AUTOMATED FEATURE ANALYSIS IN BIOLOGICAL IMAGES

By

Guneet Singh Mehta

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science (Electrical and Electronics Engineering) at the UNIVERSITY of WISCONSIN-MADISON 2017
Abstract
This thesis is comprised of three projects that I worked on during the period of my Master of Science degree. All three projects use computer vision and image processing techniques to improve microscopy image analysis workflows and develop object detection applications. This work also discusses an image analysis tool developed for imaging and analysis of collagen.

The first project is aimed at improving the current state of image acquisition by autofocusing the slide and removing artifacts from image by flat field correction. This project will serve as a stepping stone for smart microscopes where runtime analysis can be done during acquisition.

The second project was developed in collaboration with the Exploratorium Museum (San Francisco) to detect and highlight zebra fish embryos and zebrafish in a stream of video captured by a microscope as the objective is moved or zoomed by users. The aim of this project was to improve museum visitor participation by highlighting all the zebrafish in current field of view.

The third project is aimed at developing data analysis and data visualization tools which use the fiber data extracted from Second Harmonic Generation (SHG) images by CT-FIRE (Curvelet Transform - Fiber Extraction Algorithm) software. Two broad functionalities developed were: Post Processing Graphical User Interface (GUI) for fiber analysis; and Region of Interest (ROI) manager.
Introduction

This work is based on multiple projects with a common theme of finding meaningful features in biological images and converting them into numbers for inference and hypothesis testing. The ground for this work was laid in 2014 during my internship at Laboratory for Optical and Computational Instrumentation (LOCI) at University of Wisconsin-Madison under LOCI Director Dr. Kevin Eliceiri and Co-Advisor Prof Dan Negrut. My summer experience gave me a broad overview of imaging techniques, computer vision, image processing and software developed at LOCI.

My summer work also exposed me to ImageJ software, acquisition tools and MATLAB tools for data analysis. I also gained hands-on experience in developing and maintaining the MATLAB based suite of software for fiber analysis in Second Harmonic Generating (SHG) images.

Subsequently, I decided to join LOCI as a Research Assistant while pursuing my Master of Science degree at Electrical and Computer Engineering Department at UW Madison. While trying to solve the research problems, I got to learn about ImageJ software, ImageJ Macros, building standalone MATLAB projects and using libraries for software development.


This work includes details of the projects I have tackled during my degree, which are organized in three chapters, concluding remarks and appendices for further reference.
Chapter 1
Smart Microscope

One of the common workflows of biologists while imaging is to image a lot of samples in a similar manner. Current systems require human intervention to change slides, reset the objective, post process acquired images etc. Automation of these routine tasks will remove humans from the decision making loop and improve the speed of acquisition. One of the software systems at Laboratory for Optical and Computational Instrumentation (LOCI) University of Wisconsin-Madison, is called WiscScan which is an open source project for microscope control and imaging.

As initial steps to reduce human intervention and post acquisition image processing, the following two functionalities were implemented in C++:-

1. Autofocus during image acquisition
2. Flat field correction during image acquisition

Part A: Autofocus during image acquisition

Current imaging systems while imaging large tissues, acquire multiple images at the same Z-stack level. These individual images (also referred as tiles) are then stitched together to create a bigger image.

Many a times the tissue being imaged has different thickness at different points (due to improper slicing), requiring different focus at different portions of the tissue. In current systems, at the start of imaging, the scope is focussed such that the first tile is under focus and during acquisition the stage is moved only in x and y directions (in plane of slide) and not in z direction (perpendicular to slide plane). Thus the focus setting of the first tile is used for all other tiles, which can often lead to acquisition of a blurred image.

This work proposes to use a common numerical measure of focus called Focus Measure to correct the Z-position of the slide as it is moved in raster scan order while acquisition. The Focus Measure is a measure of high frequency content / Entropy in the image. An out of focus image (Figure 1) is blurry and does not have sharp edges, while a focussed image will have sharp edges(Figure 2).

Focus Measure for an image is calculated as:-

\[
\text{Focus Measure} = \text{avg} \left( \sum_i \sum_j \nabla^2_M f(i,j) \right)
\]
Figure 1: Images with different focus settings: (A) Out of focus image and (B) In focus image

where

\[ \nabla^2_M f(i,j) = \left| \frac{\partial^2 f}{\partial x^2} \right| + \left| \frac{\partial^2 f}{\partial y^2} \right| \]

And \((i,j)\) is the position of each pixel. The second derivative in x and y directions is calculated as:

\[ \left| \frac{\partial^2 f}{\partial x^2} \right|_{(i,j)} = f(i+1,j) + f(i-1,j) - 2 \times f(i,j) \]

\[ \left| \frac{\partial^2 f}{\partial y^2} \right|_{(i,j)} = f(i,j+1) + f(i,j-1) - 2 \times f(i,j) \]

Focus measure for images at different Z positions are calculated and the z position with maximum Focus Measure is the best focussed image. For example the following images are acquired at increasing z distance of the stage from the objective (in raster scan order)
Figure 2:- Images with increasing z distance: A, B, C, D, E, F, G
Focus Measure of images

<table>
<thead>
<tr>
<th>Image</th>
<th>Focus Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>29.3561</td>
</tr>
<tr>
<td>B</td>
<td>64.6141</td>
</tr>
<tr>
<td>C</td>
<td>82.5003</td>
</tr>
<tr>
<td>D</td>
<td>67.2039</td>
</tr>
<tr>
<td>E</td>
<td>39.7218</td>
</tr>
<tr>
<td>F</td>
<td>19.406</td>
</tr>
<tr>
<td>G</td>
<td>10.3512</td>
</tr>
</tbody>
</table>

*Table 1: Focus measure of images in Figure 2*

It can be seen from the above images that image “C” has maximum focus measure. It can also be visually verified that image “C” is the most focussed image with sharpest edges and features.

