Automated polyp detection in colon capsule endoscopy

by

Alexander V. Mamonov, Isabel N. Figueiredo, Pedro N. Figueiredo, Yen-Hsi Richard Tsai
Automated polyp detection in colon capsule endoscopy

Alexander V. Mamonov, Isabel N. Figueiredo, Pedro N. Figueiredo, Yen-Hsi Richard Tsai

Abstract—Colorectal polyps are important precursors to colon cancer, a major health problem. Colon capsule endoscopy (CCE) is a safe and minimally invasive examination procedure, in which the images of the intestine are obtained via digital cameras on board of a small capsule ingested by a patient. The video sequence is then analyzed for the presence of polyps. We propose an algorithm that relieves the labor of a human operator analyzing the frames in the video sequence. The algorithm acts as a binary classifier, which labels the frame as either containing polyps or not, based on the geometrical analysis and the texture content of the frame. The geometrical analysis is based on a segmentation of an image with the help of a mid-pass filter. The features extracted by the segmentation procedure are classified according to an assumption that the polyps are characterized as protrusions that are mostly round in shape. Thus, we use a best fit ball radius as a decision parameter of a binary classifier. We present a statistical study of the performance of our approach on a large data set containing over 18,900 frames. The algorithm demonstrates a solid performance, achieving 47% sensitivity per frame and over 81% sensitivity per polyp at a specificity level of 90%.

Index Terms—Capsule endoscopy, colorectal cancer, polyp detection, ROC curve.

I. INTRODUCTION

Colorectal cancer is the second most common cause of cancer in women and the third most common cause in men [12], with the mortality reaching to about 50% of the incidence. Colorectal polyps are important precursors to colon cancer, which may develop if the polyps are left untreated. Colon capsule endoscopy (CCE) [1], [4], [5], [8], [11], [15], [16], [17], [18] is a feasible alternative to conventional examination methods, such as the colonoscopy or computed tomography (CT) colonography [6].

In CCE a small imaging device, a capsule, is ingested by the patient. As the capsule passes through the patient's gastrointestinal tract, it records the digital images of the surroundings by means of an on-board camera (or multiple cameras). As the images are recorded, they are transmitted wirelessly to a recording device carried by the patient. Depending on the model of the capsule and its regime of operation, the images are captured at a rate ranging from 2 to 30 or more frames per second, with the low frame rate devices being prevalent currently. After the whole video sequence is recorded, it has to be analyzed for the presence of polyps. The video sequence of examination of a single patient may contain thousands of frames, which makes manual analysis of all frames a burdensome task. Using an automated procedure for detecting the presence of polyps in the frames can greatly reduce such burden. Thus, an efficient algorithm should not only be able to detect the polyps accurately (high sensitivity), but should also have a reasonably low rate of false positive detections (high specificity) to minimize the number of frames that have to be analyzed manually.

In this paper we provide an efficient algorithm for detecting polyps in CCE video frame sequences. The performance of the algorithm is assessed on a large data set, which ensures that no over-fitting takes place. The paper is organized as follows. The main idea and its comparison to existing approaches is discussed in section II. The steps of the algorithm are described in detail in section III. We summarize the algorithm and present the results of its performance on a large data set in section IV. Finally, we conclude with the discussion of the results and some directions for future research in section V.

II. BINARY CLASSIFIER WITH PRE-SELECTION

The proposed algorithm of polyp detection is based on extracting certain geometric information from the frames captured by the capsule endoscope’s camera. Such approach is not new, as it has been noticed before that the polyps can be characterized as protrusions from the surrounding mucosal tissue [7], [14], [19], [20], [21], which was first used in CT colonography. Thus, it is natural to compute some measure of protrusion and try to detect the frames containing polyps as the ones for which such measure is high. However, this leads to an issue that was also observed in the above mentioned works. The issue is distinguishing between the protrusions that are polyps and the numerous folds of healthy mucosal tissue. This problem can be alleviated by some form of image segmentation that takes place prior to the computation of the measure of protrusion [3], which is what we do in this work as well.

A particular choice of a measure of protrusion is of crucial importance. Many authors have proposed the use of principal curvatures and the related quantities, such as the shape index and curvedness [22], or the Gaussian and mean curvatures [7]. The main disadvantage of such approaches is that the computation of the curvatures is based on differentiation of the image, which must be approximated by finite differences.
In the presence of noise these computations are rather unstable, which requires some form of smoothing to be applied to the image first. However, even if the image is smoothed before computing the finite differences, the curvatures are still sensitive to small highly curved protrusions that are unlikely to correspond to polyps. Thus, in this work we use a more globalized measure of protrusion, the radius of the best fit ball. A similar sphere fitting approach was used in the CT colonography setting in [13]. In our approach we do not do the fitting to the image itself, but we first apply a certain type of a mid-pass filter to it. This allows us to isolate the protrusions within a certain size limit. We use the radius of the best fit ball as the decision parameter in a binary classifier. If the decision parameter is larger than the discrimination threshold, then the frame is classified as containing a polyp.

