

A Machine Learning Framework Using SOMs: Applications in the Intestinal Motility Assessment

Fernando Vilariño¹, Panagiota Spyridonos¹, Jordi Vitrià¹,
Carolina Malagelada², and Petia Radeva¹

¹ Computer Vision Center. Universitat Autònoma de Barcelona,
Bellaterra 08193, Spain

² Hospital Vall d'Hebron, Barcelona 08035, Spain

fernando@cvc.uab.es

<http://www.cvc.uab.es/~fernando>

Abstract. Small Bowel Motility Assessment by means of Wireless Capsule Video Endoscopy constitutes a novel clinical methodology in which a capsule with a micro-camera attached to it is swallowed by the patient, emitting a RF signal which is recorded as a video of its trip throughout the gut. In order to overcome the main drawbacks associated with this technique -mainly related to the large amount of visualization time required-, our efforts have been focused on the development of a machine learning system, built up in sequential stages, which provides the specialists with the useful part of the video, rejecting those parts not valid for analysis. We successfully used Self Organized Maps in a general semi-supervised framework with the aim of tackling the different learning stages of our system. The analysis of the diverse types of images and the automatic detection of intestinal contractions is performed under the perspective of intestinal motility assessment in a clinical environment.

1 Introduction

Wireless Capsule Video Endoscopy constitutes a recent technology in which a capsule with an attached camera is swallowed by the patient. The camera travels along the intestinal tract and emits a radio signal, in a ratio of two frames per second, which is recorded as a video in an external device carried by the patient [6]. Once the study is finished, the final record can be easily downloaded into a PC with the appropriate software installed for its posterior analysis by the physicians. Recently, several works have tested the performance of capsule endoscopy in multiple clinical studies. Some of these clinical scenarios include intestinal polyposis and the diagnosis of small bowel tumors, obscure digestive tract bleeding, Crohn's disease and small bowel transplant surveillance [17,11,3,1,8,15,7].

The assessment of intestinal motility using wireless capsule video endoscopy is a novel and challenging fieldwork, which has been the focus of interest in recent publications [20,21]. In this medical image modality, the intestinal contraction sequences constitute the target of the labelling process, which consists

of video sequences showing the radial contraction of the intestinal lumen. These events have a low prevalence, presenting a typical ratio about 1 frame out of 50 video frames. The clinical procedure associated with this technique implies the visualization of the whole video (1-2 hours per video), which represents a huge amount of time and efforts for the specialist. In addition to this, images from video capsule endoscopy are usually significantly noisy, due to the presence of turbid content from the intestinal juices which hinders the correct visualization of the gut, providing wide range video zones not valid for analysis. All these drawbacks hinder the manual labelling process, and drive a useful clinical routine into a not feasible one. For this reason, we tackle the problem of intestinal motility assessment in a novel approach involving classical machine learning techniques. The main traits of our approach can be listed as follows:

- We state our problem in a cascade-wise strategy, identifying the different features of interest involved in motility assessment, and associating them with different and independent stages.
- We use self organized maps in an exhaustive way in the learning stage of each part of the cascade, adding the possibility of integrating the expert knowledge in a semi-supervised context.
- Finally, we apply a classification step based on a support vector machine in order to accomplish a two-fold objective: the identification of the portion of the video not valid for analysis, and the automatic labelling of intestinal contractions.

By using this strategy we can face several challenges in a simultaneous way: On the one hand, the resulting model is open and flexible, and new features of interest can be easily included adding new stages in the cascade as new necessities appear in the clinical scenario. On the other hand, the intervention of the specialists is reduced to a common framework, implemented with the self organized maps, making the expert-system interactivity bounded to a single common application interface.

In this paper we deploy the former issues according to the following structure: Section 1 is devoted to the explanation of the proposed system as a sequential cascade. Section 2 deploys the main traits exposed in Section 1, focusing on the identification of the portion of the video which is not valid for analysis, the SOM framework for the training stages, and the final classification step for the identification of intestinal contractions. Section 4 presents the results of the system in terms of intestinal motility assessment. We finally conclude this piece of research with the conclusions exposed in Section 5.

