A COMPLETE COLOR NORMALIZATION METHOD ON PATHOLOGICAL IMAGES

AMAL KUMAR R.G[1], ARUN LAL C.K[2]

[1] *PG Scholar, Dept of ECE, KMCT College of Engineering Kozhikode, Kerala, India*
[2] *Asst Professor, Dept of ECE, KMCT College of Engineering Kozhikode, Kerala, India*

ABSTRACT

In digital histo-pathology, tasks of segmentation and disease diagnosis are achieved by quantitative analysis of image content. However, color variation in image samples makes it challenging to produce reliable results. This paper introduces a complete normalization scheme to address the problem of color variation in histo-pathology images jointly caused by inconsistent biopsy staining and non-standard imaging condition. Method: Different from existing normalization methods that either address partial cause of color variation or lump them together, our method identifies causes of color variation based on a microscopic imaging model and addresses inconsistency in biopsy imaging and staining by an illuminant normalization module and a spectral normalization module, respectively. In evaluation, we use two public datasets that are representative of histo-pathology images commonly received in clinics to examine the proposed method from the aspects of robustness to system settings, performance consistency against achromatic pixels, and normalization effectiveness in terms of histological information preservation. Results: As the saturation-weighted statistics proposed in this work generates stable and reliable color cues for stain normalization, our scheme is robust to system parameters and insensitive to image content and achromatic colors. Conclusion: Extensive experimentation suggests that our approach outperforms state-ofthe- art normalization methods as the proposed method is the only approach that succeeds to preserve histological information after normalization. Significance: The proposed color normalization solution would be useful to mitigate effects of color variation in pathology images on subsequent quantitative analysis.

Keyword : - *Color variation, illuminant normalization, histological information preservation, Histo-pathology image*

1. INTRODUCTION

Digital histo pathologyis a research field where color image processing algorithms and pattern recognition methods are exploited to enable computers to understand histo-pathology images and to make diagnosis decisions. As quantitative analysis on histo-pathology images is usually achieved through comparing numerical descriptors of a query image to prior knowledge obtained from training data or physicians, deviation of numerical descriptors of a query image from priorknowledge should only reveal the true differences between histological information conveyed by images. However, due to operational inconsistency in histo-pathology image preparation, images of biopsy samples stained by the same types of chemical dyes usually appear in different colors. Consequently, numerical features extracted directly from images may be distorted by such color variation and deviate from true values, finally resulting in inaccurate segmentation and diagnosis.

To reduce effects of color variation among histo-pathology images on numerical features and subsequent analysis, some works extracted numerical features from a grayscale version of a query image [1], [2]. However, a large amount of information mainly carried by color is ignored in these approaches. Recent research in digital histo-pathology has confirmed significance of color information in quantitative analysis on histo-pathology images with few color variation generated under tightly controlled laboratory conditions [3], [4]. To take advantage of color

information for accurate quantitative analysis within large datasets, color variation caused by operational inconsistency in histo-pathology image preparation should be removed beforehand, so that extracted features represent the real histological characteristics only. In this sense, normalization algorithms to remove color variation in images are crucial in histo-pathology.

In this paper, we propose an effective normalization scheme to address color variation in histo-pathology images generated by light-absorbing stains only. Compared to previous works, our method is significant as it addresses 3 challenges of histopathology image normalization holistically as follows.

A histo-pathology image is a final output of a biopsy processing pipeline including sectioning, staining, and imaging. Any operational inconsistency in this pipeline may cause color variation [5]. Hence, the first challenge of color normalization for histo-pathology images is to blindly identify different sources of color variation. To this end, this paper introduces a complete color normalization solution based on a microscopic imaging model. Although similar models are exploited in other methods to normalize spectral variation in stains, it should be stressed that our solution solves a more complicated problem usually occurring in clinics, where color variation in images is jointly caused by inconsistent staining and imaging conditions. By identifying the two causes of color variation, an illuminant normalization module and a spectral normalization module are proposed and concatenated to form our complete normalization pipeline.