Another feature that distinguishes our approach from the ones mentioned above is the use of texture information. The surface of polyps is often highly textured, so it makes sense to discard the frames with too little texture content in them. On the other hand, too much texture implies the presence of bubbles and/or trash liquids in the frame. These unwanted features may lead a geometry-based classifier to classify the frame as containing a polyp when no polyp is present, i.e. they lead to an increased number of false positives. Thus, in order to avoid both of the situations mentioned above, we apply a pre-selection procedure that discards the frames with too much or too little texture content. Combined with the binary classifier this gives the algorithm that we refer to as binary classification with pre-selection.

III. DETAILS OF THE ALGORITHM

In the section below we present the detailed step-by-step description of the algorithm of processing of single frames from a capsule endoscope video sequence. The algorithm makes a decision for every frame whether to classify it either as containing polyps (“polyp” frame) or as containing normal tissue only (“normal” frame).

Besides the frame to be processed, the algorithm accepts as the inputs a number of numerical parameters that have to be chosen in advance. A complete list of these parameters and their values used in the numerical experiments is given in Table II at the end of section IV. For the purpose of numerical experiments, the values of these parameters were chosen manually, but we avoided any fine-tuning. The values were chosen from the common sense considerations, not from the considerations of improving the performance for a particular data set that we used. Ideally, we would like to have a systematic way to calibrate our algorithm, i.e. to assign the optimal values to the parameters based on the algorithm’s performance for some calibration data set. Currently, we do not have such a procedure, so this remains one of the topics of future research addressed in section V.

Another choice that requires a separate study is the choice of the color space of the frame. In this work we convert the captured color frames to grayscale before processing. This choice provides good polyp detection results, as we observe from the numerical experiments in section IV. However, we believe that certain improvements in this area are possible. For example, the polyps are often highly vascularized, so one would expect them to have a stronger red color component. Thus, one may use a measure of red color content in the frame, like the a component of the Lab color space [10], in polyp detection. Here we rely mostly on the geometrical information for polyp detection, but our algorithm could still be supplemented by the use of color information.

A. Pre-processing

Since the capsule endoscope operates in an absence of ambient light, an on-board light source is used to capture the images. Because of the directional nature of the light source, the captured frames are often subject to an artifact known as vignetting, which refers to the fall-off of intensity of the captured frame away from its center. As a first step of frame pre-processing we perform the normalization of intensity using the vignetting correction algorithm of Y. Zheng et al. [23]. Performance of the intensity normalization procedure is illustrated in Figure 1 (c).

The images acquired by the endoscope are of circular shape. The area of the rectangular frame outside the circular mask is typically filled with a solid color. This creates a discontinuity along the edge of the circular mask, which may cause problems in the subsequent steps of the algorithm. To remove this discontinuity we use a simple linear extrapolation to extend the values from the interior of a circular mask to the rest of the rectangular frame. This is shown in Figure 1 (d), where the radius of a circular mask $R_{\text{mask}}$ is taken to be slightly less than half the frame size.

B. Texture computation and convolution

Computation of the texture content in the frame is an important first step of the algorithm. We use the thresholding on the texture content as a pre-selection criterion, i.e. some frames are discarded from the consideration (and labeled as “normal”) based on the texture content alone.

To separate the pre-processed frame $f$ into the texture $t$ and cartoon $c$ components

$$f = t + c,$$  

we use an algorithm of Buades et al. [2]. The algorithm is based on iterative application of low-pass filtering by convolution with a Gaussian kernel. As its input parameters it accepts the number of iterations $n_{\text{iter}}$ and the standard deviation $\sigma_t$ of the Gaussian kernel in pixels. Hereafter, we treat the frame as a matrix $f \in \mathbb{R}^{N_y \times N_x}$, where $N_x$ is the width and $N_y$ is the height of the frame in pixels. The individual pixels are denoted by $f_{ij}$, $1 \leq i \leq N_y$, $1 \leq j \leq N_x$. The quantities having the same dimensions as the frame itself ($t$, $c$, etc.) are treated the same way.

The use of texture in pre-selection is motivated by two considerations. First, the surface of polyps is often textured, so discarding the frames with low texture content helps to distinguish the polyp frames from the frames with flat mucosa. Second, when trash liquids or bubbles are present in the frame, most of $f$ ends up in $t$, so we expect the texture content to
be abnormally high in this case. Since detecting polyps in the frames polluted with trash or bubbles is not feasible anyway, we may as well discard the frames with very high texture content.

