TOPOGRAPHIC SEGMENTATION AND TRANSIT TIME ESTIMATION FOR ENDOSCOPIC CAPSULE EXAMS

M. Coimbra, P. Campos, J.P. Silva Cunha

IEETA, Department of Electronics and Telecommunications, University of Aveiro.
{miguel.coimbra,pcampos}@ieeta.pt, jcunha@det.ua.pt

ABSTRACT

The endoscopic capsule is a recent medical technology with important clinical benefits but suffering from a practical handicap: long exam annotation times. This paper shows how support vector machines can be used to segment the gastrointestinal tract into its four major topographic areas, allowing the automatic estimation of the clinically relevant gastric and intestinal transit times. According to medical specialists, this can reduce exam annotation times by up to 12%.

1. INTRODUCTION

The endoscopic capsule is the first autonomous micro-device to explore the human inner body of wide clinical application. This 11 by 30 mm device developed by researchers in Israel and the UK, includes a camera, a light source, RF transmitter and batteries. It is ingested by the patient and films the whole gastrointestinal tract during 6-8 hours, reaching places where conventional endoscopy is not capable of. The full system consists of the capsule itself, an external receiving antenna and a portable hard drive carried in the patient's belt. According to its distributor (Given Imaging, Israel), "over 230,000 patients worldwide have experienced the advantages of painless and effective PillCam™ Capsule Endoscopy" [1]. Articles in renowned scientific journals have shown the clinical importance of wireless capsule endoscopy, namely Iddan [2], Qureshi [3] and Ravens [4].

Currently, one of the main setbacks of this new technology is the long duration of the exam analysis task. A specialized doctor needs to view around 60,000 images such as the ones in Figure 1, looking for both abnormal situations (events) such as blood or ulcers, and defined topographic marks of the gastrointestinal tract (e.g. pylorum, ileo-cecal valve). This process, when performed by a trained specialist, can take about 2 hours. Ravens [4] comments that "the time a doctor needs to analyze the exam may be the most costly part of the procedure". There is a pressing need for automatic tools that can reduce these long annotation times. Even small time savings can be vital when multiplied by the 230,000 exams that, according to Given Imaging [1], have been performed worldwide.

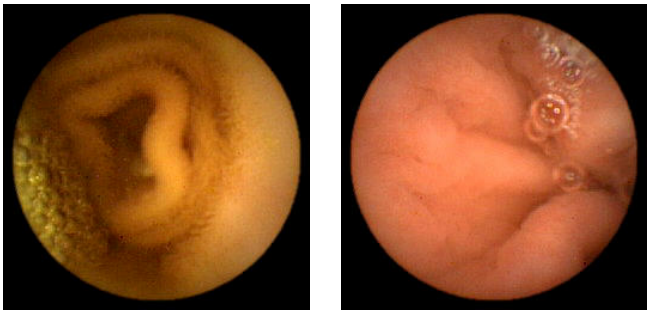

Figure 1 – Examples of endoscopic capsule images

The main contribution of this paper is the presentation of the first automatic tool that performs the topographic segmentation of the gastrointestinal tract for endoscopic capsule exams. According to doctors of Santo António General Hospital (www.hgsa.pt) in Portugal, responsible for over 100 capsule exams per year, this task takes approximately 15 minutes to complete by a specialist and allows the estimation of the clinically relevant gastric and intestinal transit times [5].

The automatic segmentation method presented is based on four support vector machine classifiers using MPEG-7 visual descriptors as feature vectors. Results are fitted to a four-section model obtaining the locations of the esogastric junction, pylorum and ileo-cecal valve. Methods are detailed in Section 2 and results presented in Section 3. Observations and conclusions are drawn in Section 4.

2. METHODS

Figure 2 – Topographic segmentation system.

Our proposed topographic segmentation system can be seen as a three stage solution. As shown in Figure 2, we have a training stage where support vector machine models are

obtained, a classification stage where each image is labeled as belonging to one of four topographic zones, and a final stage where results are fitted to a four section model representing the gastrointestinal tract.

Previous work by the authors has shown that it is possible to segment the gastrointestinal tract using simple classifiers based on Euclidean and Mahalanobis distances [6]. This paper improves such results by using a soft computing approach called *support vector machines (SVM)*. These formulate the learning problem as a quadratic optimization one whose error surface is free of local minima and has a global optimum. SVM approaches consist in first transforming the input data into a higher dimensional space using a kernel function, and then estimating an optimal hyperplane between the two classes that maximizes the margin of separation between them. For a detailed description of this method we refer to Burges' excellent tutorial [7]. The following SVM kernel functions *K(x,y)* were tested: linear (*Lin* - Equation 1), polynomial (*Poly* - Equation 2: *p*-polynomial degree), radial-base functions (*Rbf* - Equation 3: σ - Gaussian width) and sigmoid (*Sig* - Equation 4: *k*, δ - scale constants):

