

# Analysis of multispectral images of the colon to reveal histological changes characteristic of cancer

Džena Hidović-Rowe<sup>a</sup>, Ela Claridge<sup>a</sup>, Tariq Ismail<sup>b</sup> and Phillipe Taniere<sup>b</sup>

<sup>a</sup>School of Computer Science, The University of Birmingham

<sup>b</sup>Birmingham University Hospital, Birmingham  
{D.Hidovic, E.Claridge}@cs.bham.ac.uk

**Abstract:** Colon cancer alters the macroarchitecture of the colon tissue. Common changes include the formation of new blood vessels (angiogenesis) and the distortion of the tissue collagen matrix. Such changes affect the colon colouration. This paper presents the principles of a novel image analysis method capable of extracting parameters depicting histological quantities of the colon from its multispectral images. The method is based on a computational, physics-based model of light interaction with tissue. The colon structure is represented by three layers: mucosa, submucosa and muscle layer, and parametrised by the concentration of blood haemoglobins; the size and density of collagen fibres; and the thickness of the layers. Using the entire histologically plausible ranges for these parameters, a cross-reference is created computationally between the histological quantities and the spectra which can be obtained at each point of a multispectral image set. In a pilot study the method was applied to two excised colon samples comprising normal tissue and adenocarcinomas. The histological parameters extracted from the cancerous regions showed the changes characteristic of the colon cancers, namely the increase in the blood volume fraction, decrease in the collagen density and increase in the thickness of the mucosal layer.

## 1. Introduction

Development of colon cancer alters the macroarchitecture of the colon tissue. The main changes include an increase in microvascularisation and hence the blood content of the tissue [1], and distortion of its collagen matrix [2]. Given that blood and collagen act, respectively, as strong absorbers and scatterers of light in the visible range of the spectrum, those changes alter the colon colouration. At the early stages of the disease the variations in the colour between normal and abnormal tissue tend to be very subtle, and hence not easily discernible by the human eye. Standard red, green, blue (RGB) primaries used in colour imaging provide a fairly impoverished representation of the tissue colouration, and it is likely that spectra arising from different tissue histologies may produce the same RGB values [3] – a phenomenon known as metamerism. This may result in a loss of diagnostically important information. Therefore, a richer spectral data set might be more appropriate for representing subtle changes in colouration. Multispectral imaging combines imaging and spectroscopy, and enables the extraction of the whole spectral content at every pixel of the image, so providing the combined advantages of high spatial and spectral resolutions.

Raw multispectral images are difficult to interpret directly. In remote sensing, which has a long history of using multispectral imagery, image regions are classified into various terrain types based on their spectral signatures. In this paper we present a different approach, where multispectral data is interpreted in terms of the underlying tissue histology. Such interpretation is possible through the use of a physics-based model of light interaction with colon tissue. The result is a set of parametric maps which show the spatial distribution, relative quantities of blood and the density and thickness of the collagen matrix in the topmost layer of the colon. We believe that, in the long term, the explicit presentation of these histological should lead to more reliable and possibly earlier diagnosis.

Our approach differs from traditional and more common approaches to image analysis which are based on statistical analysis of the image features. Instead of classifying tissues as normal or neoplastic, we are concerned with understanding the causes of the changes in tissue spectra and hence colour, and correlating those changes to the histological alterations of the tissue. At the heart of our method is a computational model of colon reflectance. Using the Monte Carlo method of simulating light-tissue interactions, the model correlates tissue parameters which describe its optical properties with the spectral composition of the light that can be observed at the tissue surface after penetrating the tissue and interacting with its structure. The spectra obtained at each point of a multispectral image are then interpreted by the inversion of the model.

This paper presents preliminary results of interpretation of multispectral images of the colon using the above approach. Section 2 presents a method of constructing a forward model of colon. A method of extracting histological parameters from the spectra, and its validation, are presented in section 3. Section 4 is concerned with image acquisition and post-processing of multispectral images in order to extract the spectral content at each pixel. Results, in the form of parametric maps, and their analysis are presented in section 5.

