



PSEUDO-GRAYSCALING TECHNIQUE: A PRE-PROCESSING STEP TOWARDS PATHOLOGICAL SLIDE ANALYSIS

¹Sanjay Nag, ²Satabdi Bhattacharya, ³Ranita Banerjee, ⁴Dr. Indra Kanta Maitra and ^{5*}Prof. Samir Kumar Bandyopadhyay

¹University of Calcutta, Kolkata.

^{2,3,4}B.P.Poddar Institute of Management and Technology, Kolkata.

⁵University of Calcutta, Kolkata.

***Author for Correspondence: Prof. Samir Kumar Bandyopadhyay**

University of Calcutta, Kolkata.

Article Received on 03/03/2016

Article Revised on 26/03/2016

Article Accepted on 16/04/2016

ABSTRACT

This paper presents a method for converting digitized colour images in RGB format, obtained from pathological/histopathological slide to pseudo-grayscale as a pre-processing step for image segmentation. The digitized slide images exhibits colour shades in the order of 2^{24} which increases algorithmic complexity. At the same time due to the application of disparate stains for slide preparation some specific sub-domain(s) of the entire colour space is actually exhibited by these slide images. The pseudo grayscale is actually exploiting these features to make the grayscale image more enhanced and preserving the disparate colour information, making subsequent feature extraction easy and efficient. The statistical regression on the image information is followed by selection method to transform the images from RGB colour model to pseudo-grayscale images. The proposed method is tested on a large set of pathological/ histopathological slide images some of which has been demonstrated in this citation.

KEYWORDS: Histopathology, Digital Microscope, Grayscale, Bitmap, Multiple Regression.

1. INTRODUCTION

Pathology is a branch of medicinal biology where laboratory examination including microscopic study of body fluids and tissue samples are performed to study the manifestations of disease. Histopathology is the microscopic examination of biological tissues to observe the presence of infected cells and tissues in very fine detail. The origin of the word 'Histopathology' is a combination of three Greek words, 'histos' meaning tissue, pathos meaning disease and logos meaning study. Hence Histopathology is the study of microscopic changes or abnormalities in tissues that are caused as a consequence of diseases/infections.

Histopathology, as of today is practicing pathology in a totally different scenario in context to the change taking place in science and technology. Advancement in information technology, communication science has introduced tele-medicine, tele-pathology and digital imaging analysis systems. Histopathology over the years has witnessed its evolution from mainly autopsy based pathology to the current molecular histopathology.

The slides containing tissue samples that is required to be analysed by the histo-pathologist are digitized using a digital microscope. A **digital microscope** uses optics and

a digital camera to capture image and display it on to a monitor through a software running on the computer. The images are of high resolution so image analysis using computerized algorithms require higher computation complexities. The colour image matrices are of the order $M \times N \times 3$ and each pixel is of three values. The image matrix can be reduced to the order $M \times N$ with each pixel consisting of a single value, the computation for applying the techniques reduces immensely. This dimension reduction problem is the conversion of the colour image to a grayscale image.

The general problem of conversion of a colour image to a grayscale image has been subject to much research. Many such techniques ventures to match human perception of the colour and grayscale images as much as possible during conversion. However the demand in digital histopathology is slightly different. The information loss needs to be minimized for a specific image under consideration. A new method has been proposed in this paper for conversion of colour image of digital histopathological slide to its pseudo-grayscale equivalent.

The following section reviews existing work in grayscale conversion using different methods. In section three of

this citation, the proposed a pseudo-grayscale conversion technique to convert the RGB pixel values of the original image to a grayscale value in the resulting is discussed. The results of implementation of the algorithm on the dataset of histopathological slide images has been discussed in section four followed by a discussion section and conclusion in the final section.

