

Automatic Region-of-Interest Segmentation and Pathology Detection in Magnetically Guided Capsule Endoscopy

Philip W. Mewes^{1,2}, Dominik Neumann¹, Oleg Licegevic¹, Johannes Simon¹, Aleksandar Lj. Juloski¹, and Elli Angelopoulou²

¹ Siemens AG, Healthcare Sector, Erlangen, Germany

² Pattern Recognition Lab, University of Erlangen-Nürnberg, Erlangen, Germany
philip.mewes@informatik.uni-erlangen.de

Abstract. Magnetically-guided capsule endoscopy (MGCE) was introduced in 2010 as a procedure where a capsule in the stomach is navigated via an external magnetic field. The quality of the examination depends on the operator’s ability to detect aspects of interest in real time. We present a novel two step computer-assisted diagnostic-procedure (CADP) algorithm for indicating gastritis and gastrointestinal bleedings in the stomach during the examination. First, we identify and exclude subregions of bubbles which can interfere with further processing. Then we address the challenge of lesion localization in an environment with changing contrast and lighting conditions. After a contrast-normalized filtering, feature extraction is performed. The proposed algorithm was tested on 300 images of different patients with uniformly distributed occurrences of the target pathologies. We correctly segmented 84.72% of bubble areas. A mean detection rate of 86% for the target pathologies was achieved during a 5-fold leave-one-out cross-validation.

1 Introduction

Background and Purpose of This Work. Endoscopy of the upper gastrointestinal (GI) tract with flexible endoscopes is a standard clinical procedure. The main disadvantages of this procedure are high invasiveness and patient discomfort. Wireless capsule endoscopy (WCE) was introduced in 2001 and is mainly used in the duodenum. The stomach, in comparison, has large surface and volume and can not be reliably examined with an uncontrolled capsule. Endoscopic capsules that can be steered from the outside by means of magnets have been reported in [7,12,11]. In this paper we use human data from the clinical study of [11] that stems from 29 volunteers and 24 patients (Fig.1c and 1d). A single dataset from one patient contains on average 3600 images. For the MGCE procedure, the patient’s stomach is filled with water and the capsule is navigated from the outside using an external magnetic field. During the examination the operator can control the motion of the capsule so as to obtain a sufficient number of stomach-surface images with diagnostic value. The quality of the examination

depends on the skill of the operator and his ability to detect aspects of interest in real time. We present a novel two-step computer-assisted diagnostic-procedure (CADP) algorithm for detecting, during the examination, two distinct stomach pathologies: gastritis and haematin. Gastritis is one of the main causes of stomach cancer. Chronic gastritis usually appears as a reddish and blotched turgor, while acute gastritis often appears as a small circle-shaped bleeding (Fig.2b). Haematin on the other hand, accrues from blood coming in contact with stomach acid. It is a sign of gastrointestinal bleeding and is an indicator of many significant diseases. It appears as a uniform brownish shape close to the stomach wall (Fig.2a). The proposed algorithm is divided into two steps: 1) a region-of-interest (ROI) segmentation to separate medically relevant sections of the image from parts containing bubbles; 2) a contrast-normalized filtering to identify and localize possible lesions of pathologies. For this 2nd task we develop a feature vector, which is used for classifying pathologies in a machine learning approach.

State of the Art. In a typical WCE examination a large number of frames is medically irrelevant, as they either do not show pathologies or contain mainly intestinal juices, bubbles or debris. To assist the physician in reviewing up to ten hours of video material, software for computer-aided diagnosis (CAD) has been developed. In [2,14,13] different descriptors for the task of blood and ulcera detection and topographic segmentation of the GI tract are investigated. Topographic segmentation is addressed in [1] and [5] for the purpose of a more efficient and faster review. In [4] and [15] the issue of eliminating redundant frames, as well as those with intestinal juices, is addressed. In [6] a set of color and texture based features for the detection of intestinal bleedings is presented. These methods are not directly applicable to MGCE. The duodenum, when compared to the stomach, exhibits different pathologies and imaging conditions, such as texture and distance to objects of interest. In [8] a method for the automatic detection of gastritis aspects in MGCE was presented. In this paper we used the method from [8] as a starting point and developed a ROI pre-segmentation for bubbles and an improved segmentation method for different pathologies (gastritis and haematin). Compared to the aforementioned existing CAD algorithms which are only used for review, our method indicates the pathologies during the examination itself. The operator can, for instance, navigate to a suspicious region for further closer inspection. Even if 100% accuracy is not achieved, our algorithm may still point out lesions which would otherwise be missed.

