

On-the-fly Detection of Images with Gastritis Aspects in Magnetically-Guided Capsule Endoscopy

P. W. Mewes^{1,2}, D. Neumann², A. Lj. Juloski², E. Angelopoulou¹, J. Hornegger¹

¹Pattern Recognition Lab, University of Erlangen-Nuremberg, Erlangen, Germany

²Siemens Healthcare Sector, Erlangen, Germany

ABSTRACT

Capsule Endoscopy (CE) was introduced in 2000 and has since become an established diagnostic procedure for the small bowel, colon and esophagus. For the CE examination the patient swallows the capsule, which then travels through the gastrointestinal tract under the influence of the peristaltic movements. CE is not indicated for stomach examination, as the capsule movements can not be controlled from the outside and the entire surface of the stomach can not be reliably covered. Magnetically-guided capsule endoscopy (MGCE) was introduced in 2010. For the MGCE procedure the 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 in order 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 computer-assisted diagnostic-procedure (CADP) algorithm for indicating gastritis pathologies in the stomach during the examination. Our algorithm is based on pre-processing methods and feature vectors that are suitably chosen for the challenges of the MGCE imaging (suspended particles, bubbles, lighting). An image is classified using an ada-boost trained classifier. For the classifier training, a number of possible features were investigated. Statistical evaluation was conducted to identify relevant features with discriminative potential. The proposed algorithm was tested on 12 video sequences stemming from 6 volunteers. A mean detection rate of 91.17% was achieved during leave-one out cross-validation.

Keywords: capsule endoscopes, computer-assisted diagnosis, magnetically guided capsule endoscopy, gastritis

1. INTRODUCTION

Cancer in the gastrointestinal tract (GI) has the first highest mortality rate in Asia and the second highest mortality rate of all kinds of cancers in Germany and in the US.^{1,2} One of its main causes is acute gastritis, which has different acute and chronic causes. Certain anti-inflammatory drugs and abusive alcohol consumption are the main acute causes, while an infection with bacteria, primarily helicobacter pylori, is the most important cause of chronic gastritis and represents 60-70% of all gastritis cases.³ Therefore, in diagnostic and screening procedures the detection of gastritis deserves special attention. Chronic gastritis usually appears as reddish and blotched turgor, while acute gastritis usually appears as a point-shaped bleeding. Classical screening and diagnosis procedures are based on esophagogastroduodenoscopy, which involves inserting an endoscope through the mouth into the duodenum, stomach and esophagus of a sedated patient.

Capsule Endoscopes (CE, also known as passive capsule endoscopes) were introduced in 2000 by Given Imaging (Given Imaging, Yoqneam, Israel). CE is aimed primarily for small bowel examination but has since also been applied for the examination of the colon, esophagus and duodenum.⁴⁻⁸ CE is not indicated for stomach examination because the capsule can not be controlled and imaging has been random, which made it difficult to reliably cover the entire surface of the stomach. Magnetically guided capsule endoscopes (MGCE) were introduced in 2010 by Siemens Healthcare (Siemens Healthcare Sector, Erlangen, Germany) and Olympus Medical (Olympus Corporation Shinjuku, Tokyo, Japan) making it possible to navigate a capsule endoscope in a water-filled stomach. MGCE enables the diagnosis of gastritis inflammations from capsule images. The first clinical study including 53 volunteers and patients was published in 2010.⁹ MGCE appears to be feasible and sufficiently accurate for gastric examination.⁹

Send correspondence to Philip Mewes: E-mail: philip.mewes@siemens.com

In a typical CE examination a large number of frames that do not show pathologies are captured during the passive capsule passage through the GI tract. To assist the physician in reviewing up to ten hours of video material software for computer-aided diagnosis (CAD) has been developed. CAD approaches for CE as described in¹⁰⁻¹⁴ are not directly applicable or only partially applicable for MGCE. Because we are looking at different pathologies and working in a different environment where the imaging conditions are different. In¹² Coimbra assesses the potential of selected visual MPEG-7 descriptors in CE for the task of blood and ulcera detection. The methodology of Coimbra can not be applied to gastritis as this pathology does not appear in CE. In¹¹ Cuhna focuses on segmentation of the gastrointestinal tract into its four major topographic areas, allowing the automatic estimation of the clinically relevant gastric and intestinal sections and corresponding transit times. The detection method of relevant pathologies is not proposed. In¹⁴ Mackiewicz is combining both methods using color and texture based features. In¹⁰ Szczypinski investigated the reduction of the feature space for the detections of blood and ulcera. With the exception of the work of Cuhna all of the mentioned papers have investigated the reduction of the feature space. We are also considering the reduction of the feature space in section 2 for MGCE related features. In¹³ Vilarino proposed a method for automatic detection of intestinal juices by applying Gabor filters to reduce the amount of relevant video material.

