Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 1202 – 1210

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

A study on nuclei segmentation, feature extraction and disease stage classification for human brain histopathological images

Kiichi Fukuma[a], V. B. Surya Prasath[b], Hiroharu Kawanaka[a,*], Bruce J. Aronow[c], Haruhiko Takase[a]

[a]*Graduate School of Engineering, Mie University, 1577 Kurima-machiya, Tsu, Mie 514-8507, Japan*
[b]*Computational Imaging and VisAnalysis (CIVA) Lab, Department of Computer Science, University of Missouri-Columbia, MO 65211 USA*
[c]*Department of Biomedical informatics, Cincinnati Children's Hospital Medical Center, OH 45229 USA*

Abstract

Computer aided diagnosis (CAD) systems are important in obtaining precision medicine and patient driven solutions for various diseases. One of the main brain tumor is the Glioblastoma multiforme (GBM) and histopathological tissue images can provide unique insights into identifying and grading disease stages. In this study, we consider nuclei segmentation method, feature extraction and disease stage classification for brain tumor histopathological images using automatic image analysis methods. In particular we utilized automatic nuclei segmentation and labeling for histopathology image data obtained from The Cancer Genome Atlas (TCGA) and check for significance of feature descriptors using K-S test and classification accuracy using support vector machine (SVM) and Random Forests (RF). Our results indicate that we obtain classification accuracy 98.6% and 99.8% in the case of Object-Level features and 82.1% and 86.1% in the case of Spatial Arrangement features, respectively.

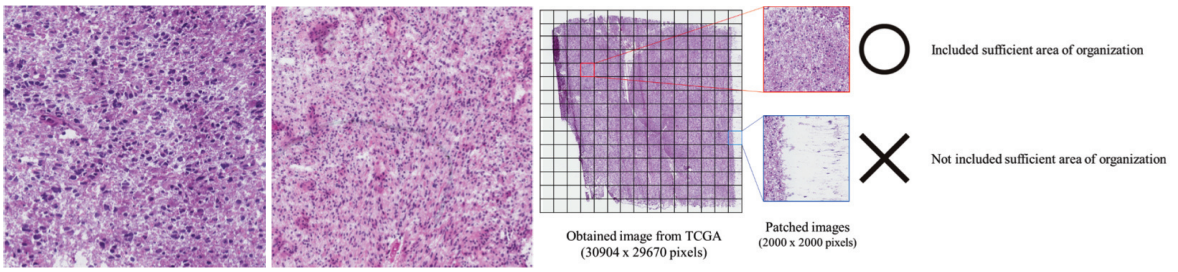
Peer-review under responsibility of KES International
Keywords: Histopathology; Human Brain; Feature Extraction; Progress Analysis.

1. Introduction

In the field of histopathology, quantitative and efficient analysis are paramount in devising precision medicine and patient specific diagnosis. Therefore, over the years many studies on evaluation methods for histopathological tissue images have been reported[1]. One of the malignant tumor is based on Glioma which occur in the brain and the prognosis is usually quite worse. Disease stage classification of Glioma is very important in the field of histopathology. However, there are various problems exists in doing manual disease stage classification. First, the number of tissue samples is enormous and this gives much burden to pathologists because they have to analysis them manually, and in some case pathologists diagnose these images during surgery. Therefore, there is a demand for quickly analysis to do that. In addition, the criteria of analysis heavily depend on each pathologists experience and perception. From this

* Corresponding author. Tel.: +81-59-231-9737 ; fax: +0-000-000-0000.
E-mail address: kawanaka@elec.mie-u.ac.jp

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.164



(a) LGG (b) HGG (c) Patch extraction and separation

Fig. 1. (a) Lower-Grade Glioma (LGG). (b) High-Grade Glioma (Glioblastoma multiforme) (GBM). (c) How to separate from big .svs format image to patched images.

background, many studies on computational pathology using computer vision techniques have been reported [1,2,3,4,5]. As discussed in [1] computer-assisted diagnosis (CAD) systems to detect tumors from histopathological images require strong feature extractions and further classification analysis to forecast patients prognosis. It also showed that the employed feature descriptors were effective for cytopathology and histopathology. However, we note that most of them were applied to only specific diseases like breast cancer, esophagitis and so on. Therefore, we cannot judge whether these descriptors are ineffective for other tissue type images or not [6].

