International Journal of Computer Vision and Signal Processing, 2(1), 17-28(2013) ORIGINAL ARTICLE

Wireless Capsule Endoscopy Segmentation using Global-Constrained Hidden Markov Model and Image Registration

Yiwen Wan*, Prakash Duraisamy**

Department of Computer Science
University of North Texas
Denton-76203, Texas, USA

Mohammad S Alam

Department of Electrical Engineering and Computer Science
University of South Alabama
Mobile-36688, Alabama, USA

Bill Buckles

Department of Computer Science
University of North Texas
Denton-76203, Texas, USA

Abstract

This paper describes about analysis of wireless capsule endoscopy (WCE) using pattern recognition and statistical analysis. Specifically, we introduce a novel approach to discriminate between oesophagus, stomach, small intestine, and colon tissue present in WCE. Automatic image analysis can expedite this task by supporting the clinician and speeding up this process. Video segmentation of WCE into the four parts of the gastrointestinal tract is one way to aid the physician. The segmentation approach described in this paper integrates pattern recognition with statistical analysis. Initially, a support vector machine is applied to classify video frames into four classes using a combination of multiple color and texture features as the feature vector. A Poisson cumulative distribution, for which the parameter depends on the length of segments, models a prior knowledge. A priori knowledge together with inter-frame difference serves as the global constraints driven by the underlying observation of each WCE video, which is fitted by Gaussian distribution to constrain the transition probability of hidden Markov model. We also used image registration method to confirm our segmentation results. Experimental results demonstrated effectiveness of the approach.

Keywords: endoscopy image, segmentation, image,

IJCVSP
International Journal of Computer Vision and Signal Processing

ISSN: 2186-1390 (Online)
http://www.ijcvsp.com

Article History:
Received: 1 July 2012
Revised: 1 September 2012
Accepted: 20 September 2012
Published Online: 25 September 2012

1. INTRODUCTION

Wireless capsule endoscopy (WCE) [14] is an emerging technology for non-invasive inspection of the entire gastrointestinal (GI) tract. An exam by WCE commences with the patient swallowing a capsule (11 $\times$ 26mm). The capsule is integrated with a battery with 8-hour life, a camera, and an optical dome

*Corresponding author
**Principal corresponding author
Email addresses: YiwenWan@my.unt.edu (Yiwen Wan), prakashduraisamy@my.unt.edu (Prakash Duraisamy), malam@usouthal.edu (Mohammad S Alam), bbuckles@cse.unt.edu (Bill Buckles)

with light emitting diodes (Figure 1a). The capsule is propelled through the GI tract by normal paristalis (Figure 1b), capturing two 256 $\times$ 256-pixel frames per second. The frames are transmitted to a video recorder (shown in both Figure 1a and c) worn by the patient. When the exam is complete, the frames (in MPEG format) are uploaded to a PC where they may be inspected. One manufacturer (Given Imaging, Ltd.) provides a rapid reader (RR) for which a screen shot is shown in Figure 1c. While colonoscopy allows direct inspection of the large intestine, lesions in the esophagus, stomach, and small intestine are generally not recognized until symptoms appear. Early detection and patient survival are correlated. The incidence of colon cancer death is roughly 30% that of the incidence of new diagnoses. By contrast, the ratio of death incidence to new diagnoses is roughly 50% elsewhere in the GI tract [1, 15, 12]. The difference can be attributed in part to the lack of clinically available diagnosis methods comparable to colonoscopy. WCE is one approach to filling that void.

Figure 1: WCE system: (a) capsule and data recorder; (b) GI transit; (c) Rapid Reader from Given Imaging, Ltd.

Additional details of the clinical application of WCE are given in Section 2. It suffices for the moment to note that an exam by WCE produces a video of the GI tract consisting of approximating 50,000 frames. The miniaturized and swallowable camera travels in the GI tract with progression through the entrance/esophagus, stomach, small intestine, and the large intestine. However, a significant disadvantage of the technology is the time that the physician must spend examining the video. An experienced physician may spend 1-1.5 hours for each video. Automating inspection is a long term goal. That goal is not likely to be attained in the near future. In the meantime, steps toward automation are desirable and will have benefits such as reducing inspection time and improving the diagnostic accuracy.

The aim of this research is to determine methods for segmenting the video according to anatomical region. It enables other technologies such as adapting classification and diagnostic techniques to the differences in tissue among the organs. Indexing the video by segment allows a physician to locate, or re-locate, an organ for additional attention. The segmentation task presented here has a simple workflow – classify the frames independently by organ; use the classification results to condition the inter-frame probability of transition from one organ to the next. Videos from 15 patients were used in training and testing. The results indicate high accuracy and great promise for the technique.

2. Background

The gastrointestinal tract consists of four major zones: Entrance (Z1) – lies between the beginning of the video exam and the capsule entering the stomach. As only seconds are required for the capsule to travel through the esophagus to the entrance of the stomach, it is a small topographic zone. Stomach (Z2) – The stomach topographic area begins at the esogastric junction and ends at the pylorus. Small intestine (Z3) – The small intestine is the most important zone of the exam because a large number of events can be detected and conventional endoscopy does not reach much of it. Large intestine (Z4) – This topographic zone is bounded by the ileocecal valve and the end of the video.

