Development of an Automatic Quantification Method
for Cancer Tissue Microarray Study

Teresa H. Sanders, *Student Member, IEEE*, Todd H. Stokes, *Student Member, IEEE*, Richard A. Moffitt, *Student Member, IEEE*, Qaiser Chaudry, R. Mitchell Parry, May D. Wang*, *Member, IEEE*

Abstract— **Clinical Histopathology is based on the analysis of immunohistochemistry (IHC) stained tissue images. Selection of antibodies for detecting the presence, type, and grade of cancerous tissue has a great influence on the diagnostic potential of IHC tests. Automated evaluation methods for tissue microarrays applied to many combinations of antibody and tissue type can speed development of new clinical assays. We present an automatic method that successfully quantifies stain intensity, fraction of cells stained and sub-cellular location of staining in tissue microarray images. The method combines an opponent color preprocessor and a novel statistical approach for identifying brown and blue staining, followed by multilevel morphological processing. We verify the capability of our method by comparing the results to manually annotated image databases. We also demonstrate cross-tissue robustness using two clinical case study data.**

I. INTRODUCTION

Antibody-based proteomics enables systematic exploration of the human proteome through analysis and quantification of cellular responses to specific antibodies. One such proteome database, the Human Protein Atlas (HPA) [1,2] contains thousands of antibody-stained tissue cross-section images. This database is a valuable resource because the images are manually annotated and curated by the human experts -- pathologists. For users who have new tissue images stained with disease biomarker candidates, a computer-based automatic processing system benchmarked to these human annotated images would be truly beneficial. Additionally, an automated, benchmarked system would enable users who have no local access to pathologists to conduct antibody-based proteome studies and compare with existing manually annotated HPA database images.

In this work, we demonstrate algorithms that automatically classify HPA tissue microarray (TMA) images by stain intensity, fraction of cells stained, and sub-cellular stain location using measures obtained at multiple levels:

pixel, cell / sub-cell, and image. While previous work has focused on either cell micro-arrays [3] or, if using TMAs, on sub-cellular pattern recognition [4,5], we calculate overall staining measures directly comparable to those provided by pathologists.

We verify the capability of our automatic method by comparing with existing HPA database annotations. The similarity of results produced by our method combined with the fact that a computer-based system is faster and more reproducible (i.e. less subject to variation caused by different human curators) makes our work a promising platform for future TMA IHC quantification guidelines.
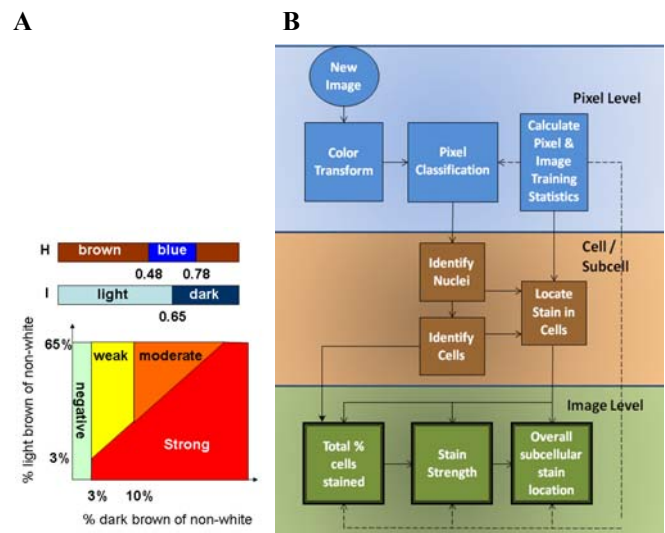

Fig. 1. A. HSI Simple classifier. B. Multilevel classification model flow diagram. The end results of the classification process are produced in the lower three blocks: total % cells stained, stain strength, and overall sub-cellular stain location.

II. SIMPLE CLASSIFICATION MODEL

Of the three image annotations tested in this paper, stain intensity was the simplest. We first developed a simple stain intensity classifier using standard image processing techniques. This classifier used the Hue and Intensity channels of the HSI color space to categorize pixels into 5 classes: white, light blue, dark blue, light brown, and dark brown. We selected two features: percent of light brown pixels and percent of dark brown pixels (of non-white pixels). The HPA image database has good consistency in contrast and high image quality, so we used basic thresholds selected using 54 training images to classify the images.

Our first classifier achieved 43.5% accuracy on the 210-

image head and neck cancer tissue database and 68% accuracy on the 91-image prostate cancer database. These results reflect that this simple classifier poorly models the process used by a pathologist to grade the image. To improve these results, we developed a multilevel classification model to use sub-cellular region identification in addition to pixel-level classification.