This paper discusses the method of nuclei segmentation using custom designed image processing and analysis steps, and describe the performance of various feature descriptors which are then used to evaluate disease stages of Glioma and classify them. In this paper, the feature descriptors were applied to Glioma images obtained from The Cancer Genome Atlas (TCGA), and we determined significant descriptors for Glioma images. We also tested support vector machine (SVM), and Random Forests (RF) [7] classification methods. We show that our image analysis pipeline can be used effectively in feature extraction and disease stage classification for Glioma histopathological images. Further our proposed pipeline is general in the sense that it is extensible to other tissue type images.

Rest of the paper is organized as follows. Section 2 and Section 3 illustrates the performance of our proposed method with other segmentation schemes. Section 4 concludes the paper.

2. Disease stage classification

In this study, we used human brain histopathological images of Glioma which were obtained from TCGA Data. Generally, Glioma tissue images can be categorized into 4 grades based their disease stages. In particular, the status of Grade 3 and 4 are so serious and have poor prognoses. Actually, TCGA Data has 2 types image of Glioma, and these are Brain Lower-Grade Glioma (LGG) and Glioblastoma multiforme (GBM), and LGG is Grade 1 and 2, GBM is Grade 3 and 4, respectively. We used these two distinct types image as the experimental materials. Figures 1 show examples of materials of those image. In this image, nucleus were stained deep purple and other tissues were stained pale purple, red and so on using H&E staining method. As you can see, features of the organization are different from each other. Generally, these differences come from disease progression, i.e. genes expressions. In the previous research, we used an image each category, LGG and GBM. However we thought that stained state is different from each images, and this caused bad influence to accuracy of disease progression. Therefore, in this paper, we used 10 LGG images and 10 GBM images respectively to avoid this problem. Obtained images from TCGA Data are .svs format. These are very big size, for example one of the image is about 30000x30000 pixels. Therefore it is unsuited for processing. For that reason, we divided original images to patched images whose sizes were 2000x2000 pixels, and then the patched images including sufficient nuclei were used as experimental materials. Figure 1 shows how to dividing original image to patched images and the criteria as experimental material. In this paper, we used 10 .svs format images in each category, and we obtained 100 patched images from a single .svs format image. Therefore, we used 1000 patched images in each category.

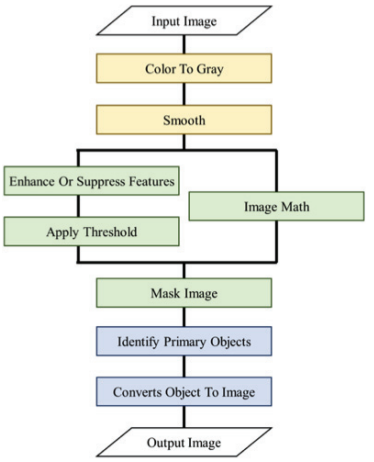


Fig. 2. Our proposed pipeline for the Glioma histopathology nuclei segmentation.

2.1. Nuclei Segmentation

In this study, we employed CellProfiler[8] to implement our nuclei segmentation pipeline. CellProfiler is software designed for histopathology image processing and supports basic methods and algorithms as libraries.

Figure 2 shows the pipeline (flow of processing) of our segmentation method. The pipeline mainly consists of 3 steps:

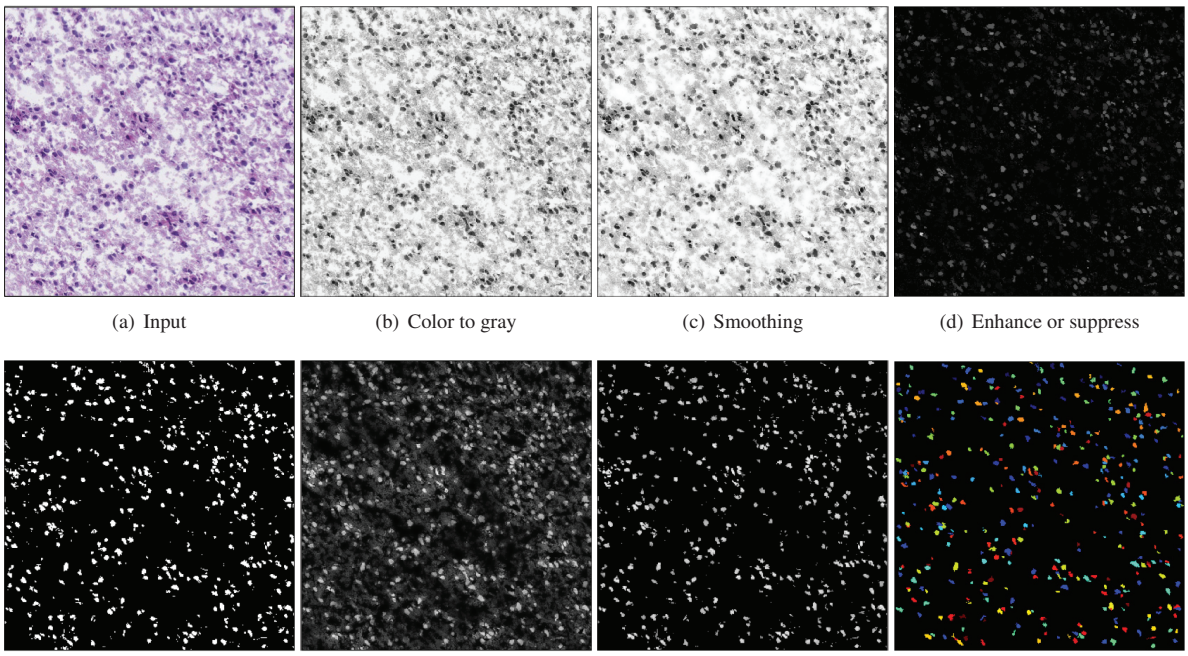
(1) pre-processing (yellow steps),
(2) rough segmentation (green ones), and
(3) nuclei separation (blue ones).

In the cases of the input images, nuclei are stained deep purple and others are red (or light purple) by H&E staining method, and nuclei can be observed as dark-colored objects compared with others. With consideration for such features, the proposed method emphasized dark areas in the image. In this step, the input image was converted to a gray-scale image by using Color To Gray module and then, Smooth module was applied to the converted image. In this step, we set parameters of these modules to illuminate red colored objects, e.g. blood cells, in the image, because we doesn't need blood cells but cell nucleus. Figures 3 (a), (b) and (c) show the input image, the converted image and applied image, respectively.

Next, nucleus in the image were roughly segmented by using the following procedure. In this step, we used Enhance Or Suppress Features module to emphasize dark areas, i.e. nuclei, and then the inverted image like Figure 3 (d) was obtained. After this, Apply Threshold module was used to get the binary image shown in Figure 4 (e). As you can see, nuclei regions could be roughly extracted from the obtained image, but there were nuclei-attached regions that should be separated each other. Further fuzzy smoothing and thresholding techniques[9,10] can be used to improve these preliminary pre-processing steps.

To separate these regions, we prepared the inverted image without enhancement (Figure 3 (f)) by using Image Math module and the smoothed image (Figure 3 (c)). Then, we obtained the rough segmentation image by superimposing the binary image (Figures 3 (e)) and the inverted image (Figure 3 (f)). In this step, Mask Image module was employed. Figure 3 (g) shows the result of rough segmentation by the above procedure.