The second challenge arises from the use of multiple chemical dyes on a single tissue sample. Despite counter colors visually, stains may mix due to co-located histological components, and thus image colors are determined by various combinations of stains. To identify contribution of each stain to color variation, a non-negative matrix factorization (NMF) based stain decomposition algorithm is developed. Compared to existing adaptive stain decompositions whose performance are prone to be affected by achromatic pixels, a key characteristic of our algorithm is its noise-resistant property contributed by the proposed saturation-weighted (SW) statistical method. The innovation of SW statistics is advantageous to limit impacts of achromatic pixels on color estimation, and thus to produce consistent stain estimation. This noise resistance is beneficial especially for stain estimation on images having more blank areas.

A third challenge is to avoid histological information loss after image normalization. The capability of preserving histological information (including tissue texture details, spatial structures, and morphology features of histological objects) is a very important criterion to evaluate normalization algorithms in digital histo pathology. However, this challenge is never explicitly examined in previous works. In this work, we pay close attention to maintaining histological information when developing algorithms. In evaluation, we particularly design an experiment to examine this capability of the proposed normalization method and existing color normalization algorithms. Our results show that only the proposed method succeeds to maintain tissue features after normalization.

2. STATE OF THE ART

2.1 Previous works on normalization

We categorize histo-pathology image centered color normalization solutions in literature into three distinct groups.

Histogram matching: The first group of normalization algorithms are based on histogram matching in the RGB color space. In [8], after image background removal, histogram matching is performed in the red, green, and blue channels, respectively. In [9], color map quantile matching, a variationof histogram matching, was proposed for stable color normalization. Since histogram matching on entire images ignores local differences of image content, color associated with one stain may be matched to irrelevant colors. Recently, tissue component segmentation followed by histogram landmark matching was proposed to remove stain variation . Ashistogram matching based methods do not distinguish causes of color variation, histological information is hardly preserved after normalization, consequently introducing unwanted bias into subsequent image analysis.

Color transfer: The second group of color normalizationsolutions relies on the so-called color transfer technique discussed in [5]. After converted to the color space the mean and variance in each color channel of a query image are matched to the statistics of a reference image [10]. Since images stained by multiple chemical dyes may have different color distributions, colors associated with different histological components may smear each other after color transfer. To address this problem, before color transfer, an image is divided into regions by segmentation

manually [7] or automatically through pixel classification [11], so that each region contains one type of histological objects only. For one thing, manual segmentation of images in large datasets is infeasible; For another, segmentation achieved by hard pixel classification is not reliable because one pixel in a histo-pathology image may belong to multiple objects due to histological component overlap. Hence, soft pixel classification by stain decomposition was proposed to precede color transfer [11] An improved normalization method based on work [11] was proposed in In this method, mean and ranking statistics in each decomposition channel of a test image are mapped to the statistics of a reference image nonlinearly. However, applying color transfer on decomposition results has two limitations. First, sources of color variation in histo-pathology images are lumped together and not addressed separately. Second, statistics in decomposition channels, which is closely related to histological information in images, is modified and thus tissue features may not be preserved after normalization.

Spectral matching: This class of color normalization solutions aims to remove stain variation only. To that end,algorithms belonging to this group first estimate stain spectra either using adaptive estimators [7] or via dedicated hardware [9]and then match the estimated quantity to a reference stain spectra. Approaches in this category have an advantage over other groups in the sense that if stain variation is the only source of color disagreements in images, with good estimation, spectral matching approaches can preserve histological features. However, if other causes also contribute, as algorithms in this category do not identify their effects, histological features may be modified after normalization.

