Local Polynomial Approximation for Unsupervised Segmentation of Endoscopic Images

Artur Klepaczko[1], Piotr Szczypiński[1], Piotr Daniel[2], and Marek Pazurek[2]

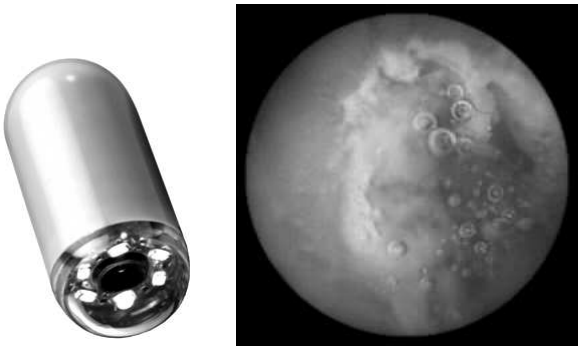
[1] Technical University of Lodz, Institute of Electronics
90-924 Lodz, ul. Wolczanska 211/215
aklepaczko@p.lodz.pl
[2] Medical University of Lodz, Department of Digestive Tract Disease
90-153 Lodz, ul. Kopcinskiego 22

Abstract. In this paper we present a novel technique for unsupervised texture segmentation of wireless capsule endoscopic images of the human gastrointestinal tract. Our approach integrates local polynomial approximation algorithm with the well-founded methods of color texture analysis and clustering (k-means) leading to a robust segmentation procedure which produces fine-grained segments well matched to the image contents.

1 Introduction

In this paper we present a novel technique for unsupervised texture segmentation applied to wireless capsule endoscopic (WCE) 2D images of the human gastrointestinal tract. We are primarily focused on improving automatic detection of ulcerations and lesions visible in the internal lumen of the small intensine and their precise delineation from normal or irrelevant regions. Our approach integrates local polynomial approximation (LPA) [1] algorithm with the well-founded color texture analysis and unsupervised classification (k-means) methods. As a first step, LPA performs pixel-wise analysis of the circular view given by a WCE camera (cf. Fig. 1a) and for each pixel it defines a corresponding region of interest (ROI) whose size and shape is adapted to this pixel local neighborhood. Then, using histogram information calculated separately for 8 different color channels, each ROI (and thus also its associated pixel) is described by a vector of color texture features. Eventually, these vectors are classified in the unsupervised manner by the k-means algorithm with the number of clusters set a priori. This combination of methods leads to a robust, three-phase fully automated segmentation procedure which produces fine-grained segments well matched to the image contents.

The motivation for our research stems from the persistent need for automation of WCE video interpretation process. It traditionally involves much effort from a human expert, is a monotonous and time-consuming task, requiring high level of concentration. Apparently, in the literature the problem of automatic



(*a*) (*b*)

Fig. 1. Wireless capsule endoscope (a) and an example image (b).

segmentation of WCE images has attracted little attention. Majority of studies (e.g. [2–5]) concentrate on WCE video sequence segmentation as a whole — the aim is to identify different sections of the gastrointestinal system. It is worth noticing that combined color and texture information is widely exploited in these works showing its potential in application to endoscopic image processing.

Another category of approaches utilize color and texture features to detect certain structures (like polyps, bleeding, ulcers and lesions) with or without performing image segmentation. In [6] WCE frames are first submitted to the smoothing procedure and then segmented by the region growing algorithm. Identified regions and their mean RGB color values are used to train a neural network classifier so that it recognizes a range of abnormal patterns in new images. Similar studies — although they involve different classifiers, color and texture models — are described in [7, 8] and [9, 10]. However, in these approaches images are divided uniformly into a predefined number of square or circular ROIs, disregarding particular shapes visible in a given frame. The focus is put on determination of characteristics of chosen patterns in endoscopic images. These characteristics can be used to build classifiers capable of distinguishing classes of ideally shaped image snippets. Although such classifiers may to some extent help indicate diagnostically valuable WCE frames, precise delineation, visualization and quantification of interesting regions remains an unsolved task.

