A Novel Approach for the Colour Deconvolution of Multiple Histological Stains

Duane Carey[1]
Dr Nagitha Wijayathunga[1]
Dr Andrew Bulpitt[1]
Dr Darren Treanor[2&3]

[1] Leeds School of Computing Leeds University
[2] Leeds Teaching Hospitals NHS Trust, Leeds
[3] Leeds Institute of Cancer and Pathology, Leeds

Abstract

Colour Deconvolution (CD) is a commonly used tool in histological medical image analysis that separates histochemical or immunohistochemical stains into their component parts. Traditional CD uses matrix inversion to change the Red, Green and Blue (RGB) channels of an image into a new domain that is representative of reference colours but this limits the total number of stains that can be separated.

This is problematic for histological staining protocols that use more than three stains, such as FAST staining. This limitation has restricted the use of multicolour staining in light microscopy. To address this issue, this paper evaluates the use of Non Negative Matrix Factorisation (NNMF) and Non Negative Least Squares (NNLS) to enable the decomposition of multistained histological sections into its source components. It will be shown that NNLS is better suited to imaging modalities such as Whole Slide Image (WSI) scanners and that the multiple staining metrics produced from a single sample are analogous to those generated by applying single reagents to contiguous tissue sections.

1 Introduction

Histology samples are routinely stained with reagents that enable the tissue structure and or function to be visualised. This plays an integral role within current diagnostic practices and can also aid biomedical engineers and scientists in their understanding of tissue samples.

The properties of the tissue can be derived from analysing the spatial and spectral relationship of the stains and this can automatically be facilitated by Colour Deconvolution (CD) with thresholding [1] . CD is a simple methodology that changes the bases of an image using matrix inversion. However, this methodology is limited to the number of dimensions that are associated with an image and so in the case of standard optical microscopy no more than three stains can be separated. Therefore, new specialised staining regimes, such as FAST **(Figure 1)** [2], which involves the application of more than three stains cannot be analysed in this way. Other approaches, such as multi-channel fluorescence imaging, provide a solution, but often lose the valuable morphological information present in light microscope images.

Solutions to this problem include Non Negative Matrix Factorisation (NNMF) [3] which estimates both the source staining colours and separates pixels into their constituent components. However, a methodology of this type is not suitable when considering Whole Slide Imaging (WSI) slides because of their large size. This imaging modality sequentially acquires high resolution image tiles of a tissue sample and stitches these back together to produce a final image which contains many hundreds of millions of pixels. Since NNMF is subject to empirical estimates the size of the data used in this paper precludes its use.

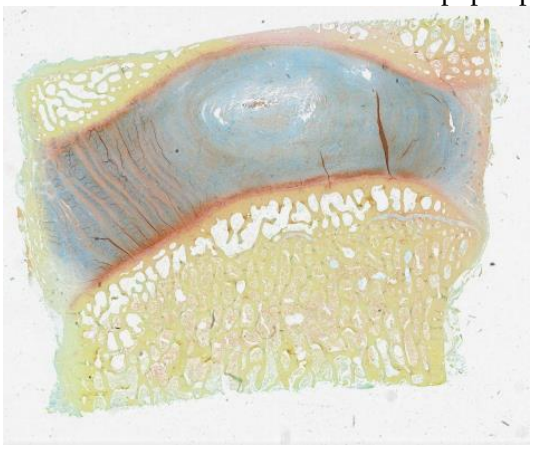

Figure 1: A FAST stained section of a bovine intervertebral disc.

Therefore, this paper presents a simple solution in the form of Non Negative Least Squares (NNLS) [4]. It will be demonstrated that since the number of stains applied to tissue samples is known beforehand that a colour gamut (look up table) can be used to enable the fast and efficient decomposition of digital microscope images. A comparison will be made against NNMF using a similar methodology and it will be qualitatively shown that this papers method produces a more informative deconvolution.

2 Methods

2.1 Datasets

Paraffin embedded bovine intervertebral disc tissue samples were serially sectioned and alternatively stained with the FAST staining protocol and single reagents designed to highlight similar molecular components. In total 408 samples were analysed by this paper. The samples were imaged using an Aperio T2 Virtual Slide Scanner and were acquired using a 20x magnification with a pixel resolution of $0.49\mu m^2$. In this paper, all analysis was conducted on the 20x magnification images. The size of these images was in between the range of around 40,000 and 80,000 pixels in both the images x and y dimensions. This work was funded through WELMEC, a Centre of Excellence in Medical Engineering funded by the Wellcome Trust and EPSRC, under grant number WT 088908/Z/09/Z.