Once we have a decomposition \( \{I_i\} \), we need to define a measure of texture content that would be appropriate for performing the pre-selection. The measure should be more sensitive to the presence of large textured regions and less sensitive to small regions even if those are strongly textured, since those typically correspond to occasional trash liquids or bubbles. Thus, we perform the following non-linear convolution-type transform of the texture

\[
T = L_\sigma(|t|^p),
\]

where the absolute value and exponentiation in \(|t|^p\) are pixel-wise, and \(L_\sigma\) is a linear operator convolving the frame with a Gaussian kernel with standard deviation \(\sigma\). The operator \(L_\sigma\) is also used in mid-pass filtering in the next step of the algorithm, and is defined as follows. First we define a one-dimensional Gaussian kernel on a stencil of \(2\lceil \sigma \rceil + 1\) pixels,
normalized so that it sums to one. Then an intermediate frame is obtained by convolving the original frame with the one-dimensional kernel using mirror boundary conditions when the stencil protrudes outside the frame. The final convolution is obtained by convolving the columns of the intermediate frame with the same stencil and mirror boundary conditions.

The usage of convolution in (2) with \( \sigma \) equal to a half of a typical polyp size, allows to emphasize the textured regions that are likely to be polyps. Adding non-linearity in the form of exponentiation with \( t \) with strong texture. This is illustrated in Figure 1 (e) and (f), where \( t \) and \( T \) are shown respectively. We observe that there are two well-pronounced bumps in \( T \). A larger in size and magnitude on the left corresponds to the polyp, a smaller one in the middle-right is due to a few bubbles present in the frame.

Once the non-linear transform (2) is calculated, we can compute the measure of texture content

\[
T_{max} = \max_{i,j} T_{ij}, \quad 1 \leq i \leq N_y, \quad 1 \leq j \leq N_x. \tag{3}
\]

Then the pre-selection criterion is a simple thresholding

\[
T_L \leq T_{max} \leq T_U. \tag{4}
\]

The lower bound filters out the frames with too little texture content that are unlikely to contain any polyps due to most polyps having a textured surface. The upper bound allows us to discard the frames polluted with trash and bubbles, since even if they contain polyps, they are likely to be obscured. To verify these considerations, we compare in Figure 2 the histograms of the distributions of \( T_{max} \) for normal and polyp frames in our test data set (see section IV-B for a detailed description). We observe that the peak of the histogram for normal frames is shifted towards the lower values of \( T_{max} \), which explains the effectiveness of the lower threshold. For high values of \( T_{max} \) the histogram for normal frames is consistently well above the one for the polyp frames, indicating a large number of frames polluted with trash and bubbles. Given such distributions of \( T_{max} \), the pre-selection criterion (4) appears quite effective. For the values of parameters given in Table II exactly 90% of the polyp frames pass the pre-selection, while only 47.84% of the normal frames do so.

C. Mid-pass filtering and segmentation

After the frame passes the pre-selection, we identify certain regions that may correspond to polyps. An essential feature of polyps is that they are protrusions or bumps on a flatter surrounding tissue. The purpose of this step is to detect such geometric features. Note that the polyps have a certain range of characteristic dimensions. Thus, in order to detect possible polyps, the geometrical processing should act as a mid-pass filter that filters out the features that are too small or too large. Here we use a mid-pass filter of the form

\[
u = H(w) \cdot w,
\tag{5}
\]

where \( w \) is defined by

\[
w = \frac{L_{\sigma_1}(f)}{L_{\sigma_2}(f)} - 1,
\tag{6}
\]

and \( H \) is the Heaviside step function

\[
H(x) = \begin{cases} 
0, & \text{if } x < 0, \\
1, & \text{if } x \geq 0.
\end{cases}
\tag{7}
\]

The application of \( H \), multiplication and division in (5) are pixel-wise. The standard deviations of the convolution operators satisfy \( \sigma_1 < \sigma_2 \). They correspond to the typical radii (in pixels) of the polyps that we expect to detect. We use a ratio in (5) instead of a difference to obtain a better scaling that depends less on the absolute prominence of the protrusion, but more on its relative prominence compared to the surrounding tissue. Since the convolution with a smaller standard deviation is in the numerator, the protrusions correspond to large positive values, hence we only use of the positive part of \( w \).

In Figure 1 (g) we show the result of mid-pass filtering \( u \). We observe several features present in \( u \) with the most prominent one corresponding to a polyp. To perform the binary classification we need to assign a numerical quantity to each of these features that would determine how likely does each of them correspond to a polyp.

