Texture analysis applied to second harmonic generation image data for disease classification& development of a multi-view second harmonic generation imaging platform

By

Lianggong “Bruce” Wen

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Medical Physics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination 5/31/2016

The dissertation is approved by the following members of the Final Oral Committee:

Paul J. Campagnola, Professor, Biomedical Engineering

Kevin W. Eliceiri, Senior Scientist, Molecular Biology

T. Rockwell. Mackie. Professor Emeritus, Medical Physics

Vikas Singh, Associate Professor, Biostatistics

Jeremy D. Rogers, Assistant Professor, Biomedical Engineering

Timothy J. Hall, Professor, Medical Physics

Wally Y. Block, Professor, Medical Physics

Abstract

Many diseases, e.g. ovarian cancer, breast cancer and pulmonary fibrosis, are commonly associated with drastic alterations in surrounding connective tissue, and changes in the extracellular matrix (ECM) are associated with the vast majority of cellular processes in disease progression and carcinogenesis: cell differentiation, proliferation, biosynthetic ability, polarity, and motility. We use second harmonic generation (SHG) microscopy for imaging the ECM because it is a non-invasive, non-linear laser scanning technique with high sensitivity and specificity for visualizing fibrillar collagen.

In this thesis, we are interested in developing imaging techniques to understand how the ECM, especially the collagen architecture, is remodeled in diseases To quantitate remodeling, we implement a 3D texture analysis to delineate the collagen fibrillar morphology observed in SHG microscopy images of human normal and high grade malignant ovarian tissues. In the learning stage, a dictionary of "textons"—frequently occurring texture features that are identified by measuring the image response to a filter bank of various shapes, sizes, and orientations—is created. By calculating a representative model based on the texton distribution for each tissue type using a training set of respective mages, we then perform classification between normal and high grade malignant ovarian tissues classification based on the area under receiver operating characteristic curves (true positives versus false positives). The local analysis algorithm is a more general method to probe rapidly changing fibrillar morphologies than global analyses such as FFT. It is also more versatile than other texture approaches as the filter bank can be highly tailored to specific applications (e.g., different disease states) by creating customized libraries based on common image features.

Further, we describe the development of a multi-view 3D SHG imaging platform. Unlike fluorescence microscopy, SHG excites intrinsic characteristics of collagen, bypassing the need for additional primary and secondary imaging labels. However, single view image collection

from endogenous SHG contrast of collagen molecules is not “a true 3D technique,” because collagen fibers oriented along the plane of the lasers used to excite them are invisible to the excitation The loss of information means that researchers cannot resolve the 3D structure of the ECM using this technique. We are developing a new, multi-view approach that involves rotation of agarose embedded sample in FEP tubing, so that the excitation beam path travels to from multiple angles, to reveal new insight in understanding the 3D collagen structure and its role in normal and diseased tissue.

Acknowledgements

During the time pursuing my Ph.D. degree, I have received so much help from so many people. I fell so grateful and so fortunate, and I want to express my appreciation to all of them. First, I would like to thank my main advisor Paul Campagnola for giving me guidance, support and also the freedom to explore directions I am interested for my research, without which this thesis would not be possible. And he also taught me sincere attitude exploring science and the pursuit of excellence in what I do, which I would always benefit from later in my career. Second, I want to thank my co-advisor Kevin Eliceiri for his constant support and collaborations. Also I learned from him how to explore and make best use of all possible resources. I would like to thank T. Rockwell. Mackie giving me advice and directions. His vision always amaze me, and his way of connecting concept inspires great ideas. I would also thank Vikas Singh who gave me tremendous help and lent me his wealth of knowledge in computer vision and machine learning as great tool for my research. Special thanks to Jeremy Rogers for teaching Zemax and his advice for my research. I also want to thank Robert Swader and George Petry for teaching me solidworks and also providing me a lot of help while designing the prototype for Multiview SHG imaging platform. Further I want to thank other committee member, Tim Hall and Wally Block.

My gratitude goes out to all my colleagues. I want to give special thanks Kirby Campbell since we have done a lot of the research together. And I always feel that we are like battle companions and brothers in the same trench helping each other out and learning from each other. I am very grateful for the collaborations and help from other colleagues Karissa Tilbury, Xiyi Chen, Gunnsteinn Hall, Visar Ajeti and Ben Cox.

At last I want to thank my family. My parents gave me tons of support and understanding during all these years in graduate school. I know that they were always on my back. I met my wife Heidi He in graduate school, and we had a daughter Clara last winter. They gave me so much joy and more motive for me to finish this thesis.

Table of Contents

Chapter 1: Introduction

Second Harmonic Generation (SHG) microscopy has already emerged as a highly sensitive/specific probe of collagen architecture changes in many kinds of diseases, including many epithelial cancers, connective tissue disorders, and fibroses. All these diseases can be characterized by changes in alterations of collagen density, fibrillar organization, collagen isoform distribution or combinations thereof in the extracellular matrix (ECM). In this thesis, we will focus on two main aspects of the SHG microscopy imaging technique. First, we are interested in developing machine learning and computer vision algorithms to understand how the ECM is remodeled in diseases, and further, adopting such algorithms to classify idiopathic pulmonary fibrosis (IPF) and ovarian cancer. Second, we are focusing on the development of a multi-view 3D SHG microscopy imaging platform to improve the current SHG imaging technique for 3D morphology measurement. The current SHG microscopy technique has much higher lateral resolution than axial resolution due to its probing focus profile. Also, due to dipolar orientations, ECM fiber structures aligned with the laser excitation are unable to be visualized. Our multi-view approach involves the rotation of agarose embedded samples in FEP tubing so that the excitation and emission beam paths may interrogate the sample from multiple angles revealing true 3D collagen structure with high fidelity. In this introduction, we will give a brief review of fundamentals of SHG microscopy imaging and explain the current challenges which lead to the significance of later chapters.

1.1 Second Harmonic Generation Microscopy Overview

1.1.1Second Harmonic Generation Physics

Second Harmonic Generation (SHG) is a coherent, non-linear physical process in which two photons with the same energy up-convert to one photon with twice the energy as the excitation photons (shown in Fig. 1.1). The first application of SHG imaging in the biomedical field was discovered by Freund in 1986. He successfully imaged rat tail tendon with low resolution around 50 μm. Campagnola and Mohler further developed a robust SHG imaging platform at high resolution for tissue around 2002. Since then SHG has been widely used as a tool for tissue imaging.

The complete emission polarization by laser excitation with electric field vector, $\boldsymbol{E}$, can be expressed [1]:

$$\boldsymbol{P} = \chi^{(1)}\boldsymbol{E} + \chi^{(2)}\boldsymbol{E}^2 + \chi^{(3)}\boldsymbol{E}^3 + \cdots \quad (1)$$

Where $\boldsymbol{P}$ is the induced polarization, and $\chi^{(n)}$ is the nth order nonlinear susceptibility tensor. In equation 1, $\chi^{(1)}\boldsymbol{E}$ describes reflection, scattering and absorption; $\chi^{(2)}\boldsymbol{E}^2$ describes SHG, sum-frequency generation and difference frequency generation; and $\chi^{(3)}\boldsymbol{E}^3$ describes the multiphoton absorption, third harmonic generation (THG), stimulated Raman process and coherent anti-Stokes Raman scattering (CARS).[2]

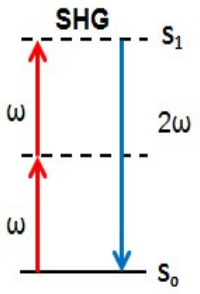


Figure 1.1, Second Harmonic Generation

The detailed derivation for the SHG term with different dipole interactions will be further discussed in chapter 5.

1.1.2 SHG and Biomedical Imaging of Collagen

The non-centrosymmetric environment at the size scale of the SHG emission wavelength provides effectent SHG production and imaging. The primary proteins that meet this requirement are type I and II fibrillar collagen and myosin within actomyosin complexes. Recent research on SHG microscopy imaging has been focusing on visualizing collagen fibers in connective tissues and internal organs. Collagen type I is the predominate component of ECM with a triple helical (~300 kD) structure having three α-chains hydrogen bonded to each other. The individual molecules self-assemble into fibrils of 20-250 nm diameters. [3] The bundle of fibrils forms individual fibers around 0.5 μm to several micrometers in diameter. Type I fibrillar collagen has now been imaged by SHG in internal organs (e.g., ovary, liver, kidney and lung), connective tissues (e.g., skin, bone, tendon), blood vessels and cornea. Other isoforms such as type II collagen, found predominantly in cartilage, also efficiently produce SHG emission.[4-6]

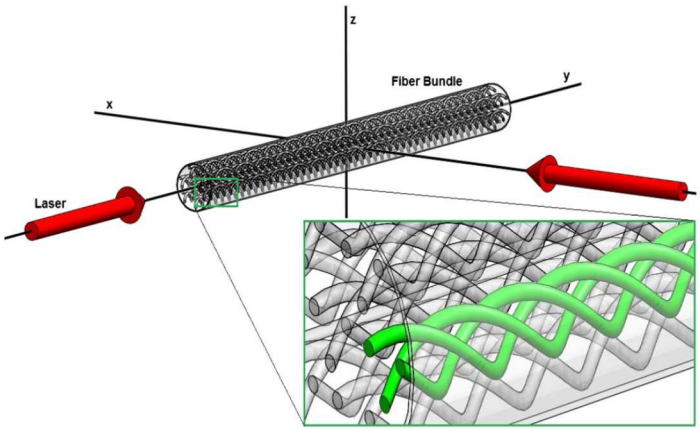


Figure 1.2. SHG imaging for mouse tail tendon.