The code for the above functionality is attached in Appendix A

**Part B: Flat Field Correction during image acquisition**

Flat Field correction is a standard calibration technique which is used to remove the effects of system defects and errors from a digital image. When a perfectly uniform background is imaged using a microscope, different regions in the acquired image may not have the same pixel values due to optical path distortion or different sensitivities of individual pixels in a camera pixel array.

Once an imaging system has been flat field corrected a uniform background will create an image with the same value for all pixels in the image. Thus any variation in the image acquired after correction will be a result of the sample being imaged and not any system error.

For example, consider the following two images - a) Image captured without flat field correction and b) image captured with a uniform background
Figure 3: Images captured without flat field correction: (A) image of a section of an H&E stained slide and (B) image of a uniform background

It can be seen that due to dust particles and non uniformity in optical paths the background image (Figure 3B) is not uniform. Image captured without compensating for this effect contains artifacts. It can be seen in the image below that such an image has dust particles and non gray border at the edges:

Figure 4: Image containing circular brown artifacts (marked in circles) and non grey background (marked in rectangle)
Artifacts like these lead to suboptimal performance of other algorithms like automated object detection. For example, the image below was stitched together without doing the flat field correction of individual images, leading to artifacts of horizontal and vertical stripes in the image below.

![Figure 5: Image obtained after stitching images not compensated by flat field correctness.](image)

This effect can be removed by flat field correction which requires the following:-

1. A Background image captured by imaging a perfectly flat and featureless area in a slide prior to any imaging.
2. The background image is converted to a grayscale image and average value of all pixels is calculated, referred as “mean_background_value”
3. Each pixel in each channel is traversed and the value is corrected by the formula:-

\[
\text{new_image}[i][j][c] = \frac{\text{raw_image}[i][j][c]}{\text{background_image}[i][j][c]} \times \text{mean_background_value}
\]

where:

a. “i,j,c” are the pixel coordinates along x y and z axis.
b. “raw_image” is the image of sample without any correction
c. “background_image” is the image of a uniform featureless region in a slide
d. “new_image” is the flat field corrected image.
e. Division is a long form division
Results
Below is the original raw image with artifacts marked in red and the flat field corrected image with the C++ code attached in Appendix B.

Figure 6: Images before and after flat field correction: (A) Raw image and (B) Flat field corrected image
Chapter 2
Detecting Zebrafish in video frames

This work led to a collaboration with the Exploratorium Museum at San Francisco to develop a computer vision system to detect zebrafish embryos and adults in a live video feed. The video was captured by a microscope mounted over a slide containing zebrafish embryo in various stages of development as shown in Figure 8.

Two detectors were created for this project: - Embryo Fish Detector and Adult Fish Detector

**Embryo Detector**
The distinguishing feature of all embryo fish is its spherical shell as seen in Figure 8 below. Thus detecting circles in the image will detect position of embryos in the image.

![Figure 8: Embryo fish in different stages of development: A Semi developed embryo and (B) Embryo in nascent stage](image)

(A)  
(B)
However some images also contain spherical artifacts like air bubbles at the boundary of the image and empty Embryo shells as seen in the figure (Figure 9) below.

![Figure 9: Spherical artifacts in acquire images: (A) Air bubble at edge of the slide and (B) Empty zebra fish shell](image)

**Method:**

As mentioned earlier, all the embryos in the acquired images were highly circular with a specific range of radius, thus reducing the problem of finding embryos to finding circles of certain radius range (35-45 pixels in steps of 2 pixels at a Magnification of 1X in image feed of size 200 x 129 pixels) in the image.

A common technique of finding regular shapes in images is using a Hough Transform. An image containing a circle is converted to an image with same size with intensity values corresponding to the probability of presence of a center of circle at that position. For example consider the following images:

![Figure 10: Image and Hough Transform: (A) Image containing two circles (B) Hough Transform of (A) with intensity proportional to the possibility of presence of a circle](image)

In practice each image (Figure 11A) is converted into an edge image containing only object outline (Figure 11B). Thereafter the edge image was converted into Hough Transform where intensity of each point is indicative of probability of presence of a circle centered at that position(Figure 11C).

Using the Hough Transformed image, at most 10 most prominent circles are selected based on intensity of Hough Transform. The resulting circles are overlaid on the original image (Figure 11D).
Additional comments:
It was observed that the instances of presence of empty embryo shell and embryo shell sized air bubbles were very less which after consulting with our client did not warrant development of code to get rid of these cases.

Adult Fish detector
An adult zebrafish has fully developed organs including well formed eyes which are dark black in color. This feature was used to find the location of hatched fish in images. It was observed that zebrafish eyes were at least of size 2000 pixels and minimum circularity of 0.3 at a Magnification of 1X in image feed of size 834 x 540 pixels.

To detect an adult fish, an image (Figure 12A) is converted to a grayscale image and all pixels below an intensity of 50, were converted to white region and the rest were converted to black to obtain a binary image (Figure 12B). Next, all white objects which did not have a minimum circularity of 0.3 and a minimum area of 2000 pixels were discarded (Figure 12C).

The image so obtained contains the eyes of adult fish in the image. These regions were then increased by dilation to highlight the upper half of an adult zebrafish(Figure 12D and 12E).
Figure 12: Workflow of adult fish detector: (A) Original image, (B) Thresholded image, (C) Adult Fish eye regions, (D) Dilated eye regions, (E) Overlaid results
Conclusion
The detectors developed using ImageJ for the project were used as part of a larger integrated system. The detectors provided the X,Y coordinates of the fish and embryo in the slide which were then used to highlight them in the display. The highlighted objects are expected to increase the user’s involvement and learning from the system.