To separate the features from each other we use a binary segmentation via thresholding

\[
s = H(u - \Theta) \in \{0, 1\}^{N_y \times N_x},
\tag{8}
\]

where \( H \) is taken pixel-wise and the scalar threshold \( \Theta \) is defined by

\[
\Theta = \max \left( \min \left( \frac{1}{2} \max_{i,j} u_{ij}, M_U \right), M_L \right),
\tag{9}
\]

with some bounds \( M_U > M_L > 0 \). This means that the threshold \( \Theta \) is taken to be a half of maximum value of \( u \), unless it goes above \( M_U \) or below \( M_L \), in which case it defaults to the corresponding bound.

An example of a binary segmentation \( s \) obtained with (8) is shown in Figure 1 (h), where four features can be seen. By features we mean the connected components of \( s \), which can
be found using an algorithm by Haralick and Shapiro [9]. It provides a decomposition
\[ s = \sum_{k=1}^{N_C} s^{(k)}, \]
where \( N_C \) is the total number of connected components in \( s \), and \( s^{(k)} \) are the disjoint connected components. Decomposition (10) is illustrated in Figure 1(h), where the four features \( s^{(k)} \) are numbered \( k = 1, \ldots, N_C \) in the order they are found by the algorithm.

D. Geometrical processing and the tensor of inertia

After the binary segmentation of the frame is decomposed into separate features (10), we process them individually to determine which of them might correspond to polyps. The simplest criterion we can apply is filtering the features by their sizes
\[ K_S = \left\{ k \in \{1, 2, \ldots, N_C\} \mid S_L \leq S^{(k)} \leq S_U \right\}, \]
where the size \( S^{(k)} \) of the \( k \)th feature is defined by
\[ S^{(k)} = \sum_{i,j} s_{ij}^{(k)}, \quad k = 1, \ldots, N_C. \]

Features that are too large, typically correspond to folds of normal mucosal tissue. Very small features are likely to be the artifacts of mid-pass filtering and the subsequent segmentation. The above feature size criterion can be used to discard such features.

A more sophisticated criterion that can help to eliminate the non-polyp features that pass the size criterion (11) is based on the computation of the features’ tensors of inertia to determine how stretched the feature is. For this we define the matrices
\[ x_{ij}^{(k)} = \begin{cases} j, & \text{if } s_{ij}^{(k)} = 1 \\ 0, & \text{if } s_{ij}^{(k)} = 0 \end{cases}, \]
\[ y_{ij}^{(k)} = \begin{cases} i, & \text{if } s_{ij}^{(k)} = 1 \\ 0, & \text{if } s_{ij}^{(k)} = 0 \end{cases}, \]
\[ 1 \leq i \leq N_y, 1 \leq j \leq N_x, \]
that allow us to compute first the centers of mass
\[ c_x^{(k)} = \frac{1}{S^{(k)}} \sum_{i,j} x_{ij}^{(k)}, \]
\[ c_y^{(k)} = \frac{1}{S^{(k)}} \sum_{i,j} y_{ij}^{(k)}, \]
\[ k = 1, \ldots, N_C. \]

Then we can define the tensors of inertia \( I^{(k)} \in \mathbb{R}^{2 \times 2} \) as
\[ I^{(k)} = \sum_{i,j} \begin{bmatrix} \left( y_{ij}^{(k)} \right)^2 & -x_{ij}^{(k)} y_{ij}^{(k)} \\ -x_{ij}^{(k)} y_{ij}^{(k)} & \left( x_{ij}^{(k)} \right)^2 \end{bmatrix}, \]
\[ k = 1, \ldots, N_C, \]
where \( x_{ij}^{(k)} = x_{ij}^{(k)} - c_x^{(k)} \) and \( y_{ij}^{(k)} = y_{ij}^{(k)} - c_y^{(k)} \) are the coordinates relative to the centers of mass. The tensors of inertia are symmetric positive definite, thus they can be used to define ellipses. Using an analogy from the classical mechanics, we refer to these ellipses as the ellipses of inertia. The eccentricities \( E^{(k)} \) of such ellipses are given by
\[ E^{(k)} = \frac{\lambda_{max}^{(k)}}{\lambda_{min}^{(k)}}, \quad k = 1, \ldots, N_C, \]
where \( \lambda_{max}^{(k)} \geq \lambda_{min}^{(k)} > 0 \) are the eigenvalues of \( I^{(k)} \). The eccentricity \( E^{(k)} \) determines how much is the \( k \)th feature stretched in one direction compared to its transversal. This information is useful for our purposes, since we expect the polyps to be more round in shape than the mucosal folds, that are often stretched.

