USE OF IMPERFECTLY SEGMENTED NUCLEI IN THE CLASSIFICATION OF HISTOPATHOLOGY IMAGES OF BREAST CANCER

Laura E. Boucheron[*]*, B. S. Manjunath*[†]*, and Neal R. Harvey*[‡]

[*] Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003
[†] Department of Electrical and Computer Engineering, University of California Santa Barbara, Santa Barbara, CA 93106
[‡] Space and Remote Sensing, Los Alamos National Laboratory, Los Alamos, NM 87545

ABSTRACT

Many features used in the analysis of pathology imagery are inspired by grading features defined by clinical pathologists as important for diagnosis and characterization. A large majority of these features are features of cell nuclei; as such, there is often the desire to segment the imagery into individual nuclei prior to feature extraction and further analysis. In this paper we present an analysis of the utility of imperfectly segmented cell nuclei for classification of H&E stained histopathology imagery of breast tissue. We show the object- and image-level classification performance using these imperfectly segmented nuclei in a benign versus malignant decision. Results indicate that very good classification accuracies can be achieved with imperfectly segmented nuclei and further that perfect nuclei segmentation does not necessarily guarantee better classification accuracy.

Index Terms— Histopathology, medical image analysis, nuclei segmentation, breast cancer, H&E

1. INTRODUCTION

The segmentation of cell nuclei on an object level in histopathology imagery is a very difficult problem. Many nuclear segmentation algorithms make use of immunostaining for more specific location of cellular structures [1], the Feulgen stain which is specific to DNA [2–4], or even the physical extraction of nuclei prior to image acquisition [3]. The use of cytology imagery is more often addressed since this provides a less complicated image structure with more instances of isolated cells and/or well-delineated cell clusters [5–7].

The use of automated nuclear segmentation algorithms on standard H&E histology imagery is rarer and has met with mixed success [8, 9]. Semi-automated or even manual nuclear segmentation is still used in some research, even for the more specific Feulgen staining protocol [2–4], immunohistochemical staining [1], and cytology imagery [6]. This indicates the inherent difficulty in segmentation of nuclei and also indicates the desire for appropriately segmented nuclei for further nuclear-based quantitative analysis (e.g., classification of cancer imagery). Automated segmentation is more desirable, however, for ultimate use in computer-aided diagnosis systems since it does not require extensive user interaction.

In this paper, we investigate the use of imperfectly segmented nuclei for the classification of H&E-stained histopathology imagery of breast cancer. We describe the imagery in Section 2. In Section 3 we desribe an automated pixel-level classifier to determine the nuclei pixels, and several automated segmentation algorithms to delineate the individual nuclei. Comparisons of object- and image-level classification performance using different segmentations are presented in Section 5. We conclude and briefly discuss future work in Section 6.

2. IMAGERY

Our dataset contains 58 Hematoxylin and Eosin (H&E) histopathology images of breast tissue from the Yale Tissue Microarray Facility. The data was captured from 5 microarrays (ytma10, 12, 49, and 55), with (6, 6, 34, and 6) images captured per array, respectively; in total we have 26 malignant images, and 32 benign (including 6 normal from ytma55). These 58 images are not microarray images in the general sense since they are single histopathology images as might be obtained from standard clinical biopsy specimens. The images were labeled benign or malignant by an experienced pathologist.

As acquired, the images are multispectral with 29 bands, ranging within the visible spectrum from 420 to 700 nm, spaced 10 nm apart, acquired with a liquid crystal tunable filter and a typical clinical pathology microscope setup with a 40x objective. Each band is represented in an image stack as an 8 bit, 768×896 grayscale image.

The 29-band multispectral images are converted to representative RGB images by converting the light wavelengths into constituent red, green, and blue values based on a modification of a standard color matching method [10]. The second lobe in the red response (due to human perception of violet as a combination of red and blue) was removed to better match the spectral responses of common 3-CCD color cameras. The success of the conversion process in maintaining spectral characteristics of the H&E dyes was further validated by qualitative observation of an experienced pathologist. A portion of an example derived RGB image is in Fig. 1 (a).

