IMAGE ANALYSIS FOR AUTOMATED ASSESSMENT OF GRADE OF NEUROBLASTIC DIFFERENTIATION

Jun Kong[1, 2], Hiroyuki Shimada[3], Kim Boyer[1], Joel Saltz[2], Metin Gurcan[2]

[1,2]The Ohio State University
[1]Dept. of Electrical and Computer Engineering
[2]Dept. of Biomedical Informatics
Columbus, OH

[3]The University of Southern California
Dept. of Pathology and Laboratory Medicine
Childrens Hospital Los Angeles
Los Angeles, CA

ABSTRACT

Peripheral Neuroblastic Tumors (pNTs) make the most commonly encountered tumor groups in children. Neuroblastoma, one of the categories in pNTs, is known to have unique biological behaviors with variable clinical prognoses of the patients. Part of the neuroblastoma prognosis is closely related with grade of neuroblastic differentiation. In this work, we present an automatic classification system that includes a novel segmentation method using the EM algorithm with the Fisher-Rao criterion as its kernel. This is followed by a classification stage with classifiers applied to the actual neuroblastoma images. The good classification accuracy suggests that the developed method is promising in automating this pathological assessment.

Keywords*: Neuroblastoma, Neuroblastic Differentiation, Image Segmentation, Pattern Classification, Color Space*

1. INTRODUCTION

Peripheral Neuroblastic tumors (pNTs) are a group of embryonal tumors of the sympathetic nervous system that can be further classified into neuroblastoma, ganglioneuroblastoma (intermixed), ganglioneuroma and ganglioneuroblastoma (nodular) in terms of such measures as maturation, malignancy and constituents [1-2]. Finding a robust yet efficient way of classifying NTs based on their pathological characteristics and clinical behavior is a challenging problem. The current prognosis is manually done by pathologists, which may lead to inter- and intra-variability. Although there are several qualitative criterion made to facilitate the prognosis and classification of NTs, it is problematic, in general, to automate this process when a large number of images are to be analyzed in real time.

Neuroblastoma prognosis is carried out using the International Neuroblastoma Classification System developed by Shimada *et al.* [2]. Grade of neuroblastic differentiation is part of this classification system. There are three categories of grades: undifferentiated, poorly differentiated and differentiating. Example images from these categories are shown in Figure 1. Clinical characteristics that are used to separate these grades are explained in Reference [1]. In this work, these features are used to automatically recognize grade of differentiation using image analysis methods.

There are few other works devoted to developing automated image analysis for neuroblastic tumors. A segmentation method based on morphological operations and the hysteresis thresholding technique is presented by Gurcan *et al.* [3] for segmenting the nuclei of the cells from neuroblastoma tissues. This work does not address detection of other cell components such as cytoplasm or neuropil. Ayres *et al.* [4] developed a method using Gaussian mixture model and the expectation-maximization (EM) algorithm to analyze the tissue composition of the tumor mass in neuroblastoma from computed tomography (CT) images. Although this method may provide pathologists with useful information on the clinical response by studying the histograms of necrotic tissue, viable tumor and calcification area, it requires accurate manual segmentation of tumor mass region, which may be difficult in a clinical setting where a large number of images need to be analyzed.

In this paper, we present a completely automated system that involves a novel segmentation method, a feature extraction technique and statistical classification.

2. SEGMENTATION AND CLASSIFICATION

2.1. Image Preprocessing

The inputs to our grading system are images of haematoxylin and eosin (H&E)-stained tumor samples digitized using ScanScope T2 digitizer (Aperio, San Diego, CA) at 40x magnification and then compressed using JPEG compression at approximately 1:40 compression ratio with the typical storage size around 1~2.5 GB each. Due to the large image sizes, we cropped 129 images at random for each grading class (i.e. 387 images in total), each of size 512×512, from sample slides associated with all the classes.

