33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Towards Integrating Temporal Information in Capsule Endoscopy Image Analysis

Qian Zhao[2], Themistocles Dassopoulos[3], Gerard Mullin[2], Greg Hager[1], Max Q–H Meng[2], Rajesh Kumar[1]

***Abstract*— Analysis of Wireless Capsule Endoscopy (CE) images has become a very active area of research since this novel technology enabled access to previously inaccessible areas of the gastrointestinal tract, especially the small intestine. Art has investigated automatic segmentation of organ boundaries, detection of lesions and bleeding as well as other supervised and unsupervised analysis. However, all of this art has focused on treating the images as individual and independent observations that contribute towards a unique and separate decision. Given the overlap between the images, this is clearly not the case. A human, by contrast, performs assessment by combining the information seen in all neighboring views of the anatomy in a study. This article makes two significant contributions. Towards combining information from multiple images, we propose a supervised classification approach using an HMM framework. Secondly, we use a weak (k-NN) classifier to prototype and evaluate such a framework for regions of the GI tract containing polyps. The combined framework significantly improves the performance of the individual classifier and experiments show promising performance with accuracy > 0.9.**

I. INTRODUCTION

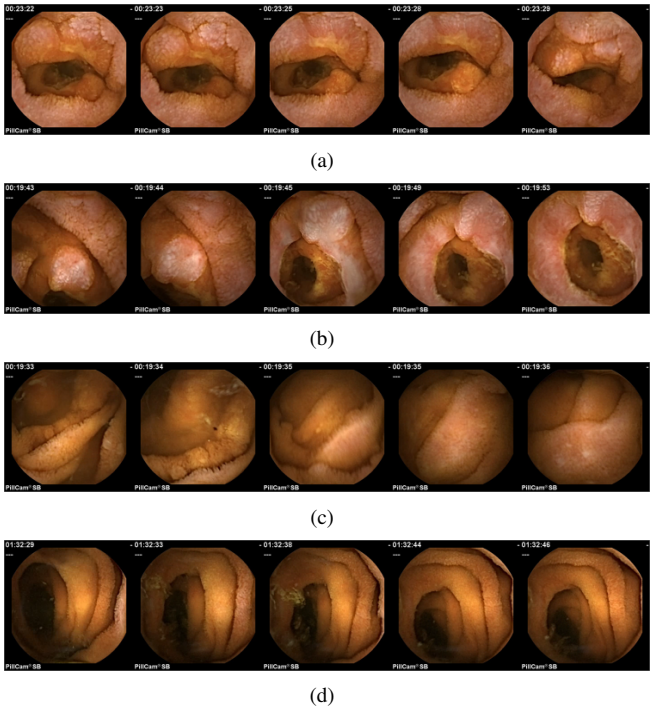
Poor gastrointestinal (GI) tract health has an impact beyond digestive health, and often causes poor overall health. As a result, capsule endoscopy (CE) images have become an interesting treasure trove of information for researchers creating tools to diagnose GI conditions such as bleeding, ulcers or other irritable bowel syndrome conditions. CE was introduced by Given Imaging in 2000 [1], and granted the US Food and Drug Administration (FDA) approval in 2001. Over a million GIVEN PillCam Small Bowel (SB) alone devices have been used worldwide by 2010, and CE is on track towards becoming a common technique for the diagnosis of many diseases of the small bowel [2], and more than 1000 peer-reviewed publications have appeared in the medical literature [3].

The disposable CE device is size about the same as a large vitamin pill measuring 26 mm $\times$ 11 mm, and consists of a video camera, an illumination source, batteries and wireless communication electronics. The GIVEN PillCam SB used in these experiments transmits 576 $\times$ 576 color images at 2 images per second with a field of view of 140 degrees, magnification of $\times 8$, and a depth of view is 1 to 30 mm.

This work was supported by Johns Hopkins University internal funds and by the Hong Kong government ITF project #6902928 awarded to Max Q.-H. Meng.

Qian Zhao was a visiting researcher at the Visual Imaging and Surgical Robotics laboratory, Johns Hopkins University during this work. She is a graduate student at the The Chinese University of Hong Kong, Shatin, N.T., Hong Kong(*qzhao@cuhk.edu.hk*).

[1] authors are with Jonh Hopkins University, Baltimore MD 21218 (*rajesh@jhu.edu*). [2] authors are with The Chinese University of Hong Kong. [3] author is with the Washington University School of Medicine.



(a) (b) (c) (d)

Fig. 1. Examples of CE image sequences - polyp (a, b) and normal (c,d) sequences.

