Color Analysis for Segmenting Digestive Organs in VCE

Hai Vu, Yasushi Yagi
The ISIR, Osaka Univ.
{*vhai,yagi*}*@am.sanken.osaka-u.ac.jp*

Tomio Echigo
Osaka Electro-Communication Univ.
echigo@isc.osakac.ac.jp

Masatsugu Shiba, Kazuhide Higuchi, Tetsuo Arakawa
Graduate School of Medicine, Osaka City Univ.

Keiko Yagi
Kobe Pharmaceutical Univ.

Abstract

This paper presents an efficient method for automatically segmenting the digestive organs in a Video Capsule Endoscopy (VCE) sequence. The method is based on unique characteristics of color tones of the digestive organs. We first introduce a color model of the gastrointestinal (GI) tract containing the color components of GI wall and non-wall regions. Based on the wall regions extracted from images, the distribution along the time dimension for each color component is exploited to learn the dominant colors that are candidates for discriminating digestive organs. The strongest candidates are then combined to construct a representative signal to detect the boundary of two adjacent regions. The results of experiments are comparable with previous works, but computation cost is more efficient.

1. Introduction

A Video Capsule Endoscopy (VCE) sequence captured over 8 hours by a Wireless Capsule Endoscopy [1] device usually contains about 57,000 images spanning the organs of the Gastrointestinal (GI) tract: esophagus, stomach, small intestine, and colon. At the beginning of an examination, doctors often browse frames to find landmarks forming boundaries between the digestive organs, such as the pylorus (the boundary between the stomach and the small intestine), or ileocecal valve (IV) (the boundary between the small intestine and colon). The segmenting results produce suggestive information to estimate the position of the capsule in the diagnostic procedure. Manually detecting the boundaries (especially the pylorus and IV), however, is time-consuming and requires a doctor's undivided attention owing to the huge number of frames to be inspected and the unknown orientation of the capsule device.

To reduce the examination time and burden on the doctors, Coimbra et al. [2] approached segmenting the landmark frames as a problem of video segmentation in a Bayesian framework. Recently, Mackiewicz et al. [3] proposed an intensive approach using combinations of multiform image features (color, texture, and motion), and robust classifiers (SVC, GMM) to find cues for the best segmentation results. In these works, the authors considered the problem using a framework consisting of two stages. In the first stage, images are classified according to the digestive region, while in the latter stage, a video segmentation technique is utilized to detect landmark frames. Unfortunately, these works suffered from high computation costs and issues regarding recognition accuracy in the first stage. In other words, obtaining a reasonable result requires a large training data set to deal with over-fitting problems and unclear patterns in the digestive organ images.

In view of these challenges, we use the color characteristics of the GI tract to obtain well-discriminated digestive organs. We make use of the fact that the color of each digestive organ has unique tones: the stomach contains pinkish colors, while the small intestine contains pinkish-yellowish colors [3]. We first construct a GI color model based on a large VCE dataset. The components of the GI color model are then separated into three groups: the GI wall, darkness and noise (with the latter two groups are referred to as non-wall components). We then exploit the appearance frequency of wall color components along the time dimension. Consequently, a series of dominant or candidate images for discriminating digestive organs can be learned. To detect landmark frames, a signal is structured by combining the candidates along the transit route of the capsule. Sharp changes in the distributions of the dominant colors implied in the signal suggest the positions of the landmarks. The detection results are comparable with recent work [3]; however, the proposed method requires less computation time as well as being more easily deployable in clinical applications.

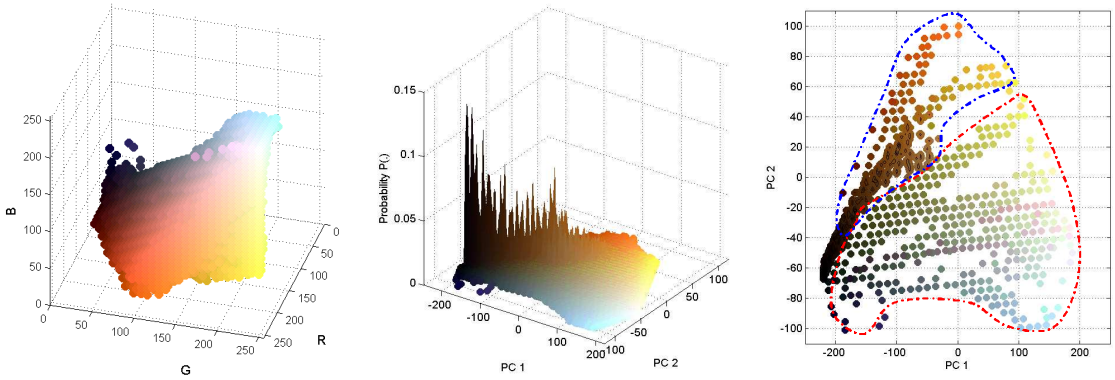

Figure 1. (a) The GI color model. (b) Distribution of $P(.)$ with the color components plotted in PCA space. (c) Initialization scheme for the component classification.

