Context Aware Colour Classification in Digital Microscopy

Derek Magee[1]
D.R.Magee@leeds.ac.uk
Darren Treanor[2]
darrentreanor@nhs.net
Phattthanaphong Chomphuwiset [1]
scpc@comp.leeds.ac.uk
Philip Quirke[2]
P.Quirke@leeds.ac.uk

[1] School of Computing, University of Leeds, UK
[2] Pathology and Tumour Biology, Leeds Institute for Molecular Medicine, University of Leeds, Leeds, UK

Abstract

With the advent of digital histopathology imaging and automatic image analysis, colour constancy across multiple microscope slides has become an important issue. Colour variation due to chemical, user or protocol inconsistency is widespread. This paper presents an approach for computationally efficient context aware colour classification. A 'context vector' derived from the colour distribution of the complete image is combined with the per-pixel information to improve pixel classification performance. The context vector implicitly encodes global image information such as whether the slide is under/over stained, or cut thinly, or thickly. The method is evaluated for segmentation accuracy on two data sets with different stains, and as a pre-processing method for a cell nuclei detection algorithm.

1 Introduction

Histopathology is the diagnosis of disease by examination of tissue. In order to visualise tissue sections (which are virtually transparent), tissue sections are prepared using coloured histochemical stains that bind selectively to cellular components. Colour constancy is a problem in histopathology based on light microscopy due to: variable chemical colouring/reactivity from different manufacturers/batches of stains, colouring being dependent on staining procedure (timing, concentrations etc.), and light transmission being a function of section thickness. Lyon *et al.* [5] outline the need for standardisation of reagents and procedures in histological practice. However, such rigorous standardisation is not practised in the majority of hospital laboratories and complete standardisation is not possible without purer (and less variable) reagents (requiring action from multiple chemical manufacturers, and an associated increase in cost). Current practises are limited to physical and procedural quality control methods, including subjective assessment of stain quality and inter-laboratory comparisons of staining, in order to minimise the visible variability in staining and its impact on diagnostic quality.

With the advent of digital imaging and automatic image analysis colour consistency in histopathology has become more of an issue. For example, many commercial automatic image analysis algorithms require parameters defining the expected colour of anatomy of interest and fail if these parameters are incorrect. This paper presents methods for taking

into account the variation in staining (and other preparation inconsistencies) using 'context aware' colour classification. We define 'context aware' to be the inclusion of global (whole image) information in local (i.e. per-pixel) analysis. This is achieved using a low dimensional embedding of a colour histogram of the whole image, in combination with per-pixel representations.

2 Methods and Materials

2.1 Slide Preparation

Data used in experiments in this paper was acquired from our existing digital repository subject to relevant ethical permissions. Tissue was formalin fixed, paraffin embedded, cut in ≈5 micron sections using a microtome, and mounted on standard glass microscope slides. Samples were stained using Haematoxylin and Eosin counterstain (H+E), or Haematoxylin and DAB (H+DAB). Virtual slides were obtained by scanning at 20x or 40x magnification using an Aperio XT scanner (Aperio, San Diego). Sets of 60 1000×1000 representative sub-images at native resolution for each stain pair (12 images × 5 batches) were extracted from these gigapixel images via the ImageServer http interface (JPEG quality=100%).