Nevertheless, physicians are not confident with CAD methods as none of this software can provide 100% accuracy.^{15,16} In MGCE, CADP can also be used during the later review. A much more compelling setup, however, is the application of CADP during the examination itself. The operator can immediately use the information provided by CADP during the examination to further enhance the diagnostic value. Even if 100 % accuracy is not achieved, our algorithm may still point out the lesions which would otherwise be missed by the operator. We present an automatic detection algorithm which is adapted to MGCE and verifies on-the-fly the stream of images in real-time.

For the MGCE procedure the patient’s stomach is filled with water. A water-filled stomach may contain debris, such as small bubbles, mucus and suspended particles, that attach to the stomach walls or to the capsule housing or swim in the capsules field of view. In contrast to classic endoscopes the frame rate as well as the resolution of the images are smaller, which makes image processing more challenging. Lighting conditions may change depending on the capsule position and the distance to the stomach wall. The aforementioned effects are rarely disturbing for a human user, but need to be considered in automatic detection.

We present an automatic detection algorithm based on a MGCE-specific feature vector, that verifies on-the-fly the stream of images and is able to warn the physician when possible gastritis lesions appear on the images. In our analysis we used 12 data sets which stem from 6 volunteers and contain 1051 images with and 355 without gastritis aspects. The images originate from different regions of the stomach to ensure that the training data is not biased for a specific region of the stomach.

In the following section we sort possible features starting with a preprocessing step. After presenting different possible features in section 2, their discriminative potential is statistically evaluated. Weak features and correlating features can be eliminated to make computation faster. Subsequently we test our new feature vector on a set of 1406 images. Results are presented in section 4. In section 6 we discuss our results and give an overview of future work.

2. METHOD

Preprocessing. The preprocessing step identifies and locates possible gastritis regions on a contrast-normalized image. The preprocessing is done as follows:

- The original dimensions of the images are denoted as x_{size} and y_{size} . The original RGB image $o_c[x, y]$ is converted to a grayscale image $o_g[x, y]$ for further processing. In the gray-scale image gastritis lesions exhibit low intensity while the surrounding tissue has high intensity.
- The original grayscale image $o_g[x, y]$ is convolved with a Laplacian-of-Gaussian kernel (LoG-kernel). The result of this convolution is a filter response map, that has peaks at possible gastritis aspects and other image regions with rapid intensity changes (e.g. shadows, particles and bubbles). The convolution result

is computed as $c[x, y] = o_g[x, y] * k[x, y]$ where $o_g[x, y]$ is the original gray-scale image and $k[x, y]$ is the convolution kernel (see figure 1). At the edges of gastritis lesion the filter response changes from low to high amplitude. Changes from low to high amplitude can also occur for other reasons as mentioned above. The convolved image $c[x, y]$ is subsequently searched for those amplitude changes, that appear as peaks on $c[x, y]$. We are hence looking for a threshold (denoted t_c) that best separates $c[x, y]$ into relevant and irrelevant amplitude changes.

- To that end we first perform contrast normalization. Such normalization is necessary because images suffer from varying contrast and lighting conditions. A fixed threshold to search for amplitude changes in $c[x, y]$ would fail on some images. A high threshold value would fail on images with low contrast and therefore with smaller peaks (e.g. capsule view on far objects). A low threshold would identify most peaks in a high-contrast image as relevant.

To compute a threshold t_c for contrast normalization, each image $c[x, y]$ is converted into a probability mass function (see equation 2). To do so, we represent $c[x, y]$ as a sequence of observations $[c(0), \dots, c(S)]$ where $S = x_{size} * y_{size}$ is the total number of pixels and the total number of filter responses. We form a set of intervals:

$$B = \{b_1 = [b_{11}, b_{12}); b_2 = [b_{21}, b_{22}); \dots; b_{N_b} = [b_{N_b1}, b_{N_b2}]\} \quad (1)$$

where

$$\begin{aligned} b_{11} &= c[x, y]_{min} \text{ and } b_{N_b2} = c[x, y]_{max} \\ &\text{and} \\ b_{i1} &= b_{11} + (i - 1) \times \frac{b_{N_b2} - b_{11}}{N_{b_i}} \text{ and } b_{i2} = b_{(i+1)1} \end{aligned}$$

