Color and Position versus Texture Features for Endoscopic Polyp Detection

Luís A. Alexandre
Department of Informatics, Univ. Beira Interior, Portugal
IT - Networks and Multimedia Group, Covilhã, Portugal

Nuno Nobre
Department of Informatics, Univ. Beira Interior, Portugal

João Casteleiro
Department of Informatics, Univ. Beira Interior, Portugal

Abstract

This paper presents a comparison of texture based and color and position based methods for polyp detection in endoscopic video images. Two methods for texture feature extraction that presented good results in previous studies were implemented and their performance is compared against a simple combination of color and position features. Although this more simple approach produces a much higher number of features than the other approaches, a SVM with a RBF kernel is able to deal with this high dimensional input space and it turns out that it outperforms the previous approaches on the experiments performed in a database of 4620 images from endoscopic video.

1. Introduction

A polyp is an abnormal growth of tissue on a mucous membrane. One of the places were polyps can appear is in the intestine. In this case, the polyps may develop into colon cancer. If they are detected in an early phase of development they can be easily removed. Unfortunately, many go undetected until far too late and can turn into colon cancer which is one of the most prevalent cancers in the developed countries: in the United States colorectal cancer is the third cancer related cause of death [11].

We are currently developing a system for detecting polyps on video captured by a device that is ingested by a patient (video capsule endoscopy). This video has typically many hours (over 12 hours) and it would be a waste of time to have a medical doctor examining it completely. The system will do a first pass and detect potential problems (in this case, polyps) so that the doctor's attention is used only on video segments were it is more probable to find a pathology.

Several previous works have considered the detection of intestine polyps [4, 3, 6], but these work mainly over Computerized Tomography (CT) images. The number of publications on the detection of polyps in video endoscopic images is very small. The ones that we found are refered bellow.
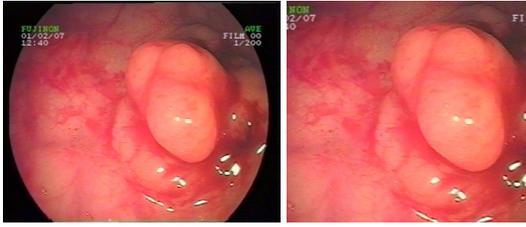
In [10] instead of trying to detect polyps in colon images, the authors are concerned with a method for evaluating if the colon status is normal or abnormal. They use PCA, texture and color features and neural networks for classification. They conclude that the best results are obtained by combining both texture and color information.

In [12] the authors were also concerned with abnormality detection from endoscopic images. They use a fusion approach to reach a final decision from sub-decisions made based on associated component feature sets. They report that the overall detectability of abnormalities using the fusion approach is improved when compared with corresponding results from the individual methods.

In [9] a general approach for image texture feature extraction called the Local Binary Pattern (LBP) was proposed. It was used in [7] as a texture feature for polyp detection in endoscopic images.

In [8], an approach called Color Wavelet Covariance (CWC) was also proposed with the same goal: estimating texture features for the detection of polyps.

In this paper we implemented the LBP and CWC approaches and compare their performance against a much simpler feature extraction: using for each pixel its RGB values and its position in an image window. Surprisingly, this simple approach yielded very good results on the experiments we made. We conducted experiments on a database of 4620 images from video endoscopy using 10-fold cross validation and a Support Vector Machine as a classifier with a Radial Basis Function kernel. The three approaches all used the same setup and were evaluated on the same data.

IEEE computer society

Figure 1. An image before and after the operation that removes the black frame.

The comparisons were made on the average area under the ROC curve values.

The paper is organised as follows: the next section explains the image pre-processing that is applied before all the feature extraction methods; these are detailed in section 3. The experiments are presented in section 4 and the final section contains the conclusions.