We illustrate the above considerations in Figure 3 where we compare the ellipses of inertia for a polyp frame and two frames with pronounced mucosal folds. The ellipses we plot are
\[ r^{(k)}(\theta) = \sqrt{\frac{S^{(k)}}{\pi \lambda_{max}^{(k)} \lambda_{min}^{(k)}}} I^{(k)} \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} + \begin{pmatrix} c_x^{(k)} \\ c_y^{(k)} \end{pmatrix}, \]
where \( \theta \in [0, 2\pi] \). The scaling term in front of \( I^{(k)} \) is chosen so that the area of the ellipse of inertia is the same as the size \( S^{(k)} \) of the corresponding feature.

As expected, we observe that the ellipses corresponding to mucosal folds (feature 2 in the second row and features 2 and 3 in the third row of Figure 3) are indeed much more stretched out than the ellipse corresponding to a polyp (feature 1 in the first row of Figure 3). Stretched ellipses imply higher eccentricity, thus we impose the following criterion
\[ K_E = \left\{ k \in \{1, 2, \ldots, N_C\} \mid E^{(k)} \leq E_{max} \right\} \]
with some threshold \( E_{max} \) to select moderately stretched features that are more likely to correspond to polyps.

The combined geometric criterion is
\[ K_G = K_S \cap K_E. \]
If none of the features in the frame passes this criterion, i.e. if \( K_G = \emptyset \), then the frame is labeled as normal. If one or more features satisfy (19), then we continue to the next step, where we compute a parameter upon which we base the decision whether the frame is classified as containing polyps or not.

E. Decision parameter and binary classifier

The final step of the algorithm is the computation of the decision parameter that we use in the binary classification. This parameter is geometrical in nature and we define it as follows.

First, we compute the centers of mass using \( u \) masked with \( s^{(k)} \) instead of just \( s^{(k)} \), which gives
\[ \tilde{c}_x^{(k)} = \frac{1}{U^{(k)}} \sum_{i,j} x_{ij}^{(k)} u_{ij}, \quad \tilde{c}_y^{(k)} = \frac{1}{U^{(k)}} \sum_{i,j} y_{ij}^{(k)} u_{ij}, \]
for \( k \in K_G \), where
\[ U^{(k)} = \sum_{i,j} u_{ij} x_{ij}^{(k)}, \quad k \in K_G. \]
Fig. 3. Comparison of the geometrical processing and the tensor of inertia calculation for a polyp frame (first row) and two normal frames (second and third rows). Columns: (a) original frames, (b) mid-pass filtering \(u\), (c) binary segmentation \(s = \sum_{k=1}^{N_G} s^{(k)}\), (d) ellipses of inertia \(r^{(k)}(\theta)\) (yellow) with the centers of mass \((c^{(k)}_x, c^{(k)}_y)\) given by green \(\times\).

and the matrices \(x^{(k)}, y^{(k)}\) are defined in (13).

Second, we place a ball with a center at \((\tilde{c}^{(k)}_x, \tilde{c}^{(k)}_y)\) and we search of the radius of such ball so that it fits best the mid-pass filtered image \(u\). The radius of this ball will be the decision parameter in our binary classification. Such definition of the decision parameter is motivated by the same considerations as the criterion (19), i.e. we expect the polyps to be the protrusions that are somewhat rounded. Note that the combined geometric criterion (19) only uses the two-dimensional information in \(u\) by only working with the binary segmentation \(s\). To utilize information about the height of the protrusions, we need to work with \(u\) itself, which is why we fit the ball to \(u\) instead of \(s\).

To compute the optimal fit ball radius we define the matrix-valued functions
\[
b^{(k)}_{ij}(R) = \frac{1}{N_x N_y} \left( R^2 - (i - \tilde{c}^{(k)}_x)^2 - (j - \tilde{c}^{(k)}_y)^2 \right),
\]
\[1 \leq i \leq N_y, \quad 1 \leq j \leq N_x,
\]
and their positive parts
\[
\tilde{b}^{(k)}(R) = H(b^{(k)}_{ij}(R)) \cdot b^{(k)}(R), \quad k \in K_G,
\]
where Heaviside step function and the multiplication are performed pixel-wise.

Then for each feature we can define the radius of the ball that provides the best fit of \(u\) as a solution of a one-dimensional optimization problem
\[
R_{opt}^{(k)} = \arg\min_{R} ||u - \tilde{b}^{(k)}(R)||_F, \quad k \in K_G,
\]
where \(||.||_F\) is the matrix Frobenius norm. Since the objective in the optimization problem (24) is cheap to evaluate, the problem can be easily solved by a simple one-dimensional search over the integer values in some interval, which we take here to be \([1, \lfloor N_x/3 \rfloor]\).

