Color model comparative analysis for breast cancer diagnosis using H&E stained images

Xingyu Li, Konstantinos N. Plataniotis

Multimedia Lab, The Edward S. Rogers Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario, M5S 3G4, Canada
xingyu.li@mail.utoronto.ca, kostas@comm.utoronto.ca

ABSTRACT

Digital cancer diagnosis is a research realm where signal processing techniques are used to analyze and to classify color histo-pathology images. Different from grayscale image analysis of magnetic resonance imaging or X-ray, colors in histopathology images convey large amount of histological information and thus play significant role in cancer diagnosis. Though color information is widely used in histopathology works, as today, there is few study on color model selections for feature extraction in cancer diagnosis schemes.

This paper addresses the problem of color space selection for digital cancer classification using H&E stained images, and investigates the effectiveness of various color models (RGB, HSV, CIE $L^*a^*b^*$, and stain-dependent H&E decomposition model) in breast cancer diagnosis. Particularly, we build a diagnosis framework as a comparison benchmark and take specific concerns of medical decision systems into account in evaluation. The evaluation methodologies include feature discriminate power evaluation and final diagnosis performance comparison. Experimentation on a publicly accessible histopathology image set suggests that the H&E decomposition model outperforms other assessed color spaces. For reasons behind various performance of color spaces, our analysis via mutual information estimation demonstrates that color components in the H&E model are less dependent, and thus most feature discriminate power is collected in one channel instead of spreading out among channels in other color spaces.
Keywords: Color spaces, discriminant power, breast cancer classification, H&E model

1. INTRODUCTION

With latest advances of whole slide imaging, high-resolution color images of biopsy samples are accessible. Under such progress, digital histo-pathology has received considerable attention in hope of accurate cancer diagnosis based on the images. By applying image processing algorithms and pattern recognition methods, computers are expected to understand histo-pathology images and to make diagnosis decisions.

Thanks to the chemical staining process in biopsy sample preparation routine, color in histo-pathology images conveys large amount of information and color processing plays a significant role in digital cancer diagnosis. For instance, to grade prostate cancer, color texture features were extracted from either RGB color space [1] or the intensity channel of HSI color space [2]. Despite promising results reported, there is no specific works on the efficiency of color spaces for digital cancer diagnosis, and the issue "what is the appropriate color model for cancer diagnosis" is rarely addressed. Though several stain-dependent color spaces such as hematoxylin and eosin (H&E) model were proposed and derived from various stain decomposition methods [3,4,5], none of them evaluated the performance in terms of cancer diagnosis. In the study of classification of three types of malignant lymphoma, RGB, $L^*a^*b^*$, and H&E spaces were selected [6]. However, the focus of this work was not color model comparison, and only classification accuracy was reported which is limited [7]. Later, the work in [8] compared RGB, HSI, CMYK, $L^*a^*b^*$, and HSD color spaces for histological component segmentation in histo-pathology images. Since segmentation is one composition block of a cancer diagnosis scheme, conclusions drawn on it have little guidance on color model selection for the disease classification scenario.

Due to the well-established methods and successful use of H&E staining in histopathology, it is believed that the H&E stains will continue to be the most common practice in histopathology in future [9]. Therefore, this paper addresses the issue of color space selection for feature extraction in digital cancer diagnosis on H&E stained histopathology images.

We analyze four most widely used color spaces, RGB, HSV, CIE $L^*a^*b^*$, and stain decomposition model [3] (or H&E model), in this research using a public breast cancer cell image set. In detail, a typical cancer classification framework is built as the comparison benchmark. The evaluation is based on discriminant power of extracted features and final classification performance. Particularly, our experiment design considers the unique issues in medical decision systems such as high sensitivity level, which makes our study complete and reliable. Our results show that H&E model outperforms other assessed color models due to its independent property between color bands.

The rest of this paper is organized as follows. H&E stained images, diagnosis scheme and evaluation methods are specified in Section 2. Simulations and discussions are given in Section 3 and Section 4, respectively. Conclusions are finally drawn in Section 5.

