

# Enhancing feature extraction from colon microscopy images using colourspace rotation

James Shuttleworth<sup>a\*</sup>, Alison Todman<sup>a</sup>, Raouf N. G. Naguib<sup>a</sup>,  
Robert Newman<sup>a</sup>, Mark Bennett<sup>b</sup>

<sup>a</sup>BIOCORE, School of Mathematical and Information Sciences, Coventry University, UK

<sup>b</sup>The Royal Victoria Infirmary, Newcastle-upon-Tyne, UK

**Abstract.** Severity of dysplasia is an important factor in the diagnosis of colorectal tumours. Currently, diagnosis is reached by visual examination of dysplasia, a time consuming and subjective process that is prone to both inter-observer and intra-observer variation.

Previously, we have shown that automated classification based on colour texture features and dual-objective colour texture features can greatly improve on grey-level texture analysis results.

In this paper, we discuss an abstract colour space developed specifically for colon microscopy images. By scaling and rotating the standard RGB (Red, Green, Blue) colour space, it has been possible to create a new model with two channels that can be used to extract textural features that have a greater correlation with classification.

## 1 Introduction

Worldwide, colorectal cancer is the third most common malignant neoplasm. In the UK, colon cancer is the second most common cancer related cause of death, and kills around 17,000 people annually, with approximately 34,000 new cases each year. After diagnosis, around 60% of patients die within 5 years [1].

With the continuing research into treatment of cancer, the likelihood of survival after diagnosis is increasing. As with most other types of cancer, early diagnosis of colon cancer can drastically increase the chances of successful treatment [1]. This has prompted schemes such as the National Health Service screening pilot scheme for earlier diagnosis of “at-risk” groups, such as the elderly. With the number of incidents of colon cancer steadily increasing and the large number of cases likely to be identified by such schemes, the volume of cases investigated by pathologists is almost certain to increase.

The classification of colon tissue samples is a particularly time-consuming task. Currently, each sample must be analysed individually by trained pathologists under a microscope to determine whether the tissue is normal or abnormal. In cases of abnormal tissue, it is graded according to severity by a trained pathologist. Research into automated classification of colon tissue samples using grey-level texture has yielded promising results [2–4] and yet diagnosis in all cases still requires human judgement, which is subject to inter-observer [5] and even intra-observer [6] variation.

Already, we have shown that using colour can increase the accuracy of a three-class texture based classification from 87%, using grey-level texture, to 96%, for regions of interest selected by a human [7]. We have also shown that by combining colour texture measurements at two scales, this can be increased to 98% [9].

When examination of colon tissue is carried out by pathologists, a finer system of grading is applied that has five possible classes. To be able to provide a useful system to pathologists, an automated system for the analysis of dysplasia must be able to classify based on such a set of grades, and be able to examine whole slides.

Using whole slide images is problematic because there is much less control over the content of the input images. Images may contain areas that have no tissue, or tissue that doesn’t clearly show a single classification. Even using colour texture at multiple scales, accuracy on such images in a five-class system falls to around 75% [8].

## 2 Colour

Previously, we have shown the value of colour [7], and scale [9] when using textural features to classify colon tissue based on dysplasia.

---

\*Contact: j.shuttleworth@coventry.ac.uk

Our earlier methods used a combination of features from the RGB and HSB (Hue, Saturation and Brightness) colour models. These models were chosen to represent the two types of colour model: mixed colour models, such as RGB and CMYK (Cyan, Magenta, Yellow, black) and chrominance/luminance models such as HSB or HCL (Hue, Chroma, Luminance). However, colour texture analysis can be affected quite severely by selecting an inappropriate colour model [10]. Unfortunately, we currently have no elegant way to determine the best colour model for a specific application, and so a system of trial and error is required to test each possible model in turn.

Rather than perform a comparative analysis of classification using textural features using each colour model, we have developed an abstract colour model by scaling and rotating the RGB model.

## 2.1 Image Acquisition

In total, 44  $5\mu\text{m}$  slices from colorectal biopsy tissue at  $\times 100$  magnification were used for the investigation. These samples were taken from routine individual biopsies and exhibit tissue in various stages of dysplastic progression, divided into normal tissue (no dysplasia), polyps (mild/ moderate dysplasia) and tumours (exhibiting severe dysplasia). Staining was performed using Haematoxylin and Eosin. The slides were digitised and classified by a qualified histopathologist with a specialism in gastro-intestinal cancers. The dimensions of the resulting images are  $768 \times 576$  pixels.

## 2.2 Colour space manipulation

Tissue in the colon has a number of components that are of interest to pathologists. These may be areas of tissue, such as the muscularis mucosae, empty areas (lumen) or parts of cells, such as the blueness signifying nuclear material. To simplify discussion, we will refer to these as subtypes.