Since success of spectral matching approaches heavily relies on the accuracy of stain spectral estimation, a very brief review on it is presented here. Spectral estimation, or stain decomposition, is a process that estimates stain spectra and corresponding stain proportions at each pixel in a histopathology image. Early works of stain decomposition can be traced back to [5], where stain spectra are pre-determined by experiments. To address spectral variation in stains, spectra were estimated at regions of interest by manual selection [8]. However, such manual operation is time-consuming for large datasets. For adaptive stain decomposition, sparseness analysis followed by relative newton method was proposed to achieve blind source separation [3]. Since this method is designed for hyperspectral images whose channel number is much larger than the number of stain types on biopsies, it is not applicable to RGB images. Based on an imaging model where image colors were linear combinations of stain spectra in the optical density domain, a plane fitting method was achieved by singular value decomposition and thresholding for images containing two stains only [7]. For accurate estimation of weak stains, prior knowledge on stains is used in the plane fitting process .However, their performance vary as the pre-fixed thresholds for stain estimation may be inappropriate for a testing image. For accurate estimation, spectra of chemical stains were detected using dedicated hardware ,whereas dependence on devices limited their adoption. Later, NMF was used for spectral estimation [4], [1]. Since both works don't address an inconsistent convergence issue, stain decomposition may converge to any local minima, leading to wrong estimation. Blind color decomposition (BCD) is achieved by performing expectation-maximization on color distributions in the Maxwell color triangle [2]. Though a heuristic randomization function tries to select stable colors for estimation, BCD method is prone to be affected by achromatic pixels when estimating weak stains' spectra. Recently, spectral estimation via supervised learning on a training set of histopathology images was proposed in [11]. As the learning model relies on statistics of images in the training set, stain estimation may not be accurate, or even fail, when spectral variation in stains occurs between a query image and training images.

3. PROPOSED COLOR NORMALIZATION SCHEME

The block diagram of the proposed normalization approach consisting of two processes is shown in Fig. 1:The off-line process defines a standard histo-pathology image preparation condition including information For operational flexibility, information on the standard condition can be defined via either predetermined quantities explicitly In the diagram, a reference image Is is used, where a standard imaging illuminant and stain spectra are obtained by illuminant estimation and spectral estimation, respectively. Note, to qualify an image to be a reference, it must contain the same types of stains occurring in a query image, and have image background and histological components stained by different chemical dyes clearly presented. Otherwise, inaccurate, or even irrelevant, quantities are estimated in this off-line process, finally affecting overall normalization performance.

The on-line process normalizes colors in a query image I. For each query image, the imaging illuminant and stain spectra are estimated and matched to reference quantities generated by the off-line process. Since side

information on query images' preparation, such as information about imaging device, is hardly guaranteed for large image sets, our work focuses on an operational scenario where no side information on a query image, other than knowledge about the stain type, is available to the system.

3.1 Illuminant normalization module

Illuminant variation, corresponding to inconsistency in biopsy imaging in this work, introduces color bias in images .To remove the color bias E(λ) of a query image should be estimated and matched to a standard Es(λ), which is either estimated from a reference image or defined by a predetermined quantity such as the CIE illuminant D65 [6]. However, estimation of E(λ) directly from an image is relatively difficult due to the integral effect of a camera. Based on the theory of metamer [2], we deduce an equivalent intensity matching algorithm to achieve illuminant normalization.

Illuminant Matching*:* Following Eqn. (1), intensities of an image generated under illuminant with standard SPD are

$$I(p, \lambda_i) = \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) E(\lambda) e^{-M_i(\lambda) D(p)} d\lambda \quad (1)$$

$$I_S(p, \lambda_i) = \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) Es(\lambda) e^{-M_i(\lambda) D(p)} d\lambda \quad (2)$$

Applying the first mean value theorem for integration [8] to Eqn. (1) and Eqn. (2), we get

$$I(p, \lambda_i) = E(\varepsilon 1) \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) e^{-M_i(\lambda) D(p)} d\lambda \quad (3)$$

$$I_S(p, \lambda_i) = E_s(\varepsilon 2) \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) e^{-M_i(\lambda) D(p)} d\lambda \quad (4)$$

$$I(p, \lambda_i) / I_S(p, \lambda_i) = E(\varepsilon 1) / E_s(\varepsilon 2) \quad (5)$$

Since backgrounds, or blank areas, of histo-pathology images correspond to tissues that are not bound by any stains, D(p) = 0. Hence,

$$I^b(\lambda_i) = E(\varepsilon 1) \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) \, d\lambda \quad (6)$$

$$I_s^b(\lambda_i) = E_s(\varepsilon 2) \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) \, d\lambda \quad (7)$$

$$I_s(p, \lambda_i) / I(p, \lambda_i) = I_s^b(\lambda_i) / I^b(\lambda_i) \quad (8)$$

Eqn. (8) is a formula to remove color disagreement caused by illuminant variation in histo-pathology images. Although it is simple, intuitively appealing and probably easy to infer, to the best of our knowledge, this work constitutes the first attempt to rigorously derive it using an imaging model.