where $B_i = [b_{i1}, b_{i2})$ are intervals of equal size. The size of each interval is determined by the range of intensity values in an image and the total number of intervals N_{b_i} which in our case is fixed to 256. The size of the intervals b_i varies from image to image. $c[x, y]_{min}$ and $c[x, y]_{max}$ are derived from the convolution $o_g[x, y] * k[x, y]$. $o_g[x, y]$ is defined between 0...255 while the discretized LoG-kernel $k[x, y]$ is defined as in figure 1. In practice, the minimal and maximal possible values $c[x, y]_{min}$ and $c[x, y]_{max}$ of $c[x, y]$ never occur. Experimental results show that in average $c[x, y]$ lies in an interval of $[-16.000, 8000]$. The average interval size of $b \in B_i$ is computed as $(|c[x, y]_{max}| - |c[x, y]_{min}|) / N_{b_i}$. Both values $c[x, y]_{min}$ and $c[x, y]_{max}$ are computed independently for each image and represent the upper and lower limits of the interval B_i . The probability mass function $f(b_i)$ is therefore defined as

$$f(b_i) = \sum_{l=0}^S (I(c(l) \in b_i)) \quad (2)$$

where I is an indicator function equaling 1 if its argument is true and zero otherwise.

- The contrast normalized threshold t_c we are looking for is then computed as follows

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

where I is again an indicator function and t_{pos} is the index of the interval b_i computed with equation 3. t_c is finally computed as

$$t_c = max(b_{t_{pos}}) \quad (4)$$

$$k[x, y] = \begin{bmatrix} 0 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 0 \\ 1 & 2 & 4 & 5 & 5 & 5 & 4 & 2 & 1 \\ 1 & 4 & 5 & 3 & 0 & 3 & 5 & 4 & 1 \\ 2 & 5 & 3 & -12 & -24 & -12 & 3 & 5 & 2 \\ 2 & 5 & 0 & -24 & -40 & -24 & 0 & 5 & 2 \\ 2 & 5 & 3 & -12 & -24 & -12 & 3 & 5 & 2 \\ 1 & 4 & 5 & 3 & 0 & 3 & 5 & 4 & 1 \\ 1 & 2 & 4 & 5 & 5 & 5 & 4 & 2 & 1 \\ 0 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 0 \end{bmatrix} \quad (5)$$

Figure 1. A discretized 9×9 LoG kernel (with Gaussian of $\sigma = 1.4$)

A binary filter response map $c_t[x, y]$ is computed using the variable threshold t_c . $c_t[x, y]$ is 1 for $c[x, y] > t_c$ and 0 for $c[x, y] < t_c$.

Subsequently, a spatial distribution of filter responses for further localization of relevant spots is defined by overlaying the image with an uniform fixed grid over the binary image $c_t[x, y]$, separating the image into spatial bins $c_t bin_i[x, y]$ with $x = 1..15$ and $y = 1..15$. The number of bins within each image is denoted as N_{bin} and can be derived from the bin size and the size of the image $c[x, y]$.

Regions of fast intensity changes as they appear in a gastritis lesion usually lead to an accumulation of filter responses with high changes in amplitudes. High amplitude changes also occur for high frequency noise components, but only in an isolated fashion or in small numbers. A non-maximum suppression is conducted by counting the number of threshold filter responses $Nb(c_t bin_i)$ for all x, y within each $c_t bin_i[x, y]$ for each bin. For further processing only those satisfying the heuristically determined threshold of $Nb_{bin} > 10$ are considered. Other bins are discarded for the remainder of the preprocessing. The result from the preprocessing steps is a number of spatial bins in the static grid that fulfill the aforementioned conditions. We define two classes: the *positive class* $\mathcal{C}1$ with a gastritis aspect or to the *negative class* $\mathcal{C}2$ without any abnormalities. For training purposes, all $c_t bin_i[x, y]$ which are chosen for further processing are hand-labeled by an expert so that

$$l_i = \begin{cases} 1 & \text{for all } c_t bin_i[x, y] \in \mathcal{C}1 \\ 0 & \text{for all } c_t bin_i[x, y] \in \mathcal{C}2 \end{cases} \quad \text{for } i = 1 \dots N_{bin} \quad (6)$$

and $\vec{l} = (l_1, l_2, \dots, l_{N_{bin}})$ is a vector containing labels for each bin.