Finally, we can define the decision parameter by taking the maximum of the optimal fit ball radii over all the features that pass the combined geometric criterion (19) as
\[
R_{max} = \max_{k \in K_G} R_{opt}^{(k)},
\]
where \(K_G = \emptyset\) we set \(R_{max} = 0\). To account for the pre-selection criterion (4), we also set \(R_{max} = 0\) for the frames with \(T_{max} < T_L\) or \(T_{max} > T_U\).
Since we expect the polyps to correspond to more pronounced round protrusions, we define the binary classifier as

$$BC(f) = \begin{cases} 
\text{"polyp"}, & \text{if } R_{\text{max}} \geq R_P \\
\text{"normal"}, & \text{if } R_{\text{max}} < R_P 
\end{cases}$$  \quad (26)

Obviously, the performance of the classifier depends dramatically on the choice of the \textit{discrimination threshold} value $R_P$. This choice is discussed in detail in section IV-C where we use statistical analysis to determine the value of $R_P$ that provides the desired performance.

In Figure 4, we show the circles of radius $R_{\text{max}}$ corresponding to the features that were correctly classified as polyps by (26). We observe that the classifier was able to identify the polyps of a variety of shapes even in the presence of small amounts of trash liquid (first row) or when the polyps are located next to mucosal folds (rows two to four in column (c)).

The good performance of the classifier can be explained by statistical considerations. In Figure 5, we show the histogram of the distribution of $R_{\text{max}}$ for normal and polyp frames in our test data set (see section IV-B for the detailed description). We observe that the peak of the distribution of $R_{\text{max}}$ for
polyp frames is shifted to the right compared to the peak of the distribution for “normal”, non-polyp frames. Note that overall the distribution for normal frames is well below the distribution for the polyp frames. This is due to setting $R_{\text{max}}$ to zero for the frames that do not pass the pre-selection or that do not satisfy the combined geometric criterion.

### IV. Numerical results

In this section we provide the results of a statistical test of performance of our algorithm on a large data set. The algorithm’s implementation and all the computations were performed with Matlab and the Image Processing Toolbox. The values of all the parameters used in the trial runs are given in Table I.

#### A. Summary of the algorithm

Below we summarize the processing of a single frame in a video sequence. The algorithm accepts the frame as an input and gives the decision “polyp” or “normal” as an output. The flow of data in the algorithm is illustrated in Figure 6:

**Algorithm 1** (Binary classification with pre-selection).

1) Pre-processing: convert the frame to grayscale, perform the intensity normalization, apply the linear extension outside the circular mask of radius $R_{\text{mask}}$ to obtain the pre-processed frame $f$.

2) Compute the texture $t$ from $f$ and its non-linear convolution-type transform $T$ using (2). Find the maximum $T_{\text{max}}$ of $T$ and apply the pre-selection criterion (2). If the frame fails the criterion, set $R_{\text{max}} = 0$ and go to step 6 otherwise continue.

3) Apply the mid-pass filtering $\mathcal{T}(k)$ to obtain $u$ and perform the binary segmentation $\mathcal{B}(u)$ of $u$ to obtain $s$. Decompose the binary image $s$ into connected components $s(k), k = 1, \ldots, N_C$ that correspond to $N_C$ “features” present in the frame.

4) For each of $N_C$ features compute the tensor of inertia $I^{(k)}$ via (15) and the eccentricity $E^{(k)}$ of the corresponding ellipse of inertia (16). Apply the eccentricity criterion (18) and the feature size criterion (17) to obtain the features $K_G$ that satisfy both criteria. If $K_G = \emptyset$, set $R_{\text{max}} = 0$ and go to step 6 otherwise continue.

5) For each of the features passing the combined geometric criterion (19) compute the radius $R^{(k)}_{\text{opt}}$ of the best fit ball. Take the maximum $R_{\text{max}}$ of these radii over $k \in K_G$ (25).

6) Apply the binary classifier $\mathcal{C}(s)$ to $R_{\text{max}}$ to classify the frame as either “normal” or “polyp”.

Note that the algorithm is quite inexpensive computationally. The computational cost of processing a single frame is of the order of $O(N_x N_y)$ operations. None of the steps requires a solution of expensive PDE or optimization problems, that are often used in image processing. The only minimization sub-problem is a simple one-dimensional search (24), which can be done very efficiently. Even with a very crude Matlab implementation, the algorithm takes less than one second per frame on a regular desktop. Obviously, one would expect a proper C/C++ implementation to be a lot more efficient. This gives an advantage to our algorithm in the common situations where the captured video sequence contains thousands of frames, which makes the processing time an important issue.

