COLOUR AND TEXTURE BASED GASTROINTESTINAL TISSUE DISCRIMINATION

Michal Mackiewicz, Jeff Berens, Mark Fisher, Duncan Bell

School of Computing Sciences, University of East Anglia, Norwich, UK

ABSTRACT

Wireless Capsule Endoscopy is a colour imaging technology that enables close examination of the interior of the entire small intestine. The Wireless Capsule Endoscope (WCE) operates for $\sim$ 8 hours and captures $\sim$ $40,000$ useful images. The images are viewed by a clinician as a video sequence, generally taking over an hour to analyse. In this paper we present a method of automatically discriminating stomach and intestine tissue which can significantly speed-up one key part of the video analysis time, namely the process of locating the Pylorus- the valve between the stomach and the intestine. We divide the WCE image into 28 sub-regions and process only those regions where tissue is clearly visible. We create a feature vector using colour and texture information. The colour features are derived from Hue Saturation chromaticity histograms of the useful regions, compressed using a hybrid transform, incorporating the Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA). The texture features are derived by Singular Value Decomposition of the same tissue regions. After training the Support Vector classifier, we apply a discriminator algorithm, which scans the video with an increasing step and builds up a classification result sequence. By minimizing the number of misclassifications within this sequence, we predict the most probable position of the Pylorus. We present experimental results that demonstrate the effectiveness of this method.

1. INTRODUCTION

Standard endoscopy enables a physician to view both ends of a patient's digestive tract including the foodpipe, stomach, duodenum, colon and terminal ileum. Examination of the remainder of the small intestine was until very recently a difficult procedure. The solution to this problem first appeared in [1], and involves the use of wireless transmission to send images from inside the small intestine to the outside world. The 11mm x 26mm capsule is swallowed and propelled through the food tract by normal peristalsis. One end of the capsule contains an optical dome with white light LEDs and a colour camera that takes 2 pictures a second. These images are relayed via a transmitter to a data-recorder worn by the patient on a belt. At the end of the 8 hours (the battery lifetime), the data-recorder is removed and the image data uploaded to a workstation for later viewing. The stored data consists of $\sim$ $50,000$ images, and is viewed as a video sequence using special software provided by the manufacturers[2]. An important factor with regards to using the WCE system is that viewing each video requires a significant time-commitment and the close concentration of an expert clinician. Even for an experienced viewer using "fast-forward" mode, it can take over an hour to analyse a WCE video[3].

In order for certain parts of the viewing software to work, it is necessary for the clinician to find and mark the location in the video of the Pylorus. Finding the point in the video where the WCE leaves the stomach and enters the intestine can be difficult and time-consuming, even for an experienced viewer, as images from the stomach and intestine regions around the pylorus appear visually highly similar. Providing this location automatically would significantly reduce the amount of time taken by a doctor to analyse a WCE video.

In [4] we showed that colour based stomach/intestine discriminators, which are briefly described in Section 2.1, can be used to locate the pylorus. In this paper we show that incorporating textural features into these discriminators improves their accuracy. In Section 2.2 we describe our new method of feature extraction for both colour and texture. Section 3 describes the four classifiers used in this work. Section 4 describes the discriminator algorithm, and Section 5 shows and discusses the results.

2. FEATURE EXTRACTION

2.1. Review of the previous feature extraction method - Entire Image colour features

The distribution of colours in an image often provide a useful cue for image indexing and object recognition. The most commonly used method of representing image colour for image indexing is the colour distribution histogram, which is relatively invariant to image scale changes, translation and rotation about the viewing axis, and to partial occlusion [5]. WCE images might be indexed using this method since colour information is the primary feature analysed by the clinician.

The authors would like to thank National Association of Colitis and Crohn's Disease (UK charity) for its financial support towards this research and Dr Crawford Jamieson from the Norfolk & Norwich University Hospital for providing the WCE videos.

