How to Automatically Identify Regions of Interest in High-resolution Images of Lung Biopsy for Interstitial Fibrosis Diagnosis

Oscar Cuadros Linares
Institute of Mathematics and Computer Sciences
University of São Paulo
São Carlos, Brazil
ocuadros@icmc.usp.br

Bruno S. Faiçal
Institute of Mathematics and Computer Sciences
University of São Paulo
São Carlos, Brazil
bsfaical@alumni.usp.br

Paulo Renato C. Barbosa
Institute of Mathematics and Computer Sciences
University of São Paulo
São Carlos, Brazil
paulorcb@usp.br

Bernd Hamann
Department of Computer Science
University of California
Davis, California, U.S.A.
hamann@cs.ucdavis.edu

Alexandre T. Fabro
Ribeirão Preto Medical School
University of São Paulo
Ribeirão Preto, Brazil
fabro@fmrp.usp.br

Agma J. M. Traina
Institute of Mathematics and Computer Sciences
University of São Paulo
São Carlos, Brazil
agma@icmc.usp.br

***Abstract*—Airway-centered Interstitial Fibrosis (ACIF) is a histological pattern of Interstitial lung diseases. Its diagnosis requires a multidisciplinary approach, in which diverse information, such as clinical data, computed tomography data, and lung biopsy data, is analyzed. Biopsy samples are digitized at high-resolution. Of crucial interest are broncho- and bronchiolocentric remodeling with extracellular matrix deposition. To analyze an image, specialists have to explore it at low microscope magnification, select a region of interest and export a smaller specified sub-image to be interpreted at higher magnification. This process is performed several times, requiring hours, becoming a tiresome task. We propose a method to support pathologists to identify specific patterns of ACIF in high-resolution images from lung biopsies. This can be done by a) automatic microscope magnification reduction; b) computing the probability of pixels belonging to high-density regions; c) extracting Local Binary Patterns (LBP) of the high- and low-density regions; and d) visualizing them in color. We have evaluated our method on nine high-resolution lung biopsies. We have tested the LBP features of high- and low-density regions with the kNN algorithm and obtained a classification accuracy of 94.4%, which is the highest one reported in the literature for this type of data.**

***Index Terms*—Airway-centered Interstitial Fibrosis, high-resolution lung biopsy, image analysis, quantitative assessment of microscopic images, interstitial fibrosis diagnosis.**

I. Introduction

Pulmonary fibrosis belongs to a heterogeneous group of interstitial lung diseases in which lung tissue becomes thick, stiff and scarred over time. The most serious consequence is that it progressively hinders the proper functioning of the lungs, leading to shorter of breath and hypoxia [1]. So far, these diseases could not be cured, often with poor prognosis.

Airway-centered Interstitial Fibrosis (ACIF) is a recently (2002) described histological pattern of ILD with variable etiology and outcome, which can be idiopathic or with a history of environmental exposure, such as smoke, birds, chalk dust, agrochemicals, and cocaine [2]. It is mainly characterized by broncovascular remodeling and metaplastic bronchiolar. Clinically patients present cough and progressive dyspnea. ACIF is a disease with a variable prognosis of 40% mean survival in 10 years [3].

To diagnose ACIF, a multidisciplinary approach is required, in which biopsies are analyzed in three progressive steps: a) clinical data such as patients age, symptoms, and environmental exposure; b) radiological imaging, X-ray and Computed Tomography (CT). Typical signs of ACIF are often present on high-resolution CT, which can include bronchiectasis, airway-centered reticular infiltrates, ground-glass opacity, and parabronchial interstitial thickening [2]; c) histopathology, usually indicated as a surgical lung biopsy when CT patterns are indeterminate or inconsistent to diagnose Idiopathic Pulmonary Fibrosis (IPF) or when clinical data suggest an alternative diagnosis [4]. Biopsy samples are analyzed at low and high microscope magnification. In ACIF the key histopathological finding is the airway interstitial fibrosis centered on membranous and respiratory bronchioles [3]. At higher magnification, patterns, such as chronic inflammatory infiltrates, and muscle hyperplasia in the wall of bronchioles, can be found.