Features. For the sake of simplicity we refer to $c_t bin_i[x, y]$ as $c_t bin_i$. When $c_t bin_i$ is validated as a possible location of gastritis through the above described preprocessing steps, the following features are extracted from the corresponding region in RGB colorspace. We refer to this 3-channel spatial bin as $c_{c_k} bin_i$ with $k = \{R, G, B\}$.

Color Features and Color Histogram Features (CHF). Color based features aim to distinguish between color characteristics of bins actually containing gastritis lesions and bins of class $\mathcal{C}2$. Color based features consist of mean and maximum values of each channels of the RGB color spaces within a spatial bin. Furthermore a set of intervals is formed as described in equation 1. Each bin $c_{c_k} bin_i$ for $k = 1$ is subsequently represented as a probability mass function (see equation 2). The operation is repeated for three different channels after converting $c_{c_k} bin_i$ in HSV and Lab color spaces. We refer to each probability mass function (pmf) and color channel as

- $f(R_i)$ for the R-channel of the RGB color space
- $f(V_i)$ the 3rd channel of the HSV(hue, saturation and brightness) color space
- $f(L_i)$ the L-channel of the Lab colorspace

For each function the number of intervals ($N_{R_i}, N_{V_i}, N_{L_i}$) is determined independently. The number of intervals is experimentally determined from a representative number of bins from all datasets. Pixel values within one bin may not cover the complete range of values of a particular channel. The goal is to cover areas of large variety

within the range of values with a high density of consecutive intervals. Therefore the number of intervals within $f(R_i, V_i, L_i)$ was iteratively increased in areas of large variety. If intervals within $f(R_i, V_i, L_i)$ become zero the process was stopped. The number of intervals is specified in table 1. Each interval represents a single feature.

Color Structure Feature (CSF). This feature set tries to distinguish between the quantitative and spatial distributions of color characteristics within a spatial bin. The aim of this feature is to describe the distribution and occurrence of color characteristic within one image region. Possible gastritis aspects often appear as round spots with irregular borders. Objects identified as possible gastritis but labeled as a negative $\mathcal{C}2$ class object often exhibit different color structures. To define a feature set based on color structure each spatial bin from the static grid $c_{c_k} bin_i$ is again divided into a small sub grid of 5×5 sub-bins. Within each sub-bin the presence of a particular color characteristic is detected. The number of appearances within a bin $c_{c_k} bin_i$ represents a feature. This step is done for five color channels: R, G, B, L (of LaB) and V (of HSV).

Texture Features (TF). To further investigate textures with each $c_{c_k} bin_i$, a Difference-of-Boxes filter (DoB) is applied on each grayscale converted $c_{c_k} bin_i$ spatial bin at different scales and orientations. Filter responses have high negative and positive amplitudes for spatial changes of intensity from low-to-high and high-to-low respectively. Three groups of observations are obtained from this feature: Statistical observations (mean, minimum, maximum) within each bin, spatial distribution within each bin similar to CSF, and intervals of a probability mass function for each bin such as CHF.

Filter Features (FF). This feature group includes two features: The number of positive filter responses $Nb_{cbin_{t_i}}$ within a bin $cbin_{t_i}[x, y]$ chosen as a possible location of a gastritis through the above described preprocessing steps. As a second feature the discretized LoG-kernel is replaced with a boxplot filter to compute filter responses. This Laplacian of Gaussian approximation is computationally less expensive. The two filters are compared in¹⁷ and it was concluded that the boxplot representation has a negligible loss in accuracy.

An overview of all features and their number is summarized in table 1 on page 5.

Feature group	Feature	No. of initial Features	No. of Features after NZR	No. of Features after VDR/MDR
CHF	statistical observations of RGB features	6	6	0
	RGB Histogram observation	8	8	3
	Lab Histogram observation	11	10	6
	HSV Histogram observation	101	98	44
CSF	color distribution and occurrence	5	5	0
TF	statistical observations of DoB filter	6	6	3
	spatial distribution of DoB	4	4	0
	absolute number of occurrences of DoB filter responses	44	44	10
FF	$Nb_{cbin_{t_i}}$	1	1	0
	$Nb_{cbin_{t_i}}$ (boxplot)	1	1	0
\sum of features	-	187	183	66