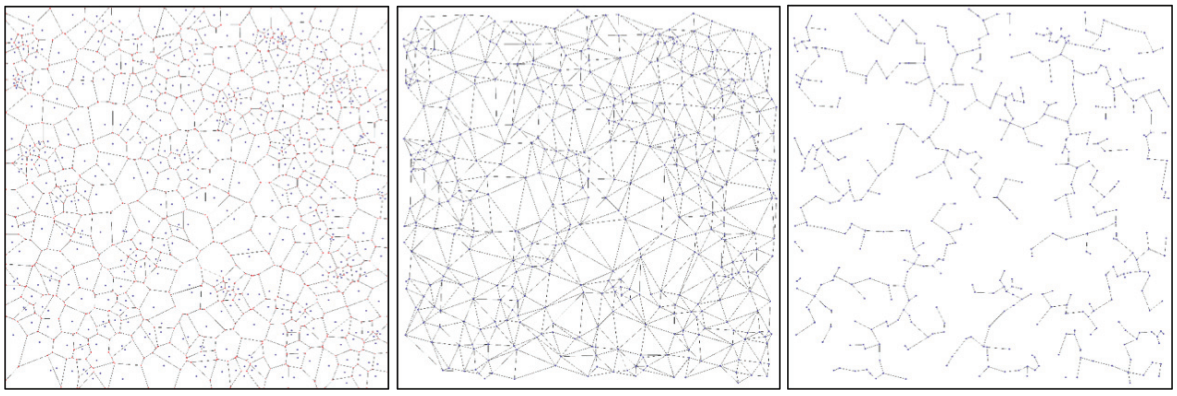
As the last step, we separated nuclei attached regions into each nucleus based on pixel values. Generally, a nuclei attached region has some peaks of pixel values like a multimodal distribution. In addition, the ideal distribution of pixel values of a single nucleus is unimodal. Thus, we first detected local minimum points of pixel values in the region and then traced the points using Identify Primary Objects and Converts Object To Image modules. In these step, we also removed nuclei overlapping with boundaries of the image, because such nuclei could not be extracted correctly



(a) Input (b) Color to gray (c) Smoothing (d) Enhance or suppress

(e) (f) (g) (h)

Fig. 3. Example patch image and its corresponding steps of proposed pipeline for obtaining Glioma histopathology nuclei segmentation and labelling process. (a) Input histopathology image, (b) color to gray conversion, (c) image smoothing, (d) enhancing or suppressing operation, (e) automatic thresholding, (f) image math operations, (g) after masking, and (h) nuclei identification and labeling.



(a) Voronoi Tessellation (b) Delaunay Triangulation (c) Minimum Spanning Tree

Fig. 4. Sample spatial-arrangement features extracted from example histopathology image given in Figure 3 (a).

and it harmed segmentation accuracy. Finally, segmentation result was obtained by coloring each nucleus (Figure 3 (h)).

2.2. Feature extraction

In the literature, feature descriptors can be categorized into

Table 1. List of object-level feature descriptors for histopathological image analysis.

	Type	Features
Object-level Features	—	# of Nuclei, Area
	Elliptical	Major Axis Length, Minor Axis Length, Eccentricity, Orientation, Elliptical Deviation, Extent, Aspect Ratio
	Convex Hull	Convex Area, Convex Deficiency, Solidity
	Bounding Box	Extent, Aspect Ratio
	Boundary	Perimeter, Radii, Perimeter Curvature
	Other Shape	Equivalent Diameter, Sphericity, Compactness, Inertia Shape

Table 2. List of spatial-arrangement feature descriptors for histopathological image analysis.

	Type	Features
Spatial-arrangement Features	Voronoi Tessellation	# of Nodes, # of Edges, Edge Length, Cyclomatic #, # of k-Walks, Degree, Spectral Radius, Randic Index, Voronoi Area, Voronoi Area Disorder, Perimeter, Roundness Factor, Roundness Factor Homogeneity
	Delaunay Triangulation	# of Nodes, # of Edges, Edge Length, Cyclomatic #, # of Triangles, # of k-Walks, Degree, Spectral Radius, Randic Index
	Minimum Spanning Tree	# of Nodes, # of Edges, Edge Length, Degree, Spectral Radius, Randic Index

(1) Object-Level, and
(2) Spatial-Arrangement features.

Tables 1 and 2 show the extracted features list. Object-Level features are features related to sizes and shapes of the nuclei. Spatial-Arrangement features describe topological features based on graph theory. We used Voronoi Tessellation (VT), Delaunay Triangulation (DT) and Minimum Spanning Tree (MST)[11]. Samples of these graphs were showed in Figure 4, respectively. To calculate them, we applied the features defined in the literature to the Glioma images, and their statistical significances were evaluated by using Kolmogorov-Smirnov (K-S) test. In this paper, the significance level was set to 0.01.