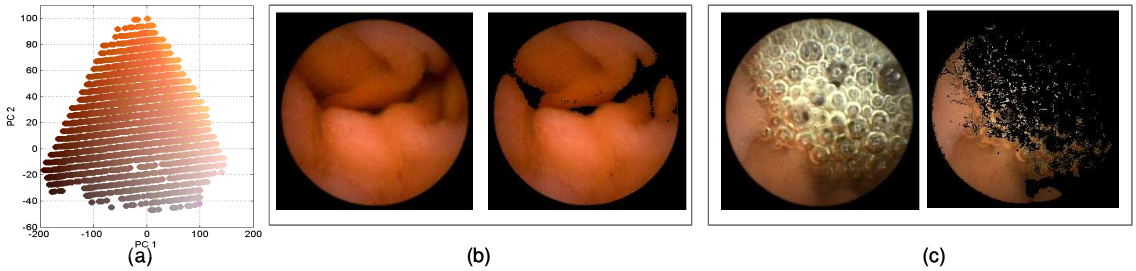

Figure 2. (a) Wall color components. (b)-(c) Two examples of meaningful segmented regions. (In each panel, the original image is on the left, with the segmented image on the right.)

2. The GI Color Model

2.1 Material and the generic GI color model

Intuitively, constructing a consistent GI color model requires a very large dataset of VCE sequences to cover a variety of patient data. For this study, we collected 30 sequences from a database containing data for 300 patients, being careful to select a wide range of ages, gender and length of video sequences. Subsequently, a large dataset containing $\approx 107 \times 10^9$ images was quantized using the popularity color algorithm. For an original RGB image (256 x 256 pixels), each color channel is divided into N bins to construct a global color histogram. The popularity algorithm simply selects the K color components with the highest histogram values. The components included in the model are uniquely collected from fully quantized data. The probability P(.) of a color component c_{rgb} is defined by:

$$P(c_{rgb}) = \frac{H(c_{rgb})}{T}, \qquad (1)$$

where $H(c_{rgb})$ is the total number of pixels belonging to color c_{rgb} and T is the total number of pixels in the dataset. $N = 32$ and $K = 256$ are optimal values. Fig. 1(a) shows the constructed GI color model in RGB space, while Fig. 1(b) gives the distribution of P(.), where PC 1 and PC 2 are the primary components transformed from the RGB space using the PCA. Fig. 1(b) is also rendered using the true colors of the components.

2.2 Color component classification

The generic GI color model constructed in Fig. 1(a) contains components classified as GI wall or non-wall regions. The wall consists of homogenous regions and visualizing surfaces of rugae in the GI tract, whereas non-wall regions are the dark lumen regions or contaminated noise such as water bubbles, gas, food, etc. Intuitively, Fig. 1(b) verifies that with a large dataset, the probability P(.) of noise is very small, whereas P(.) of wall regions and dark regions is more robust. Therefore, our proposed scheme includes two steps for training and classifying the color components into three groups: wall components, darkness and noise. First, we initialize a small set including components that are strong representatives of both the GI wall and non-wall using two thresholds $Thresh1$ and $Thresh2$, respectively. $P(.) > Thresh1$ strongly defines the components of wall or darkness regions, whereas $P(.) < Thresh2$ suggests noise data. Fig. 1(c) shows the blue contours of high P(.) components in Fig. 1(b).