Identify and categorize ILDs is a very complex, time-demanding and tiresome task, since a pathologist could take many hours analyzing a lung biopsy sample. When a biopsy

The authors would like to thank FAPESP for the financial support to the project "Mining, indexing and visualizing Big Data in clinical decision support systems (MIVisBD)" (Grants 16/17078-0, 18/06074-0, and 18/06228-7). The authors also thank CNPq for the financial support grant 144908/2018-2.

sample is digitized a high-resolution file is generated. For example, a digitized biopsy with $40x$ magnification can produce an image of about $(110,000 \times 110,000)$ or $12,100,000,000$ pixel resolution, which generates a file of approximately 2.0 GB. To analyze such an image, the specialist has to find ACIF patterns, or ROIs, seen in a very low microscope magnification (digitized reduction) and export them to "smaller" images (usually in Tagged Image File Format, (TIFF)) to be analyzed later. This process is performed several times, and it requires many hours, not only because the pathologist takes time finding ROIs, but also exporting them requires long computational processing. In addition, there does not exist a software tool or method capable of automatically finding ROIs in low microscope magnification or discovering ACIF patterns at higher magnification.

We present a method to automatically segment regions of interest (ROI) and extract features in order to identify ACIF patterns found in a low microscope magnification. The main goal is to *accurately* and *quickly* highlight the most remodeled regions (mainly fibrosis centered and extending around the bronchioles), candidates being indicative of the diseases. Thus, the specialists can analyze them in high microscope magnification looking for more specific ACIF patterns. By doing so the specialist's time and effort is best utilized.

II. Airway-centered Interstitial Fibrosis Method

During the process of airway-centered interstitial fibrosis (ACIF) bronchi may exhibit loss of structural integrity, such as muscularized bronchioles, bronchiolar metaplasia, proliferative bronchiolitis, and interstitial thickening. These patterns can be seen, using low microscope magnification, as high-density pixel regions. Examples of these regions are shown in Figure 1.

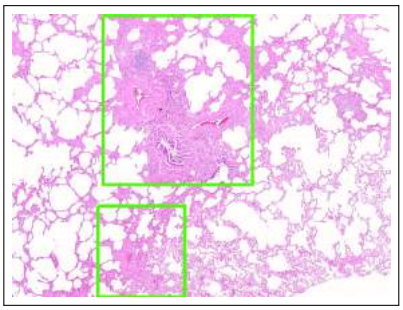


Fig. 1. ACIF patterns in a lung biopsy, indicated by green frames. Using low microscope magnification ($< 5x$) ACIF patterns can be seen as high-density pixel regions. In this sample, one can see peribronchial remodeling (large frame), and fibroblastic extension (small frame). Both patterns present a high concentration of pixels. We consider only colored pixels to define high-density regions.

A specialist traditionally searches for important patterns in large microscopic images, which is an extremely time-consuming and tiring task. Our main goal is to highlight regions of interest with relevant information to direct the specialist's attention to them.

Relevant information is defined by high-density pixel regions, which are automatically searched, thereby leading to a significant reduction in time invested by the specialist, and enabling a more accurate diagnosis. Our method is divided into four main steps:

1) Automatic high-to-low microscope magnification reduction, producing a lower pixel resolution image.
2) Computation of probability of a pixel belonging to a high-density region, defined as ROI.
3) Extraction of features of high- and low-density regions.
4) Visualization of ROIs using a color.

It is important to note that the original image is not affected. Our method creates a new image with reduced resolution. Thus, it is possible to extract the ROIs, found in low magnification, from the original image based on the original resolution/magnification.

A. Magnification reduction

In this step, we reduce the microscope magnification to $< 5x$, which produces a new, smaller image (in terms of pixel resolution). The goal of this step is to reduce the huge number of pixels in the input image, and subsequent steps can be performed in an acceptable computational time.

Let Ω be an image consisting of n pixels, where the value of the i^{th} pixel is defined as Ω_i. The spatial location of pixel i is $\Omega_i(\mathbf{x}, \mathbf{y})$, the physical isotropic pixel sizes, in micrometers, are defined as $\{PhysicalX, PhysicalY\}\mu m$, in the $X-$ and $Y-$axis, respectively. The sizes, in pixel units, of Ω are defined as $\{SizeX, SizeY\}$.