3. SEGMENTATION OF NUCLEI

In this section, we describe the nuclei segmentation methods that we apply to our data. We also briefly describe the pixel-level classification of nuclei pixels and the metric of segmentation performance used to analyze the segmentation algorithms. We define *segmentation* very specifically as the *delineation of individual objects* (nuclei) in an image; we refer to the partitioning of the image into nuclei and non-nuclei regions as pixel-level classification of nuclei.

3.1. Pixel-Level Classification of Nuclei

In [11] we presented the pixel-level classification of nuclei for this same dataset. For the work presented here, we use the results of the Fisher linear discriminant analysis classifier, since it had one of the highest overall performances. This supervised classification method uses solely spectral information from training data to determine a classifier. We found the pixel-level classification performance on both training and test data to be approximately equivalent, so we use both training and test images for subsequent segmentation. An example pixel-level classification is shown in Fig. 1 (b).

(a) Derived RGB image.

(b) Original binary image (pixel-level classification).

(c) WSCDT, $P = 0.24$.

(d) WSGran, $r = 4$, $P = 0.28$.

(e) CB, $P = 0.24$.
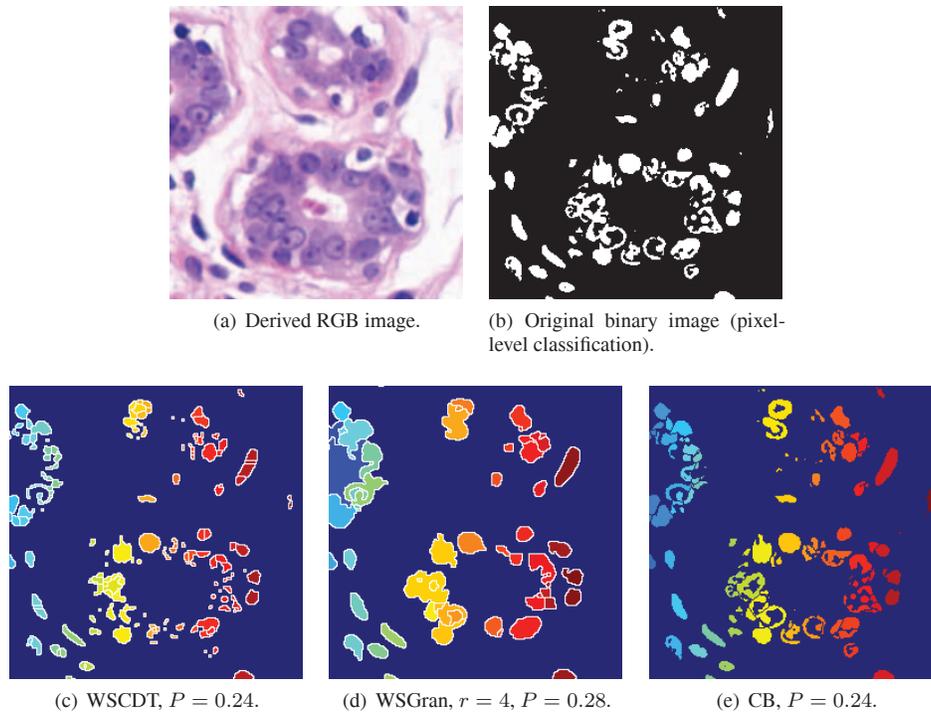
Fig. 1. (a) Derived RGB image, (b) binary image, and example (c) WSCDT, (d) WSGran, and (e) CB segmentations with associated segmentation performance metric P.

3.2. Segmentation Performance Metric

As a measure of segmentation performance, we use the measure described in [12]. Briefly, this measure penalizes, on an object-by-object basis, deviations in size and shape between the segmented objects and the ground truth objects. As ground truth, the nuclei are individually delineated manually in a $\sim 200 \times 200$ window chosen to be the most representative of the range of nuclei appearances within the entire image. The measure P lies in the range [0,1], with 1 being a perfect segmentation.

3.3. Watershed-Based Segmentation of Nuclei

For the watershed-based segmentation of nuclei, we use the watershed transform of the complemented Euclidean distance transform on a binary image of nuclei pixels. The binary image is the output of a pixel-level classifier for nuclei pixels as discussed above. This method yields an average performance of 0.15 ± 0.17. An example WSCDT segmentation is shown in Fig. 1 (c).