Illuminant Estimation: According to Eqn. (8), image intensities in blank areas need to be estimated for illuminant normalization. As D(p) = 0 in image blank areas,

$I^b(\lambda_i) \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) \, d\lambda > \int_{\lambda_i-\delta}^{\lambda_i+\delta} f_i(\lambda) E(\lambda) e^{-M_i(\lambda) D(p)} d\lambda$ Therefore, after removing image noise (e.g. shot noise) by a N-by-N square-shaped median filter hN, the largest pixel value in the ith color channel is picked to estimate $I^b(\lambda_i)$ .We formulate this illuminant estimation approach as

$$I^b(\lambda_i) = \max[I(p, \lambda_i) * h_N] \quad (9)$$

where * denotes a convolution operation. Though most histo-pathology images have blank areas, to ensure system robustness, we propose the use of a threshold to identify images .

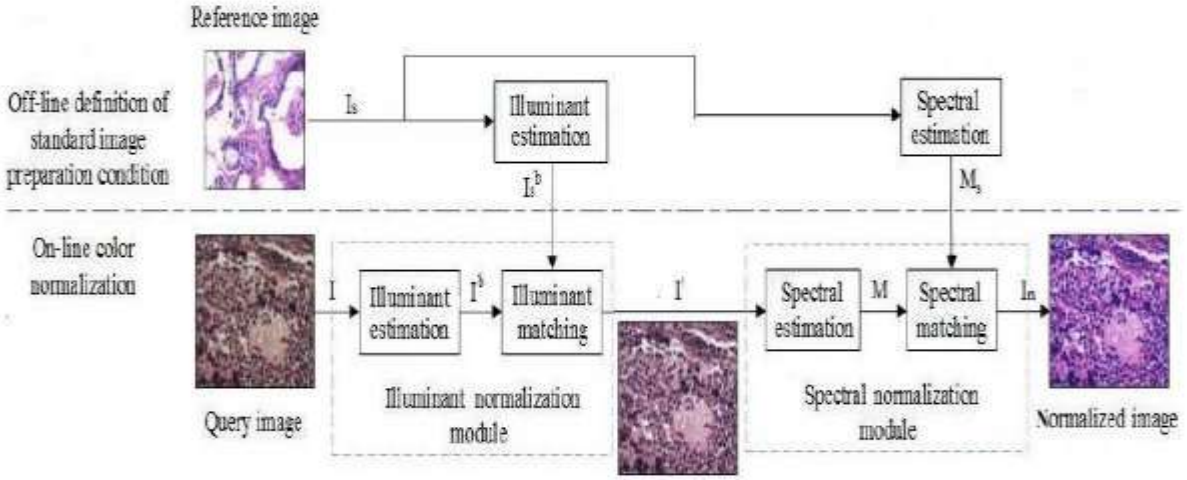

Fig -1 : Diagram of the proposed normalization pipeline

3.1 Spectral normalization module

If a set of biopsy samples stained by the same types of chemical dyes are imaged using one scanner, color variation in images are mainly caused by disagreement in biopsy staining. Though stains have their own diagnostic colors, such as hematoxylin usually appearing in blue while eosin in pink visually, due to inconsistency in stain manufacture, stain concentration, or storage condition, absorption spectra of stains, denoted as Mi(λ) in Eqn. (1), may differ, resulting color variation in images. Therefore before quantitative analysis, color variation caused by inconsistent stain spectra Mi(λ) should be removed.

When normalizing spectral variation in stains, care should be taken to preserve histological information. Though color variation among images is caused by different factors, within one image, these factors are constant and histological information is conveyed by colors. As various colors in an image are generated by different combinations of stains, histological information is actually conveyed by stain depths at each pixel, which is denoted by D(p) in Eqn. (1). To normalize inconsistent stain spectra Mi(λ) while to maintain stain proportion D(p) unchanged, we concatenate a NMF based spectral estimation and spectral matching.