Lin: $$K(x,y) = x.y + 1 \quad (1)$$

Poly: $$K(x,y) = (x.y + 1)^p \quad (2)$$

Rbf: $$K(x,y) = e^{-\|x-y\|^2/2\sigma^2} \quad (3)$$

Sig: $$K(x,y) = \tanh(kx.y - \delta) \quad (4)$$

To reduce development time, we've decided to use MPEG-7 visual descriptors as our low-level features. These constitute a set of well known and studied features with available free source code that we could use. We were especially interested in color and texture features (please see Manjunath [8] for definitions), and used the reference software freely available at the website of the Institute for Integrated Systems of TU Munich. *(www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html)*. Instead of testing all descriptors, we've selected the most useful ones using previous relevance studies [9]. Following this, two are used for the experiments in this paper: *Scalable Color* (SC) and *Homogenous Texture* (HT). They not only obtained best results but have a complementary nature evaluating both color and texture. Three-fold cross validation was used in all experiments: data was divided into three sets where two were used for training and one for testing. Average results for all permutations were then obtained.

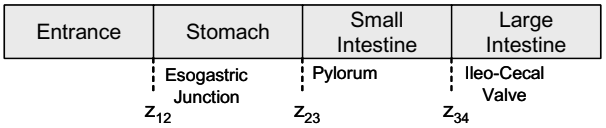

Figure 3 – Four-stage model for topographic segmentation.

The problem of topographic segmentation requires the estimation of three parameters: z_{12}, z_{23} and z_{34}, as shown in Figure 3. Before this is accomplished, we need to train four SVM classifiers: one for each topographic zone. Sets of 30,000 images per zone were randomly selected out of 60 different annotated exams. These positive and negative examples were used to train each classifier to estimate if an image belongs to a single topographic zone or not. Single image classification (Figure 2) consists in assigning a topographic location *TC* to each image as belonging to one of the four zones. This is accomplished by selecting the SVM classifier with the largest positive distance to its corresponding hyperplane. For performance analysis of this stage we've used typical accuracy and recall measures:

$$\text{Accuracy} = \frac{\text{Correct detections}}{\text{Total automatic detections}} \quad (5)$$

$$\text{Recall} = \frac{\text{Correct detections}}{\text{Total manual annotations}} \quad (6)$$

Assuming our three topographic parameters $z_{x,y}$ are estimated and that our individual topographic classification *TC* of image *t* has been obtained, we can define the single image classification error *E* for each image *t* as:

$$E(t) = \begin{cases} 0, \Leftarrow z_{x-1,x} < t < z_{x,x+1}, \text{where } x = TC(t) \\ 1, \text{otherwise} \end{cases} \quad (7)$$

This means we can define the total error TE of our parameter estimation as a linear combination of these individual errors for N images:

$$TE = \frac{1}{N}\sum_{t=1}^{N} E(t) \quad (8)$$

Iteration is used to minimize TE by consecutively varying one segmentation parameter in each cycle. These are initialized using the mean values of the corresponding training sub-set. Cross-validation is used to avoid over-fitting of results. Numerical quantification of the quality of the estimation is accomplished by measuring the segmentation error *SEz* for each parameter as the difference between an estimated parameter *z* and the manually annotated ground truth *z'*. Our total segmentation error SE is equal to the sum of the three individual errors.

$$SE_z = |z - z'| \quad (9)$$

$$SE = \sum_{x=1}^{3} |z_{x,x+1} - z'_{x,x+1}| \quad (10)$$

3. RESULTS

All results have been obtained on a Pentium 3,2GHz with 1GB RAM. The well-known SVM-light software package (http://www.cs.cornell.edu/People/tj/svm_light/) was used for SVM training. Segmentation results were obtained by a SVM-light module integrated into the CapView annotation software (www.capview.org), developed within our group. Each image has 256x256 resolution with 24 bits of color data per pixel and a full exam averages 60,000 images.

Figure 4 illustrates some results for the single image classification stage. *Scalable Color* clearly outperforms

Homogenous Texture, with best results obtained by SC using a polynomial kernel. Better results can be obtained by combining descriptors, reaching a maximum accuracy of 86% when using Gaussian kernels. Although higher accuracies are still desirable, one should notice that the 'low-pass' filter effect of the global model fitting stage rejects most of these outliers.