2 A General Framework: The Cascade System

Our system is deployed in a sequentially modular way, namely, a cascade, as pictured out in Figure 1. 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

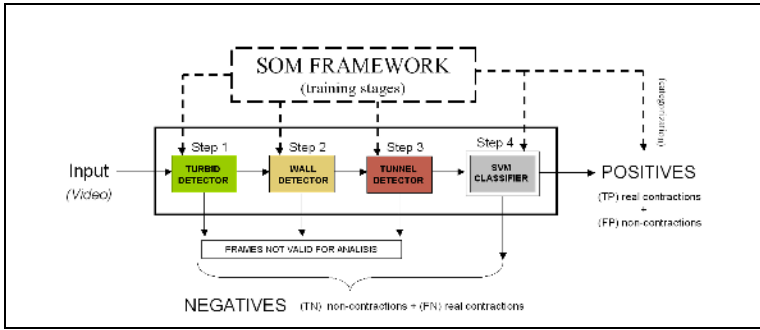


Fig. 1. Sequential Cascade System for Intestinal Motility Assessment

frames are distributed in several classes: turbid frames not valid for analysis, wall frames, and tunnel frames. All the learning stages of each step of the cascade involve a self organized map. The final classification step consists of a support vector machine trained with frames previously labelled by the experts.

In the feature extraction step we used the same set of six features proposed by Vilariño *et al.* [21], [18], based on widely known image descriptors such as, mean intensity, Laplacian response, and statistical descriptors obtained from the concurrence matrices. In addition to this, we included two color descriptors based on the *color opponent* decomposition, using only the two chrominance components *R-G* and *B-Y* [16], [10]. These eight descriptors constitute the feature vector to be associated to each single frame. Since intestinal motility is a dynamic process, not only static information (frame) should be taken into account, but dynamic information also. For this reason, for each frame we construct a feature vector containing the descriptors of the 4 previous and 4 following frames, i.e., containing the descriptors of a whole sequence of 9 frames. The span of 9 frames is based on well known physiological outcomes which bound the minimum time interval between two intestinal contractions in about ten seconds [14].

The choice of the cascade system is underpinned by the fact that each step is designed in order to reject an amount of frames which mainly include images which are not valid for analysis, letting pass through the sequential stages those frames related to intestinal contractions. Finally, the last step of the cascade, consisting of the support vector machine classifier, 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 imbalanced ratio is shown to be an effective way of tackling the problem of classification in imbalanced data sets. Moreover, the modular shape of the system lets the expert identify new features in order to include them as a new filter stage, adding domain knowledge in a natural way.

Each step of the cascade is trained independently from the others. For each of them, a reduced set of features is utilized, and a self organized map is used to build up the training sets. Self Organizing Maps (SOMs) or Kohonen networks [12] are a specific type of neural networks which provide the possibility of reducing the

dimensionality of complex multivariate data sets, allowing their visualization in a 2-dimensional representation. By producing easily comprehensible low-dimensional maps of informative features, SOMs offer a technique for visual understanding and interpretation of hidden structures and correlations in the input dataset. Several works have been published referring medical applications using SOMs in a wide range of fieldworks, including classification of craniofacial growth patterns [13], extraction of information from electromyographic signals [5], magnetic resonance image segmentation [9], cytodiagnosis of breast carcinoma [22], classification of renal diseases [4], and drug design [2], among others. In our case, the use of SOMs is fully justified, since for each step of the cascade, a multidimensional representation of the frames is provided, but the selection of the training sets is strictly related to the visual patterns which the specialist identifies for each case. The SOM shows a clustering of the video sequences based on the information specifically needed for each step, grouping them into cells which correspond to similar descriptors. Each cell is identified by a representative element of the underlying sequences, which gives the specialist a first impression about the visual aspect of the cluster, and all the underlying sequences can be pictured for each cell. Now, the expert can select those cells which best adapt to the sought patterns in order to conform the optimal training sets.

3 System Implementation

3.1 Identification of the Region Not Valid for Analysis

The video frames not valid for analysis are constituted by frames presenting turbid liquid hindering the visualization of the whole image, frames with the camera focusing the intestinal wall, and frames with the camera focusing the intestinal lumen trough a time interval in which no intestinal motion is registered.

The motivation of the first category is straightforward: since the turbid liquid is hindering the visualization of the intestinal tube, no motility information can be inferred, since no intestinal contractions can be visualized. From a visual point of view, these frames are presented as a green-yellowish turbid bath which covers the whole field of view. The second category -frames focusing the intestinal wall- follows a similar reasoning: since the camera is focusing the intestinal wall (due to a lack of contractile activity or a transversal location in the intestinal tube), no intestinal contractions can be visualized. The last category, is related to a lack of motility activity, since the camera is focusing the gut as a tube, appearing as a steady tunnel, with no apparent motion in the intestinal walls.