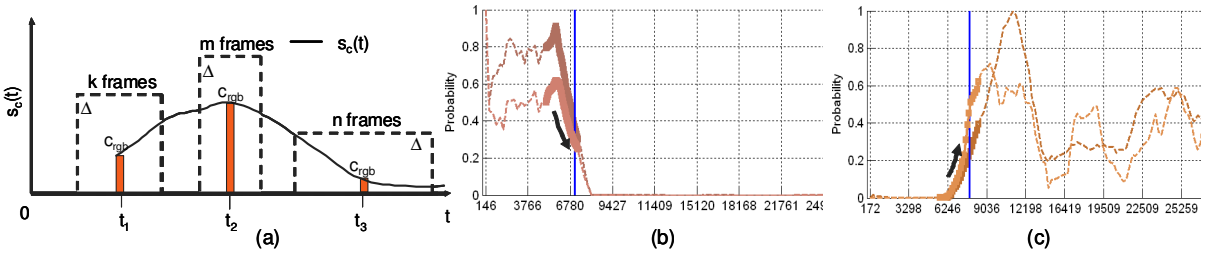

Figure 3. (a) Prototype to generate $s_c(t)$. (b)-(c) Two $s_c(t)$ candidates with high distributions in the stomach and the small intestine, respectively. The blue line marks the pylorus position .

These contours suggest an estimated $Thresh1$ value (e.g., $Thresh1 = 0.01$). The noise components, plotted within the red dotted lines, have $P(.) < Thresh2$ with $Thresh2 <= 0.01 \times Thresh1$ value. Based on the appearance of these components (Fig. 1(c)), meaningful groups are labeled manually. We then deploy the K-NN classifier using R G B features to assign labels for the remaining components.

Figure 2(a) shows the wall color components separated from the model in Fig. 1(a). Wall regions in an image can therefore be segmented. Fig. 2(b)-(c) shows two examples, the results of which are robustly even despite the inclusion of hard contaminations such as water bubbles (in the right panel). Moreover, to detect landmark frames, color components that clearly discriminate the digestive organs can be learned based on the segmented images along the capsule's route. Such components are referred to as dominant components.

3. Detecting the landmark frames

3.1 Learning the dominant color components

Assume that at time t_i, a meaningful region with area $\triangle$ concatenated from k consecutive frames is extracted. The probability of a color c_{rgb} at time t_i is:

$$s_c(t_i) = \frac{H(c_{rgb})}{\triangle} \tag{2}$$

A signal $s_c(t)$ denoting the distribution of color c_{rgb} along the time dimension is constructed. Fig. 3(a) illustrates a prototype for generating signal $s_c(t)$ (at t_1, t_2, t_3). Note that $\triangle$ is composed from several frames around t_i so that $\triangle$ reaches the predetermined value (in pixels). Using Eq. (2), at each time t_i the appearance frequencies of the color components are normalized. Therefore, the dominant components can be learned by exploiting the shape of the signal $s_c(t)$ around the boundary of two adjacent regions.

For the learning phase, ten sequences with already marked landmark frames, were collected as the training data set. The signals $s_c(t)$ of all wall color components were generated and their shapes around landmark frames were observed. Fig. 3(b)-(c) shows the components that are suggestive for detecting pylorus. In Fig. 3(b), two signals show high distributions in the stomach region, yet barely appear in the small intestine. Contrary to those, Fig. 3(c) shows two signals whose distributions are high when the capsule enters the small intestine. To detect these signals, we simply examine the derivation of the first order of $s_c(t)$. Consequently, a series of candidates for detecting stomach/small intestine are shown in Fig. 4. Thus, the strongest candidates are selected based on their appearance frequency in the training sequences.

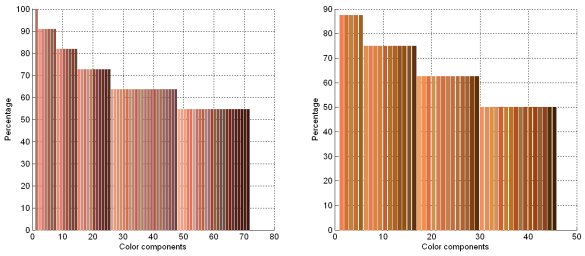

Figure 4. Dominant colors in the stomach (left) and the small intestine (right).