2.3. Disease stage classification

The significant feature descriptors were employed to generate a feature vector for disease stage classification. In this study, all of feature values obtained by significant descriptors were used as coefficients of the feature vector. The patched images were classified by using support vector machine (SVM) and Random Forests. In the SVM case, one-leave-out cross-validation method was used for evaluation. In this study, these classification was used R which is software for statistical analysis.

Table 3. Kolmogorov-Smirnov (K-S) test result of object-level features.

Type	Features	*p*-value
——	# of Nuclei	< 0.01
	Area	
Elliptical	Major Axis Length	< 0.01
	Minor Axis Length,	
	Eccentricity	
	Orientation	
	Elliptical Deviation	
	Extent	
	Aspect Ratio	
Convex Hull	Convex Area	< 0.01
	Convex Deficiency	
	Solidity	
Bounding Box	Extent	< 0.01
	Aspect Ratio	
Boundary	Perimeter	< 0.01
	Radii	
	Perimeter Curvature	
Other Shape	Equivalent Diameter	< 0.01
	Sphericity	
	Compactness	
	Inertia Shape	

Table 4. Kolmogorov-Smirnov (K-S) test result of spatial-arrangement features.

Type	Features	*p*-value
Voronoi Tessellation	# of Nodes	< 0.01
	# of Edges	
	Edge Length	
	Cyclomatic #	
	# of k-Walks	
	Degree	
	Spectral Radius	
	Randic Index	
	Voronoi Area	
	Voronoi Area Disorder	
	Perimeter	
	Roundness Factor	
	Roundness Factor Homogeneity	
Delaunay Triangulation	# of Nodes	< 0.01
	# of Edges	
	Edge Length	
	Cyclomatic #	
	# of Triangles	
	# of k-Walks	
	Degree	
	Spectral Radius	
	Randic Index	
Minimum Spanning Tree	# of Nodes	< 0.01
	# of Edges	
	Edge Length	
	Degree	
	Spectral Radius	
	Randic Index	

Table 5. Classification results of SVM.

Type	Vector	Accuracy%
Object-level features	# of Nuclei, Area	99.87
	Major Axis Length, Minor Axis Length	98.46
	Perimeter, Radii	98.76
	Convex Area, Solidity	97.26
	Equivalent Diameter, Sphericity	99.03
Voronoi Tessellation	Voronoi Area, Voronoi Perimeter	82.48
	Randic Index, # of Edges	80.72
	# of Nodes, Cyclomatic #	82.88
	# of 3 Walks, Roundness Factor	81.11
	Voronoi Area Disorder, Roundness Factor Homogeneity	80.62
Delaunay Triangulation	# of 3 Walks, # of 5 Walks	81.92
	# of Triangles, Average Edges Length	82.66
	Cyclomatic #, Randic Index	81.21
Minimum Spanning Tree	Randic Index, Average Edge Length	83.05
	# of Nodes, Degree	82.47