Figures 2 a) and b) show the SOM resulting from the clustering of all the frames of two different videos, using only the two chrominance features described in Section 1. The number of cells and its shape has been set to a common value for visual comparison purposes. It can be seen how the cell structure is graded in color. After selecting in the SOM the cells corresponding to turbid frames, the feature space is partitioned into two different areas: turbid and not turbid. Figures 2 c) and d) show the two dimensional plots corresponding to the videos in Figures 2 a) and b). The horizontal axis represents the R-G chrominance

component, and the vertical axis the B-Y chrominance component. The green points correspond to the frames selected in the SOM as turbid, while blue points are related to the remaining frames, i.e., the frames which are valid for analysis. The red points represent contractions labelled by hand by the specialist for each video. It can be seen that some real contractions fall into the selected not valid for analysis area, resulting as missed contractions (false negatives) of the system.

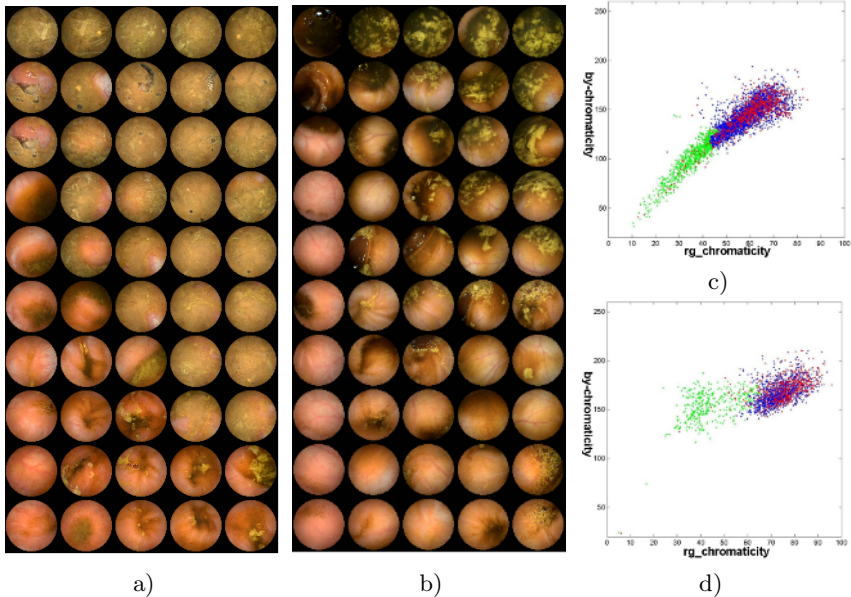


Fig. 2. a) and b): Two SOMs using color features for two different videos. c) and d): Chrominance plot for both videos.

Figure 3 represents a SOM for the wall detection using as features the number of pixels with positive value of the Laplacian filter, related to the size of the lumen (in order to deal with the size of the hole), and their overall sum, related to the contrast (in order to distinguish between real lumen and shadows). On the left side, we present a SOM where the different cells have been delimited with red lines for visualization purposes. On the right side, an example of a cell content is shown in a), with four video sequences which do not correspond to the wall paradigm, and in b), with seven sequences matching the wall paradigm. For this case, the specialist is to select those sets of cells which best fit the wall paradigm. The same procedure is applied for the case of tunnel sequences.

The classification step for turbid, wall and tunnel is performed with three different support vector machines (SVM) [19] with a radial basis function kernel and $\gamma = 0.1$ for turbid and $\gamma = 0.05$ for wall and tunnel. The choice of the kernel and the γ parameter was obtained in a heuristical way using an exhaustive analysis.

The output of the classification stages is used to construct interactive mosaics. These mosaics deploy the whole video in a sequential way from left to right and

