Features for Optical Biopsy of Colorectal Polyps

Lucas Hadjilucas[1]
lucas.hadjilucas05@imperial.ac.uk
Anil A. Bharath[1]
a.bharath@imperial.ac.uk
Ana Ignjatovic[2]
James E. East[2]
Brian P. Saunders[2]
David Burling[3]

[1] Department of Bioengineering
Imperial College London
London, United Kingdom

[2] Wolfson Unit for Endoscopy
St Mark's Hospital
London, United Kingdom

[3] Department of Radiology
St Mark's Hospital
London, United Kingdom

Abstract

Despite current advances in endoscopic image acquisition, strong reliance is still placed on resecting and histologically examining colorectal polyps to assess their malignancy potential. In this study, we analyze the performance of a computer aided polyp classification system that uses features extracted from high magnification narrow band images to describe the density and irregularity of polyp pit patterns and separate polyps as either neoplastic or non-neoplastic. Our main features are based on the magnitude and angle of orientation dominance fields constructed with the use of a wavelet filter bank and a careful scale selection strategy. The features were tested using images from an Olympus™ Evis Exera endoscopic system and achieved a classification accuracy of 86.44% using a non-linear classifier with an *n*-fold cross validation strategy. The relatively high classification rate is a good starting point towards an automated optical biopsy system designed to decrease the miss rate of potentially malignant lesions.

1 Introduction

Colorectal cancer is the second most common cause of cancer related death in developed countries [1] with the majority of cases arising from adenomatous colorectal polyps. A polyp is defined as an abnormal growth of tissue on the surface mucosa of the colon. Polyps are not necessarily dysplastic or malignant and therefore not all give rise to cancer.

Colorectal polyps, such as those found on the mucosal lining of the colon, are histologically classified as neoplastic or non-neoplastic. The majority of non-neoplastic polyps, such as hyperplastic, have little malignancy potential and in some clinical protocols are treated as benign. Adenomatous polyps, or adenomas, on the other hand can have a high degree of dysplasia (villous adenomas) and hence greater potential for malignant change. An example of an adenomatous and a hyperplastic polyp under narrow band imaging (NBI) are shown in Figures 1(a) and 1(b) respectively.

The risk of cancer development can be reduced by up to 80% [2] by surgically removing or resecting the dysplastic polyp. The multiple classes and varied morphological appearance of polyps, however, make this task increasingly difficult. It is crucial to accurately discriminate between neoplastic and non-neoplastic lesions to avoid missing possibly malignant tumours and risking patient overtreatment. There are various methods of polyp detection and classification employed by clinical practice such as fecal occult blood tests (FOBT), sigmoidoscopy, colonoscopy and double-contrast barium enema.

(a) (b) (c)

Figure 1: (a) Example of an adenomatous polyp; (b) Example of a hyperplastic polyp; (c) visualisation of the local dominant orientation field (LDO) on part of polyp (b); the length of field lines corresponds to the LDO magnitude and their direction to the LDO orientation

Although studies to date have been inconclusive as to which is the best screening method [3] , it has been shown that colonoscopy can detect polyps that would otherwise be missed by sigmoindoscopy or FOBT [4]. In addition, chromoendoscopy, in the hands of an expert gastroenterologist, allows for precise optical diagnosis with accuracies ranging from 85-96% [2] for two classes. This advantage is offset by the time, cost and associated learning curve to achieve expertise. This means inexperienced surgeons often run the risk of incurring large time and comfort penalties on patients by needlessly resecting all polyps they are unable to visually classify.

Another limiting factor in polyp recognition is the inadequate pattern classification scales such as the Kudo's pit pattern [5] or Vascular Colour Intensity (VCI) scale that, although widely accepted by the medical community, are by no means refined enough to be absolute. For example, polyps with a light VCI are classified by colonoscopists as non-neoplastic but in practise 19% of adenomas have a light VCI [6]. The end result of this inadequacy to classify by optical means is that a large number of polyps are resected and sent for histopathology to decide their class. This introduces not only delays and added cost for the patient but also increased risk, as resecting a polyp can lead to perforation and blood loss in 0.1-0.2% of patients [7].

This highlights the need for an efficient, automated optical biopsy system to aid the *in vivo* classification of polyps with minimal invasive actions. By providing confidence on which are the neoplastic regions, it will also reduce the unnecessary resection of hyperplastic polyps for biopsy, cutting down further on cost and patient overtreatment.