2. Image pre-processing

Before the feature extraction stage, the images were pre-processed. This includes a frame removal step, a division of each original image into sub-images and the determination of the class (normal or polyp) of each sub-image to produce the dataset used in the experiments.
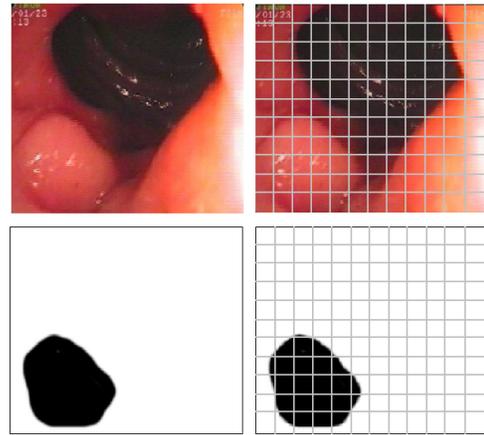
2.1. Frame removal

The videos were captured with PAL (768×576) resolution. The images have a black frame around the useful region of image as in figure 1. This black frame is removed leaving each image with a resolution of 514×469. This approach discards some of the useful area, but since we are working with video, we can recover the lost data from other video frames.

2.2. Image division

Our approach considers the division of the original images into smaller sub-images, that is, we will not classify directly an input image but, we subdivide it and classify each sub-image individually as containing a polyp or not. Then this information can be used to classify the original image.

The idea consists in processing sub-images that can sometimes be completely contained within the polyp region. This means that ideally we should use sub-images of the size of a single pixel. Of course this would not produce enough data to have statistically significant results on the sub-image level. So, we define the sub-image area with

Figure 2. From left to right and top to bottom: original image, its subdivision, the corresponding manual classification mask (black means polyp) and its subdivision.

dimensions of 40×40 pixels. This is small enough such that the sub-images are frequently completely contained in the polyp region but are also big enough to produce significant feature results.

The sub-images were obtained by sliding a window with a 40 pixels step, both horizontally and vertically through the original image. This means that there is no overlap between the sub-images.

Given the dimensions of the input images after the black frame removal, the sub-division process generates 132 sub-images for each original image.

2.3. Classification masks

The procedure we will describe in this section was developed to simplify (automate) the manual classification of each of the sub-images, for the purpose of creating a dataset that will be used in the experiments section.

For each image, we produced a binary image that was used as a classification mask. This mask is a manual painted image the size of the original image, that has the polyp region painted black and the remaining portion is white (an example is shown in figure 2). The masks were validated by a grastroentrologist.

The division of the original image into sub-images is done also on the classification mask, yielding a sub-image that contains only black and white pixels. To assign a class label to the original sub-images we look at the corresponding classification mask sub-image and count the number of black pixels it contains. If this number is higher than a given threshold, in this case 1300, we consider that the sub-image

'contains' a polyp. The value of 1300 was obtained after experimenting with different values.

3. Feature extraction

In this section we describe the three approaches used for feature extraction. All features from these different approaches were reduced and centered.

3.1. Color and position

Our approach to feature extraction is quite simple and produced very interesting results. Given the capabilities of the SVMs in dealing with high-dimensional input data, we chose as features only color and position information for each pixel: RGB + XY. Each pixel in a sub-image is represented by five values: its RGB components and its co-ordinates in the sub-image. So each sub-image is in fact represented by a total of 8000 features: 5 features for each of the 1600 pixels.

3.2. Color Wavelet Covariance

The CWC method [7, 8] starts by converting the image from RGB to the I1I2I3 color space. Then each image channel is decomposed using a Discrete Wavelet Frame Transform (DWFT). The difference between the Discrete Wavelet Transform (DWT) and the DWFT is that the later does not downsample the images from one level to the next. The decomposition is made up to two levels and only the detail images of the second level are used for the subsequent processing. Coocorrence matrices are calculated for directions $0°, 45°, 90°$ and $135°$. These matrices are quantified to have only 64 levels. Four Haralick texture features are computed using these matrices [5]: angular second moment, correlation, inverse difference moment and entropy. Finally the covariance between pairs of these features is found giving 72 features for each image.

In the experiments section we changed the number of decomposition levels used and also the number of quantization levels of the coocorrence matrix while searching for a performance improvement.

3.3. Local Binary Pattern

In [9] a method is proposed for texture extraction from images. In this case the texture features are estimated using a 3×3 neighborhood. The value of each pixel in the neighborhood is compared with the value of the center pixel. The LBP value for this neighborhood is obtained by summing the binomial coefficients shown in figure 3 that correspond to the position of the pixels with value larger than the value

of the center pixel. A histogram of these LBP values is built for the image.