Using the two detectors for embryo detector and adult fish detector, we were able to get good and acceptable precision and recall as can be seen in the following video:-
https://www.youtube.com/watch?v=xL1TINUJ_Gs

Source Code Repository
The source code for the repository can be obtained at:-
https://github.com/uw-loci/exploratorium.git

Detailed Analysis
A detailed analysis of the performance of developed detectors on individual images can be found at:-
https://github.com/uw-loci/exploratorium/blob/master/test_results.pdf
Chapter 3
Curvelet Transform and Fiber Extraction (CT FIRE) Modules

The extracellular collagen matrix (ECM) has been found to be an indicator of progression of many types of cancer. The mechanism of interaction of collagen matrix and cancer cells has been a subject of intense medical research. Development of Second Harmonic Generation (SHG) Imaging technique has enabled acquisition of high resolution images of individual fibers and subsequent analysis. For example, Conklin et al. [5] showed that patterns of arrangement of collagen fibers can predict outcome of a breast cancer patient. Nadiarnykh et al [6] discovered that patterns of collagen fibers in SHG image provided quantitative discrimination between benign and tumor tissues. Consequently, there was a need to develop an automated tool to extract individual fiber properties like fiber angle, length, straightness, width etc, from SHG images.

Consequently, researchers at Laboratory of Optical and computational Instrumentation (LOCI) at UW Madison developed CT-FIRE (Curvelet Transform and Fiber Extraction Algorithm). CT-FIRE is a software tool developed in MATLAB, that allows user to automatically extract collagen fibers in an image and quantify fibers with descriptive statistics, such as fiber angle, length, straightness, and width. The program reads in image files supported by MATLAB and extracts the individual collagen fibers via a combined method we call CT-FIRE (named ctFIRE in the software implementation). The approach of CT-FIRE is described in a publication [1], which combines the advantage of the fast discrete curvelet transform [2] (CT) for denoising the image and enhancing the fiber edge features and the advantage of fiber extraction (FIRE) algorithm [3] for extracting individual fibers. The output may be displayed on the screen and written into .csv, .xlsx, .mat data file or .tif image files for further analysis, such as statistical test or image classification. This tool does not require prior experience of programming or image processing and can handle multiple files, enabling efficient quantification of collagen organization from biological datasets.

Prior to the start of this thesis, the CT-FIRE functionality of extracting fibers from an image, existed in form of a MATLAB software. However, there was a lack of analysis tools to convert the extracted information to meaningful data. The thesis involved developing 2 main tools:-

1. Advanced Output Module
2. Region of Interest (ROI) Manager

The above tools were developed in MATLAB programming environment since the basic functionality of fiber extraction (CT-FIRE) was already developed in MATLAB.
Figure 13 (A): main GUI of CT-FIRE to extract individual fiber properties including fiber length, angle, width, and straightness; (B): advanced output GUI of CT-FIRE to combine the results from multiple images with or without thresholding; and (C): region of interest (ROI) manager of CT-FIRE to do ROI annotation and analysis for a single image.

Advanced Output Module
This module is used in the analysis of fiber data generated by the CT-FIRE program. This module can be used to:-

1. Remove fibers from analysis
2. Visualise Fibers on image
3. Select fibers on a combination of conditions
4. Generate statistics on selected fibers
5. Visualise fibers color coded by properties

This module plots fibers with a randomly generated color along with fiber number (Figure 14). During the analysis process, it is useful to be able to visualize individual fibers plotted in the image. Individual fiber number can thus be specified for visualisation.
Users also have an option to remove fibers from their analysis. Users can further select fibers on a combination of conditions on 4 properties of length, width, straightness and angle. This selection can be done in either comparative terms (percentage) or in absolute terms of (pixels and degrees).

After thresholding, statistics on fiber properties are generated and individual fiber properties are saved in a .xls file as shown in figures (Figure 15 A and 15 B) below.
Region of Interest (ROI) Manager

Region of Interest in an image is an area with interesting features which a user is interested in analyzing. A Region of Interest allows a user to focus on a specific area and dedicate computation time and resources to the ROI instead of the entire image, which is especially useful when working with large stitched images.

The ROI Manager module contains the following components:-

- **Filename**: displayed on the top left corner of the GUI.
- **ROI table**: Panel on the left contains the names of ROIs associated with the image.
- **New ROI**: Dropdown menu to select the shape of ROI to be drawn. The shapes included are – rectangle, freehand, ellipse and polygon and specify. Specify option allows the user to place a rectangular ROI of specified dimensions at the specific position.
- **Save ROI(s)**: Saves the recently drawn ROI.
- **Combine ROIs**: combine multiple ROIs into one ROI. (multiple selection of ROIs can be done by pressing “Ctrl” key and selecting the new ROI. ROIs in series can be selecting the first ROI, pressing shift and selecting the last ROI.)
- **Rename ROI**: change ROI name.
- **Delete ROI**: delete the selected ROIs.
- **Measure ROI**: for the selected ROIs, calculate minimum and maximum pixel values in the ROI, Area of each ROI (units – pixels) and mean pixel value.
- **Load ROI from text**: allow the user to load a ROI coordinates saved in a .csv file.
- **Load ROI Mask**: allow the user to load a ROI saved as a mask image.
- **Save ROI Text**: save the selected ROI coordinates into a .csv file.
- **Save ROI Mask**: save the selected ROI into a .tif mask image file.
- **ctFIRE ROI Analyser**: Launch an advanced ROI operations window.
- **Apply ctFIRE on ROI**: apply ctFIRE on the selected ROI(s) to extract fibers within them.
- **Show All**: display all ROIs on the image.
- **Labels**: display the name of ROIs along with the boundaries. Combined ROI is not labeled.
- **Message Window**: provide ongoing message.
Some of the functions are displayed below:

Figure 16: ROIs plotted on image

Figure 17: Fibers in defined ROI with histogram of properties

A detailed description of the tools developed for the CT-FIRE software can be found in [4].
Conclusion and Future Direction

This work is based on multiple projects which helped me gain a great breadth of knowledge in Image Processing, Computer Vision, Data Analysis and Machine Learning. However, as with any project there is a room of improvement in each.