Table 1. Number of feature before and after feature space reduction

At the end of the feature extraction process a feature matrix F of size $(N_{c1} + N_{c2}) * (n + 1)$ is built, where n is the number of features and $N_{c1, c2}$ the number of spatial bins from class $\mathcal{C}1$ and $\mathcal{C}2$. F contains row vectors $\vec{v}(x_{(i,j)})$ for each bin with $j = 1 \dots (N_{c1} + N_{c2})$. $x_{i,j}$ denotes a single feature of a particular bin for $i = 1 \dots n$. We refer to $(N_{c1} + N_{c2})$ as N . To determine the class affiliation of each $\vec{v}(x_{(i,j)})$ the label vector \vec{l} (see equation

6) is added to the feature matrix. The feature matrix is therefore composed as follows:

$$F = \begin{pmatrix} l_1 & (x_{1,1}) & \cdots & (x_{1,n}) \\ l_2 & (x_{2,1}) & \cdots & (x_{2,n}) \\ \vdots & \cdots & \ddots & \vdots \\ l_N & (x_{N,1}) & \cdots & (x_{N,n}) \end{pmatrix} = \begin{pmatrix} l_1 & \vec{v}(x_{(i,1)}) \\ l_2 & \vec{v}(x_{(i,2)}) \\ \vdots & \vdots \\ l_N & \vec{v}(x_{(i,N)}) \end{pmatrix} \quad (7)$$

For the sake of simplicity we refer to a spatial bin as an observation for the rest of this paper. Within F a feature is denoted as a column vector $\vec{u}(x_{(i,j)})$ which is split by the class sizes N_{C1} and N_{C2} into two column vectors denoted as $\vec{x}_{(N_{C1},j)}$ and $\vec{x}_{(N_{C2},j)}$ for all $j = 1 \dots N$. $\vec{u}(x_{(i,j)})$ contains one particular feature for all observations.

Feature space reduction. Feature space reduction is performed in order to reduce the computational cost by decreasing the amount of data and to identify relevant features with strong discriminative potential. Dimensionality reduction is conducted in three steps. If all elements of a feature vector $\vec{u}(x_{(i,j)})$ are zero for both of the classes, then the feature is ignored in further training and testing. The criterion is denoted as NonZeroRelevance (NZR).

Four statistical measurements for the relevance of the feature vectors are introduced:

- the $C1$ -class related mean $\mu_{\vec{x}_{N_{C1},j}}$
- the $C2$ -class related mean $\mu_{\vec{x}_{N_{C2},j}}$
- the variance for both classes $\sigma_{x_{(N,j)}}^2$
- the $C1$ -class related variance $\sigma_{x_{(N_{C1},j)}}^2$

If the difference of $\sigma_{x_{(N,j)}}^2$ and $\sigma_{x_{(N_{C1},j)}}^2$ is not significant, or $\sigma_{x_{(N_{C1},j)}}^2$ is higher than the variance $\sigma_{x_{(N,j)}}^2$, then this feature is not relevant. We denote this relevance as VarianceDifferenceRelevance (VDR) where the function

$$\text{VDR}(\vec{u}(x_{(i,j)})) = \sigma_{\vec{x}_{N_{C1},j}}^2 - \sigma_{\vec{x}_{(N,j)}}^2 < c_{\sigma^2} \quad (8)$$

assigns a boolean value to each $\vec{x}_{(N,j)}$ depending on the threshold c_{σ^2} .

The mean is a measurement for the average location of each class. If the difference between the mean of $C1$ -class related observations within the feature vector is significantly higher than the mean of all class-specific elements in the feature vector it can be assumed that this feature is able to distinguish very well between $C1$ and $C2$. The measurements for this criteria is denoted as MeanDifferenceRelevance (MDR) where

$$\text{MDR}(\vec{u}(x_{(i,j)})) = \left| \frac{\mu_{\vec{x}_{N_{C1},j}} - \mu_{\vec{x}_{(N_{C2},j)}}}{\mu_{\vec{x}_{(N_{C2},j)}}} \right| < c_{\mu} \quad (9)$$

assigns a boolean value to each $\vec{x}_{(N,j)}$ depending on the threshold c_{μ} .

c_{μ} and c_{σ^2} are thresholds defined through an objective criteria for the particular data set and the underlying diagnostic requirements as follows: The diagnostic value for the automated detection of gastritis aspects becomes worse if an observation belonging to $C1$ is wrongly classified and changes the classification result of an entire image. Changing the classification result of an image within a dataset would worsen the overall classification rate ε_{c1} described in chapter 3. Avoiding a poorer classification rate ε_{c1} is therefore our criteria after which the reduction of features is conducted and evaluated.

