Combining Color with Spatial and Temporal Position of the Endoscopic Capsule for Improved Topographic Classification and Segmentation

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

***Abstract*—Capsule endoscopy is a recent technology with a clear need for automatic tools that reduce the long exam annotation times of exams. We have previously developed a topographic segmentation method, which is now improved by using spatial and temporal position information. Two approaches are studied: using this information as a confidence measure for our previous segmentation method, and direct integrating of this data into the image classification process. These allow us not only to automatically know when we have obtained results with error magnitudes close to human errors, but also to reduce these automatic errors to much lower values. All the developed methods have been integrated in the CapView annotation software, currently used for clinical practice in hospitals responsible for over 250 capsule exams per year, and where we estimate that the two hour annotation times are reduced by around 15 minutes.**

***Index Terms*— Endoscopic capsule, image classification, biomedical engineering, medical imaging**

I. Introduction

The clinical importance of the endoscopic capsule is now solidly established in literature: Iddan [1], Oureshi [2], etc. Due to space limitations, we refer to our previous work [3,4], for more extensive capsule details and clinical importance information. All this attempts to solve an important limitation of the endoscopic capsule, excessively long annotation times. Currently it takes about 2 hours to fully view, annotate an exam and write its corresponding report. Our clinical studies show that the task of topographic segmentation is both difficult (the median error performed by three senior capsule specialists was about 400 images) and time-consuming (around 15 minutes can be saved by automation).

The main contribution of this paper is the improvement of our previous topographic segmentation methods using color and texture, by incorporating not only temporal but also spatial position information in the image classification process.

Manuscript received July 11, 2006. This work was supported in part by the IEETA institute and the Fundação para a Ciência e Tecnologia (grant nr. SFRH/BPD/ 20479/2004/YPH2). The authors would like to thank Dr. José Soares of the gastroenterology department of Santo António General Hospital in Porto, Portugal (www.hgsa.pt) for providing and annotating all the anonymous data that has made this work possible and for contributions regarding the medical importance of capsule endoscopy.

M. Coimbra, J. Kustra, P. Campos are with the IEETA institute, Campus Universitário de Santiago, 3810-193 Aveiro - Portugal (email: {miguel.coimbra, jacek,pcampos}@ieeta.pt).

J.P. Silva Cunha is with the Departments of Electronics, Telecommunications and Informatics of the University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro – Portugal (email: jcunha@det.ua.pt).

II. Methods

The ultimate objective of the presented methods is to reliably divide the video of the gastrointestinal tract into its 4 constituent parts (entrance, stomach, small intestine, large intestine) and thus determine its corresponding junctions (eso-gastric junction, pylorus, ileo-cecal valve).

A. Capsule Position and Velocity

We can theoretically estimate the spatial position of a capsule via antenna signal triangulation. We have selected 47 capsule exams where a clinical specialist manually annotated the temporal location of the pylorus (t_{PYL}) and of the ICV (t_{ICV}) in the video using the CapView annotation software. We've then used our automatic topographic segmentation algorithm to determine these same temporal locations. Using our 2D position information, we can then obtain the corresponding spatial locations: x_{PYL}, y_{PYL}, x_{VIC}, y_{VIC} etc. For comparison purposes, these were normalized. Besides analysing 2D position information, we have looked at average capsule displacement velocity (module of the displacement vector between two points with temporal references t and $t+1$).

B. Topographic Segmentation Algorithm

Our previously developed automatic topographic segmentation method, from now on referred as *TSA*, is described in Coimbra [4].

C. Spatial Information as a Confidence Measure

Two high-confidence areas were defined, one for the pylorus and another for the *ILC*. We have measured the median segmentation error SE_z for all marks (z_{12} - eso-gastric junction; z_{23} – pylorus; z_{34} – ileo-cecal valve), and for all exams *SE*, whose junctions are inside and outside these areas,

and results presented in section 3 have showed that this information is indeed useful as a confidence measure for automatic segmentation results.

D. *Integrating Spatial and Temporal Information for Classification*

An alternative way of using this information is to use it directly for individual image classification. Our previous method trained 4 SVM classifiers, one for each zone, which determine the topographic section each image belongs to as the classifier with the highest positive distance to the SVM hyperplane (see [4] for details). We can however, use these distances to build a feature vector for each image, along with additional information such as spatial and temporal location. Our new feature vector F is now defined as:

$$\vec{F} = [x, y, Z_1, Z_2, Z_3, Z_4, V, t] \quad (1)$$

where x and y are the normalized spatial location coordinates (1,2), Z_1, Z_2 Z_3 Z_4 are the SVM classifier results [4] (distances to SVM hyperplanes), V is the spatial velocity, and t the temporal location in number of frames. The combination of these different features into a single vector requires that all coefficients are previously normalized.