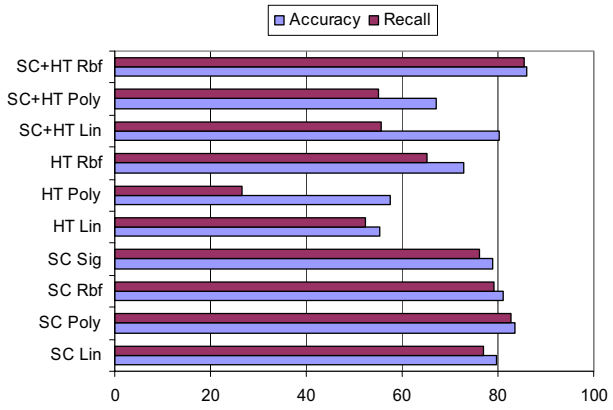

Figure 4 – Individual classification performance using various descriptors (SC-*Scalable Color*; HT-*Homogenous Texture*) and kernel functions (Equations 1-4).

We can now observe the behavior of these classifiers near zone boundaries and understand why the final segmentation obtains reasonably small errors. Figure 5 shows an example of a correct segmentation: blue represents the distance of each image to the Zone 1 *SC Poly* SVM hyperplane, and green the distance to Zone 2. The solid black line shows the manually annotated transition (esogastric junction). Although these distances have strong fluctuations, they usually produce good classifications (negative distances mean negative classification results) and there is a clear transition near the annotated boundary. As results will show, a model fitting stage such as the one described in Section 2 produces an accurate topographic segmentation.

Another example of a topographic segmentation can be seen in Figure 6. Careful observation shows a slight difference between the manually annotated position of the pylorum and the results from our automatic system. This reflects a small instability of the manual annotation itself: doctors tend to annotate the pylorum (and other topographic marks) when they see them. The automatic system is trained to detect variations in image characteristics (e.g. color, texture) so it annotates the moment that the capsule actually crosses the valve. Sometimes this difference is quite large since the capsule tends to 'bounce' several times on such valves before crossing them. This is especially serious in the ileo-cecal valve where the capsule can stay for more than 30 minutes. We are currently conducting clinical studies in cooperation with several hospitals and private clinics to measure these 'bouncing' times and the subsequent instability of the manual annotation.

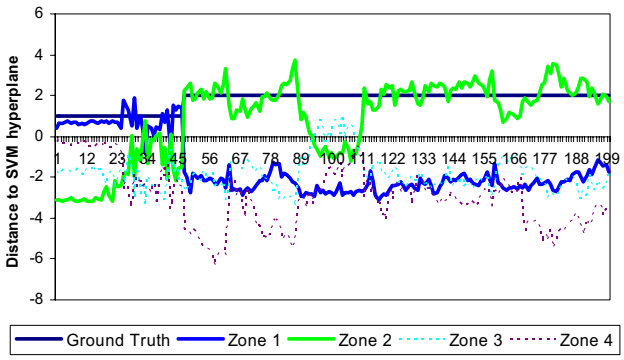

Figure 5 – Example of the behavior of Zone 1 and Zone 2 SC Poly classifiers near the esogastric junction.

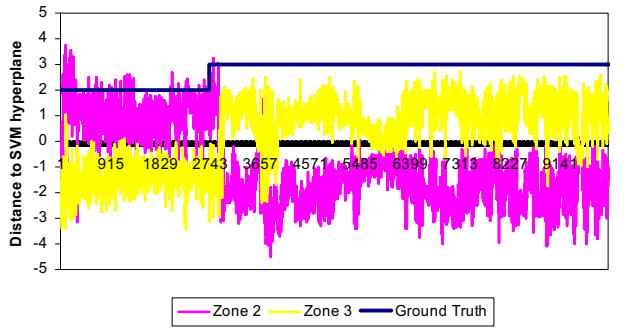

Figure 6 – Example of the behavior of Zone 2 and Zone 3 SC Poly classifiers near the pylorum.

	Median Error (images)			
	E(12)	E(23)	E(34)	Total
SC Lin	6.0	326.3	2689.3	3021.7
SC Poly	2.8	211.8	1070.7	1285.3
SC Rbf	9.2	704.3	1902.0	2615.5
SC Sig	8.0	205.0	1916.0	2129.0
HT Lin	47.7	5396.5	10322.0	15766.2
HT Poly	1212.3	23076.2	9676.0	33964.5
HT Rbf	15.5	647.0	1815.3	2477.8
SC+HT Lin	2379.0	1386.2	2046.8	5812.0
SC+HT Poly	11.0	10006.8	7913.7	17931.5
SC+HT Rbf	11.5	295.8	1322.8	1630.2

Table 1 – Median errors of the topographic segmentation.

For numerical assessment of the quality of the topographic segmentation we've used median SE_z and SE errors (Equations 9-10) instead of the more obvious mean error. The reason for this choice is that segmentation outliers tend to be rather drastic, placing all junctions at the beginning or end of the exam and thus generating very large mean errors. Since a single drastic error (which is easily detected

anyway) might mask all the smaller segmentation errors, we've found the median error to be much more informative for evaluating and therefore improving our system. Table 1 and Figure 7 summarize results for all descriptors and kernels.