An outpatient procedure typically generates data for up to 8 hours limited by battery life, and typically produces more than 50,000 images for each patient. The video is reviewed offline on a workstation, where a physician can perform diagnosis and assessment [2].

Literature [4]–[6] reports CE image and video analysis for automation of assessment for conditions such as bleeding [7], lesions [8], polyps [9]–[12] as well as topographical segmentation [13]. For example, [14] et al describe a cascade method for informative frame detection which uses local color histogram to isolate highly contaminated non-bubbled frames, and [13] et al perform topographic segmentation of the gastrointestinal tract to detect organ boundaries. We have also previously investigated CE studies for lesion detection and classification, duplicate detection and other assessment [8], [15]–[18].

Literature has so far treated CE images as individual independent observations. The information in one image is not related to another image, even though there is significant overlap in the neighboring images (Figure 1) in most images. Several neighboring images (a *sequence*) capture clinically relevant information for the same anatomy and these are viewed by a clinician together in a multi-image (2,or 4 frame grid) view for diagnosis.

This paper, therefore, significantly extends the art by combining the information in CE image sequences for a common diagnosis. To our knowledge, there is no comparable art on supervised classification of *CE image sequences*. Below, we describe a Hidden Markov Model (HMM) framework to classify CE polyp image sequences.

II. METHODS

We use a supervised classification framework. Supervised classification requires generation of a set of features representing the image information in fewer dimensions, to be used for training of and validation of the statistical classifier.

Polyps result in variation of shape, color, and texture in CE images. Literature reports many color, texture, shape and edge features [8], [13], [15], [19], [20]. Here, we used customized color, texture, and edge features reported in [8], and used in that work for representing lesion characteristics.
Color Features: The MPEG-7 [21] dominant color descriptor (DCD) clusters image color information in LUV space with a generalized Lloyd algorithm. The descriptor contains the representative colors, their percentages in the image and several other statistical measures. Our customized version uses 6 colors and their percentages in 24 dimensions as the feature vector.
Edge Features: The MPEG-7 edge histogram descriptor (EHD) extracts the spatial distribution of five types of edges(0, 45, 90, 135 degrees and remaining non-directional edges). Edge features play an important role in detection of the circular bumps and erosions resulting from polyp boundaries. Our customized descriptor uses a histogram of edge distributions for 16 image blocks in 80 dimensions.
Texture Features: The MPEG-7 homogeneous texture descriptor (HTD) uses Gabor filters. We limit our implementation to 3 channel filters each in two radial scales for a compact representation of 12 dimensions.

This feature extraction results in a 24 dimension color, a 80 dimension edge, and a 14 dimension texture representation. We combined our features by concatenation (118 dimensions), and then employed a Laplacian Eigenmap method to reduce the total dimensions to 13 suitable for further analysis.

In [8], [22], [23], we reported several individual and combined classifiers for assessment of individual CE images. This includes the common support vector machines [8], Bayes, and k-NN classification as well as combined classifiers using Adaboost [16], [23]. We use a k-NN classifier here. The aim is to combine the information in many images rather than focusing on the classification of an individual image or a particular classifier, and classifier(s) may be replaced or added as appropriate in subsequent investigations.

A. k-NN Classifiers

The k-NN classifier is based on non-parametric density estimation. It is an intuitive method because examples are classified based on their similarity with training data. The k-NN method only requires an integer k, a set of labeled examples and a measure of "closeness". k-NN is also a stable classifier [24], which means that the parameter k does not greatly vary the performance of the classification. We use a 3-nearest neighbor implementation in the experiments below.

B. Hidden Markov Models

HMMs are widely used in temporal pattern recognition in speech, handwriting, gesture recognition, and bioinformatics [25]–[29]. For example, [29] presents a system that segments a soccer video into parts, representing active playing sequences and breaks, and [30] segments a CE study into GI organs.

Before a model is trained, we specialize the HMM for abnormality (polyp) detection. Our model is summarized below:

- Hidden states (Figure 2), denoted by $S = \{\text{Polyp}, \text{Normal}\}$ ($N = 2$) annotated by the ground truth are defined as polyp (P, state 1) images labeled as "polyp" in the ground truth, and normal (N, state 2) images labeled as "normal" in the ground truth.
- Observations, denoted by $O = \{\text{polyp}, \text{normal}\}$ ($M = 2$), are the outputs of the binary k-NN classifier.
- A state transition matrix characterizes the temporal relationship between hidden states for the polyp sequences in our study.
- The Observation matrix represents the performance of the classifier, i.e., whether images are classified correctly by the current classifier. As usual, its four entries are true positive, false positive, false negative and true negative, respectively.
- An initial state distribution means the probability of each hidden state (polyp or normal) occurring at the beginning in any polyp sequence.