2 Automatic Pathology Detection in MGCE

2.1 Region of Interest Segmentation

Endoscopic images obtained via MGCE may contain bubbles and mucus (Fig.1a). Such a region within an image usually contains no medically relevant information. One should, thus, segment such regions to exclude them from further processing. In [15] and [10] two methods for automatic detection of intestinal juices are presented that exclude entire frames containing intestinal juices and bubbles for the review process. The location of bubbles in the stomach is relatively stable

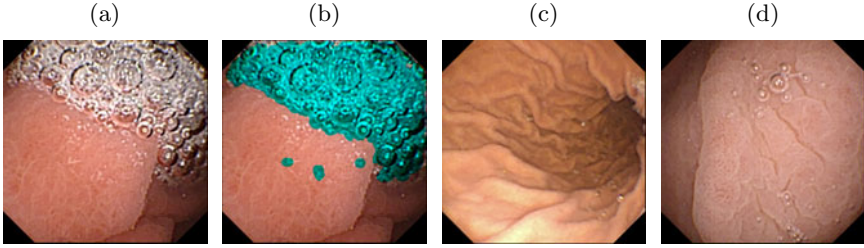


Fig. 1. (a) sample image with bubbles. (b) segmented bubble area (c) lesser curvature and body (d) mucosa close-up B.

and localized. Rejecting an entire frame with bubbles could therefore eliminate images of whole stomach regions together with possible pathologies.

The segmentation of bubble regions is not as straightforward as it may initially appear. The edges of most bubbles appear bright, but the translucent part is dominated by the color of the underlying tissue (Fig. 1a). Therefore, an intensity based approach is not effective. One has to combine geometry and color cues in order to detect the entire bubbles region. To that end, we investigated a large variety of robust feature descriptors initialized with different key point detectors. We randomly chose as our training data 100 healthy and 100 diseased images from our dataset containing different amounts and spatial distributions of bubbles. Regions with bubbles were hand-labeled. We denote one pixel within an image as a tuple (x, y, v) , where x and y are the pixel coordinates, and v is a pixel value. An image \mathcal{I} (432×432 pixels) is defined as a set of pixels, $\mathcal{I} = \{(x_1, y_1, v_1), \dots, (x_N, y_N, v_N)\}$ where N is the total number of pixels in the image. We define $\mathcal{O}_1 \subseteq \mathcal{I}$ as the set of all pixels in a bubble area and $\mathcal{O}_2 \subseteq \mathcal{I}$ as the area without bubbles. In addition $\mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset$, and $\mathcal{O}_1 \cup \mathcal{O}_2 = \mathcal{I}$.

We considered 5 descriptors (SIFT, steerable filters, GLOH, SPIN, Cross Correlation), each of which can be initialized with one of 5 different key point detectors (Hessian-Laplacian, Harris-Laplacian, Harris-Affine, Hessian-affine, Harris)¹. For each combination of descriptor type and key point detector we obtained n descriptors (feature vectors) $D_k, k = 1, \dots, n$, which were calculated on the corresponding image patches $P_k \subset \mathcal{I}, k = 1, \dots, n$ around the detected key points. We used the labeled data to train (using Adaboost) a 2-class classifier that could distinguish between descriptors corresponding to \mathcal{O}_1 and \mathcal{O}_2 . We evaluated each descriptor / key-point-detector combination for bubble detection using a 5-fold leave-one-out cross-validation (LOO CV).

Training phase: An image patch P_k is characterized by its center point $c = (x, y)$ and an ellipse centered at c which is defined by one or two radii $r_{1,2}$ respectively. For the supervised training input, a label

¹ Affine covariant features [Online]. Available: <http://www.robots.ox.ac.uk/~vgg/research/affine/>

$$l(P_k) = \begin{cases} 0 & \text{if } P_k \subset \mathcal{O}_1 \\ 1 & \text{if } P_k \subset \mathcal{O}_2 \end{cases} \quad (1)$$

was assigned. The remaining P_k s (i.e. image patches which do not completely belong to either of the classes) were discarded from training. For training the feature vector D_k was extended by 10 features based on HSV and RGB histograms. More specifically, the mean, variance, skew, kurtosis and entropy were computed for the S and R channel histograms for each patch P_k . These features enhance the classifier’s ability to distinguish structures with geometric similarities to bubbles (such as a round shape) which are, however, part of the stomach mucosa or a target pathology.