A variety of well-known distances was used for classification (L1 Norm, Euclidean, Mahalanobis). Finally, we have measured the relevance of each coefficient for the segmentation process using a step-wise elimination analysis.

III. RESULTS

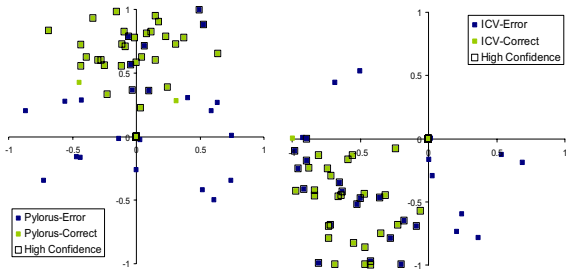

Fig. 1. Spatial distribution of correct (green) and incorrect (blue) estimations. Points in high confidence areas are highlighted with a black bounding box. We can observe that most correct detections fall into high-confidence areas while incorrect ones are more distributed over the whole 2D space.

Table 1. Numerical analysis of the spatial distribution of automatic topographic estimations. Accuracy = correct estimations / total estimations; recall = correct estimations / total annotations; mean and median segmentation errors are given in number of images.

	Pylorus		**ICV**	
	Accuracy	**Recall**	**Accuracy**	**Recall**
Correct	80 %	93 %	58 %	96 %
Incorrect	89 %	70 %	92 %	39 %
	Mean Err.	**Median Err.**	**Mean Err.**	**Median Err.**
All	2157	158	3966	844
High-confidence	493	45	2096	246

An analysis of Table 1 and Figure 1 shows that high-confidence areas contain almost all correct estimations, and low-confidence areas mainly contain incorrect estimations.

Table 2. Segmentation results using various distances and classifiers for feature vector *F*. L1, L2, and Full Multivariate were previously defined. Max Z corresponds to our previously used classification method [7], which is the maximum positive distance to the SVM hyperplanes. Finally, we use Mahalanobis distances on a reduced feature vector F = [Z_1, Z_2, Z_3, Z_4] – Multivariate Color.

	Accuracy	**SE**	**SE**-EGJ	**SE**-PYL	**SE**-ICV
L1	82.2%	2285	7	50	2228
L2	83.1%	1730	5	23	1702
Max Z	79.4%	3063	5	16	3042
Multivariate Color	77.4%	1052	5	22	1025
Full Multivariate	79.7%	2285	6	433	1846

Table 3. SE values as coefficients are removed from the feature vector F in a step-wise elimination process. In each step we eliminate the coefficient that produces the minimum SE when removed from the vector. These areas are marked in light grey in the table. The corresponding individual classification accuracy is presented instead. Discrepancies with are highlighted in dark grey.

	Median Segmentation Error						
Maximum	83.6%	83.1%	83.5%	82.8%	82.2%	82.1%	79.2%
x	82.3%	82.3%	83.1%	82.8%	82.2%	82.1%	79.2%
y	83.6%	83.1%	83.1%	82.8%	82.2%	82.1%	79.2%
Z1	81.8%	81.2%	81.7%	80.8%	79.9%	79.2%	79.2%
Z2	81.9%	81.0%	82.8%	82.8%	82.2%	82.1%	79.2%
Z3	80.9%	82.6%	83.4%	79.3%	70.3%	69.8%	68.7%
Z4	80.8%	82.5%	83.5%	82.2%	82.2%	82.1%	79.2%
V	83.2%	82.4%	83.1%	82.8%	82.1%	82.1%	79.2%
t	80.4%	79.9%	79.9%	79.5%	66.4%	63.0%	54.2%

IV. DISCUSSION

Results show that doctors can trust that automatic segmentation errors in high-confidence areas are as low as human ones. Including other information has allowed us to improve segmentation results significantly. Step-wise elimination analysis has shown us that the most relevant features for segmentation are capsule temporal position, and the color recognition of the entrance and the small intestine topographic sections. It has also shown us that spatial location is not a relevant factor for individual image classification.

REFERENCES

[1] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless Capsule Endoscopy", in Nature, 2000, 405, pp. 417.

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

[3] 4. M. Coimbra, and J.P. Silva Cunha, "MPEG-7 visual descriptors – Contributions for automated feature extraction in capsule endoscopy", in IEEE Transaction on Circuits and Systems for Video Processing, vol. 16/5, 2006, pp. 628-637.

[4] 6. M. Coimbra, P. Campos, and J.P. Silva Cunha, "Topographic Segmentation and Transit Time Estimation for Endoscopic Capsule Exams", in Proc. of IEEE ICASSP 2006, Toulouse, France, 2006.