## 2. Forward modelling of colon reflectance: from tissue parameters to spectra

We have developed a computational, physics-based model of light interaction with tissue, which predicts reflectance spectra associated with specific instances of colon tissue. Here we give just a brief description of the model which is described in more detail in our previous paper [4].

The colon structure is represented by three layers: mucosa, submucosa and smooth muscle, which interact with the light incident on the surface of the colon. In all colon layers, the strongest absorbers of light are blood haemoglobins, while scattering is mainly due to collagen and subcellular organelles. These components are characterised by different properties and content in colon layers, which as a consequence have different optical properties, and cause different light-tissue interactions. Our analysis have shown that the parameters characterizing the first layer, the mucosa, have the most significant effect on behaviour of the remitted spectra, whereas the deeper layers determine the relative magnitude of the spectra and produce small changes in the red end of the spectrum [4]. Therefore the colon parametrisation includes variable parameters ( $N=5$ ) characterizing the mucosal layer only: blood volume fraction, haemoglobin saturation (the ratio of oxygenated and deoxygenated blood), the size of collagen fibres, their density and the layer thickness. The parameters characterizing the submucosa and the muscle have been assumed to have constant values, which have been set to the median of their relative ranges. Although this restricts the analysis to the mucosal tissue, it should be still of diagnostic value, as the early changes of the colon normally happen in this layer. For detailed description of parameters and the sources of the parameter values see [4].

Once the histological parameters and their normal ranges of values are defined, it is possible to create a cross-reference between these parameters and the reflectance spectra, and subsequently, a correspondence between the histology and images. The tissue reflectance is defined by  $N$  ( $=5$ ) variable parameters. Each *specific instance* of tissue can thus be defined by an  $N$ -dimensional parameter vector  $\mathbf{p} = [p_1 \dots p_n]$ . The range of each parameter is discretised to  $k_n$  levels, giving in total  $K = k_1 \times k_2 \times k_n$  parameter vectors which, together, define *all the possible instances* of the colon tissue. Each parameter vector  $\mathbf{p}$  can be associated with a spectrum by simulating light propagation in the tissue depicted by  $\mathbf{p}$  using one of the methods for approximating the solution to the light transport equation, such as for example Monte Carlo (MC) [5]. Subsequently, multispectral image values can be obtained from the computed spectra by convolving each spectrum with a set of appropriate bandpass filters,  $f_m$ ,  $m=1, \dots, M$ . Through such modelling of the light interaction with tissue and of the image acquisition process, we associate with each parameter a spectrum, and also an  $M$ -dimensional image vector  $\mathbf{i} = [i_1 \dots i_M]$ . All the parameter vectors  $\mathbf{p}$  together with the respective image vectors  $\mathbf{i}$  form the *tissue reflectance model*. This model is used to derive parameters from multispectral images of tissue (see section 4).

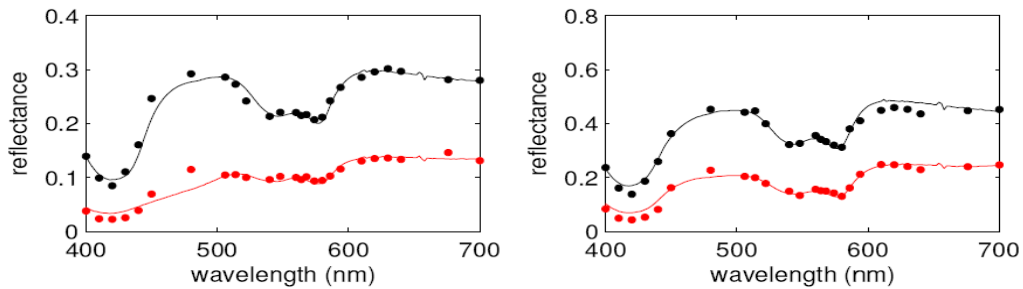
Reflectance spectra of the colon tissue were calculated for a discrete set of the ( $K=$ ) 20 wavelengths chosen to best describe the colon spectra: {450, 480, 506, 514, 522, 540, 548, 560, 564, 568, 574, 580, 586, 594, 610, 620, 630, 640, 676, 700}. Each spectrum was obtained by MC simulation using 200,000 photons. The modelled spectra were validated by comparison with the experimental data acquired *in vivo* using a fibre optics probe during colonoscopy procedures [6]. If the model is correct, it must be able to produce the spectra obtained by measurements (this, however, is just a necessary, but not also a sufficient condition). The validation data comprised 50 spectra from histologically confirmed non-neoplastic tissue (normal mucosa) and 7 spectra from cancerous tissue. Parameter values extracted from cancerous tissue were compared to the values extracted from normal colon tissue of the same patient. The method of parameter extraction and the validation results are described in the next section.