Testing: After the training was completed, the new labels $l(P_k)$ were computed based on the obtained classifier. We define two measures for the quality of segmentation:

$$E(\mathcal{O}_i) = \frac{\text{card}\left(\left(\bigcup_{\{P_k|l(P_k)=1\}} P_k\right) \cap \mathcal{O}_i\right)}{\text{card}(\mathcal{O}_i)} \quad (2)$$

with $i = \{1, 2\}$ ² $E(\mathcal{O}_1)$ measures the proportion of the ground-truth bubble area covered by image patches classified as $l = 1$, while $1 - E(\mathcal{O}_2)$ measures the area without bubbles, that is wrongly covered by image patches classified as $l = 1$. We found that the combination of a Hessian-affine key point detector and a steerable filter descriptor yields the best results (see Table 1 (first row)). We refer to the union $\{\cup(P_k|l(P_k) = 0)\}$ as the binary mask \mathcal{I}_b and to bubble free area as $\mathcal{I}_{bn} = \mathcal{I} \setminus \mathcal{I}_b$. Note that, the use of an elliptic shape for the image patches P_k yields small areas between the image patches which are not classified as bubbles area. To overcome this problem we introduced a circular morphologic dilation element with a radius r_s . The circle radius adapts to the size of the surrounding image patches to ensure gap closing between large image patches and prevent the dilation of non-bubbles areas. A neighborhood around the dilation element is defined as a window \mathcal{W} of 30×30 pixels around the center of the structuring element. The radius r_s at a position x, y within \mathcal{I}_b is computed from the average radii of all $P_k|P_k \subseteq \mathcal{W}$.

The improvement obtained with the dilation can be seen in Table 1. The quantitative measurements were obtained by equation 2. After the segmentation we checked if some pathologies were wrongly classified as bubbles and confirmed that this was not the case (Visual results in Fig. 1b).

2.2 Contrast Normalization and Region Localization

For the following steps only \mathcal{I}_{bn} (i.e. areas without bubbles) are considered. The pathological lesions exhibit low intensity while the surrounding tissue has high intensity values. In order to detect such lesions a LoG-edge detection, as proposed by [8], is performed. Thereby, \mathcal{I}_{bn} is convolved with a Laplacian-of-Gaussian

² $\text{card}(\cdot)$ denotes the cardinality, i.e. the number of pixels in the image subset.

Table 1. Results of Bubble Segmentation

Method	$E(O_1)$	$1 - E(O_2)$
Direct use of descriptor	81.53%	4.24%
With adaptive morphological operation	84.72%	5.74%

kernel \mathcal{K} . The resulting image \mathcal{I}_c is subsequently searched for prominent edges whose magnitude is above a certain threshold. Because images suffer from varying contrast and lighting conditions, a contrast normalized variable threshold is introduced. To that end, \mathcal{I}_c is converted into a probability mass function $f(b_i)$ with N_{b_i} intervals. The contrast normalized threshold t_c is then computed as follows. We first define the index t_{pos} of an interval b_i

$$t_{pos} = \sum_{i=N_{b_i}}^1 I \left(\left(\sum_{j=N_{b_i}}^i f(b_j) \right) < \left(\frac{\sum f(b_i)}{4} \right) \right) \quad (3)$$

where I is an indicator function which is equal to 1 if its argument is true and zero otherwise. t_c is then computed as $t_c = \max(b_{t_{pos}})$.

A binary edge image $\mathcal{I}(t_c)$ is then computed using the variable threshold t_c . $\mathcal{I}(t_c)$ is 1 for $\mathcal{I}_c \geq t_c$ and 0 otherwise. Given such a binary image our goal is to merge connected pixels into areas representing possible locations of pathologies. A morphologic closing operator with a disc-shaped structuring element of a radius of 5 pixels is applied to $\mathcal{I}(t_c)$. Subsequently, using 4-connectivity, pixels are grouped into structures that we refer to as S_j with $j = 1, \dots, m$ where m is the total number of structures per image. On average, images contain 45 S_j 's with a mean size of 20×14 pixels. Ultimately, we want to classify each structure S_j into one of the three classes: the *gastritis class* $\mathcal{C}1$, the *haematin class* $\mathcal{C}2$ and the *negative class* $\mathcal{C}3$ without any abnormalities. For training purposes, all S_j are superimposed on to the original RGB-image. Visual inspection shows that all possible lesions of $\mathcal{C}1$ and $\mathcal{C}2$ are detected by the above described region localization method. All computed S_j are therefore directly hand-labeled by an expert so that a label L is assigned to each structure S_j

$$L(S_j) = \begin{cases} 0 & \text{if } S_j \in \mathcal{C}1 \text{ (gastritis class)} \\ 1 & \text{if } S_j \in \mathcal{C}2 \text{ (haematin class)} \\ 2 & \text{if } S_j \in \mathcal{C}3 \text{ (negative class)} \end{cases} \quad (4)$$

2.3 Feature Extraction

The following sets of features are extracted for each structure S_j (Fig. 2c).