3. Experimental result and discussion

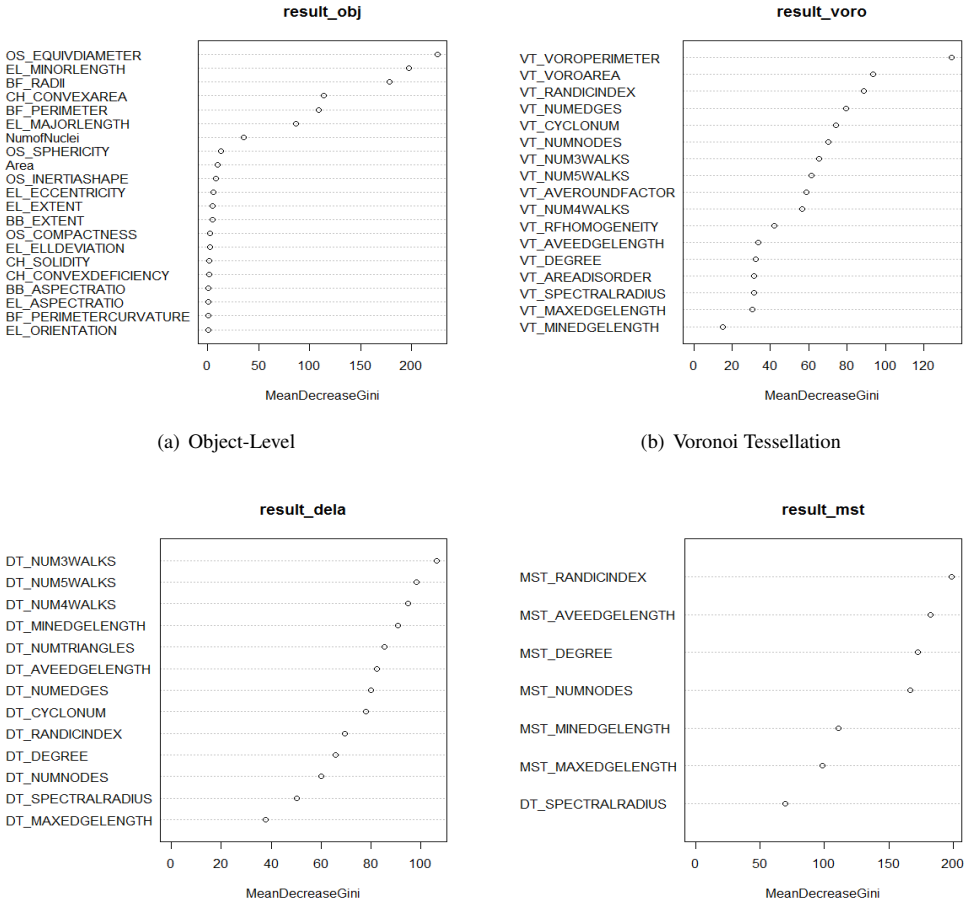
Tables 3 and 4 show the summarized Kolmogorov-Smirnov (K-S) test results on Object-Level features and Spatial-Arrangement features, respectively. From the results, p-values of all features in Object-Level features and Spatial-Arrangement features were effective for evaluation of disease progression of Glioma. Because p-values of these features were less than 0.01.

Table 5 shows the results of disease stage classification using SVM. The results of SVM using Object-Level Features show that 98.6% of Glioma images were classified correctly. And, the results of SVM using the features of VT, DT and MST show 81.6%, 81.9% and 82.8%, respectively. Especially, in the case of Object-Level Features, these combination were consisted from same type features. Actually, in the previous research have used only 1 .svs format image in each category, the result of accuracy of SVM using Object-Level features was 98.7%. This accuracy was not difference from the accuracy in this paper. However, the result of SVM using VT, DT and MST features were 99.3%, 98.8% and 99.3%. These accuracy ware vary difference from the accuracies in thi paper. From these result, we thought that Object-Level features is better than Spatial-Arrangement features to classify the progression of Glioma image in the TCGA Data.

Figures 5 show the graphs of Mean Decrease Gini in the case of each feature using Random Forests. As a result, in the case of Object-Level features, classification accuracy is 99.8% and we used mtry = 4. The value of mtry is the parameter of Random Forest in the "R" package. And mtry is the number of variables ($\sqrt{n}$: n is number of features). On the other hand, in the case of VT which is Spatial-Arrangement features, classification accuracy is 84.38% and we used mtry = 4. Similarly, in the case of DT and MST, classification accuracy is 87.06% and 86.75%, and mtry = 3, and = 2, respectively. These values of mtry were obtained by tuning step. Mean Decrease Gini is importance for classification of LGG and GBM, and these graphs indicate the importance of each features in disease classification. Therefore, the features which have high Mean Decrease Gini value are significant features for disease stage classification in this study.