Topographic segmentation research in the context of WCE

video is guided by two concepts. One, the event-detection approach [17], takes the video as a serial time signal. The other focuses on labeling all the frames in a WCE video into four classes. Typically, either concept involves feature extraction. Naturally, the performance of the second approach partially depends on the choice of classifier.

Approaches based on the first concept require that video frames be taken in time order to form a temporal signal. Signal processing techniques are used for detection of abnormality and events such as segment transitions. In [3], the authors deploy energy-based boundary detection for the classification of events. General and special features are extracted from video frames for classification. In this paper, the authors focus on two main events, bleeding and organ transitions. The latter segments the video. The classification phase involves color signal processing in which the hue component of the HSV color model is used to characterize peristalses and the color signal are processed after a fast Fourier transform.A rule-based assessment system is constructed for final classification using a high frequency content (HFC) function. Characteristics of color tones of the digestive organs are utilized to distingush between digestive organs in [13]. Dominant colors for every organ is learned and are combined to construct a representative signal to detect transitions between organs.

Approaches based on the second concept depend on the classification of WCE frames by organ. In [5], present a solution that considers the three boundaries – esogastric junction (B1), pylorus (B2), and ileocecal valve (B3). Homogeneous texture [28] and scalable color [4] from the MPEG-7 standard are the low-level features employed. The performance of a Bayesian classifier and a support vector machine are compared based on the single classification results.Believing that capsule exhibits motion pecurior to different organs, Mackiewicz in [23] developed a motion descriptor. Additionally, features were extracted from subimages within a frame that had the least obscuration. Because feature extraction is critical for classification-based approaches in topographical segmentation, many color and texture descriptors and even shape related descriptors have been applied to represent WCE frames informatively [21, 10, 2]. A color texture descriptor extended LBP was introduced specifically for WCE images in [10]. A custom set of Haar features are extracted from WCE images in [9].Beyond single classification scheme a classification casacade is proposed in [25] to classify capsule endoscopy images into semantic and topological categories.A unsupervised learning approach based on SIFT features is introduced in [26] to segment WCE video into four regions due to the availability of large enough labelled databases.

The three segments boundaries are esogastric junction (B1), pylorus (B2), and ileocecal valve (B3). Precisely locating them using only the classification results of the four topological zones is difficult. Cunha [5] defines an error function to be minimized in order to estimate boundary positions. In [23], a naive segmentation algorithm based on converging search, sliding windows, and a hidden Markov model is described for analyzing a sequence of single frame classifications. Yet, all three methods above are based on classification results. None systematically combine frame classification with boundary detection.Certain efforts have been made for WCE vidoe segmentation by detecting significant changes with respect extracted features as in [8, 27] As we understand whenever there are significant changes of color and texture there are changes for the incidence of events such as boundaries transitions.Incorporating global, model-based knowledge with classification will enhance performance.

To briefly summarize the contents of the remainder of this paper: In Section 3, we introduce single classification of WCE frames, temporal statistics models, and global-constrained hidden Markov models. These are incorporated into a methodology

that integrates classification with boundary detection. In Section 4, we describe the experimental design and describe the results. In Section 5, we draw conclusions and suggest directions for future research.

3. METHODOLOGY

The novelty of the approach described here is incorporating statistical analysis with pattern recognition for segmenting gastrointestinal tract videos into the four zones. We take advantage of the temporal statistics, consisting of model-based prior knowledge and temporal inter-frame changes of visual features. Prior knowledge is primarily based on the length of each zone. (In more general contexts, a zone is a scene.) Based on the understanding that different zones of the GI tract have distinguishable visual characteristics such as color and texture, we utilize concepts of pattern recognition to extract multi-component features from images to train a classifier that labels them into corresponding classes or zones. Choices of features and classifier are major factors. Each method – temporal statistics or classification – can alone be applied to segmentation. The combination of the two is better. Here we utilize a hidden Markov model as a basic framework to incorporate statistical knowledge with pattern knowledge in order that the three transition boundaries be located automatically and accurately.

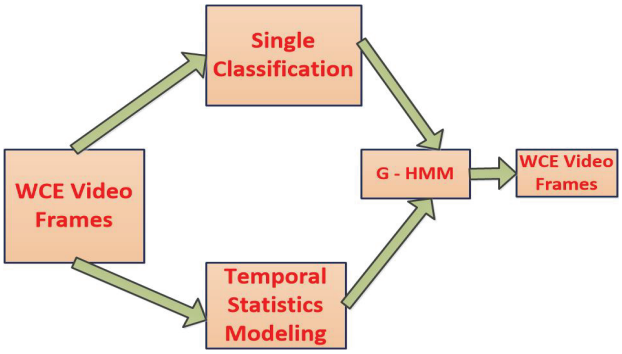

Figure 2: Framework

The choice of features and classifier influences the step that follows – temporal analysis. Thus, it is necessary to describe and justify the choices incorporated into the method.