2. Literature Review

There have been significant research work for converting a coloured image to grayscale with minimal loss of information. Strickland et al.,^[1] proposed a local colour image enhancement technique based on saturation feedback. Amy A.Gooch, Sven C. Olsen^[2] devised an extension of the CIE Lab colour model (S-CIE Lab) that is useful for measuring colour difference between the images. Using the pattern-colour separable transformation, the S-CIE Lab difference measure reflects both spatial and colour sensitivity. Bala and Eschbach^[3] represented spatial colour to grey transformation that locally preserves the chrominance edges by introducing high-frequency chrominance information into the luminance channel. The method applies a spatial high-pass filter to the chromatic channels, weighs the output with a luminance dependent term, and finally adds the result to the luminance channel. Bala and Braun^[4] proposes a new method for converting colour graphics to grayscale in a manner that preserves distinctiveness of the regions. Colours are first sorted according to their original lightness values. They are then spaced equally in grey, or spaced according to their 3-D colour difference from colours adjacent to them along the lightness dimension. This is most useful when maximum differentiation is desired in images containing a small number of colours, such as pie charts and bar graphs. Jashojit Mukherjee^[5] presented a global mapping scheme such that the same colour does not map to two different grayscale values. This method adds colour information to the brightness information to reduce information loss and the weights for the added colour are dependent on the distribution of the pixels in a particular image.

The techniques that Amy A.Gooch, Sven C. Olsen^[2] and BalaEsbach proposed were both based upon the luminance and the chrominance of the image. While BalaEsbach's method dealt about the preservation of adjacent colours by applying high pass filter to the prominence channels, further adding the results to the luminance channel, resulting in the luminance variation introducing only in that region of the image which contains high frequency chrominance statistics, AmyA.Gooch, Sven C. Olsen^[2] used the chrominance and luminance differences to create grayscale target differences from its neighbouring image pixels. The methods by Strickland et al^[1] and Jashojit Mukherjee^[5] provide two divergent paths in converting the image into grayscale from coloured images. The citation^[1] proposed a colour enhancement technique. This research citation proposes a new method based on statistical analysis of

the colour image and based on multivalued regression of the colour space.

3. Proposed Method

The digitized histopathological/ pathological slides are obtained as Bitmap images. The images are coloured in RGB colour model. For better segmentation results and to reduce the complexity associated with multiple colour channel the image needs to be registered as grayscale image through conversion. The input medical image of histo-pathological slide is more often a coloured image and to reduce algorithmic complexity it is needed to reduce the number of colours with minimal loss of significant information. In this paper, a pseudo-grayscale conversion method is proposed which is different from the true grayscale method, and is effective for pathological and histopathological slides since it is preserving most of its features. The general and most accepted grayscale conversion method adopted by most image processing software is the weighted model. The grayscale image obtained using the given formulation results in an image that can perceptually distinguish colour regions within the gamut of grayscale. The process of conversion is initiated by reading the image in RGB colour space into a bitmap matrix containing all the three colour channel intensity values.

$$\text{Bitmap} = \sum_{i=0}^r \sum_{j=0}^c I(i, j)$$

The Red, Green, Blue values of each pixel of the colour image are extracted and stored in an image matrix. Then using the proposed method of Statistical Regression the image is converted into a pseudo grayscale image. A grayscale image containing only intensity information for each pixel location, varying from black as the weakest to white as the strongest intensity. It results in consequential reduction of information. Pseudo-grayscale is not grayscale in true sense. It is eliminating the number of colour shades without significant loss of RGB colour information, making the feature extraction faster and efficient.

In this research paper multivariate Regression has been used to convert the colour image to grayscale. In case of bivariate distribution where the value of one variable is dependent on the value of another variable, a simple correlation can measure the degree of association between the two variables. A linear regression equation can then be applied to obtain the values for the dependent variable. It may so occur as in the case of RGB model there are more than one variable. Such a distribution where there is joint distribution of three or more variable values, this is statistical known as a multivariate distribution. To determine the value of one variable among several others, it is required to apply multiple regression equation that attempts to determine the strength of the relationship between one dependent variable (Y) and a series of other variables (independent variables). When the values of one variable are influenced by those of another, the correlation coefficient r yields a measure of the degree of association between

these two variables, and estimates of the values of a variable are obtained from the appropriate linear regression equation.

Linear Regression: $Y=a+bx$ (i)

For application of multiple regression on three variables the following equation will estimate the value of a dependent variable over the other two independent variables.