1.1.3 Advantages of SHG Imaging

The SHG microscopy demonstrates multiple advantages over other types of imaging modalities. First, like all other multi-photon imaging, it enables intrinsic optical sectioning as only the focus of the excitation will generate sufficient emission signal. The emission intensity is largely proportional to the second order of the excitation intensity. In confocal microscopy, on the other hand, the signal of the emission is proportional to excitation energy therefore introducing background noise. Upon signal collection, confocal microscopy implores a pinhole conjugate to the focus of the excitation laser to exclude background noise and thus increase sensitivity. On the other hand, SHG has intrinsic optical sectioning with high resolution without the use of a pinhole. [7]

Another advantage of SHG microscopy imaging is that it is a label-free imaging method. The non-centrosymmetrical environment of collagen I provide endogenous SHG contrast, and collagen I assembling is an important component for ECM structures. However, common linear fluorescence imaging or two-photon excitation fluorescence (TPEF) mostly relies on exogenous labels to extract structural information through polarization or directional resolved methods. THG is also a coherent label-free imaging method, but it is restricted to regions in which a large change of refractive index occurs such as interfaces between tissue layers. CARS and stimulated Raman scattering (SRS) probie chemical bonds (primarily C-H stretches) and are useful for measuring lipids around collagen [8]. Optical coherence tomography (OCT) is widely used clinically for cornea disease and cardiovascular disease and is powerful for measuring optical properties of tissue like refractive index, density, and the scattering coefficient. However, OCT contrast is not collagen specific and is not sensitive to collagen structure. Also, OCT has a relatively low resolution of about 10-20 μm [9, 10].

1.1.4 Limitations of SHG Imaging

Several limitations exist regarding SHG microscopy. First, there are only a small number of structural proteins that this imaging modality is capable of detecting including type I and II collagen and actomyosin complexes within skeletal muscle. [11] Second, the penetration depth of SHG imaging is relatively low. Most SHG imaging applies near infrared (IR) excitation (700—1000 nm) wavelength. The penetration depth is around 5 to 10 scattering lengths which is several hundred microns depending on the tissue under examination. We could tune the

excitation wavelength towards 1200 nm, but we would not achieve penetration depth beyond 500 μm due to the photon scattering and absorption in the tissue. If longer excitation wavelengths go beyond that, water absorption will start to dominate the light attenuation, and SHG signal will get weaker. Therefore, the in vivo imaging applications would be restricted by the penetration depth.[12] It must be noted, for ex vivo imaging, depth of light penetration can be aided by use of optical clearing agents to minimize the scattering effect from tissue. Finally, the field of view for SHG microscopy imaging is relatively small. For ex vivo imaging we could stitch individual image segments together to form a mosaic of a larger field of view. However, for in vivo imaging we need to combine SHG imaging with micro-endoscopy or perhaps OCT to be more clinical impactful. The scanning speed for large scale imaging might be an issue with photon-counting PMTs as the choice of detector. However, there are emerging technologies for enhancing SHG imaging sensitivity for better signal noise ratio (SNR).

1.2 Second Harmonic Generation Microscopy Experimental Setup

The schematic of the optical layout of the SHG microscope is given in Fig. 1.3. We used a Nd:YVO4 (532 nm; 5–18 W) pumped Ti:sapphire oscillator with tuning range from 700 to 1,000 nm at repetition rate of 80 MHz, average power of 1–1.5 W and pulse width of ~100 fs, which corresponds to a bandwidth of about 10 nm full width at half-maximum (FWHM). We used 890 nm excitation as a compromise between imaging depth, viability and Ti:sapphire performance. A short wave pass (SWP) dichroic mirror following the laser filters the residual pump background at 532 nm. An optical isolator prevents back reflections from subsequent optics from re-entering the

oscillator and potentially disrupting mode locking. The setup provided ~40 dB of isolation, rejecting light of all polarization states. We used a combination of lenses to achieve best collimation and to properly fill the back aperture of medium-and high-numerical-aperture (NA) objectives to yield the best possible resolution.

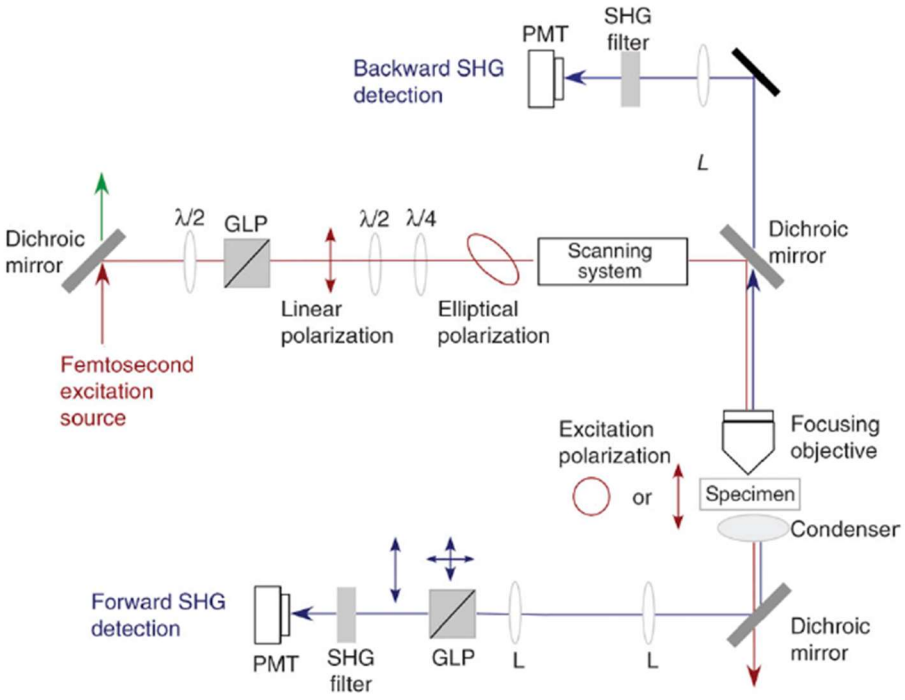


Figure 1.3. Schematic of the optical layout of the SHG microscope, showing the optical components before the scan head and the detection pathways. L, lens; λ/2 and λ/4 are half- and quarter-wave plates, respectively. Reprinted with permission from Nature Protocols. [13]

We used a Glan-laser polarizer (GLP; first set of polarization optics (shown in Fig. 2) as a linear polarizer. By changing the voltage of GLP, we can change the polarization angle. The half wave-plate behaved as a rotator of the polarization. Therefore, due to Malus's Law, the combination of

the GLP and half wave plate could adjust excitation laser power over ~100 fold, which was able to control a dynamic range of SHG intensity by about 10000 fold.. This was a relative inexpensive way for electro-optic control of excitation laser power.

In our experiments, we used circular polarization for SHG imaging as the excitation, and thus emission signal, varies with different orientations between collagen fiber and laser polarization. The use of circular polarization eliminated the directional dependence of the sample alignment. The quarter wave plate converted linear polarization to circular polarization. And another half wave plate was applied for compensation of ellipticity introduced by non-45 degree reflections in scanning system, birefringence and strain in dichroics and other optics. The distortion was calibrated using a specimen of cylindrical symmetry and achieving a 'ring stain'.

We utilized the Olympus Fluoview 300 laser scanning system and Olympus BX 61 upright microscope, which provides high efficiency of throughput for both excitation and emission of the SHG process. We further customized the acquisition beam path using two PMTs for forward and backward propagating SHG signal. Moreover, this setup allowed the use of long working distance water-immersion objectives. Most of the microscope companies now offer dipping lenses over a range of 20X to 60X, with reasonable NA (~0.5–0.9) with long working distances (e.g., 3 mm for 40X, 0.8 NA), and are optimized for transmission of the near-IR laser excitation. These lenses are ideal for imaging tissues of several hundreds of micrometers in thickness at good spatial resolution. For example, by using sub-resolution fluorescent beads and TPEF, we

have measured the lateral and axial point spread functions at 890 nm of the 0.8-NA lens to be about 700 nm and 2.5μm laterally and axially, respectively. The upright setup enabled us to use long working distance water-immersion objectives, which is helpful for thick tissue imaging. The system was optimized for 40X water immersion long distance 0.8 NA objectives. However, a wide selection of objectives is available with NA ranging from 0.25 to 1.0 and magnification ranging from 5X to 60X.