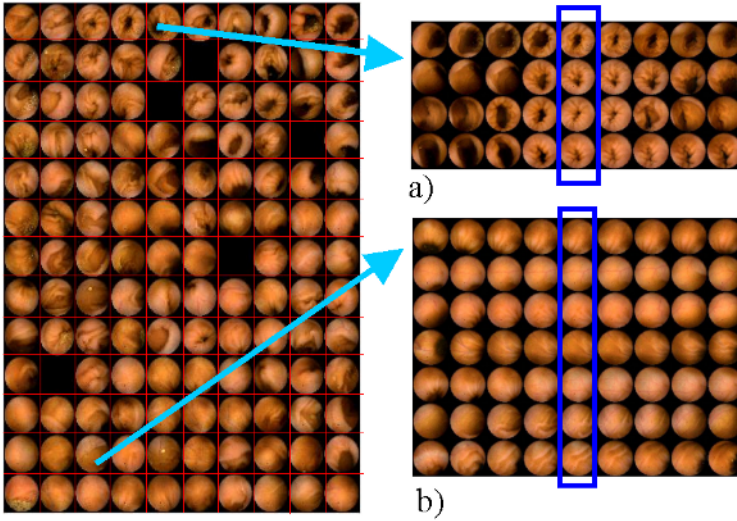


Fig. 3. a) Sequences underlying a cell which does not represent the wall paradigm. b) The wall paradigm is present in this cell. The central frame of each sequence is surrounded by the blue frame.

top to down, associating one pixel to each frame, and showing a different color depending on the the classification performed by the system. Figure 4 shows the interactive mosaics related to the videos in Figure 1. Each frame is assigned with a different color, so the specialist can visualize the valid area (in blue) directly and avoid visualizing the turbid (green), wall (beige) and tunnel frames (brown). For some cases, as it is straightforward to visualize from the mosaic of the video in Figure 4 b) the reduction in visualization time becomes drastic (about 50%).

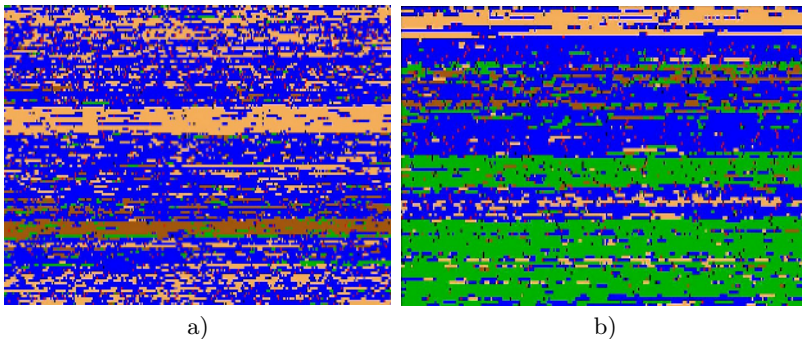


Fig. 4. a) and b): Two color mosaics. Green represents turbid, beige wall, brown tunnel, red labelled contractions and black labelled contract falling into the not valid for analysis area.

3.2 The SVM Classification Step

The aim of the final module of the cascade is to determine in an automatic way whether a frame is to be labelled as a contraction sequence. The input of this stage consists of the frames resulting valid for analysis, and its output consists of the frames which the system suggest as intestinal contractions. This last step is performed by a support vector machine with a kernel of radial basis functions and $\gamma = 0.1$, obtained in a heuristical way using exhaustive search. The feature vectors are constructed using all the existing features except the chrominance values used for turbid detection. The training set consists of a pool of all the previous labelled contractions from the remaining videos for the positives, and a random sample of the same number of non-contractions (under-sampling the negative class).

One of the main difficulties concerning the problem of intestinal motility assessment lies on the different patterns of intestinal contractions present in each video. From a physiological point of view the different patterns can be gathered into *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 of a closed lumen in an undefined number of frames. The analysis of the typology of contractions provides the specialist with helpful information for the evaluation of the presence of malfunctions. We used SOMs in order to provide the specialist with a useful tool for the validation of the different typology of intestinal contractions. In this case, the SOM is constructed with the output of the system. The feature set for the SOM is the same used for the final SVM classifier. Figure 5 a) Shows the SOM of the contraction frames corresponding to the video in Figure 2 a). Figure 5 b) Shows the sequences associated to the cell (9,1) (ninth row, second column) which are to be mainly associated with the pattern of phasic contractions. Figure 5 c) Shows the sequences associated with the cell (4,1), which are to be mainly associated with the pattern of tonic contractions.

