J Pathol Inform

Editor-in-Chief:
Anil V. Parwani , Pittsburgh, PA, USA
Liron Pantanowitz, Pittsburgh, PA, USA

OPEN ACCESS
HTML format

For entire Editorial Board visit : www.jpathinformatics.org/editorialboard.asp

Research Article

Mapping stain distribution in pathology slides using whole slide imaging

Fang-Cheng Yeh[1,2], Qing Ye[2], T. Kevin Hitchens[2], Yijen L. Wu[2], Anil V. Parwani[3], Chien Ho[1,2]

Departments of [1]Biomedical Engineering, and [2]Biological Science, Pittsburgh NMR Center for Biomedical Research, Carnegie Mellon University, [3]Department of Pathology, Division of Pathology Informatics, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

E-mail: *Chien Ho - chienho@andrew.cmu.edu
*Corresponding author

Received: 01 July 13 Accepted: 03 December 13 Published: 31 January 14

This article may be cited as:
Yeh F, Ye Q, Hitchens TK, Wu YL, Parwani AV, Ho C. Mapping stain distribution in pathology slides using whole slide imaging. J Pathol Inform 2014;5:1.

Available FREE in open access from: http://www.jpathinformatics.org/text.asp?2014/5/1/1/126140

Abstract

Background: Whole slide imaging (WSI) offers a novel approach to digitize and review pathology slides, but the voluminous data generated by this technology demand new computational methods for image analysis. **Materials and Methods:** In this study, we report a method that recognizes stains in WSI data and uses kernel density estimator to calculate the stain density across the digitized pathology slides. The validation study was conducted using a rat model of acute cardiac allograft rejection and another rat model of heart ischemia/reperfusion injury. Immunohistochemistry (IHC) was conducted to label ED1$^+$ macrophages in the tissue sections and the stained slides were digitized by a whole slide scanner. The whole slide images were tessellated to enable parallel processing. Pixel-wise stain classification was conducted to classify the IHC stains from those of the background and the density distribution of the identified IHC stains was then calculated by the kernel density estimator. **Results:** The regression analysis showed a correlation coefficient of 0.8961 between the number of IHC stains counted by our stain recognition algorithm and that by the manual counting, suggesting that our stain recognition algorithm was in good agreement with the manual counting. The density distribution of the IHC stains showed a consistent pattern with those of the cellular magnetic resonance (MR) images that detected macrophages labeled by ultrasmall superparamagnetic iron-oxide or micron-sized iron-oxide particles. **Conclusions:** Our method provides a new imaging modality to facilitate clinical diagnosis. It also provides a way to validate/correlate cellular MRI data used for tracking immune-cell infiltration in cardiac transplant rejection and cardiac ischemic injury.

Key words: Stain distribution image, stain recognition, whole slide imaging

Access this article online
Website: www.jpathinformatics.org
DOI: 10.4103/2153-3539.126140
Quick Response Code:

INTRODUCTION

Optical microscopy has been a primary tool in pathology diagnosis for decades. The recent development of whole slide imaging (WSI) is rapidly making many advances in pathology and enabling the creation of novel tools and applications for the pathology community, including virtual microscopy,[1,2] and computer-aided quantification and diagnosis.[3-6] In contrast to conventional light microscopy that allows only a view of a fraction of a specimen at a time, WSI offers a digital replica of an entire histopathology slide. The data can be processed with a pattern recognition algorithm to identify features of interest.[7] This technique, called whole slide pattern

recognition, can provide more information than traditional pathology slide analysis, minimizes tedious visual inspection by trained pathologists and has the potential to improve clinical diagnosis. Implementing WSI analysis, however, faces the following challenges. First of all, a typical WSI dataset can contain a number of pixels at tera-scale (10^{12}) and previous computational methods used to identify targets on traditional histopathology images,[8,9] including texture-analysis,[10,11] nuclei pattern classification,[12,13] and edge detection,[14] are impractical for WSI due to computational time requirements. Parallel processing strategies have been developed to handle WSI;[10,15,16] however, boundary artifacts can jeopardize the integration of the analysis results and introduce a substantial amount of re-processing time. In addition, the abundant information provided by the whole slide analysis also requires a method to provide a quantitative assessment and to present meaningful results in a manner that is easy to interpret.[17]

In this paper, we report a WSI analysis method that conducts automatic stain recognition to estimate the density distribution of the recognized stains across an entire pathology slide. The whole slide images were tessellated into smaller image blocks with overlapping regions at the image boundary to enable parallel processing. This strategy ensures that the analysis result is free from the boundary effect. Pixel-based stain recognition with morphological smoothing[10,18,19] was applied to identify stains in the tessellated images. The density distribution of the identified stains was calculated by the kernel density estimator,[20,21] yielding a distribution image of a feature or a particular cell type labeled by the stains.

To validate our method, we applied it to specimen slides from an ongoing study using a rat cardiac allograft rejection model and also another study using a rat heart ischemia/reperfusion injury model. In both studies, the macrophage infiltration was revealed by immunohistochemistry (IHC) with anti-rat ED1 antibody. The accuracy of the stain recognition algorithm was first validated by comparing its results with manual counting without involving the WSI technique. The pathology images were obtained with a standard light microscope and the results of our stain recognition algorithm were compared with those obtained from manual counting. Then, our method was further validated/correlated with the results obtained from *in-vivo* cellular magnetic resonance imaging (MRI), an imaging technique that can detect macrophage infiltration when labeled with ultrasmall superparamagnetic iron-oxide (USPIO) or micron-sized iron-oxide (MPIO) particles.[22,23] Our method was also validated/correlated with the results obtained from *ex-vivo* cellular MRI, which were conducted using magnetic resonance microscopy (MRM) to achieve a much higher resolution for our final examination.[23]

To demonstrate the potential utilities of our method, we trained the algorithm to recognize different targets under hematoxylin and eosin (H and E) stain and to present their distribution across the tissue sections. We further calculated the diameter of the recognized targets and estimated the spatial distribution of cell nuclear diameter, which may have potential application in characterizing tissue in histopathology slides.