3.0.1. Feature Extraction

Color and texture are generally conceded as best for distinguishing images from different organs. Four features are extracted from each frame. For color, the descriptors are scalable color (SC) [20] and hue saturation component (HS) [21]. For texture, the descriptors are local binary pattern (LBP) [22] and homogeneous texture (HT) [28].

RGB and HSV are frequently used color models. Previous research has indicated that HSV model is better for classification and in [4] it was concluded that the MPEG-7 standard descriptors, scalable color and homogeous texture, stand out as best for WCE images. HSV performs better because varying lighting conditions (as present in WCE) affect only the V componet. The scalable color descriptor is a histogram scaled into 256 bins – H component of 16 bins, S component of four bins, and V component of four bins. The HS color descriptor [21] is also based on the HSV model but the histogram only considers the H and S components. A DCT transform was applied (in [21] to compress the HS histogram. Both were deployed for the experiments and a performance comparison is described in Section 4.

Texture is another essential feature and the number of texture descriptors proposed is increasing. Some of them are specific gray-level images and others are on color images. Homogeneous Texture (HT) [28] is one of MPEG-7 standard fundamental tools [4] for describing multimedia content. The HT descriptor provides a quantitative characterization and actually a combination of a bank of Gabor filtered Fourier transform coeffients of gray-level images. Besides HT descriptor local binary pattern (LBP) is a color texture feature which is extracted from color images instead of gray images. In this paper, we extracted 3-D LBP histograms recently introduced by Connah and

Finlayson [6]. 3-D means LBP features are calculated individually for three different color channels such as R, G and B. The 3-D LBP histogram is in fact the joint histogram of the three independent histograms.

3.0.2. Classifier

Support vector machines (SVMs) have become a stable technology and provide good performance in general pattern recognition tasks. For WCE video analysis researchers often utilize SVMs for classification; in [21], Mackiewicz applied multivariate Gaussian classifier and SVMs to classify feature vectors into one of the four classes and concluded they provide superior performance. For the SVM, four widely used kernel functions were chosen and all were tested. Both [21] and [16] suggest that the radial basis kernel provides the best classification result. Thus, in this study we also applied a support vector machine on single image classification and radial basis is chosen to be the kernel function.

3.1. Modeling of Temporal Statistics

The modeling of Temporal Statistics consists of two major component: Prior knowledge and temporal inter-frame difference.The four major segments in a whole GI tract are entrance, stomach, small intestine and large intestine. The travel time of the capsule camera in different segments are different,however, segments of the same location are within a proportion range for most of normal people. Normally, it only takes the capsule camera minutes to travel in the entrance, 0.5-1 hours in stomach, 4-5 hours in the small intestine and 2-3 hours in the large intestine. Thus the knowledge of the approximate transition time for each part is what we first need to model statistically, which is called here the a priori knowledge. Temporal inter-frame difference is used here to quantitatively capture the change of color and texture between continuous frames so that based on the prior knowledge and the temporal inter-frame difference varying transition probability with time elapsing is constructed for each boundary individually, which will serve as the input for GHMM together with the single classification results.

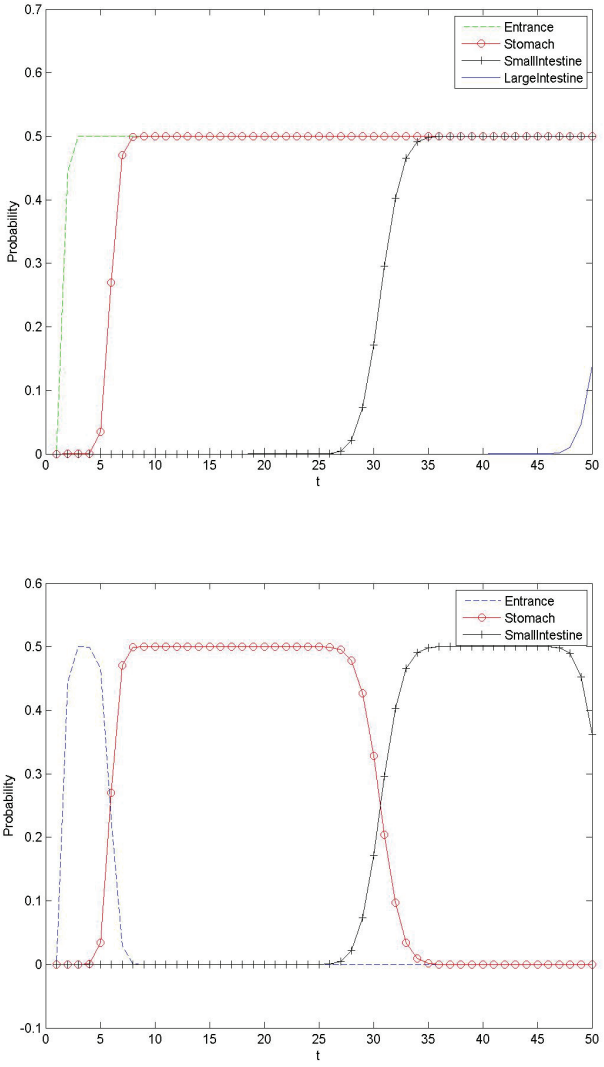

Figure 3: A Priori Knowledge for Each Segments: (a) Individually; (b) Correlation Considered;