2.2 Context Aware Colour Classification Using Low Dimensional Histogram Projections

Classification of individual pixels by colour into different classes relating to different stains, or background, is an important first step in many digital pathology applications. Per-pixel colour classification was performed based on a feature vector constructed from the RGB values at a given pixel concatenated with a whole image specific 'context vector'. Pixels were classified into 3 classes; 'Stain 1' (Haematoxylin), 'Stain 2' (Eosin, or DAB), and background (i.e. no stain). The context vector consisted of a low dimensional projection of a histogram of colour prototypes. The histogram was constructed using the highly computationally efficient Oct-tree based colour quantisation method of Gervautz [3]. This method works by iteratively partitioning a 3D colourspace into 8 equal sized regions to form a tree of regions of increasingly small size. In practise, a sparse tree is built containing only nodes relating to colours which occur in the training set. Leaves in this sparse tree are labelled with a unique 'prototype identifier'. The number of leaves can be reduced by subsuming multiple leaves of tree by their common parent node (which then becomes a leaf). In our implementation subsumption is based on the node with the fewest associated pixels until there are only 256 prototypes (a common heuristic and quantisation level often used in colour palette construction). This prototype histogram generation process is related to the widely used bag-of-words family of methods (e.g. [2]); the 'words' in this case being colour prototypes. The process of feature generation for a given pixel is illustrated in figure 1. The low dimensional projection of each image histogram (H_n) is performed using Principal Components Analysis (PCA) on a training set of histograms from different images (equation 1).

$$\hat{H}_n = E_h^T (H_n - \bar{H}) \tag{1}$$

Using equation 1, the histogram of any image (whether in the training set or not) can then be represented as it's projection in a truncated Eigenspace ($\hat{H}_n$), where E_h is the truncated eigenvector matrix, and $\bar{H}$ is the mean of, the training histogram set. Dimensionality reduction of the context vector is necessary to avoid the (potentially high dimensional) context vector dominating the 3 dimensional colour vector in classification. An added advantage is the computational saving of performing classification using a low-dimensional feature. Use of a class unaware dimensionality reduction method (i.e. PCA), rather than a class aware

dimensionality method (e.g. Linear Discriminant Analysis - LDA), allows projections to be performed once per image, rather that once per pixel, as a single image histogram typically represents multiple pixel classes. There is an obvious computational saving in this.

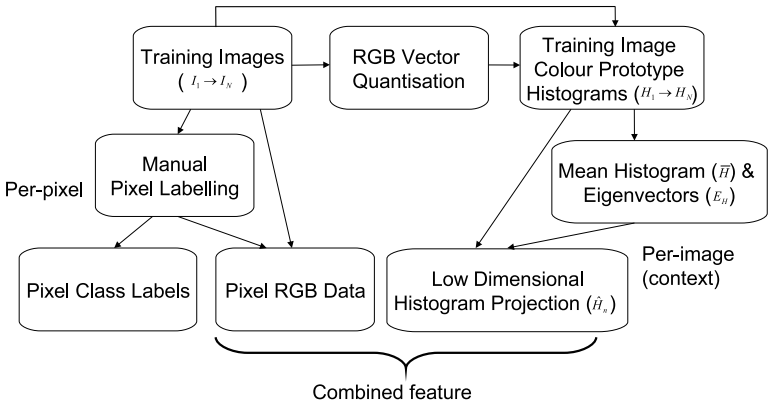

Figure 1: Overview of Pixel Feature Generation in Training Stage: A per-pixel RGB colour value is combined with a per -image colour histogram projection in eigenspace. Manual labelling of pixels (using an interactive tool) assigns class labels to particular pixels in the image, which form the training set. Test set generation follows the same protocol, except that the Oct-tree and Eigenvectors estimated at training time are used.

Classification was performed using supervised learning by the Relevance Vector Machine (RVM) method [7, 8] (*http://dclib.sourceforge.net/*). This method can classify a 1000x1000 image in a fraction of a second, whereas a Support Vector Machine trained on the same data takes many 10s of seconds. Additionally, the RVM provides a probabilistic (rather than binary) output. As we have a 3-class problem, and the RVM is a 2-class classifier, classification was implemented using the 'one against all' approach [4], training 3 RVM models per data set and normalising as in equation 2.

$$P(Class_n|C) = \frac{P_{Class_n}(Class_n|C)}{P_{s1}(s1|C) + P_{s2}(s2|C) + P_{bgd}(bgd|C)} \tag{2}$$

Where $Class_n \in \{s1, s2, bgd\}$, $P_{Class_n}(Class_n|C)$ is the probabilistic output of the RVM model trained with pixels from $Class_n$ as positive examples and pixels from the other two classes as negative examples, and C is the combined feature vector (figure 1). Other classifiers (e.g. Random Forests [1]) were also evaluated, as were alternative colour representations (e.g. LAB). The method was evaluated for per-pixel colour classification accuracy, and as a method for estimating image specific colour deconvolution matrices as a pre-processing method prior to nuclei detection (following section).