MATERIALS AND METHODS

Animal Models

The rat cardiac allograft rejection model, as described in a study,[23] was established using Dark Agouti and Brown Norway rats (Harlan, Indianapolis, IN) as the donor-recipient transplantation pairs. The superior vena cava of the graft heart was anastomosed to the recipient inferior vena cava (IVC) and the aorta of the graft heart was anastomosed to the recipient abdominal aorta, with the recipient proximal IVC partially obstructed to increase the pre-load. The graft, with intact pulmonary circulation, received proper pressure and volume loading without detectable atrophy over time. On day 7 after the transplantation, the graft hearts were scanned using *in-vivo* cellular MRI and then harvested for pathology inspections.

The rat myocardial ischemia/reperfusion injury model was established without thrombolysis component. BN rats were used in this experiment. A suture was tied around the left anterior descending artery for 45 min and then was cut to allow reperfusion. After 2 days, the heart was harvested and scanned by *ex-vivo* MRM.

Cellular MRI

Cellular MRI experiments were conducted by using dextran-coated USPIO particles[24] or polystyrene-coated MPIO particles (Bangs Laboratories, Fishers, IN) as an MRI contrast agent to image macrophage distribution.[22,23,25] The iron-oxide particles are taken up by macrophages in circulation and create signal-voids in T_2*-weighted images to reveal the macrophage infiltration *in-vivo*, offering a unique way to correlated the IHC stain distribution obtained from our method.

In our study, the iron-oxide particles were given to the animals before surgery. The rats, undergoing heterotopic heart transplantation, received USPIO particles and were scanned by *in-vivo* cellular MRI. The *in-vivo* MRI scan was conducted on a Bruker 7-T scanner (Bruker, Billerica MA). The rats were intubated and ventilated during the MRI scan and electrocardiogram (ECG) leads were placed on the abdomen using a gating and monitoring system (SA Instruments, Stony Brook, NY). T_2*-weighted magnetic resonance (MR) images, an imaging modality known to be sensitive to the signal void artifact created by iron-oxide particles,[23] were acquired

using a FLASH sequence with ECG and respiratory gating. TR = the respiration cycle (~1 s), TE = 5 ms, field of view (FOV) =4 cm, slice thickness = 1.5 mm, in-plane resolution = 156 μm.

The ischemia/reperfusion model animals received MPIO particles and the hearts were harvested. The *ex-vivo* MRI scans were conducted using a Bruker 11.7-Tesla scanner (Bruker, Billerica MA). T_2*-weighted MR images were also acquired using a FLASH sequence with TE = 5.8 ms, FOV = 1.2 cm, and isotropic resolution = 23 μm.

Specimen Preparation and IHC

The hearts were fixed in 4% paraformaldehyde for 1-2 weeks and embedded in paraffin. 5-μm transverse heart tissue sections were taken. Before staining, the sections were deparaffinized. Rehydrated through graded alcohols and rinsed with distilled water. Antigen retrieval was performed in a digital decloaking chamber (Biocare Medical, Cat. #DC2002), which streamed slides in preheated high pH target retrieval solution (Dako #S3308, Carpinteria, CA, USA) for 15 min. This was followed by cooling of the slides at room temperature for 20 min and rinsing the slides in distilled water before commencing the staining steps. To stain the slides, the tissue sections were incubated with 10% horse serum for 10 min at room temperature to block non-specific binding. The slides were then incubated with 1:200 anti-rat ED1 monoclonal purified immunoglobulin G (AbD SeroTec, Oxford, UK) for 2 h. After PBS wash, the slides were incubated with biotinylated secondary antibodies (LSAB kit, Dako, Carpinteria, CA, USA) for 30 min and labeled with the streptavidin-horseradish peroxidase (Dako, Carpinteria, CA, USA) for 30 min. Then the slides were stained with aminoethyl carbazole chromogen (SkyTek Laboratories, Inc., Logan, UT) and counterstained with Mayer's hematoxylin. After the slides were dehydrated, the coverslips were mounted. A rat spleen tissue section was used as a positive control section, whereas a normal rat cardiac tissue was used as a negative control section.

The stained slides were scanned on a whole slide scanner (Nanozoomer 2.0-HT, Hamamatsu, Japan) to acquire WSI at ×20 magnification. The resulting pathology images were then processed by our whole slide pattern recognition method, as detailed in the following section.

Whole Slide Pattern Recognition

The whole slide image was tessellated to enable the whole slide pattern recognition, as shown in Figure 1. The tessellation lines divide the whole slide image into several *n*-by-*n* sized blocks and the choice of *n* depends on the application scenario and the applied recognition algorithm. A higher value offers a larger field of view, but it may exceed the maximum allowable image size