In the following we present details of our approach to WCE image segmentation. Section 2 provides concise review of the materials and methods used in this research. In Sect. 3 we present preliminary results of segmentation obtained for sample WCE images taken from several different examinations. We compare these results with segmentation performed manually by two independent experts in WCE-based diagnosis. High level of correspondence between both types of segmentation techniques (manual and the automatic one) makes us believe that the proposed method can be successfully introduced to the clinical practice. Conclusions and possible improvements are outlined in Sect. 4.

2 Materials and methods

2.1 Wireless Capsule Endoscopy

Wireless capsule endoscopy (WCE) [11, 12], is a technique that facilitates the imaging of the human gastrointestinal system including small intestine. The WCE system consists of a pill-shaped capsule (cf. Fig. 1a) with built-in video camera, light-emitting diodes, video signal transmitter and battery, as well as a video signal receiver-recorder device. The wireless capsule endoscope used in this study produces color images of the internal lumen (cf. Fig. 1b). The images cover a circular 140° field of view. A patient under investigation ingests the capsule, which then passes through the gastrointestinal tract. When the capsule goes through the small bowel it is propelled by peristaltic movements. The capsule transmits video data at a rate of two frames per second for approximately 8 hours. Investigation of the recorded video, usually numbering several tens of thousends of frames, is performed by a trained clinician. It is a tedious task that usually takes more than an hour. The video interpretation involves viewing the video and searching for abnormal-looking entities like bleedings, erosions, ulcers, polyps and narrowed sections of the bowel.

In this study we examined four WCE video sequences among which we selected 20 sample images comprising of various forms of ulcerations, bleedings and erosions. Within these images, pathology regions were manually delineated by the cooperating gastroenterologists. Example images included in our experiments with their corresponding regions of interest are depicted in Fig. 2.

2.2 Local Polynomial Approximation

The task of the LPA algorithm invoked in the first step of the segmentation procedure is to define for each pixel a region that is adapted to the shape of this pixel local neighborhood. It is presumed that such a neighborhood exhibits homogenous color and texture characteristics. Consequently, it can be expected that in the subsequent texture analysis step pixels gain credible description, representative for a wider range of points. This in turn allows making the resulting texture description independent from local fluctuations in pixel values caused by noise and other imaging artifacts.

The LPA algorithm itself is a technique of non-parametric regression recently adopted in various image processing applications. Using low order polynomial function, LPA models a non-linear relationship between an independent variable X and a dependent variable Y. Data are fitted to a modeled polynomial function within a sliding window positioned at subsequent observations (X, Y) — e.g. measured values of a sampled signal. In a window, a signal is convolved with a kernel function of a known form. This enables estimating values of the Y signal in the neighborhood of a given data point X. Window size h is a key parameter of the method. It is defined as a number of data samples beyond which it becomes impossible to estimate signal Y basing on values measured in the proximal neighborhood of X.

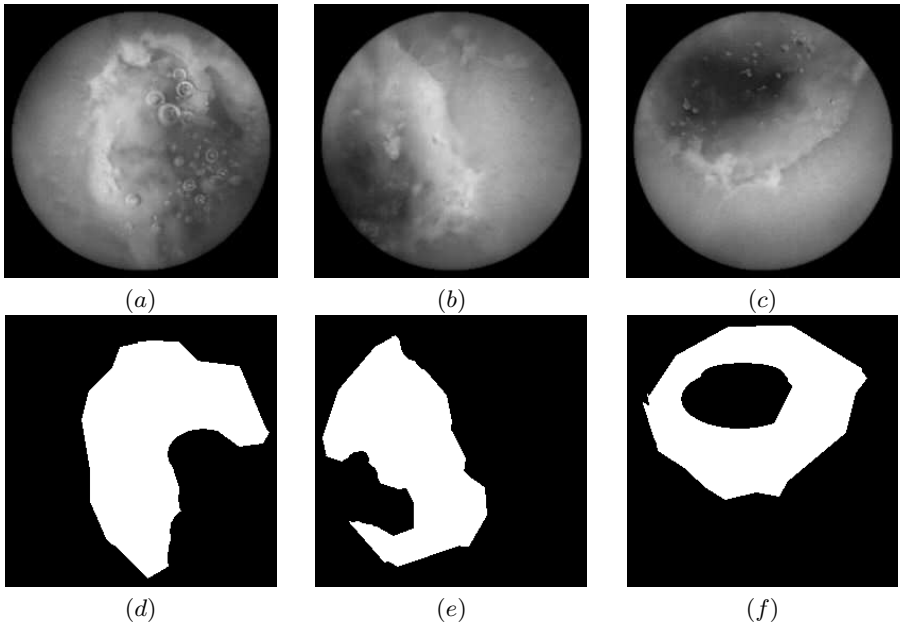


Fig. 2. Example WCE video frames (a-c) and their corresponding manually delineated regions of interest containing ulcerations with bleedings and narrowings (d-f).