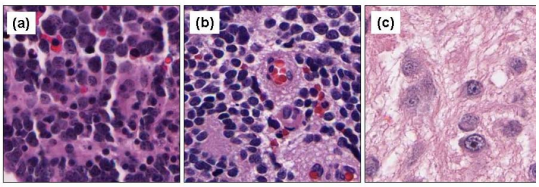

Fig. 1 Typical H&E stained images associated with (a) Undifferentiated; (b) Poorly-differentiated; (c) Differentiating neuroblastoma subtypes.

Tissues stained by H&E staining process tend to have blue-purple color and darker intensity in nuclei and cytoplasm regions, while the colors associated with neuropil and red blood cell are usually in hues of pink and red. The significant differences in color among different tissue components suggest that we can use color information to partition different tissue components using clustering methods in a well formulated feature space.

Each pixel represented by R, G and B channels in the RGB color space was converted to the LA^*B^* space developed by Commission Internationale d'Eclairage (CIE). When compared to other color spaces such as HSI, YIQ and YUV, LA^*B^* color space provides the color perceptual uniformity and the Eucledian distance, thus, can be used. The LA^*B^* space is also a good choice in terms of its ability to represent luminance and chrominance information separately. To be more precise, the channel L carries the information for the light intensity while color information is described by A^* and B^* components.

The patterns of the color change within image regions associated with different tissue structures vary considerably. Hence, textural information from the RGB color space was used to enrich the resulting feature space so as to make textural feature components as much uncorrelated as possible with the color features obtained from the LAB color space. Specifically, we move a 9×9 window across the image and compute the entropy of the center pixel in each of the R, G and B channels using:

$$E(y_{ij}^k) = -\int_{\Omega} p_{y_{ij}^k}(\omega)\log_2 p_{y_{ij}^k}(\omega)d\omega \qquad (1)$$

where $E(y_{ij}^k)$ is the entropy associated with pixel y_{ij} in the k-channel image and $k\in\{R,G,B\}$. In Equation (1), the notation ω represents any integer whose value is in the set Ω, i.e. $\omega\in\Omega=\{0,1,...,255\}$ and $p_{y_{ij}^k}(\omega)$ is the estimated PDF of ω in the given window centered at y_{ij} in the k-channel image. As a result, the feature vector is composed of six components, three of which are extracted from the color information and the other three from textural information.

2.2. Image Segmentation

Color image segmentation is one of the most fundamental yet difficult problems in computer vision and image analysis. There are many approaches available in addressing this problem, such as histogram thresholding, feature clustering, edge detection, region-based, artificial neural network and fuzzy logic-based methods [5].

After testing standard K-means and Fuzzy C-Means (FCM) segmentation algorithms, we found that their convergence speeds were slow for this particular problem. In order to have a method with competitive accuracy and efficiency, we developed a novel segmentation method that uses the Fisher-Rao criterion as the kernel of the EM-algorithm, named *EMLDA algorithm*. The EM-algorithm is typically applied to classification problems where the parameters governing the true underlying distributions are to be estimated given data with missing features. More frequently, the EM-algorithm is used to estimate the parameters of some parameterized distributions, such as the popular Gaussian models, and assign labels to data in an iterative way [4]. Instead, the EMLDA algorithm uses the Linear Discriminant Analysis (LDA) [6], a supervised classification technique, as the kernel of EM-algorithm and iteratively group data points projected to a reduced dimensional feature space in such a way that the separability across all classes is maximized.

In the generic EM-algorithm, there are two iterative steps known as Expectation (E-step) and Maximization (M-step). Let us denote $X=\{x\,|\,x\in R^p\}$, as the data set in the feature space constructed in the way stated above, where p is the dimension of the feature space (p=6 in our case), the E and M steps of our proposed algorithm (EMLDA) at each single iteration can then be described as follows.