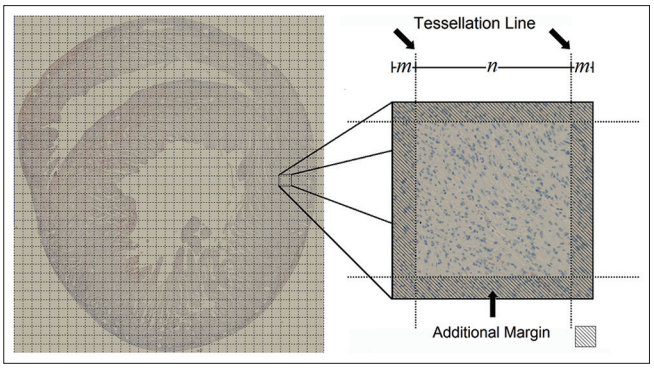

Figure 1: The tessellation scheme for the whole slide pattern recognition. The whole slide image is tessellated into n-by-n sized blocks to enable parallel recognition. Each block is patched with additional m-width margin to consider stains bisected by the tessellation boundary

for the algorithm. After tessellation, each block was padded with an additional margin, *m*-pixel wide, to the neighboring block. The padding setting was devised for accurate accounting of stains near the block boundary, since a strict boundary could bisect a stain, incorrectly assigning it to both neighboring blocks. The choice of *m* considers the span of individual stain and requires prior information about their morphology. An *m*-pixel-wide margin allows stains with a maximum span of *m* to be fully recognized within the tessellated image. In our study, we used $n = 1000$ and the value of *m* was set to 100.

The center of each recognized stain can be used to determine whether it should be counted or ignored to avoid double counting of stains, as illustrated by the example scenario in Figure 2. One should note that the tessellation borders are present at all sides of the tessellated image and Figure 2 shows only the border between two adjacent images blocks to simplify the illustration. As shown in Figure 2a, our stain recognition algorithm is applied to the image block to the left and the recognized stains with center points inside the tessellation line (annotated by yellow) are counted, whereas those outside the line (annotated by red) are ignored. Similarly, in Figure 2b, our stain recognition algorithm is applied to the image block to the right. The recognized stains with center points inside the tessellation line (annotated by yellow) are counted, whereas the rest are ignored (annotated by red). This approach avoids any double counting and ensures the consistency of the results.

Stain Recognition Algorithm

The flow chart of our stain recognition algorithm is shown in Figure 3. The method is composed of three modules, including a classification module that performs pixel-wise color classification, a morphology smoothing module that eliminates the fragments considered to be false positives and an analysis module that identifies

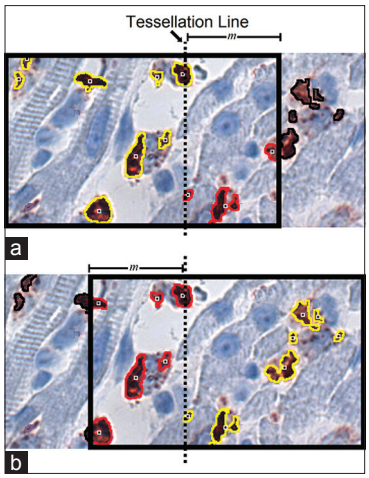

Figure 2: The recognition scheme that avoids double counting. (a) The stain recognition is applied to the image block to the left to recognize the stains. The center locations of the stains (marked by dot points) are used to determine whether a recognized stain should be counted (annotated by yellow) or ignored (annotated by red). (b) The stain recognition is applied to the image block to the right, showing how double counting is avoided. The counted stains (annotated by yellow) have center points located inside the tessellation line

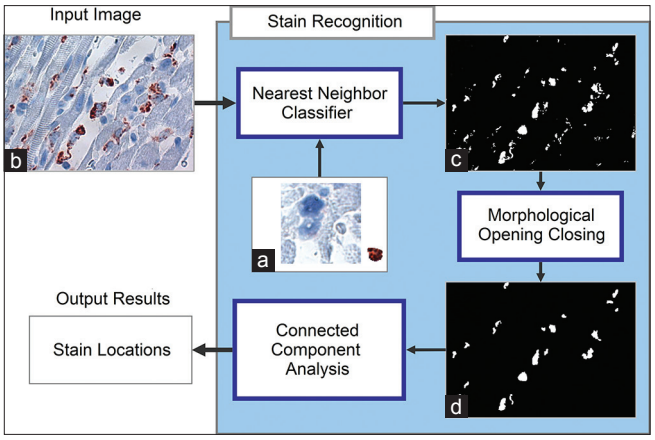

Figure 3: The flow diagram of our automatic stain recognition algorithm. The pixel-wise classification is trained by a foreground stain image and also a background stain image (a). Each pixel of the input image (b) is then classified using nearest neighbor classification to obtain a binary image (c). The small fragments in the binary image are then eliminated by morphological closing and opening, resulting in a smoother output (d). This image is then analyzed by connected component analysis to obtain the center locations of the recognized stains

each isolated stain region. It is noteworthy that a variety of stain recognition algorithms can be applied to this problem, and in this study, we have used the simplest pixel-wise color recognition approach for its low computation time and stable performance. This algorithm viewed the stain recognition as a classification problem. The red-green-blue color channels were used as the classification features and the output (classification result) was either 0 (background) or 1 (foreground). The stain classification was conducted using nearest neighbor method, a popular method in data mining.[26] Nearest neighbor method required a training dataset of background and foreground stain image. In this study, we manually cropped the background and foreground stain from the WSI data as the training dataset. Alternatively, one may also get training data from the positive and negative controls. The nearest neighbor method classified each image pixel based on its color difference with the pixels in the training images. If the minimum difference was from the foreground stain, then the pixel was classified as foreground; otherwise, it was classified as background. This classification was applied to each pixel in the input image [Figure 3b] to obtain the binary recognition output [Figure 3c]. This output image might have small fragments due to error classification or the partial slice coverage of the stained cells. To eliminate fragments in the binary images and to facilitate analysis, morphological opening and closing were applied 3 times using a 3-by-3 square structuring element to smooth the contours of the recognized stains,[27] as shown in Figure 3d. Finally, the connected component analysis[28] identified each isolated stain region in the image and the center location of each stain was calculated for further density estimation.