3.2 Detecting the landmarks

The strongest candidates suggest good features for discriminating two adjacent regions. In other words, constructing a signal $S(t)$ that combines the strongest candidates will significantly cover distributions of dominant color tones for each digestive region. A simple version of signal S can be defined as follows:

$$\begin{aligned} S(t) \quad = \quad & max(s_{c_i}(t)|i = 1..m) + \\ & (sign)max(s_{c_j}(t)|j = 1..n) \end{aligned} \tag{3}$$

where s_{c_i} and s_{c_j} are, respectively, signals of the m and n strongest candidates obtained from the learning scheme described above. To detect the boundary of two adjacent regions, the $sign$ operators in (2) reverse sign of the candidates from one region with respect to the other. The landmark frames are then detected using a zero-crossing technique. (See details in Fig. 5)

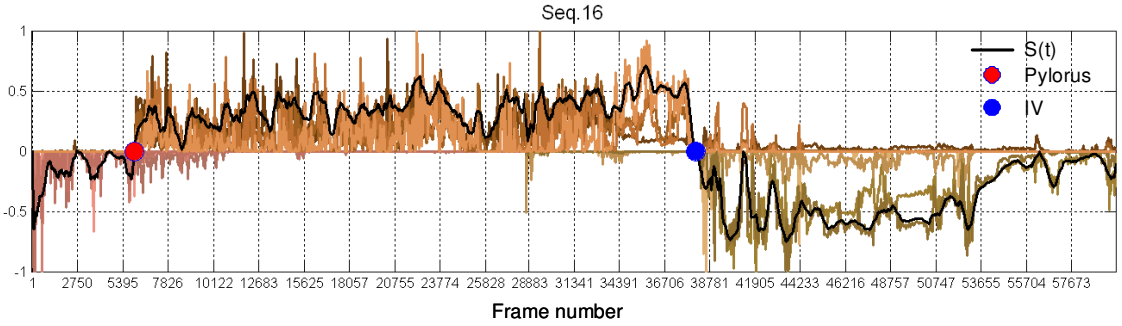

Figure 5. The result of a sequence with detected points is marked. The components comprising $S(t)$ are plotted in their true colors.

Table 1. The distance errors of the testing data

Results	Pylorus Distance Errors					IV Distance Errors						
						Undetected		Detected				
	<1 min	1-2 min	2-5 min	>5 min	Total	FAR	Corrected	<1 min	1-2 min	2-5 min	>5 min	Total
Number of Seq	19	11	11	9	50	3	9	12	8	9	9	50
Percentage (%)	38	22	22	18	100	25	100	32	21	24	24	100
Mean (frames)	45	156	390	1878				68	176	353	1758	
Median (frames)	**105**							**319**				

4. Experimental results

To evaluate the proposed method, we randomly selected 50 sequences from the patient database. The signal $S(t)$ in (4) is generated from the four strongest candidates ($m = n = 4$) in each region. The signal $s_c(t)$ of each component c_{rgb} (by (2)) is generated by randomly selecting 1200 data points in a full video data. The value of $\triangle = 200,000$ pixels was pre-determined.

The $S(t)$ of an example sequence is shown in Fig. 5, in which the components $s_c(t)$ are expressed in their true colors. Two landmark frames for pylorus and IV are detected on the smoothed $S(t)$ signal. We then calculated the absolute distance between the detected frames and ground-truth data, in which the ground-truth data were labeled by experts at the Osaka City University Hospital. The distance errors of the testing data are classified into four groups (within 1 min., 1-2 min. , 2-5 min., and $>$ 5 min., with 1 min. = 120 frames), as given in Table 1. The undetected column shows the detected sequences that do not locate the IV due to unreachable colon region in the route of the capsule. Similarly to [3], the median is mainly used for statistical measurement. The medians of 80% of the sequences (with distance $<$ 5 min) are comparable with the final results in [3] (91 and 288 frames for pylorus and IV, respectively). (In this study, the esogastric junction was not detected, because in practice it can be detected clearly and quickly by manual inspection). For each sequence, generating signal $S(t)$ and detecting the landmark frames takes less than three minutes. This is obviously more faster than the approaches in [2, 3] because of complexity of computations in their learning algorithms.

5. Conclusion

In this paper, we proposed an efficient method for automatically segmenting digestive organs from VCE sequences. The color characteristics of the digestive organs were exploited according to the proposed GI color model. To detect landmark frames, a signal representing the distributions of the dominant colors was constructed. The experimental results are comparable with previous works. Moreover, the proposed GI color model also suggests further applications such as image enhancement and abnormal region analysis.

References

[1] G. Iddan et al. . Wireless capsule endoscope. *Nature*, 405:417, May 2000.

[2] J. P. S. Cunha et al. Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE Trans. on Medical Imaging*, 27(1):19–27, 2008.

[3] M. Mackiewicz et al. Wireless capsule endoscopy color video segmentation. *IEEE Trans. on Medical Imaging*, 27(12):1769–1781, 2008.