2 Materials and Methods

2.1 Image Datasets

Our study dataset comprises of 118 images (59 adenomatous and 59 hyperplastic) obtained from an Olympus™ Evis Exera endoscopic system mainly used in the US and Japanese markets. This dataset, kindly provided by Douglas K. Rex (Indiana University School of Medicine, Indianapolis, USA) will be referred to as the 'Exera set'. All images are high magnification NBI images and are histopathologically cross checked to ensure they are correctly labelled. The polyp surface, used in feature generation and classification was manually segmented out of every image with the help of experienced endoscopists.

2.2 Image Pre-processing

Due to the reflective properties of the mucus present in the colon epithelium, a large number of specular reflections occur in the images. These could induce artifacts into spatial filter responses generated at a later stage so they had to be marked and removed.

We identified specularities by constructing a bivariate Intensity-Saturation histogram for each image. Specular reflections correspond mainly to regions that have high intensity and low saturation [8] and hence by selecting that area of the 2D histogram we were able to isolate specular regions.

2.3 Feature Construction

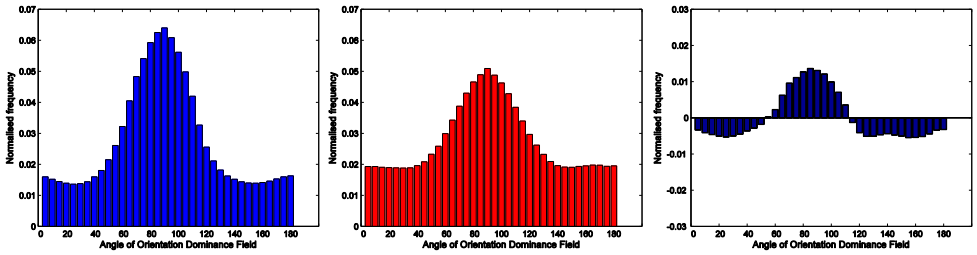
The vascularity appears in NBI images as dark lines on the polyp surface making its texture locally oriented along one direction. In order to quantify the strength and the direction of this orientation, we construct a complex field of local dominant orientations (LDOs), using the wavelet filtering technique proposed by Bharath and Ng [9]. In brief, an isotropic lowpass filter and a set of four oriented bandpass complex analysis filters, with orientations at angles $\phi_k = 0\,, \pi/4\,, \pi/2\,, 3\pi/4$, are used for decomposition at a given scale. Thus, the orientation dominance at scale l is computed as:

$$\boldsymbol{O}^{(l)}(m,n) = \frac{\Sigma_{k=1}^{4} \left| \boldsymbol{f}_k^{(l)}(m,n) \right| e^{j2\phi_k}}{p + \sqrt{\Sigma_{k=1}^{4} \left| \boldsymbol{f}_k^{(l)}(m,n) \right|^2}} \qquad (1)$$

where m, n are the image coordinates, $f_k^{(l)}$ is the output of the k^{th} oriented filter at scale l, j is defined as $\sqrt{-1}$ and p is a normalisation parameter set at 1% of the maximum value in the original image. Scale l is defined as $\left(\sqrt[4]{2}\right)^q$ where $q = 0,1,2,\ldots,q_{max}$. A visualization example of this field is shown in Figure 1(c).

To optimally characterise the vessels using Eq.(1) we must first find the optimal value of scale l, and hence q, at which the LDO field will have maximum response. This is directly related to the distance of the polyp from the camera, as well as the width of the vascularity. This is a parameter that other studies in polyp characterisation have often neglected [10] or assumed constant by fixing the zoom factor of the endoscope [11]. Since this information is not known a priori for every image we take an iterative approach to selecting the best possible scale. We therefore compute the LDOs for a number of finer and coarser scales, starting with the original image and each scale separated by a factor of $\sqrt[4]{2}$ up to a factor of $\left(\sqrt[4]{2}\right)^{q_{max}}$. For each scale, we compute the complex LDO field and apply a low pass filter to eliminate abrupt changes. Finally we select the scale, l, with the maximum response by keeping the field with the largest mean scalar value of the magnitude of orientation dominance.

As indicated by Kudo's pit pattern [5], routinely used by endoscopists, the vascularity of adenomas is denser, bolder and more irregular when compared to hyperplastic polyps. Ignjatovic et *al.* [12] made a good attempt to quantify this (using a similar approach) by examining histograms of the magnitude of the LDO field of the polyp surface. These features mainly encapsulate the difference in the prominence or boldness of the vessels but not their irregularity. We address this limitation by looking at how the angle of the orientation dominance field, $O^{(l)}(m,n)$ changes over the polyp surface.