Density Estimation and Feature Mapping

Since the locations of the recognized stains were discrete points scattered across the WSI, a regression analysis method was used to calculate their density distribution. This was carried out by the kernel density estimator[20,21] as shown in Figure 4, where Figure 4a shows an exemplary result of the discrete stain locations and Figure 4b shows their density distribution estimated by the kernel density estimator. We had a total of k stains obtained from our whole slide pattern recognition and each of them was located at coordinate x_i. The kernel density estimator estimated the density distribution function by using the following formulae:

$$I_h(x) = \frac{1}{k}\sum_{i=1}^{k}\varphi\left(\left|\frac{x-x_i}{h}\right|\right) \quad (1)$$

where h is the kernel bandwidth and is the kernel function. The scalar feature (e.g. diameter or area) of the k recognized stains can also be estimated using kernel regression.

$$f_h(x) = \frac{\sum_{i=1}^{k} y_i\varphi\left(\left|\frac{x-x_i}{h}\right|\right)}{\sum_{i=1}^{k}\varphi\left(\left|\frac{x-x_i}{h}\right|\right)} \quad (2)$$

where y_i is a scalar feature of the i^{th} stain. Equation 2 can be used to estimate the spatial distribution of the feature across the entire pathology slide.

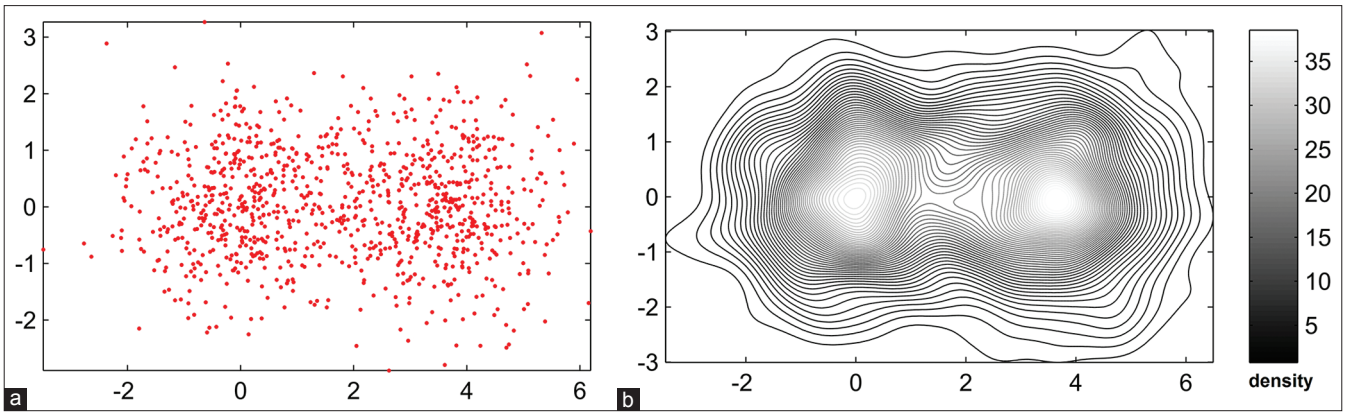

Figure 4: The exemplary analysis result of the kernel density estimator. (a) An example image shows the discrete locations of the recognized stains. (b) Their density distribution can be estimated by the kernel density estimator

The bandwidth of the kernel determines the variability of the density estimation. A higher value of the bandwidth generates a smoothing effect, leading to reduced variability. On the contrary, a lower value creates a sharper estimation, but may cause over-fitting and larger variability. In this study, we used a data-driven method to automatically determine the bandwidth.[29] We used Equation 1 to estimate the ED1 stain distribution images and Equation 2 to estimate the spatial distribution of the stain feature. The source codes of our whole slide recognition program (WS Recognizer) are publicly available at http://ws-recognizer.labsolver.org.

Manual Validation

We first validated our stain recognition algorithm without involving WSI. The validation was conducted using a total of 4 rats from the acute cardiac allograft rejection study in our laboratory. For each rat, one tissue section was obtained from the mid portion of the excised heart. For each slide, 16 pathology images were acquired on a transmitted light microscope (Olympus Proves AX 70) at ×40 objective lens. The field of view covered different areas in the slide to acquire images with different macrophage density. In each pathology image, manual counting and automatic stain recognition were conducted separately to avoid a spurious correlation between the manual counting and the automated recognition. Manual counting was conducted without knowing the stain recognition results and the training data for automatic recognition were selected without knowledge of the manual-counting results. The final comparison was conducted by performing regression analysis to calculate the correlation coefficient.

Cellular MRI Validation

To validate our method on WSI data, we compared the ED1 stain distribution image with both *in-vivo* and *ex-vivo* cellular MR images. To facilitate a visual comparison, the orientations of the MR images were manually rotated to match those of the pathology images. The hypointensity patches shown in the T_2*-weighted images were due to the accumulation of the iron-oxide labeled macrophages,[23] a feature that allowed us to examine whether our ED1 stain distribution image was in good agreement with the distribution shown on cellular MRI.

RESULTS

The sample result of the stain recognition algorithm is shown in Figure 5a, where the recognized stains are annotated by the yellow contour. The training images are presented in Figure 5b and c, which are the background and foreground IHC stain images, respectively. The image shows the infiltration of $ED1^+$ macrophages due to acute cardiac allograft transplant rejection and the brownish IHC stains were identified by our recognition method. The total numbers of macrophages were counted by summing up the number of annotated regions in the image and the center point of each annotated region can be calculated to facilitate further stain density estimation.