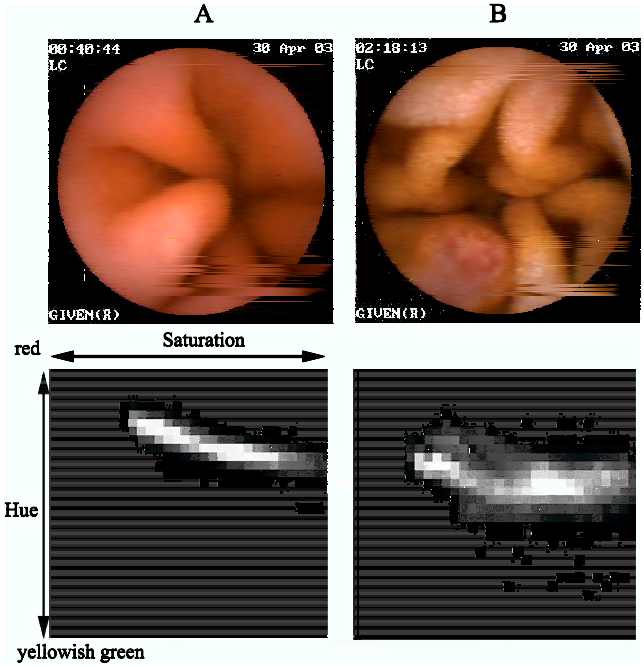

Fig. 1. A figure showing WCE images taken from A) stomach; B) small intestine; and below their corresponding equalized HS histograms. A visible shift in hue (vertical axis of the histogram) between these examples is clearly visible.

Visually, the stomach contains pinkish colours whereas the small intestine contains pinkish to yellowish colours. Often the tissue is partially or entirely occluded by varying amounts of saliva, bile or faeces.

We convert the WCE RGB images into the HSI colour space and, ignoring the intensity information, we form 2-D HS histograms. There is a great deal of intensity variation in WCE images as the distance between the WCE and the intestine surface constantly varies. By ignoring intensity information we force intensity invariance, and also reduce data size. The range of colours present in WCE images is relatively small, mapping to a region covering just around 20% of the possible colour space, and so we equalise the histograms to this subset (of red to yellowish-green) colours. Figure 1 shows typical WCE images taken from the stomach and intestine regions, and their histograms. It can be seen that the colour distribution of the stomach is shifted towards red compared to the intestine distribution.

The 2-D histogram feature vector is a large structure, and for practical applications compression is necessary. Consequently, we apply a hybrid transform consisting of DCT compression followed by PCA, which was shown to perform better than DCT or PCA alone [6]. The hybrid transform allows a small feature vector that provides a fast and accurate model of the histogram data.

2.2. Region based feature extraction

In order to improve on our earlier (satisfactory) results, we have added two stages to the feature extraction method described in 2.1. Firstly, we exploit the fact that there is a just visible textural difference between stomach and intestine tissue. Intestine tissue contains villi, small finger-like projections responsible for food absorption, which are not present in the stomach, a difference we can represent using textural analysis. Secondly, we address the issue of WCE images often being obscured (to a varying degree) by strong shadows, or by air-bubbles and other artifacts - such as saliva, bile, faeces, food etc. Histograms built using the entire image will contain any visual contamination present in the image. In order to minimise the affect of visual contamination, we divide each WCE ($256 * 256$ pixel) image into a grid of 28 sub-images, 32*32 pixels each, covering most of the image area, as shown in Figure 2 A. We derive five parameters for each of the sub-images - Mean Intensity, Saturation, Hue, and Standard Deviation of Intensity and Hue. The values for these parameters were set by experiment so that sub-images containing visual contamination (i.e. outside the expected colour range for the tissue type) are rejected. We test each sub-image against the five parameters and discard any sub-image that falls outside the range of values typical for visually clear images of stomach or intestine tissue.

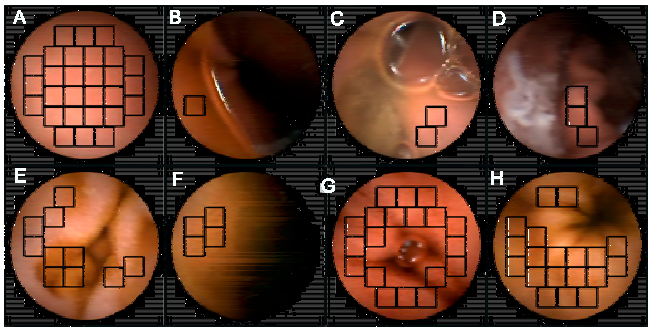

Fig. 2. A-D stomach; E-H intestine WCE images. The texture of the villi can be seen in the intestine images.

2.2.1. Colour

The colour features are derived by building an HS histogram for each usable sub-image, and compressed using the hybrid transform, as described in Section 2.1.

2.2.2. Texture

For our texture features, we use a texture measurement method based on Singular Value Decomposition, introduced by Ashjari in [7]. Each usable sub-image block is converted to grey scale and treated as a 32*32 matrix. Next, the amplitude ordered set of singular values is computed for each block and

an average of the singular values across all the pre-selected regions is recorded. Figure 3 shows eight singular value spectra, four for intestine and four for stomach, corresponding to the images in Figure 2. It can be clearly seen that the plot of the stomach spectra for the mid components lies below the intestine spectra.