Current fiber analysis tools include Curvelet Transform - Fiber Extraction algorithm (CT-FIRE) and CurveAlign which have different workflows and different interfaces. A future direction for could be to integrate all fiber analysis tools into one intuitive and easy to understand Graphical User Interface. The community of researchers who use the fiber analysis tools can be leveraged to improve these tools, which will reduce the development cycle.

Another major push of LOCI lab is to build a smart microscope. Current imaging systems do not leverage the advances in hardware like Graphical Processing Unit (GPU) or Machine Learning during image acquisition. A considerable time is spent to acquire a stitched image of entire sample which is then analyzed by other tools. A large size image warrants a considerable time to acquire and process the image. Most of the time, a sample contains interesting phenomenon viz tumor cells in a small portion, due to which time can be saved if only those regions are acquired. Machine Learning can be used to train the imaging systems to capture regions with interesting features. Currently, researchers at LOCI are working on a project to use low resolution images which are quickly captured to find out regions with features of interest. These regions will then be imaged using a high resolution lens.

It is hoped that this work will help the reader to not only tackle similar imaging problems but also serve as a guide on how to develop framework for imaging solutions.
References:


Appendix A (Autofocus during image acquisition)

```c
#include <stdio.h>
#include <opencv2/opencv.hpp>

using namespace cv;
using namespace std;

float main(int argc, char** argv )
{
    //argument check
    if ( argc != 2 )
    {
        printf("usage: DisplayImage.out <Image_Path>\n");
        return -1;
    }
    //image reading
    Mat image;
    image = imread( argv[1], 1 );
    if ( !image.data )
    {
        printf("No image data \n");
        return -1;
    }
    //channels and rows and columns
    int nChannels=image.channels();
    int nRows=image.rows;
    int nCols=image.cols;

    long entropy=0,fx,fy;
    for( int y = 1; y < image.rows-1; y++ )
    {
        for( int x = 1; x < image.cols-1; x++ )
        {
            for( int c = 0; c < nChannels; c++ )
            {
                fx=(2*(long)image.at<Vec3b>(y,x)[c]-(long)image.at<Vec3b>(y,x-1)[c]-(long)image.at<Vec3b>(y,x+1)[c]);
                fy=(2*(long)image.at<Vec3b>(y,x)[c]-(long)image.at<Vec3b>(y-1,x)[c]-(long)image.at<Vec3b>(y+1,x)[c]);
                entropy=entropy+fx*fx+fy*fy;
            }
        }
    }
    float avg_entropy=(float)entropy/(nRows*nCols);
    cout<<argv[1]<<" entropy=\""<<avg_entropy<<endl;
    return entropy;
}```
Appendix B (Flat Field Correction)

#include <stdio.h>
#include <opencv2/opencv.hpp>

using namespace cv;
using namespace std;

int main(int argc, char** argv )
{
    /*
        This function takes in a backgroundImage and a rawImage. Finds the mean value of
        the backgroundImage. Finds the field corrected image as =
        rawImage/backgroundImage*mean_value.
        The corrected image is saved as corrected_<raw_image_name>
    */

    if(argc!=3){
        printf("Correct Usage is DisplayImage.out <BackgroundImage_Path> <Raw_Image_Path>");
        return -1;
    }

    Mat backgroundImage=imread(argv[1],1);
    Mat rawImage=imread(argv[2],1);
    if(!rawImage.data||!backgroundImage.data)
    {
        printf("No image data in either backgroundImage or rawImage");
        return -1;
    }

    //displaying images
    namedWindow("Background Image", WINDOW_AUTOSIZE );
    imshow("Background Image", backgroundImage); //waitKey(0);
    namedWindow("Raw Image", WINDOW_AUTOSIZE );
    imshow("Raw Image", rawImage);
    waitKey(0);

    //error check- both images have same dimensions
    if(backgroundImage.channels()!=rawImage.channels()||backgroundImage.rows!=rawImage.rows||backgroundImage.cols!=rawImage.cols){
        printf("raw and background images have different dimensions");
        return -1;
    }

    //storing image dimensions
    int nChannels=rawImage.channels();
    int nRows=rawImage.rows;
int nCols=rawImage.cols;

//find mean value of backgroundImage
long meanBackgroundValue=0;
for( int c = 0; c < nChannels; c++ ){
    for( int y = 0; y < nRows; y++ ){
        for( int x = 0; x < nCols; x++ ){

            meanBackgroundValue=meanBackgroundValue+backgroundImage.at<Vec3b>(y,x)[c];
        }
    }

    meanBackgroundValue/=nRows*nCols*nChannels;
    //printf("mean value=%li\n",meanBackgroundValue);
    for( int c = 0; c < nChannels; c++ ){
        for( int y = 0; y < nRows; y++ ){
            for( int x = 0; x < nCols; x++ ){

                rawImage.at<Vec3b>(y,x)[c]=(int)((float)rawImage.at<Vec3b>(y,x)[c]/(float)backgroundImage.at<Vec3b>(y,x)[c]*(float)meanBackgroundValue);
            }
        }
    }
}

namedWindow("Raw Image", WINDOW_AUTOSIZE );
imshow("Raw Image", rawImage);
waitKey(0);

//Need to save the image too
char correctedImagename[200];
strcpy(correctedImagename,"corrected_ ");
strcat(correctedImagename,argv[2]);
imwrite(correctedImagename,rawImage);
return 0;
Comparison of Picrosirius Red Staining With Second Harmonic Generation Imaging for the Quantification of Clinically Relevant Collagen Fiber Features in Histopathology Samples


Department of Biomedical Engineering (CRD, AK, WJK, KWE), Laboratory for Optical and Computational Instrumentation (LOCI) (CRD, KM, GM, AK, YL, WJK, KWE), Department of Surgical Pathology (AGL), Department of Urology (SL, WAR), Department of Surgery (SMW, WJK), University of Wisconsin–Madison, Madison, Wisconsin; George M. O’Brien Research Center of Excellence, Madison, Wisconsin (WAR); University of Wisconsin Comprehensive Carbone Cancer Center, Madison, Wisconsin (AGL, WAR, SMW, WJK, KWE); and Morgridge Institute for Research, Madison, Wisconsin (CRD, KWE)

Summary
Stromal collagen alignment has been shown to have clinical significance in a variety of cancers and in other diseases accompanied by fibrosis. While much of the biological and clinical importance of collagen changes has been demonstrated using second harmonic generation (SHG) imaging in experimental settings, implementation into routine clinical pathology practice is currently prohibitive. To translate the assessment of collagen organization into routine pathology workflow, a surrogate visualization method needs to be examined. The objective of the present study was to quantitatively compare collagen metrics generated from SHG microscopy and commonly available picrosirius red stain with standard polarization microscopy (PSR-POL). Each technique was quantitatively compared with established image segmentation and fiber tracking algorithms using human pancreatic cancer as a model, which is characterized by a pronounced stroma with reorganized collagen fibers. Importantly, PSR-POL produced similar quantitative trends for most collagen metrics in benign and cancerous tissues as measured by SHG. We found it notable that PSR-POL detects higher fiber counts, alignment, length, straightness, and width compared with SHG imaging but still correlates well with SHG results. PSR-POL may provide sufficient and additional information in a conventional clinical pathology laboratory for certain types of collagen quantification. (J Histochem Cytochem 64:519–529, 2016)

