35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Exploiting Gastrointestinal Anatomy for Organ Classification in Capsule Endoscopy using Locality Preserving Projections

*Carl Azzopardi, *Member,* Yulia A. Hicks, *Member,* and Kenneth P. Camilleri, *Senior Member, IEEE*

***Abstract* – Capsule Endoscopy is a technique designed to wirelessly image the small intestine within the gastrointestinal (GI) tract. Its main drawback is the vast amount of images it generates per patient, necessitating long screening sessions by the clinician. Previous studies have proposed to partially facilitate this process by automatically segmenting the GI tract into its constituent organs, thus identifying the region of interest. In this work, we propose to exploit the anatomical structure of the GI tract when carrying out dimensionality reduction on visual feature vectors that describe the capsule images. To this end, we suggest a novel adaptation of a technique called Locality Preserving Projections, and results show that this achieves an improved performance in organ classification and segmentation, at no additional computational or memory cost.**

I. Introduction

Wireless Capsule Endoscopy (WCE) was introduced by Iddan *et al*. as a non-invasive medical procedure which allows physicians to examine the small intestine within the gastrointestinal (GI) tract [1]. The procedure simply involves the patient swallowing a small capsule, such as that shown in Figure 1, which then travels down the GI tract by peristalsis [2]. The capsule, equipped with a battery powered camera and two miniature LED light-sources, takes two to four pictures per second and transmits them wirelessly to a belt strapped to the waist of the patient. A collection of image frames showing the internal structures of the small intestine are thus collated into a video, and the physician then examines these to diagnose any present pathologies such as internal bleeding, Crohn's Disease, ulcers or polyps [3].

Over the course of the eight-hour examination, the capsule generates approximately 50,000 images [4], which need to be examined one by one by the physician. Needless to say, automation in this regard would be greatly appreciated, especially on two particular fronts. Firstly, a quick and automatic discrimination between organs in the gastrointestinal tract would allow the physician to home in on the organ of interest. Secondly, automatically identifying

The research work disclosed in this publication is funded by the Strategic Educational Pathways Scholarship (Malta). The scholarship is part-financed by the European Union – European Social Fund (ESF) under Operational Programme II – Cohesion Policy 2007-2013, "Empowering People for More Jobs and a Better Quality of Life". The publication itself is funded by the Institute of Medical Engineering and Medical Physics, School of Engineering, Cardiff University, and by the Centre for Biomedical Cybernetics, University of Malta.

*Carl Azzopardi and Yulia. A. Hicks are with the Institute of Medical Engineering and Medical Physics, School of Engineering, Cardiff University, Wales, (e-mail: AzzopardiC@cardiff.ac.uk; HicksYA@cardiff.ac.uk).

Kenneth P. Camilleri is with the Centre for Biomedical Cybernetics, University of Malta, Malta, (e-mail: kenneth.camilleri@um.edu.mt).

Figure 1 - The capsule used for the procedure. Source: [7]

images that are suspect positives for pathologies reduces the chances of physicians missing out on important details, and further speeds up the process of diagnosis.

Physicians presently spend roughly 2 hours to examine capsule endoscopy videos [5], and the software solutions supplied by manufacturers provide only a very basic assistive feature to identify frames having increased redness as suspect candidates of internal bleeding [6]. Previous studies have proposed image processing techniques to facilitate the screening process, both in terms of GI organ segmentation and pathology detection. Unfortunately however, such techniques have not yet found their way onto physicians' computers in the form of commercially available software.

When considering GI organ segmentation in particular, studies by Mackiewicz *et al.* [3] and by Cunha *et al.* [8] have shown that algorithms based on feature extraction and classification achieve a good level of accuracy when classifying images. The same authors also propose boundary detection methods based on local topographic features, which locate the junctions between the stomach and small intestine (pylorus valve) and between the small intestine and colon (ileo-cecal valve). Their reported results were also satisfactory, although the authors confirm that there is room for improvement [3] [8], particularly since physicians were noted to be more accurate (by up to ~ 1500 frames) than their automated counterparts at distinguishing between organs.

Studies such as those by Mackiewicz *et al.* and Cunha *et al.* base their classification primarily on the local *topographic* features (such as edges and colour distributions) of the capsule images. While such methods certainly yield promising results, they do not exploit the anatomical topology of the gastrointestinal tract. In this work, we postulate that this topology imposes an intrinsic ordering on

the capsule images, and that there may be benefits in exploiting this. The method known as Locality Preserving Projections (LPP) provides a mechanism to encode adjacency relations between data [9], and may therefore be adapted to encode the physical adjacency between the capsule images. Such adjacency would reflect the anatomical topology and feature-based similarities that are present between them, and LPP may then be used to reduce the feature dimensionality while preserving this encoded local structure. Our results indicate that taking advantage of the information encoding the anatomical structure of the GI tract leads to an improved accuracy of classification, and in improved boundary detection, regardless of the type of feature descriptor tested.