3.1.1. A Priori Knowledge

The a priori knowledge is dependent on the individual lengths of the four zones. In order to fulfill this purpose, the a priori knowledge should satisfy the following criteria. One is for each major zone the a priori probability should be a function which monotonously, increases with time. The other criteria is

that while starting at the value 0 (zero) immediately after the previous boundary and ending at 0 (zero) and in between the value should converge towards the value 0.5. The value 0.5 indicates that within a range of time the a priori probability has no influence on the transition boundary detection. The a priori distribution can be modeled as cumulative probability. Poisson distribution was found in studies [24] to match the distribution of shot lengths. Since there are four major segments in the GI tract with different lengths, it should be treated as four different scenes. There should be a different a priori distribution for each scene respectively.

$$P_s = \frac{1}{2}\sum_{f=\lambda_s}^{C(s)} \frac{\lambda^f}{f!} e^{-\lambda} \quad (1)$$

The parameter λ_s of the Poisson distribution represents the average length of each scene s (s=1, 2, 3, 4) where 1, 2, 3, 4 represent entrance, stomach, small intestine, and large intestine respectively. f_{λ_s} is the frame counter and for each scene it starts from the end of previous scene. And C(s) is the current scene length at the frame f.

3.1.2. Temporal Inter-Frame Difference

Inter-frame difference has been used for shot detection or transition detection to capture where the most significance changes happen. Various methods were generated for measuring the difference between continuous frames such as template measurements, color histogram measurement and χ^2 [11]. However, in this study we utilized single classification results to measure the difference of continuous frame, which actually measures the similarity between continuous frames in terms to color and texture.

With the classification results that each frame of a video at time t (counter of frames) was assigned with four probabilities of belonging to four major segments respectively [21] denote as P_{t1}, P_{t2}, P_{t3}, P_{t4}

$$D_{12}(t, t+1) = U(P_{t,1} - P_{t+1,1}) + U(P_{t+1,2} - P_{t,2}) \quad (2)$$

$$D_{23}(t, t+1) = U(P_{t,2} - P_{t+1,2}) + U(P_{t+1,3} - P_{t,3}) \quad (3)$$

$$D_{34}(t, t+1) = U(P_{t,3} - P_{t+1,3}) + U(P_{t+1,4} - P_{t,4}) \quad (4)$$

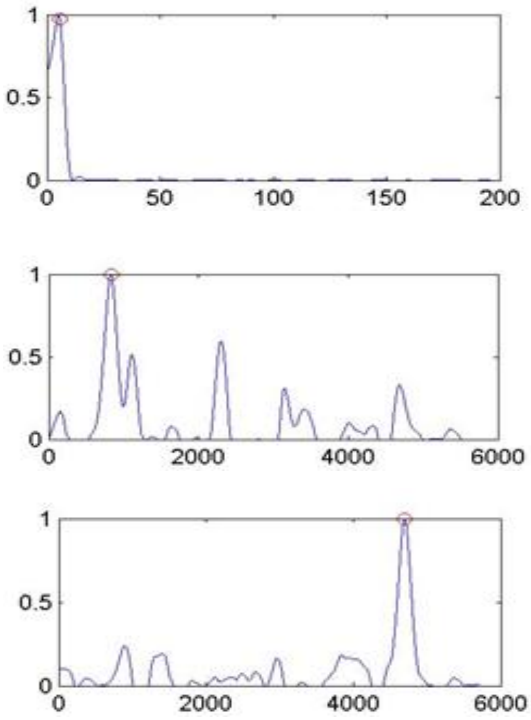

Figure 4: Inter-Frame Difference: (a) D12; (b)D23 ; (c) D34

D_{ij} is a inter-frame difference curve to capture the possibility of transition from segment i to segment j. Intuitively, whenever capsule camera transits from one segment to the next probability of belonging to segment i at moment t $P_{t,i}$ will drop off when it comes to moment $t+1$. In the same way probability of belonging to segment j at moment t $P_{t,j}$ will arise up when it comes to moment $t+1$. D_{ij} were calculated to add two difference measured together and thus gives us an indicator of where there is significant changes of color and texture with respects to characteristics in four different segments.