In our study we apply the LPA algorithm following the approach presented in [13]. Firstly, a color WCE video frame is converted into a gray-scaled image. Next, we filter each pixel neighborhood in 8 distinct directions θ_i deviated from the horizontal East-oriented axis at angles $0°, 45°, 90°, 135°, 180°, 225°, 270°$ and $315°$. For a given pixel X we calculate

$$\mu^{(h)} = \sum_{j=1}^{h} g_j^{(h)} I(X + (j-1)\theta_i), \tag{1}$$

where $g^{(h)}$ is a discrete convolution kernel of scale h (window size), $g_j^{(h)}$ with $j = 1, \ldots, h$ denote kernel weights which sum to unity and decrease with the increasing distance from a center pixel X. The exact procedure of weights generation is described in [1]. Eventually, I is a matrix of image intensity values.

Adjusting the window size to local image contents is performed using the *intersection of confidence intervals* (ICI) rule. The idea is to test several values of scale h, i.e. $h \in \{h_1 \ldots, h_k\}$ and $h_1 < h_2 < \ldots < h_k$ and for each of them evalute (1) as well as local standard deviation value

$$\sigma_{\mu^{(h)}} = \sigma \| g^{(h)} \|, \tag{2}$$

where σ is the global standard deviation determined for the whole image. Then for each direction θ_i and scale h one calculates confidence intervals

$$\mathcal{D}_h = [\mu^{(h)} - \Gamma \sigma_{\mu^{(h)}}, \mu^{(h)} + \Gamma \sigma_{\mu^{(h)}}], \tag{3}$$

in which $\Gamma > 0$ denotes a global parameter that allows controlling noise tollerance. The lower Γ, the stronger requirement for local homogeneity is, and thus fewer pixels are included in the resulting neighborhood regions. The ICI rule states that for each direction one should choose a maximum value of h that ensures nonempty intersection of all previous confidence intervals, i.e.

$$h_{\max,i} = \max_{h \in \{h_1,\ldots,h_k\}} \{h : (\mathcal{D}_1 \cap \mathcal{D}_2 \cap \cdots \cap \mathcal{D}_h) \neq \emptyset\}. \tag{4}$$

In our experiments, we arbitrarily set $h \in \{1, 2, 3, 5, 7, 9, 12, 15, 18, 21\}$, hence the upper bound for the window size in any direction amounts to 21 pixels. On completion, pixels determined by relations $X + h_{\max,i}\theta_i$ constitute a set of polygon vertices whose interior determines a locally adapted ROI of a pixel X. Although we invoke LPA for a gray-scaled image, at the current stage of the research, for performance reasons, the resulting ROIs are used for each color channel in the texture analysis step.

2.3 Color and texture analysis

In order to exploit color information inherent in a WCE frame we convert it into 8 gray-scaled images using following color channels:

- brightness (according to the CCIR Recommendation 601-1),
- R,G, B, U and V color componentrs,
- saturation and hue.

For each color channel and every pixel we then calculate its associated ROI first-order histogram (256 bins). The histogram is computed from the intensity of pixels, without taking into consideration any spatial relations between the pixels within the image. Features are simply statistical parameters of the histogram distribution such as: mean brightness, variance, skewness, kurtosis and percentiles [14]. However, distinct regions of WCE images differ mainly by color and intensity and second-order regularities appear less important. Moreover, determination of histogram features is time-efficient and thus reduces the computational load.