4. Conclusions

This paper discussed nucleus segmentation method, feature descriptors and disease stage classification for Glioma histopathology images. In this paper, we discussed nucleus segmentation method for LGG and GBM images of TCGA Data. And, the feature descriptors defined in the literature were applied to Glioma images, and statistical significance of them were discussed too. We also tried to distinguish LGG from GBM images by using the significant descriptors, based on support vector machines, and Random Forests based classifiers. After the experiments, more than 98.6% of images were classified correctly using SVM which was consist of Object-Level features, and 82.1% of images were classified correctly using SVM which was consist of Spatial-Arrangement features. In addition, more than 99.8% of



(a) Object-Level (b) Voronoi Tessellation

(c) Delaunay Triangulation (d) MST

Fig. 5. Importance of various features in random forests based classifier.

images were classified correctly using Random Forests in the case of Object-Level features, and more than 86.1% of images were classified correctly using RF in the case of Spatial-Arrangement features. From these result, we thought that the Object-Level features is better than Spatial-Arrangement features to classify the progression of Glioma image in TCGA Data.

We are currently improving the overall image analysis pipeline with fuzzy entropy thresholding[10], and active contour for improving the accuracy of nuclei segmentation[12]. We will further investigate to discover disease subtypes using feature matrices and confirm relationship of the disease stage and gene expression available within the TCGA data. We hope our automatic image analysis method will be of help in medical decision analysis in general, and brain tumors in particular.

References

1. M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, Histopathological image analysis: A review, IEEE Reviews In Biomedical Engineering, vol. 2, pp. 47171, 2009.
2. A. M. Marchevsky and M. R. Wick, Evidence-based medicine, medical decision analysis, and pathology, Human Pathology, vol. 35, no. 10, pp. 1179–1188, 2004.

3. L. E. Boucheron, Object-and spatial-level quantitative analysis of multispectral histopathology images for detection and characterization of cancer, Ph. D. dissertation, University of California, Santa Barbara, CA, 2008.
4. K. Fukuma, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, H. Takase. Feature extraction and disease stage classification for glioma histopathology images, IEEE International Conference on E-health Networking, Application & Services (HealthCom), Boston MA, USA, Oct 2015. Proc. IEEE, pp. 429–430.
5. K. Fukuma, H. Kawanaka, V. B. S. Prasath, B. J. Aronow, H. Takase. A study on feature extraction and disease stage classification for glioma pathology images," IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), July 2016.
6. K. Tamaki, K. Fukuma, H. Kawanaka, H. Takase, S. Tsuruoka, B. J. Aronow, and S. Chaganti. Comparative study on feature descriptors for brain image analysis. International Conference on and Advanced Intelligent Systems (ISIS), Proc. IEEE, pp. 679–682, 2014.
7. Tin Kam Ho, Random decision forests, Third International Conference on Document Analysis and Recognition, vol. 1. Proc. IEEE, pp. 278–282, 1995.
8. M. R. Lamprecht, D. M. Sabatini, A. E. Carpenter, CellProfiler: free, versatile software for automated biological image analysis, Biotechniques, vol. 42, no. 1, pp. 71–75, 2007.
9. V. B. S. Prasath and R. Delhibabu, Image restoration with fuzzy coefficient driven anisotropic diffusion, Joint International Conference on Swarm, Evolutionary and Memetic Computing (SEMCCO), Visakhapatnam, India. Proc. Springer LNCS 8947. (eds. B.K. Panigrahi, P. N. Suganthan, S. Das), pp. 145–155, Dec 2014.
10. V. B. S. Prasath, R. Delhibabu, Automatic mucosa detection in video capsule endoscopy with adaptive thresholding, International Conference on Computational Intelligence in Data Mining (ICCIDM), Bhubaneswar, India. Proc. Springer SIST 410 (Eds.: H. S. Behera, D. P. Mohapatra), pp. 95–102, Dec 2015.
11. B. Weyn, W. A. Tjalma, P. Vermeylen, A. van Daele, E. Van Marck, and W. Jacob, Determination of tumour prognosis based on angiogenesis-related vascular patterns measured by fractal and syntactic structure analysis, Clinical Oncology, vol. 16, no. 4, pp. 307–316, 2004.
12. V. B. S. Prasath, K. Fukuma, B. J. Aronow, and H. Kawanaka, Cell nuclei segmentation in glioma histopathology images with color decomposition based active contours, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA. Proc. IEEE, pp. 1734–1736, Nov 2015.