(1) E step (Expectation):
Compute:

$$V^* = \arg\underset{V\in R^{p\times s}}{Max} J(V\,|\,\theta) = \arg\underset{V\in R^{p\times s}}{Max}\frac{\left|V^T S_B(\theta)V\right|}{\left|V^T S_w(\theta)V\right|} \qquad (2)$$

where $J(V\,|\,\theta)$ is the Fisher-Rao criterion to be maximized; V^* is a $p\times s$ matrix whose each column is a discriminant vector and $s\le C-1$. In addition, θ is the labeling configuration determined from the previous step, while S_B and S_W are the between- and within-class scatter matrices that are symmetric and positive-definite:

$$S_B(\theta) = \sum_{i=1}^{C} n_i(\theta)\left(m_i(\theta)-m(\theta)\right)\left(m_i(\theta)-m(\theta)\right)^T \qquad (3)$$

$$S_W(\theta) = \sum_{i=1}^{C}\sum_{j=1}^{n_i(\theta)}\left(x_{ij}-m_i(\theta)\right)\left(x_{ij}-m_i(\theta)\right)^T \qquad (4)$$

where C is the number of classes and n_i is the number of members in class i, $i=1,2,...,C$; m_i and m are the class and overall means respectively; Furthermore, x_{ij} represents the j^{th} member in class i.

(2) M step (Maximization):

The matrix V^* maximizing the Fisher-Rao criterion $J(V \mid \theta)$ is composed of the column discriminant feature vectors with which the set of data X are projected, by Equation (5), to $\widetilde{X} = \{\widetilde{x} \mid \widetilde{x} \in R^s\}$ in a lower dimension space where the C class data can be best discriminated:

$$\widetilde{x} = x^T V^* \tag{5}$$

Next, find:

$$\widetilde{\theta} = \arg \underset{i \in \Theta}{Max} \left(- \left\| \widetilde{x} - \widetilde{m}_i \right\|_2 \right) \tag{6}$$

where $\Theta = \{1,2,...C\}$ is the label set and $\widetilde{m}_i = m_i^T V^*$.

After finding the labels $\widetilde{\theta}$ for all data points, we substitute θ with $\widetilde{\theta}$ and repeat steps (1) and (2) again, until $J(V^* \mid \theta)$ starts decreasing after a minimum number of iterations (10 in our application). As compared to the label change, $J(V^* \mid \theta)$ is a better termination condition since it is both an easily-computed scalar and an accurate measure of how well data associated with different classes are separated.

To initialize this iterative process, we can compute the initial discriminant vectors V_0^* either from a small portion of manually labeled pixels or data classified by the classical K-means clustering method. We can simply skip the feature normalization step if the process is started with manually labeled pixels due to the theorem shown in the Appendix.

From *a priori* knowledge of the structures in neuroblastoma slides, we know there are four major components to be detected: nuclei, cytoplasm, neuropil and background [1]. In addition to these four classes, red blood cells (RBCs) are also frequently observed in our testing images. Instead of increasing the class number C from four to five, we used simple thresholding techniques to remove those image regions containing RBCs so as to keep the computational cost as low as possible. In our experiments, the ratio of R (red) component to the sum of R, G and B components at each pixel is employed as the measure for the RBC detection. After analyzing the ratio histograms, we fix the threshold at 0.45.

2.3. Feature and Classifier Design

Due to the fact that no single classifier can be inferior to any other for any given data, it is worthwhile testing the given data with multiple classifiers when there is a lack of *a priori* knowledge of the data distribution. Therefore, we tested two different classifiers with the same feature set.

For each of the L, A^* and B^* channels, we only compute the entropy, mean and variance of the range of values within a local neighborhood, and the homogeneity degree of the co-occurrence matrix associated with each pixel within the cytoplasm and neuropil regions, since these two components bear the most discriminating textures after a careful study on the segmentation results. As a result, a feature vector with 24 entries, *i.e.* 12 components derived from cytoplasm and the other half from neuropils, are extracted from each image. We assumed are that these features are good representations for different subtypes of neuroblastoma and that feature points associated with different classes form well-shaped and well-separated clusters in the resulting feature space. These assumptions are verified by the experimental results that follow.