2. METHODS

2.1 Materials

We use the UCSB breast cancer cell dataset [10] as analysis images in this work. This histopathology image set was published by center for bio-image informatics, University of California, Santa Barbara. The objective of this data set was accurate cell segmentation for subsequent classification in benign and malignant cells. The dataset consists of 26 cancerous cell images and 32 normal cases cut from 10 H&E stained breast cancer biopsies prepared by David Rimm's Laboratory at Yale. All images are stored in 24-bit RGB format with resolution of 896*768. As shown in Fig.1, the nuclei are stained blue by hematoxylin while cytoplasm and extra-cellular components are in pink due to eosin staining. As images in this dataset were prepared by the same pathology laboratory, there is no spectral variation in imaging illuminant observed. This simplifies our analysis since no illuminant normalization is required to process images in this dataset.
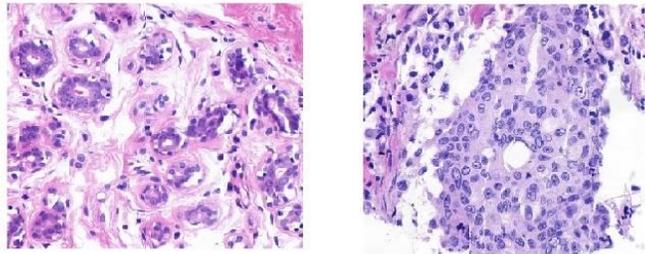

Fig. 1. Histopathology image samples in the UCSB breast cancer cell dataset. The image in the left panel is associated with benign case, while the right panel shows a malignant breast cancer cell image.

2.2 Cancer diagnosis pipeline

Digital cancer diagnosis is essentially a classification problem. By extracting a set of discriminant features from a color histo-pathology image and comparing it to information in the training set, a classification system gives a positive or negative diagnosis. In this work, we set up a diagnosis framework as shown in Fig. 2, which consists of

- Pre-processing module. The nuclei segmentation block is achieved by a watershed based algorithm for subsequent morphology analysis and structural analysis. As images in the UCSB dataset are stored in RGB format, the color space transform block converts a query image to other color spaces for texture analysis. In specific, the color models compared in this work includes the RGB, HSV, $L^*a^*b^*$, and stain decomposition model, namely H&E model achieved by the decomposition algorithm proposed in [3].

- Feature extraction module. The features extracted include (1) nuclei density, proximity, and number of nuclei in neighborhoods in morphology analysis, (2) statistics (mean, variance, disorder, and min-to-max ratios) of polygon/triangle areas and perimeters in voronoi graph/Delaunay triangulation, and of path length in minimal spanning tree calculated in structural analysis, and (3) greylevel features(mean, variance, median, mode, and range of intensity in blocks of various sizes), gabor filtering features(mean, variance, mode, and power of values in blocks of various sizes from each filter bands) and haralick indexes (energy, contrast, correlation, homogeneity, and entropy of co-occurrence matrix constrained by different directions and distances) by texture

analysis. In texture analysis, each color channel is considered as an independent image, and texture features extracted are concatenated as a long feature vector.

- Dimension reduction module. Similar to the work in [11], a non-linear projection technique, so-called spectral clustering or graph embedding [12], is adopted to reduce the feature space to a low-dimension domain.

- Classification module. Considering its good classification performance in complicated problems, a support vector machine (SVM) is used to classifier benign and malignant breast cancer cell images.
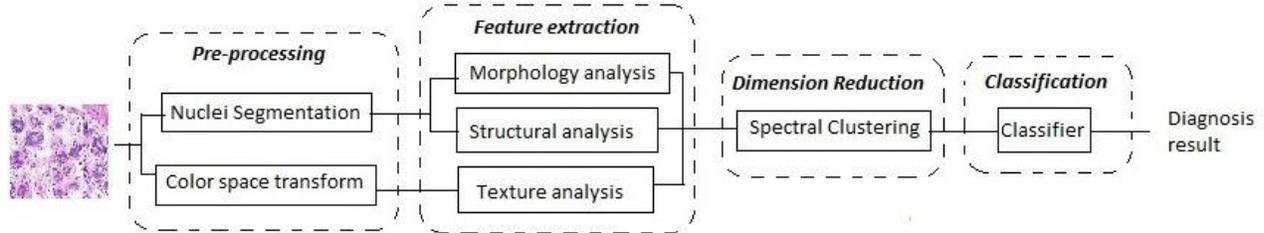

Fig. 2. Breast cancer diagnosis pipeline used in this work.

2.3 Evaluation methodology

The evaluation of the effectiveness of different color spaces in H&E stained breast cancer diagnosis is divided into two parts. In the first experiment, a classifier-independent comparison is performed to examine discriminant power of feature sets extracted from the four color spaces. The more separable the feature sets between normal and breast cancer images, the larger discriminant power of feature set have.

