Color Swapping to Enhance Breast Cancer Digital Images Qualities Using Stain Normalization

Izzati Muhimmah¹, Dhina Puspasari Wijaya², Indrayanti³

Department of Informatics Engineering, Universitas Islam Indonesia¹
Magister of Informatics Engineering, Universitas Islam Indonesia²
Faculty of Medicine and Health Science, Universitas Muhammadiyah Yogyakarta³

Jl. Kaliurang Km 14 Yogyakarta 55510
Phone (0274) 895287 ext. 122, fax (0274) 895007 ext. 148
izzati@uii.ac.id¹, dhina.puspa@gmail.com²

Abstract. Histopathology is the disease diagnosis by means of the visual examination of tissues under the microscope. The virtually transparent tissue sections were prepared using a number of colored histochemical stains bound selectively to the cellular components. A variation of colors comes to be a problem in histopathology based upon the microscope lighting for the range of factors. This research aimed to investigate an image enhancement by applying a nonlinear mapping approach to stain normalization and histogram equalization for contrast enhancement. Validation was carried out in 59 datasets with 96.6% accordance and expert justification.

1. Introduction
Histopathology is a disease diagnosis by the visual examination of tissue under the microscope. To examine the tissue sections (which are virtually transparent), they are first prepared using colored histochemical stains that bind selectively to cellular components[4]. A variation of colors can be a problem in histopathology commonly generated from microscope lighting, different scanners, staining chemical variables/reactivity, and transmission of light as a function of the thickness of the part.

Most of the applications that have been developed for color enhancement in microscopic images at this time are concentrated on the features of shapes and ignore the color between slides. When the color information is utilized, the easiest method is by using the color values obtained from scanning. The most popular staining methods for medical diagnosis are stain hematoxylin (H) and stain eosin (E)[6]. The improvement of image quality is a fundamental process of image processing applications; for images taken under bad conditions, the images often become sharper but with the reduction of the staining in the images[6]. Detail images provide the important information for analysis and thus image enhancement is desired[1].

Histogram equalization methods are often employed to correct low contrast images[3]. Rong & Dongnan (2015) developed a color correction from histogram equalization. They used the greyscale
images and their work has never been used in microscopic images with more than one layer. Lin et al. (2015) proposed an approach to stretch the layer colors with histogram averaging and remapping. Their method did not target any enhancement; thus we cannot compare the results for image enhancement standards. An efficient algorithm for image enhancement is needed to improve the image quality in image visualization and image analysis using the computer.

This research investigated a number of methods to enhance image color quality by applying a non-linear mapping approach to stain normalization[4]. We also improved the contrast in the color image of each stain with the histogram equalization of each RGB layer[7]. This research is expected to produce a better model. Testing results have been measured by the distance averages for the RGB layers between the target and purposed images. We aimed to improve breast cancer histology color image by swapping the colors to for the more detail access such as the nucleus and membrane.

2. Materials & Methods

2.1. Color Deconvolution

In the original image, there is a wide, heterogeneous and complex range of colors. Therefore, color separators are used to group the colors of the same type. Ruifrok (2001) mentioned that color deconvolution transforms the red, green, and blue (Ψ) layers into a new space layers of colors red, green, blue (Ψ') using a tissue staining color selection. The original (Ψ) and new (Ψ') image has three stain color descriptors used to represent an optical density. Ruifrok's (2001) stain has some specific characteristics in the three-layer RGB; each of RGB value is represented in the matrix vector with the optical density of 3x1. The implementation of optical density is symbolized with P. P values include the types of solutions: Hematoxylin (H), Eosin (E) and Diaminobenzine (DAB). Here is a stain for one type of solution:

\[
\begin{bmatrix}
\hat{p}_{11} & \hat{p}_{12} & \hat{p}_{13} \\
\hat{p}_{21} & \hat{p}_{22} & \hat{p}_{23} \\
\hat{p}_{31} & \hat{p}_{32} & \hat{p}_{33}
\end{bmatrix}
\]  

(1)

Normalization P is achieved by dividing each vector of optical density (OD)[8]

\[
\hat{p}_{ij} = p_{ij}/\sqrt{p_{ij}^2 + p_{ij}^2 + p_{ij}^2}
\]

(2)

This results in a normalized OD matrix, M:

\[
\begin{bmatrix}
\hat{p}_{11} & \hat{p}_{12} & \hat{p}_{13} \\
\hat{p}_{21} & \hat{p}_{22} & \hat{p}_{23} \\
\hat{p}_{31} & \hat{p}_{32} & \hat{p}_{33}
\end{bmatrix}
\]

(3)

If C is the 3x1 vector for the three stains at a particular pixel, then the vector of the OD levels detected at the pixel is y=CM.

\[
C = M^{-1}[y]
\]

(4)

This implies that multiplying the OD image with the inverse of the OD matrix, defined as color deconvolution matrix D, results in an orthogonal representation of the stains forming the image:

\[
C = D[y]
\]

(5)

The stain value was used to determine the optical density of each layer. In each solution of Hematoxylin, Eosin, and Diaminobenzine, the stain values are unique in each of RGB layer[8].