We stress that forward SHG detection is not Kohler illumination, and the height of the condenser is optimized for the maximum SHG intensity. Further, we employed a 45° long-wave pass (LWP) dichroic mirror (~100 nm bandwidth) to re-direct SHG emission onto the photomultiplier and filter the excitation laser. A narrow bandpass filter (20 nm) was also implemented to optimize the SHG emission range of the photon counting PMTs (Hamamatsu, 7421). For counting, PMTs ran in saturation mode, and individual pulses were pre-amplified, discriminated and converted into a digital signal, where the GaAsP photocathodes have quantum efficiencies of ~40% for blue wavelengths (400–500 nm). These devices have fixed gain, amplification and discrimination, and output transistor-transistor logic (TTL) pulses (logic level, 0–5 V) that can be plugged directly into the Fluoview inputs used normally for regular PMTs (internal or external). Backward SHG detection was implemented in an epi-geometry. But this setup differed from confocal detection as the desired light did not return along the excitation path, nor was it passed through a pinhole before hitting the PMT. The LWP dichroic for SHG was placed in the infinity space as for conventional fluorescence detection. Next, the arc lamp was replaced with the PMT and band-pass filter, where these were identical to those used for the forward detection. To minimize stray light, we enclosed

the microscope stand and detectors in light-tight boxes. Weak positive lens was placed before the detector to direct the light into the detector acquisition window.

1. 3 ECM Morphology Feature Extraction as a Biomarker for Disease Diagnosis

1.3.1 ECM Structure and SHG Microscopy

The studies revealed that most of the SHG contrast comes from collagen, which consists of non-symmetrical molecules with triple helical structure. Collagen is the most abundant protein in the body and is a major structural component in most types of tissues, as well as forming the ECM in many organs. Although traditional pathology focuses on cellular architecture, many recent studies have demonstrated that there is a close correlation between disease initiation/progression and remodeling of the ECM in the tissue microenvironment. For example, changes in collagen composition and morphology in the ECM have been documented for many cancers, connective tissue disorders, and organ fibrosis [13-15] (shown in Table 1.1). Traditional biomedical imaging modalities, such as CT, MRI, PET, and ultrasound lack of the sensitivity and specificity for detecting ECM variations in a multitude of diseases. However, SHG imaging microscopy has been successful as a collagen specific modality for detecting changes in tissue micro-environment. (Table 1.1). In this thesis, we will focus our study on imaging and analysis of ECM structure for idiopathic pulmonary fibrosis (IPF) and ovarian cancer using SHG microscopy imaging.

Table 1.1. Representative example applications of SHG

Application	Subspecialty	Key observation
Cancer	Breast	SHG can delineate cancer of different stages
	Ovary	SHG shows an increase in collagen fibril/fiber organization
	Skin	SHG can delineate tumor boundaries in different types of skin cancer
Fibrosis	Liver	SHG results agree with standard pathology
	Kidney	SHG results agree with pathology
	Lung	SHG morphology can delineate IPF and normal tissue
Connective tissues and disorders	Osteogenesis Imperfecta	SHG delineates normal and mutation states
	Sjogren's syndrome	SHG shows disorganized collagen in this disease SHG
	Cornea	SHG can delineate the stroma from other corneal components
	Skin damage	SHG uniquely shows changes in collagen assembly upon thermal damage
Antherosclerosis		SHG shows that collagen plaques intermingle with elastin
Model tissues		Self-assembled fibrillar gels can be imaged by SHG

1.3.2 Collagen I and Other Collagen Isoforms

The primary component of the ECM is Collagen I, but other isoforms (III, V, VI) are part of the normal ECM architecture and their relative abundance can change in the initiation/progression of different diseases. Different types of collagen are generally defined as structural molecules in the ECM containing a triple-helix domain. The procollagen molecules are covalently linked into fibrils with a diameter of ~20–250 nm, which then self-assemble to form fibers on the order of ~500 nm in diameter. The hierarchical structure of the collagen fiber, the structure observed in the SHG microscope, aligns well with the size scale of λ_{SHG}, allowing both molecular and supramolecular information to be encoded within the SHG signal. The morphology changes in ECM structure also indicate the variation of different collagen abundance during disease initiation/progression. There are researches using SHG microscopy to interrogate the structure of mixed ColI/ColV and also ColI/ColII fibrillary gels, which are models for invasive breast carcinoma and ovarian carcinoma. The higher Col III and Col V concentration led to shorter and more randomly distributed fibers and lower SHG intensity. [27.28]

1.3.3 Extract morphology features as biomarker for disease diagnosis

In pathology, slides are usually stained with hematoxylin and eosin (H&E), which labels nuclei and protein content, respectively, for ready detection of alterations present primarily in cancer cells. Pathologists have focused on cell morphology. However, recent studies have demonstrated that the ECM composition and structure directly relates to the initiation and progression of a number of diseases. The quantitative evaluation of ECM remodeling might have significant value for disease diagnosis and prognosis. A number of groups have investigated the potential of

SHG microscopy using histology slides or intact tissues from human patients. SHG imaging has demonstrated the ability to distinguish normal and abnormal tissue using both cellular and collagen features in many diseases. A number of image processing techniques were developed focusing on morphological variation evaluation of ECM structure between diseased and non-diseased tissues. Here we will discuss some of the most popular algorithms for ECM structure feature extraction.

a. Fourier Transforms

The Fourier transform is usually applied on a global scale to evaluate the fiber variation and alignment, which is effective in analyzing spatial frequencies over different scales. For instance, collagen fibers that are randomly distributed would have a circular distribution for the resulting transform. However, more aligned fiber assemblies have elliptical distributions. A combination with support vector machine (SVM) learning algorithm is adapted with breast and ovarian tissue classification. Principal component analysis (PCA) combined with 2D FFT coefficients were used to quantitatively describe the collagen fiber shape within normal, benign, and malignant breast cancer tissues. Similarly, in ovarian cancer tissues, 2D FFT analysis was able to differentiate normal stroma from cancer but was unable to differentiate types of ovarian cancer (ie, serous, mucinous, endometrioid, and mixed adenocarcinomas). The 2D FFT applied on ovarian and breast cancer tissue demonstrate its ability to discern well-organized ECM structure from less organized ECM structures. [16-20]

b. Curvelet Transform

Recent studies also have shown a relationship between collagen fiber alignment relative to the cell boundary and disease progression for many different tissue types, but especially in breast cancer. “Tumor-associated collagen signatures” (TACS) was introduced by Provenzano to describe the collagen alignment variation with the cell boundary. The collagen density and collagen fiber orientation around the breast tumor boundary were carefully evaluated using the curvelet transform. Also Conklin et al showed the “tumor-associated collagen signatures” (TACS) also retains prognostic value using a three-person-panel time-intensive manual approach [21]. This method has been further developed as Matlab software by researchers in LOCI combining the fiber identification algorithm (FIRE) [22] and the curvelet transform to provide frequency components within segmented areas around the boundary of the tumor.

Other studies adapted slide-based multi-photon emission (MPE) and SHG imaging systems for rapidly imaging standard histologically prepared H&E slides, allowing overlapping of classical bright-field and MPE/SHG microscopy images. This overlap can identify the cellular boundary required for TACS analysis, which is invisible by SHG. The tumor–stromal boundary is readily detectable in breast cancer, making TACS a straightforward and powerful assessment; however, in ovarian cancer this boundary is often not easily identified. Despite this difficulty, Adur et al, with the guidance of an expert pathologist, identified TACS signatures in normal and cancerous human ovarian tissues. Normal ovarian tissues demonstrated a TACS-2 cellular–stromal interface with long, straight, and parallel fibers, whereas malignant tissues primarily had TACS-3 fiber alignment.

c. Texture Analysis

The popular algorithms of texture analysis have also been applied for ovarian cancer and breast cancer classification. An image texture is a set of metrics calculated in image processing designed to quantify the perceived texture of an image. Image texture quantifies perceptual qualities (e.g., roughness, smoothness, or orientation of signal gradient) and repetitive patterns as a function of the spatial variations in pixel intensities around small individual regions in the image. Image features are extracted from regions of interest and such areas are treated equally. Comparing to CT-fire, texture analysis is not specific tothe relationship between the region of interest (ROI) boundary and collagen fiber alignment.