To that end, we first need to sort all features vectors $\vec{u}(x_{(i,j)})$ regarding their discriminative potential defined

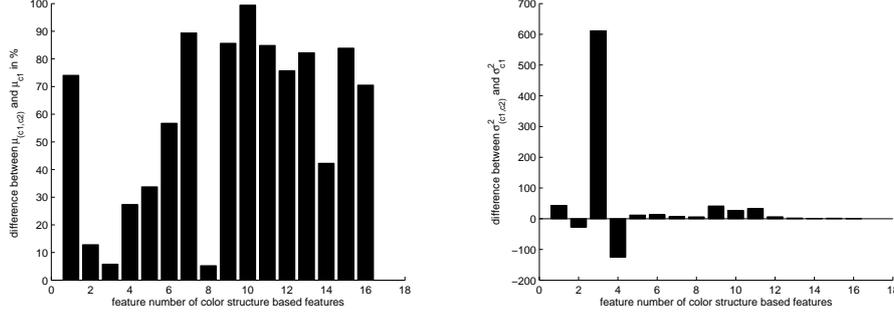


Figure 2. Left: absolute difference between $\mu_{c1,c2}$ and μ_{c1} in % for color-structure based features. Right: absolute difference between $\sigma^2_{c1,c2}$ and σ^2_{c1} for color-structure based features.

through eq.(9) and eq.(8). The value c_μ and c_{σ^2} is iteratively increased. Thus more and more features will fail to fulfill eq.(9) and eq.(8). The sorted indexes j of the features within $F_{i,j}$ are stored in the $2 \times n$ matrix.

$$A = \begin{pmatrix} VDR(j) & MDR(j) \\ \vdots & \vdots \\ VDR(j) & MDR(j) \end{pmatrix} \quad (10)$$

A contains therefore the features indexes sorted after the discriminative potential in increasing order. This process is described in algorithm 1.

Algorithm 1 Pseudo code algorithm for sorting features depending on $VDR(\vec{u}(x_{(i,j)}))$

```

 $c_{\sigma^2} = \max_{j=1 \dots n} (\sigma^2_{(\vec{x}_{N_{c1,j}})} - \sigma^2_{(\vec{x}_{N,j})})$ 
 $i = 1$ 
while  $i \leq n$  do
  if  $((\sigma^2_{(\vec{x}_{N_{c1,j}})} - \sigma^2_{(\vec{x}_{N,j})}) < c_{\sigma^2})$  for all  $j = 1 \dots n$  then
    add  $j$  to  $A_{j,1}$ 
     $i = i + 1$ 
     $c_{\sigma^2} = c_{\sigma^2} + \varepsilon$ 
  else
     $c_{\sigma^2} = c_{\sigma^2} + \varepsilon$ 
  end if
end while

```

The process is repeated for $MDR(\vec{u}(x_{(i,j)}))$. After sorting the features according to their discriminative potential the number of features is iteratively decreased. The number of wrongly classified $\mathcal{C}1$ -observations is monitored at each step and the process is stopped as soon as a $\mathcal{C}1$ -observation that influences the detection rate is wrongly classified (See algorithm 2).

Boosting We are using the real adaptive boosting algorithm (adaboosting) from Open CV with boosted decision trees as weak classifiers.^{18,19} Adaboost is a supervised learning algorithm, and as such it needs hand-labeled ground-truth data. Each spatial bin belonging to the positive class $\mathcal{C}1$ is labeled by hand. All others are considered to belong to the negative class $\mathcal{C}2$. A parameter search is conducted to find the optimal number of weak learners and the optimal depth of decision trees.

Computational cost We ran all code on an Intel Core i7 Q840 @1,87 Ghz. Feature extraction is performed with MATLAB©(2010a, The MathWorks, Natick, MA). Classification is performed with Visual C++ (2010

Algorithm 2 Pseudocode algorithm of feature space reduction

```
 $\varepsilon_{c1,R} = \varepsilon_{c1}$   
 $i = 1$   
while  $\varepsilon_{c1} == \varepsilon_{c1,R}$  do  
  delete column  $F(A(i, 1), j)$  and  $F(A(i, 2), j)$   
  Adaboost  $\rightarrow$  train( $F$ )  
   $\varepsilon_{c1,R} =$  Adaboost  $\rightarrow$  test( $F$ )  
   $i = i+1$ ;  
end while
```

Express, Microsoft Corporation, Redmond, WA).