3.4. Marker-Based Watershed Segmentation of Nuclei

Granulometries are a common technique to elicit information about the size distribution of objects in an image. Granulometries are calculated by applying successively larger structuring elements in a morphological image operation and analyzing the residue image. More specifically, we use the morphological opening of a binary image with a disk-shaped structuring element (SE).

We compute watershed transforms with minima of the input image imposed according to foreground and background markers:

- Foreground markers: We use image residues from structuring elements of varying radii. These markers serve as an approximation of the center of the binary objects.

- Background markers: We use the erosion of the complement of the original binary image with a discrete circle of radius 3 These markers impose a minima in all parts of the background of the binary image. The erosion operator creates a separation between the background markers and the object boundaries.

The performance of the watershed using granulometry-based markers (WSGran) method increases with SE radius up to a $r = 4$ pixels, followed by a decrease in performance. The maximum average performance of the WSGran method is 0.28 ± 0.19. Fig. 1 (d) shows an example segmentation for SE radius $r = 4$. We note qualitatively here that the WSGran segmentation yields fewer small regions compared to the WSCDT segmentation. This can be seen in the many merged nuclei in Fig. 1 (d) compared to Fig. 1 (c).

3.5. Concavity-Based Segmentation of Nuclei

Motivated by observations that shape is a large factor in humans' ability to properly discriminate individual nuclei, we use the method in [13] which uses concavities as the basis for segmentation lines.

The concavity-based (CB) segmentation method of Kumar et al. [13] uses a rule-based approach for the segmentation of binary objects, beginning with a measure of concavity depth. In this method, the segmentation of one binary blob into multiple constituent objects is accomplished with the use of a "split line." A split line may occur between two object concavities, or between a single concavity and the opposite object boundary. Binary objects are split recursively until no further valid split lines are found.

667

Using our implementation of the concavity-based segmentation method of Kumar et al. [13], we find very poor performance (0.03 ± 0.08). Referring to the example segmentation shown in Fig. 1 (e), it appears that the CB method is penalized for splitting single nuclei into multiple objects and for including small regions as objects.

4. FEATURE EXTRACTION, FEATURE SELECTION, AND CLASSIFICATION

4.1. Feature Extraction

The extraction and use of relevant image features for automated analysis of cancer imagery is a topic of great interest. Research on useful features for cancer classification and diagnosis has often been approached by the definition of features flagged by clinicians as important features for the diagnosis process. The vast majority of these features are nuclear features. We use a comprehensive set of object-level features as described in [14] for the classification of histopathology imagery, resulting in a total of 1035 features per image object (nucleus). We apply feature extraction and selection to the segmented nuclei of Section 3.

4.2. Feature Selection

Feature selection is a means to select the relevant and important features from a large set of features, many of which may be redundant, irrelevant, or possibly detrimental to the classification performance. While humans have innate abilities to process and understand imagery, they do not tend to excel at explaining **how** they reach their decisions. As such, large feature sets are generated in the hopes that a subset of features incorporates the information the human expert is using for analysis.

Grafting (from "gradient feature testing") [15] is used as the feature selection method for this work. Grafting is based on a formulation of the feature selection problem whereby the classification of the underlying data and the feature selection process are not separated. Within the grafting framework, a loss function shows preference for classifiers with larger margins. We use a linear classifier in this work. Feature extraction and feature selection is computationally intensive for training, but feature selection allows for a streamlining of the feature extraction process for test imagery.

5. RESULTS

5.1. Object-Level Classification Performance

The grafting method of feature selection was applied to the WSCDT, WSGran ($r = 4$), and CB nuclear segmentations. Additionally, grafting was applied to a simple connected-components analysis (CC), and to the ground truth markup (GT). The area of ground truth markup within the image is used for analysis of object-level performance. This $\sim 200 \times 200$ window is the same window used for the pixel-level classification in [11], and encompasses an average of 50 nuclei per image. It should be noted, however, that the number of segmented objects may be more or less than the total number of ground truth nuclei in each image. The resulting feature subsets contain between 51 and 88 features out of the 1035 total.

The performances (percentage of segmented objects correctly classified) on the training and test data are presented in Table 1; training data are images used for the feature selection process, while test data are images unseen in the feature selection process. Each training and test dataset contains a randomly chosen half of the benign and

Table 1. Object-level feature subset performance (% correctly classified objects), when all objects in malignant imagery are considered malignant.