There are many algorithms for tissue classification using texture analysis. One of the most commonly implemented methods is grey-level co-occurrence matrix (GLCM) [23,24], which determines the texture via gray levels of pixels in different orientations. The gray level neighbors are recorded in matrixes. The second statistics are derived from these mastics such as entropy, homogeneity, energy, and variation which are the extracted image features. Researchers have demonstrated that GLCM is sometimes effective for the fiber structure evaluation. Watson et al has combined GLCM and 2D FFT to evaluate the ECM structure variation of a mouse ovarian cancer model with ~80% sensitivity and specificity. [24]

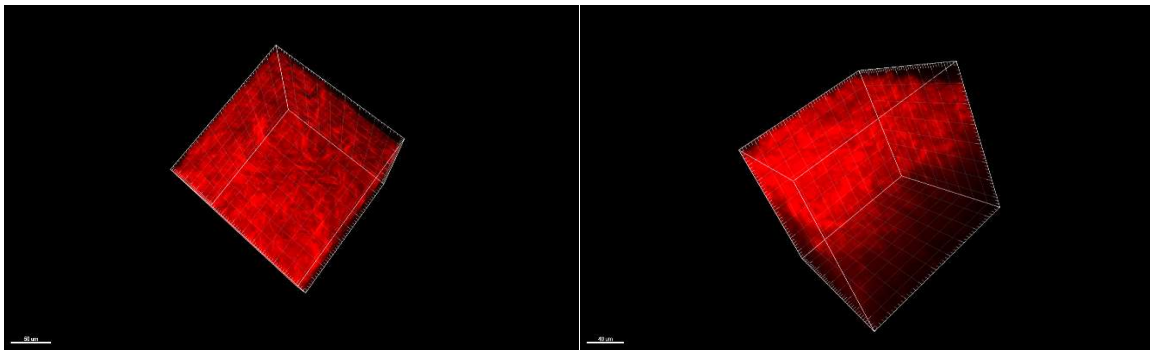
In chapter 3 and 4, we introduce a more sophisticated texton analysis for ovarian cancer classification. The idea was to first develop a set of “hand crafted” filters that would be a good

representation of collagen fibers in different size-scales. Then we convolved this set of filters with the training images. The filter responses were clustered and extracted as image features or represented image patterns. Based on the generated image features, we further model the training images with statistical distributions across the entire training image set. Then we employed one vs. all multi-class classification and nearest neighbor classification algorithm to differentiate different types of ovarian cancer tissues. We achieved 97% accuracy for 2D texton analysis with two classes, normal and high grade serious ovarian tissue. [25] We further implemented the 3D texton analysis method achieving 83-90% accuracy for 6 different types of ovarian tissues.

1.4 3D SHG Microscopy with High Fidelity

As stated previously, SHG imaging has been established as a powerful modality for visualizing the collagen assembly in a wide range of normal and diseased tissue types. Applications for imaging structural changes in many pathologic conditions, including cancers, fibroses, and connective tissue disorders have received considerable attention, as changes in the collagen rich ECM are often revealed by SHG imaging via changes in fibrillar morphology. As we discussed earlier, many different morphological image features were introduced as biomarkers for different diseases. Later in Chapter 5, we will discuss how 3D textons were generated for ovarian cancer classification with six different classifications. At the same time, it is significantly important to acquire 3D SHG imaging with high fidelity for data acquisition. However, current 3D SHG microscopy techniques are not “true” 3D imaging but constructed with a series of 2D imaging

stacking at different depth. With such imaging acquisition, there are three main factors that would affect the fidelity of the 3D imaging. First, the excitation laser focus is a probe for the collagen structure, but instead of isotropic probe shaping like a soccer, the scanning laser focus profile looks more like American football, which has smaller axial size (~0.75 μm) and bigger longitudinal size (~2 μm). Therefore, the voxel size differs from x,y and z axis to satisfy the Nyquist frequency. In Fig. 1.4 the 3D rendering of ovarian cancer tissue (using Imaris for reconstruction) demonstrate the z-axis resolution is much lower than x-y axis. Secondly, due to the electric dipole interaction, excitation of collagen fibers fully aligned parallel to the excitation laser is dipole forbidden and will not produce any SHG emission. This phenomenon is not commonly seen in fluorescence imaging as probe molecules (either dyes or fluorescent proteins)



(a) (b)

Figure 1.4. Imaris reconstructed 3D ovarian cancer tissue from stacking 2D(a) image stack with x-y scanning plan facing up (b) image stack rotated showing x-z axis with lower imaging fidelity in z axis.

typically have rotational freedom and absorb at all angles. Further discussion as well as a detailed derivation will be provided in chapter 5. Thirdly, as excitation laser penetrates though

tissue, the deep tissue imaging is affected by the photon scattering and absorption. The effect of this is significant as the penetration depth is limited to 300-500 μm (shown in Fig. 1.4) depending on the type of tissue and its optical properties.

Considering the above limiting factors, we introduced a 3D SHG imaging modality in chapter 5 which allows laser illumination and data acquisition from multiple views. The multi-view imaging system comparing to old 2D stacking system can be analogous to a CT system and X-ray system. The system would improve the 3D SHG imaging fidelity and solve some of 3D imaging challenges caused by the limiting factors mentioned previously.

At the end of the introduction, a brief review about the layout of the rest of the chapters will be provided. A brief introduction about the SHG physics, experimental setup, imaging features as biomarkers for classification and a new 3D SHG imaging platform has been discussed. In chapter 2, we will discuss in detail about how to use wavelet transform and principle component analysis (PCA) to classify idiopathic pulmonary fibrosis (IPF) and normal tissue.[26] While this algorithm was successful for differentiating IPF and normal lung tissue, it did not provide good enough classification accuracy for high grade serous ovarian cancer tissue (HGS, the most common ovarian cancer type) against normal ovarian tissue. Therefore, in Chapter 3, we further discuss the development of the texture analysis algorithm for ovarian tissue. By extracting image features from a set of 8 maximum response filters, we modeled training and testing images and further classified HGS from normal ovarian tissue in which we achieved 97% of recognition accuracy. [25] In chapter 4, we further developed 3D texture analysis for ovarian tissue stacks, which accomplished clinical impactful classification accuracy from 83%-91% among six

different types of ovarian tissues. This method could be of significant diagnostic and prognosis value as different types of ovarian cancer would require different treatment. Secondly, the high risk tissue from patients with BRCAI/II gene mutations was successfully classified using SHG imaging for the first time, a feat other SHG analysis metrics were unable to achieve. In chapter 5, we further explore the 3D SHG imaging motivation, proof of concept using reflective micro prisms. Then we conceptualize, iterate and optimize the commercially compatible platform for deep tissue imaging. Chapter 6 will be the conclusion of the thesis, in which I will discuss about the future work and clinic impact of my work.

1.5 References

1. R. W. Boyd, Nonlinear Optics (*Academic Press*, 2008).
2. Y. R. Shen, The Principles of Nonlinear Optics (Wiley-Interscience, 1984).
3. H. R. C. Screen, D. L. Bader, D. A. Lee, and J. C. Shelton, "Local Strain Measurement within Tendon," *Strain* 157–163 (2004).
4. Campagnola, P.J. & Dong, C.Y. Second harmonic generation microscopy: principles and applications to disease diagnosis. *Laser Photonics Rev.* 5, 13–26 (2011).
5. Campagnola, P.J. & Loew, L.M. Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nat. Biotechnol*. 21, 1356–1360 (2003).
6. Campagnola, P. Second harmonic generation imaging microscopy: applications to diseases diagnostics. Anal. Chem. 83, 3224–3231 (2011).

7. Zoumi, A., Yeh, A. & Tromberg, B.J. Imaging cells and extracellular matrix in vivo by using second-harmonic generation and two-photon excited fluorescence. Proc. Natl. Acad. Sci. USA 99, 11014–11019 (2002).
8. Le, T.T., Langohr, I.M., Locker, M.J., Sturek, M. & Cheng, J.X. Label-free molecular imaging of atherosclerotic lesions using multimodal nonlinear optical microscopy. J. Biomed. Opt. 12, 054007 (2007).
9. Brezinski, M.E. et al. Correlation of collagen organization with polarization sensitive imaging of in vitro cartilage: Implications for osteoarthritis. J. Rheumatol. 28, 1311–1318 (2001).
10. Liu, B. et al. Characterizing of tissue microstructure with single-detector polarization-sensitive optical coherence tomography. Appl. Opt. 45, 4464–4479 (2006).
11. Plotnikov, S.V., Millard, A.C., Campagnola, P.J. & Mohler, W.A. Characterization of the myosin-based source for second-harmonic generation from muscle sarcomeres. Biophys. J. 90, 693–703 (2006).
12. LaComb, R., Nadiarnykh, O., Carey, S. & Campagnola, P.J. Quantitative SHG imaging and modeling of the optical clearing mechanism in striated muscle and tendon. J. Biomed. Opt. 13, 021108 (2008).
13. X. Chen, O. Nadiarynkh, S. V. Plotnikov, and P. J. Campagnola, "Second harmonic
1. generation microscopy for quantitative analysis of collagen fibrillar structure," Nat. Protoc.7, 654–69 (2012).
14. Barsky, S.H., et al., Increased content of Type V Collagen in desmoplasia of human breast carcinoma. Am J Pathol, 1982. 108(3): p. 276-83.