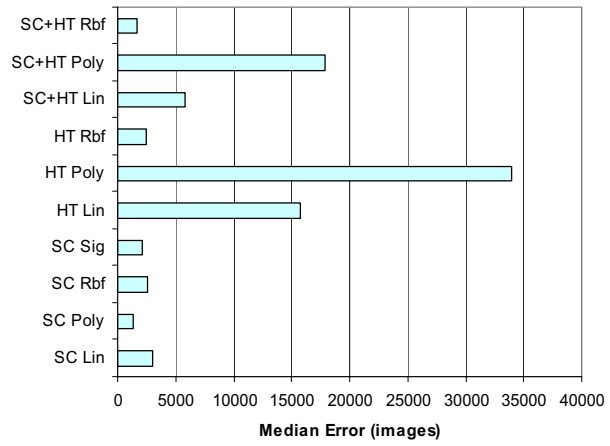

Figure 7 – Total median segmentation errors for various descriptors and kernel functions.

Following the results of the single image classification stage, best results are obtained by the *SC Poly* and *SC+HT Rbf* classifiers. Although the latter exhibits slightly better individual classification accuracy, it proves slightly worse in detecting the location of the ileo-cecal valve (E_{34} value in Table 1) and thus a slightly higher total median error. This directly affects transit time estimation where *SC Poly* is clearly the best classifier obtaining relative errors of 8% and 6% for gastric and intestinal transit times as seen in Table 2.

Transit Time Median Error	Zone 2		Zone 3	
	Total (m)	Relative	Total (m)	Relative
SC Lin	2.0	11.8 %	34.5	12.6 %
SC Poly	1.4	8.0 %	14.6	5.9 %
SC Rbf	3.4	15.2 %	27.2	10.9 %
SC Sig	2.9	14.6 %	28.8	11.4 %
HT Lin	38.3	100.0 %	168.6	68.5 %
HT Poly	217.1	479.6 %	206.9	100.0 %
HT Rbf	5.4	21.7 %	33.8	12.6 %
SC+HT Lin	10.6	75.0 %	24.0	11.6 %
SC+HT Poly	18.2	86.5 %	137.4	56.1%
SC+HT Rbf	3.3	12.4 %	22.3	9.2 %

Table 2 – Gastric and intestinal transit time errors.

4. DISCUSSION

A full topographic segmentation system for endoscopic capsule exams has been presented. This is accomplished using support vector machine classifiers that label each image as belonging to one of four zones, followed by a global model fitting stage that estimates zone transitions by minimizing an error function. Best results were obtained using the MPEG-7 *Scalable Color* descriptor as the feature vector of SVM classifiers with polynomial kernels. A total median error of 1285 images (out of 60,000) was obtained generating relative errors smaller than 10% for transit time estimation.

Future work will expand this methodology to detect abnormal events in capsule endoscopy exams, thus reducing annotation times even further.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. José Soares of the gastroenterology department of Santo António General Hospital in Porto, Portugal for providing all the anonymous data that has made this work possible and for contributions regarding the medical importance of capsule endoscopy. We would also like to thank the IEETA institute and the Fundação para a Ciência e Tecnologia for their vital support (SFRH/BPD/ 20479/2004/YPH2).

REFERENCES

[1] Given Imaging Home Page - www.givenimaging.com.

[2] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless Capsule Endoscopy", in *Nature*, pp. 405-417, (2000)..

[3] W.A. Qureshi, "Current and future applications of the capsule camera", in *Nature*, vol.3, pp. 447-450, (2004).

[4] A.F. Ravens, C.P. Swain, "The wireless capsule: new light in the darkness", in *Digestive Diseases*, vol. 20, pp. 127-133, (2002).

[5] M.N. Appleyard, A. Glukhovsky, J. Jacob, D. Gat, S. Lewkowicz, and P. Swain, "Transit times of the wireless capsule endoscope", in *Gastrointest. Endosc*., Vol. 53, AB122, (2001).

[6] M. Coimbra, P. Campos, and J.P. Silva Cunha, "Extracting clinical information from endoscopic capsule exams using mpeg-7 visual descriptors", in *IEE EWIMT 2005*, London, Uk, (2005).

[7] C.J. Burges, "A tutorial on support vector machines for pattern recognition", in *Knowledge Discovery Data Mining*, vol.2, no.2, 1998, pp.1-43.

[8] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, and A. Yamada, "Color and texture descriptors", in *Special Issue on MPEG-7, Trans. Circ. Syst. for Video Tech.*, vol. 11/6, pp. 703-715, (2001).

[9] M. Coimbra, and J.P. Silva Cunha "MPEG-7 Visual Descriptors – Contributions for Automated Feature Extraction in Capsule Endoscopy", second round of reviews for *IEEE Trans. Circuits and Systems for Video Technology*.