For analysing colon tissue, an ideal colour model would provide the greatest separation between subtypes across the colour space. That is, for each pixel in the image plotted against the colour space axes, each subtype would be optimally clustered, with minimal intra-group variance and the greatest possible inter-group variance.

Discriminant analysis scales and rotates a feature vector to give exactly this optimal group separation.

Using a set of five images, a mask was prepared highlighting clear examples of the kinds of staining density characteristic of the most visually dissimilar subtypes. These were muscularis mucosae, lumen, nuclear material and full mucosal cells.

In total, this produced a set of over one million classified pixels. From this set, the first in every thousand pixels was used, giving a working set of 1385 cases, with the same distribution of subtypes as the original image.

Discriminant analysis was applied to this set to produce a set of discriminant functions that maximise the separation between groups. The functions produced are:

$$df_1 = 0.13r + 0.7g + 0.68b$$

$$df_2 = 0.63r + 0.7g + 0.43b$$

Where  $r, g, b$  are the red, green and blue values of a given pixel.

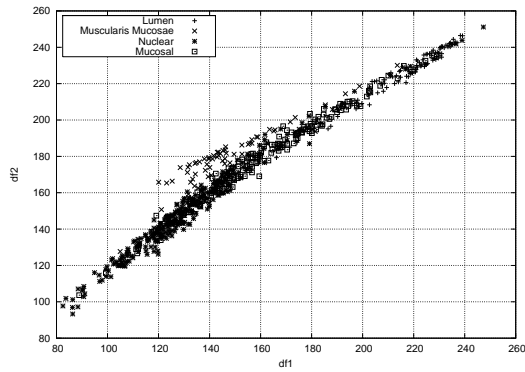
These functions distribute the highlighted pixels as shown in Figure 1.

There is a strong and unwanted correlation between the two functions produced by the equal weighting of  $g$  in  $df_1$  and  $df_2$ . This is not a problem while using discriminant analysis for classification, but for a colour space, we require the axes to be as orthogonal as possible. By subtracting one function from the other, we obtain:

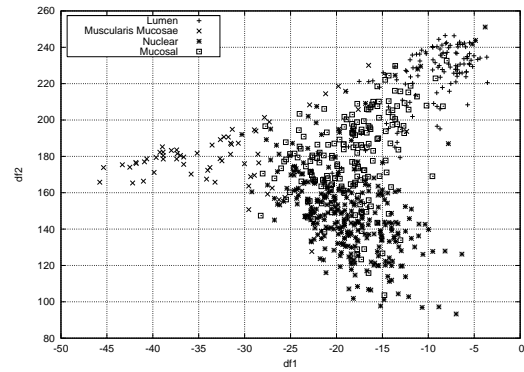
$$df'_1 = df_1 - df_2 = 0.25b - 0.5r$$

Replacing  $df_1$  with  $df'_1$ , the separation between groups appears much more clearly, as shown in Figure 2.

Further shifting and scaling is required to be able to express images in the new colour space in standard image formats - values must fit into eight bit integers. Finally, the axes of the new colour space are:



**Figure 1.** Functions produced by discriminant analysis



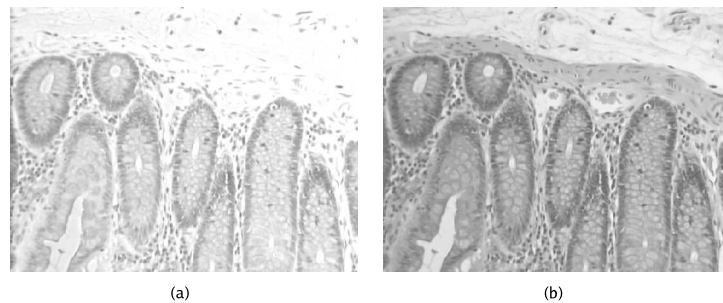
**Figure 2.** Modified functions

$$A = 1.330(0.25b + 0.00g - 0.50r)$$

$$B = 0.529(0.63r + 0.70g + 0.43b)$$

Discriminant analysis does not yield a scaling suitable for a colour model without manipulation, which is to be expected since this is not its purpose. While the output of discriminant analysis is not limited by range or orthogonality, colour models must be. That is, the scaling and further simplification does not rectify a shortcoming in discriminant analysis, but extends the technique for this purpose.

A real test of the efficacy of the new colour space requires an attempt at classification, discussed in the remainder of this paper, but there are other benefits to the new colour model that can be seen simply by looking at an example image. Desaturating colour colon slides usually makes it difficult to distinguish between neighbouring areas, such as in Figure 3a. The new colour model, however, provides much clearer desaturated images, Figure 3b.