2.2.1 Image specific colour deconvolution for Nuclei Detection

Colour deconvolution (C.D.) [6] is a method for 'unmixing' different stains in RGB microscopy images. In this work we extract an image relating to the Haematoxylin stain (a nuclear stain) as a precursor to Nuclei detection with a Hough Transform based method [Self citation]. C.D. requires a representation of the absorption factors of the mixed stains (a 'colour deconvolution matrix'), the accuracy of which effects the accuracy of the unmixing. C.D. is based on a subtractive model of image formation (equation 3).

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 255 \times \prod_{n=1}^{3} e^{-A_n c_{r,n}} \\ 255 \times \prod_{n=1}^{3} e^{-A_n c_{g,n}} \\ 255 \times \prod_{n=1}^{3} e^{-A_n c_{b,n}} \end{bmatrix} \tag{3}$$

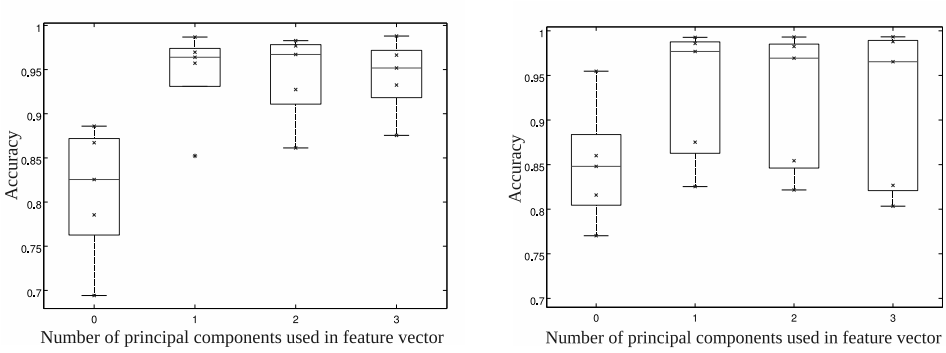

Figure 2: Boxplots of 5-fold cross validation of colour+context based segmentation with different context vector dimensionalities. Left: H+E Data, Right: H+DAB data

A_n relates to the amount of stain n at that pixel, and $c_{r|g|b,n}$ defines the 'colour deconvolution matrix' (absorption factors) for a particular stain of interest. $c_{r|g|b,n}$ may be calculated from an example set of colours for each stain. The mean colour of regions of interest (one for each stain n) was used in this work $(\bar{r}_n, \bar{g}_n, \bar{b}_n)$ (Equation 4).

$$\begin{bmatrix} c_{r,n} \\ c_{g,n} \\ c_{b,n} \end{bmatrix} = \begin{bmatrix} -log((\bar{r}_n+1)/256) \\ -log((\bar{g}_n+1)/256) \\ -log((\bar{b}_n+1)/256) \end{bmatrix} / \left| \begin{bmatrix} -log((\bar{r}_n+1)/256) \\ -log((\bar{g}_n+1)/256) \\ -log((\bar{b}_n+1)/256) \end{bmatrix} \right| \tag{4}$$

A complement (cross product of $[c_{r,1}, c_{g,1}, c_{b,1}]$ and $[c_{r,2}, c_{g,2}, c_{b,2}]$) is used to complete the (3×3) colour deconvolution matrix if only 2 stains are used (as in the experiments in this paper). We estimate example colours for each channel using the colour classification method presented in section 2.2 to define pixels of interest for each stain. Pixels assigned a probability greater than some threshold (0.99 is used in experiments) are used to calculate the example colours for each channel. Nuclei detection is applied to the resultant Haematoxylin image, and to greyscale, and probability images for comparison.