2.4 K-means clustering

As a last step in our segmentation technique we perform cluster analysis for unsupervised grouping of image pixels described by their corresponding ROI texture features. For that purpose we employ k-means algorithm [15] due to its low complexity and ease of implementation. An important drawback of k-means is its intrinsic tendency to get stuck in a local minimum of the optimization criterion (sum of squared distances between feature vectors and their corresponding cluster centers). Thus, we call a clustering procedure 5 times, each time with different initialization. Then, the result with the best score is chosen to represent final segmentation.

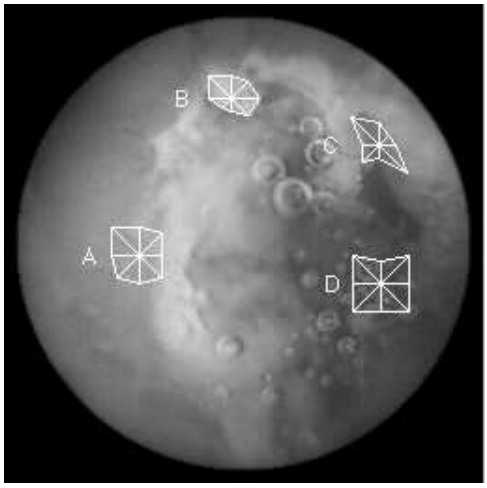


Fig. 3. Example ROIs defined by the LPA.

3 Results

In this preliminary research we have submitted all analyzed images into our segmentation procedure presuming arbitrarily that number of segments is equal to 6 in each case. By dividing an image into 6 segments we expect to encapsulate six possible patterns: ulcerations and accompanying lesions, bleedings, digestive residues, normal tissue and gastrointestinal lumen. This may of course lead to further subdivisions of some patterns if there are fewer distinguishable regions in a given frame.

In Fig. 3 we present example ROIs produced by the LPA algorithm for the image from Fig. 2a. Note, how regions A-C are well fitted to the surrounding tissue. On the other hand, region D placed in the interior of the intestinal canal embraces bubble-like structures whose borders emerge to be unimportant details in their local neighborhood. This observation shows that the algorithm — remaining local in nature — posses capability of reflecting general characteristics of a given pixel neighborhood.

Segmentation results obtained for three images viewed in Fig. 2a-c are depicted in Fig. 4a-c. Analysis of the segments borders reveals their fine matching to physical boundaries of ulcerations and bleedings. Moreover, it is apparent that identified image segments correspond well to the manually delineated regions of interest (cf. Fig. 4d-f). Although the latter occupy more than one automatically computed segments, those which are included to a large extent agree with the human expertise. In order to quantify a degree of this agreement, for each analyzed image we have calculated the Jaccard similarity coefficient between manually specified ROI and its respective automatically found and aggregated segments. The mean value obtained in these calculations amounted to 89% (with 4% of standard deviation).

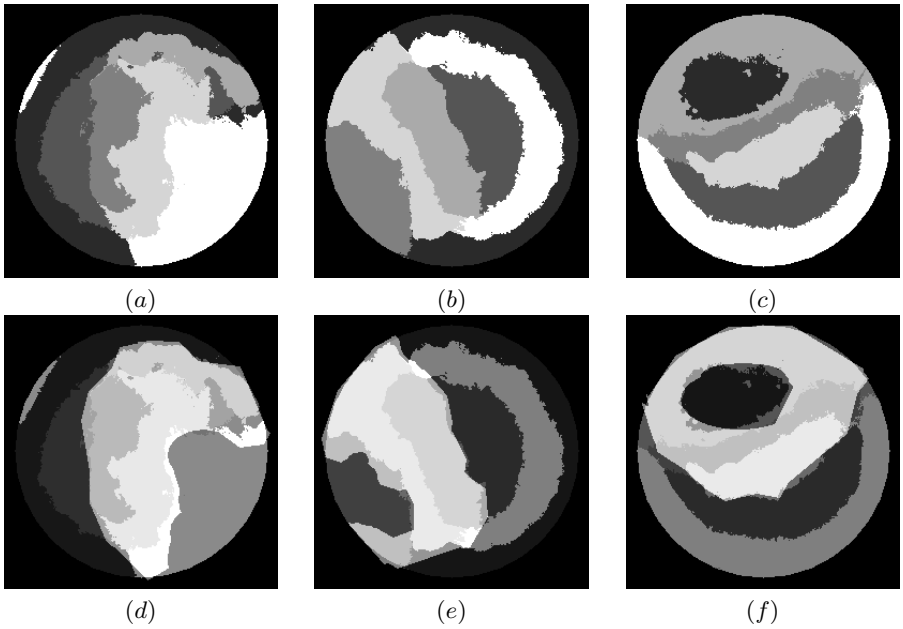


Fig. 4. Segmentation results (a-c) obtained for images from Fig. 2; d-f — identified segments overlaid on the manually delineated regions of interest.