(a) (b) (c)

Figure 2: Pooled histograms of angle of orientation dominance for Exera set: (a) adenomatous polyps; (b) hyperplastic polyps; (c) the difference of adenomatous and hyperplastic histograms

To obtain this angle, we split the complex field into its real and imaginary parts, $O^{(l)}(m,n) = u(m,n,l) + jv(m,n,l)$, and from complex arithmetic it follows that the angle of the orientation field, θ, can be obtained as in Eq. (2).

$$\theta(m,n,l) = arctan\left[\frac{v(m,n,l)}{u(m,n,l)}\right] \qquad (2)$$

The histogram of orientations is then computed, for each polyp, using 36 bins of 5^o width that cover the interval [0^o-180^o). In addition, the histograms are weighted by the magnitude of the orientation dominance fields so that only locations with a strong magnitude (and hence possibly vasculature) will contribute to the orientation histograms. Lastly, to achieve rotation invariance, the orientation histograms are also registered with respect to the highest bin. This allows us to process all histograms together, irrespective of the angle the polyp had relative to the capturing device. The normalised pooled histograms of orientations for the adenomatous and hyperplastic polyps for the Exera set are shown in Figure 1(a) and 1(b) respectively. As illustrated in Figure 1(c), the difference between the two histograms is indicative of the irregularity of the vessel structure between polyp classes.

To exploit this difference in a feature that can be used in a classifier, we select the bin interval that maximises the Fisher criterion, J in Eq.(3), between the two classes:

$$J = \frac{(\mu_1 - \mu_2)^2}{({\sigma_1}^2 + {\sigma_2}^2)/2} \qquad (3)$$

where μ_1, μ_2 are the means of the two classes and σ_1, σ_2 is the standard deviation. Essentially maximising J is equivalent to maximising the distance between the means of the two classes whilst minimising within class variance. By selecting the bin intervals, we avoid including poor features that would dilute the accuracy of our classifier.

In the case of Exera images, the bins that maximise J lie in the interval of [112.5^o-117.5^o] along the orientation dominance axis. A number of other statistical features were also extracted from the angle of orientation histograms including kurtosis (the adenomatous polyp histogram has a more leptokurtic shape than the hyperplastic case), skewness, radius of curvature, entropy and the Kolmogorov-Smirnof statistic. For our classification, we grouped the newly formed orientation features along with the features obtained by Ignjatovic et *al.* [12] in an attempt to increase overall classification accuracy.

2.4 Classifier

For polyp classification, we used a non-linear classifier by applying the kernel trick as described by Boser et *al.* [13] on a Support Vector Machine (SVM). Our SVM uses a non-linear radial basis function kernel. Due the small amount of data available, n - fold cross validation (leave-one-out) was used where $n = 118$ (118 images in our dataset) to assess how the classifier performance will generalise in an independent data set.

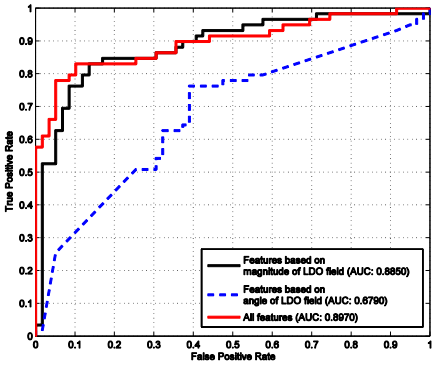


Figure 3: Receiver Operating Characteristic curves for Exera dataset

	Sensitivity	Specificity	Accuracy
Features based on magnitude of LDO field	83.05 %	86.44%	84.75%
Features based on angle of LDO field	76.27%	61.02%	68.64%
All features	**83.05%**	**89.83%**	**86.44%**

Table 1: Classification rates for different set of features for Exera

3 Results and Discussion

The results of SVM classification for each group of features (magnitude and angle of LDO field) for the Exera dataset are shown in Table 1. The corresponding Receiver Operating Characteristic (ROC) curve is shown in Figure 3. From the ROC curve we can conclude that each relevant feature can have a complimentary contribution to the overall classification accuracy of polyps. With overall accuracy of 86.44% in the case of Exera images the results are good for such a small feature vector. This is also comparable to the accuracy levels obtained by expert clinicians using chromoendoscopy (85-96%) [2]. We are confident that these metrics will improve on expansion of the datasets and feature vectors. Larger datasets will give greater generalisation and limit the irregularities between results of different datasets. Our future plan is incorporate a separate scale selection for the normal background mucosa and the foreground polyp (currently both use the same scale) and to build in better image optimisation methods to remove interlacing and chromatic aberration artifacts inherent to the mechanics of endoscopic system's image acquisition.