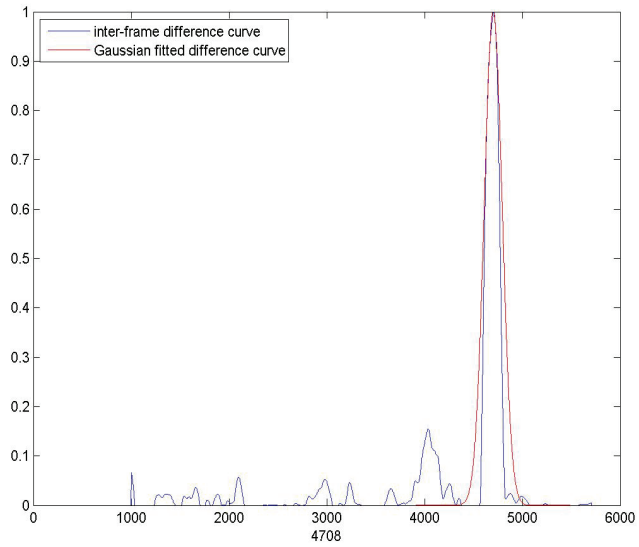

Figure 5: Gaussian Fitting for Global Constraints

3.2. Global-Constrained Hidden Markov Model

Hidden Markov models [18] have been applied in video segmentation. In [21, 29] HMM is used as a framework to refine classification results but transition probabilities are fixed at any time. However, transition probability shouldn't stay the same since we have prior knowledge that traveling time in entrance is about minutes and traveling in small intestine costs the most always. And we know that transition from one segment to the next should happen at the time when there are significant changes of contiguous frames in terms of color and texture characteristics. Thus, a priori knowledge and inter-frame difference should help model a constrained transition probability for HMM.The transition probability used in our method varies when time elapses instead of being fixed, thus it is called in this paper global-constrained hidden Markov model (GHMM).

After the a priori knowledge and inter-frame difference were calculated as what is described in the previous section,Poisson fitting was finally applied on the combined curve of a priori knowledge and inter-frame difference to acquire the global constraints for the transition probability.

In figure 5 segmentation results of one WCE video compared with the ground truth is presented and we see that two boundaries were perfectly found and there were two errors for the boundary between stomach and small intestine.

3.3. Image Registration for classification of Endoscopic Images

Image registration is the process of mapping the geometric features from multiple images into a common coordinate system. It is the basic step for integrating several overlapping images into a large composition of a 3D scene. It plays a very important role in many computer vision applications like image fusion, medical imaging, map updating, and multichannel image restoration.

In general, image registration methods consist of four basic steps: feature extraction, feature matching, transform modeling and image resampling .In the first step, (feature extraction) salient and distinct features are extracted. In the second step (feature mapping), the common features extracted are mapped to establish a correspondence between the overlapping images. From the recent literatures, various features are used to form the correspondence between the overlapping images such as SIFT descriptors (scale invariant feature transform), Harris corners, ICP algorithm (Iterative closest point), Mutual Information, Least Square Error (LSE), and normalized cross correlation (NCC). In the third step (transform modeling),mapping function parameters are computed from the available feature correspondence computed in the second step. Finally, (image re-sampling) the sensed image is registered through the mapping function [7].

In our work, we used image registration to classify the endoscopic images from discriminating oesophagus, stomach, intestine and colon tissue. We use different sample of endoscopic images for our experiment as shown in Fig 6. We use SIFT algorithm for registration purpose. In our experiment, we used two sample images as input to SIFT algorithm. SIFT algorithm

gives number of feature points and number of matches between two images. In our experiment, we classify the regions based number of feature points present in the images. If the difference between number of feature points are less on a sample set, then it belongs to same region of endoscopic images. If the differences are higher, then it belongs to a different region of endoscopic images or it belongs to a border region.

The Table 1, confirms the above discussion. Among all the data sets, fifth data set ($I5, I6$) have higher difference between feature points which indicates both images belongs to different regions. The remaining data sets (1-3) corresponds to oesophagus and data sets (6-8) are belongs to stomach region based on feature point difference. In our work, we used image registration to classify the endoscopic images for classifying oesophagus, stomach, intestine and colon tissue. We use different sample of endoscopic images for our experiment.

4. Experiments and Results

The 15 WCE videos were divided into a training dataset (10 videos) and a testing dataset (five videos). Training dataset were built up by selecting 5% images data from videos in the training dataset. Feature: Scalable color, hue saturation component, homogeneous texture and local binary pattern color texture were extracted and comparison of the performance between these features were accomplished based on training dataset. Principal component analysis was applied on each feature vector and the length of each feature vector is 10. We trained the support vector machine classifier with cost=1 and gamma=1 and 10-fold cross validation were tested on downsampled dataset of original training dataset for the comparison of the performance of four features. As what is presented in Fig.7, classification accuracy using four features individually were 50.94% (HT), 76.86% (SC), 79.39% (HS), and 87.97% (LBP), respectively. Apparently HS is better than SC as a color feature and LBP is much better than HT as a texture descriptor. And among the four, LBP performed the best based on our dataset. Expectedly the combination of four features gave us better classification accuracy than any single feature.