3 Results and Conclusions

Figure 2 presents the accuracy of per-pixel colour classification with, and without, our context vector. A clear (statistically significant) increase in accuracy is observed when using context. It is interesting to observe that most of the context seems to be encoded in the first context dimension and there is little increase in accuracy when using >1D context. Similar trends were observed when using a Random Forest classifier and other colourspaces, although the RGB+RVM combination performed best (Random Forests tended to over-fit).

Figure 3 presents precision-recall curves for nuclei detection using a Hough transform based nuclei detector (with varying edge detector threshold) applied to a greyscale version of images, and pre-processed images representing the Haematoxylin stain channel extracted in various ways. The first observation is that methods based on Haematoxylin channel extraction result in higher precision over using a greyscale image over a range of recall values. Image specific deconvolution (based on our classifier) outperforms all other methods at high recall values. Deconvolution with 'standard vectors' (H=[0.644,0.717,0.267],E=[0.093,0.954,0.283]) supplied with standard implementations of [6] performs well at low recall (i.e. there are few false positives), but results in a significantly lower maximum recall than other methods as weakly stained nuclei are not represented in this image. Image specific colour deconvolution

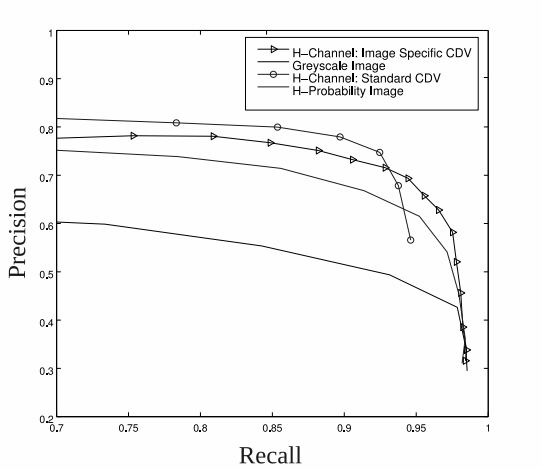

Figure 3: Average Precision-recall curves for Hough based nuclei detection over 8 Haematoxylin and Eosin stained liver images (2059 nuclei) with different pre-processing methods.

seems to be a better approach to extracting a stain-specific image than using a probability image in this context. This is unsurprising as this representation allows us to represent the degree of staining, whereas the probability image is simply a probability that the dominant stain at that pixel is Haematoxylin (independent of degree of staining).

In conclusion, the method presented demonstrates that inclusion of whole image context information (a low-dimensional embedding of a colour histogram) can aid per-pixel colour based classification/segmentation. We have demonstrated the utility of this to colour deconvolution matrix estimation, and that this can improve the performance of nuclei detection. We continue to evaluate new applications of these ideas.

References

[1] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[2] C. Dance, J. Willamowski, L. Fan, C. Braya, and G. Csurka. In *ECCV International Workshop on Statistical Learning in Computer Vision*, page Visual categorization with bags of keypoints, 2004.

[3] Michael Gervautz and Werner Purgathofer. A simple method for color quantization: Octree quantization. In *New Trends in Computer Graphics*. Springer Verlag, Berlin, 1988.

[4] C-W. Hsu and C-J. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002.

[5] H. Lyon, A. De Leenheer, and et. al. Standardization of reagents and methods usid in cytological and histological practice with emphasis on dyes, stains and chromogenic reagents. *Histochemical Journal*, 26:533–544, 1994.

[6] A. Ruifrok and D. Johnston. Quantification of histochemical staining by color deconvolution. *Analytical & Quantitative Cytology & Histology*, 23:291–299, 2001.

[7] M. Tipping. Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

[8] M. Tipping and A. Faul. Fast marginal likelihood maximisation for sparse Bayesian models. In *Proc. International Workshop on Artificial Intelligence and Statistics*, 2003.