**Figure 3.** (a) Desaturated RGB image (b) Desaturated AB image

### 3 Texture Analysis

To compare the new colour model with those used in previous experiments, the same process has been applied using the new AB colour model and the combined RGB/HSB models.

Co-occurrence matrix texture features [11] were extracted, of which the contrast of channel A and the mean and contrast of channel B (at 6, 1 and 6 pixel distances respectively) were found to be useful. As described in [7], co-occurrence matrices were calculated with 32 bins, which is equivalent to reducing the number of colours in the image to 32, making the matrix less sparse and more useful.

### 4 Classification

Using features extracted from all channels of the RGB and HSB colour models, the cross-validated classification accuracy was 65%, and required four of the extracted features.

Using the same features extracted from the A and B channels of the new colour model, accuracy increases to 68%, using only three features.

## 5 Discussion

Although the increase in accuracy is just 3%, cross-validation gives us confidence that it is a real increase and not a result of this data set.

Apart from the obvious benefit of any increase in accuracy reached by using a bespoke colour model, gains have also been made in the simplification of the process.

Firstly, experiments in colour texture analysis in colon tissue are simplified because of the reduced number of channels. Previously, it has been necessary to extract features from six channels to be able to select those that are useful discriminators, while now there are only two.

Secondly, the reduction in the number of features required to calculate the discriminating functions has a direct effect on the speed of processing of new images. In this case, using only three features instead of four results in a roughly 25% improvement in processing time per image.

Using a tailored colour model is also expected to improve more complex classification systems - systems with more classes, with features extracted at multiple scales. By using a colour model developed specifically for the problem domain, so that areas of tissue with different properties are clustered in the colour space, we expect to find that the features extracted for classification can be related to those used by pathologists during visual examination of colon tissue.

Further, we expect that it would be possible to apply the process of creating tailored colour models in other cases - in similar microscopy image analysis on other types of tissue, such as liver or oesophagus, or using different staining techniques, and possibly in many areas of image analysis. The ability to define a colour model that best expresses differences in image content has been shown here to be a useful tool, and it would be expected that the same technique could be successfully applied in many other areas.

## Acknowledgements

Funded by the Department of Health Access Directorate and Newcastle-Upon-Tyne Hospitals NHS Trust.

## References

1. M. Quinn, P. Babb, A. Brock et al. *Cancer trends in England and Wales 1950-1999*. Office for National Statistics, 2001.
2. A. N. Esgiar, R. N. G. Naguib, B. S. Sharif et al. "Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa." *IEEE Transactions on Information Technology in Biomedicine* **2(3)**, pp. 197-203, 1998.
3. P. W. Hamilton, D. C. Allen, P. C. H. Watt et al. "Classification of normal colorectal mucosa and adenocarcinoma by morphometry." *Histopathology* **11**, 1987.
4. P. W. Hamilton, P. H. Bartels, D. Thompson et al. "Automated location of dysplastic fields on colorectal histology using image texture analysis." *Journal of Pathology* **182**, pp. 68-75, 1997.
5. J. Eaden, K. Abrams, H. McKay et al. "Inter-observer variation between general and specialist gastrointestinal pathologists when grading dysplasia in ulcerative colitis." *Journal of Pathology* **194**, pp. 152-157, 2001.
6. D. Coppola & R. C. Karl. "Barrett's esophagus and barrett's-associated neoplasia: Etiology and pathologic features." *Cancer Control, Journal of the Moffit Cancer Center* **6(1)**, 1999.
7. J. K. Shuttleworth, A. G. Todman, R. N. G. Naguib et al. "Colour texture analysis using co-occurrence matrices for classification of colon cancer images." In *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering*. 2002.
8. J. K. Shuttleworth, A. G. Todman, R. N. G. Naguib et al. "Fine grading of colorectal biopsy images using colour texture analysis." In *Proceedings of the Medical Image Understanding and Analysis (MIUA) conference*. 2003.
9. J. K. Shuttleworth, A. G. Todman, R. N. G. Naguib et al. "Multiresolution colour texture analysis for classifying colon cancer images." In *Proceedings of the joint 4th Annual International Conference of the EMBS and Annual Fall Meeting of the BMES*. October 2002.
10. A. Drimbarean & P. F. Whelan. "Experiments in colour texture analysis." *Pattern Recognition Letters* **22**, pp. 1161-1167, 2001.
11. R. M. Haralick, K. Shanmugam & I. Dinstein. "Textural features for image classification." *TransSMC* **3(6)**, pp. 610-621, November 1973.