One limitation of the proposed algorithm is that it is not real time. In particular it takes 6 minutes to produce the LDO fields and do scale selection for high definition polyp images on a 2.6Ghz processor. In addition, the polyp boundaries were manually segmented in each image. We plan to address these limitations in the future by incorporating an automatic segmentation algorithm and using graphics processing units for LDO field generation.

4 Conclusion

We have presented an initial set of algorithms that can be used for *in vivo* optical biopsy of colorectal polyps, using high magnification NBI images. The features have been tested against a commonly used endoscopic system. Although many parts of the project are still

in progress, we have been able to acquire good polyp classification metrics with 86.44% for Exera systems using an SVM based non-linear classifier, results comparable to the performance of expert endoscopists. We have also developed a scale selection strategy to use polyps acquired at varied magnification in the same dataset. In the future we plan to also examine the stability of features across the polyp surface for robust classification.

Although the system has its limitations, we believe it is a good first step towards an automated polyp classification system. Such a system could decrease the miss rate of potentially malignant lesions. When coupled into a review system or to automated camera acquisition systems this could lead to savings in time, cost and reduced patient risk.

References

[1] B. Morson, "President's address. The polyp-cancer sequence in the large bowel.," *Proceedings of the Royal Society of Medicine*, vol. 67, no. 6 Pt 1, pp. 451–7, Jun. 1974.

[2] A. Ignjatovic, J. E. East, N. Suzuki, M. Vance, T. Guenther, and B. P. Saunders, "Optical diagnosis of small colorectal polyps at routine colonoscopy (Detect InSpect ChAracterise Resect and Discard ; DISCARD trial): a prospective cohort study," *Lancet Oncology*, vol. 10, no. 12, pp. 1171–1178.

[3] J. M. E. Walsh, "Colorectal Cancer Screening: Scientific Review," *JAMA: The Journal of the American Medical Association*, vol. 289, no. 10, pp. 1288–1296, Mar. 2003.

[4] D. A. Lieberman and D. G. Weiss, "One-time screening for colorectal cancer with combined fecal occult-blood testing and examination of the distal colon.," *The New England journal of medicine*, vol. 345, no. 8, pp. 555–60, Aug. 2001.

[5] S. Kudo, S. Tamura, T. Nakajima, H. Yamano, H. Kusaka, and H. Watanabe, "Diagnosis of colorectal tumorous lesions by magnifying endoscopy," *Gastrointestinal Endoscopy*, vol. 44, no. 1, pp. 8–14, Jul. 1996.

[6] J. N. Rogart, D. Jain, U. D. Siddiqui, T. Oren, J. Lim, P. Jamidar, and H. Aslanian, "Narrow-band imaging without high magnification to differentiate polyps during real-time colonoscopy: improvement with experience.," *Gastrointestinal endoscopy*, vol. 68, no. 6, pp. 1136–45, Dec. 2008.

[7] J. H. Bond, "Polyp guideline: diagnosis, treatment, and surveillance for patients with nonfamilial colorectal polyps. The Practice Parameters Committee of the American College of Gastroenterology.," *Annals of internal medicine*, vol. 119, no. 8, pp. 836–43, Oct. 1993.

[8] M. D. Levine and J. Bhattacharyya, "Detecting and removing specularities in facial images," *Computer Vision and Image Understanding*, vol. 100, no. 3, pp. 330–356, Dec. 2005.

[9] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 14, no. 7, pp. 948–59, Jul. 2005.

[10] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf, "Classification of colon polyps in NBI endoscopy using vascularization features," in *Medical Imaging 2009: Computer-Aided Diagnosis, Proceedings of the SPIE*, 2009, p. 72602S–72602S–12.

[11] M. Hafner, C. Kendlbacher, W. Mann, W. Taferl, F. Wrba, A. Gangl, A. Vecsei, and A. Uhl, "Pit Pattern Classification of Zoom-Endoscopic Colon Images using Histogram Techniques," in *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, 2006, pp. 58–61.

[12] A. Ignjatovic, A. Varnavas, J. E. East, A. A. Bharath, J. N. S. Kwong, B. P. Saunders, and D. Burling, "Computer aided polyp characterization through the quantification of vascularity and colour differences," *Gastrointestinal Endoscopy*, vol. 71, no. 5, pp. AB240–AB241, 2010.

[13] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*, 1992, pp. 144–152.