15. Ricciardelli, C. and R.J. Rodgers, Extracellular matrix of ovarian tumors. Semin Reprod Med ,2006. 24(4): p. 270-82.
16. Kirkpatrick ND, Brewer MA, Utzinger U. Endogenous optical biomarkers of ovarian cancer evaluated with multiphoton microscopy. Cancer Epidemiol Biomarkers Prev. 2007;16(10):2048–57.
17. Williams RM, Flesken-Nikitin A, Ellenson LH, et al. Strategies for high resolution imaging of epithelial ovarian cancer by laparoscopic nonlinear microscopy. Transl Oncol. 2010;3(3):181–94
18. Adur J, Pelegati VB, de Thomaz AA, et al. Optical biomarkers of serous and mucinous human ovarian tumor assessed with nonlinear optics microscopies. PLoS One. 2012;7(10):e47007.
19. Watson JM, Marion SL, Rice PF, et al. In vivo time-serial multi-modality optical imaging in a mouse model of ovarian tumorigenesis. Cancer Biol Ther. 2014;15(1):42–60.
20. Adur J, Pelegati VB, de Thomaz AA, et al. Second harmonic generation micros- copy as a powerful diagnostic imaging modality for human ovarian cancer. J Biophotonics. 2014;7(1–2):37–48.
21. Provenzano PP, Eliceiri KW, Campbell JM, Inman DR, White JG, Keely PJ. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 2006;4:38.
22. Stein AM, Vader DA, Jawerth LM, Weitz DA, Sander LM. An algorithm for extracting the network geometry of three-dimensional collagen gels. J Microsc. 2008;232(3):463–75

23. Watson JM, Marion SL, Rice PF, et al. In vivo time-serial multi-modality optical imaging in a mouse model of ovarian tumorigenesis. Cancer Biol Ther. 2014;15(1):42–60.
24. Watson JM, Rice PF, Marion SL, et al. Analysis of second-harmonic-generation microscopy in a mouse model of ovarian carcinoma. J Biomed Opt. 2012; 17(7):076002.
25. Wen BL, Brewer MA, Nadiarnykh O, et al. Texture analysis applied to second harmonic generation image data for ovarian cancer classification. J Biomed Opt. 0001;19(9):096007. doi:10.1117/1.JBO.19.9.096007
26. Karissa Tilbury and others, 'Second Harmonic Generation Microscopy Analysis of Extracellular Matrix Changes in Human Idiopathic Pulmonary Fibrosis.', Journal of Biomedical Optics, 19.8 (2014), 086014
27. Visar Ajeti and others, 'Structural Changes in Mixed Col I/Col V Collagen Gels Probed by SHG Microscopy: Implications for Probing Stromal Alterations in Human Breast Cancer.', *Biomedical Optics Express*, 2.8 (2011), 2307–16.
28. Karissa Tilbury and others, 'Differentiation of Col I and Col III Isoforms in Stromal Models of Ovarian Cancer by Analysis of Second Harmonic Generation Polarization and Emission Directionality.', *Biophysical Journal*, 106.2 (2014), 354–65.

Chapter 2: Wavelet and PCA Texture Analysis Applied to Second Harmonic Generation Microscopy Analysis for Idiopathic Pulmonary Fibrosis

2.1 Forward

This chapter focuses on using nonlinear microscopy techniques to probe the morphological alterations of the extracellular matrix (ECM) of idiopathic pulmonary fibrosis (IPF) patients. Specficiallym MPE and SHG microscopy techniques, both the elastin and collagen signal are simultaneously collected. Alterations of the ECM were assessed using a ratio of the elastin/collagen signal, intensity of the SHG signal, collagen coverage, and morphological alterations of the collagen fibers. The elastin/collagen ratios of IPF diseased tissues indicated that the tissues were less elastic than normal tissues, consistent with the known mechanical consequences of the disease. Morphological collagen fiber alterations were assessed using Wavelet transforms, specifically the Daubachies filter family. Using the principal components of the wavelet coefficients, a simple K-Nearest Neighbor classifier provided accurate >95% classification, far exceeding the classification of simpler metrics based on SHG intensity and collagen coverage. This chapter was adapted from the paper entitled Second Harmonic Generation microscopy analysis of ECM changes in human idiopathic pulmonary fibrosis in the Journal of Biomedical Optics.[1]

2.2 Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic interstitial lung disease with unknown pathological etiology. It accounts for 34,000 deaths in the U.S. each year and the daily life of thousands more are affected by its symptoms including dyspnea, daily cough, limited exercise capacity, and fear of shortness of breath. Generally, the prognosis of IPF patients is poor in the aggregate with a median survival of 3-4 years post diagnosis, whereas a lower percentage of patients (10-15%) live 5 or more years, highlighting the heterogeneity of disease progression.[2, 3] Prognosis is poor due to both the lack of effective therapeutic options, and also due to limited knowledge of the disease pathology and underlying molecular and temporal changes associated with disease progression.

Currently IPF is thought to be due to alveolar injury leading to focal activation and proliferation of fibroblasts accompanied with mild inflammation followed by the accumulation of new extracellular matrix (ECM) and its subsequent destruction.[4] Areas of dense collagen accumulation (old scar) are juxtaposed with fibroblastic foci (new scar formation). Collagen I and other minor isoforms (type III, V) are the primary components of the new abnormal matrix,[4-7] where the balance changes during progression. For example, previous studies have found that collagen III is characteristic of early IPF, whereas collagen I dominates in late stage disease. [5, 8, 9] Changes in elastin also contribute to the ECM remodeling, where the proportion of collagen/elastin determines the elastic recoil of the lungs and airway patency.[10, 11] Elastin in normal alveolar septa is found as an organized epithelial layer of mature elastin fibers

providing the elasticity required for proper lung function, however in early IPF these fibers are degraded by MMP-9 and elastase that are released from the inflammatory cells and compromise lung patency. As IPF progresses and elastin is degraded, fibroblasts respond through synthesis of not only collagen but also elastin, however the new elastin is highly disordered and results in poor mechanical properties of the new lung matrix.[10, 11]

High resolution CT (HRCT) scans can be used to diagnose IPF when a classic radiographic pattern is present,[12] however, in many cases the appearance may not be sufficient to establish a positive diagnosis. In these cases, the gold-standard for IPF diagnosis still remains surgical biopsy followed by pathology. However, this surgery procedure carries significant morbidity due to typical IPF patient characteristics (i.e. older patient population, and often with multiple medical ailments (diabetes, heart disease etc.)) along with risk of worsening of the disease in the post-operative period.[13, 14] Despite the diagnostic utility of HRCT, the resolution is not sufficient to probe the remodeling of collagen and elastin components of the matrix, further limiting its ability to understand the disease pathology.

Non-linear microscopy techniques, including Second Harmonic Generation (SHG) and two-photon excited fluorescence (TPEF) are attractive solutions to this problem as they are able to probe the collagen and elastin, respectively of the matrix in a label-free manner. Here we posit that their combined use may contribute to enhanced diagnosis/prognosis of IPF and also further the understanding of the disease etiology and progression. SHG directly probes the structure of

collagen and has been used to describe ECM alterations in several diseases such as cancers, fibroses, and connective tissue disorders. [15-28] Multiphoton microscopy of elastin has also been used for several applications, including imaging skin and cardiovascular tissues, often in conjunction with SHG and Coherent anti-Stokes Raman Scattering (CARS).[29-31] The use of SHG and TPEF microscopy has not yet been used extensively for lung tissues and has been limited to mouse models. For example, Abraham[32] and Schanne-Klein[33] both have used SHG and TPE to study the remodeling in the lung matrix in COPD and a bleomycin-mouse model of IPF respectively. Both these studies were successful in differentiating diseased-remodeled lungs from normal lungs using a pixel-based measure of collagen coverage and a voxel ratio of the collagen/elastin balance. However additional structural information is encoded within the collagen SHG signal that was not utilized. For example, the fiber pattern observed in the SHG images in normal and IPF tissues can be used as a machine learning classification system, affording the collagen fibrillar pattern to be used as a label-free biomarker of IPF. This is important as, surprisingly, the fibrotic changes in the IPF matrix are considerably less characterized than the cellular aspects.