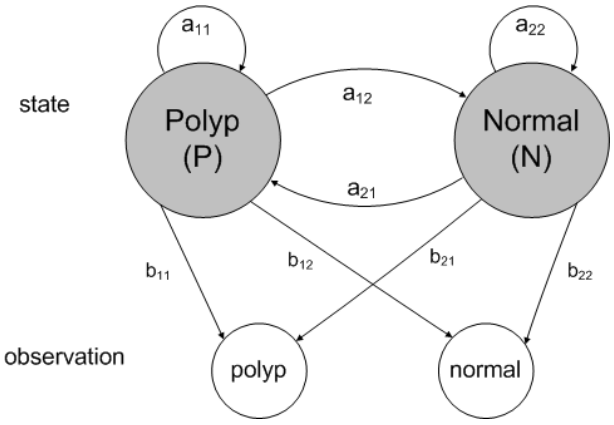


Fig. 2. The HMM for Polyp Analysis

We use multiple first-order discrete HMMs which perform classification based on $b_j(k)$ and apply transition constraints a_{ij}.The classification process is divided in a training (learning stage) and evaluation stage as usual. We first train an HMM model with the labeled dataset. Our experimental data contains annotated polyp and non-polyp (normal) sequences. Each image in a sequence is classified using a binary classifier trained on the features described above. The output of the classifier forms the observation for the HMM. In the

experiments below, we report on both constant and varying lengths of sequences.

Given a set of sequences $\{O_i\}$, an HMM is trained using the Baum-Welch algorithm [25], [26] which determines the parameters λ that maximize the probability $P(\{O_i|\lambda\})$. A Generalized Expectation Maximization algorithm was used to uncover the best structure of the hidden states. In the evaluation stage, the trained HMM uses the standard Viterbi algorithm [31] to estimate the most probable sequence of states for the test dataset for classification of sequences as usual.

III. EXPERIMENTS AND RESULTS

The Johns Hopkins CE archive contains 47 CE studies [15], and over 80,000 annotated images collected using a Johns Hopkins Institutional Review Board (IRB) protocol. We constructed two relevant data sets from these images. A dataset with 200 polyp and 200 normal CE images was built for k-NN classifier training and validation. These images are not necessary consecutive. A second dataset consists of 1120 images (560 polyp images and 560 normal images) in which there are 224 representative polyp (112) and normal (112) sequences, each consisting of 5 consecutive images. Images containing polyps are labeled as a positive sample. A single expert reviewer assigned the ground truth, and annotation by multiple clinical experts is ongoing in related work.

We then trained and tested both k-NN image classification and HMM sequence classification on this data. All classification methods were implemented and tested using MATLAB (MathWorks Inc., NATICK, MA) on a Windows dual core workstation with 4 GB RAM. The UBC HMM Toolbox (Murphy et al, http://www.cs.ubc.ca/∼murphyk/Software/HMM/hmm.html, 1998) provided the HMM framework. As usual, we computed the accuracy measure as:

$$Accuracy = \frac{\text{Correct Classifications}}{\text{Total Number of Sequences}} \quad (1)$$

$$Precision = \frac{\text{Correct Positive Classifications}}{\text{Positive Classifications}} \quad (2)$$

$$Recall = \frac{\text{Correct Positive Classifications}}{\text{Number of Positives}} \quad (3)$$

On individual images, our trained k-NN classifier produces an accuracy of 0.833. We then trained an HMM using variable length image sequences, i.e. $T \in \{2,3,4,5\}$ using the trained k-NN classifier. The state transition matrix generated during the training phase was:

$$A = \begin{bmatrix} 0.9871 & 0.0129 \\ 0.0157 & 0.9843 \end{bmatrix} \quad (4)$$

We can see (from (4)) that a polyp sequence may contain some normal views($a_{12} = 0.0871$), and vice versa, although the variation in length 5 sequences is not very high. The corresponding observation matrix, and the initial state distribution were:

$$B = \begin{bmatrix} 0.9129 & 0.0871 \\ 0.0329 & 0.9671 \end{bmatrix} \quad (5)$$

$$\pi = \begin{bmatrix} 0.4865 & 0.5135 \end{bmatrix} \quad (6)$$

which also agree with our expectation. The probabilities that a sequence begins with polyp or normal image are almost equal.

