifier, each time all the videos except one, were used for learning and the one kept out was used for testing. For each video the performance was evaluated in terms of sensitivity, specificity and False Alarm Ratio (FAR) (Table II). We used the latter as a false positive error measure since specificity rate by itself is not a good metric for skewed datasets. The definitions for sensitivity, specificity and FAR are the following:

$$\text{Sensitivity}=\frac{TP}{TP+FN}*\%, \quad (4)$$

$$\text{Specificity}=\frac{TN}{FP+TN}*\% \quad (5)$$

$$\text{FAR}=\frac{FP}{TP+FN}*\% \quad (6)$$

Table II. Truth table construction

	Manual Identification	
System Identification	Contractions	Non Contractions
Contractions	True Positives (TP)	False Positives (FP)
Non Contractions	False Negatives (FN)	True Negatives (TN)

3 Results and discussion

CE technology offers a safe, painless and effective method of diagnosing abnormalities in the SI. Current methods can be uncomfortable, or might be of limited diagnostic ability [2]. Although CE provides an excellent view of inaccessible parts of intestine, the amount of information registered during the capsule's transport time through the gastrointestinal tract is huge. Consequently, analyzing a CE study visually and qualitatively is a difficult and time consuming procedure, coupled with subjective interpretations. In the present study, we introduced a CE video analysis system based on computerized image analysis techniques. According to this system, data were analyzed in a cascade way in which redundant information was removed gradually. In this way, the reading time shorten significantly, without considerable loss of diagnostic information related to intestine motility. The detection of contractions is a primary feature, assessed visually by the experts during the CE visual inspection.

At the first level, the system reduced drastically the amount of data by removing 89.17% of them as redundant and missing 3.04% of the labelled findings (Table III).

Table III. First level processing of CE data

	Sequences passing the first stage	Lost Contractions	Number of Contractions	Number of Non-Contractions
Video_1	3220/29424 (10.94%)	26/716 (3.63%)	690	2530
Video_2	3072/28783 (10.67%)	25/487 (5.13%)	462	2610
Video_3	3194/27796 (11.49%)	11/524 (2.09%)	513	2681
Video_4	4056/38865 (10.43%)	15/718 (2.08%)	703	3353
Video_5	1869/17599 (10.61%)	7/347 (2.01%)	340	1529
Video_6	2950/27156 (10.86%)	30/911 (3.29%)	881	2069
MEAN	***10.83%***	***3.04%***		

At the second level the system refined the recognition of contractions by receiving only 10.83% of the initial data volume. SVM classifier exhibited a performance of 75.81% in correct recognizing contractions and of 85.69 % in correct identifying non-contraction sequences. In Table IV are given the results from the second processing level.

Table IV. Second level processing of CE data

	TP	TN	FP	FN
Video_1	580	2082	448	110
Video_2	319	2272	338	143
Video_3	369	2395	286	144
Video_4	575	2711	642	128
Video_5	268	1343	186	72
Video_6	610	1796	273	271

The system yielded an overall sensitivity of 73.53% in correct detecting events of contractions (Table V). This percentage might be considered moderate in comparison with the overall specificity which was of the order of 98.76%. However, in highly skewed classification problems with a very small number of positive instances the specificity rate, by itself, is a rather obscuring measure of accuracy, which might mislead the understanding of system performance. For this reason we have used the FAR which is more indicative measure of the system accuracy. The overall FAR of the system was of 59.92%. Concluding, the results from the present study might be promising in the development of assisting tools for computer based CE video analysis, reducing drastically the physician's time spent in image evaluation and enhancing the diagnostic potential of CE examination by introducing qualitative descriptors in diagnostic assessments.

Table V. Overall system performance in correct identifying sequences of SI contractions

	Overall sensitivity	Overall specificity	Overall FAR
Video_1	580/716 (81.00%)	28976/29424 (98.47%)	448/716 (62.56%)
Video_2	319/487 (65.50%)	28445/28783 (98.82%)	338/487 (69.40%)
Video_3	369/524 (70.41%)	27510/27796 (98.97%)	286/524 (54.58%)
Video_4	575/718 (80.08%)	38223/38865 (98.34%)	642/718 (89.41%)
Video_5	268/347 (77.23%)	17413/17599 (98.94%)	186/347 (53.60%)
Video_6	610/911 (66.95%)	26883/27156 (98.99%)	273/911 (29.96%)
MEAN	***73.53%***	***98.76%***	***59.92%***

4. REFERENCES

[1] Z.Fireman, A.Glukhovsky, H.Jacob, A. Lavy. S. Lewkowicz, E.Scapa , "Wireless Capsule Endoscopy" IMAJ, vol.4,pp. 717-719,2002.

[2] K.Schulmann,S.Hollerbach, K. Kraus, J.Willert, T.Voleg, G.Moslein, C.Pox, M.Reiser, A. Reinacher-Schick, W. Schmiegel, " Feasibility and Diagnostic Utility of Video Capsule Endoscopy for the detection of Small Bowel Polyps in Patients with Hereditary Polyposis Syndromes",American Journal of Gastroenteroscopy, vol.100,pp. 27-37,2002.

[3] JD.Waye, "Small-intestinal endoscopy", Endoscopy, vol.33(1),pp. 24-30,2001.

[4] J-F.Rey, G.Gay, A. Kruse, R.Lambert, "European Society of Gastrointestinal Endoscopy Guideline for Video Capsule Endoscopy", Endoscopy, vol.36,pp. 656-658,2004.

[5] D.G.Alder, C.J. Gostout, "Wireless Capsule Endoscopy",Hospital Physician, pp.14-22, ,May 2003

[6] M.P. Tjoa, S.M.Krishman, "Feature extraction for the analysis of colon status from the endoscopic images",Biomedical Engineering OnLine, vol.2,pp. 3-17,2003.

[7] M.M.Zheng, S.M.Krishman, M.P. Tjoa, "A fusion-based clinical support for disease diagnosis from endoscopic images",Computers in Biology and Medicine, Article in Press

[8] S.A, Karkanis, G.D, Magoulas, D.K. Iakovidis, D.E. Maroulis, N. Theofanous, "Tumor recognition in endoscopic video images"26th EUROMICRO Conference, Maastricht, Netherlands, pp.423-429,2000.

[9] M.B.Hansen, "Small Intestinal Manometry", Physiological Research,

vol.51,pp. 541-556,2002.
[10] http://www.givenimaging.com
[11] G.Karakoulas, J.S. Taylor, “Optimizing classifiers for imbalanced training sets”, Proceedings of Neural Information Processing Workoshop NIPS’98, pp.253-259
[12] N.V. Chawla, K.W.Bowyer, L.O.Hall, W.P.Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique”, Journal of Artificial Intelligence Research, vol.16,pp. 321-357,2002
[13] P. Viola and M. Jones, "Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade," Advances in Neural Information Processing System 14, MIT Press, Cambridge, MA, 2002.
[14] V.Vapnick, Statitistical Learning Theory, John Wiley&Sons.
[15] S.Theodoridis, K.Koutroumbas , Feature Generation II. In Pattern Recognition 1998: Academic Press.
[16] T.Ojala, M.Pietikainen, “Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns”,IEEE Transactions on Pattern Recognition Analysis and Machine Intelligence, vol.24(7),pp. 971-987,2002.
[17] J.C.Russ. The Image Processing Handbook. CRC Press. 2nd Edition,1994
[18] Theodoridis S, Koutroumbas K.: Feature Selection. Preprocessing. In Pattern Recognition 1998: Academic Press.