Keywords
fibrillar collagen, histopathology, picrosirius red, polarized microscopy, second harmonic generation

Introduction
Basic research and clinical evaluation of cancer tissue focuses primarily on identification of malignant epithelial cells; however, the importance of the stroma is now appreciated as a key mediator of carcinogenesis.\(^1,^2\) A major component of the tumor-associated stroma, fibrillar collagen, has been shown to impact cancer cell behavior through biochemical and biophysical cues. Collagen normally provides the underlying structural framework of homeostatic tissues, but alterations in specific properties such as alignment have been observed in cancer and other pathological processes.
For example, researchers have identified a unique “tumor-associated collagen signature” (TACS) in breast cancer defined as bundles of straightened, aligned collagen fibers that are oriented perpendicular to the tumor boundary.\(^3\) Mechanistically, it is hypothesized that aligned collagen fibers form pathways that facilitate cancer cell migration away from the tumor and toward vasculature during the metastatic process.\(^3,4\) While there is significant research work ongoing on the role of collagen in live, dynamic in vitro and in vivo animal models of a wide array of cancers,\(^5\)–\(^10\) many groups, including ours, are also focused on investigating the potential clinical utility of collagen properties in routinely fixed and processed human tissues. Toward clinical translation, it has already been shown that the detection of TACS in routine histopathological evaluation of breast cancer can serve as an optical biomarker and be predictive of disease recurrence and patient survival.\(^11\) The clinical importance of collagen organization has also been investigated in pancreatic,\(^12\) esophageal,\(^13\) and ovarian cancers.\(^14\) Accordingly, there is great interest by basic and clinical researchers alike in examining and quantifying properties of stromal collagen fibers such as overall alignment as well as individual metrics including length, width, and straightness.

Many alternative methods exist to visualize collagen in the context of routine histopathological sections. For example, established histology dyes (Masson’s trichrome, Movat’s pentachrome, picrosirius red [PSR]),\(^15\)–\(^18\) antibody detection,\(^19\) polarization-based widefield imaging,\(^20\) and second harmonic generation (SHG) imaging\(^21,22\) have all been employed. Other imaging methods often associated with collagen imaging either do not offer the appropriate resolution for the assessment of collagen organization, such as optical coherence tomography,\(^23,24\) or require non-routine sample preparation methods and offer limited field of view, such as transmission electron microscopy,\(^25\) scanning electron microscopy,\(^26\) and atomic force microscopy.\(^27\) Of the imaging methods that are compatible with histopathological samples, SHG is highly specific to the non-centrosymmetric structure of fibrillar collagen.\(^28\) SHG has become a particular focus for this reason and a variety of others, including high resolution, good signal-to-noise ratio, ability to work nondestructively on stained and unstained tissues, and power to extract additional information such as polarization properties and forward–backward scattering ratios. However, there is concern over the distinction whether SHG can be used for routine collagen visualization and quantification in more routine pathology tissues versus whether it is required. Undoubtedly, SHG has great advantages in terms of imaging depth and higher-order information, but for most of the reported histopathological studies done with standard 5-µm-thick pathology sections, these advantages are not of primary interest.\(^11\)–\(^13,29\) In addition, SHG is not readily accessible in the clinical setting due to high cost, lack of clinical interfaces, and relatively slow acquisition workflows.

To this end, we wanted to determine whether a more accessible technique could provide similar quantitative information about collagen alignment and fiber structure in histopathological samples as to what SHG can provide. Collagen fibers are composed of molecules that are hierarchically ordered in a parallel orientation resulting in inherent birefringent properties and also contain a large proportion of basic amino acid residues. Catering to both of these properties, sirius red is an elongated, acidic dye molecule that efficiently aligns and binds parallel to collagen fibers, thereby enhancing its innate birefringence. Due to its ability to specifically visualize collagen, as well as its ready availability in many research and pathology laboratories, PSR was chosen as a potential candidate. In this study, we systematically imaged collagen in tissue regions from 92 pancreatic cancer patients using PSR with a standard polarization widefield microscope followed by subsequent SHG imaging and then quantified collagen properties from both sets of data using well-established software tools for collagen quantification.\(^29,30\) Pancreatic cancer was decided on as a rational cancer model because it remains one of the most deadly human malignancies and has been the subject of recent studies correlating collagen organization to disease progression.\(^31\) In addition, in the context of the tumor microenvironment, an extreme collagen-rich stromal response is a hallmark of pancreatic cancer and can account for up to 90% of the tumor mass.\(^32\) Furthermore, characteristic changes in collagen alignment have been documented,\(^12\) and other stromal cues have been correlated to patient prognosis.\(^33\) Consistent with previous SHG trends,\(^12\) we found that key collagen properties, such as alignment, were computed to be higher in pancreatic cancer tissues compared with normal adjacent tissues using picrosirius red stain with standard polarization microscopy (PSRPOL). Interestingly, all of the collagen measurements that we considered were computed to be significantly higher in the PSR-POL images relative to SHG images. While the exact reasons underlying the differences remain to be determined, our data suggest that PSRPOL can extract similar, if not additional and complementary, information about collagen compared with SHG. Due to a relatively low clinical adaptability barrier, PSR could serve as a viable alternative technique for interrogating and quantifying certain collagen properties in histopathological tissues in conjunction with established fiber tracking algorithms.