2.2 Colour Deconvolution

Colour Deconvolution (CD) is a simple methodology [1] which changes the bases of histological images into those that are representative of reference stains by matrix inversion, **Equation 1**.

$$\mathbf{S} = \mathbf{M}^{-1}\mathbf{I}_{\mathbf{OD}} \quad (1)$$

Where **S** represents the deconvolved staining values, **M** is the mixing matrix (three by three) composed of reference staining colours and $\mathbf{I}_{\mathbf{OD}}$ is the Optical Density converted image of interest. The OD, log, colour space is used as there is a nonlinear relationship between the observed Red, Blue and Green (RGB) values of an image and staining intensity (Beer Lamberts Law) [1]. In this instance, the reference staining colours used with this method were generated by Singular Value Decomposition (SVD) [5].

2.3 Non Negative Matrix Factorisation (NNMF)

Non Negative Matrix Factorisation (NNMF) [6] factorises a matrix, **V**, into two nonnegative matrices **W** and **H**, **Equation 2**.

$$\mathbf{V} = \mathbf{WH} \quad (2)$$

Where **V** is the multistained image and **W** and **H** represent the mixing and the source contribution matrices. It is commonly referred to as a Blind Source Separation (BSS) methodology as it can simultaneously estimate the number of components that make up a signal and provide estimates on their contribution to the observation. **W** and **H** are estimated via alternately minimising a cost function of the form presented in **Equation 3**.

$$D(\mathbf{V}||\mathbf{WH}) = \frac{1}{2}\|\mathbf{V} - \mathbf{WH}\|^2 \quad (3)$$

In this instance, a multilayer Hierarchal NNMF algorithm (HALS NNMF) [6] was used whereby the output from previous layers was used to construct the source contribution matrix for the next layer. To initiate the algorithm a mixing matrix was randomly assigned with dimensions of three, one for each RGB channel, by the number of desired signals, four in this instance as FAST is comprised of four staining colours. Further experiments that used reference staining colours as an initialisation, provided from an expert, were also trialled but the end result was the same.

However, it was not feasible to apply NNMF directly to a digital microscope image because of the vast quantities of data that are associated with this modality. Therefore, the unique RGB triplets associated with a digital microscope images were collected and converted to the OD colour space before use with the multilayer HALS NNMF algorithm.

2.4 Non Negative Least Squares (NNLS)

Similar to NNMF, Non Negative Least Squares (NNLS) [4] seeks to minimise the same function presented in **Equation 3** and again results in non negative source contribution estimates. However, instead of estimating the mixing matrix of source components, this algorithm needs it to be explicitly defined.

In this instance, this was achieved via manual assessment of the acquired virtual slide. Representative colours of stains of interest were selected, converted to the OD space and used to form a three by four mixing matrix. The method by Mackenco et al. could not be used with NNLS as the FAST images of interest contained more objects than it had spectral dimensions. For this investigation, the active set NNLS method was applied to the colour gamut of the RGB colour cube in Matlab.

This resulted in a 256*256*256 vector which was representative of the contribution that the mixing matrix made to all possible RGB triplets. Since this papers tissue sample

was serially sectioned the lookup table could be used for all sections and this ensured consistent deconvolution and enabled a parallelised architecture to be implemented.

2.5 Validation

The validation of this papers methodology is difficult since it is impossible for a human to exactly discern the individual staining components of the pixels that make up a digital microscope image. However, since the serially stained sections also contained single stained equivalents to the FAST solutions, the staining metrics acquired from contiguous sections of different stains that highlight similar molecular components should be equivalent.

In this instance, traditional colour deconvolution was used to analyse the single stained slides and area metrics were derived by thresholding the staining channel of interest using Otsu's method. The same or similar staining channel to that of the single reagent in the FAST deconvolved image channels was also thresholded using Otsu's method and any correlation that existed between the two was quantified via r values.

3 Results and Discussion

In the first instance, Non Negative Matrix Factorisation (NNMF) was used with the unique Red, Blue and Green (RGB) colour gamut triplets generated from a digital microscope image. In addition to this, a single region of interest from the digital microscope image was selected and the source components estimated by NNMF were applied to it by consulting the deconvolved colour gamut. The result of this is displayed in **Figure 2** and it is evident that estimating the source signals in this fashion with NNMF is ineffective. This can be attributed to the use of the colour gamut, which is required for faster processing, as this causes a reduction in the colour space provided to the NNMF algorithm and has an effect on the source contribution estimation.