Multiple Regression: $x_1=a+bx_2+cx_3$ (ii)

In this research work the problem involves three variables, R, G and B values of each pixel of the colour image. The mean of all the values of R, G and B were calculated separately as (mean R), (mean G) and (mean B) using the formula

Mean= $\sum n(i) / n$ (iii)

The greatest among the three mean value represents the most prominent colour dependency factor, was taken as the dependent variable and the other two as the independent ones. The independent variable affects the change over the dependent variable colour. The objective of using multiple regression is to obtain a regression value for the dependent colour based on the influence of the independent variable colours channel information. Then using the values the independent colour value are calculated employing the following equation:

$(\bar{R})=b_{12.3}(G-(\bar{G}))+b_{13.2}(B-(\bar{B}))$ (iv)

where

$b_{12.3} = ((\sigma_1/\sigma_2)((r_{12}-r_{13}r_{23})/(1-r_{23}^2)))$ (v)

$b_{13.2} = ((\sigma_1/\sigma_3)((r_{13}-r_{12}r_{23})/(1-r_{23}^2)))$ (vi)

are called partial regression coefficients. σ_1 , σ_2 , σ_3 are the standard deviations of the variables R, G, B which are required to measure intra colour variations. The values of r_{12} , r_{13} , r_{23} denote the product moment

correlation coefficients between R and G, between R and B, between G and B respectively.

Finally the grayscale value for each pixel was calculated. To calculate the unique colours the following relation is computed:

Unique colour= $\begin{cases} \text{selected} & , \text{if not present} \\ \text{redundant} & , \text{if present} \end{cases}$

A pseudo grayscale palate is constructed based on the multiple regression values obtained for the prominent colour in relation to the independent colour channel value. It may be observed that the number of grayscale generated by using the regression model may be restricted to a limited zone on the grayscale window. The probability of such a distribution is highest when using digital slide images as they are mostly stained to highlight region of interest. While doing so, the images tend to pick up the stain colour as the most dominant colour. The image histogram will be highly skewed on one side while a vast region of the histogram will remain almost empty. So to achieve an evenly distributed grayscale and to differentiate the closest of the regions the obtained grayscale palate is equally stretched. A ratio of difference value is generated based on the number of unique grayscale value obtained from the image and the maximum possible grayscale possible. This difference value is added to the individual grayscale values of the existing palate to obtain a new distribution of grayscale value on the grayscale palate. This palate is used for conversion to grayscale by mapping the dominant colour value i.e. the dependent colour value to the grayscale value from the palate created.

The image is again read and the RGB values are replaced with the newly obtained regression values as a representative pseudo-grayscale values whose image representation of colour space is given in Fig 1b.



1(a).Original Colour palette



1(b).Pseudo-grayscale palette



2(a). Original Colour palette



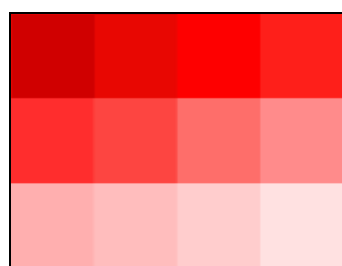
2(b). Pseudo-grayscale palette

4. RESULTS

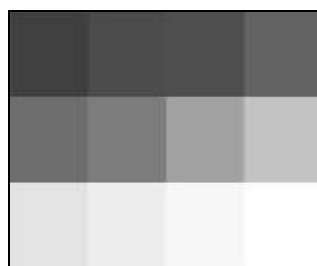
The proposed algorithm was implemented using C# and tested with various images that were typically designed to differentiate colours based on those expected colours found on histopathological slide samples. Further tests were carried out on images using standard grayscale conversion algorithms and some of the results have been compared with the proposed method.

A number of test images from histopathology and blood smear slide images were used to test the proposed methods. The figure below compares the result of the proposed method against the standard grayscale method obtained using commercial software like Adobe Photoshop [6]. In figure (a) a colour swatch was used

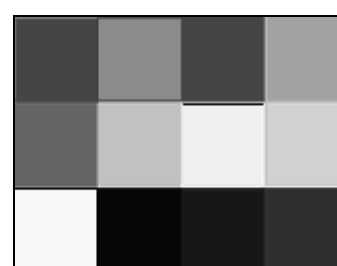
where there were variations in hue and saturation with mild variation in luminosity. Figure (b) shows the comparative results for a histopathological image of cells exhibiting mitotic division. Figure (c) The image of a malaria slide smear. The image is stained with Giemsa and have the typical bluish tinge and the RBC cells are darker in contrast because of bio-chemical reaction of the stain with haemoglobin of RBC cells. Figure (d) Biopsy slide image showing histopathology of muscles are images that are also stained to highlight the different tissue regions. Figure (e) Biopsy slide image showing histopathology of tumor of breast. This image also show a region of high cellular activity indicated by extensive mitotic division.