The rest of this paper is organized as follows. Section II describes the visual features extracted, as well as the proposed method for dimensionality reduction. It then describes the methods employed for image classification and segmentation. Section III will present and discuss the results obtained for classification and segmentation, particularly when compared to those of previous studies. Section IV will then provide some concluding thoughts.

II. Methodology

A dataset of six patient videos, or approximately 360,000 capsule endoscopy images, were available for this study. The videos were split into three sub-groups in order to implement three-cross-fold validation and thus to maximize the significance of the results. In each experiment therefore, two sub-groups were used for training, and the third was used for testing. For *classification* tests, random subsets of 1200 images were taken for each of the stomach, small intestine and colon and used for training. Similarly, random subsets of images taken from two organs in the third sub-set were then collated into one clip for testing. For *boundary detection* tests, the same training sets were used, and short clips showing the capsule traversing the junction of interest were then taken from the third subgroup and used for testing.

A. Feature Extraction

This work proposes to use LPP to reduce the dimensionality of data by exploiting the anatomical topology of the GI tract. It is intended as a direct replacement for principal component analysis (PCA), a standard technique which was used by Mackiewicz *et al.* for dimensionality reduction and data compression in [3]. To this end, the features chosen for data representation in our work are similar to those used in [3], and a direct comparison of dimensionality reduction techniques was carried out by simply replacing the use of PCA with that of LPP.

Hue-Saturation Histograms

Hue-Saturation (HS) histograms are colour-based features which were first applied to capsule endoscopy by Berens *et al.* in [10]. HS histograms primarily exploit the difference in hue between stomach, small intestine and colon images, which varies between pink and green respectively for these organs. Images are first converted from the RGB to the HSI domain, and intensity information is then immediately discarded, since this varies significantly in relation to the capsule's distance from the intestinal lumen. A joint 2D

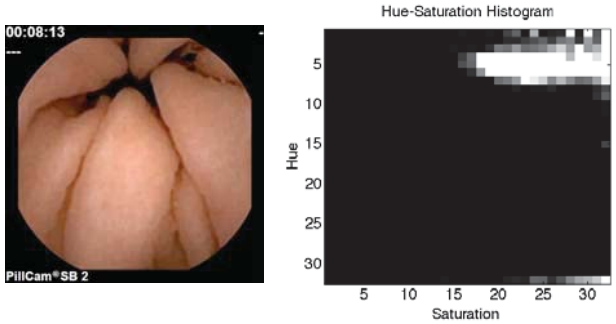

Figure 2 – [Left] A typical capsule image taken from the stomach. [Right] The corresponding HS histogram for the said image.

histogram is then constructed for each image by plotting 32 bins for hue against 32 bins for saturation, as shown in Figure 2. The histogram structure is then compressed using the Discrete Cosine Transform, and the first 136 coefficients corresponding to the lower frequencies are retained for later dimensionality reduction and classification.

Local Binary Patterns

Local Binary Patterns is a technique used for texture analysis and classification. It evaluates the local variation around a particular pixel in the form of binary number. To calculate the LBP value for a particular pixel p, its N neighbours, n_i, $i = 0\ldots, N-1$, are compared to its own value, to establish whether they are greater or smaller than p [3]. Each n_i is therefore mapped onto a value b_i as follows:

$$b_i = \begin{cases} 1, & \text{if } n_i \geq p \\ 0, & \text{if } n_i \leq p \end{cases} \quad (1)$$

The LBP value for the said pixel is then formed as an N-bit binary number by concatenating all binary values b_i together as follows:

$$LBP(x_0, y_0) = \sum_{i=0}^{N-1} b_i 2^i \quad (2)$$

LBP values may be plotted to display textural features such as edges, as shown in Figure 3. However, a 7-bin histogram of the LBP values is typically used as a feature vector to characterize the textural content for a particular image [3]. In [11], Connah and Finlayson extended the technique to consider colour *and* texture, by separately applying LBP on all the colour channels and then combining their LBP histograms into a 1D or 3D feature vector. Both types of histograms were tested by Mackiewicz *et al.* in [3], though in this work, our chosen feature vectors are 1D histograms made of 21 bins (7 bins x 3 colour channels), due to the latter performing better on our dataset.

