Automatic detection of colonoscopic anomalies using capsule endoscopy

Limamou Gueye,
and Sule Yildirim-Yayilgan
and Faouzi Alaya Cheikh
Faculty of Computer Science
Gjvik University College, Norway

Ilangko Balasingham
Biomedical Sensor Network Research Group,
Oslo University Hospital, Norway

Abstract—Colon cancer and precancerous colon lesions are major health problems. The most frequent precancerous cancer lesions are polyps characterized by abnormal tissue growth in the human bowel. There are also anomalies such as inflammation and bleeding area. Colon capsule endoscopy (CCE) is a recent promising technology which enables to obtain videos of the inside of the intestine via an on board digital camera. The small capsule ingested by the patient is recuperated after it passes through the whole gastrointestinal track. Expert gastroenterologists analyze the video sequence visually in order to find out frames containing abnormalities related to cancer. This process is labor intensive and time consuming.

In this paper we propose an algorithm that provides automatic video analysis in order to classify tissue regions in images into two categories: normal and abnormal. In order to achieve high correct classification rates, we first pre-process the image data by removing the noise in it and by normalizing the intensity values of the image pixels. Next, we use the well-known SIFT descriptor algorithm with Bag of Feature (BoF) approach for feature extraction. For training, Support Vector Machine (SVM) is trained on the extracted features using the training dataset. Finally a testing data set is used to assess the performance of the proposed algorithm. The early experimental results are very encouraging and show high correct classification rates, reaching up to 98.25% for images with polyps.

The novelty of the proposed algorithm is the combination of using specific SIFT features with BoF and the vignetting correction to capture the local characteristics of the abnormalities in capsule video endoscopy images.

Keywords—Anomalies, cancer, polyp, inflammation, video endoscopy, classification, colon, bowel, intestine, normal, abnormal, SVM, pill-camera, SIFT.

I. INTRODUCTION

Each year, more than one million of men and women are diagnosed with colon cancer. Colon cancer represents the third leading cancer death [8]. Some abnormal areas in the colon such as inflammation, bleeding and particularly polyps occur in the early phases of the disease before the occurrence of colon cancer [2]. To prevent the occurrence of cancer, it is important to find and remove the precancerous polyps and other abnormalities when they occur in order to decrease the risk of mortality. In order to detect those precancerous pathologies early enough, colonoscopy is the most sensitive method [8]. However, this procedure requires a lot of time to be used by specialized physicians, in addition to the discomfort and risks inherent in the procedure. Therefore, an alternative method called colon capsule endoscopy was established and was compared to the colonoscopy gold standard in several studies [8]. Endoscopic capsule is a pill-size camera that is swallowed by a patient and the endoscopy procedure consists of the pill traveling through the gastrointestinal track, capturing and transmitting frames wirelessly to a wearable data recorder. Those still images of the digestive tract can then be downloaded to a computer and viewed for diagnosis of gastrointestinal diseases by physicians. An example of a pill camera is shown in Figure 1.

Figure 1: An example of a capsule camera.

The results of the diagnosis obtained from this new technology are quite satisfactory. Unfortunately the frames taken during the camera-transition through the bowel must still be viewed and analyzed by expert doctors. Therefore in this paper, we focus on developing an automatic classification scheme in order to differentiate frames containing abnormalities from normal frames. Our database is made of 4 different video sequences with a total of around 10000 frames. Each video has a tag file containing labels done via manual classification of the image contents. The abnormalities present in the videos are polyps, inflammation, cancer and some bleeding areas. The first video sequence contains only frames with inflammation and normal frames; the second one contains normal frames but with bubbles; the third one contains frames containing polyps, some bleeding frames and some normal frames and the last video sequence contains cancer. The ground truth classification of the frames in these four sequences is provided with the video data which allows us to understand which parts of the videos are normal and which parts are not. In Figure 2, we present examples of polyp, cancer, inflammation and normal tissue.
complicated during the classification step when there are some normal frames containing bubbles. It would be complicated to make the differentiation between a normal frame containing one single bubble and an abnormal frame containing a single polyp.

Papers such as [8] related to colonoscopy image classification used statistics of Local Binary Pattern (LBP) (e.g. histogram) as texture features. LBP is very robust to noise and illumination changes. It is a very efficient texture operator which labels each pixel of an image by comparing its value with its surrounding neighborhood pixels and considers the comparison result as a binary number. The LBP features have been used for polyp [6] and bleeding [7] detection. However, in our method, we have not only taken the texture problem into account, but also the shape, the texture and the color in the frames are taken into account.