We guide the adequacy of the original image magnifications using the physical isotropic pixel size. The magnification reduction is performed based on the following equations:

$$\begin{aligned} \Delta M &= ((M_{in} \times ep) - (M_{out} \times ep)) \\ PhysicalX_{out} &= (PhysicalX_{in} \times \Delta M) + PhysicalX_{in} \\ PhysicalY_{out} &= (PhysicalY_{in} \times \Delta M) + PhysicalY_{in} \\ SizeX_{out} &= \frac{SizeX_{in}}{PhysicalX_{out}} \\ SizeY_{out} &= \frac{SizeY_{in}}{PhysicalY_{out}} \end{aligned} , \tag{1}$$

where ΔM is the microscope magnitude reduction, M_{in} is the original magnitude, M_{out} is the target magnitude, $ep = 10x$ is the microscope eyepiece, $\{PhysicalX_{in}, PhysicalY_{in}\}\mu m$ are the original physical pixel sizes in $X-$ and $Y-$ directions, $\{PhysicalX_{out}, PhysicalY_{out}\}\mu m$ are the resulting physical pixel sizes after reduction, $\{SizeX_{in}, SizeY_{in}\}$ are the original sizes in pixel units, and $\{SizeX_{out}, SizeY_{out}\}$ are the resulting sizes.

Due to the high resolution of the input image, we divide it into tiles for reading and processing. We reduce the size of every tile using an interpolation method, called "INTER_AREA" provided in OpenCV [5]. The INTER_AREA method receives the $\{SizeX_{out}, SizeY_{out}\}$ (Equation 1) values as input parameters and uses the pixel area relation to re-sample the image. Finally, we construct a reduced image Ω' as a mosaic

of reduced tiles. The pixel values of the reduced image Ω' are given in gray-scale. Pixel values of the input image Ω exhibit color (blue, violet and red) of Hematoxylin and Eosin stain (H&E), used to color the biopsy. In order to categorize texture and morphological patterns, we use the gray-scale value instead of the H&E value (usually represented as an RGB triple).

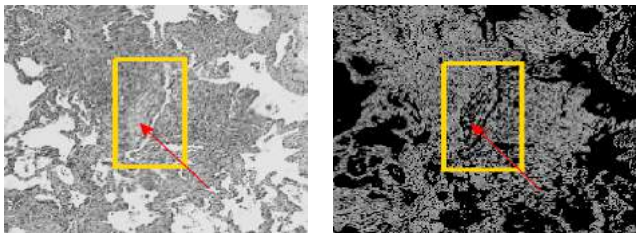
B. High-density region identification

In order to identify high-density regions, we first segment the white background, by applying a threshold-based approach to Ω' using the Otsu algorithm [6]. This step produces a binary representation Ω^b of Ω'. We define a circular mask $\omega \in \Omega^b$ of radius r and, for each pixel $i \in \Omega^b$, we compute the probability P_i of it belonging to a high-density region, as follows:

$$P_{i \in \Omega^b} = \frac{(\sum_{j \in \omega} j) \times 100}{\pi r^2}, \tag{2}$$

where $j \in \omega$ is a neighbor pixel of i. We use a threshold t to divide the high (H) and low (L) density regions. We consider $P_i > t$ as the condition to separate them.

The pixel probabilities P_i can be visualized via a colormap, where all high-density regions are represented with a red color. We use the "Jet" colormap, which employ a color range from blue to red. Examples of the generated colormaps are shown in Figure 4.



(a) Input Image (b) Otsu

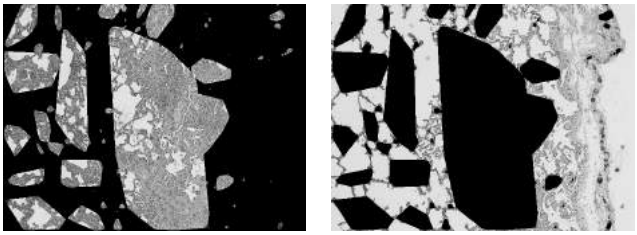
Fig. 2. Regions can be wrongly labeled as background (black color) when using Otsu, see yellow frames in (a) and (b). Proliferative bronchiolitis is pointed out with red arrows. Otsu misses the remodeled tissue.

When using Otsu to remove the background in Ω', some regions may wrongly be labeled as background, as shown in Figure 2. This problem can also arise when using other threshold-based methods (simple threshold methods, triangle or statistical moment-based methods).

