Colour Deconvolution: stain unmixing in histological imaging

(Article begins on next page)

Colour Deconvolution – stain unmixing in histological imaging

Gabriel Landini [1,*], Giovanni Martinelli [2], Filippo Piccinini [2]

1 School of Dentistry, Institute of Clinical Sciences, University of Birmingham, B5 7EG Birmingham, UK; 2 Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS, 47014 Meldola (FC), Italy.
* To whom correspondence should be addressed.

Abstract

Motivation: Microscopy images of stained cells and tissues play a central role in most biomedical experiments and routine histopathology. Storing colour histological images digitally opens the possibility to process numerically colour distribution and intensity to extract quantitative data. Among those numerical procedures is colour deconvolution, which enables decomposing an RGB image into channels representing the optical absorbance and transmittance of the dyes when their RGB representation is known. Consequently, a range of new applications become possible for morphological and histochemical segmentation, automated marker localisation and image enhancement.
Availability and implementation: Colour deconvolution is presented here in two open-source forms: a MATLAB program/function and an ImageJ plugin written in Java. Both versions run in Windows, Macintosh, and UNIX-based systems under the respective platforms. Source code and further documentation are available at: https://blog.bham.ac.uk/intellimic/g-landini-software/colour-deconvolution-2/
Contact: Gabriel Landini, corresponding author, e-mail: G.Landini@bham.ac.uk
Supplementary information: Supplementary data are available at Bioinformatics online.

1. Introduction

Quantitative histological imaging is a relatively new area of research combining microscopy and computational techniques aimed at extracting information from biomedical and diagnostic pathology specimens. The growing availability of image digitisation and virtual slide technologies has driven the search for new ways of algorithmically analyse large image data sets. Numerous experiments rely on visualisation of results using histological staining techniques, due to the ability of light-absorbing dyes to selectively bind to molecules and complexes of interest. Commonly, 'counter stains' are added to reveal further structural morphology of otherwise 'invisible' components of specimens and such dye combinations appear to the human eye as subtractively-mixed colours. Since colour images can be conveniently stored digitally as additive combinations of red, green and blue (RGB format), it is possible to extract and process colour information numerically and reproducibly (*e.g.* distribution, hue, saturation, texture, *etc.*) rather than assess it on a visual, subjective basis. Among the numerical procedures available is *colour deconvolution*, introduced by Ruifrok and Johnston (2001) which, based on the Beer-Lambert law, it allows unmixing brightfield images into up to three 'channels' representing the absorbance of the individual dyes, when the RGB components of each dye are known (**Supplementary Note 1**). This has found several applications for non-destructive unmixing of pigments or inks in document analysis, restoration and forensics, but it is has become of particular interest in histological imaging because brightfield histological stainings mix

subtractively and dyes commonly colocalise (i.e. tissue components uptake more than one dye). Therefore colour deconvolution can resolve to what extent structures are stained by one or another dye, which is something that cannot be achieved by simply processing colour commonly used space channels like RGB, Lab or HSB. In addition unmixing digital RGB images has obvious practical advantages over methods for spectral decomposition that require more complicated image capture through sets of narrow band interference filters (Zhou *et al.*, 1996).

2. Implementation and functionality

We present two open-source implementations of colour deconvolution: a MATLAB (The MathWorks, Inc., USA) function and an ImageJ (Rasband, 1997) Java plugin, Colour Deconvolution 2 (CD2) which run on any computer operating systems supporting MATLAB or ImageJ platforms.