The result of the manual validation is shown in Figure 6, where the numbers of IHC stains counted by our automated stain recognition algorithm were correlated with manual counting. The analysis included 64 observations obtained from 64 pathology images. The regression line and 95% of confidence band are also plotted in the figure. The analysis shows an $R^2 = 0.8029$ and a correlation coefficient of 0.8961 between the number of cells counted manually and that recognized by our algorithm, suggesting that the results obtained from our automatic recognition algorithm was in good agreement with the manual counting results. The narrow bank of the confidence interval further suggests that the high correlation result is highly reproducible. One may note that the coefficient of the regression line was 0.445, meaning that the automatic recognition algorithm tends to count twice as many as that by the manual counting. The systematic overestimation of the algorithm could be due to the fact that the algorithm counted the IHC stains regardless of their size, whereas manual counting

may have ignored small stains. A feasible approach to improve consistency is adding a size filter to the stain recognition algorithm. Nonetheless, this discrepancy does not invalidate our stain distribution images, since the high correlation coefficient ensured that our recognition algorithm revealed the same distribution pattern and only differed by a scaling factor.

The *in-vivo* cellular MRI is shown in Figure 7, where the T_2*-weighted image (left) was compared with the ED1 stain distribution image (right). The animal from our rat heart transplantation model developed acute cardiac rejection and USPIO was used as a macrophage-labeling agent in the cellular MRI. As shown in Figure 7a, the signal-void artifact in the T_2*-weighted MR image, as illustrated by the arrow, suggested the accumulation of USPIO-labeled macrophages in the endocardium of the left ventricle. In comparison, our ED1 stain distribution image shown in Figure 7b illustrated a similar pattern of accumulation of ED1$^+$ macrophages in the endocardium. One may note that the T_2*-weighted image does not match perfectly with the ED1 stain distribution image. This may be due to the fact that there could be tissue distortion during the fixation procedure and that the images may not align perfectly with the pathology slides. Despite this difference, the highly similar pattern suggested that our method can be used to reveal ED1$^+$ macrophages within pathology slides.

The *ex-vivo* MRM is shown in Figure 8, where the T_2*-weighted image of the *ex-vivo* tissue (left) was

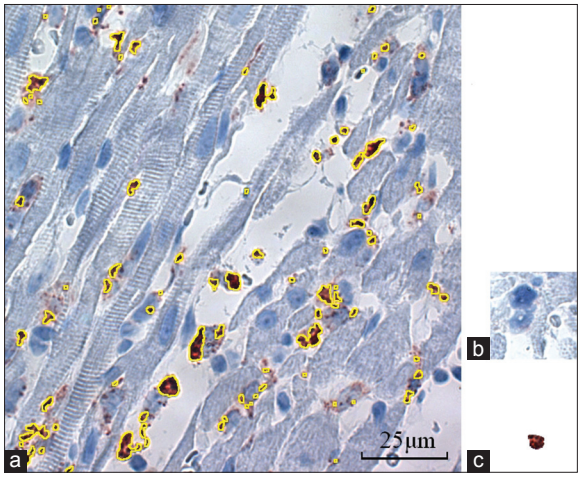

Figure 5: The result of our stain recognition and the training images used in the stain classification. (a) The stain recognition is applied to a pathology image of myocardium under acute cardiac allograft rejection, where macrophage accumulation revealed by immunohistochemistry (IHC) stains can be recognized by our algorithm. The stain recognition algorithm is trained by using one background image (b) and one foreground IHC stain image (c) and the recognized stains are annotated by yellow to illustrate the performance of our algorithm

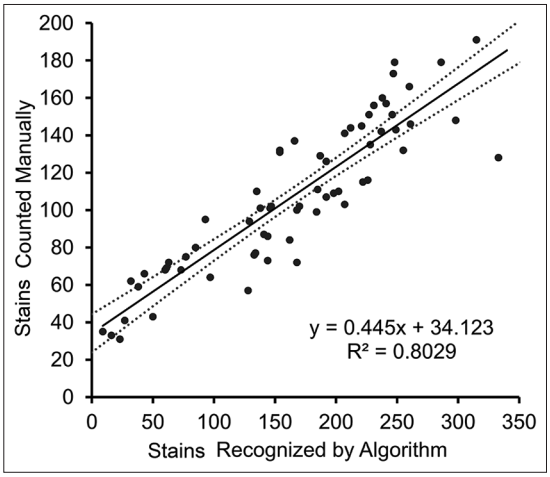

Figure 6: The regression analysis shows the correlation between the number of stain recognized by our stain recognition algorithm and that by the manual counting. The analysis has included 64 observations and the regression line is presented between the 95% confidence band. The regression analysis shows an R^2 = 0.8029 and a correlation coefficient of 0.8961, suggesting a high correlation between manual counting and our stain recognition algorithm

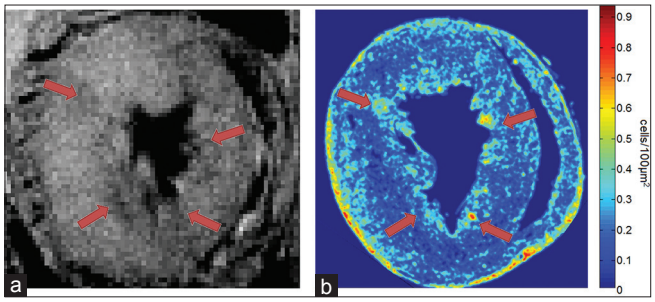

Figure 7: The in-vivo cellular magnetic resonance (MR) images correlates with the ED1 stain distribution image. (a) The in-vivo T_2*-weighted cellular MR image, where the macrophages labeled with ultrasmall superparamagnetic iron-oxide particles create hypo-intensity patches (annotated by arrows) in the myocardium that develops acute allograft rejection. (b) The corresponding ED1 stain distribution image shows a similar pattern of the macrophages accumulation in the endocardium. The hypo-intensity patches in (a) correspond to the hyper-intensity locations in (b)