Scale-invariant feature transform (SIFT) is well used to extract local features [10]. Its particularity compared to the other techniques is its robustness to illumination, orientation and scale changes. In the following sections we will discuss our proposed method, the experiments and then comment on the different results.

II. RELATED WORK

Different image features such as color, shape or their combination have been used in capsule video endoscopy. In [1], texture and color rotation features have been used to detect ulcer regions in the stomach, which is different from the abnormalities in our case because we are dealing with human colon not stomach. The authors’ goal in [1] was to get a color rotation matrix with the red, green and blue components of the RGB color space and then to remap chromatic content towards specific color planes. The main objective of doing this is to meliorate and optimize the projection of the content-of-interest (ulcer region) on more informative color planes and promote the extraction of intrinsic characteristics that will facilitate the discrimination process. This method can also provide good results on polyp detection because most of the time polyps are highly vascularized (a vascularized polyp in a frame has more vessels and is more red than the rest of the area), so the polyps will have a stronger red component than the other parts of the area. In this case, color feature can be used to extract characterizing features of the regions in frames [2]. In our case, as in Figure-2b in the frame containing the polyp, we see that there is not much difference in terms of color between the polyp and the rest of the frame. All other frames contains polyps in our database are roughly similar to frame in Figure-2a in terms of size, shape and texture. This is explained by the fact that polyps are in their first stage (new born polyps). Hence the color-based method may not lead to satisfying results in our case.

In [3], the authors used directly the transformation from the RGB color space to CIELab color space through which the color and luminance components can be processed and analyzed individually. Each component is divided into overlapping patches in order to obtain the color histogram of each patch as features.

Shape-based approach is discussed in [2] in order to extract certain geometric information from the frames captured by the capsule endoscopic camera. The shape features can be easy to compute when the frame contains only polyps. This can be

III. METHODOLOGY

The images are captured during the capsule endoscopy of the human bowel. After swallowing the capsule, the duration of the filming is around 8 hours and during this time, the patient might do a lot of activities such as running, walking, drinking, eating or lying down. In addition to that, in the bowel, there is no ambient light but only capsule on-board light source. Under these conditions and also due to the optical properties of the camera lens, the frames may be of different visual quality (e.g. artifacts, illuminations, vignetting, etc). They are often subject to some artifacts known as vignetting which refers to an irregular distribution of the pixel intensities in the captured frame from its center [2]. Therefore a pre-processing step
is needed in order to reduce the image degradations due to illumination changes and high reflectance factors. Here the technique in [11] is used to solve the problem as follows:

Let's consider $F_i(x, y)$ be the i-th frame of the video and $I(x, y)$ be the illuminated image with the vignetting effects of the same frame i. We assume that there is an unknown original frame without degradations; let's call that original frame as $F_{oi}(x, y)$.

So we can write:

$$F_i(x, y) = I(x, y)F_{oi}(x, y),$$  \hspace{1cm} (1)

where $I \in [0, 1]$.

Given a variable $\beta$ such that $\beta \in [0, 1]$, the final original frame without degradations is defined as:

$$F_{oi}(x, y) = \frac{F_i(x, y)}{\beta I(x, y) + (1 - \beta)}$$

B. Feature extraction

Regarding the characteristics of the training dataset discussed in the previous section, the selected descriptor should be robust, able to handle rotation, scale and intensity. Therefore, we used SIFT which is a well-known descriptor [10]. Each frame from the training set is divided into several local patches and using SIFT, each patch is converted to a 128-dimensional vector. This yields a collection of descriptor vectors with the same size (128).

Consequently, we combined SIFT with the BoF method. It is one of the most popular visual descriptors technique to visually classify data. Its concept is based on the BoW (Bag of Words) which consists of obtaining a set of bags of features from a large dataset and clustering is applied on the set of bags in order to build the visual vocabulary. The algorithm composes of two parts and is given as follows:

- A- Creation of the bags of features:
  - Selection of a large training dataset.
  - SIFT feature points detection and the computation of corresponding descriptors.
  - Clustering the descriptors to a number of words to build the visual vocabulary using k-means.
- B- Obtaining the BoF descriptor for any given image or image set:
  - Extracting the SIFT features points from the image.
  - Computing the descriptors of the obtained points. - Matching the descriptors with those in the visual vocabulary to build the histogram.

C. Classification

Now we have a big matrix where all the descriptors of our dataset are stored and are used to distinguish different abnormal frames from normal ones. We have in this case only two categories because polyp, inflammation and cancer frames are considered as one class (abnormal frames).