#### B. Data set

A key to developing an efficient and robust algorithm for polyp detection is being able to test it on a sufficiently rich data set. Using a small number of sample frames can easily lead to overfitting that may create an illusion of good performance, but in realistic conditions such an algorithm can easily fail.

In this work we were able to use a large data set courtesy of the Hospital of the University of Coimbra. The data set contains $N = 18968$ frames, out of which $N_N = 18738$ are normal frames and $N_P = 230$ frames contain polyps. The frames are taken from the full exam videos of five adult patients. The videos were captured with PillCam COLON 2 capsule (manufactured by Given Imaging, Yoqneam, Israel, http://www.givenimaging.com) in the native resolution of $512 \times 512$ pixels and were downsampled to $N_x = N_y = 256$ before processing.

The “normal” part of the data set is organized into short video sequences of around 100 frames each. These sequences contain mucosal folds, diverticula, bubbles and trash liquids, which allows us to test our algorithm in realistic conditions. The sheer number of non-polyp frames in the data set ensures that our algorithm not only has a high sensitivity (high true positive rate), but also high specificity (low false positive rate).

The part of the data set containing the frames with polyps is organized into sequences corresponding to each polyp. A total of 16 polyps are present in 230 “polyp” frames. The lengths of these sequences are given in Table I. Grouping the frames into sequences allows us to study the performance of the algorithm not only on per frame basis, but also on per polyp basis. The second row of Table I presents the results of such study, which is described in detail in the next section.

![Histogram of the distribution of $R_{\text{max}}$ for normal and polyp frames after passing the pre-selection and satisfying the combined geometric criterion.](image)
Then we can define the true positive rate of every frame in the data set as either “polyp” or “normal”. To define the ROC curve we need to introduce the following quantities. Suppose that we know a true classification of each frame in the data set. The measures of performance should also capture the statistical nature of the testing, which leads us to the consideration of receiver operator characteristic also referred to as the ROC curve.

ROC curves are a standard tool for evaluating the performance of binary classifiers. They quantify the change in performance of a classifier (in our case $BC(f)$ defined in (26)) as the discrimination threshold value (in our case $R_P$) is varied. To define the ROC curve we need to introduce the following quantities. Suppose that we know a true classification $TC(f)$ of every frame in the data set as either “polyp” or “normal”. Then we can define the true positive rate $TPR$ and the false positive rate $FPR$ as

$$TPR = \frac{1}{N_P} \left( \begin{array}{c} \text{number of frames } f \text{ s.t. } TC(f) = BC(f) = \text{“polyp”} \\ \text{number of frames } f \text{ s.t. } TC(f) = \text{“normal”}, \text{ but } BC(f) = \text{“polyp”} \end{array} \right), \quad (27)$$

$$FPR = \frac{1}{N_N} \left( \begin{array}{c} \text{number of frames } f \text{ s.t. } TC(f) = \text{“normal”}, \text{ but } BC(f) = \text{“polyp”} \end{array} \right). \quad (28)$$

The true positive rate is also known as sensitivity

$$SENS = TPR \cdot 100\%, \quad (29)$$

which measures how likely the classifier is to correctly label a non-polyp frame as “normal”.

Obviously, we would like both the sensitivity and the specificity to be as high as possible. However, there is always a trade off between the two. To visually represent this trade off we use the ROC curve. It is a parametric curve in the space $(FPR, TPR)$, where the changing parameter is the discrimination threshold $R_P$. A ROC curve connects the points $(0,0)$ and $(1,1)$. If the classifier makes a decision randomly with equal probability, then the ROC curve will simply be a diagonal $TPR = FPR$. For any classifier that behaves better we expect to have a concave ROC curve that deviates far from the diagonal.

We show the ROC curve for the binary classifier (26) in Figure 7 (a). Note that it does not go all the way to $(1,1)$. The reason for that is the use of pre-selection and combined geometric criteria in steps 2 and 4 of Algorithm 1 respectively. For the frames that do not satisfy those criteria we set $R_{max} = 0$, but for plotting the ROC curve we only use positive $R_P$. However, this limitation is not important, since the portion of the ROC curve that is of most interest to us is the one corresponding to the small values of $FPR$. This is due to the fact that an overwhelming majority of the frames in the endoscopy video sequences are non-polyp frames (all of them for healthy patients). The frames that the algorithm labels as “polyp” have to be inspected manually by a doctor. Thus, to minimize the work that the doctor has to do, the specificity has to be high.