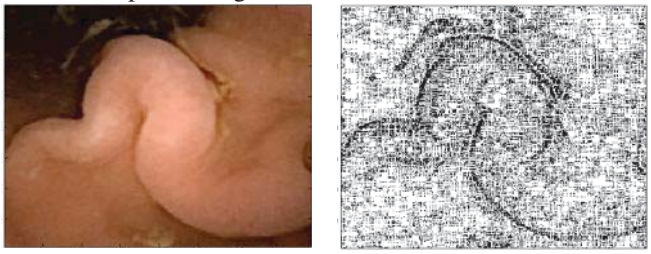

Figure 3 – [Left] Original colour image containing red, green and blue components; [Right] Textural features extracted from red component after application of LBP operator.

B. Dimensionality Reduction

Following extraction of descriptive features from the images in the form of HS and LBP histograms, Mackiewicz *et al.* applied dimensionality reduction using PCA to the feature vectors in each case [3]. Dimensionality reduction has a two-fold effect; firstly, it allows for compression of the data, since it permits redundant variables that contain uninformative information to be identified and discarded. Secondly, it provides an alternative representation of the data in the hyperspace, and this may further accentuate the differences between images from one organ and another. The end result could then be an improved accuracy for the classification algorithm. Thus, similarly to [3], we applied PCA to both HS and LBP histograms, and to combinations of both, prior to classification. The results formed a baseline against which to compare our proposed technique.

Locality Preserving Projections

Locality preserving projections is a technique which may be considered similar, and thus, an alternative to PCA. The important difference however is that LPP optimally preserves the local neighbourhood structure of the data set through the creation of linear projective maps [9]. Similarly to PCA, it attempts to solve the problem of dimensionality reduction by finding a transformation matrix $\mathbf{A}$ that maps a set of m points, $\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_m$, in $\mathbf{R}^n$ to another set $\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_m$, in $\mathbf{R}^l$, whereby $l << m$ [9]. The local neighbourhood structure of the data is defined *a-priori* by constructing an adjacency graph, where two nodes are connected by an edge if they are close to each other. This 'proximity' may be defined in standard ways such as ε-neighbourhoods or k-nearest neighbours [9]. Alternatively however, one might define a more utilitarian form of adjacency, which bests suits the data at hand.

In our work, we take a heuristic approach to constructing the adjacency matrix, and exploit the anatomical topology of the gastrointestinal tract. Consider for a moment, the various capsule images belonging to different individual organs, and thus sharing 'global' features. Taking colour for instance, previous studies have shown that stomach images have a common pinkish hue; small-intestine images share a pinkish-yellow hue, whereas colon images share a greenish hue. One might therefore propose that images belonging to a particular organ have similar global characteristics, and thus share a *class adjacency*. Similarly then, if one considers a brief time-window of images, representing a short distance being traversed by the capsule, one notices that these would share similar local characteristics since the change in scene between one image and the next would be minimal. Thus we might propose that such images share a certain *temporal adjacency*. These concepts are aptly depicted in Figure 4.

We propose to describe class adjacency and temporal adjacency mathematically, by defining two separate and symmetric $N \times N$ square matrices, W_{CA} and W_{TA}. Multiplying these matrices together yields the adjacency matrix, which itself defines which nodes (or capsule images) in the graph are connected by an edge, and thus 'neighbours'.

$$W_{CA}(i,j) = \begin{cases} 1 & \text{if } [1 \le i \le N/2] \text{ and } [1 \le j \le N/2] \\ 1 & \text{if } [N/2 < i \le N] \text{ and } [N/2 < j \le N] \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

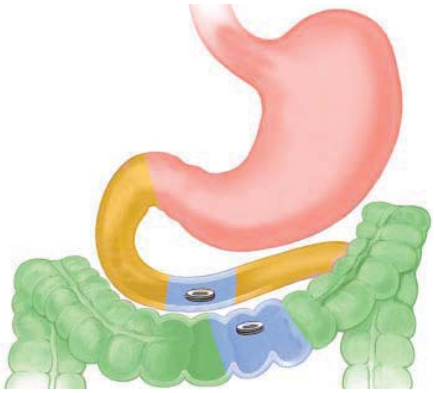

Figure 4 – Depiction of class adjacency within organs (red, yellow and green sections), and temporal adjacency as the capsule travels through the said organs (blue 'time-windows'). Source: Modified from [13].

$$W_{TA[\sigma]}(i,j) = \begin{cases} 1 & \text{if } |i-j| \le \sigma \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In equations (3) and (4), i and j are the matrix row and column indices, while σ is the temporal adjacency window size. N is the size of the square matrix, and also the number of images making up a training set. Both matrices are multiplied together, and weighted using a standard heat-kernel matrix. This assigns the graph edges a weight, based on the similarity of content between two different images.