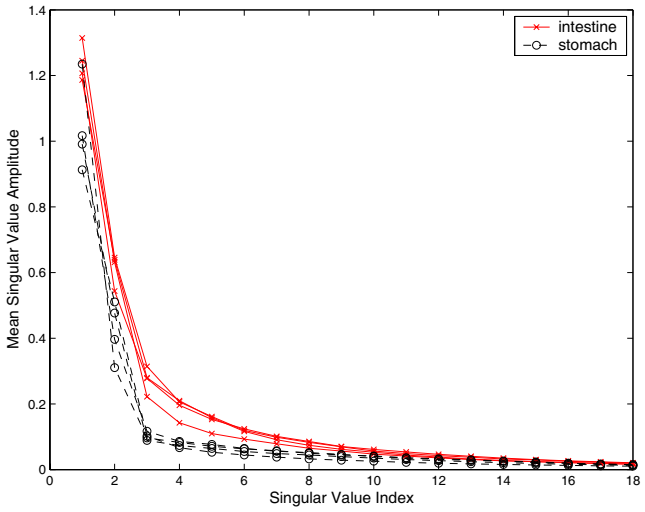

Fig. 3. Stomach and intestine singular value texture spectra.

3. CLASSIFIER

Each training set was built from 1000 images (500 stomach and 500 intestine) selected from 70 different videos. We trained four Support Vector classifiers:

1. Entire Image Colour based classifier: The feature vector consists of 64 Principal Components (PCs) derived from the histogram of the entire image.

2. Region based Colour classifier. The feature vector contains 64 PCs derived from the HS histograms of valid sub-images.

3. Region based Texture classifier. The feature vector contains 32 singular values representing texture spectra obtained from valid sub-images.

4. Region based Colour + Texture classifier. The feature vector consists of 64 PCs derived from HS histograms and 32 singular value spectra taken from valid sub-images.

4. DISCRIMINATOR ALGORITHM

A WCE video contains $\sim 40,000$ useful frames, too many frames to classify in a reasonable time. We have demonstrated in our previous work that it is not necessary to classify each frame in the WCE video - for the purpose of tissue classification a sample of frames suffices. In [4], we proposed a simple jump algorithm. Having classified the first frame, it moves forward or backward depending on the outcome of the classification, gradually decreasing the step until it converges at a boundary between stomach and intestine tissue types. However, this method was found to be very sensitive to a single classification error. We have developed a new algorithm that scans the video from the start of the stomach (which is found automatically using our algorithm described in [8]) to the end of the video. Analysis has shown that the mean gastric transit time (the time spent in the stomach) is less than 20 minutes, and the mean pylorus position is around frame 2500. We have modified the step of the algorithm to take advantage of this feature, by making the value of the step proportional to the square root of the frame number. As a result, the algorithm becomes more precise in the region where the pylorus is statistically likely to be, and also keeps the number of frames to be classified low - around 200. Moreover, some videos contain long sequences in which tissue is completely occluded, and which disadvantages the classifiers, since no region could be selected for the colour and/or texture information extraction. If this occurs, we revert to the Entire Image colour classifier. However, as any information contained in these images may be distorted, an output from this classifier is given half of the weight of the other classifiers. Formally, the result sequence $S = \{S(m) : S(m) \in \{-1, -\frac{1}{2}, 1, \frac{1}{2}\} \wedge m \in \mathbf{N} \wedge m \leq M\}$ where M denotes a length of the classification sequence. A typical result for the Region based Colour + Texture classifier can be seen in Figure 4. From the result sequence S, we calculate the estimate classification error E as a sequence:

$$E(i) = \sum_{k=1}^{i} S_{pos}(k) - \sum_{k=i+1}^{M} S_{neg}(k)$$

where

$$S_{pos}(k) = \begin{cases} S(k) & S(k) > 0 \\ 0 & S(k) < 0 \end{cases}$$

and

$$S_{neg}(k) = \begin{cases} S(k) & S(k) < 0 \\ 0 & S(k) > 0 \end{cases}$$

The minimum of sequence E determines the most probable position of the pylorus, as seen in Figure 4. Next, we run the second pass of the discriminator, centred on the first pass result: Let x denote a frame number in the video such that $E(x) = min(E)$, then the second pass of the discriminator classifies the sequence of frames F_n, as in Equation 1 where n denotes the frame number in the video.