We solve this problem by computing the convex hulls of all components in H, filling them with the respective pixel values in Ω', which includes background pixels. Finally, the image L can be viewed as the input Ω' with all convex hulls removed. The resulting images are shown in Figure 3.

III. Feature extraction

Several features can be extracted from histopathological images, e.g., gray-scale histograms, statistical moments, co-occurrence matrices, or Local Binary Patterns (LBP). We consider LPB features as suggested by Kumar et al. [7], where a comparison of LBPs, Convolutional Neural Networks (CNNs),



(a) High density (b) Low density

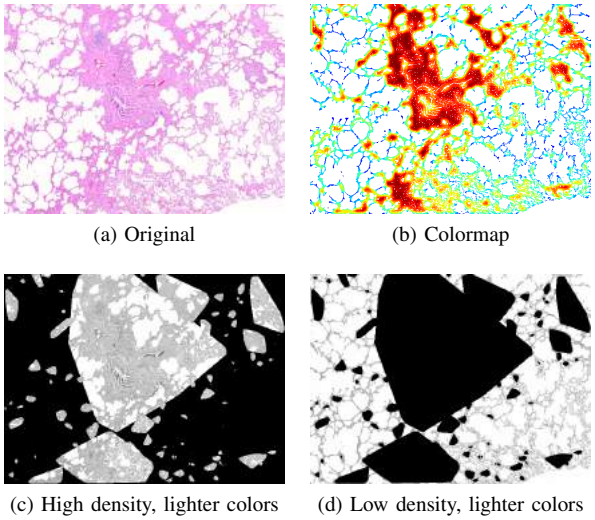
Fig. 3. Image Ω' divided into high- and low-density regions. See lighter colors. B

and a Bag of Visual Words (BoVWs) showed that LBP features led to high-accuracy results for classifying histopathological images. The main goal of this step is to compare the LBP features between the high- and low-density regions identified before, in order to quantify the accuracy of our high-density identification method.

IV. Results

We have evaluated our method for nine histopathological images in Virtual Slide Image (VSI) format of the lung, of resolutions $(110,019 \times 100,196)$ and $(128,018 \times 56,878)$, physical isotropic pixel sizes of $(0.172\mu m \times 0.172\mu m)$, and microscope magnifications of $40x$ and $20x$. We reduced the original image magnifications to $5x$, leading to a resolution of (1611×1467) and physical isotropic pixel size of $(68.309\mu m \times 68.309\mu m)$. All images were obtained with the Olympus BX 61 VS microscope with an eyepiece of magnification $10x$, using biopsies from patients with confirmed ACIF.

We have performed our experiments on a Linux workstation, with Intel Core i7-2600 CPU 3.40GHz x 4 and 16GB memory. The main average processing time was 15.13 minutes. This short processing time is due to the magnification reduction step and the fact that we implemented our method using multiple threads.



(a) Original (b) Colormap

(c) High density, lighter colors (d) Low density, lighter colors

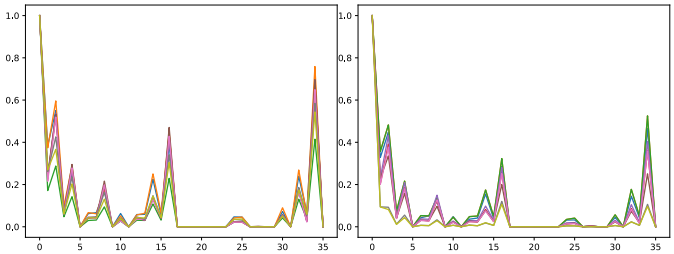
Fig. 4. High-density regions are shown in dark red (a), extracted as ROIs by computing their respective convex hulls, see (c) and (d).

A. Colormap

After identification of high-density regions in all samples, results are shown in high- and low-density regions, using a color. Pathologists, trained in ACIF identification, analyzed these results and were able to identify ACIF patterns in the high-density regions, demonstrating that our segmentation correctly identifies and captures all ACIF patterns in high-density regions. According to the specialists, color helps to rapidly focus on important areas and increase microscope magnification for them, to search for more specific patterns. Figure 4 shows a colormap together with the patterns identified by the pathologist.