## 3. Extraction of parameters from spectral data

Extraction of the parameter values from the spectra was implemented as an optimisation procedure. The goal was to find a set of parameter values such that a spectrum generated from the parameters using the Monte Carlo simulation provides the best match to the measured spectrum. The criterion for the best match is the minimum distance between the two spectra defined as the mean absolute difference between the spectral values at corresponding wavelengths. The optimisation was implemented using an evolutionary strategy [7] which can be briefly summarised as follows. At each step, a set of parameters is chosen and the corresponding spectrum calculated. If the distance between the modelled and measured spectrum is smaller than some fixed threshold, the procedure is terminated. Otherwise, a new set of parameters is chosen according to a continuous version of the Gray-code neighbourhood distribution [7] and the process repeated.

This procedure was applied to the 57 experimentally obtained colon spectra described in section 2. In all the cases the optimisation procedure found a well matching spectrum in the reflectance model (figure 1). The matching error, measured as the mean absolute distance between the spectra, was  $0.01 \pm 0.0019$ . Parameters

extracted from the measured spectra were well within the histologically plausible ranges. Encouragingly, parameters extracted from the spectra of the abnormal tissue were characterised by increased blood content and decreased collagen density in comparison to the normal tissue of the same patient [8]. This is consistent with known histological differences between normal and abnormal colon tissue.



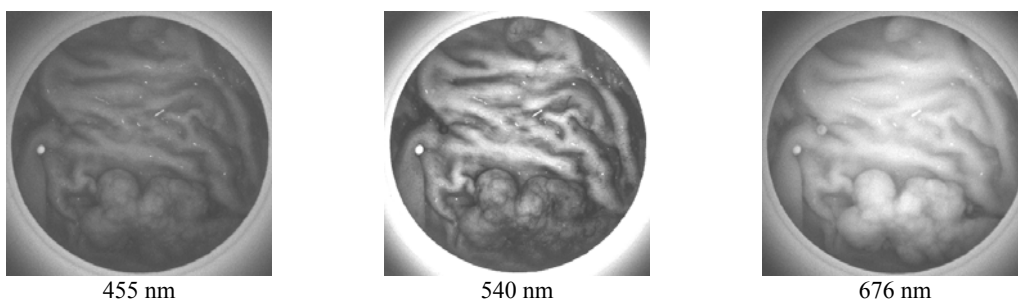
**Figure 1.** Model verification. Two examples of the measured spectra and the corresponding model spectra. Measured spectra are shown as solid lines, and modelled spectra as filled circles. Spectra of the normal tissue are shown in black, and of the cancerous tissue from the same patient are shown in red (grey in a monochrome reproduction).

#### 4. Extraction of parameters from multispectral images of the colon

Following validation on the spectral data, we have carried out a preliminary experiment on the data obtained from multispectral images of *ex-vivo* colon tissue. *Ex-vivo* samples were used as currently there is no established endoscopic technology for multispectral imaging of the intact colon. The objective of the experiment was to investigate whether the parametric maps obtained from the multispectral image data show the spatial distribution and the magnitudes of the individual histological parameters consistent with the histology of a given sample.