4. EXPERIMENTAL RESULTS

We design three experiments to evaluate the proposed normalization approach. First, we assess the robustness of our normalization method to system settings, and demonstrate that our method is stable under various parameter settings. In the second experiment, since spectral estimation plays a significant role in spectral normalization, effectiveness and consistency of the proposed stain decomposition is examined and comparedto state-of-the-art adaptive stain estimation algorithms. In the last experiment, performance of the proposed normalization approach ,especially the capacity of histological information preservation, is evaluated. This experiment is also performed on 4 representative normalization approaches for comparison. Two public histo-pathology image sets are used as experimental data in this work. They are selected in different experiments depending on experimental objectives.

The NIA malignant lymphoma dataset can be accessed from the IICBU Biological Image Repository [6]. It contains 20x magnified H&E stained images with resolution of 1388 * 1040, including 113 chronic lymphocytic leukemia images, 140 follicular lymphoma cases, and 122 mantle cell lymphoma slides. Because biopsies in this dataset were prepared by different pathology laboratories, significant variations in image colors are observed. This dataset is believed more representative of histo-pathology images commonly received in clinics [4].

The UCSB breast cancer cell dataset [3] was published for an objective of cell segmentation in both benign and malignant cell images. The dataset consists of 26 cancerous cell images and 32 normal cases cut from 10 H&E stained breast cancer biopsies. All images are scanned in the same laboratory, with resolution of 896 * 768.

4.1 Experiment 1:System robustness to parameter settings

There are two sets of parameters pre-determined in our normalization method: median filter dimension in the illuminant normalization module, and NMF convergence condition in the spectral estimation algorithm. Usually, performance of an algorithm varies with parameter selection. In medical-related applications, we want this performance variation small. Hence, we perform two tests in this experiment and assess sensitivity of the proposed approach to the two sets of parameters. The objective is to assess sensitivity of the illuminant normalization module to the median filter with neighbor-supporting in N by N square area.

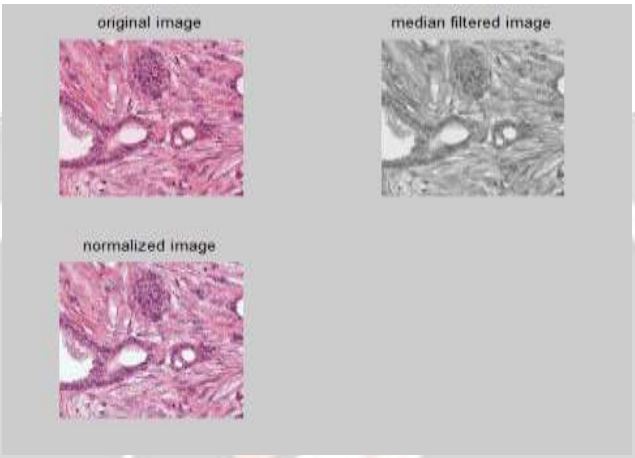

Fig -2 : Illuminant normalized image

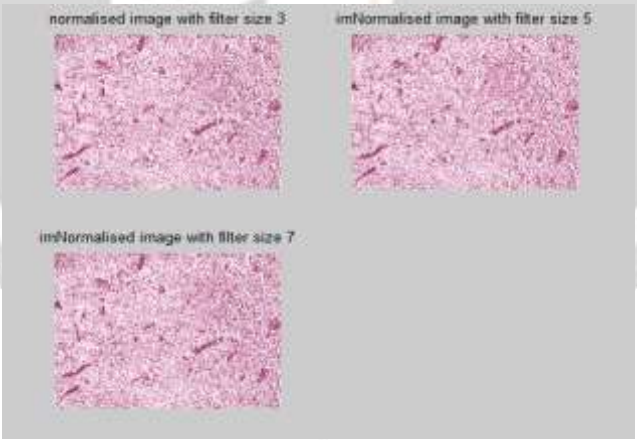

Fig -3 : An example of illuminant normalization with median filters

4.2 Consistency of spectral estimation

To identify contribution of each stain to color variation, accurate spectral estimation is crucial. As histo-pathology images may contain different contents, we want the spectral estimation is not affected by content difference. Therefore, this experiment evaluates consistency of our spectral estimation against different image contents.