III. MULTILEVEL CLASSIFICATION MODEL

A. System Overview

Fig. 1 compares the design of the naïve classifier with the multilevel classifier. Pixel-level classification occurs first, followed by cell and sub-cellular segmentation, and finally image-level processing and classification. At each level, features are computed which feed into the image-level classification. Training of this system was done using 400 randomly-selected images from head and neck cancer tissues. None of the original 210 head and neck cancer or 91 prostate cancer test images were used in training.

B. Pixel Level Processing

1) Color

For pixel color classification, we extract a set of HPA training images for each of the four image strength classes annotated in the HPA (strong, moderate, weak, negative). We transform the pixels in each training image from the standard *red-green-blue* (rgb) color space into a 2-dimensional opponent color space [6] (*rg, by*) using the following equations:

$$rg = \text{red} - \text{green}$$
$$by = \tfrac{1}{2}(\text{red} + \text{green}) - \text{blue}$$

We chose this color space because it closely represents the way the human vision system experiences color [7], a matter of importance since human pathologists annotated our HPA TMA training images. For comparison, we tested the pixel classification described in Section II.B.3 using a 2-dimensional color space derived using Principal Component Analysis (PCA) in addition to the (rg, by) color space.

2) Intensity

The 2-dimensional color transform described above does not include an intensity component. This characteristic of the transform is advantageous when differentiating solely on hue. Hematoxylin stains nuclei dark blue while cytoplasmic staining may range from light to dark. The diaminobenzidine (DAB) staining also has a wide intensity range.

We use the achromatic pixel value Y, from the YIQ color space [6], for our intensity component since it is available as the MATLAB rgb to grayscale intensity conversion.

Y Intensity = 0.2989*red + 0.5870*green + 0.1140*blue

The resulting intensity ranges from 0 – 255 with grayscale white having a value of 255, and grayscale black a value of 0. The darker pixels then have *lower* grayscale intensities.

3) Pixel Classification

For pixel color classification, we form a nonparametric estimate of the 2-dimensional conditional probability density function for each of the four image strength classes by summing integer-binned 2-dimensional (*rg, by*) histograms of the training images. The maximum likelihood estimate of the strength class for each (*rg, by*) integer pair, $\hat{\theta}$, is then obtained using

$$\hat{\theta} = \arg\max_{\theta} f(x|\theta)$$

where x is the observed (*rg, by*) integer pair, Θ is the strength class (strong or negative), and f($x|\Theta$) is the corresponding probability density function estimate.

We use only images classified as negative or strong to train the pixel classifier, because moderate and weak images contain a complex mixture of brown and blue pixels, while strong and negative are dominated by either blue or brown. Fig 2 shows the RG-BY classifier used to distinguish between brown and blue pixels.
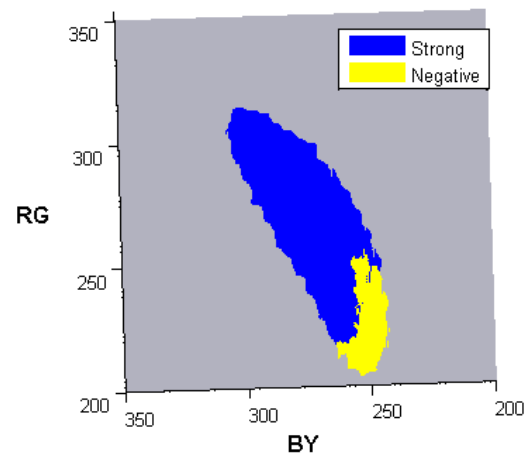

Fig. 2. Pixels classified as strong (blue), and negative (yellow), in the (rg,by) color space. The strong pixel region corresponds to brown pixels. The negative pixel region contains blue pixels.

We select fixed Y intensity thresholds based on visual inspection of a subset of the training images. These are 160 for brown and 142 for blue. Fig. 3 compares portions of original prostate cancer TMA images (top row) and the corresponding classified pixel images (bottom).
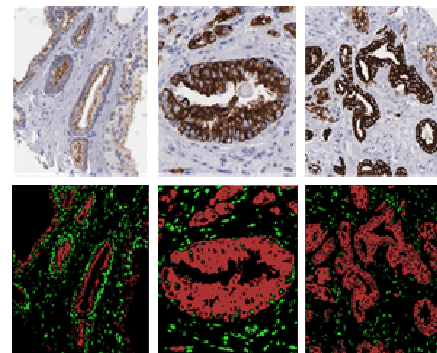