Materials and Methods

Patient Tissue Staining With Picrosirius Red

To quantitatively compare collagen in multiple patients, a tissue microarray (TMA) available at the University of Wisconsin Comprehensive Carbone Cancer Center was utilized. This TMA was constructed (IRB No. 2011-0840-CR003) from pancreatic cancer surgical tissues acquired at the University of Wisconsin Hospital and Clinics between 1987 and 2012. All patients underwent curative-intent resection and were not treated in the neoadjuvant setting. Hematoxylin and eosin (H&E) sections were prepared from formalin-fixed paraffin embedded (FFPE) blocks originally used for diagnosis and reviewed by board-certified pathologists to identify different representative tumor regions. Blocks were then punched using a 1-mm-diameter core and inserted into a recipient FFPE block using a Beecher Instruments (Sun Prairie, WI) MTA-1 manual tissue arrayer. 5-μm-thick sections were cut from the TMA blocks on a microtome and adhered to standard glass slides. A PSR solution was formulated by dissolving 0.1-g Fast Green FCF (F7258; Sigma-Aldrich, St. Louis, MO) and 0.1-g Sirius Red F3B (S03695; Pfaltz & Bauer, Waterbury, CT) in 100 ml of saturated picric acid. After deparaffinization and rehydration, the TMA sections were covered with one to two drops of the PSR solution for 60 min at 40°C. The slides were then rinsed with acidified water, dehydrated through three changes of 100% ethanol and three changes of xylene, and coverslipped with toluene-based Richard-Allan mounting medium (Thermo Fisher Scientific, Waltham, MA). Representative human breast and prostate cancer tissues were also stained using the same protocol.

Picrosirius Red Microscopy

The PSR-stained tissues were digitalized using an Aperio slide scanner (Leica Biosystems, Vista, CA) to enable easy pathology review for region of interest (ROI) selection. For the pancreatic cancer TMA, two to three ROIs containing cancer cells and associated stroma were annotated on 92 patient tumors. All PSR imaging was performed on a traditional transmission pathology microscope (BX53; Olympus Corp.) using a 20x objective (UPlanFL N NA = 0.50; Olympus Corp.) without polarizers (PSR) and with linear polarizers (PSR-POL). Halogen lamp intensity was kept constant for both image types, and an exposure time that optimized the signal-to-noise ratio was chosen and kept constant within each image type. All images were digitally captured using cellSens acquisition platform (Olympus Corp.). To control for intrinsic tissue birefringence, we also imaged unstained pancreatic tissue without and with polarized light (Supplementary Fig. S1).

Second Harmonic Generation Microscopy

The same ROIs were imaged using a custom-built SHG microscope developed by the Laboratory for Optical and Computational Instrumentation known as CAMM (Compact Automated Multiphoton Microscope). The CAMM was designed to screen histology samples for cancer-associated changes in collagen fiber properties. Tissue landmarks were used to align the field of view and assure the exact same ROIs were imaged as on the PSR-stained sections. SHG signal was generated using 780-nm excitation light from a Ti:sapphire femtosecond laser and detected in the forward direction after spectrally filtering (390/18 BP; Semrock, Inc., Rochester, NY) using a 20x objective (S Fluor NA = 0.75; Nikon Corp.). All images were acquired at 1024 x 1024 resolution using WiscScan software (http://loci.wisc.edu/software/wiscscan) with consistent parameters. Fifteen of the pancreatic cancer ROIs were also imaged using a backward-detection SHG microscope. For this, the SHG signal was generated using 890-nm excitation light and collected with the same objective after filtering (445/20 BP; Semrock, Inc.).

Fiber Segmentation and Measurement

PSR-POL images were converted to 8-bit, cropped, and downsampled in the ImageJ-based analysis package Fiji to align with the field of view and pixel resolution of the SHG images. CT-FIRE fiber analysis software (http://loci.wisc.edu/software/ctfire) was then applied to the SHG and PSR-POL images using default settings after thresholding background (20–255 for SHG, 5–255 for PSR-POL). Quantitative information computed by CT-FIRE for each image type included the individual fiber metrics of length, width, and straightness. Collective fiber properties of alignment and total count were also computed for the entire image.

Statistics

For every CT-FIRE metric, a patient value was calculated by averaging the mean values for the two to three ROIs. All data were verified to follow Gaussian sampling using the Shapiro–Wilk normality test. Values obtained using SHG and PSR-POL were compared using a repeated-measures paired t-test and Pearson correlation analysis in GraphPad Prism (GraphPad Software, Inc., La Jolla, CA). A value of p<0.05 (two-tailed) was considered statistically significant.
Figure 1. Panel of commonly used histological dyes and microscopy methods to visualize collagen complementary to standard hematoxylin and eosin staining in thin sections. Also shown is second harmonic generation, which remains the experimental “gold standard.” All staining was done on serial sections of pancreatic cancer tissue. Scale bars = 100 µm.
Figure 2. Representative human pancreatic tissue visualized with PSR-POL and SHG. Under brightfield transmission microscopy, PSR stains collagens with a deep red color. When paired with polarization microscopy, PSR enables visualization of highly birefringent thick (orange-red) and thin (green-yellow) collagen fibers. Collagen organization has previously been shown to be characteristically aligned in the pancreatic cancer stroma using SHG. Abbreviations: PSR-POL, picrosirius red stain with standard polarization microscopy; SHG, second harmonic generation; PSR, picrosirius red. Scale bars = 100 µm.
Results and Discussion