C. Image Classification and Boundary Detection

Classification was implemented using a Support Vector Machine (SVM) having a radial basis kernel, as in previous studies [3], [4], [10]. As described in Section I, three-cross-fold validation was carried out, and in each run, sequential feature selection was used to identify the best features from the reduced hyperspace returned by the dimensionality reduction techniques. Boundary detection was implemented using the method devised by Igual *et al.* in [14]. The labelled sequence returned by the classifier was compared to a step function, and the best fit gave an indication of the likely position of the boundary between two organs.

III. Experimental Results

The results for image classification are shown in Table 1, whereby the accuracies obtained when describing images using HS and LBP histograms and combinations thereof, are shown. The table compares the performances obtained when using PCA and LPP to reduce the dimensionality of the feature vectors prior to classification. For LPP, the results show a consistent increase in accuracy when classifying images across a stomach – small intestine interface, and a very similar accuracy when classifying images across a small intestine – colon interface. Such an improvement is attributed to LPP being able to achieve tighter and more distinct clusters for different classes, thus providing a discriminating advantage to the classifier that follows. Experimental results further suggest that the improvement in classification accuracy is achieved using only a few dimensions. Figure 5 shows such an example for LBP, whereby the improvement in accuracy is noticed to peak and remain roughly constant after the second or third dimension.

Table 1 – Accuracy of Image Classification (%)

	Stomach – Small Intestine		Small Intestine - Colon	
Feature	**PCA**	**LPP**	**PCA**	**LPP**
HS	90.28	92.14	99.33	97.42
LBP	67.42	86.33	58.98	94.42
HS_LBP	90.47	92.58	99.33	97.86

Table 2 – Deviation in Boundary Detection (Mean and Median Error)

	Pylorus Valve				Ileo-cecal Valve			
Feature	**Mean**		**Median**		**Mean**		**Median**	
	PCA	LPP	PCA	LPP	PCA	LPP	PCA	LPP
HS	186	68	215	81	13	27	8	31
LBP	168	182	119	185	243	23	278	13
HS _LBP	148	112	134	90	101	36	3	29

Locality preserving projections also shows an improvement in the trials for boundary detection, since it led to smaller mean and median deviations from the expert-annotated boundaries when compared to principal component analysis, as shown in Table 2. The use of LPP showed an improved performance specifically for the pylorus valve, whereas similar performance was achieved for the ileo-cecal valve. This increased performance may likely be attributed to the increased accuracy achieved with dimensionality reduction prior to classification, since the improvements appear to closely mirror those of the first stage image classification trials, where LPP outperformed PCA.

When comparing the results obtained from both the image classification and boundary detection trials to the results reported by Mackiewicz *et al*. in [3], we note that there is general agreement in the image classification trials, particularly for the HS histograms and their combination with LBP histograms, since the accuracies obtained were in excess of 90%. Considerable variation was however noted when comparing the accuracies obtained with 1D LBP histograms on their own, as a difference of over 20% was present when using PCA for dimensionality reduction. In view of this, one should question whether some textural bias is present in the results, potentially derived from the use of this particular, much smaller data set.

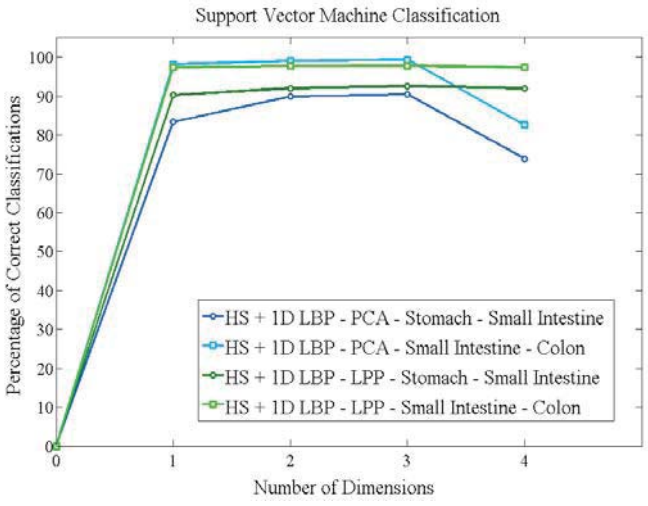

Figure 5 – Classification accuracy of an SVM classifier, after dimensionality reduction of HS + 1D LBP feature vectors is carried out with PCA and LPP. Dark green graph: LPP window size = 7. Light green: LPP window size = 2.