##### 4.1. Imaging

In a pilot study, images of two tissue samples were acquired within one hour of the surgical extraction performed at the Queen Elizabeth Hospital, Birmingham. In each sample, images were obtained for the apparently healthy tissue, at a distance more than 2 cm away from the histologically confirmed cancerous tissue (not shown), and the cancerous site itself (Fig. 2). The system used for acquiring the images consisted of a Retiga EXi (QImaging, Canada) 12bit monochrome camera, VariSpec (CRI, U.S.A.) liquid crystal tunable filters, and an Integrating Sphere (ProLite, U.K.) which produces a smooth spatial and spectral illumination. The VariSpec filters allow the selection of Gaussian-shaped filters of halfwidth 5 nm in the range from 400 to 700 nm. The images were acquired through a 3 inch aperture of the sphere. The total acquisition time per sample was under 5 seconds. An example set of images of the colon including cancerous tissue (adenocarcinoma) is shown in Figure 2. To ensure the spatial correspondence, the images in each set were registered.



**Figure 2.** Images of the cancerous colon tissue at wavelengths 455, 540 and 676 nm. Blood vessels and smaller features of the tissue can be seen in the image 540 nm where blood absorption is strong. At 455 nm the features are less distinct because high blood absorption coefficient causes even small amounts of blood to absorb strongly thus reducing differentiation between high and low blood levels. At 676 nm those features seem again to disappear because of low absorption of light by haemoglobin derivatives present in the blood and increased scatter by collagen, cells and organelles in the mucosal layer.

##### 4.2. Extraction of the spectra and construction of the parametric maps

Each pixel  $(x,y)$  in a multispectral image set can be represented by an image vector  $\mathbf{i} = [i_1 \dots i_M]$  where the  $m$ -th component of the vector is the image value  $i_m(x,y)$  corresponding to  $m$ -th spectral band. This value is a function of the tissue properties, the incident light and the characteristics of the image acquisition system:

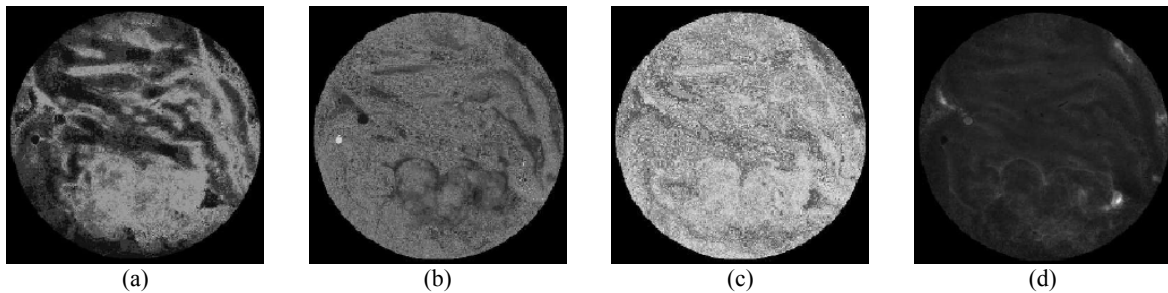
$$i_m(x,y) = \int_{\lambda} R(x,y,\lambda) \cdot I_0(x,y,\lambda) \cdot q(\lambda) \cdot f_m(\lambda) d\lambda \quad (1)$$

where  $R(x,y,\lambda)$  is tissue reflectance at the pixel  $(x,y)$  at the wavelength  $\lambda$ ,  $I_0(x,y,\lambda)$  is intensity of incident light at  $(x,y)$  and wavelength  $\lambda$ ,  $q(\lambda)$  is the quantum efficiency of the camera at  $\lambda$ , and  $f_m(\lambda)$  is the spectral response of the  $m$ -th filter at the wavelength  $\lambda$ . It can be seen that the image vector  $\mathbf{i}$  depends on the spectral reflectance of the tissue,  $R(x,y,\lambda)$ , as well as on the parameters characterising the imaging system. By deconvolving  $i_m(x,y)$  with  $I_0(x,y,\lambda)$ ,  $q(\lambda)$  and  $f_m(\lambda)$ , which are available as a function of wavelength from instrument calibration, it is possible to obtain the “pure” tissue reflectance spectrum at each point of the multispectral image.