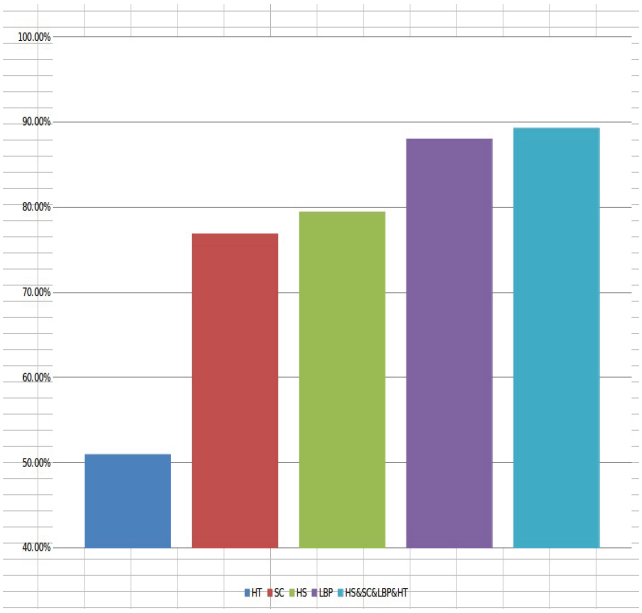

Figure 7: Feature Performance Comparison Based on Classification Accuracy

Table 2: Results of the segmentation – SVM:classification accuracy when SVM used only; GHMM: refined accuracy after GHMM was applied; E_{12}: Esogastric junction error; E_{23}: Pylorus error; E_{34}: Ileocecal valve error

Video	SVM	GHMM	E_{12}	E_{23}	E_{34}
V_1	71.80%	99.95%	0	22	8
V_2	91.29%	99.23%	6	430	4
V_3	93.97%	97.41%	4	226	1250
V_4	57.26%	98.14%	32	838	188
V_5	59.06%	98.84%	2	0	644
Median	71.80%	98.84%	4	430	188
Mean	74.68%	98.71%	8	303	419

From comparison, we used a support vector machine as a classifier to label WCE frames into four classes (entrance,stomach,small intestine and large intestine). Predicted probabilities of four classes related with each frame were also provided by SVM. For

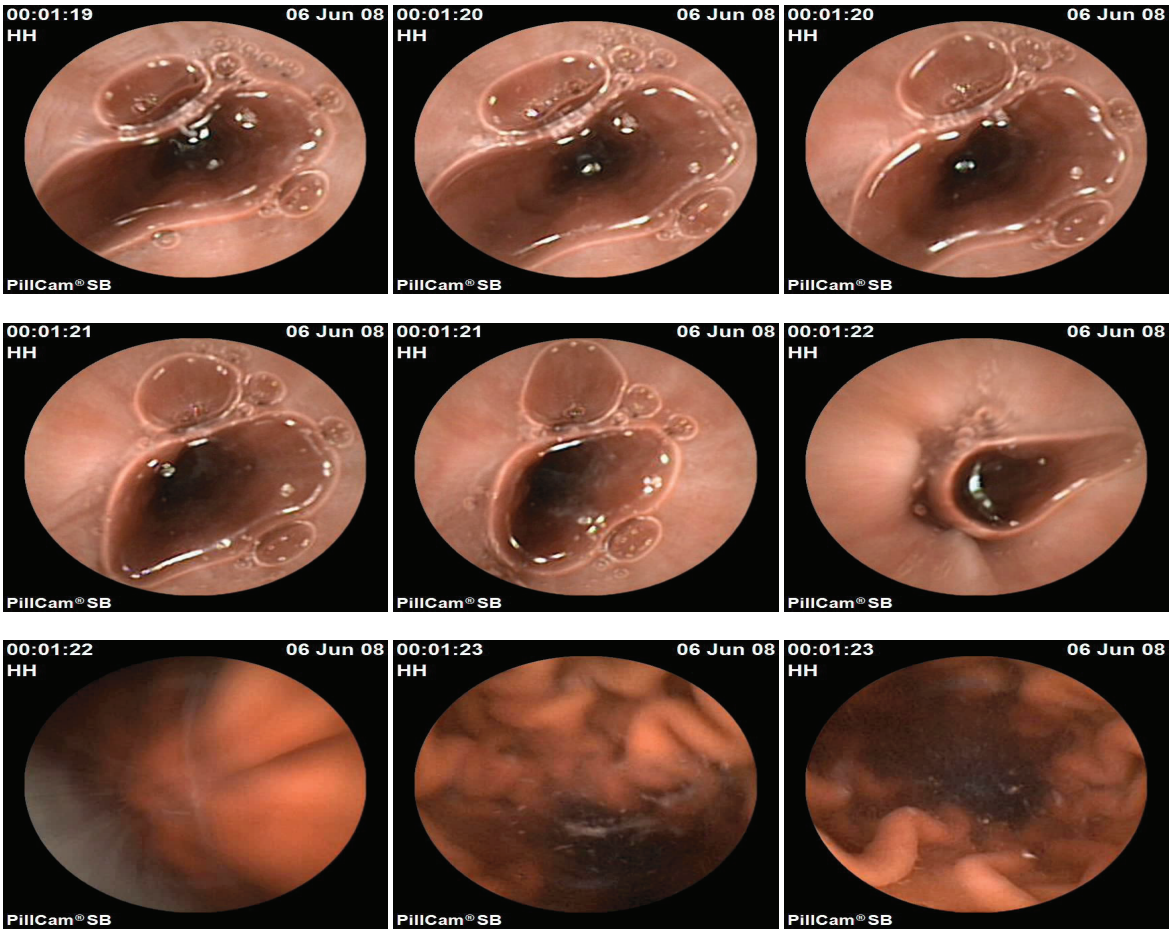

Figure 6: Classification using Image Registration: (a) first sample image from Oesophagus; (b) second sample image from Oesophagus;(c) third sample image from Oesophagus;(d) fourth sample image from Oesophagus;(e) fifth sample image from Oesophagus; (f) first sample image from stomach; (g) second sample image from stomach; (h) third sample image from stomach; (i) fourth sample image from stomach.