Based on the above considerations, we assess the performance of Algorithm 1 in the following way. First, we fix a desired level of specificity. Here we take $SPEC = 90\%$. Then we select the minimal value of $R_P$ that provides at least
that much specificity. Finally, we compute the corresponding level of sensitivity of the binary classifier. On our data set with $R_P = 37$ we were able to achieve a specificity of 90.2% while having a sensitivity of 47.4%. This is a good performance, especially considering the fact that it is achieved on such a large and diverse data set. However, such level of sensitivity for single frames can actually imply even better performance for video sequences, which is a more relevant way of evaluating the usefulness of the algorithm in the real clinical setting, as explained below.

In the clinical setting the main purpose of the automated processing of the capsule endoscope videos is not to detect the individual polyp frames, but to find the actual polyps. Typically one polyp will be visible not just in a single frame, but in a sequence of frames, even for the cameras with low frame rates. Since the frames classified by the algorithm as “polyp” are going to be inspected manually afterwards, it is enough for a single frame in the sequence to be classified as “polyp” for the actual polyp to be detected.

To measure the sensitivity of the algorithm to actual polyps instead of the single polyp frames, we can use a modified definition of $TPR$. For each of $N_S$ sequences corresponding to a single polyp we define the detection flags

$$D^{(p)} = \begin{cases} 1, & \text{if for at least one } f \text{ in the } p^{th} \text{ sequence } \\
0, & \text{if for all } f \text{ in the } p^{th} \text{ sequence } \end{cases}$$

Then we can define a per polyp $TPR$ as

$$TPR = \frac{1}{N_S} \sum_{p=1}^{N_S} D^{(p)}. \quad (32)$$

We also call the corresponding sensitivity and the ROC curve as defined on a per polyp basis. This is to distinguish from those based on (27), that we may refer to as being defined on a per frame basis.

In Table I we show the values of the detection flags $D^{(p)}$ for all $N_S = 16$ sequences. These values are obtained for the value of the discrimination threshold $R_P = 37$, which gives the per frame sensitivity of 47.4% as we established earlier. However, the sensitivity per polyp in this case is 81.25%, i.e. the algorithm correctly detects 13 out of 16 polyps in at least one frame of each corresponding sequence. Thus, as expected, the algorithm has a much better performance when a single polyp is present in a number of consecutive frames. This is further confirmed by Figure 7 (b), where we show a per polyp ROC curve. In fact, we observe that we can obtain a specificity of 93.47%, while still maintaining the same per polyp sensitivity of 81.25% if we take $R_P = 40$.

V. CONCLUSIONS AND FUTURE WORK

In this paper we developed an algorithm for automated detection of polyps in the images captured by a capsule endoscope. The problem of polyp detection is quite challenging due to a multitude of factors. These include the presence of trash liquids and bubbles, vignetting due to the use of a non-uniform light source, high variability of possible polyp shapes and the lack of a clear cut between the geometry of the polyps and the folds of a healthy mucosal tissue. We attempt to overcome these issues by utilizing both the texture information and the geometrical information present in the frame to obtain a binary classification algorithm with pre-selection. We perform a thorough statistical testing of the algorithm on a rich data set to ensure its good performance in realistic conditions. While our approach is by no means an ultimate solution of the automated polyp detection problem, the achieved performance makes this work an important step towards a fully automated polyp detection procedure.

Throughout the paper we identified some directions of future research and the possible areas of improvement of the algorithm. We summarize them in the list below.
Currently, the frame is converted to grayscale before processing, so the algorithm only utilizes the information about the texture and the geometry. It would be beneficial to also utilize the color information present in the frame. For example, the amount of red color can point to polyps that are highly vascularized.

The algorithm relies on a number of parameters. For the purpose of this study the values of most parameters were chosen manually. Alternatively, one may use an automated calibration procedure to choose these parameters. Developing such a procedure remains a topic of future research.

In order to properly assess the size and shape of the protrusions, the algorithm should be able to correctly infer the actual height map of the object from the intensity information in the image. This presents a particular challenge when an object of interest is located in the dark section of the image. Using the mid-pass filtering provides an adequate solution to inferring the height map from the image. However, we would like to investigate possible alternatives to the approach used here.

ACKNOWLEDGMENTS
This work was partially supported by CoLab, the UT Austin | Portugal International Collaboratory for Emerging Technologies (http://utaustinportugal.org), project UTAustin/MAT/0009/2008 and also by project PTDC/MATNA/0593/2012, by CMUC and FCT (Portugal, through European program COMPETE/FEDER and project PEst-C/MAT/UI0324/2011). The work of Y.-H. R. Tsai was partially supported by Moncrief Grand Challenge Award.

REFERENCES