Considering all initially extracted features the classification of a single bin takes 0.0021ms on average. The average number of bins chosen through the previously described preprocessing steps over all images within the test set is 88, the maximum is 166 and the minimum 9 bins. Classification therefore takes between 0.0189ms and 0.3486ms with an average time per classification of 0.1848ms per image. Feature extraction takes 2.6s per image on average using all 187 features. After feature space reduction the computation takes 1.9s on average and the average time of classification for an image decreases by 28%.

3. EXPERIMENTS

A leave-one-out (LOO) cross-validation was performed considering each of the 12 datasets as a complementary subsets of all $\mathcal{C}1$ and $\mathcal{C}2$ images. Each bin within one image is individually classified. Most images contain more than a single bin with or without gastritis so that different combinations of correctly and incorrectly classified bins can occur within a single image. The different combinations are summarized in table 2. Each of this combinations contribute to one or both of the following performance measurements:

1. $\varepsilon_{\mathcal{C}1}$: The percentage of images belonging to $\mathcal{C}1$ that are correctly classified out of the number of all images belonging to $\mathcal{C}1$.
2. $\varepsilon_{\mathcal{C}2}$: The percentage of images belonging to $\mathcal{C}2$ that are correctly classified out of the number of all images belonging to $\mathcal{C}2$.

$\varepsilon_{\mathcal{C}1}$ measures the sensitivity of the algorithm while $\varepsilon_{\mathcal{C}2}$ is a measurement for the specificity. Both are important to evaluate the performance of our CADP algorithm.

Image Ground Truth	Number of False Positives	Number of False Negatives	Contributes negatively(-) or positively (+) to the performance measurements
$\mathcal{C}1$	0	$< N_{\mathcal{C}1}$	$+\varepsilon_{\mathcal{C}1}$
$\mathcal{C}1$	0	$= N_{\mathcal{C}1}$	$-\varepsilon_{\mathcal{C}1}$
$\mathcal{C}1$	> 0	$< N_{\mathcal{C}1}$	$+\varepsilon_{\mathcal{C}1}$
$\mathcal{C}1$	> 0	$= N_{\mathcal{C}1}$	$-\varepsilon_{\mathcal{C}1}$
$\mathcal{C}1$	> 0	0	$+\varepsilon_{\mathcal{C}1}$
$\mathcal{C}1$	0	0	$+\varepsilon_{\mathcal{C}1}$
$\mathcal{C}2$	-	0	$+\varepsilon_{\mathcal{C}2}$
$\mathcal{C}2$	-	> 0	$-\varepsilon_{\mathcal{C}2}$

Table 2. Occurrences of classification errors within a single image and their contribution to sensitivity and specificity measurements

4. RESULTS

Two benchmarks $\varepsilon_{\mathcal{C}1}$ and $\varepsilon_{\mathcal{C}2}$ as described in section 3 were evaluated on our dataset. The results are summarized in Table 3. The initial feature vector of 187 features could be reduced to 66 features without losing diagnostic

relevance. In summary a set of features is presented that is able to detect a gastric lesion with a mean detection rate (ε_{C1}) of 80.56%. Images without gastric lesion are correctly classified at 91.17% (performance measurement (ε_{C2})).

DataSet No.	Class Occurrence	ε_{C2} (in %)	ε_{C1} (in %)	DataSet No.	Class Occurrence	ε_{C2} (in %)	ε_{C1} (in %)
1	$C1/C2$	93.33	90.625	4	$C1/C2$	67.5	33.33
1a	$C2$	100	-	4a	$C2$	100	-
2	$C1/C2$	86	77	5	$C1/C2$	80.16	76.77
2a	$C2$	90.16	-	5a	$C1/C2$	92	75
3	$C1/C2$	94.66	100	6	$C1/C2$	90.79	88.46
3a	$C2$	100	-	6a	$C1/C2$	100	100

Table 3. Results of LOO validation in % of correct detections for 8 Datasets (Each containing two series of images from the same patient).

5. NEW OR BREAKTHROUGH TO BE PRESENTED

We presented a classification method for on-the-fly detection of images with gastritis aspects in magnetically-guided capsule endoscopy. A first study of suitable features for automatic detection of gastritis in MGCE was conducted. Our method yielded an overall detection rate of 91.17% for images from both classes and 80.56% for images containing gastritis lesions.