In this study, we take a step in this direction by using a combination of the Wavelet transform, principle component analysis (PCA), and K-Nearest-Neighbor algorithm (KNN) to more specifically probe the alterations of the collagen structure observed by SHG in IPF diseased tissues from normal lung ECM architecture. The wavelet/PCA/KNN classifier algorithm is able to accurately classify normal from IPF diseased lung tissues, potentially ushering in a non-invasive clinical technique to probe the remodeling of the ECM in this disease. We also

characterized the change in elastin/collagen balance as an additional biomarker and find the optical method is consistent with mechanical consequences of IPF on breathing. Understanding the remodeling process may enhance our ability to differentiate patients who will have rapid progression from those with slow progression and thus provide patients with a better prognosis.

2.3 Methods

2.3.1 Tissues

All lung tissues were obtained from lung transplant recipients at UW hospital Madison WI under a current IRB approved protocol. The normal tissues were from pathologist-defined normal adjacent tissue from biopsies of patients without fibrotic lung disease. Tissues were fixed in formalin and sectioned using a Vibratome (Leica VT1200) to approximately 150 μm thickness. After sectioning the tissues were stored in PBS at 4℃ until they were imaged. During imaging they were mounted on glass slides in PBS with #1.5 coverslips and Vaseline to seal the slide while imaging. A total of six normal and three IPF independent patient samples were prepared and imaged.

2.3.2 Imaging Parameters

The imaging system is described in Chapter 1, Figure 1.1. All imaging (SHG and TPEF) was performed with an excitation wavelength of 890 nm with an average power of ~ 20mW at the specimen using a water immersion 40x 0.8 NA objective. Circular polarization at the focus was used to equally probe all fiber orientations. The SHG wavelength (445 nm) in the forward channel and the elastin autofluorescence signal in the epi-fluorescence channel using a 22 nm bandpass filter (583 nm; Semrock) were collected simultaneously.

2.3.3 Wavelet/PCA/KNN Analysis

We use a Wavelet transform to obtain texture features with PCA and KNN analysis for our classification system.[34-36] The Wavelet transform decomposition provides both spatial and frequency domain information, which is intricately related to the scale and orientation of the texture we seek to characterize in the image data. The wavelet function is placed on a specific location on the image to determine the correlation coefficients. At that location in the image, the shape of the wavelet function is anisotropically scaled in two dimensions, which then captures both the width and the orientation of the fibers. This process is then translated to different regions in the image and the local Wavelet coefficients are calculated. In terms of SHG images, these coefficients correspond to the edges of the collagen fibers at different scales and angles. A pictorial diagram of the process going from raw single optical sections to wavelet coefficients is shown in Figure 5.1.

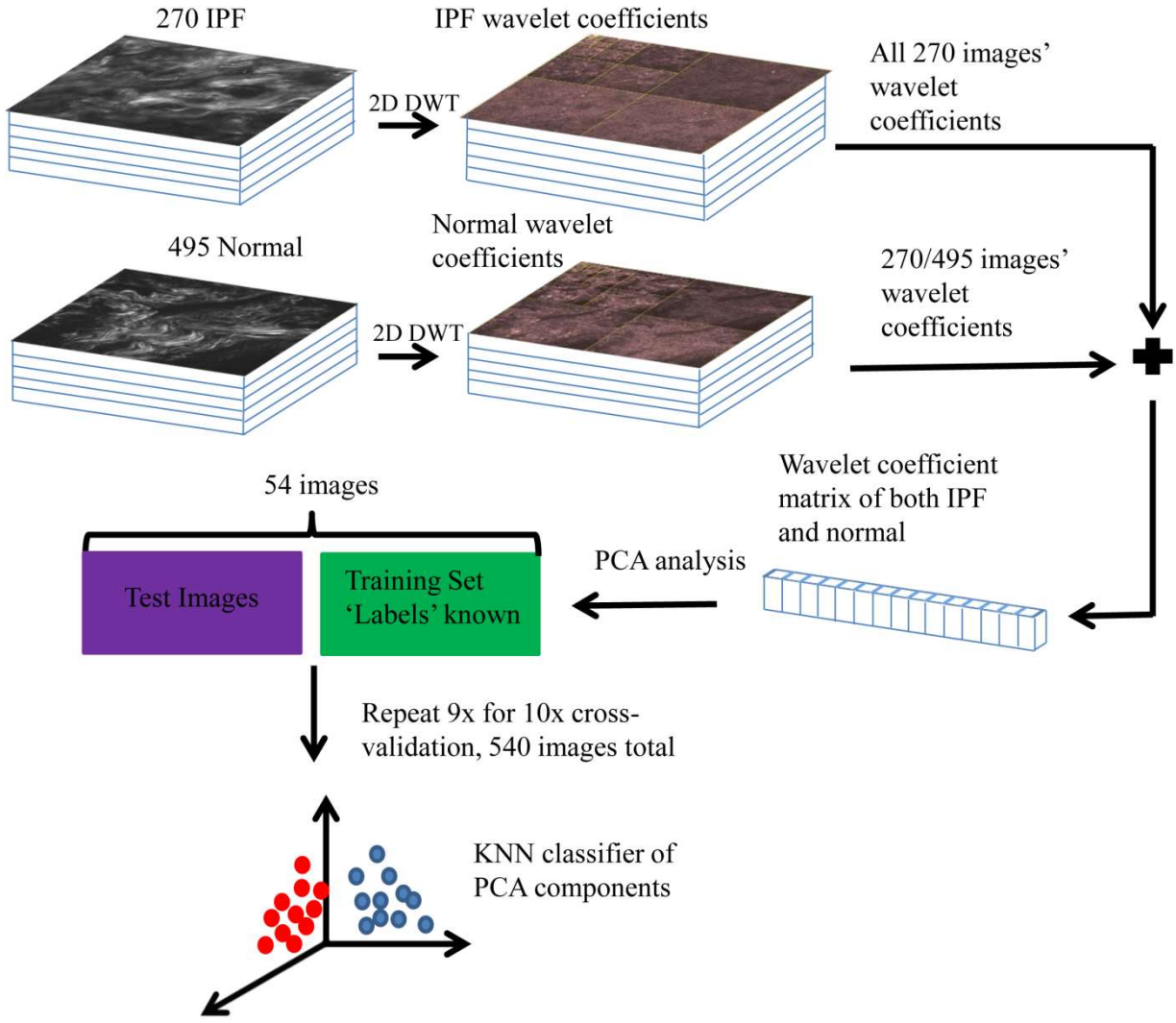


Figure 5.1 Flowchart of wavelet/PCA/KNN algorithm, beginning with the raw image data, calculating Wavelet coefficients, and then performing classification using KNN analysis of the extracted PCAs. DWT=discrete wavelet transform

Since the study size of six IPF and three normal lungs is small, working with the full set of Wavelet coefficients (which characterize the input image in terms of the chosen Wavelet basis, here the 9 filter Daubachies basis) is problematic. In particular, the number of Wavelet

coefficients one chooses directly corresponds to the dimensionality of the statistical inference problem that needs to be solved in downstream analysis. If the dimensionality of this space is large, one invariably needs to provide the model a larger number of images to make inference well-posed. The solution to this problem is to instead analyze the distribution of the Wavelet coefficients in terms of their projections on the principal components (PCs). This corresponds to the axes that explain the maximum variance, describing the full set of images with a low-dimensional representation that is more amenable to traditional statistical analysis.

Once these principal components are obtained (via the covariance matrix of the Wavelet coefficient distribution), we setup a machine learning task which constitutes two main steps. First, we use a set of "training" images where the class labels of the images are known to learn the pattern which best distinguishes one group from the other (in a space defined by treating the principal axes as the basis). Second, this pattern is used to classify test images whose class label is not known. For classification, we use a simple K-Nearest Neighbor classifier, a non-parametric method that works under the assumption that the class of each example is similar to the class of its neighbors in the space of PCA axes (see Figure 5.1). In other words, for each test image, we consider the majority of votes of its neighbors which determines the class label of the test image.

The images used in the analysis all came from three IPF and six normal patient samples that were available to us. Each patient sample had several imaging locations providing different optical stacks. Then 15 individual optical sections were selected from the middle 60% from each

optical stack. The middle regions were chosen to avoid any edge effects where the surfaces can be uneven, and also to avoid any effects of attenuation on the signal intensity. As a result of the different numbers of normal and diseased patient samples, there were 270 IPF and 495 normal available optical sections for a total of 765 images. For the wavelet/PCA/KNN analysis the sample size of the normal and IPF tissues were sized-matched, where 270 of the 495 normal images were randomly selected for a total of 540 images (270 IPF, 270 normal). The PCA dimensions resulting from the wavelet transform of 540 images were randomized and partitioned into 10 subgroups, each with 54 images for KNN classification and cross-validation. Ten KNN cross-validation trials were run, in which 9 groups served as the training set and 1 group was the testing set, each subgroup served as the test group once, as is common in cross-validation experiments.