Histological examination of biopsied or resected tissue by surgical pathology remains the current “gold standard” for cancer diagnosis, staging, and patient prognostication. In practice, several large tissue sections are routinely examined, but only the epithelial elements are taken into consideration in determining grade and stage of the tumor, and consequently, only these inform decisions regarding clinical management. Collagen has long been qualitatively appreciated as a mainstay tissue component of tumors, but its features have yet to be incorporated into pathologists’ overall assessment. With the burgeoning evidence for collagen structure and organization as important players in progression of multiple cancer types, there is growing interest by pathologists to integrate quantitative information about collagen and other features of the tumor microenvironment with traditional qualitative observations to provide a more nuanced characterization of the tissue under examination.

To date, basic research of collagen has been greatly accelerated by SHG imaging, a laser scanning microscopy technique that can provide high-resolution, quantifiable images of discrete collagen fibers without the need for exogenous staining. However, the technology currently is poorly suited for pathology workflows due to an array of issues including expense, instrumentation complexity and footprint, technically demanding operation, and clinical acceptance. Furthermore, collagen assessment needs further clinical validation. To adopt assessment of collagen structure and organization to pathological practice, a surrogate technique that provides certain types of collagen quantification information similar to that of SHG imaging and can be used routinely in the pathology laboratory needs to be identified. Although H&E is the most common histological stain used in clinical pathological practice, several complementary dyes are routinely used in surgical pathology labs that are capable of visualizing stromal collagen (Fig. 1). Masson’s trichrome and Movat’s pentachrome are used to visualize connective tissue on the basis of color, but their specificity to collagen is limited. PSR also differentially stains connective tissue with a rich red color, and it can be combined with conventional widefield polarization microscopy to greatly enhance the birefringence of collagen fibers.

Figure 3. Quantification of different collagen fiber metrics (alignment, length, straightness, and width) in normal and pancreatic cancer-associated stroma using picrosirius red stain with standard polarization microscopy imaging. Images were acquired from matched normal and cancer tissues from 75 patients.
SHG and Picrosirius Red Morphometric Comparison

525

(Supplementary Fig. S1) and specifically visualize them with a green–yellow–orange–red spectrum.\(^3\)\(^5\) PSR has already been exploited to understand collagen changes in cancer and other pathologies associated with fibrosis.\(^1\)\(^6\),\(^1\)\(^8\),\(^3\)\(^7\)–\(^3\)\(^9\) Consistent with our previous observations,\(^1\)\(^2\) stromal collagen appears more elongated and aligned in the immediate stroma surrounding malignant ducts compared with normal pancreas ducts when examined using PSR-POL (Fig. 2). Highly remodeled collagen was also observed in breast (Supplementary Fig. S2)\(^1\)\(^1\) and prostate (Supplementary Fig. S3)\(^3\)\(^8\) cancers using PSR-POL.

To validate the utility of PSR-POL to detect collagen changes, we acquired images around histologically confirmed normal and malignant pancreatic ducts. Using SHG, we previously established that collagen is characteristically reorganized and shows increased alignment, length, straightness, and width around malignant pancreatic ducts compared with benign ducts.\(^1\)\(^2\) Here, we determined that PSR-POL imaging yields similar data: collagen fibers are more aligned, elongated, and thick in the stroma adjacent to the malignant epithelium (Fig. 3, \(p<0.0001\)). Fiber straightness was not significantly different between normal and cancer stroma when examined with PSR-POL \((p=0.4391)\).

We next quantitatively compared images acquired in the same ROIs of 92 pancreatic cancer patient tumors using PSR-POL and SHG. For each technique, image acquisition parameters were set, so that the best signal-to-noise ratio was obtained and images were therefore representative of optimal visualization capabilities. We decided to employ forward-detection SHG in this comparison study due to our experience in achieving brighter signals in thin samples compared with backward-detection SHG (Supplementary Fig. S4).\(^4\)\(^0\) We also quantitatively confirmed that forward- and backward-detection SHG configurations yield statistically comparable fiber measurements (Supplementary Fig. S5). Although many of the same collagen features are seen by both PSR-POL and SHG, it was apparent that the images did not perfectly mirror each other (Figs. 2, S2, and S3). We quantified the images using an open-source, validated software program, CT-FIRE, developed by our group, which is capable of reducing noise, extracting, and measuring important fiber metrics, including count, alignment, length, straightness, and width.\(^3\)\(^0\) The quantitative scoring ability of CT-FIRE has several advantages over standard qualitative examination including increased objectivity, reproducibility, and sensitivity. We observed that CT-FIRE was compatible with grayscale images generated by both SHG and PSR-POL modalities (Supplementary Fig. S4). As depicted in Fig. 4, we determined that PSR-POL produced higher fiber count, length, width, and straightness values than SHG.

Figure 4. Left column: Quantification of different collagen fiber metrics (count, alignment, length, straightness, and width) in images acquired from pancreatic cancer tissues using SHG and PSR-POL. Right column: Pearson correlation analysis between metrics generated by SHG and PSR-POL imaging. Data represent 92 patient tumors. Abbreviations: SHG, second harmonic generation; PSR-POL, picrosirius red stain with standard polarization microscopy.
(p<0.0001). Alignment was also higher in PSR-POL images (p=0.0154). Similar quantitative trends were observed in prostate tissues (Supplementary Fig. S6), suggesting that the results were not cancer-type specific. Using Pearson correlation analysis, we determined that length (r = 0.159, p=0.129) and width (r = 0.174, p=0.097) were not significantly correlated between the two techniques. Straightness (r = 0.430, p<0.0001), alignment (r = 0.565, p<0.0001), and count (r = 0.656, p<0.0001) did show significant positive correlation between PSR-POL and SHG.