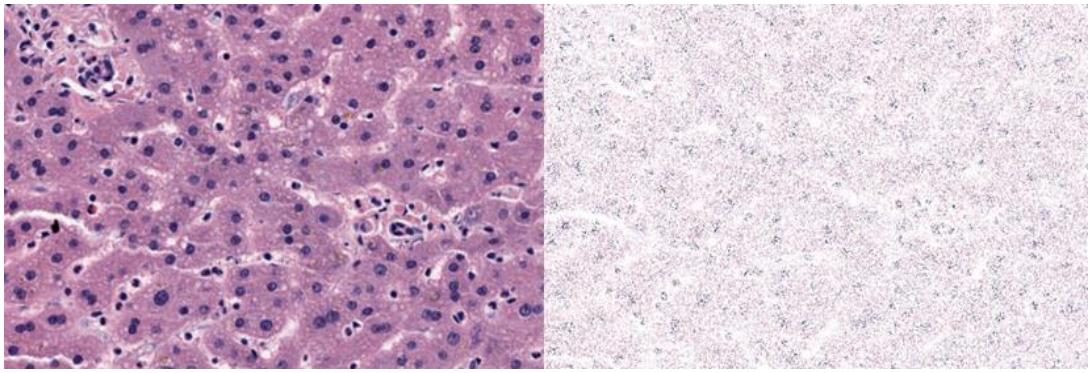

Figure 2: The result of using NNMF to estimate Haematoxylin and Eosin (H&E) source contributions for pixels from a region of interest in a Virtual Scanned slide of normal human liver. Left) The original image. Right) The recombined RGB image after the source estimates from NNMF had been applied.

The erroneous source estimation in **Figure 2** is reflective of the fact that only the unique RGB triplets of the digital microscope images were used with the multilayer HALS-NNMF algorithm. The source staining colours would have been weakly represented in this and so the chance of the algorithm finding the correct mixing matrix was minimised. It should be noted that when NNMF was conducted just on the region of

interest, the left of **Figure 2**, the source contribution estimation produced reasonable results, i.e. similar pixel staining relationships were evident.

However, this is not appropriate as the analysis of virtual slides on a per tile basis is complicated because there will be inconsistent source contributions estimated for each tile. Therefore, Non Negative Least Squares (NNLS), which is a generalisation of NNMF, was used as an alternative. In this algorithm, a mixing matrix must be defined beforehand and this was facilitated by an expert manually selecting regions representative of specific FAST stains. To enable efficient deconvolution of the FAST digital microscope image slides, the selected reference colours were used to construct a 256*256*256 lookup table for faster computation times.

To validate this methodology, comparisons were made against contiguous sections that were stained with single reagents that identified similar biomolecules to FAST. FAST is comprised of four stains that identified glycoproteins via Alcian Blue (AB); glycosaminoglycans via Safranin-O; histones via Fast Green and mucin associated extracellular matrix proteins by Tartrazine. For this research, contiguous single reagent slides were made using AB and Millers Elastin (ME) for comparative purposes and these highlighted similar features to the FAST staining regime. The FAST AB is the same as standard AB and ME identifies similar features to Tartrazine. Therefore, it should be expected that the spatial distribution between these stains should be correlated as they are essentially highlighting the same molecular components.

This is evident from **Figure 3** where it can be seen that there is a strong positive relationship between the different stains area metrics. The r, correlation, value obtained for these plots was 0.81 and 0.79 for AB and Tartrazine respectively. The outliers in this data can be explained by the damage which occurred to serial sections during processing.

4 Conclusion

This paper presents a methodology for deconvolving histology images that are stained with multiple components. In the first instance, Non Negative Matrix Factorisation (NNMF) was used in conjunction with the unique Red, Blue and Green (RGB) colour gamut triplets generated from a virtual slide. The resultant source contribution estimates were found not to be satisfactory and this was attributed to the dilution of the pixels that were representative of staining colours of interest.

Therefore, Non Negative Least Squares (NNLS) was used as an alternative. It was found to be better suited to the task of multi stain colour deconvolution as it enabled consistency between sections and facilitated the faster processing of digital microscope via the use of a lookup table.

To validate this approach, metrics obtained from contiguous single stained sections to those of FAST were correlated and r values indicated a highly positive relationship, **Figure 3**. However, when considering digital microscope images the fact that there is an associated variation in recorded staining intensity of between ten and twenty percent and that serial sectioned samples will have varying thickness means that these results could be biased. However, the large sample size used in this paper, 408, gives credence to the use of this method in analysing multi stained tissue samples as the metrics derived from these sections are comparable to their single stained counterparts.

References

[1] A. Ruifrok and D. Johnston, “Quantification of Histochemical Staining by Color Deconvolution.,” Anal Quant Cytol Histol., vol. 4, pp. 291–299, 2001.