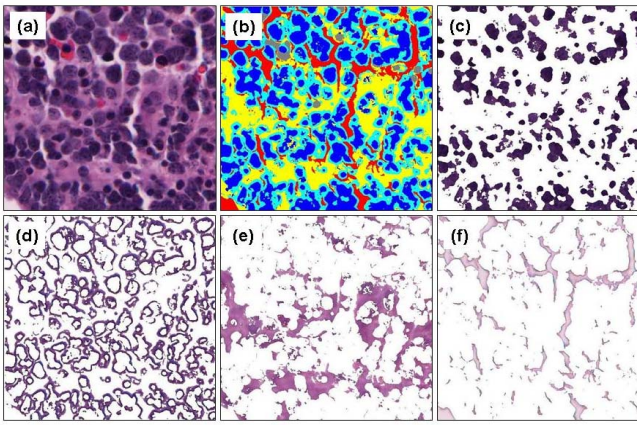

Fig. 2 A typical segmentation result of an image from undifferentiated class with components detected by EMLDA is shown. (a) Original image; (b) Partitioned image shown in color; (c)Nuclei; (d)Cytoplasm; (e)Neuropil; (f)Background component.

3. EXPERIMENTAL RESULTS

Typical segmentation results using the EMLDA are shown in Figure 2. To assess the performance of the overall classification system, we employed the leave-one-out validation process with two different classifiers (i.e. KNN, LDA) over 129 images cropped from slides of each class. The classification performances associated with three different segmentation methods (EMLDA, K-means and FCM) are shown in Table 1. In particular, when EMLDA segmentation with the LDA technique is used, the 2D scatter plot of the projected data in the lower dimensional space spanned by the resulting discriminant vectors is shown in Figure 3. This figure demonstrates good separation of different classes with a few outliers.

Seg. / Classifier	EMLDA	KMEANS	FCM
KNN	97.7%	**98.7%**	98.7%
LDA	**98.7%**	98.5%	**99.0%**

Table 1. Classification rates with different combinations of segmentation techniques and classifiers.

From Table 1, we can see that the best recognition rate associated with the EMLDA segmenation is comparable to those of KMEANS or FCM, yet EMLDA segmentation yields much faster convergence rate, as shown in Figure 4.

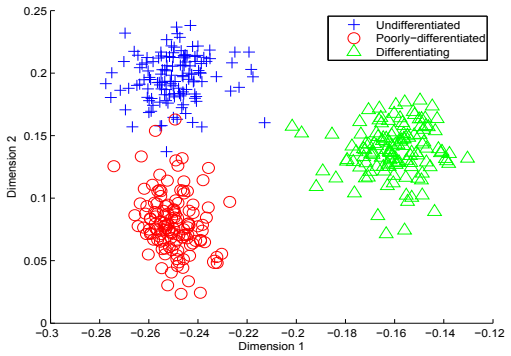

Fig. 3. A typical scatter plot of the feature data associated with testing images in the 2D space spanned by the discriminant vectors when the combination of EMLDA and LDA is used.

In our experiments, the mean and standard deviation of the computational cost of the EMLDA is 13.6 ± 4.93 in seconds, while those for KMEANS and FCM are 29.6 ± 9.92 and 71.5 ± 7.72 seconds respectively.

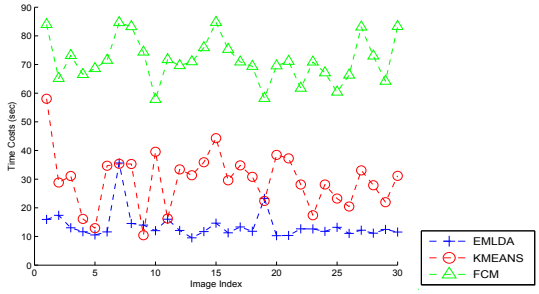

Fig. 4. Computational costs for EMLDA, KMEANS and FCM segmentation methods on a randomly selected subset of testing images.

4. CONCLUSIONS

In this paper, we have presented a new automated system for the assessment of grade of neuroblastic differentiation. The proposed system achieves high degree of accuracy with computational efficiency. A novel segmentation method modeled after the expectation-maximization algorithm and Fisher-Rao criterion has been developed and shown to be effective. The misclassification rate is relatively low. When the EMLDA segmentation followed by the LDA classification is used, there are only three and two misclassified cases from undifferentiated and poorly-differentiated classes, respectively. The main reasons for misclassification are relatively small size of cropped images, which sometimes do not contain all four major components of interest, and the existence of heterogeneous region associated with different grades of differentiations within a single image. In our future work, we will address these issues and extend this analysis to the whole slide images.