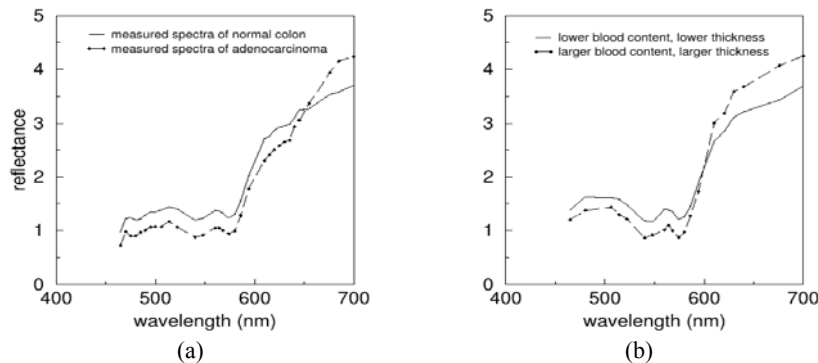
Using the parameter extraction method described in section 3 above, the histological parameters were derived from the tissue reflectance spectra  $R(x,y,\lambda)$  at each pixel, and their magnitudes were recorded in the five parametric maps representing the mucosal blood volume fraction, haemoglobin saturation, the size of collagen fibres, their density and the thickness of the mucosal layer. In addition to these parameters, two additional useful quantities were stored in the parametric maps: the scale factor, which is a function of the distance between the camera sensor and the imaged tissue (and reflects the surface topography); and the match error, which captures the degree of discrepancy between the modelled and the measured spectrum. This latter may have a diagnostic significance, as discussed below.

## 5. Results

Figure 3 shows three out of five parametric maps of a colon tissue sample shown in figure 2 (the other two maps did not show significant variations). A tumour, histologically confirmed as an adenocarcinoma, is clearly visible in the lower part of the image. The maps show the increased blood volume fraction, the decrease in the collagen density and increase in the mucosal thickness in the region of the cancer, as compared to the tissue further away from the tumour. This is in agreement with known histological changes which occur with development of the cancer. The error map shows a slightly increased error between the measured spectra and the spectra found in the reflectance model. This could indicate that the tissue in the cancerous region is distorted beyond the limits represented in the model, and as such, may provide an additional sign of abnormality.



**Figure 3.** Parametric maps derived from a multispectral image set illustrated in Fig. 2. The maps show (a) blood volume fraction; (b) collagen density; and (c) mucosal layer thickness (d) Error map. In all the maps brighter = a higher value.



**Figure 4.** Comparison of spectra of normal and cancerous colon tissue from figure 3. (a) The measured spectra: solid line – normal tissue; dashed line - adenocarcinoma. (b) Analogous spectral differences are visible in the modelled spectra if blood content and thickness are increased (dashed line) with respect to normal tissue (solid line).

A comparison of the spectra of cancerous and normal tissue is shown in figure 4. The main spectral variations seem to be due to larger blood content of the cancerous growth (corresponding spectrum is lower in blue and green bands), and larger mucosal thickness (the red end of the spectrum is higher). The spectra are consistent

with known histological variations that occur with development of malignancy. Analogous spectral variations were reproduced by our model, and shown in the plot in figure 4(b).

## 6. Conclusions and future work

The model correctly predicts the spectra, and thus colours, of colon tissue, indicating that histological quantities can be computed from multispectral images acquired by a digital camera equipped with suitable filters. The model validation [4] suggested that the model should be capable of differentiating between normal and abnormal tissue, which would potentially make it a useful non-invasive aid in the early detection of colon cancer. A preliminary study described in this paper has extended the parameter extraction method from the pure spectra [4] to the multispectral images of the colon. These very first results obtained for *ex-vivo* colon samples indicate that parametric maps of the blood volume fraction, the collagen density and thickness of the mucosa show the differences between the normal tissue and the adenocarcinoma which are consistent with the colon histology.