2.2. Normalization Stain Matrix Algorithm

The proposed algorithm (see Figure 1) consisted of four modules: stain matrix estimation, CD, nonlinear mapping of channel statistics, and reconstruction. There were two input images used, which were served as the source and target images. These images have some different qualities in terms of their stain solution. The nonlinear correction (mapping) was applied to each channel. Then, data were normalized and separated from the target image (based on statistics calculated from the target image).
Finally, this research reconstructed the normalized source image using the normalized stain channel of the target. The following sections present the details of each of these modules.

![Figure 1 Normalization Stain Algorithm](image)

2.2.1. Stain Matrix Generation

This study used the CD framework to convert both source and target images from RGB color space to a new one defined by the constituent stains. This required the estimation of image-specific stain matrix (S) each for source and target images. The estimation of image specific stain matrices used a global (preimage) SCD (Stain Color Descriptor) and local pixel-level color information in a supervised color classification framework. Figure 2 shows the diagram of the stain color descriptor process.

![Figure 2 Stain Color Descriptor](image)

Figure 2 presents the overview of the stain matrix estimation method consisting of two phases: learning and evaluation. Learning, performed offline, essentially involved two steps: 1) deriving principal color (PCH) from the training set of quantized image histogram to obtain SCDs and 2) the learning classification models by utilizing RGB (pixel information) and SCD (whole-image color information) in a supervised classification framework to generate the stain-specific probability maps (one for each stain and background). These probability maps were used to estimate the color of each stain for a particular image.

The following rules were used to categorize the images into three classes: background, stained, and other (see Table 1).

| Condition | Class |
|-----------|-------|
| $P(bgd|F) > T_{bgd} \hat{\zeta}(c)$ | BACKGROUND |
| $P(s_n|F) > T_{fgd} \hat{\zeta}(c)$ | STAINED |
| Otherwise | OTHER |

2.2.2. Nonlinear Mapping of Channel Statistics
Each channel of deconvolved target and source image, the spline-based nonlinear mapping was used to calculate the statistics set and their corresponding probability map of both images. These values were then utilized to map the statistics of each source channel to replace the corresponding target image channel.

2.2.3. Reconstruction
Once each of the stain channels of source image was normalized independently, they were then recombined on each pixel basis as follows:

\[
\begin{align*}
X_r^{\text{norm}}(c) &= 255X \prod_{n=1}^{3} e^{-X_n^{\text{norm}}(c) s_{r,n}} \\
X_g^{\text{norm}}(c) &= 255X \prod_{n=1}^{3} e^{-X_n^{\text{norm}}(c) s_{g,n}} \\
X_b^{\text{norm}}(c) &= 255X \prod_{n=1}^{3} e^{-X_n^{\text{norm}}(c) s_{b,n}}
\end{align*}
\]

where \( c \in C \) refer to a pixel on a 2-D grid \((C, x^{\text{norm}})\) refers to the normalized stain channel \( n \) and \( s_{\alpha,n} \) (where \( \alpha \in \{r,g,b\} \)) is the stain vector associated with \( n \) channel of the normalized \( s^{3} \) stain matrix[4].

2.3. Histogram Equalization
Each color pixel was represented by a vector with as many components as the color components in a proper color space (i.e., three components: Red, Green, and Blue in the RGB space). It was enhanced by smoothing the multi-level techniques derived from the statistical language techniques.

Greyscale histogram equalization attempted to uniformly distribute the pixel gray levels of an image to all available gray levels \( L \) (e.g. \( L = 256 \), when 8 bits were used to represent each gray level). Let consider the image pixel gray level to be an rv \( x \). this histogram of a grayscale image is the probability density function (PDF) of \( x \) defined as

\[
f_x(x_k) = P\{x = x_k\} = \frac{N(x_k)}{\sum_{m=0}^{L-1} N(x_m)} \quad \forall \ k = 0,1,...,L-1
\]

(7)

where \( N(x_k) \) is the number of pixels with gray level value \( x_k \). The Cumulative Distribution Function, CDF, \( f_k(x_k) \) of the rv \( x \) is given by

\[
y_k = F_k(x_k) = P\{x \leq x_k\} = \sum_{m=0}^{k} f(x_m) \quad \forall \ k = 0,1,...,L-1
\]

(8)

This is the simplest approach to color the histogram equalization. Since many color images have three color bases, the color on each pixel is represented by 3-dimentional vector and greyscale histogram equalization is performed in each of the three color components separately. For color image, the color of each pixel was assumed to be a random vector= (XR, XG, XB), where XR, XG, XB were rvs modeling the Red, Green and Blue components, respectively. Thus, by applying equation (8) the equalized histogram of each color component was estimated[2].