4 Results

In order to validate our system, several considerations must be taken into account. On the one hand, the validation of the turbid, wall and tunnel detectors is hard to perform in a direct way, since the large number of frames in video (typically 50,000) makes the manual labelling of all of them not feasible. On the other hand, we must assume the loss of some frames labelled as contractions both in the intermediate stages of the cascade, and the final SVM classifier (these are the false negatives). In addition to this, we must also assume that the final SVM classifier identifies some frames as contractions in a wrong way (these are the false positives). One usual way of assessment for this kind of problems is tackled by using performance measures such as sensitivity and specificity. Sensitivity is defined as the number of the positives correctly detected over all the existing positives, while specificity is defined as the number of negatives correctly detected over all the existing negatives. Since for imbalanced problems, where

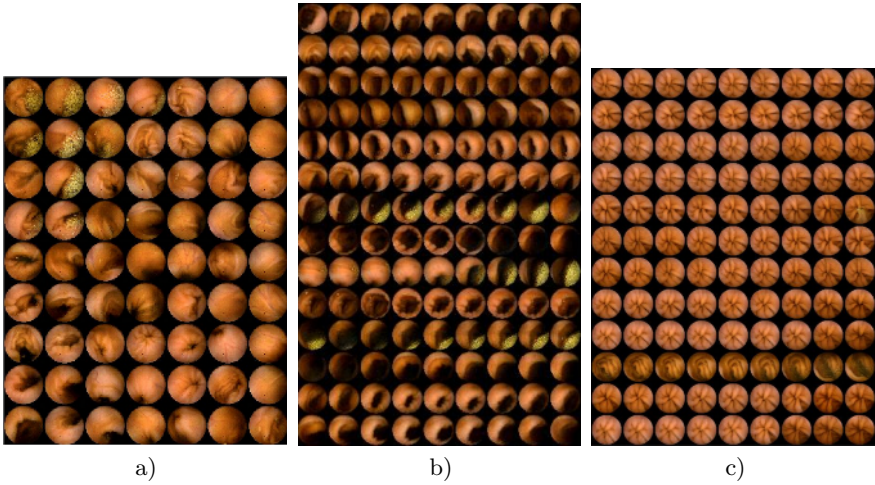


Fig. 5. a) SOM constructed exclusively with sequences of contractions provided by the system. b) Phasic contraction sequences underlying the cell (9,1). c) Tonic contraction sequences underlying the cell (4,1).

the negative samples outnumber the positives, specificity is not a useful metric, because it is generally close to 1. In our case we define a performance measure as false alarm ratio (FAR) as the number of wrongly classified frames over all the existing frames (false positives / (true positives + false negatives)). Any other performance measure, such as precision (true positives / (true positives + false positives)) could be used as well. For the specialist, the FAR as defined above adds information in the sense that It normalizes the number of wrongly detected frames by the number of existing contractions. In this sense, the FAR measure is not bounded as precision is, and may be greater than one. The pair sensitivity-FAR provides the operation point of our system.

The validation of the system was performed using ten different videos. For each video, the specialists were asked to visualize the whole video and to label all the existing contractions. Then, the specialist was asked to perform the classification stages for the turbid, wall and tunnel. The final SVM was trained for each video with all the contractions of the remaining videos and the same number of a random sample of negatives, in a leave-one-out strategy. The performance results are shown in Table 1.

The resulting overall sensitivity is 74.78 with $\sigma = 7.85$, and the overall FAR is 42.75 with $\sigma = 23.00$. The columns of sensitivity and FAR in Table 1 deserve a separate analysis. On the one hand, the values of sensitivity imply that our system is shown to be able to detect about the 75% of all existing intestinal contractions, with a pick of performance of 85.22% in Video 3. In the specific context of motility assessment, it may not be mandatory to achieve higher levels of accuracy, since the aim of the procedure is to reconstruct the pattern of intestinal contractions along the time. On the other hand, it is important to notice the high value of FAR for Video 10 -a video which has shown to present

Table 1. Performance results for 10 videos

Video	Sensitivity %	FAR%
Video 1	66.01	35.38
Video 2	65.25	25.35
Video 3	85.22	69.58
Video 4	73.32	17.78
Video 5	70.97	41.47
Video 6	75.35	32.36
Video 7	65.36	43.95
Video 8	82.07	37.56
Video 9	79.77	29.05
Video 10	84.52	95.01

a large amount of turbid frames-. If Video 10 is not taken into account, the overall FAR is reduced to 36.94 with $\sigma = 14.67$, while sensitivity does not suffer a significant variation. This means that, as an average, for each three existing contractions in the video, the system is to provide us with a false positive.