Therefore, for the classification step, we used the linear SVM. SVM separates data into different classes by using hyperplanes in a high dimensional feature space. In the training process, decisions are given a class label (true or false) to form feature-class pairs $(x, y)$. Given a training data of n decisions $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ for p-dimensional feature space $x_i \in \mathbb{R}^p$ and $y_i \in [1, -1]$, one can write the optimized hyperplane separating the different classes as:

$$f(x) = w^T \phi(x) + b = 0,$$  \hspace{1cm} (2)

where $w$ is known as the weight vector and b is the bias; $\phi(x)$ is a function to map vector x into a higher dimensional space. The vectors $x_i$ are the patterns with their associated labels $y_i$.

For separating two different classes, the following optimization problem needs to be solved:

$$\min \left( \frac{1}{2} w^T C \sum_{i=1}^{N} \xi_i \right)$$  \hspace{1cm} (3)

subject to $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i$, $\xi_i \geq 0$ where C is the penalty parameter. This optimization goal is to minimize the distance of all the training points. After obtaining the hyperplane, the decision function for the classification is defined as:

$$h(x) = \text{sign}(f(x))$$  \hspace{1cm} (4)

This indicates the side of the hyperplane that a new pattern x is classified as part of. An illustration is shown in Figure 3.

![SVM hyperplane](image)

**Figure 3:** SVM hyperplane

Two input parameters are needed in order to run the SVM function: the first one is a vector that will contain all the SIFT descriptors of the training data and the second parameter is another vector which contains the class labels of the training data. In the training data, normal frames are labelled to 1 and abnormal frames to -1.

After the training function, an SVM model is automatically generated. This model is used for testing and will classify a frame as containing polyp, inflammation, bleeding or cancer.

IV. Experiments

The different videos provided in the data set have varying combinations of noise and illumination changes. For training, different representative subsets of the data were formed with different number of images. As discussed earlier, the SIFT-based technique is used to extract the local features.

The classification is performed to identify images with polyps or inflammation separately by using two different classifiers. Firstly, we trained an SVM classifier with 300 normal and 300 abnormal frames, containing only inflammation. The classification rate is 88.50% in this case. Later we trained the classifier with 150 normal and 150 abnormal frames. The classification rate increased to 92.6%. Our results are shown in Table 1.
Table I: Normal vs. Inflammation

<table>
<thead>
<tr>
<th>Data size</th>
<th>Classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 normal-300 abnormal</td>
<td>88.50%</td>
</tr>
<tr>
<td>150 normal-150 abnormal</td>
<td>92.67%</td>
</tr>
</tbody>
</table>

Next we trained an SVM, with 400 normal and 400 abnormal frames where the abnormal frames contained polyps only. Table 2 shows the classification rate is 61.83%. When the number of training frames is reduced to 200 normal and 200 abnormal, the classification rate increases to 98.25%.

Table II: Normal vs. Polyps

<table>
<thead>
<tr>
<th>Data size</th>
<th>Classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>400 normal-400 abnormal</td>
<td>61.83%</td>
</tr>
<tr>
<td>200 normal-200 abnormal</td>
<td>98.25%</td>
</tr>
</tbody>
</table>

Last we trained one more SVM classifier with 300 normal and 300 abnormal frames to identify two of the anomalies at the same time. This time the abnormal frames include either polyps or inflammations whereas the normal frames have neither of them. The classification rate obtained is 88.50% as shown in Table 3.

Table III: Normal vs. (Polyps+Inflammation)

<table>
<thead>
<tr>
<th>Data size</th>
<th>Classification rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>300 normal-300 abnormal</td>
<td>88.50%</td>
</tr>
</tbody>
</table>

In each case above, the size of the testing data was the same as the size of the training data.

V. CONCLUSION

In this paper we proposed an automatic video frame analysis algorithm, which is able to classify anomalies in the images into two categories: normal and abnormal. In order to achieve that objective, we propose to use the SIFT descriptor. A binary classifier is used in order to label a frame as containing abnormalities or not based on the proposed features. The early experimental results are very encouraging and show high correct classification rates.

When the abnormalities are classified separately, we can see that the polyp classification has better results with 98.25% compared to 92.67% for the inflammation classification. This can be explained by the fact that the polyps are easier to detect thanks to their characteristic shape features.

This project is still ongoing and in the near feature. The local binary pattern (LBP) technique will be tested for the feature extraction and will be compared to the SIFT technique.

VI. REFERENCES


[4] A Vedaldi (Department of Engineering Science Oxford University, UK) and B. Fulkerson (Computer Science DepartmentUniversity of California at Los Angeles, CA, USA” An Open and Portable Library of ComputerVision Algorithms”