3. Research Methodology

The image data examination of HE breast cancer cells was taken from anatomic pathology laboratory of medicine, Universitas Muhammadyah Yogyakarta. The process of getting the image started from sampling tissue in breast cancer patients seen under the lens Olympus CX31 microscope with a magnification of 40X. Cells were taken in the paint using Hematoxylin (H) and generating substrate Eosin (E). Reaction bonding that occurred between cells and dye caused the cell to be blue and magenta. Image had 1920x1080 dimensions.

Figure 3 presents the research scheme in which an image of the target would map the source images with different color. The reconstruction results in Figure 3 used the method of histogram equalization on each layer RGB.

4. Results

Target image and source image were selected from the image data of HE breast cancer. The target image was as a reference image and the source image was the image that would be color corrected. Figure 3 presents the results of the repair process. Each source image and target image did the method of color deconvolution, and then normalization stain mapping to stain H, stain E, and background. The results reconstructed the normalization stain (Khan normalization method results[4]). The proposed method was conducted by applying the histogram equalization of the image reconstruction results normalized stain.

![Figure 3. Chart Research Process](image)

Tests performed image calculated distance method with pdist2 function in Matlab version 2015b. The result on Khan’s method calculated the distance methods resembling to the image target as well as the methods proposed. The closer the image, the more it resembles. The result of the calculation of distance is displayed in Figure. 4.

The evaluation of the quality of the image enhancement was also justified by the pathologist. We had provided the pathologist with 59 sets of source image, target image, Khan’s, and the proposed results. We asked her to visually compare which the enhancement results provided the detail information on cell and membranes better. We took the notes on her scoring.

![Figure 4. Figure result color swapping](image)
Figure. 5. Results of Distance Testing

Figure 5 shows the results of the average distance of methods Khan[4] for each layer $R= 0,770371667$, $G= 0,546687$, $B= 0,519965083$. On the other hand, the average distances of the proposed method each layer were $R= 0,349969167$, $G= 0,505668$, $B= 0,43282475$. From the calculation of the distance it indicated a distance between the image of the proposed method and the image of the target closer.

Testing models used 59 datasets. The results of proposed method declared 57 dataset in accordance with the criteria of experts. Whilst, there were 2 dataset declared better than Khan’s by the experts. The percentage of testing results declared that the proposed method reached an agreement with the expert criteria by 96.6%.

5. Conclusion

This study is the first process (preprocessing) for detecting the presence of breast cancer in the image HE. From the results of the implementation method, it is found that the proposed method has more similarities than previous methods proven by using the method of calculation of the distance.

Acknowledgement

A research grant from PUPT from Ministry RISTEK-DIKTI of Indonesia for the period of 2015-2016, contract no. 001/ HB-LIT/III/2015, is gratefully acknowledged.

6. References

[1] Bai X 2015 Morphological Feature Extraction for Detail Maintained Image Enhancement by Using Two Types of Alternating Filters and Threshold Constrained Strategy Elsevier Optic 126 pp 5038-43.
[2] Bassiou N and Kotropoulos C 2007 Color Image Histogram Equalization by Absolute Discounting Back-Off Compt. Vis. Image Underst 107(1) pp 108-22.
[3] Jiang G, Lin S F, Wong C Y, Rahman M A, Ren T R, Kwok N, Shi H, Yu Y H and Wu T 2015 Color Image Enhancement with Brightness Preservation Using Histogram Specification Approach Elsevier Optic 126 pp 5656-64.
[4] Khan A M, Rajjpoon N, Treanor D and Magee D 2014 A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution Biomed. Eng. Vol 61 6 pp 1729-38.
[5] Lin S F, Wong C Y, Jiang G., Rahman M A, Ren T R, Kwok N, Shi H, Yu Y H and Wu T 2016 Intensity and Edge Adaptive Unsharp Masking Filter for Color Image Enhancement Elsevier Optic 127 pp 407-14.
[6] Macenko M, Niethammer M, Marron J S, Borland D, Woosley J T, Guan X, Schmidt C, and Thomas N E 2009 A Method For Normalizing Histology Slides for Quantitative Analysis Proc. of the Sixth IEEE Int. Conf. on Symposium on Biomedical Imaging: From Nano to
Macro pp 1107-10.

[7] Rong Z, Li Z and Dong-nan, L 2015 Study of Color Heritage Image Enhancement Algorithms Based on Histogram Equalization Elsevier Optic 126 pp 5665-67.

[8] Ruifrok A C and Johnston D A 2001 Quantification of Histochemical Staining by Color Deconvolution The international academy of cytology analytical and quantitative cytology and histology pp 291-9.