In the second experiment, we evaluate the effectiveness of color space in terms of classification performance. In specific, the entire image set is split into two parts, with 1/10 as testing cases, and 9/10 as training images. To avoid data drift in evaluation, we adopt the data split methodology of stratified cross-evaluation, where both training and testing sets maintain the same cancer prevalence as the original dataset. In SVM training, grid-search [13] is performed to select parameters of SVM Gaussian kernel function for optimal classification. After finding the best set of parameters, the whole training set is used to generate a final SVM classifier. This classification process will be repeated 5 times, therefore yielding 50 performance indexes. The agreements between ground truth and diagnosis results are estimated and compared in terms of two metrics as follows:

- Classification accuracy (ACC). ACC represents the probability of a correct diagnosis for a query image. It is estimated by the number of accurate diagnosis over all query images in a testing set. Due to its intuition and easy computation, ACC is conventionally adopted in literature [2,11,14].

- pAUC. pAUC is defined as the area under the ROC curve given an acceptable minimal hit rate p [15]. It is usually in the range of $[(1+p)(1-p)/2,1]$, where the lower bound corresponds to a random classification. In diagnostic medicine research, disease detection systems usually require high hit rates as a penalty of missing a disease may be a human's life. Considering the high sensitivity requirement in cancer diagnosis particularly, we use $_{90}AUC$ as another performance metric.

3. RESULTS

3.1 Experiment 1: discriminant power of feature sets

Evaluation of effectiveness of assessed color models is first based on a classifier-independent comparison. As the dimensions of extracted feature are as high as to 3000 for the RGB, HSV, $L^*a^*b^*$ color models and 2000 for the H&E decomposition domain, to visualizes discriminant power of feature sets extracted from the four color spaces, we use spectral clustering to reduce the feature set dimension to 2, and depict the projected feature sets in Fig. 3, where the axes of the plots correspond to the two dominant eigenvectors found by spectral clustering for the high-dimension feature set. As labeled in the figure, stars and circles denote benign breast cancer images and malignant cases, respectively.

As shown in Fig. 3, manifolds of numerical features of normal and breast cancer images extracted from the RGB, HSV, and $L^*a^*b^*$ models are mixed in various degrees, while features exacted from H&E decomposition channels show a good separation between benign and malignant cancer classes. This suggests that the numerical features extracted from the H&E decomposition domain reflects more histological information conveyed by images.
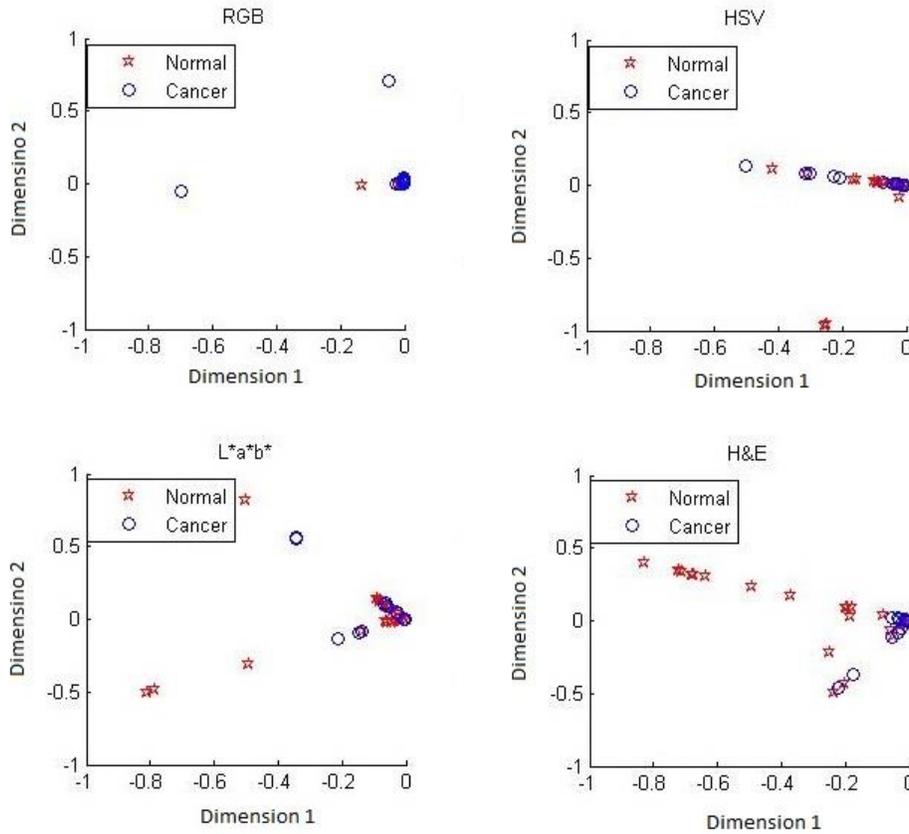