Fig. 3. Top Row: Snippets of HPA IHC stained prostate TMA images
Bottom Row: Pixels classified dark blue (green) and dark brown (red)

C. Sub-cellular Processing

1) Nuclei Segmentation

After transforming the pixels in the test images to (*rg*, *by*, intensity) space, we locate brown-stained nuclei by extracting brown pixels from the image, clustering dark brown pixels into "blobs", and filtering the blobs by area and aspect ratio to produce a set of "nuclei-like" objects. A parallel procedure is performed to locate the hematoxylin-stained blue nuclei. Fig. 4 shows the brown and blue nuclei detected for one of the head and neck cancer tissue samples.
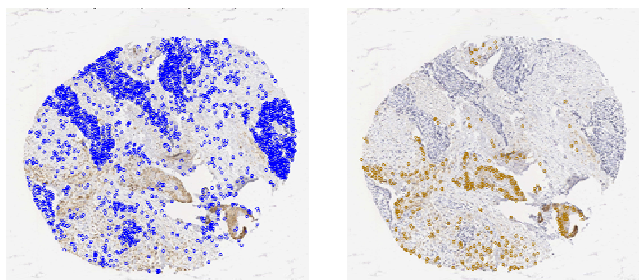

Fig. 4. Segmented blue (left) and brown (right) nuclei overlaid on image

2) Cell Identification

We then sample a cell-shaped and cell-sized region around each blue nucleus to obtain an overall estimate of the number and intensity of brown stained cytoplasmic and membranous pixels, as well as the number of blue nuclei with brown cytoplasmic/membranous stain in the image. We collect all the pixels within a fixed width border around the blue nuclei for cell sampling.

We do not address overlapping cells or clusters in this work; we assume these effects will be statistically similar for the various tissue samples so that the overall image classifications will be correct for most images. We do not observe overlapping nuclei in the test or training images.

D. Image Level Processing

1) Percent Staining

The percent of cells stained is calculated as follows from the number of nuclei stained brown, the number of blue nuclei with adjacent brown cytoplasmic or membranous stain, and the total number of blue stained nuclei:

% stain = (N_{brown_nuclei} + $N_{blue\_nuclei\_w\_brown\_cyt}$) / N_{total_nuclei}

2) Sub-cellular Localization

The sub-cellular localization of staining is determined to be either nuclear or cytoplasmic/membranous by comparing the number and intensity of brown nuclei pixels to the number and intensity of brown cytoplasmic and membranous pixels.

3) Stain Strength Classification

a) Features used for image strength classification

Six features are used:

- Number of brown nuclei,
- Average brown nuclei intensity,
- Number of blue nuclei with adjacent cytoplasmic or membranous brown stain,
- Average intensity of cytoplasmic and membranous brown stain,
- Total number of brown stained pixels in the image, and
- Average intensity of brown image pixels.

These six features are normalized, and their sum is used to classify the image stain strength as strong, moderate, weak, or negative.

b) Composite Scoring Method

The composite scoring method compares the features extracted from the test image to the statistics of the features for a training set of images as follows:

1. Midpoints between the means of adjacent classes (e.g. strong and moderate) are computed for each feature from the training data.
2. Incoming test image features are computed, compared to the midpoints, and assigned a score from 0 - 4 based on variance normalized distance from the closest mean.
3. Individual scores are summed to form the composite score. The composite score determines the overall image strength class using thresholds set to optimize the classification results for the training data.

c) Other Scoring Methods Tested

Two alternate overall classification methods using the features described in a) were tested:

1. a Support Vector Machine (SVM) and
2. a standard Mahalonobis distance method.

However, since neither approach showed improved classification and did not easily support the flexible scoring described in Section b) 2 above, we did not pursue them further in this work.

IV. PERFORMANCE RESULTS OF TWO CASE STUDIES

A. Head and Neck Cancer Images

In general, negatively stained images were straightforward to classify. However, the large number of antibodies tested introduced a significant amount of variability in brown staining. The moderate and weak stain strengths were more difficult to discriminate. The algorithm results for stain strength classification matched 80% of the HPA stain strength annotations for the 210 test images. Table I presents the confusion matrix between our classification and the pathologist annotations from HPA. The automatic sub-cellular localization algorithm matched the HPA annotation in 83% of cases. The average difference in % cell staining between the algorithm and the HPA annotation was 6%.