B. Local Binary Pattern

In order to verify and quantify the precision of our method, we computed the LBP pattern of both high- and low- density regions. Figure 5 shows the resulting curves of the LBP histograms. One observes that all LBP curves for high-density and low-density are very close to the other of the same category. This result is independent of the size and shape of the ROIs.



(a) High density (b) Low density

Fig. 5. LBP histograms of high- and low-density regions of nine biopsies. All nine curves for high- and low-density histograms closely tie.

Kumar's study showed good results when using LBP in combination with a Support Vector Machine (SVM). However, in our case, the number of available images does not allow one to train an SVM. Therefore, we used the LBP histograms as input of a k-Nearest Neighbor (Knn) algorithm. We obtained a classification accuracy of 94.44%, which is the highest one reported to date in the literature, for this type of data. Table I provides all parameters used in our experiments.

TABLE I
PARAMETERS USED IN OUR EXPERIMENTS.

Description	Value	Units
Density radius	7	pixels
Density threshold	60	grays-cale
Magnification	5x	microscope power
LBP radius	1	pixels
Number of LBP points	8	integer
KNN	3	number k of neighbors
Cross validation	9	number of folds

V. Conclusions and Future Work

We have presented a method to automatically identify high-density regions in extremely high-resolution lung biopsies. During the process of airway-centered interstitial fibrosis deceased bronchi may suffer severe remodeling. This pattern can be seen, in a low microscope magnification, as a collection of regions of high concentration of pixels. Our method is well-suited for the identification of those regions in biopsies of patients with confirmed ACIF.

We compute local binary pattern histograms of high- and low-density regions in order to quantify the accuracy of the identification process. We can obtain accuracy results up to 94.44%, showing our method's effectiveness to highlight the most remodeled regions, enabling the specialist to analyze them in a subsequent step, looking for more patterns to diagnose ACIF at higher microscope magnification.

Since our method is fully automatic, user interaction with our system is a straightforward task. Our implementation shows to the pathologist the high-density regions via colormaps, and it extracts them (shown with convex hulls) from the original input image Ω as separate images in Tiff format. Having the images separated, as is usually done by specialists, makes it possible to analyze images at higher magnification, as the process demands.

This work is, to the best of our knowledge, the first effort to visually analyze images of ACIF, automatically highlighting regions of interest for the pathologist in support of a disease report. This approach greatly improves analysis accuracy, and it significantly reduces analysis time. A procedure that typically takes hours to perform with a traditional approach can be done in minutes with our method.

Concerning future research, we plan to investigate various strategies for automatic identification of ACIF patterns in histopathological images at high microscope magnification.

References

[1] K. Ley and A. Zarbock, "From lung injury to fibrosis," *Nature Medicine*, vol. 14, no. 1, p. 20, 2008.

[2] E. Silbernagel, A. Morresi-Hauf, S. Reu, B. King, W. Gesierich, M. Lindner, J. Behr, and F. Reichenberger, "Airway-centered interstitial fibrosis–an under-recognized subtype of diffuse parenchymal lung diseases," *Sarcoidosis Vasculitis and Diffuse Lung Disease*, vol. 35, no. 3, pp. 218–229, 2018.

[3] A. Churg, J. Myers, T. Suarez, M. Gaxiola, A. Estrada, M. Mejia, and M. Selman, "Airway-centered interstitial fibrosis: a distinct form of aggressive diffuse lung disease," *The American Journal of Surgical Pathology*, vol. 28, no. 1, pp. 62–68, 2004.

[4] D. A. Lynch, N. Sverzellati, W. D. Travis, K. K. Brown, T. V. Colby, J. R. Galvin, J. G. Goldin, D. M. Hansell, Y. Inoue, T. Johkoh *et al.*, "Diagnostic criteria for idiopathic pulmonary fibrosis: a fleischner society white paper," *The Lancet Respiratory Medicine*, 2017.

[5] G. Bradski and A. Kaehler, "Opencv," *Dr. Dobbs Journal of Software Tools*, vol. 3, 2000.

[6] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[7] M. D. Kumar, M. Babaie, S. Zhu, S. Kalra, and H. R. Tizhoosh, "A comparative study of cnn, bovw and lbp for classification of histopathological images," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2017, pp. 1–7.