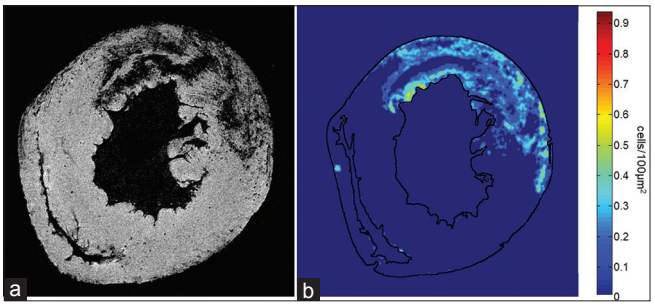

Figure 8: The ex-vivo cellular magnetic resonance microscopy (MRM) correlates with the ED1 stain distribution image. (a) The ex-vivo T_2*-weighted cellular MRM, where the macrophages labeled with micron-sized iron-oxide particles create hypo-intensity patches in the myocardium that has ischemia/reperfusion injury (b) The corresponding ED1 stain distribution image shows highly similar pattern of macrophages accumulation in the myocardium, suggesting that the ED1 stain distribution image can reveal the macrophage infiltration across an entire pathology slide

compared with the ED1 stain distribution image (right). The rat was from our model of ischemia/reperfusion injury, with MPIO particles used as the macrophage labeling agent in the cellular MRM. The tissue section is outlined (black) in the ED1 stain distribution image to facilitate comparison. As shown in Figure 8a, the MPIO-labeled macrophages generate dark patches and spots in the MRM images, suggesting massive macrophage infiltration in the myocardium with ischemia/reperfusion injury. Similarly, our ED1 stain distribution image shown in Figure 8b showed the same pattern of $ED1^+$ macrophage infiltration in the left ventricle. The high similarity of both imaging approaches confirmed that our method can reliably identify $ED1^+$ macrophages within pathology slides.

Figure 9a shows the spatial distribution of the cell nucleus, lipofusin and red blood cell in 3 rats with heart ischemia/reperfusion injury. The stain recognition algorithm was trained to recognize cell nuclei (N), lipofusin (L) and red blood cells (R) under H and E stain, as shown in Figure 9b. The distribution of cell nuclei can be used to investigate the increase of immune cells during the inflammation process, whereas the distribution of lipofusin may be used to investigate oxidative stress in the myocardium. The distribution of red blood cells indicates hemorrhage due to ischemic injury. Both three distribution images offer paranomic pathology information of the ischemia/reperfusion injury across tissue sections.

Figure 10 is an example of feature distribution image showing the spatial distribution of the cell nuclear diameter. The diameter of the recognized cell nuclei was estimated using the kernel regression formulated in Equation 2. The feature distribution image can be used to provide a quantitative evaluation of a particular feature and to assist tissue characterization in histopathology.

DISCUSSION

In this paper, we report on an automated method that conducts whole slide pattern recognition to obtain the stain distribution. The results of our stain recognition algorithm correlate well with manual counting. Regression analysis showed a correlation coefficient of 0.8961 between manual counting and our stain recognition algorithm, suggesting that the labor intensive manual counting may be replaced by our automated recognition algorithm, which is critical for further analysis of the image data from WSI. Moreover, the ED1 stain distribution image was in good agreement with both *in-vivo* cellular MRI and *ex-vivo* cellular MRM. The high density areas revealed in the ED1 stain distribution images correlated well with the hypo-intensity areas generated by iron-oxide-labeled macrophages, suggesting

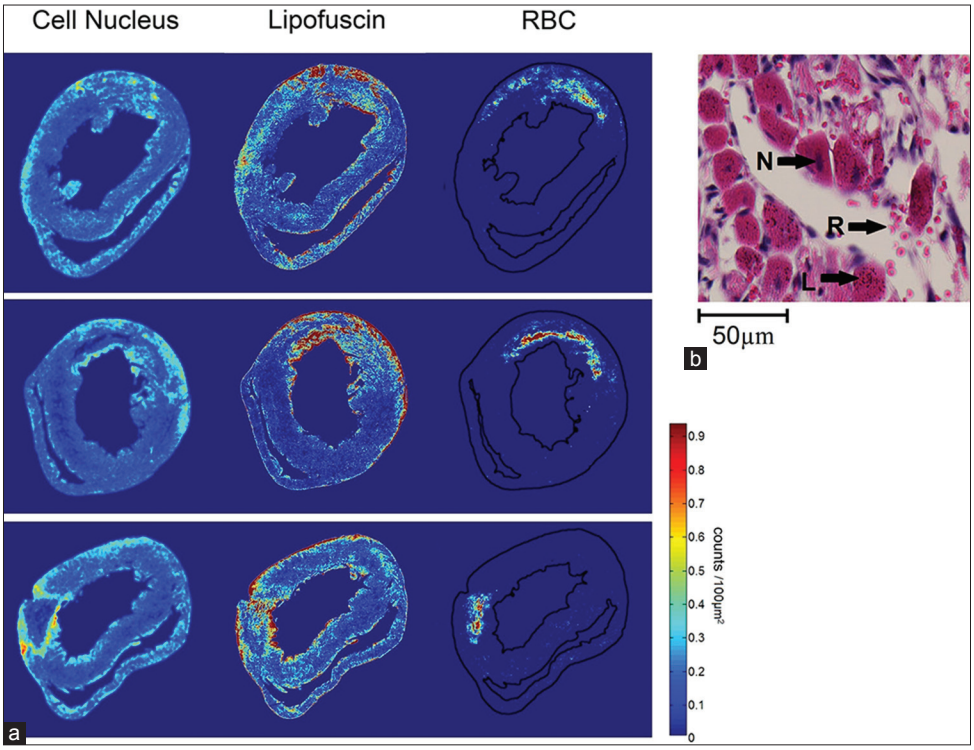

Figure 9: The spatial distribution of cell nucleus, lipofusin and red blood cell in rat hearts with ischemia/reperfusion injury. (a) The spatial distribution of cell nucleus, lipofusin and red blood cell in 3 rats. (b) The stain recognition algorithm was trained to cell nuclei (N), lipofusin (L) and red blood cells (R) under H and E stain. These three distribution images offer paranomic pathology information that facilitates quantitative characterization in histopathology