We evaluated the trained HMM model using sequences of varying lengths (1 to 5) by omitting trailing images in the evaluation. Table I shows the comparison of accuracy, precision and recall for different lengths. The length=1, or single image result represents the the performance of the k-NN classifier on individual images. All three measurements increase when classification uses multiple images, and no significant improvement is observed after 4 images in a sequence.

TABLE I

PERFORMANCE OF CLASSIFICATION USING SEQUENCES OF LENGTH 1 TO 5. THE FIRST RESULT SHOWS THE k-NN CLASSIFIER PERFORMANCE.

Sequence Length	1	2	3	4	5
Accuracy	0.833	0.792	0.875	0.917	0.917
Precision	0.910	0.833	0.917	0.945	0.945
Recall	0.833	0.810	0.893	0.917	0.917

Figure 3 shows a polyp sequence that was classified correctly. The estimated classification results by kNN is {P P N P P} (P for polyp and N for normal). The HMM result improves the accuracy as it integrates the relationship between consecutive images.

Fig. 3. A correctly classified polyp sequence.

By contrast, Figure 4 shows a polyp sequence {P P P P P} that was misclassified as normal. The k-NN classified the images as {P N N P P}, and after the HMM classification, the estimated most probable state sequence is {P N N P P}.

Fig. 4. An incorrectly classified polyp sequence

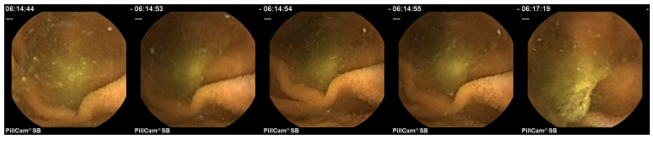


Fig. 5. A misclassified normal sequence containing extraneous matter.

Fig. 5 shows an example of a normal sequence that was misclassified. The output of k-NN classifier in this case is {N P P P N}, while the output of the HMM is {P P P P P} which classifies it to polyp sequence.

As is expected, the HMM accuracy depends on the individual k-NN classifier performance. In the misclassification above (Figure 4) the misclassified images have no weak polyp characteristics. Similarly, the misclassified normal sequence contains yellowish extraneous material that creates edge distributions and texture disruption similar to the polyps. Additional training, and combined classifiers [23] can easily remove these individual classifier limitations.

IV. DISCUSSION

We describe a novel framework for CE image sequence classification and its application to polyp detection. A k-NN classifier is embedded in a HMM framework which can recognize polyp or normal image sequence of various length. Preliminary experiments show promising performance of the framework, with an accuracy > 0.9.

In ongoing work, we are expanding our experimental dataset to include additional images, abnormalities(lesions, and bleeding), and multiple expert assessments. We will also integrate our work in [23] by using alternative features, and individual, and combined classifiers in this framework.

Our end goal is to create a semi-automated tool for quantitative assessment of pathologic findings using only CE imaging. We aim to continue further development of our statistical classification methods, combining individual classifiers, and validation on the larger datasets towards this goal.

REFERENCES

[1] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain. Wireless capsule endoscopy. *Nature*, 405(6785):417, 2000.

[2] P Swain. Wireless capsule endoscopy. *Gut*, 52(4):48–50, 2003.

[3] Miguel Munoz-Navas. Capsule endoscopy. *World journal of gastroenterology*, 15:1574–6, 2009 Apr 7.

[4] S.A. Karkanis, D.K. Iakovidis, D.E. Maroulis, D.A. Karras, and M. Tzivras. Computer-aided tumor detection in endoscopic video using color wavelet features. *Information Technology in Biomedicine, IEEE Transactions on*, 7(3):141 –152, 2003.

[5] M Tjoa and S Krishnan. Feature extraction for the analysis of colon status from the endoscopic images. *BioMedical Engineering OnLine*, 2(1):9, 2003.

[6] D.K. Iakovidis, D.E. Maroulis, S.A. Karkanis, and A. Brokos. A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. In *Computer-Based Medical Systems, 2005. Proceedings. 18th IEEE Symposium on*, pages 575 – 580, 2005.

[7] Phooi Yee Lau and P. L. Correia. Detection of bleeding patterns in wce video using multiple features, 2007. ID: 1.

[8] S. Bejakovic, R. Kumar, T. Dassopoulos, G. Mullin, and G. Hager. Analysis of crohn's disease lesions in capsule endoscopy images. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pages 2793 –2798, May 2009.

[9] L.A. Alexandre, N. Nobre, and J. Casteleiro. Color and position versus texture features for endoscopic polyp detection. In *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, volume 2, pages 38 –42, May 2008.