1	2	4
8		16
32	64	128

Figure 3. LBP binomial mask.

A contrast measure can also be used to build a two-dimensional histogram by combining the information of LBP with the contrast information. The contrast is the difference between the average value of the pixels that have a value larger then the center pixel and the average value of the pixels that have smaller value than the center pixel. These values are placed on a histogram together with the LBP values making a bi-dimensional histogram.

In [9] the LBP histogram has 256 bins. We found that better results can be obtained with a smaller number of bins. This is because if the size of the image under analysis is small, the number of neighborhoods may not be sufficient for a good characterization of the histogram. We will see in the experiments section the improvements that can be obtained by considering a smaller number of bins, for the size of the images used in our dataset.

Similar considerations can be made regarding the number of bins to consider for the contrast: in [9] it is proposed the use of 8 bins (although the authors say that 4 or 16 may give similar results), so we tested the method with 2 and 4 bins.

We also made experiments with and without the contrast.

4. Experiments

We evaluated the performance of the three analyzed methods using the values of the areas under the ROC curves (AUC), following the advice in [1]. The points in these curves are the average accuracies of 10-fold cross validations over the 4620 images. The value of γ (a SVM parameter, see below) was varied to produce the different points on the ROC curves.

4.1. Dataset

The dataset was obtained from 35 video frames recorded with a Fuji 410 video endoscope system at the Hospital Cova da Beira, Portugal, during the year of 2007. The images were subdivided into smaller images after the pre-processing described in section 2. Each image produced

132 sub-images. Each sub-image was defined as polyp or not polyp according to the correspondent sub-image obtained from the manual generated classification mask described in section 2.3. The resulting data set contained 4620 images each with a dimension of 40×40 pixels.

The features were centered and reduced such that, for each feature, the mean value is 0 and the standard deviation is 1.

4.2. Classifier

The classifier used was a support vector machine (LIB-SVM) [2]. The kernel type used was the radial basis function (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2),\ \gamma > 0$$

where $\mathbf{x}_i$ is an input and γ is a parameter inversely proportional to the kernel width. The SVM with this kernel has two free parameters to be set: $C > 0$ that corresponds to the penalty parameter of the error and γ.

In all experiments performed we used $C = 16$. It produced good results for all feature sets and the results were not sensitive to its exact value: similar results were obtained for values of $C = 8$ and 32.

A different weight can be assign to each class when the prior probabilities for each class are not equal. This is the case in our dataset given that there are more non-polyp images than polyp ones. This weighting was done using the svm-train parameters $w_0 = 1$ and $w_1 = 32$. These values were obtained experimentally.

4.3. LBP

We varied the number of bins in the texture histogram from 16 to 256 in powers of two, first, without using the contrast. We found that very similar results were obtained for 32 and 64 bins. So we decided to test 2 and 4 bins for contrast with these two values of the texture histogram.

The values of the areas under the ROC curve (AUC) curves obtained on these experiments are shown in table 1.

LBP	C=0	C=2	C=4
16	79.08	-	-
32	82.17	81.74	82.40
64	82.57	82.39	82.91
128	80.80	-	-
256	79.36	-	-

Table 1. Values of the AUC for LBP with the different number of bins for texture and contrast tested.

The best results were obtained with 64 bins for the texture histogram and 4 for the contrast, yielding a Area Under the Curve (AUC) value of 82.91.

Notice that the best result uses the contrast which is not a texture feature but a color feature. It can be seem from the table 1 that the use of the contrast only improved the results when 4 bins were used. For 2 bins, the performance decreased.

4.4. CWC

For CWC, we tried varying the number of detail images used to produce the features and also tried changing the number of levels used in the coocorrence matrices.

The results are shown in table 2. The best value for the AUC was obtained using the detail images from both first and second decomposition levels and using 64 levels in the coocorrence matrices. Since the number of images in this case is the double of the original proposal, the number of generated features is also the double: 144.

CWC version	64	256
second level only	74.49	72.78
both levels	76.50	73.99

Table 2. Values of the AUC for CWC with the different approaches tested.