IV. Conclusions

LPP may indeed appear to be a suitable candidate which could replace the use of PCA for dimensionality reduction of visual descriptors in the field of capsule endoscopy. This study has shown that the anatomical topology of the GI tract appears to contain important information that may be taken advantage of within the framework of LPP. The use of an adapted adjacency matrix that takes into account the class adjacency and temporal adjacency of capsule images, seems to have improved the accuracy of image classification and reduced the error in boundary detection. Furthermore, in comparison to PCA, the said improvements appear to come at no additional memory or computational cost, since the number of dimensions required to maximize classification accuracy appear to be similar to those of PCA. However, further validation of the test results should take place by applying the proposed technique on larger data sets, in order to test robustness against larger amounts of data variability.

Acknowledgements

The authors thank Saint James Hospital for providing copies of capsule data. The data received for processing at Cardiff University was already in anonymized form and was deemed by the Institutional Review Board not to require approval.

References

[1] G. Iddan, G. Meron, A. Glukhovsky, P. Swain; "Wireless Capsule Endoscopy", *Nature,* vol. 405, no. 6785, pp. 417 – 418, **2000**.

[2] G. Gallo, E. Granata, E. Scarpulla; "Wireless Capsule Endoscopy Video Segmentation", *MeMeA*, pp. 236-240, **2009**.

[3] M. Mackiewicz, J. Berens, M. Fisher; "Wireless Capsule Endoscopy Color Video Segmentation", *IEEE Trans. Med. Imag.,* vol. 27, no. 12, pp. 1769-1781, **2008**.

[4] M. Mackiewicz, J. Berens, M. Fisher, D. Bell; "Colour and Texture Based Gastrointestinal Tissue Discrimination", *ICASSP*, vol. 2, pp. II597 - II600, **2006**.

[5] J. Lee, J. Oh, S. K. Shah, X. Yuan, S. J. Tang, "Automatic Classification of Digestive Organs in Wireless Capsule Endoscopy Videos", *Proc. ACM Symp. Applied Computing*, pp. 1041–1045, **2007.**

[6] B. Li, M. Q.-H. Meng, "Capsule Endoscopy Video Boundary Detection", *Proc. Int. Conf. on Inf. and Auto.*, pp. 373 – 378, **2011**.

[7] "Capsule Endoscopy", The London Clinic, http://www.thelondonclinic.co.uk/specialties/endoscopy/capsule_endoscopy.aspx, Accessed: March 2012.

[8] J. P. S. Cunha, M. Coimbra, P. Campos, J. Soares, "Automated Topographic Segmentation and Transit Time Estimation in Endoscopic Capsule Exams", *IEEE Trans. on Medical Imaging*, vol. 27, no. 1, pp. 19-27, **2008**.

[9] P. Niyogi, "Locality preserving projections". *Advances in neural information processing systems*, vol. 16, pp. 153-160, **2004**.

[10] J. Berens, M. Mackiewicz, D. Bell; "Stomach, Intestine and Colon Tissue Discriminators for Wireless Capsule Endoscopy Images", *Proceedings for SPIE*, vol. 5747, no. 1, pp. 283-290, **2005**.

[11] D. Connah, G. D. Finlayson, "Using Local Binary Pattern Operators for Colour Constant Image Indexing", *Proc. CGTV,* pp. 60-64, **2006**.

[12] M. Petrou, P. Bosdogianni, "Image Processing – The Fundamentals", John Wiley & Sons Ltd. **1999**, Chichester, pp. 34 – 44.

[13] K. D. Gonzales, G. Douglas, K. Pichakron, D. Kwiat, S. G. Gallardo, J. L. Encinas, S. Hirose, M. R. Harrison, "Magnamosis III: delivery of a magnetic compression anastomosis device using minimally invasive endoscopic techniques", *J. Pediatric Surgery*, vol. 47, no. 6, pp. 1291 – 1295, **2012**.

[14] L. Igual, J. Vitria, F. Vilarino, S. Segui, C. Malagelada, F. Azpiroz, P. Radeva, "Automatic Discrimination of Duodenum in Wireless Capsule Video Endoscopy", *IFMBE Proceedings*, vol. 22, pp. 1536 – 1539, **2008**.