5 Conclusions

Intestinal motility assessment in video capsule endoscopy is a tedious task due to the large amount of time needed for the visualization of each studio, the variability of the images present in the video, and the low prevalence of intestinal contractions. We presented a novel approach based on a machine learning system using classical techniques, which overcomes these main drawbacks. The main contribution of our work lies on driving this useful but not feasible technique into a feasible clinical technique. We showed the efficiency of our system designed in a sequential way (cascade), in which each step is trained independently. We introduced a general framework based on self organized maps, which are used in the different training stages of the cascade, providing the specialist with unified way of including expert knowledge. Finally, we presented the classification results obtained by a support vector machine classifier, which represents, as far as we know, the most efficient approach present in the literature for this clinical field. In addition to this conclusions, we expect to continue working in several points for future pieces of research. The seek of optimal descriptors, the development of advanced techniques for dealing with the imbalanced ratio, and the investigation on more efficient classification techniques are some of our current lines of work.

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 FIS-G03/1085, FIS-PI031488, TIC2003-00654 and MI-1509/2005. The technology and methods embraced by this disclosure has been filed for patent protection.

References

1. A. Ali, J. M. Santisi, and J. Vargo. Video capsule endoscopy: A voyage beyond the end of the scope. *Cleveland Clinic Journal of Medicine*, 71(5):415–424, 2004.
2. S. Anzali and S. Gasteiger. The use of self-organizing neural networks in drug design. Perspectives. *Drug Disc. Des.*, 11:273–299, 1998.
3. TEC Assessments. Wireless capsule endoscopy. Technical Report 17, Blue Cross and Blue Shield Association, 2003.
4. W. Van Biesen and G. Sieben. Application of Kohonen neural networks for the non-morphological distinction between glomerular and tubular renal disease. *Neph. Dial Transp.*, 13:59–66, 1998.
5. C. I. Christodoulo and C.S. Pattichis. Unsupervised pattern recognition for the classification of EMG signals. *IEEE Trans. on Biom. Eng.*, 46(2):169–178, 1999.
6. R. Eliakim. Wireless capsule video endoscopy: three years of experience. *Wd. Journ. Gastro.*, 10:1238–1239, 2004.
7. Z. Fireman and A. Glukhovskiy. W. capsule endoscopy. *IMAJ*, 4:717–719, 2002.
8. Z. Fireman and E. Mahanja. Diagnosing small bowel crohns disease with wireless capsule endoscopy. *Gut*, 52:390–392, 2003.
9. J. O. Glass and W.E. Reddick. Hybrid artificial neural network segmentation and classification of dynamic contrast-enhanced MR imaging (DEMRI) of osteosarcoma. *Magn. Res. Im.*, 16(9):1075–1083, 1998.
10. R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall. 2nd. ed., 2002.
11. C. Gostout and D. G. Adler. State of the art, wireless capsule endoscopy. *Hospital Physician*, 39(5):14–22, 2003.
12. T. Kohonen. *Self-Organized Maps*. Springer, Heidelberg Berlin, 1995.
13. C. L. Lux and A. Atellzig. A neural network approach to the analysis and classification of human craniofacial growth. *Grow. Devel. Assign*, 62(3):95–106, 1998.
14. E. M. Quigley. Gastric and small intestinal motility in health and disease. *Gastroenterology Clinics of North America*, 25:113–145, 1996.
15. J. F. Rey. European society of gastroenterology. Guideline for video capsule endoscopy. *Endoscopy*, 36:656–658, 2004.
16. J. C. Russ. *The Image Processing Handbook*. IEEE Press. 2nd. ed., 1994.
17. K. Schulmann and S. Hollerbach. Feasibility and diagnostic utility of video capsule endoscopy for the detection of small bowel polyps in patients with hereditary polyposis syndromes. *Am. Jour. Gastro.*, 100(1):27, 2005.
18. P. Spyridonos, F. Vilariño, et al. Identification of intestinal motility events of capsule endoscopy video analysis. *LNCS.*, 3708:531–537, 2005.
19. V. Vapnik. *The nature of Statistical Learning Theory*. Springer Verlag, 1995.
20. F. Vilariño, L. Kuncheva, et al. ROC curves and video analysis optimization in intestinal capsule endoscopy. *Pat. Recog. Let.*, 27(8):875–881, 2006.
21. F. Vilariño, P. Spyridonos, et al. Experiments with SVM and stratified sampling with an imbalanced problem: Detection of intestinal contractions. *LNCS.*, 3687(2):783–791, 2005.
22. A.J. Walker and S.S. Cross. Visualisation of biomedical datasets by use of growing cell structure networks: A novel diagnostic classification technique. *Lancet*, 354:1518–1521, 1999.