Dataset: The UCSB dataset is selected as an evaluation set for reasons as follows. The UCSB images were cut from 10 biopsies. For images from the same biopsy, a spectral estimation algorithm should generate consistent, or at least similar, spectrum matrices. Thus, we use consistency of spectrum matrices estimated from images cut from the same biopsy as a metric to measure performance of a stain estimation algorithm. **:** We form 10 sets of images in the

UCSB dataset, each containing all images from the same biopsy. Then standard deviation _ of estimated matrices in each set is computed. A smaller indicates more consistent estimation. This experiment is also performed on state-of-the art blind stain estimation algorithms: the plane fitting approach [7] and BCD approach [10].

Spectral estimation results of the first image set. Since similar results were obtained, estimation results of other 9 sets were omitted here. As the first set contains six H&E stained images cut from a biopsy labeled as ytma10-010704, each approach generates 6 pairs of spectrum vectors, among which spectra of hematoxylin are depicted in the left panel and eosin spectra are in the right panel. In the figure, both hematoxylin and eosin spectrum vectors estimated by the proposed method form tight clusters, while estimations computed by the plane fitting approach show the largest diversity. Summarizes standard deviations of spectral estimation over the 10 image sets, where the smallest of each matrix element is marked black. The data illustrate that our spectral estimation method generates more consistent results than the other two methods. The relatively inconsistent performance of the plane fitting method [7] is attributed to the thresholding mechanism in stain estimation, where only image pixels whose colors are close to pre-determined thresholds contribute to spectral estimation.

However, for images containing different contents, the predetermined values may be inappropriate, resulting inaccurate and inconsistent spectral estimation. The estimation consistency of the BCD algorithm [25] lies between the plane fitting method and our method. In the BCD approach, a heuristic randomization function assigns large weights to colors in high optical density values. Then colors with large weights are selected for stain estimation. Compared to colors associated with hematoxylin that are usually in larger optical density values, eosin-stained tissues often have weak stains, and thus colors mainly associated with eosin usually have smaller weights. For images containing large blank areas, significance of eosin-stained pixels may be overwhelmed by achromatic colors. This explains the less consistency in eosin spectrum vectors estimated by the BCD algorithm. The consistent performance of our estimation method is attributed to the proposed saturation-weight statistical method. Since SW-statistics identifies saturated colors implicitly, impacts of achromatic pixels and noise on spectral estimation are alleviated. The strong noise-resistant capability of the proposed algorithm inherited from the SW-statistics ensures consistency of spectral estimation, and will benefit effectiveness and reliability of the proposed normalization method.

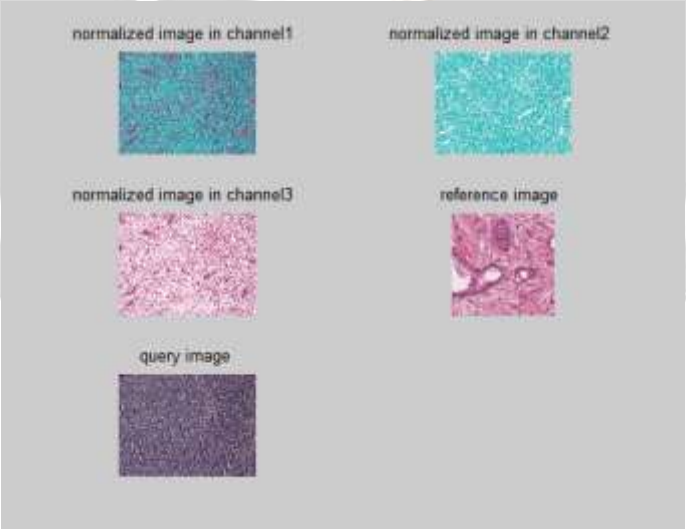

Fig -4 : Normalized image in three channels

5. CONCLUSION AND FUTURE SCOPE

This work introduced a robust and complete color normalization approach capable of addressing color variation in histopathology images. Based on an imaging model, the introduced solution was able to identify the source of color variation, and addressed effectively both illuminant variation and stain variation using an intensity matching algorithm and a spectral normalization module respectively. Extensive experimentation on publicly

available datasets indicates that the proposed solution outperforms state-of-the-art color normalization solutions, while preserving histological information.

The saturation-weighted stain estimation method introduced by this work limited impacts of achromatic spectra on stain estimation and robustified the estimation process. The experimental results of stain estimation indicate that the proposed method delivers superior consistent performance compared to state-of-the-art blind stain decomposition solutions.