Based on our data, we believe that PSR-POL may be a practical option to investigate stromal collagen in thin tissue sections depending on the application and particular metric under examination. For applications that require absolute measurements predetermined by SHG, PSR-POL needs to be considered carefully because the technique appears to generate statistically higher values than SHG. The reasons for these differences are currently unknown. It has been suggested that PSR-POL can be sensitive to fiber thickness,16 maturity,18 and collagen type.37 Further investigations will be of great interest to determine which collagen types and characteristics are being visualized or missed by each imaging technique.41 However, for applications comparing the relative magnitudes of fiber properties, PSR-POL is positively correlated to SHG for count, alignment, and straightness and therefore might be a viable alternative for assessing these in routine pathological practice. Of particular interest is fiber alignment, which out of all collagen metrics assessed here has been demonstrated to correlate most strongly with disease progression and yields slightly higher measurements in PSR-POL images compared with SHG. As the emerging clinical value of other fiber metrics (ie, length, width) become clearer, imaging methods and technologies will likely need to be reconsidered to determine which would be most appropriate for facile visualization and analysis.

There are advantages and disadvantages to SHG and PSR-POL as outlined in Table 1. SHG is optimal for directly visualizing fibrillar collagens with submicron resolution; however, its use mandates costly equipment and technical expertise. On the other hand, PSR-POL is already utilized in many histology labs and therefore has a low cost and barrier to implementation. Although PSR dye alone does not specifically stain collagen,43 it can be used to directly visualize anatomic structures within tissue sections, which is important for spatial orientation and analysis of stromal collagen relative to a defined tumor boundary, as demonstrated in breast cancer.3,29 When PSR is paired with polarized light, collagen can then be visualized in the same field of view due to specific enhancement in birefringence by certain types of fibrillar collagen.41 PSR-POL can also be combined with fluorescence microscopy44 or more advanced polarimetry setups45 to gain additional information about the tissue. Although not encountered in this study, PSR-POL may also have limitations including potential reproducibility issues due to variations in staining protocols46 and will need to be assessed in other tissues. Until further validation studies are performed, SHG remains the gold standard for assessing collagen alignment in the research setting, particularly for studies requiring deep three-dimensional intravital imaging,4 forward–backward scattering ratios,40 and circular versus linear polarization dependencies.47

| Table 1. Advantages and Disadvantages of Each Imaging Modality for Collagen Visualization. |
|---------------------------------|---------------------------------|
| SHG                             | PSR-POL                         |
| **Contrast mechanism**          | Intrinsic signal from light interaction with non-centrosymmetric fibrillar collagen structure | Collagen affinity for Sirius Red F3B |
| **Specificity**                 | Direct visualization of fibrillar collagens (ie, types I, III). Also sensitive to other harmonophores (actomyosin, microtubules) | PSR dye alone is not specific for collagen, but when paired with polarized light, fibrillar collagens undergo specific enhancement in birefringence enabling visualization |
| **Equipment**                   | Extensive equipment             | Standard transmission pathology scope with two accurately crossed polarization filters (polarizer and analyzer) |
| **Cost**                        | Expensive                       | Inexpensive |
| **Resolution**                  | Easier to visualize individual collagen fibers. Forward–backward scattering ratio | Easier to visualize bundles of collagen. Detected straightness and width significantly higher |
| **Utility**                     | No staining required and can be done on unlabeled sections making retrospective studies possible. Submicron resolution allows visualization of individual fibers. Forward–backward scattering ratio. Can optically section thick (500–1000 µm) tissues | Requires staining of sections which might not be possible for retrospective studies. Resolution limited by visible light illumination. Can visualize counterstained cell structures directly without polarizer in place enabling examination of the stroma relative to a boundary |

Abbreviations: SHG, second harmonic generation; PSR-POL, picrosirius red stain with standard polarization microscopy; PSR, picrosirius red.
Given accumulating literature that points to a correlation between collagen properties in peritumoral stroma to stage and patient outcome, there is desire by clinical and experimental pathologists to characterize tumor tissue in a manner that includes assessment of collagen organization in the spatial context of malignant cells. PSR-POL is an attractive method for visualizing collagen fibers in histopathological sections, along with routine H&E, due to its convenient availability in most clinical laboratories and ability to specifically isolate collagen contrast with linear polarizers on a brightfield microscope. PSR-POL generates images that are quantifiable by existing software packages and comparable with SHG, which has so far served as the gold standard for quantitative assessment of collagen. PSR-POL could therefore be a practical, easy-to-implement method for improving the evaluation of cancerous tissues in the clinical pathology setting.

Acknowledgments
We thank Mina Alfi for useful discussions, Dr Ruth Sullivan for pathology advice, Dr Lee Wilke and Dr Patricia Keely for access to human breast cancer tissues, and Peter Crump for guidance on statistical analyses.

Author Contributions
CRD, AGL, WAR, SMW, WJK, and KWE conceived the study and drafted the manuscript. AGL provided pathology guidance. CRD, KM, GM, and SL performed all of the imaging. AK and YL assisted in image segmentation and quantification. CRD carried out all data analysis. All authors have read and approved the final manuscript.

Competing Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The tissue microarray resource used in this study was made possible by a generous donation to the University of Wisconsin Comprehensive Carbone Cancer Center by Teresa’s Foundation for Pancreatic Cancer. We thank National Institutes of Health for their financial support P30 CA014520 (University of Wisconsin Comprehensive Carbone Cancer Center facilities and services) and U54 DK104310 (to WAR and KWE).

Literature Cited
14. Nadiamyth O, LaComb RB, Brewer MA, Campagnola PJ. Alterations of the extracellular matrix in ovarian...
40. Abraham T, Kayra D, McManus B, Scott A. Quantitative assessment of forward and backward second harmonic three dimensional images of collagen type I matrix