Fig. 3. Features extracted from different color models are projected into a 2-dimension spaces by spectral clustering to visualize their discriminate power. In the figure, a star represents a benign breast cancer case, and a circle corresponds to a malignant breast cancer image.

3.2 Experiment 2: Classification of normal and breast cancer images

Table 1 lists the comparative results based on classification performance over the UCSB dataset. For both performance indexes, ACC and $_{90}$AUC, the best diagnosis of breast cancer are marked black. Though the classification accuracy associated with color spaces RGB and $L^*a^*b^*$ are close to the ACC of H&E model, it should be note that for a medical diagnosis, even a 1% increase of diagnosis accuracy is significant. With respect to $_{90}$AUC which addresses the high sensitivity required in medical diagnosis and thus is considered as a more comprehensive metric, the H&E model achieves more than 20% increase and thus outperforms other assessed color models in this work.

4. DISCUSSIONS

To investigate the reasons behind better performance of H&E decomposition model over other color spaces, we trace back to the formation of an H&E stained image. According to Beer-Lambert law, colors in an H&E stained image are generated by superposition of two independent vectors, the absorbing spectra of hematoxylin and eosin, in optical

density domain [3]. Therefore, to convert a color represented in the RGB color to H&E decomposition model, a reverse operation, called color decomposition, is needed to separate contributions of hematoxylin and eosin. As the two chemical dyes attach different histological components, color components in H&E model should share small amount of histological information. To verify our analysis, mutual information among color channels are estimated over the UCSB dataset and shown in Fig. 4. The much smaller mutual information in H&E model implies that the channels in H&E model are less dependent on each other than the channels in other color spaces. Consequently, discriminant power of features computed from H&E color model tends to collect in one dimension; whereas considering other color spaces, discriminant power of features spreads in all channels due to a convolution effect in optical density domain during image capture, leading to worse classification performance.

Table 1. Classification performance of breast cancer diagnosis using numerical features extracted from different color spaces. The best diagnosis in terms of performance indexed are marked black.

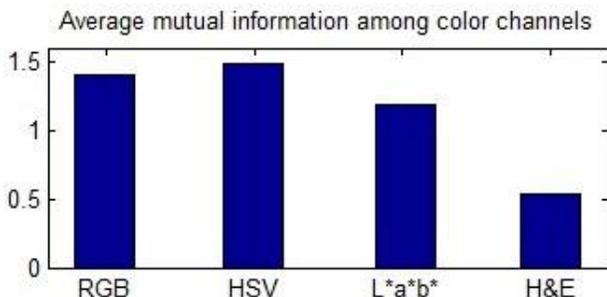| Color model	ACC		$_{90}$AUC	
	mean	std	mean	std
RGB	0.780	0.0075	0.411	0.0901
HSV	0.677	0.0184	0.356	0.0931
L*a*b*	0.763	0.0043	0.409	0.0905
H&E	**0.802**	0.0284	**0.6594**	0.1139

Fig. 4. Average mutual information between color bands of different color spaces computed from the UCSB dataset.

Based on our experimental results and observation, we find that H&E decomposition is a more appropriate color model for feature extraction in cancer diagnosis on H&E stained images. To generalize our observations to other large H&E stained histopathology image set, two issues should be noted as follows.

- First, different from the UCSB dataset whose images were prepared in the same laboratory, histopathology images in large H&E stained image set may come from different pathology laboratory and color variations may be observed. To obtain an accurate H&E decomposition, advance adaptive stain decomposition methods, such as algorithms in [5,16,17], should be used instead.

- Second, this work is performed on the UCSB breast cancer dataset containing 58 images. Further experiments can be carried on other H&E dataset associated with other cancers. For example, we will confirm the observation in this work using a public malignant lymphoma image set provided by NIH [18].