4 Summary

In this paper we presented a novel method for automatic segementation of the WCE video frames. We demonstrated robustness of the proposed technique in application to several WCE images taken from real examinations. The obtained results are exhibit high rate of agreement between automatically identified segments and regions manually delineated by gastroenterologists. Naturally, these results must be further validated with the use of larger set of sample images.

It must be noted that our study is still in the development phase. Further research is required to reduce time complexity associated with determination of pixels local neighborhoods. Number of LPA executions while processing a single WCE frame ranges the order of 10^5. Fortunatelly, the LPA algorithm can be relatively easily parallelized and implemented efficiently eg. on a multicore GPU. Moreover, postprocessing should be added after clustering step in order to cope with segmentation holes and islands, and also to identify clusters which include distal regions and thus should be subdivided.

Acknowledgments. This work was supported by the Polish Ministry of Science and Higher Education grant no. 3263/B/T02/2008/35.

References

1. Katkovnik, V., Egiazarian, K., Astola, J.: Local Approximation Techniques in Signal and Image Processing. SPIE Press (2006)
2. Coimbra, M., Cunha, J.: MPEG-7 visual descriptors–contributions for automated feature extraction in capsule endoscopy. Circuits and Systems for Video Technology, IEEE Transactions on **16**(5) (May 2006) 628–637
3. Mackiewicz, M., Berens, J., Fisher, M., Bell, G.: Colour and texture based gastrointestinal tissue discrimination. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP. Volume 2. (2006) 597–600
4. Mackiewicz, M., Berens, J., Fisher, M.: Wireless capsule endoscopy video segmentation using support vector classifiers and hidden markov models. In: Proc. International Conference Medical Image Understanding and Analyses. (June 2006)
5. Mackiewicz, M., Berens, J., Fisher, M.: Wireless capsule endoscopy color video segmentation. Medical Imaging, IEEE Transactions on **27**(12) (2008) 1769–1781
6. Bourbakis, N.: Detecting abnormal patterns in WCE images. In: 5th IEEE Symposium on Bioinformatics and Bioengineering (BIBE'05). (2005) 232–238
7. Lau, P.Y., Correia, P.: Detection of bleeding patterns in WCE video using multiple features. In: Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE. (aug. 2007) 5601 –5604
8. Li, B., Meng, M.Q.H.: Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. Computers in Biology and Medicine **39**(2) (2009) 141–147
9. Szczypinski, P., Klepaczko, A.: Selecting texture discriminative descriptors of capsule endpscopy images. In: Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on. (2009) 701 –706
10. Szczypinski, P., Klepaczko, A.: Convex hull-based feature selection in application to classification of wireless capsule endoscopic images. Advanced Concepts for Intelligent Vision Systems. Lecture Notes in Computer Science **5807/2009** (2009) 664–675
11. Iddan, G., Meron, G., Glukhowsky, A., Swain, P.: Wireless capsule endoscopy. Nature **405**(6785) (2000) 417–418
12. Swain, P., Fritscher-Ravens, A.: Role of video endoscopy in managing small bowel disease. GUT **53** (2004) 1866–1875
13. Bergmann, Ø., Christiansen, O., Lie, J., Lundervold, A.: Shape-adaptive DCT for denoising of 3d scalar and tensor valued images. Journal of Digital Imaging **22**(3) (06 2009) 297–308
14. Szczypinski, P., Strzelecki, M., Materka, A., Klepaczko, A.: MaZda - a software package for image texture analysis. Computer Methods and Programs in Biomedicine **94** (2009) 66–76
15. Duda, R., Hart, P., Stork, D.: Pattern Classification. John Wiley & Sons (2001)