example one frame is labeled as stomach with the four probabilities 0.05,0.90,0.05,0, which means it has the highest probability 0.9 to be stomach. In order to improve classification result to the best, it is necessary to tune appropriately the parameters of SVM: gamma and cost. Grid search was used to search for the best combination of these two parameters. We processed every tenth frame of each video taking time consumption into consideration. After SVM classification, GHMM was applied to get the final segmentation result. We compared our segmentation result of each video with the ground truth from the expert. As what is shown in Fig.8 green segments represent ground truth and blue ones is the results of our approach. Besides 5 testing videos were tested and single classification result after SVM,refined relative error after GHMM and exact errors of three boundaries Z_{12},Z_{23} and Z_{34} were reported for all testing videos in Table 2.

Classification accuracy from the SVM for testing videos is not that satisfying since we have only 10 videos as the training dataset. Two testing videos have classification more than 90% and the others are not that good. But after applying GHMM the median and mean accuracy was refined to 98.84% and 98.71%. What's more, mean error around Esogastric junction is 8, which is within the error tolerance 10 since frame processing rate in this experiment is 10 fps. More errors are detected around bound-

Table 1: Results of the classification using image registration

Data Sets	No of feature points	Number of Matches	Feature point difference	Endoscopic Region
$I1, I2$	$533, 603$	14	70	Oesophagus
$I2, I3$	$603, 486$	17	117	Oesophagus
$I3, I4$	$486, 490$	16	4	Oesophagus
$I4, I5$	$490, 491$	14	1	Oesophagus
$I5, I6$	$491, 206$	4	285	Oesophagus,Stomach
$I6, I7$	$206, 96$	0	110	Stomach
$I7, I8$	$96, 155$	0	59	Stomach
$I8, I9$	$155, 131$	0	24	Stomach

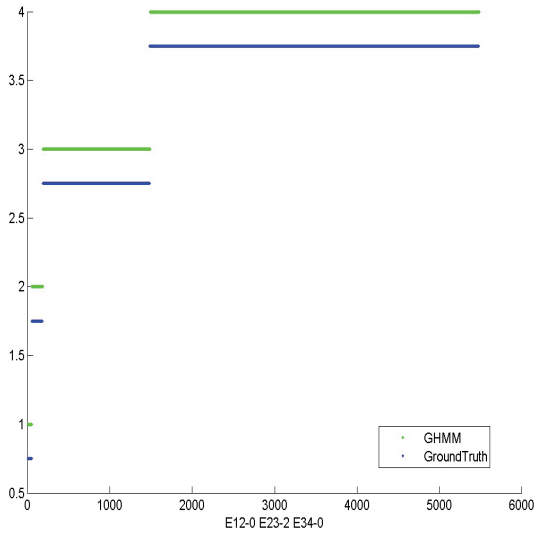

Figure 8: Segmentation Results Compared with Ground Truth

aries Pylorus and ileocecal,however, there are 50,000-60,000 frames in each video so that $Z_{23} = 303$ and $Z_{34} = 419$ is still satisfying. However, more efforts should be put on how to reduce errors around Pylorus and ileocecal valve.

Comparing to the recent work in analysis of WCE in [19], eventhough the accuracy was greater than 0.9, this approach need lot of training set for classification. In contrast, our approach we are able to classify different classes from less number of training sets with the same accuracy.

5. CONCLUSIONS AND FUTURE WORK

We have proposed a new method to segment a WCE video into four meaningful segments, which combining pattern recognition and statistical analysis. It has simple work flow and presents promising results. HS is a better color descriptor than SC; and LBP performed better than HT as a texture descriptor based on our dataset.Our proposed approach improved the overall classification accuracy and reduced errors around the three boundaries to an acceptable level. GHMM improved the classification accuracy by combining classification results and statistical knowledge.

Besides that we will apply more features to improve the single classification results so that final segmentation result would be refined further. Our research focus in the future will also include reducing redundancy between frames since it happens that the capsule camera gets stuck somewhere for a long time or moves forwards and backwards within the same region. When redundancy can be reduced appropriately, prior knowledge and statistical analysis will be more reliable and thus final segmentation results will be better.

In future, we plan to register different frames of WCE video based on their overlap and build 3D images for more accurate segmentation. 3D views of scene will aid segmentation and it

will help a clinician make better decisions.

References

[1] R. Siegel A. Jemal, J. Xu E. Ward, T. Murray, and M. J. Thun. Cancer statistics. *A Cancer Journal for Clinicians*, 57:43–66, 2007.

[2] A.Klepaczko, P.Szczypi, P.Daniel, and M.Pazurek. Local polynomial approximation for unsupervised segmentation of endoscopic images. pages 33 – 40, 2010.