2.3 Results

2.3.1 SHG Imaging of Normal and IPF Large Airway and Parenchymal Tissues

The ECM structure of normal and IPF lung display significant visual morphology differences in both the large airway spaces and parenchymal areas (Figure 5.2). Panels 1(a) and 1(c) are representative SHG optical sections of the collagen architecture surrounding large airways of IPF and normal lungs, respectively. Panels 2(b) and 2(d) are representative optical sections of the collagen architecture of parenchymal areas of IPF and normal lungs respectively. By visual inspection, the collagen in the diseased lungs in both the large airway and parenchymal regions is packed into denser regions than in the normal tissue. Images at 10x magnification were also acquired and were not visually different than smaller fields of view.

For translational purposes we need to develop objective quantitative methods. The simplest approach is to apply a threshold, and calculate the average pixel intensity and collagen area covered, as has been previously reported.[27] The threshold used to eliminate the background signal was determined by measuring the background signal in 15 different locations per image stack and finding the average plus the standard deviation. The applied level was specific to each optical section within the image stack was determined at each optical section to account for signal attenuation at increasing depths.

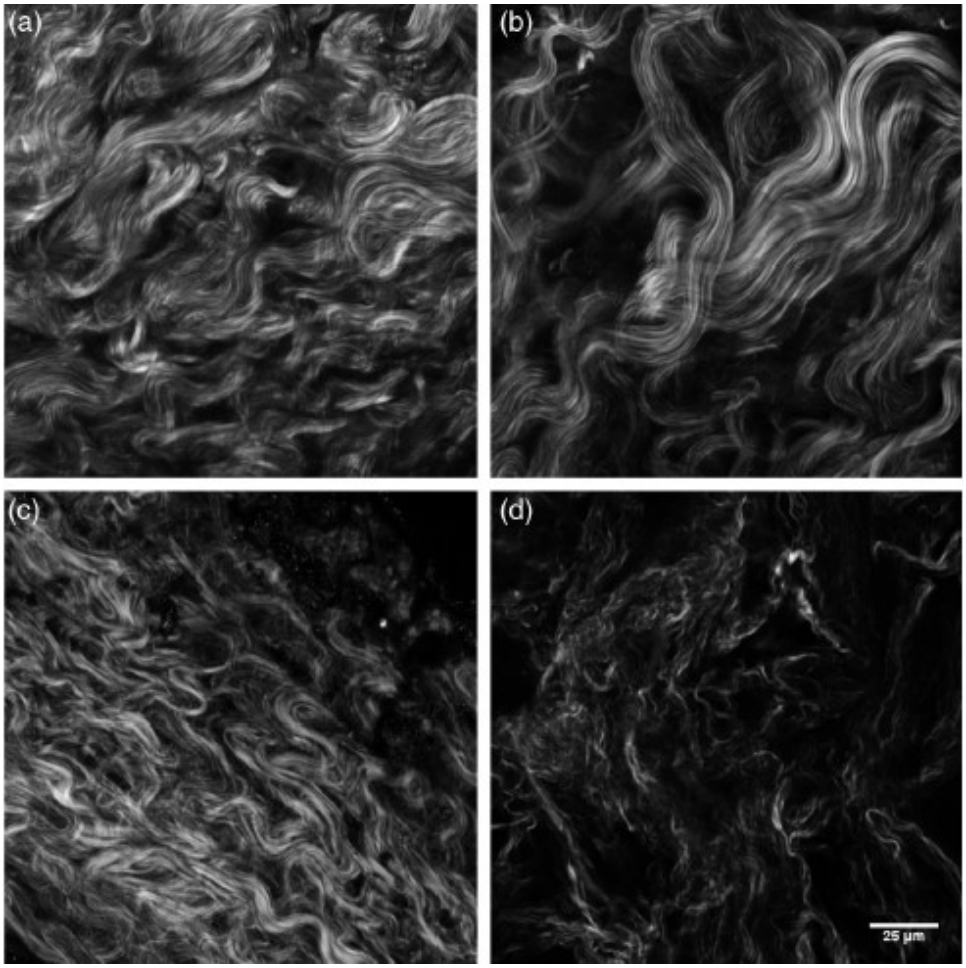


Figure 5.2 Panels are representative single SHG optical sections of a) large airway of an IPF lung; b) parenchymal region of IPF lung; c) large airway of normal lung and d) parenchymal region of normal lung. Field of view is 180 µm.

This approach showed that there were no statistical differences in SHG intensity between either normal and IPF parenchyma or large airway. The area covered, was statistically different between normal and IPF parenchyma ($p=0.008$), where the latter was higher, as might be

expected for fibrosis. However, there were no differences in coverage between normal and IPF large airway.

2.3.2 Wavelet/PCA/KNN Classification of SHG Normal and IPF Tissues

The largely insignificant results in the previous section demonstrate the need for more in depth quantitative image analysis and classification. The Wavelet transform uses the edges of the collagen fibers to provide quantification of the qualitative differences our eyes naturally detect providing a robust means independent of human visual biases of classifying tissues. The Wavelet/PCA/KNN classifier was developed for this purpose, and reliably differentiates the diseased from normal lung tissues. In this analysis, both large airway and parenchymal regions of the lungs were combined for the classification. Figure 5.3 shows plots of the receiver operator characteristic (ROC) curves (true positive vs false positive) for the classification of IPF and normal lung tissues for a few combinations of different PCA dimensions and KNN to demonstrate the dependence on the PCA and KNN parameters in classification of tissue. An area under the curve (AUC) of 1.0 is perfect classification, where 0.5 is a random outcome and provides no discrimination. In practice, values >0.9 suggest excellent test accuracy in clinical applications. Table 5.1 lists the area under the ROC curve (AUROCC) for all the combinations of PCA dimensions and KNN applied to classify the images.

The optimal classification was obtained using 20 PCA dimensions and 5 nearest neighbors with a resulting AUC of 0.998. The accuracy of classification is similar using 5 NN at all PCA dimensions. As the number of nearest neighbors is increased the accuracy of classification still

remains high even at low PCA dimension, but the accuracy increased as the PCA dimensions are increased noting that as more neighbors are included, the slight changes between the higher order PCA dimensions allow more accurate classification. We found that all combinations of PCA dimensions and KNN provided excellent classification as the worst obtained accuracy was 94%. In general, it is desirable to use as few as possible PCA dimensions and nearest neighbors to avoid overfitting errors, and this is enabled here due to the significant change in collagen fibrillar morphology.

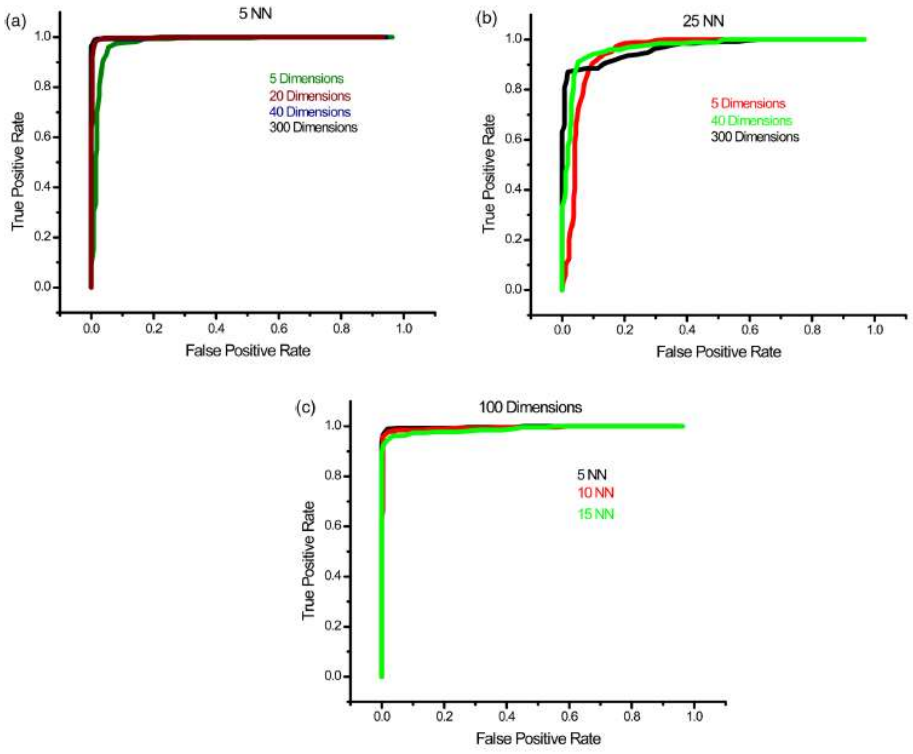


Figure 5.3 ROC curves for different combinations of PCA dimensions and KNN used to classify the normal and IPF tissues. a) Varying the number of PCA dimensions from 5-300 while constraining KNN=5. b) Varying the number of PCA dimensions from 5-300 while constraining KNN=25. c) Varying the number of KNN from 5-15 while constraining PCA dimensions to 100.