[2] V. Y. L. Leung, W. C. W. Chan, S.-C. Hung, K. M. C. Cheung, and D. Chan, “Matrix Remodeling during Intervertebral Disc Growth and Degeneration Detected by Multichromatic FAST Staining.,” J. Histochem. Cytochem., vol. 57, no. 3, pp. 249–56, Mar. 2009.

[3] I. Kopriva and A. Cichocki, “Blind Decomposition of Low-Dimensional Multi-Spectral Image by Sparse Component Analysis,” J. Chemom., vol. 23, no. 11, pp. 590–597, Nov. 2009.

[4] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems. 1974.

[5] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, “A method for normalizing histology slides for quantitative analysis,” in Biomedical Imaging: From Nano to Macro, 2009, pp. 1107–1110.

[6] A. Cichocki and R. Zdunek, “NMFLAB - MATLAB Toolbox for Non-Negative Matrix Factorisation.” .

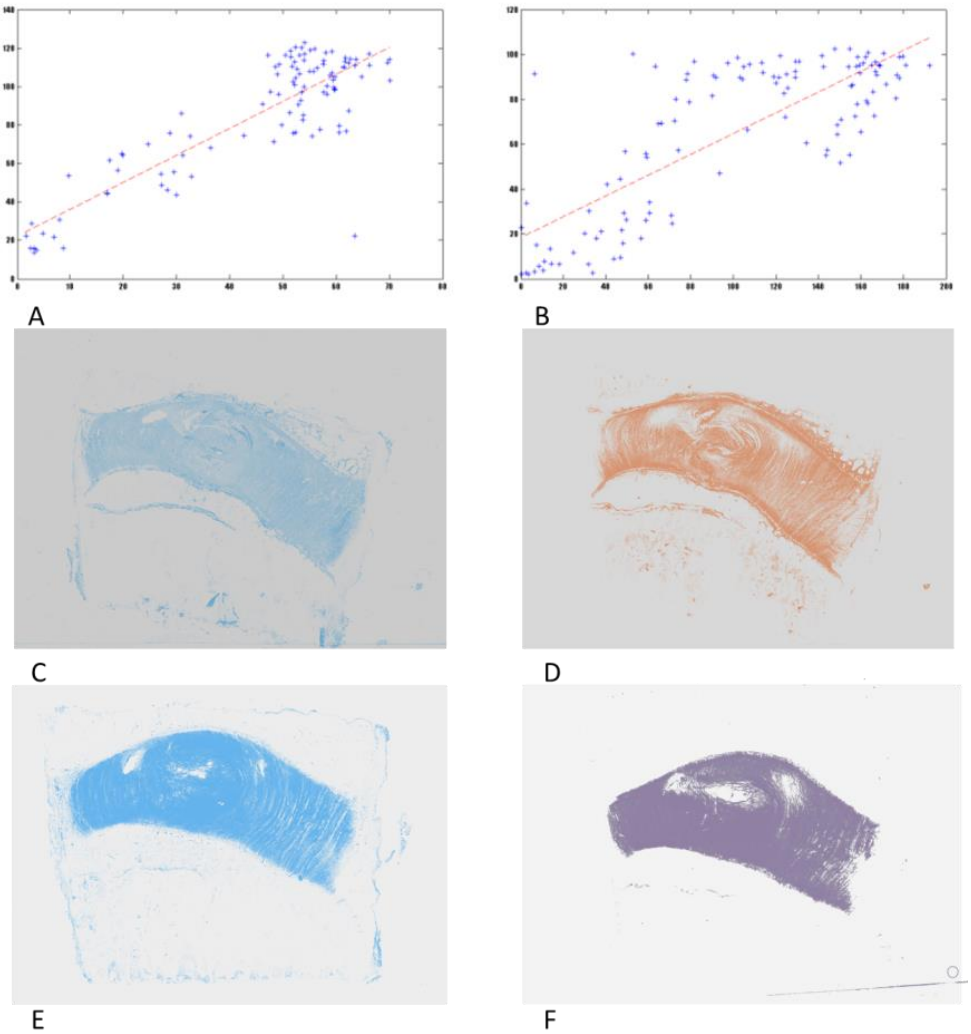

Figure 3. The correlation that exists between coloured deconvolved sections stained with a single reagents and multistained sections processed with Non Negative Least Squares (NNLS). A) The correlation between FAST AB and standard AB. B) The correlation between Tartrazine and Millers Elastin. C) A visual representation of the AB deconvolved from a FAST stained virtual slide. D)

Tartrazine deconvolved from FAST. E) AB channel from single staining with AB. F) Millers Elastic stain generated by traditional colour deconvolution. It should be noted that all processing was conducted on the original 20x magnification images and that C-F are heat map visualisations of the larger analysis.