[3] Y. Chen, W. Yasen, J. Lee, D. Lee, and Kim. Y. Developing assessment system for wireless capsule endoscopy videos based on event detection. In *Proceeding of SPIE Conference.*, volume 7260, 2009.

[4] M. Coimbra, P. Campos, and J. P. S. Cunha. Mpeg-7 visual descriptors-contributions for automated feature extraction in capsule endoscopy. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(5), 2006.

[5] M. Coimbra, P. Campos, and J. P. S. Cunha. Topographic segmentation and transit times estimation for endoscopic capsule exams. In *in Proceeding of IEEE International Conference on Acoustics, Speech and Signal Process*, volume 2, pages 1164–1167, 2006.

[6] D. Connah and G. D. Finlayson. Using local binary pattern operators for colour constant image indexing, 2006.

[7] Prakash Duraisamy, Kamesh Namuduri Stephen Jackson, Bill Buckles, and Ye Yu. Map updation using image registration for different sensors by camera calibration. *in Proc.WCSP 2011*.

[8] G. Gallo and E.Granata. Wce video segmentation using textons. 7623, 2010.

[9] G.Gallo and A.Torrisi. Random forests based wce frames classification. *25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1– 6, 2012.

[10] M. Hafner, M. Liedlgruber, A. Uhl, A. Vecsei, and F. Wrba. Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Medical image analysis*, 16(1):75 – 86, 2012.

[11] H.Jiang, A.Helal, A.K.Elmagarmid, and A.Joshi. Scene change detection techniques for video database systems. *Multimedia Systems*, 6:186–195, 1998.

[12] M.J. Horner, L.A.G. Ries, M. Krapcho, N. Neyman, R. Aminou, N. Howlader, S.F. Altekruse, E.J. Feuer, L. Huang, A. Mariotto, and et al. SEER cancer statistics review, 1975-2006. Technical report, National Cancer Institute, Bethesda, MD, 2006.

[13] H.Vu, Y.Yagi, T.Echigo, M.Shiba, K.Higuchi, T. Arakawa, and K.Yagi. Color analysis for segmenting digestive organs in vce. *2010 20th International Conference on Pattern Recognition*, 2010.

[14] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain. Wireless capsule endoscopy. *Nature*, 405:725–729, 2000.

[15] A. Jemal, R. Siegel, E. Ward, Yongping Hao, Jiaquan Xu, Taylor Murray, and Michael J. Thun. Cancer statistics. *A Cancer Journal for Clinicians*, 58:71–96, 2008.

[16] J.P.S.Cunha, M. Coimbra, P. Campos, and J.M. Soares. Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE Transactions on Medical Imaging*, 27(1), 2008.

[17] J. Lee, J. Oh, S. Shah, X.Yuan, and S.Tang. Automatic classification of digestive organs in wireless capsule endoscopy videos. In *Proceedings of the ACM symposium on Applied computing.*, pages 1041–1045, 2007.

[18] L.R.Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *in Proc. IEEE*, 77(2):257–286, 1989.

[19] M. Mackiewicz, J. Berens, and M. Fisher. Wireless capsule endoscopy color video segmentation. *IEEE Transactions on Medical Imaging, Vol. 27, No. 12, December 2008*, 2008.

[20] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, and A. Yamada. Color and texture descriptors. *Transaction of Circuit Systems for Video Technology*, 11/6:703–715, 2001.

[21] M.Mackiewicz, J.Berens, and M.Fisher. Wireless capsule endoscopy color video segmentation. *IEEE Transactions on Medical Imaging*, 27(12), 2008.

[22] M.P.T. Ojala and T. Maenpaa. Multiresolution grey-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002.

[23] J. Vitria P. Spyridonos, F. Vilarino and P. Radeva. Identification of intestinal motility events of capsule endoscopy video analysis. In *Proc. ACIVS*, pages 531–537, 2005.

[24] A. Papoulis. Probability, random variables and stochastic processes. In *International Editions ed. New York: McGraw-Hill*, 1984.

[25] P.W. Mewesand P.Rennert, A.L. Juloski, A. Lalande, E. Angelopoulou, R.Kuth, and J.Hornegger. Semantic and topological classification of images in magnetically guided capsule endoscopy. V- 8315:83151A–11, 2012.

[26] Y. Shen, P.P. Guturu, and B.P. Buckles. Wireless capsule endoscopy video segmentation using an unsupervised learning approach based on probabilistic latent semantic analysis with scale invariant features. *IEEE transactions on information technology in biomedicine*, 16(1):98–105, 2012.

[27] S.R. Stanek, W. Tavanapong, J.Wong, J. Oh, and P.de Groen. Automatic real-time detection of endoscopic procedures using temporal features. *Computer methods and programs in biomedicine*, 108(2):524 – 35, 2012.

[28] M. Kim Y. Ro and H. Kang. Mpeg-7 homogeneous texture descriptor. *ETRI Journal*, 23(2), 2001.

[29] Q. Zhao, T. Dassopoulos, G. Mullin, G. Hager, M. Meng, and R.Kumar.

Towards integrating temporal information in capsule endoscopy image analysis. *IEEE Engineering in Medicine and Biology Society*, pages 6627–30, 2011.