The input expected by the plugin is a background-corrected, brightfield image or stack in 24bit RGB format of histo/cytological specimens stained with light absorbing dyes (which are assumed to have negligible light scattering). Obviously other images such (prints, paintings, transparencies) can also be processed as long as they were generated with up to three transparent pigments, inks or watercolours. The user must provide a three-colour vector matrix *M* representing the R, G, B absorbances in saturated regions of each dye on their own. The vector values can be input in three ways: *a*) through a macro-recordable dialogue ('*User values*' option); *b*) by determining interactively regions of interest (option '*From ROIs*') from collages of images stained with a single dye at a time (**Supplementary Note 2**) or *c*) from a list of pre-computed vectors for common staining methods. In the case of two-dye stains, the non-existent third vector is defined computationally. Once the vectors are input, the program normalises the matrix *M* and inverts it into a translation matrix which is used to compute the contribution of each dye based on the R, G and B values of the image pixels.
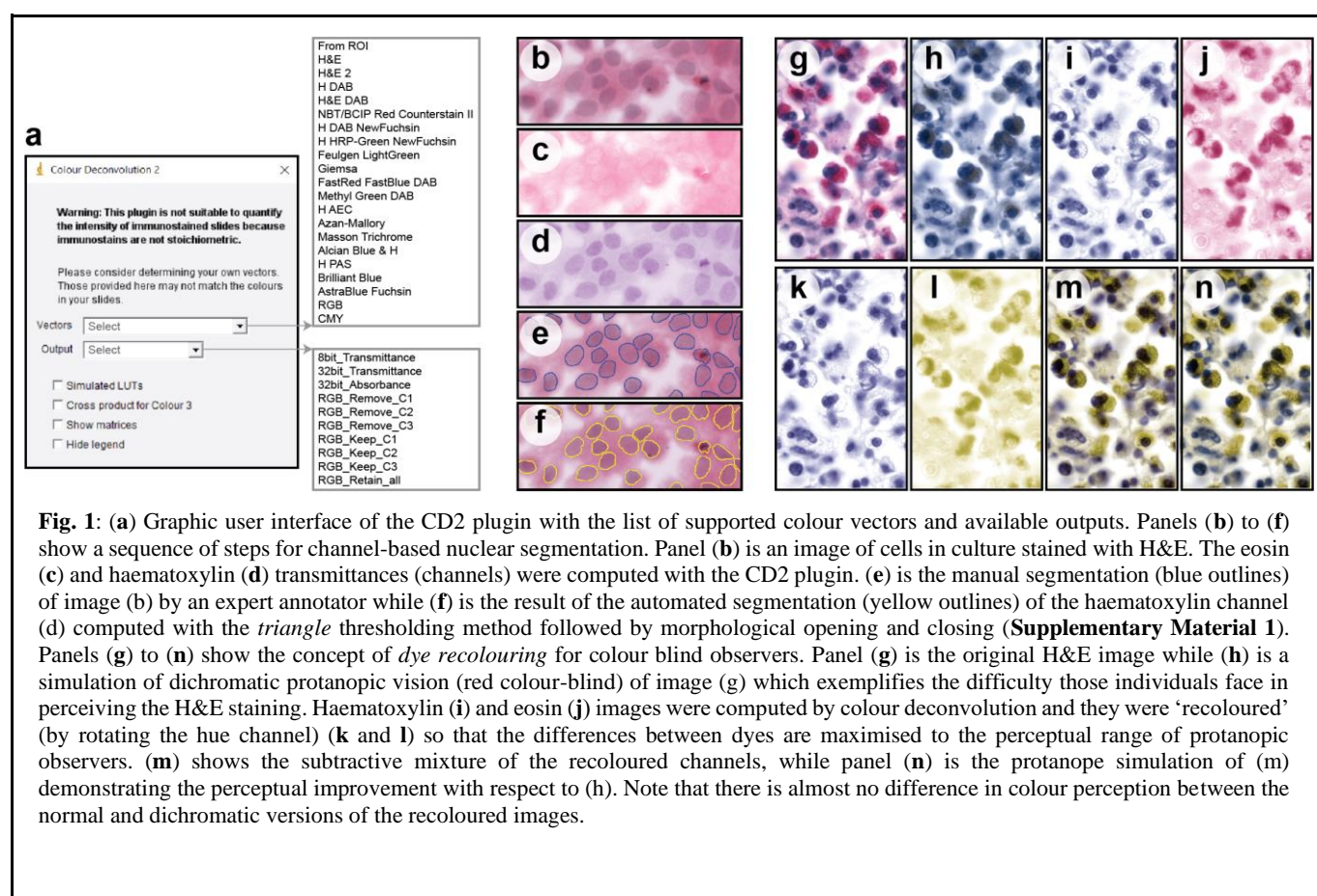
An early version of the ImageJ plugin has been enhanced as follows: *1*) The execution time was improved ~3.2 times by pre-computing formulae that are applied to every image pixel. *2*) The Moore-Penrose pseudoinverse (cross product) method was incorporated to determine the third undefined vector in two-dye stains. *3*) The invertibility of the normalised matrix *M* is verified (required for the unmixing step). *4*) Output options now include: 8bit and 32bit channel *transmittance* images (with or without colour look-up tables approximating the colour vectors) and 32bit *absorbance* images. *5*) Several RGB *intensity* image operations enable removing or retaining the contribution of a chosen dye. *6*) Macro code is auto-generated to store the colour vectors determined using ROIs. *7*) The ImageJ plugin can be executed directly from other plugins or scripting languages through a public run method without requiring dialogue interaction.

3. Examples and Results

Figure 1a shows the user interface of the plugin with the list of precomputed vectors and options available. Users are, however, encouraged to determine their own vectors because variations and modifications of staining methods can result in subtle colour differences. The ability to process dyes independently enables a range of segmentation procedures originally designed for monochromatic images (*e.g.* most histogram-based thresholding and texture segmentation algorithms) to be applied with an important advantage: instead of using pixel intensities of the colour space encoding the images, users can exploit differential dye affinities to tissues and histochemical properties for more biologically-relevant segmentation results. An example of such approach is shown in **Figure 1b-f** where colour deconvolution makes it possible to perform nuclear segmentation at a comparable level as expert human annotators (**Supplementary Note 3**). Since the staining method used in a preparation is always known, applying colour deconvolution as a pre-processing step might be worth exploring as a means of generalising segmentation methods so differences in staining techniques used are not critical.