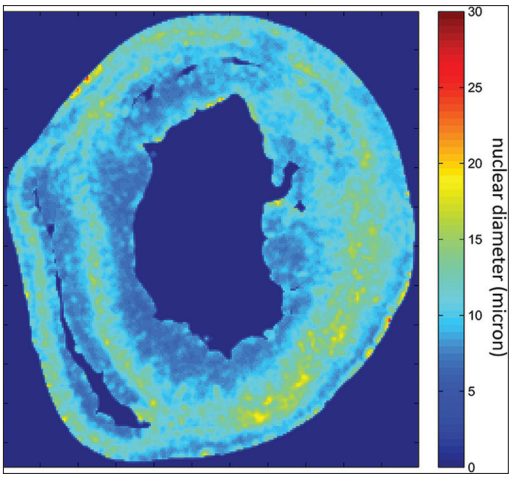

Figure 10: A feature distribution image showing the spatial distribution of the cell nuclear diameter. The cell nuclei were recognized under H and E stain and the spatial distribution of their diameter can be estimated using kernel regression. This feature distribution image may assist histopathologic characterization and quantification

that our method can serve as a surrogate to reveal the distribution of macrophages across tissue sections.

The novelty of our approach includes the following: First, we have implemented a parallel processing approach that makes whole slide pattern recognition possible. We have demonstrated the performance using a pixel-wise stain recognition algorithm. It is possible that the same processing approach can be combined with more advanced pattern recognition algorithms to extract meaningful features from pathology/IHC digital slides. Second, the stain distribution images obtained from our method offered a panorama of pathology information instead of a fraction of a view under the light microscope. This panoramic view is less susceptible to errors caused by subjective selection of the data. In contrast, traditional pathology validation often requires expert selection of the lesion site, making the process subjective and potentially biased. Whole slide pattern recognition leaves no bias and presents all the information across the pathology slide to offer more objective results. Third, the stain distribution images are quantitative information that can be further analyzed statistically. This unique feature is provided by the kernel density estimator, which is a widely-used regression approach in the field of statistics and has been used in a variety of real world applications.[30] Lastly, our method can be fully automated and no human intervention is needed to obtain the stain distribution mapping. This offers a highly objective and efficient method that may enable a high throughput processing of data and massive screening.

Despite the many advantages mentioned above, there are several possible pitfalls in our method that are noteworthy of discussion. The accuracy of our method depends on the performance of the recognition algorithm and it is possible that an algorithm may give false positive results secondary to poor slide preparation or artifacts in the preparation. Similarly, the stain variability may alter the recognition results and may create a false perception of the changes in the stain distribution. Both pitfalls may lead to misleading results that are hard to identify and thus, a post-processing inspection may be needed for quality control.

There are also limitations in our experiment design. First, we did only a qualitative comparison and did not conduct rigorous registration to compare our stain distribution imaging with MRI. This limitation is due to the fact that the tissue specimen deforms substantially in the fixation process and the spatial information between MRI and WSI cannot be perfectly aligned. Furthermore, the *ex-vivo* MRI of the tissue-embedded paraffin block cannot be acquired because the MR spectrum is dominated by the paraffin resonance. Another limitation of our experiment is that we did not examine the positive and negative predicting values of our stain recognition algorithm. Although the accuracy of the morphology-based stain recognition has already been examined,[10,18,19] it is unclear whether the same performance can be achieved in counting macrophages due to their highly pleomorphic shapes.

Our stain recognition algorithm also has limitations. The current setting is optimized to identify stains with an oval shape, but there can be other morphological presentations in the pathology slides. To overcome this limitation, the algorithm can be customized to accommodate different morphology of the stains. Moreover, although there is a positive correlation between the number of the recognized stains and manual cell counts, our stain recognition algorithm still tends to count more than that of manual counting. This may be due to the fact that macrophages have cytoplasmic projections that can lead to separated regions and result in duplicated counts. In addition, the nuclei of macrophages are often lobulated, further adding the complexity of the cell recognition in our use case scenario. Counting the number of macrophages is still a challenging task that requires more future studies. Last, the stain distribution value can be affected by several factors, including the thickness of the sections, the z-depth and the bandwidth used in the kernel density estimator. Therefore, a comparison based on the stain-distribution value should be conducted in a carefully controlled manner.

CONCLUSIONS

A method is proposed to obtain stain distribution across a pathology slide. The obtained stain distribution image represents a panorama of the pathology information and can be used in both qualitative and quantitative analyses. This imaging approach may also assist pathologists in

clinical diagnosis by offering them a toolkit for multiple applications in diagnostic pathology as well as yield newer methods to use image analysis to quantitate and analyze biomarkers, which are both diagnostic and prognostic. In addition, the algorithm described in this study may serve as a new and systematic approach to validate other imaging modalities. The ED1 stain distribution image confirms that the cellular MRI technique is able to detect the macrophage infiltration of cardiac transplantation rejection or cardiac ischemic injury in our rat models.