	CC	WSCDT	WSGran	CB	GT
Training	0.68	0.72	0.70	0.69	0.85
Test	0.73	0.69	0.69	0.73	0.74

Table 2. Object-level feature subset performance (% correctly classified objects), when only pathologist delineated regions in malignant imagery are considered to contain malignant objects.

	CC	WSCDT	WSGran	CB	GT
Training	0.60	0.69	0.66	0.66	0.78
Test	0.54	0.68	0.65	0.68	0.72

malignant imagery. Within Table 1, the assumption is that all objects in the benign (malignant) images are benign (malignant). Object-level performances are also presented in Table 2, with the assumption that only objects within a pathologist delineated region of malignant imagery are malignant. All objects outside the pathologist-specified region of malignancy are excluded from further analysis. Object-level classification for the whole 768×896 pixel image yield very similar performance to those of the $\sim 200 \times 200$ windows.

Interestingly, in all these results there is no clearly superior segmentation method. This would seem to indicate that a feature selection scheme can make use of the information contained in many different (and possibly poor) segmentations. Overall, however, it does not appear that too much information is lost or obscured with improperly segmented nuclei. Our results also compare favorably with the results on Feulgen-stained nuclei in breast cancer [4, 16].

5.2. Image-Level Classification Performance

It is the image-level classification performance that is often ultimately of interest. For this level of classification, however, there must be a threshold on the percentage of malignant objects before the entire image is considered malignant. Image-level performances are thus demonstrated with Receiver Operator Characteristic (ROC) curves, varying the threshold of malignant objects for which an image is considered malignant. Thus, an image is considered malignant for some percentage of malignant objects present, where this percentage ranges between 0 and 100. This allows for a tradeoff between the True Positive Rate (TPR) and False Positive Rate (FPR).

ROC curves for the test imagery are shown in Fig. 2, for the segmented objects in the $\sim 200 \times 200$ pixel ground truth window. Image-level results for segmented objects in the entire 768×896 image yield very similar image-level performance to the performance computed over the $\sim 200 \times 200$ window. The area under the curve (AUC) for each ROC curve is also shown in Fig. 2. The image-level performance using object-level features is very good, resulting in AUC values above 0.80 in most cases. The CB method provides the largest AUC value of 0.93, while the lowest is for with CC segmentation with an AUC of 0.80. A perfect nuclear segmentation (GT) yields an image-level performance with an AUC of 0.92.

It is interesting that many automatic segmentations provide image-level performances closer to that of GT as compared to the object-level performances in Section 5.1. This is most likely due to a different trade-off between TPR and FPR provided by the various segmentations due to different total number of segmented objects.
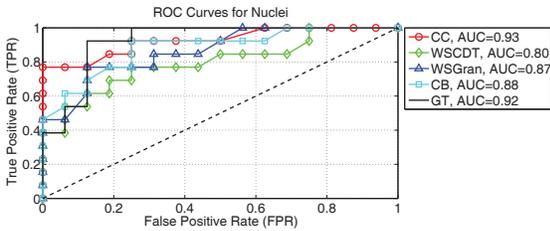
(a) Assuming all objects in malignant imagery are malignant.
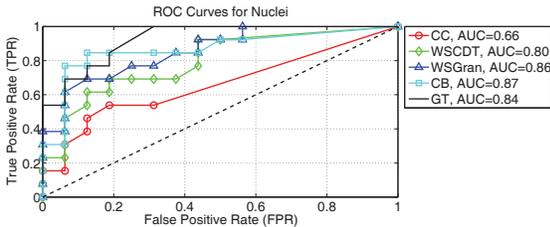
(b) Assuming objects within pathologist-specified region are malignant.

Fig. 2. ROC curves for image-level performance on test images.

6. CONCLUSION

It has been demonstrated that the comprehensive set of object-level features used herein are versatile and general enough to elicit important information from imperfectly segmented objects. This was demonstrated with object- and image-level classification performance for ground truth nuclei and for several nuclear segmentations.