Whereas the forward model has been satisfactorily validated [4], we do not yet have such evidence for the parameter recovery method. The results described in this paper are promising, but they need to be supported by results from a much larger data set. This is the subject of the ongoing work. Work is also being undertaken in order to optimise the choice of filters to use in the acquisition of multi-spectral images of tissue. Earlier work by our group has shown that under certain conditions it is possible to recover parameter values relatively accurately from a small number of optimally selected spectral bands [9]. Four parameters characterising the mucosa (thickness, blood volume fraction, size, and density of collagen - see section 2) require four bandpass filters, making the image acquisition with the existing endoscope technology feasible.

The next major step will be the acquisition and the interpretation of the multispectral colon images *in-vivo*. The changes to the model will be trivial, as only the parameter ranges will need to be suitably altered to reflect the changes in histology. However, the recovery of the true tissue reflectance spectra ( $R(x,y,\lambda)$  in Eq. 1) under the unknown illumination conditions (magnitude of the incident light, its angular distribution, etc.) pose a major challenge. One possible approach has already been developed by our group [10], but it will need to be extended to take account of the colon movement and of the complex illumination in the closed space of the colon.

The ultimate goal of this research is to enable the clinician to obtain histological information non-invasively, during an *in vivo* endoscopic examination. This may prevent unnecessary biopsies in a growing population of patients undergoing colon cancer screening under a recently introduced NHS programme. It may also be possible to detect early cancerous changes before they become apparent through a standard endoscopic examination.

## Acknowledgements

Kevin Schomacker (MediSpectra, Inc.) is gratefully acknowledged for providing *in vivo* colon tissue spectra. We thank Dr Jon Rowe for his assistance with optimisation algorithms; and Dr Nigel Suggett and Dr Emma Hamilton (all from Birmingham University) for providing the tissue samples.

## References

1. S. A. Skinner, G. M. Frydman, P. E. O'Brien, "Microvascular structure of benign and malignant tumors of the colon in humans," *Digest. Dis. Sci.* **40**, 373-84, 1995.
2. Y. Furuya, T. Ogata, "Scanning electron microscopic study of the collagen networks of the normal mucosa, hyperplastic polyp, tubular adenoma and adenocarcinoma of the human large intestine," *Tohoku J. Exp. Med.* **169**, 1-19, 1993.
3. E. Angelopoulou, R. Molana, K. Danilidis, "Multispectral skin colour modelling," *IEEE Conference on Computer Vision and Pattern Recognition*, 2001, 635-642.
4. D. Hidovic, E. Claridge, "Modelling and validation of spectral reflectance for the colon". *Physics in Medicine and Biology* **50**, 1071-1093, 2005.
5. L. Wang S. L. Jacques, "Monte Carlo modelling of light transport in multi-layered tissues in standard c," *Univ of Texas, MD Anderson Cancer Center*, 1998.
6. Z. Ge, K. T. Schomacker, N. S. Nishioka, "Identification of colonic dysplasia and neoplasia by diffuse reflectance spectroscopy and pattern recognition techniques," *Applied Spectroscopy* **52**, 833-9, 1998.
7. J. E. Rowe, D. Hidovic, "An Evolution Strategy using a continuous version of the Gray-code neighbourhood distribution," in *Proc. GECCO 2004*, K. Deb, ed., LNCS **3102**, 725-736, Springer-Verlag, 2004.
8. D. Hidovic, E. Claridge, "Model based recovery of histological parameters from multi-spectral images of the colon". *Medical Imaging 2005: Physics of Medical Imaging*, Proceedings of SPIE Vol. 5745. Flynn MJ (Ed), 127-137, 2005.
9. E. Claridge, S. J. Preece, "An inverse method for the recovery of tissue parameters from colour images" *Information Processing in Medical Imaging*, C. Taylor and J. A. Noble, eds., LNCS **2732**, 306-317, Springer, 2003.
10. S.J. Preece, I.B. Styles, S. Cotton, E. Claridge, A. Calcagni. Model- based parameter recovery from uncalibrated optical images. *Medical Image Computing and Computer Assisted Intervention*. LNCS vol. 3750, 509-516, 2005.