Original

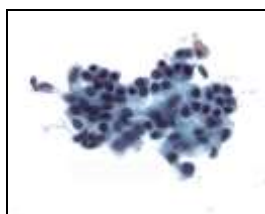


Existed Method



Proposed method

Figure (a) A Colour Swatch



Original

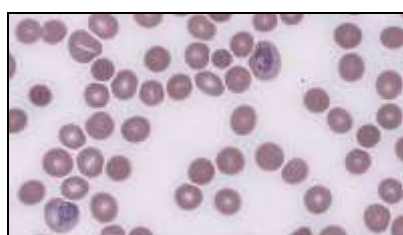


Existed Method

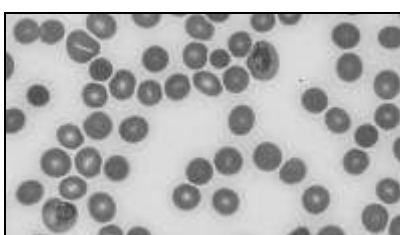


Proposed method

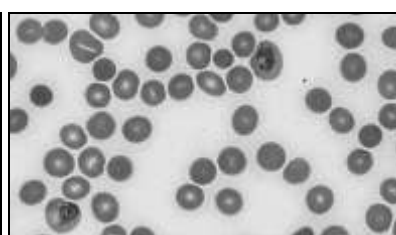
Figure (b) Histopathological image of cells exhibiting mitotic division