6. CONCLUSIONS

Automatic detection of gastritis lesion in MGCE appears to be feasible. We showed that the proposed algorithm performs with a satisfactory sensitivity and specificity on the aforementioned dataset of 6 volunteers. We noticed that the classification, even if effective for the majority of the dataset, has difficulties for some image sets. These datasets contain pathologies captured from a very acute camera angle or under extremely difficult lighting conditions. We showed that five out of seven feature groups are important for achieving good results in gastritis image classification using a LOO-validation. From initially chosen feature groups the number of features could be reduced by 64% in average without any loss of accuracy in the detection of gastritis lesions for the given dataset of 1406 images. If more ground truth data is available, it should be included in the training process to build a classifier with an improved sensitivity and specificity over all available datasets. For further work we aim to extend our algorithm to detect more pathologies that occur in the MGCE. Depending of these pathologies the choice of different features should be considered.

REFERENCES

- [1] Heidelberg, D., "Krebsatlas der Bundesrepublik Deutschland," (2010).
- [2] Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M., "Cancer statistics, 2009," *CA: a cancer journal for clinicians* **59**(4), 225 (2009).
- [3] Böcker, W., [*Pathologie*], Elsevier, Urban&FischerVerlag (2008).
- [4] Iddan, G. and Swain, P., "History and development of capsule endoscopy," *Gastrointestinal endoscopy clinics of North America* **14**(1), 1–9 (2004).
- [5] Iddan, G., Meron, G., Glukhovsky, A., and Swain, P., "Wireless capsule endoscopy," *Nature* **405**, 417 (2000).
- [6] Food and Drug Administration (USA), "November 2004 510(k) clearances no: K042960(special)," (Nov. 2004).
- [7] Heresbach, D. et al., "Diagnostic accuracy of esophageal capsule endoscopy versus conventional upper digestive endoscopy for suspected esophageal squamous cell carcinoma," *Endoscopy* **42**(02), 93–97 (2010).
- [8] Swain, P., Iddan, G., Meron, G., and Glukhovsky, A., "Wireless capsule endoscopy of the small bowel: development, testing, and first human trials," in [*Proceedings of SPIE*], **4158**, 19 (2001).

- [9] Rey, J. F., Ogata, H., Hosoe, N., Ohtsuka, K., Ogata, N., Ikeda, K., Aihara, H., Pangtay, I., Hibi, T., Kudo, S., and Tajiri, H., “Feasibility of stomach exploration with a guided capsule endoscope,” in [*J. of Endoscopy*], **42**(7).
- [10] Szczypinski, P. and Klepaczko, A., “Selecting Texture Discriminative Descriptors of Capsule Endoscopy Images,” in [*Proceedings of the 6th International Symposium on Image and Signal Processing and Analysis*], 701 (2009).
- [11] Cunha, J., Coimbra, M., Campos, P., and Soares, J., “Automated topographic segmentation and transit time estimation in endoscopic capsule exams,” *IEEE Transactions on Medical Imaging* **27**(1), 19–27 (2008).
- [12] Coimbra, M. and Cunha, J., “MPEG-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy,” *IEEE transactions on circuits and systems for video technology* **16**(5), 628–637 (2006).
- [13] Vilarino, F., Spyridonos, P., Pujol, O., Vitria, J., and Radeva, P., “Automatic detection of intestinal juices in wireless capsule video endoscopy,” in [*Proceedings of the 18th International Conference on Pattern Recognition*], **4**, 719–722 (2006).
- [14] Mackiewicz, M., Berens, J., and Fisher, M., “Wireless capsule endoscopy color video segmentation,” *IEEE Transactions on Medical Imaging* **27**(12), 1769–1781 (2008).
- [15] Signorelli, C., Villa, F., Rondonotti, E., Abbiati, C., Beccari, G., and de Franchis, R., “Sensitivity and specificity of the suspected blood identification system in video capsule enteroscopy,” *Endoscopy* **37**(12), 1170–1173 (2005).
- [16] Li, B. and Meng, M., “Computer-Aided Detection of Bleeding Regions for Capsule Endoscopy Images,” *IEEE Transactions on Biomedical Engineering* **56**(4), 1032–1039 (2009).
- [17] Bay, H., Tuytelaars, T., and Van Gool, L., “Surf: Speeded up robust features,” *Computer Vision–ECCV 2006*, 404–417 (2006).
- [18] Freund, Y. and Schapire, R., “A decision-theoretic generalization of on-line learning and an application to boosting,” in [*Computational learning theory*], 23–37, Springer (1995).
- [19] Bradski, G., “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 120–126 (2000).