4.5. Color and position

Figure 4 contains the ROC curves for the color and position method together with the best ROCs of the two other methods.
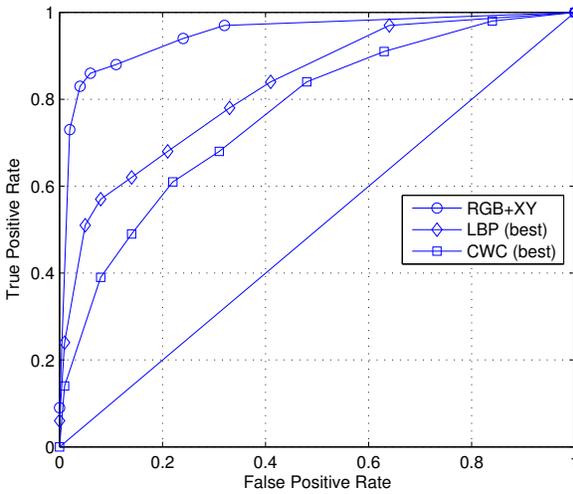
The areas under the Receiver Operating Characteristic (ROC) curve, (AUC), are 94.87, 82.91 and 76.50, for the RGB+XY, the LBP and the CWC respectively.

5. Conclusions

This paper presented a comparison between two texture based methods and a color and position method for the detection of polyps in video endoscopic images.

Despite the simplicity of the color and position approach, the results show a significant improvement when compared with the two other methods that have shown good results in previous studies.

The goal was to be able to develop a method that could do a first automatic screening of a endoscopic video and warn the physician of frames were attention is needed. Videos in which the method does not detect any polyp will

Figure 4. ROC curves for the color and position method and the best of the two other methods.

be given a smaller priority then others were possible polyps are detected.

Our method subdivides each image into smaller images (with 40×40 pixels). These are the images that are searched for polyps. Of course if a sub-image is considered a polyp, its parent image is also considered to have a polyp. (Other approaches can be used like the need for a number of detected sub-images with polyp in order to consider that the parent image contains a polyp. Our approach is the most cautious: it perhaps implies the existence of some false positives but will minimize the false negatives). We did not analyse this aspect in this paper though. We focused on the correct classification of the sub-images.

We used a very simple approach for feature extraction, relying only on color and pixel position. Although it creates many features per image (8000) the SVM was able to deal with this high dimensionality.

The results we obtained are quite satisfactory: in a database with 4620 images we were able to obtain an AUC value of 94.87 using 10-fold cross validation.

Future work will concern the application of this method to video images obtained from video capsule endoscopy instead of video from a colonoscope.

Acknowledgements

References

[1] A. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[2] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[3] T. Chowdhury, O. Ghita, and P. Whelan. A statistical approach for robust polyp detection in CT colonography. In *27th Annual International Conference of the Engineering in Medicine and Biology Society*, pages 2523–2526, September, 2005. IEEE.

[4] S. Gokturk, C. Tomasi, D. Paik, C. Beaulieu, and S. Napel. A learning method for automated polyp detection. In *4th International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001*, volume LNCS 2208, pages 85–92, Utrecht, The Netherlands, October 2001.

[5] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 13(1):610–621, 1973.

[6] A. Huang, R. M. Summers, and A. K. Hara. Surface curvature estimation for automatic colonic polyp detection. In A. A. Amini and A. Manduca, editors, *Medical Imaging 2005: Physiology, Function, and Structure from Medical Images. Proceedings of the SPIE*, volume 5746, pages 393–402, April 2005.

[7] D. K. Iakovidis, D. E. Maroulis, S. A. Karkanis, and A. Brokos. A comparative study of texture features for the discrimination of gastric polyps in endoscopic video. In *18th IEEE Symposium on Computer-Based Medical Systems (CBMS'05)*, pages 575–580. IEEE Computer Society, June 2005.

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

[9] T. Ojala and M. Pietikainen. Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32:477–486, 1999.

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

[11] U.S. Cancer Statistics Working Group. United states cancer statistics: 2003 incidence and mortality. Atlanta, USA, 2006. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute.

[12] M. Zheng, S. Krishnan, and M. Tjoa. A fusion-based clinical decision support for disease diagnosis from endoscopic images. *Computers in Biology and Medicine*, 35(3):259–274, March 2005.