Processing dyes independently also provides an opportunity to improve data visualisation, *e.g.* for dichromatic (colour blind) observers, where certain combinations of dyes can be easily confused. **Figure**

1g-n presents one method of digital recolouring of haematoxylin and eosin (H&E) images for dichromatic viewers so they can better perceive differential staining of perceptually confusing dyes more clearly (Landini and Perryer, 2009).

Fig. 1: (**a**) Graphic user interface of the CD2 plugin with the list of supported colour vectors and available outputs. Panels (**b**) to (**f**) show a sequence of steps for channel-based nuclear segmentation. Panel (**b**) is an image of cells in culture stained with H&E. The eosin (**c**) and haematoxylin (**d**) transmittances (channels) were computed with the CD2 plugin. (**e**) is the manual segmentation (blue outlines) of image (b) by an expert annotator while (**f**) is the result of the automated segmentation (yellow outlines) of the haematoxylin channel (d) computed with the *triangle* thresholding method followed by morphological opening and closing (**Supplementary Material 1**). Panels (**g**) to (**n**) show the concept of *dye recolouring* for colour blind observers. Panel (**g**) is the original H&E image while (**h**) is a simulation of dichromatic protanopic vision (red colour-blind) of image (g) which exemplifies the difficulty those individuals face in perceiving the H&E staining. Haematoxylin (**i**) and eosin (**j**) images were computed by colour deconvolution and they were 'recoloured' (by rotating the hue channel) (**k** and **l**) so that the differences between dyes are maximised to the perceptual range of protanopic observers. (**m**) shows the subtractive mixture of the recoloured channels, while panel (**n**) is the protanope simulation of (m) demonstrating the perceptual improvement with respect to (h). Note that there is almost no difference in colour perception between the normal and dichromatic versions of the recoloured images.

4. Limitations

Six important considerations to achieve best results with this technique are: *1*) background correction is essential to provide a neutral colour background and a uniformly illuminated field. *2*) Neutral (grey) dyes cannot be unmixed with this method. *3*) Staining vectors should be determined in one's own lab from single-dye stained sections. *4*) Staining intensity quantification is meaningful only when dye uptake is stoichiometric (*e.g.* Feulgen reaction for DNA). Immunohistochemical methods are unfortunately non-stoichiometric, making them unsuitable to measure antigen expression (this, however, does not preclude using colour deconvolution for segmentation or cell counting of immunostained preparations). *5*) Some dyes scatter rather than absorb light (*e.g.* silver impregnation methods) and therefore cannot be unmixed reliably. *6*) Despite an emerging trend in high resolution colour applications (with up to 16bit/channel) the software is restricted to 24bit RGB format. Consequently the analysis of images in higher bit formats must be preceded by a conversion to 24bits.

5. Conclusions

Colour deconvolution makes possible a range of applications by processing dye channels independently without the influence of dye co-localisation and enables several innovative processing strategies, such as morphological segmentation based on tissue differential stain uptake, unsupervised marker localisation and filtering artefactual material (Fouad *et al.*, 2017, 2019, Randell *et al.*, 2017).

ImageJ and MATLAB source code, documentation and examples are available from: blog.bham.ac.uk/intellimic/g-landini-software/colour-deconvolution-2/

References

Fouad,S. *et al.* (2017) Unsupervised morphological segmenta-tion of tissue compartments in histopathological images. *PLoS One*. **12** (11): e0188717.

Fouad,S. *et al*. (2019) Imaging and machine learning methods for assessing HPV in situ hybridisation patterns in oropharyngeal carcinomas. *J Pathology Inform.* **10**: 32, S13.

Landini,G. and Perryer,G. (2009) Digital enhancement of haematoxylin- and eosin-stained histological images for red-green colour-blind observers. *J Microsc*. **234** (3): 293-301.

Rasband,W.S. (1997) ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, https://imagej.nih.gov/ij/, 1997-2020.

Randell,D.A. *et al*. (2017) Mereotopological correction of segmentation errors in histological imaging. *J Imaging* **3** (4)*: 63.*

Ruifrok,A.C. and Johnston,D.A. (2001) Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol*. **23** (4): 291-299.

Zhou,R. *et al.* (1996) A multiple wavelength algorithm in color image analysis and its applications in stain decomposition in microscopy images. *Med Phys*. **23** (12): 1977-1986.