B. Prostate Cancer Images

The images tested were HPA sets of cancerous prostate tissue treated with stain-conjugated antibodies for proteins expressed by the following four genes mentioned in the literature as potential biomarkers for prostate cancer: AMACR [9], CTSL1, IAFP, and CD99. We selected five training images (independent of the 91 test images) from each strength class to compute the means, midpoints, and variances used for the overall classification. The pixel-level classification was done using the same classifier as head and neck with no additional training.

The algorithm results for stain strength classification matched 86% of the HPA stain strength annotations for the 91 test images. The average difference in % cell staining between the automatic annotation algorithm and the HPA human annotation was only 5%.

TABLE II
STAIN STRENGTH CONFUSION MATRIX FOR PROSTATE CANCER DATA

Algorithm Classification	HPA Stain Strength Label			
	Strong	Moderate	Weak	Negative
Strong	0.83	0.04	0.14	0.00
Moderate	0.00	0.86	0.14	0.00
Weak	0.17	0.03	0.72	0.12
Negative	0.00	0.07	0.00	0.88

V. DISCUSSION

This paper demonstrates a robust approach for multilevel classification of TMA images. The ability to accurately classify blue and brown pixels is one of the keys to automated analysis of DAB and hematoxylin stained tissue images. A second important factor for automated analysis is successful automated determination of nuclei locations. The algorithm presented accomplishes with reasonable accuracy. In particular, histiocytes were observed to pass the filtering criteria in several images. Future work will include template discrimination (SDF filters or Residual Vector Quantization [10] methods) for detection of nuclei and rejection of known nuclei-like objects.

Third, the imprecise sampling of the cytoplasm and membrane regions used in this study caused some errors in the classification results. More accurate sampling of these regions should produce better classification results. Level sets and Delauney triangulation are under consideration for improving this part of the system [11].

Although the stain classification for prostate cancer is typically based on staining close to the perimeter of the glands, the classification proposed in this paper does not limit the measure of staining to a particular region. Stained pixel samples are taken surrounding any valid nuclei. In spite of this obvious limitation, the algorithm performed reasonably well on a test case of prostate cancer tissue samples. Better region of interest selection and gland segmentation should provide improved results.

VI. CONCLUSION

Automatic annotation of stain strength and location is possible for a broad range of tissue and antibody types with the multilevel process described in this paper, which is not possible with standard image processing techniques. The software developed provides a faster, reproducible, computer-based automatic annotation system benchmarked to human annotation standards for antibody-based proteome studies. Other important applications include: 1) screening IHC TMA databases to identify images which statistically do not match similarly classified images in the database, and 2) automatically narrowing down the search for potential biomarkers in a large field of candidates.

REFERENCES

[1] http://proteinatlas.org/.

[2] M. Uhlen, E. Bjorling, C. Agaton, C. Szigyarto, et. al., "A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics", Molecular & Cellular Proteomics, vol. 4.12, pp. 1920-1932, August 2005.

[3] S. Stromberg, M. Bjorklund, C. Asplund, A. Skollermo, et. al., "A high-throughput strategy for protein profiling in cell microarrays using automated image analysis," Proteomics 2007, vol. 7, pp. 2142-2150, March 2007.

[4] J. Newberg and R. Murphy, "A Framework for the Automated Analysis of Subcellular Patterns in Human Protein Atlas Images", Journal of Proteome Research., vol. 7, pp2300-2308, April 2008.

[5] Camp, Chung, & Rimm, "Automated subcellular localization and quantification of protein expression in tissue microarrays", Nature Medicine, vol 8, 11, pp. 1323-1328, November 2002.

[6] P. K. Kaiser, R. M. Boynton, Human Color Vision, Optical Society of America, Washington, DC, 1996.

[7] G. Qui, J. Morris, X. Fan, "Visual guided navigation for image retrieval," Pattern Recognition 40, pp. 1711-1720, 2007.

[8] W. B. Schaming, "Adaptive Gate Multifeature Bayesian Statistical Tracker", RCA ADVANCED TECHNOLOGY LABS, 1982.

[9] K. Hammerich, G. Ayala, T. Wheeler, "Application of Immunohistochemistry to the Genitourinary System (Prostate, Urinary Bladder, Testis, and Kidney)", Archives of Pathology and Laboratory Medicine: Vol. 132, No. 3, pp. 432–440.

[10] C. F. Barnes, S. A. Rizvi, and N. M. Nasrabadi, "Advances in residual vector quantization: a review", IEEE Transactions on Image Processing, 5(2):226-262, February 1996.

[11] S. Naik, S. Coyle, S. Agner, A. Madabhushi, "Automated Gland and Nuclei Segmentation for Grading of Prostate and Breast Cancer Histopathology", IEEE ISBI, pp. 284-287, 2008.