Using object-level features and the grafting method of feature selection, we have shown object-level classification accuracies above 0.70 for test data. These results are comparable to others presented in the literature, namely [16] and [4], which relied on the more specific Feulgen staining protocol and the interactive selection of well-segmented nuclei. We hypothesize that it is the use of a comprehensive set of features that allows for the use of imperfectly segmented objects for this particular application. We have also shown image-level performances with AUC values greater than 0.80.

Ongoing work is investigating the analysis of feature subsets for insight into underlying characteristics of benign and malignant imagery. Future work will look at the application of these object-level feature extraction and selection methods to other histopathology datasets, as well as datasets from other application domains, e.g. imaging calorimeters. Furthermore, we are using cross-validation techniques and a simple tile-based "segmentation."

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] V. Sharifi-Salmation, B. Pesquet-Popescu, J. Simony-Lafontaine, and J. P. Rigaut, "Index for spatial heterogeneity in breast cancer," *J Microsc-Oxford*, vol. 216, no. 2, pp. 110–122, Nov. 2004.

[2] E. C. M. Mommers, N. Poulin, J. Sangulin, C. J. L. M. Meijer, J. P. A. Baak, and P. J. van Diest, "Nuclear cytometric changes in breast carcinogenesis," *J Pathol*, vol. 193, pp. 33–39, 2001.

[3] T. Mairinger, G. Mikuz, and A. Gschwendter, "Nuclear chromatin texture analysis of nonmalignant tissue can detect adjacent prostatic adenocarcinoma," *Prostate*, vol. 41, pp. 12–19, 1999.

[4] B. Weyn, G. van de Wouwer, A. van Daele, P. Scheunders, D. van Dyck, E. van Marck, and W. Jacob, "Automated breast tumor diagnosis and grading based on wavelet chromatin texture description," *Cytometry*, vol. 33, pp. 32–40, 1998.

[5] H.-S. Wu, J. Barba, and J. Gil, "A parametric fitting algorithm for segmentation of cell images," *IEEE T Bio-med Eng*, vol. 45, no. 3, pp. 400–407, Mar. 1998.

[6] K.-M. Lee and W. N. Street, "An adaptive resource-allocating network for automated detection, segmentation, and classification of breast cancer nuclei topic area: Image processing and recognition," *IEEE T Neural Networ*, vol. 14, no. 3, pp. 680–687, May 2003.

[7] C. Herrera-Espiñeira, C. Marcos-Muñoz, and J. Esquivias, "Automated segmentation of cell nuclei in fine needle aspirates of the breast," *Anal Quant Cytol*, vol. 20, no. 1, pp. 29–35, Feb. 1998.

[8] D. Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P.-A. Dadioti, and G. Nikiforidis, "Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models," *Anal Quant Cytol*, vol. 26, no. 6, pp. 331–340, Dec. 2006.

[9] L. Latson, B. Sebek, and K. A. Powell, "Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy," *Anal and Quant Cytol*, vol. 25, no. 6, pp. 321–331, Dec. 2003.

[10] W. S. Stiles and J. M. Burch, "NPL colour-matching investigation: Final report.," *Opt Acta*, vol. 6, pp. 1–26, 1959.

[11] L. E. Boucheron, Z. Bi, N. R. Harvey, and B. S. Manjunath, "Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery," *BMC Cell Biol*, vol. 8(Suppl 1):S8, 10 July 2007.

[12] L. E. Boucheron, N. R. Harvey, and B. S. Manjunath, "A quantitative object-level metric for segmentation performance and its application to cell nuclei," in *Proc: ISVC, LNCS*, Nov. 2007, vol. 4841, pp. 208–219.

[13] S. Kumar, S. H. Ong, S. Ranganath, T. C. Ong, and F. T. Chew, "A rule-based approach for robust clump splitting," *Pattern Recogn*, vol. 39, pp. 1088–1098, 2006.

[14] 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, Mar. 2008.

[15] S. Perkins, K. Lacker, and J. Theiler, "Grafting: Fast, incremental feature selection by gradient descent in function space," *J Mach Learn Res*, vol. 3, pp. 1333–1356, 2003.

[16] G. van de Wouwer, B. Weyn, P. Scheunders, W. Jacob, E. van Marck, and D. van Dyck, "Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei," *J Microsc*, vol. 197, no. 1, pp. 25–35, Jan. 2000.