APPENDIX

Theorem 1 If the initial discriminant vectors are computed from manually labeled pixels, feature normalization by linear scaling to unit range does not change the end classification result.

Proof. Suppose the mapping functions are linear scaling ones that map each feature to a specific unit range, *i.e.* $\hat{x} = Mx + D$, where $M = diag(m_{11} \cdots m_{pp})$ and $D = (d_1 \cdots d_p)^T$, then we have:

$$\hat{S}_B = \sum_{i=1}^{C} n_i (\hat{m}_i - \hat{m})(\hat{m}_i - \hat{m})^T$$

$$= \sum_{i=1}^{C} n_i ((Mm_i + D) - (Mm + D))((Mm_i + D) - (Mm + D))^T$$

$$= \sum_{i=1}^{C} n_i M (m_i - m)(m_i - m)^T M^T = M S_B M^T = M^T S_B M$$

Similarly, we have $\hat{S}_W = M^T S_W M$.

Furthermore, we have:

$$\hat{J}(\hat{V}) = \frac{\left|\hat{V}^T \hat{S}_B \hat{V}\right|}{\left|\hat{V}^T \hat{S}_W \hat{V}\right|} = \frac{\left|\hat{V}^T M^T S_B M \hat{V}\right|}{\left|\hat{V}^T M^T S_W M \hat{V}\right|} = J(V) \qquad (7)$$

where $V = M\hat{V}$.

Equation (7) reveals the fact that J and $\hat{J}$ can be simultaneously maximized when $V^* = M\hat{V}^*$. The projected data and the mean in the lower dimension space spanned by the discriminant vectors for $X = \{x \mid x \in R^p\}$ are $x^T V^*$ and $m_i^T V^*$, while those associated with $\hat{X} = \{\hat{x} \mid \hat{x} \in R^p\}$ are:

$$\hat{x}^T \hat{V}^* = (Mx + D)^T \hat{V}^* = x^T M \hat{V}^* + D^T \hat{V}^* = x^T V^* + D^T \hat{V}^* \qquad (8)$$

$$\hat{m}_i^T \hat{V}^* = (Mm_i^T + D)^T \hat{V}^* = m_i^T M \hat{V}^* + D^T \hat{V}^* = m_i^T V^* + D^T \hat{V}^* \qquad (9)$$

i.e. the projected data derived from X and $\hat{X}$ are related by a simple translation. Therefore, the end classification results in X and $\hat{X}$ feature space are conserved to be the same. ■

5. REFERENCES

[1] H. Shimada, et al., "Terminology and Morphologic Criteria of Neuroblastic Tumors", *American Cancer Society*, Vol. 86, No. 2, pp.349-363, 1999.

[2] H. Shimada et.al., "The International Neuroblastoma Pathology Classification (the Shimada System)", *American Cancer Society*, Vol. 86, No. 2, pp.364-372, 1999.

[3] M.N. Gurcan, T. Pan, H. Shimada, J. Saltz, "Image Analysis for Neuroblastoma Classification: Segmentation of Cell Nuclei", 28th International Conference of the IEEE Engineering in Medicine and Biology Society Conference, New York, NY, August 2006.

[4] F.J. Ayres, M.K. Zuffo, R.M. Rangayyan, G.S. Boag, V.O. Filho, and M.Valente, "Estimation of the tissue composition of the tumor mass in neuroblatoma using segmented CT images", *Medical and Biological Engineering and Computing*, Vol. 42, pp.366-377, 2004.

[5] N.R. Pal, and S.K. Pal, "A Review on Image Segmentation Techniques", *Pattern Recognition*, Vol. 26, pp.1277-1294, 1993.

[6] G.J. McLachlan, "Discriminant Analysis and Statistical Pattern Recognition", Wiley-Interscience, 1992.