Geometric features (GF). Geometric features describe the specific shape of pathologies. We extract the maximal vertical and horizontal dimensions of S_j and

Table 2. Number of features computed for each S_j

Feature Group	Feature Extractor	No. of Features	Composition of Feature Vector
GF	Aspect ratio of S_j	1	width/height
	dimensions of S_j	2	width,height
	bounding box fill factor	1	%
	Hu moments	7	invariant moments
TF	lbp H	5	local binary pattern on three channels of HSV colorspace
	lbp S	5	
	lbp V	5	
HF	RGB	15	Mean, variance, skew kurtosis and entropy of each channel Histogram
	HSV	15	
	CIE	15	
Total no. of features		71	

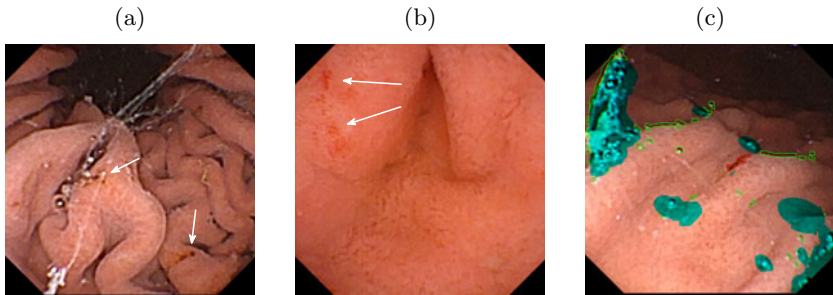


Fig. 2. Sample images of target pathologies. (a): haematin. (b): gastritis lesions. (c): Computed structure S_j (green border). S_j with red border contains a gastritis lesion. Segmented bubble area in turquoise.

of its minimal bounding box and their corresponding aspect ratio. Furthermore the ratio between the area covered by S_j and the area of the minimal bounding box is computed. We refer to this feature as *fill factor*. For further geometric analysis we also extract the Hu moments [3] for each S_j .

Texture features (TF). To further investigate textures within each S_j local binary patterns (lbp) [9] are computed for each channel of the HSV colorspace. The resulting lbp histogram is computed for each S_j .

Histogram features (HF). Color based features aim to distinguish between color characteristics of structures containing gastritis and structures of class $\mathcal{C}3$. Thus, we compute the mean, variance, skew, kurtosis and entropy of each histogram-channel of the RGB, HSV and CIE colorspaces. The entire collection of features is summarized in Table 2.

Table 3. Results of Pathology Detection using a SVM classifier (in %). Results of AdaBoost in parentheses.

Actual \ Predicted	Healthy Image (C1)	Gastritis Image (C2)	Haematin Image (C3)	Gastritis and Haematin C2/C2
Healthy Image C1	92 (95)	6(5)	2(0)	0(0)
Gastritis C2	8(11)	72(55)	6(12)	14(22)
Haematin C3	7(11)	7(12)	76(70)	10(7)

3 Experiments and Results

To evaluate the automatic detection of pathologies we used 300 images stemming from 44 patients. The images were analyzed by a medical expert and each detected structure S_j was manually labeled. The dataset of 300 images consisted of 100 images for each of the two target pathologies and a set of 100 healthy images. For the supervised learning approach we compared results of the Adaboost and SVM classifiers implemented in a 5-fold LOO CV. Most images contained more than a single structure S_j with or without a pathology. An image was attributed to one of the pathology classes if at least one structure S_j was classified as pathologic. An image was counted as healthy if none of the structures S_j was classified as pathologic. The best classification results were obtained using a SVM classifier³. A confusion matrix was obtained for each image test-set of the 5-fold LOO CV. The average confusion matrix of all 5 runs can be seen in Table 3. Out of 100 images with gastritis, 86% were correctly classified (Table 3). From the 100 images containing haematin, 86% were correctly classified. However, we also detected an average of 12% of pathological images, which exhibited both, haematin and gastritis. Finally out of the 100 healthy images 92% were correctly classified.