5. CONCLUSION

This work addresses the problem of color model selection for digital cancer diagnosis on H&E stained images. By taking the issue of high diagnosis sensitivity in cancer classification into account, this work evaluates the effectiveness of four color spaces, RGB, HSV, $L^*a^*b^*$, and stain-dependent H&E decomposition model, in breast cancer diagnosis. The experimentation suggests that H&E model demonstrates good performance in terms of both feature discriminant power and final diagnosis results. To investigate the reasons of why H&E model outperforms other color spaces, we trace back to the formation of H&E stained image and estimate mutual information between color channels in different color spaces. Our analysis shows that H&E model possess significantly smaller mutual information and thus enhances discriminate power of extracted features.

REFERENCES

[1] Tabesh,. A, Teverovski, M., Pang, H.-Y., Kumar, V.P. , Verbel, D. , Kotsianti, A., Saidi, O., "Multifeature prostate cancer diagnosis and gleason grading of histological images," IEEE Trans. Medical Imaging, 26(10), 1366-1378 (2007).

[2] Doyle, S., Feldman, M.D., Tomaszewski, J.E., and Madabhushi, A., "A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies," IEEE Trans. Biomed. Eng., 59(5), 1205-1218 (2012).

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

[4] Rabinovich, A., Agarwal, S., Laris, C., Price, J. H., and Belongie, S. J., "Unsupervised color decomposition of histologically stained tissue samples," *Advances in Neural Information Processing Systems (2003).*

[5] Gavrilovic, M., Azar, J. C., Lindblad, J., Wahlby, C., Bengtsson, E., Busch, C., and Carlbom, I. B. "Blind color decomposition of histological images, " *IEEE transactions on medical imaging*, *32*(6), 983-994 (2013).

[6] Orlov, N. V., Chen, W. W., Eckley, D. M., Macura, T. J., Shamir, L., Jaffe, E. S., & Goldberg, I. G., "Automatic classification of lymphoma images with transform-based global features, "*, IEEE Transactions on Information Technology in Biomedicine*, *14*(4), 1003-1013 (2010).

[7] Huang, J., & Ling, C. X., "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299-310 (2005).

[8] Bueno, G., Déniz, O., Salido, J., Fernández, M. M., Vállez, N., and García-Rojo, M., "Colour model analysis for histopathology image processing," In *Color Medical Image Analysis*, Springer, 165-180 (2013).

[9] Fox, H., "Is H&E morphology coming to an end?" Journal of Clinical Pathology, 53(1), 38-40 (2000).

[10] Gelasca, E.D., Obara, B., Fedorov, D., Kvilekval, K., and Manjunath, B.S., "A biosegmentation benchmark for evaluation of bioimage analysis methods," Proc. BMC Bioinformatics, 10:368 (2009).

[11] Doyle, S., Agner, S., Madabhushi, A., Feldman, M.D., and Tomaszewski, J.E., "Automated grading of breast cancer histopathology using spectral clustering with texture and architectural image features," Proc. IEEE Int. Symp. Biomed. Imag., Nano Macro, 496-499 (2008).

[12] Madabhushi, A., Shi, J., Rosen, M., Tomaszeweski, J. E., and Feldman, M. D., "Graph embedding to improve supervised classification and novel class detection: application to prostate cancer, " Proc. MICCAI, 8(pt. 1), 729-737 (2005).

[13] Chang, C.C., and Lin. C.J., "LIBSVM: a library for support vector machines," ACM Trans. Intrll. Syst. Technol., 2(3), 1-27 (2011).

[14] Ozdemir, E., and Gunduz-Demir, C. "A hybrid classification model for digital pathology using structural and statistical pattern recognition," IEEE Trans. Med. Imag., 32(2), 474-483 (2013).

[15] Dwyer, A.J., "In pursuit of a piece of the ROC," Radiology, 201(3), 621-625 (1996).

[16] Niethammer, M., Borland, D., Marron, J. S., Woosley, J., and Thomas, N. E., "Appearance normalization of histology slides," Machine Learning in Medical Imaging, 58-66 (2010).

[17] Khan, A.M., Rajpoot, N. ; Treanor, D. ; Magee, D., "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," IEEE Trans. Biomed. Eng., 61(6), 1729-1738 (2014).

[18] Shamir, L., Orlov, N., Eckley, D. M., Macura, T. J., and Goldberg, I. G., "IICBU 2008: a proposed benchmark suite for biological image analysis," Med. Biol. Eng. Comput., 64(9), 943-947 (2008).