ACKNOWLEDGEMENT

We would like to thank Ms. Lily Chen and Mr. Brent Barbe for manual counting of the macrophages in the IHC slides. Our research is supported by NIH grants (P41EB-001977 and R01HL-081349)

REFERENCES

1. Weinstein RS, Graham AR, Richter LC, Barker GP, Krupinski EA, Lopez AM, *et al.* Overview of telepathology, virtual microscopy, and whole slide imaging: Prospects for the future. Hum Pathol 2009;40:1057-69.
2. Wilbur DC, Madi K, Colvin RB, Duncan LM, Faquin WC, Ferry JA, *et al.* Whole-slide imaging digital pathology as a platform for teleconsultation: A pilot study using paired subspecialist correlations. Arch Pathol Lab Med 2009;133:1949-53.
3. Velez N, Jukic D, Ho J. Evaluation of 2 whole-slide imaging applications in dermatopathology. Hum Pathol 2008;39:1341-9.
4. Feldman MD. Beyond morphology: Whole slide imaging, computer-aided detection, and other techniques. Arch Pathol Lab Med 2008;132:758-63.
5. Gilbertson JR, Ho J, Anthony L, Jukic DM, Yagi Y, Parwani AV. Primary histologic diagnosis using automated whole slide imaging: A validation study. BMC Clin Pathol 2006;6:4.
6. Fallon MA, Wilbur DC, Prasad M. Ovarian frozen section diagnosis: Use of whole-slide imaging shows excellent correlation between virtual slide and original interpretations in a large series of cases. Arch Pathol Lab Med 2010;134:1020-3.
7. Isse K, Lesniak A, Grama K, Roysam B, Minervini MI, Demetris AJ. Digital transplantation pathology: Combining whole slide imaging, multiplex staining and automated image analysis. Am J Transplant 2012;12:27-37.
8. Kayser K, Görtler J, Bogovac M, Bogovac A, Goldmann T, Vollmer E, *et al.* AI (artificial intelligence) in histopathology – From image analysis to automated diagnosis. Folia Histochem Cytobiol 2009;47:355-61.
9. Mulrane L, Rexhepaj E, Penney S, Callanan JJ, Gallagher WM. Automated image analysis in histopathology: A valuable tool in medical diagnostics. Expert Rev Mol Diagn 2008;8:707-25.
10. Sertel O, Kong J, Shimada H, Catalyurek UV, Saltz JH, Gurcan MN. Computer-aided prognosis of neuroblastoma on whole-slide images: Classification of stromal development. Pattern Recognit 2009;42:1093-103.
11. Muthu Rama Krishnan M, Shah P, Chakraborty C, Ray AK. Statistical analysis of textural features for improved classification of oral histopathological images. J Med Syst 2012;36:865-81.
12. Rohde GK, Ribeiro AJ, Dahl KN, Murphy RF. Deformation-based nuclear morphometry: Capturing nuclear shape variation in HeLa cells. Cytometry A 2008;73:341-50.
13. Wang W, Ozolek JA, Rohde GK. Detection and classification of thyroid follicular lesions based on nuclear structure from histopathology images. Cytometry A 2010;77:485-94.
14. Loukas CG, Wilson GD, Vojnovic B, Linney A. An image analysis-based approach for automated counting of cancer cell nuclei in tissue sections. Cytometry A 2003;55:30-42.
15. DiFranco MD, O'Hurley G, Kay EW, Watson RW, Cunningham P. Ensemble based system for whole-slide prostate cancer probability mapping using color texture features. Comput Med Imaging Graph 2011;35:629-45.
16. Samsi S, Krishnamurthy AK, Gurcan MN. An efficient computational framework for the analysis of whole slide images: Application to follicular lymphoma immunohistochemistry. J Comput Sci 2012;3:269-79.
17. Kothari S, Phan JH, Stokes TH, Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. J Am Med Inform Assoc 2013;20:1099-108.
18. Veta M, van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JP. Automatic nuclei segmentation in H and E stained breast cancer histopathology images. PLoS One 2013;8:e70221.
19. Gurcan MN, Pan T, Shimada H, Saltz J. Image analysis for neuroblastoma classification: Segmentation of cell nuclei. Conf Proc IEEE Eng Med Biol Soc 2006;1:4844-7.
20. Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Stat 1956;27:832.
21. Parzen E. On estimation of a probability density function and mode. Ann Math Stat 1962;33:1065.
22. Kanno S, Wu YJ, Lee PC, Dodd SJ, Williams M, Griffith BP, *et al.* Macrophage accumulation associated with rat cardiac allograft rejection detected by magnetic resonance imaging with ultrasmall superparamagnetic iron oxide particles. Circulation 2001;104:934-8.
23. Wu YL, Ye Q, Foley LM, Hitchens TK, Sato K, Williams JB, *et al. In situ* labeling of immune cells with iron oxide particles: An approach to detect organ rejection by cellular MRI. Proc Natl Acad Sci U S A 2006;103:1852-7.
24. Dodd SJ, Williams M, Suhan JP, Williams DS, Koretsky AP, Ho C. Detection of single mammalian cells by high-resolution magnetic resonance imaging. Biophys J 1999;76:103-9.
25. Chang HH, Moura JM, Wu YL, Ho C. Automatic detection of regional heart rejection in USPIO-enhanced MRI. IEEE Trans Med Imaging 2008;27:1095-106.
26. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, *et al.* Top 10 algorithms in data mining. Knowl Inf Syst 2008;14:1-37.
27. Dougherty ER. An Introduction to Morphological Image Processing. 1st ed. Bellingham, Washington: SPIE-International Society for Optical Engine; 1992.
28. Dillencourt MB, Samet H, Tamminen M. A general approach to connected-component labeling for arbitrary image representations. JACM 1992;39:253-80.
29. Botev ZI, Grotowski JF, Kroese DP. Kernel density estimation via diffusion. Ann Stat 2010;38:2916-57.
30. Sheather SJ. Density estimation. Stat Sci 2004;19:588-97.