Breast Cancer is one of the severe diseases causing large number of deaths in women. So there is a need for efficient technique that diagnoses such cells without the involvement of human, with high accuracies. In the first phase, a complete normalization scheme is introduced to address the problem of color variation in images jointly caused by inconsistent biopsy staining and non-standard imaging condition. This scheme is robust to parameters and insensitive to content of the image and achromatic colors. In second phase I would like to modify this project to do, a Feed forward back propagation neural network classify benign and malignant tumor and also classify breast cancer tumor in type1, type2 and type3. It can be conclude that it will works as promising tool for classification of cancer cells. Breast Cancer is one of the severe disease causing large number of deaths in women. So there is a need for efficient technique that diagnoses such cells without the involvement of human, with high accuracies. In the first phase, a complete normalization scheme is introduced to address the problem of color variation in images jointly caused by inconsistent biopsy staining and non-standard imaging condition. This scheme is robust to parameters and insensitive to content of the image and achromatic colors. Then In second phase, a Feed forward back propagation neural network classify benign and malignant tumor and also classify breast cancer tumor in type1, type2 and type3. It can be concluded that it will works as promising tool for classification of cancer cells.

6. ACKNOWLEDGEMENT

I thank God almighty for all the blessings received during this endeavour. I would like to offer my deepest gratitude to everybody who helped me directly and indirectly for the successful completion of my work. Last, but not the least, I thank all my friends and my family for the support and encouragement they have given me during the course of my work.

6. REFERENCES

[1]. Xingyu Li, Student Member, IEEE, and Konstantinos N. Plataniotis, Fellow, IEEE,"A Complete Color Normalization Approach to Histo-pathology Images Using Color Cues Computed From SaturationWeightedStatistics",IEEE Transactions on Biomedical Engineering

[2]. A. Basavanhally et al., "Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology," IEEE Trans. Biomed. Eng., vol. 57, no. 3, pp. 642 – 653, Mar. 2010.

[3]. K. Nguyen et al., "Structure and context in prostatic gland segmentation and classification," in Proc. MICCAI, 2012

[4]. L. Gorelick et al., "Prostate histopathology: Learning tissue component histograms for cancer detection and classification,"IEEE Trans. Med. Imag., vol. 32, no. 10, pp. 1804 – 1818, Oct. 2013

[5] M. Veta et al., "Breast cancer histopathology image analysis: A revier ,"IEEE Trans. Biomed. Eng., vol. 61, no. 5, pp. 1400 – 1411, May 2014.

[6] M.N. Gurcan et al., "Histopathological image analysis: a review,"IEEE Rev. Biomed. Eng., vol. 2, pp. 147 – 171, 2009.

[7] M. Macenko et al., "A method for normalizing histology slides for quantitative analysis," in Proc. ISBI, 2009.

[8] A. Tabesh et al., "Multifeature prostate cancer diagnosis and gleason grading of histological images," IEEE Trans. Med. Imag., vol. 26, no.10, pp. 1366 – 1378, Oct. 2007.

[9] S. Kothari et al., "Automatic batch-invariant color segmentation of histological cancer images," in Proc. ISBI, 2011.

[10] Y. Wang et al., "A color-based approach for automated segmentation in tumor tissue classification," in Proc. EMBC, 2007.

[11] D. Magee et al., "Color normalization in digital histopathology Images," in Proc. MICCAI OPTMHisE, 2009.

BIOGRAPHIES

AMAL KUMAR R.G received the BTECH degree in Applied electronics and instrumentation engineering from University of Calicut in 2015.He is currently working towards the MTECH degree in Digital signal processing at the A.P.J Abdhul Kalam Technological University, Kerala, India .His research interests include biomedical image processing, pattern recognition ,human computer interface, cyber forensics etc.

ARUN LAL C.K received the BE degree in Electronics and communication engineering from Anna University in 2009.He received the MTECH degree in Applied electronics from Anna University in 2014.He is currently an Assistant professor at the department of Electronics and Communication at KMCT college of engineering, Calicut, Kerala, India .His research interests include biomedical signal processing, pattern recognition.