[10] Baopu Li and M. Q.-H Meng. Computer-aided detection of bleeding regions for capsule endoscopy images. *Biomedical Engineering, IEEE Transactions on*, 56(4):1032–1039, 2009. ID: 1.

[11] Baopu Li and Max Q. H Meng. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. *Computers in biology and medicine*, 39(2):141–147, 2 2009.

[12] Baopu Li and Max Q.-H Meng. Texture analysis for ulcer detection in capsule endoscopy images. *Image and Vision Computing*, 27(9):1336–1342, 8 2009.

[13] J.P.S. Cunha, M. Coimbra, P. Campos, and J.M. Soares. Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *Medical Imaging, IEEE Transactions on*, 27:19 –27, 2008.

[14] Md Khayrul Bashar, Kensaku Mori, Yasuhito Suenaga, Takayuki Kitasaka, and Yoshito Mekada. Detecting informative frames from wireless capsule endoscopic video using color and texture features. In *MICCAI '08: Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II*, pages 603–610, Berlin, Heidelberg, 2008. Springer-Verlag.

[15] R. Kumar, P. Rajan, S. Bejakovic, S. Seshamani, G. Mullin, T. Dassopoulos, and G. Hager. Learning disease severity for capsule endoscopy images. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, pages 1314 –1317, 282009-july1 2009.

[16] S. Seshamani, P. Rajan, R. Kumar, H. Girgis, G. Mullin, T. Dassopoulos, and G.D. Hager. A boosted registration framework for lesion matching. In *Medical Image Computing and Computer Assisted Intervention(MICCAI)*, volume 5761, pages 582–589, 2009.

[17] R Kumar, S Seshamani, G Mullin, G Hager, and T Dassopoulos. Assessment of crohn's disease lesions in wireless capsule endoscopy images. *IEEE Transactions on Biomedical Imaging*, 2011 (in revision).

[18] S Seshamani, R Kumar, P Rajan, G Mullin, T Dassopoulos, and G Hager. A meta method for image matching. *IEEE Transactions on Medical Imaging*, -:accepted, 2011.

[19] Hwang S., Oh J., Cox J., Tang S.J., and Tibbals H.F. Blood detection in wireless capsule endoscopy using expectation maximization clustering. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series; Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 6144, pages 577–587, 2006.

[20] J Lee, JH Oh, SK Shah, X Yuan, and SJ Tang. Automatic classification of digestive organs in wireless capsule endoscopy videos. In *SAC '07: Proceedings of the 2007 ACM symposium on Applied computing*, pages 1041–1045, New York, NY, USA, 2007. ACM.

[21] BS Manjunath, JR Ohm, VV Vasudevan, and A Yamada. Color and texture descriptors. *IEEE Transactions on circuits and systems for video technology*, 11(6):703–715, 2001.

[22] S Seshamani, R Kumar, T Dassopoulos, G Mullin, and G Hager. Augmenting capsule endoscopy diagnosis: A similarity learning approach. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 454–462, 2010.

[23] Q Zhao, T Dassopoulos, G Mullin, M QH Meng, and R Kumar. A decision fusion strategy for polyp detection in capsule endoscopy. In *IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2011 (submitted).

[24] Leo Breiman. Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6):pp. 2350–2383, 1996.

[25] J.S. Boreczky and L.D. Wilcox. A hidden markov model framework for video segmentation using audio and image features. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 6, pages 3741 –3744 vol.6, May 1998.

[26] J.S. Boreczky and L.D. Wilcox. A hidden markov model framework for video segmentation using audio and image features. In *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 6, pages 3741 –3744 vol.6, May 1998.

[27] Jia Li, A. Najmi, and R.M. Gray. Image classification by a two dimensional hidden markov model. In *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, volume 6, pages 3313 –3316 vol.6, March 1999.

[28] L. Satish and B.I. Gururaj. Use of hidden markov models for partial discharge pattern classification. *Electrical Insulation, IEEE Transactions on*, 28(2):172 –182, April 1993.

[29] Lexing Xie, Peng Xu, Shih-Fu Chang, Ajay Divakaran, and Huifang Sun. Structure analysis of soccer video with domain knowledge and hidden markov models. *Pattern Recogn. Lett.*, 25:767–775, May 2004.

[30] M. Mackiewicz, J. Berens, and M. Fisher. Wireless capsule endoscopy color video segmentation. *Medical Imaging, IEEE Transactions on*, 27(12):1769 –1781, 2008.

[31] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, 13(2):260 – 269, Apr 1967.