$$F_n : n = p + (k-1) * 20 \wedge k \in \mathbf{N} \wedge n \leq q \qquad (1)$$

where

$$p = x - 200 - \sqrt{x}$$

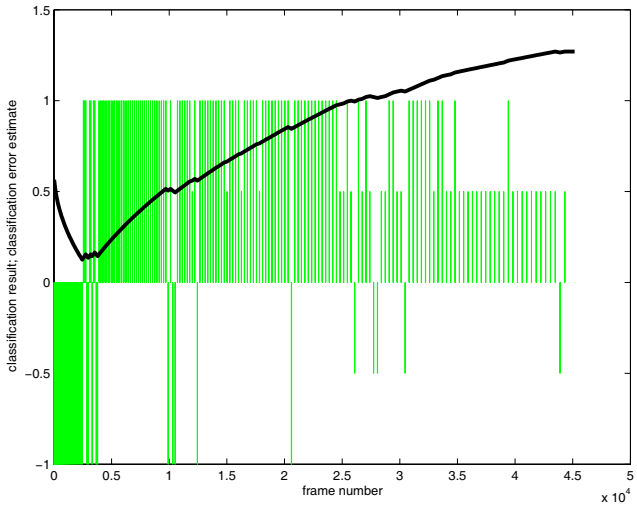

Fig. 4. Sample video classification results.

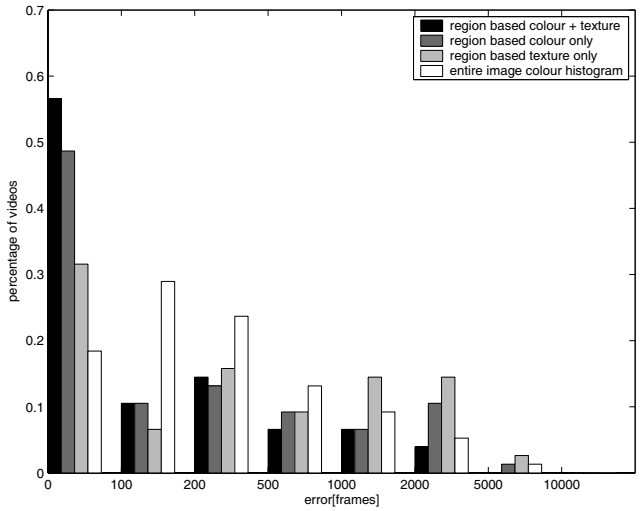

Fig. 5. Results for stomach/intestine discrimination on 76 videos.

Table 1. The median and mean errors of stomach/intestine discrimination.

method	median error [frames]	mean error [frames]
region based texture + colour	69	381
region based colour only	117	687
region based texture only	387	1803
entire image based colour	209	659

and

$$q = x + 200 + \sqrt{x}$$

The square root term in p and q controls the length of the video region to be classified in the second pass. The further from the mean pylorus position the first pass prediction is, the greater the length of the video region processed in the second pass.

5. RESULTS & DISCUSSION

The accuracy of the algorithm was assessed by the frame difference (error) between the point in the video where the boundary has been manually annotated (by a clinician) and the point selected by our algorithm. The error distribution of four tissue discriminators can be seen in Figure 5. Table 1 shows the mean and the median error of those discriminators. It can be clearly seen that the best performance is achieved by the Region based Colour and Texture discriminator, followed by the Region based Colour and then Entire Image Colour discriminator. Texture information alone does not provide an accurate discrimination, as the results reflect.

This research shows that an automatic pylorus detector using combined colour and texture features is feasible, which could lead to a reduction in the WCE video viewing time.

6. REFERENCES

[1] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, pp. 725–9, 2000.

[2] ," Given Imaging Ltd., Israel, http://www.givenimaging.com.

[3] A. Maieron et all, "Multicenter Retrospective Evaluation of Capsule Endoscopy in Clinical Routine," *Endoscopy*, vol. 36, pp. 864–868, 2004.

[4] J. Berens, M. Mackiewicz, and G. D. Bell, "Stomach, Intestine and Colon tissue discriminators for Wireless Capsule Endoscopy (WCE) images.," in *Proceedings of SPIE*, 2005, vol. 5747, pp. 283–290.

[5] M.J. Swain and D.H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, 1991.

[6] J. Berens, *Image Indexing using Compressed Colour Histograms*, Ph.D. thesis, School of Information Systems, University of East Anglia, Norwich, UK, 2002.

[7] B. Ashjari, *Singular value decomposition texture measurements for image classification*, Ph.D. thesis, Department of Electrical Engineering, University of Southern California, 1982.

[8] J. Berens, M. Mackiewicz, M. Fisher, and G. D. Bell, "Using colour distributions to discriminate tissues in Wireless Capsule Endoscopy images.," in *Proceedings of Medical Image Understanding and Analyses 2005 Conference*, Bristol, UK, July 2005.