4 Discussion and Conclusion

Our experiments have shown that our computer-assisted diagnostic-procedure algorithm can be used for indicating gastritis and gastrointestinal bleedings in MGCE. The presented algorithm includes a preprocessing step that discards areas with bubbles. This step is crucial for all following image processing steps in the presented method and may have implications for the development of future applications on this imaging modality. Pre-segmentation performs accurately in detecting areas with bubbles without hiding pathologies or large amounts of non-bubble tissue areas. Based on the results from the above pre-processing step, a method was presented that is able to automatically detect two kinds of pathologies on MGCE images. We achieved sensitivity and specificity results well over 80% for healthy and diseased images. The moderate sensitivity between the two target pathologies is due to the similarity between gastritis and haematin

³ LIBSVM [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

aspects in terms of color and texture. Still, the algorithm performs well, especially within the context of real-time warnings for certain gastritis aspects.

References

1. Berens, J., Mackiewicz, M., Bell, D.: Stomach, intestine, and colon tissue discriminators for wireless capsule endoscopy images. In: Fitzpatrick, J., Reinhardt, J. (eds.) *Proceedings of SPIE*, vol. 5747, p. 283 (2005)
2. Cunha, J., Coimbra, M., Campos, P., Soares, J.: Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE T. Med. Imaging* 27(1), 19–27 (2007)
3. Hu, M.: Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory* 8, 179–187 (1962)
4. Karargyris, A., et al.: A video-frame based registration using segmentation and graph connectivity for Wireless Capsule Endoscopy. In: *Life Science Systems and Applications Workshop LiSSA 2009, IEEE/NIH*, pp. 74–79. IEEE, Los Alamitos (2009)
5. Mackiewicz, M., Berens, J., Fisher, M.: Wireless capsule endoscopy color video segmentation. *IEEE T. Med. Imaging* 27(12), 1769–1781 (2008)
6. Mackiewicz, M., Fisher, M., Jamieson, C.: Bleeding detection in wireless capsule endoscopy using adaptive colour histogram model and support vector classification. In: Joseph, M., Josien, P. (eds.) *Proceedings of SPIE*, vol. 6914, p. 69140R (2008)
7. Menciassi, A., Valdastrì, P., Quaglia, C., Buselli, E., Dario, P.: Wireless steering mechanism with magnetic actuation for an endoscopic capsule. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009*, pp. 1204–1207. IEEE, Los Alamitos (2009)
8. Mewes, P., Neumann, D., Juloski, A., Angelopoulou, E., Hornegger, J.: On-the-fly detection of images with gastritis aspects in magnetically-guided capsule endoscopy. In: Ronald, M., van Ginneken, B. (eds.) *Medical Imaging - Computer-Aided Diagnosis. Proceedings of the SPIE*, vol. 7963, p. 79631I3. SPIE, San Jose (2011)
9. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* 29(1), 51–59 (1996)
10. Reingruber, H.: Intestinal content detection in capsule endoscopy using robust features @ONLINE (2009), <http://www.recercat.net/handle/2072/43221>
11. Rey, J., Ogata, H., Hosoe, N., Ohtsuka, K., Ogata, N., Ikeda, K., Aihara, H., Pangtay, I., Hibi, T., Kudo, S., Tajiri, H.: Feasibility of stomach exploration with a guided capsule endoscope. *Endoscopy* 42(7), 541–545 (2010)
12. Swain, P., Toor, A., Volke, F., Keller, J., Gerber, J., Rabinovitz, E., Rothstein, R.: Remote magnetic manipulation of a wireless capsule endoscope in the esophagus and stomach of humans. *Gastrointest Endosc* (2010)
13. Szczypinski, P., Klepaczko, A.: Selecting texture discriminative descriptors of capsule endoscopy images. In: Zinterhof, P. (ed.) *Proceedings of ISPA 2009*, pp. 701–706. IEEE, Los Alamitos (2009)
14. Szczypinski, P., Sriram, P., Sriram, R., Reddy, D.: Model of deformable rings for aiding the wireless capsule endoscopy video interpretation and reporting. *Comp. Imag. Vis.*, 167–172 (2006)
15. Vilarino, F., Spyridonos, P., Pujol, O., Vitria, J., Radeva, P., De Iorio, F.: Automatic detection of intestinal juices in wireless capsule video endoscopy. In: Tang, Y., Wang, S., Lorette, G., Yeung, D., Yan, H. (eds.) *ICPR 2006*, vol. 4, pp. 719–722. IEEE, Los Alamitos (2006)