Original



Existed Method



Proposed method

Figure (c) The image of a malaria slide smear with Giemsa staining

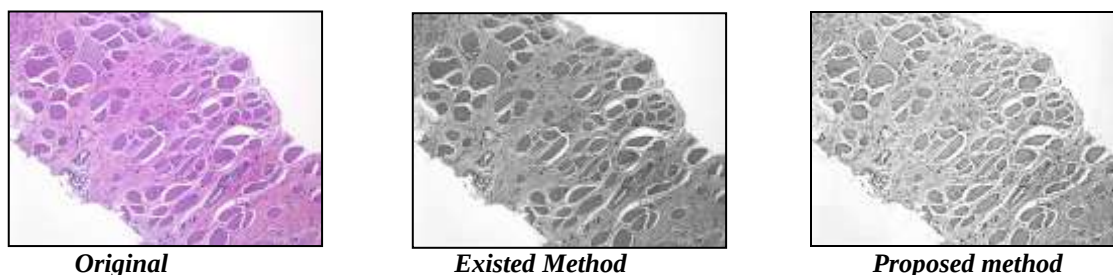


Figure (d) Biopsy slide image showing histopathology of muscles

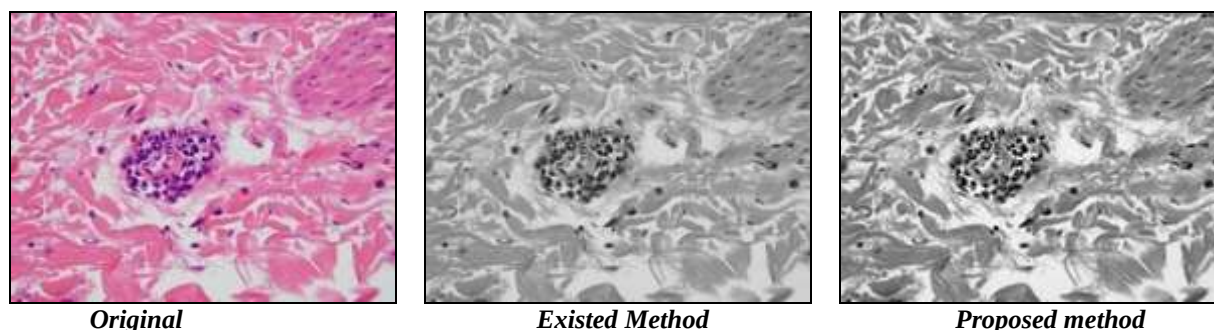


Figure (e) Biopsy slide image showing histopathology of tumor of breast

5. DISCUSSIONS

The conversion of a colour image to grayscale image is considered as a pre-processing step in image processing for suppression of colour information before image segmentation. For the purpose of histopathological slide image analysis, conversion of image to grayscale is absolutely vital for better segmentation results. The objective of this conversion is to differentiate the regions in such a manner that segmentation will segregate each distinct region as a separate object. The standard algorithm used for grayscale conversion, for standard colour television and video systems such as PAL, SECAM, and NTSC, uses a nonlinear formulation on a compressed primary intensities as a weighted sum of colour channel used for colorimetric grayscale calculations.

$$\text{Grayscale Value} = 0.229 R + 0.587 G + 0.114 B$$

The grayscale image obtained using the above formulation can perceptually differentiate the colours and preserves the brightness/luminosity of the colour image. The proposed algorithm depends on the relation of each channel and their impact on the colour formation. In most cases the histopathological images are stained with a particular chemical that imparts a distinctive colour shade on the image. Therefore, the standard conversion formula will not be able to impart different shades on very close colour distances. A pseudo-grayscale algorithm, like the one proposed in this work will be able to resolve and differentiate such small colour distances. The algorithm obtains the most prominent colour channel statistically and then determines the relationship between the other colour channels. A multiple regression model then determines a new value based on the relationship of the other two colour channels over the prominent colour channel. The number of unique grey shades obtained is distributed across the

spectrum of grayscale for better differentiation and so that the entire gamut of grayscale shade is available for the image. The results obtained above shows a better contrast with the proposed algorithm. Segmentation or edge detection algorithms will provide better results when using the grayscale conversion proposed in this citation. The effectiveness of this algorithm is best with histopathological slide images. But normal images that use a bigger spectrum of the colour space will not yield good results. In such cases the standard algorithm will produce images that are better perceptually.

6. CONCLUSION

The transformation from colour to grayscale is a dimension reduction problem. Thus loss of information one way or another cannot be avoided. However many subsequent algorithms used in image processing for processes like edge detection and image segmentations become very complex and time consuming when performed on colour images. To reduce the complexity this conversion becomes necessity. So it is required to solve the problem based on the trade-off between information loss and reduction of time complexity. This paper attempts to minimize information loss for histopathological slide images and blood smear slide images that exhibit colour space restricted to small region on the colour spectrum. The images are converted to a grayscale image that preserves the distinctiveness of region and hence can be exploited by subsequent image processing algorithms like edge detection and segmentation. Further work is viable in this area and this method is an effort in that direction.

ACKNOWLEDGEMENT

The authors would like to thanks Dr. Pradip Saha, MD for his active participation and expert opinion on medical science. The authors are also thankful to Mrs. Roshni

Dasgupta, Mr. Sayan Dutta and Mr. Sumit Das, B.P.Poddar Institute of Management and Technology, Kolkata for their contribution related to algorithm development, coding, testing and documentation.

REFERENCES

1. STRICKLAND R. N., KIM C.-S., MCDONNELL W. F.: Digital color image enhancement based on the saturation component. *Optical Engineering* 26 (July 1987), 609–616.
2. Amy A. Gooch, Sven C. Olsen, Jack Tumblin, Bruce Gooch, 2005. Color2Gray: Saliency-Preserving Color Removal. *ACM Transactions on Graphics* 24,3, 634-639.
3. R.Bala, R.Eschbach, 2004, Spatial Color-to-Grayscale Transform Preserving Chrominance Edge Information, *Proc. Color Imaging Conference 2004*; 82-86.
4. R.Bala, K.M.Braun, 2004, Color-to-grayscale conversion to maintain discriminability, *Proc. SPIE Vol. 5293, Color Imaging IX: Processing, Hardcopy, & Applications*, Reiner Eschbach, Gabriel G. Marcu, Eds., 2004; 196-202.
5. Jashojit Mukherjee, Indra Kanta Maitra, Kashi Nath Dey and Samir K. Bandyopadhyay 2015. Grayscale Conversion of histopathological slide images as a preprocessing step for image segmentation.
6. Generic Photoshop software, 2004. Photoshop: Converting color images to black and white.