Table 5.1 Area under the receiver operator curves (ROC) for combinations of PCA dimensions and nearest neighbors.

	Number of nearest neighbors				
Number of PCA dimensions	**5**	**10**	**15**	**20**	**25**
5	0.977	0.976	0.971	0.96	0.948
10	0.997	0.986	0.986	0.964	0.962
20	0.998	0.993	0.984	0.973	0.964
40	0.998	0.991	0.978	0.966	0.953
100	0.997	0.993	0.988	0.981	0.972
200	0.998	0.994	0.987	0.98	0.969
300	0.997	0.995	0.982	0.976	0.962

2.3.3 Determination of Collagen/Elastin Balance in Normal and IPF Tissues

Elastic fiber formation is also increased in IPF[11] and the elastin/collagen ratio may be impacted during disease progression.[10] Initially in IPF patients there is an increase in the collagen deposition but late stage IPF is described as having an increase presence of elastin. This balance is important in determining the mechanical properties of the lung matrix, such as ECM stiffness and associated elastic recoil forces. We specifically probed both the collagen (SHG) and elastin (TPEF), where these contrasts were simultaneously excited at the same wavelength (890 nm) and spectrally isolated in separate channels. As the elastin contrast is linearly proportional to the concentration, and SHG is a merged effect of the square of the collagen concentration and its organization, it is not possible to determine their actual molecular ratios. The collected signal of

both the collagen and elastin signals is further confounded by the different scattering phenomena by imaging relatively thick tissue samples. However, we analyzed the volumetric ratio of elastin and collagen using the well-documented method: [EV-CV]/[EV+CV] where EV and CV represent elastin and collagen voxel volumes, respectively,[32, 37] where the limiting cases of this ratio of +1 and -1 correspond to all elastin or collagen in the pixel, respectively.

Volume fraction estimates were completed on all imaging stacks taken from the parenchyma of three normal and four IPF patient samples. The TPEF spectrum of elastin is broad and overlaps with other autofluorescence signals (e.g. cellular); therefore segmentation was required to spatially isolate the elastin signal. The intensity of the autofluorescence cellular components was much weaker than that of the elastin, which allowing thresholding for successful spatial separation as seen in raw (Figure 5.4(a)) and segmented (Figure 5.4(b)) images for normal, and raw (Figure 5.4. 4(c)) and segmented (Figure 5.4(d)) images of IPF tissues. The threshold was set using the average signal intensity of the cellular components for each stack where all the pixels with grey levels above the thresholded value were summed to calculate the volumetric fractional coverage of elastin. Similarly, the background signal of the SHG images was eliminated allowing all the pixels with signal above the threshold to be summed to calculate the volumetric fractional coverage of collagen.

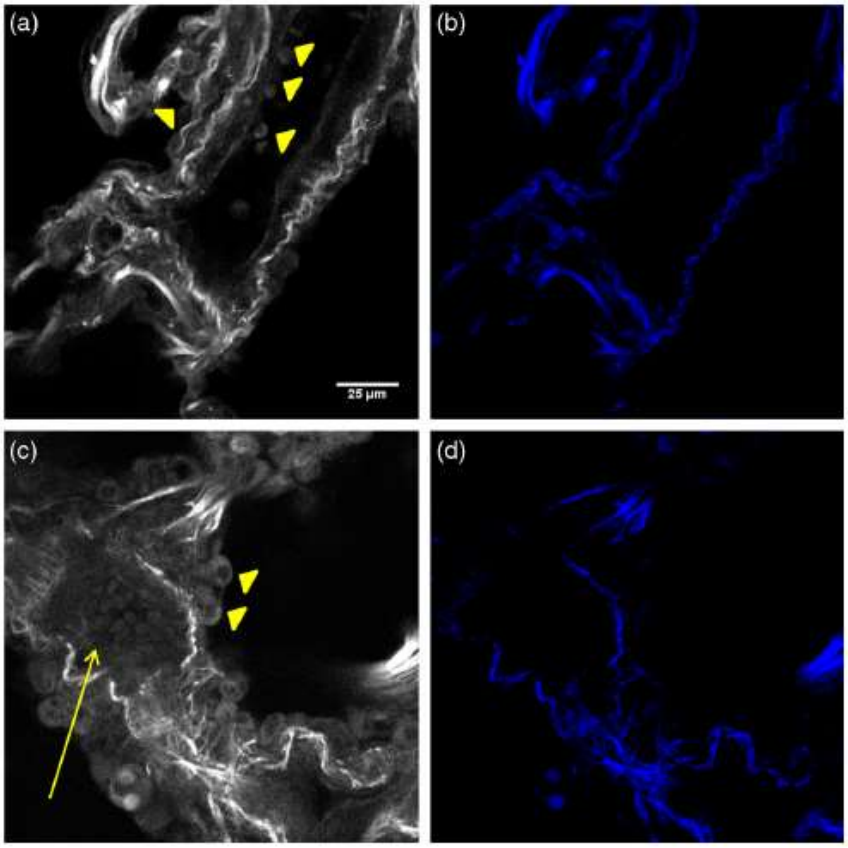


Figure 5.4 Panels 3(a) and 3(c) are representative individual TPEF optical sections of a normal and IPF diseased lung, respectively. The triangular arrows delineate cells, and the arrow indicates mucous which both are removed via thresholding techniques. The corresponding resulting segmented elastin images are shown in panels 3(b) and 3(d).

Representative background corrected two color images of the SHG (green) and elastin (blue) TPEF for normal and IPF tissues are shown in Figure 5.5(a) and Figure 5.5(b), respectively, where the organization of the collagen and the elastin are both dramatically different in these cases. Specifically, the elastin in the normal tissues is organized within the confines of the collagen, whereas in the diseased tissue the elastin is disorganized and not exclusively

intermingled with the collagen fibers. The elastin/collagen index derived from all parenchymal imaging stacks in normal and IPF tissues is shown in Figure 5.5(c), where the resulting ratios were -0.48 and -0.63, respectively, where these values were statistically significant at the p=0.07 level using a student's paired t-test. The normal tissues were more elastic relative to collagen than the diseased tissues indicative of an altered composition of the matrix. This finding determined by optical microscopy is consistent with known mechanical consequences of the disease. We note that there was no discrimination of large airways through this method.

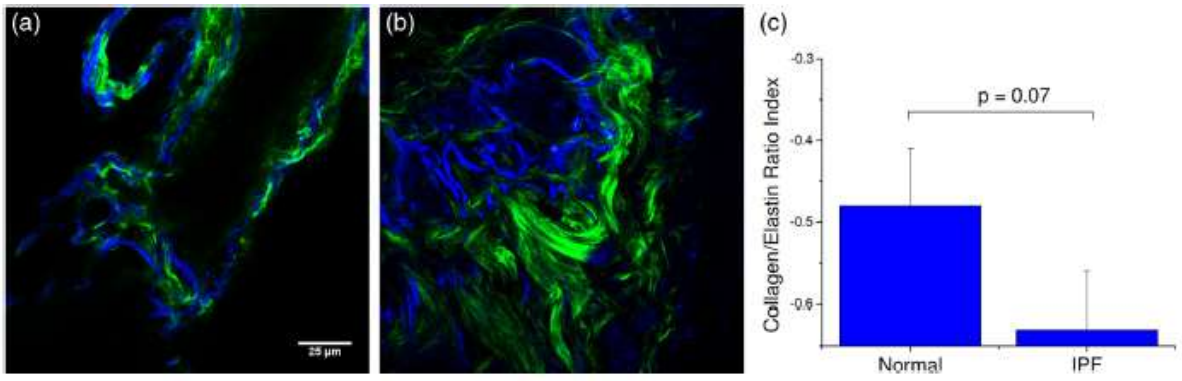


Figure 5.5 Single optical sections of a) normal and b) IPF parenchyma, where blue and green is the TPE autofluorescence from elastin and SHG from collagen, respectively. Field of view is 180µm. c) Averaged collagen/elastin ratio $[E_V\text{-}C_V]/[E_V+C_V]$ of normal (-0.48) and IPF (-0.63) parenchymal tissues where limiting values of +/- 1 are indicative of all elastin and all collagen respectively. Normal and IPF parenchyma are statistically different (p=0.07).

2.4 Discussion

The operative pathways in IPF and also the concomitant changes in the ECM are poorly understood. However the combination of SHG and TPEF microscopy affords label-free, sub-micron resolution probes of the ECM changes. Specifically, this approach may provide insights into the disease pathology and potentially guide the development of more effective therapeutic treatments. Moreover, the methods could ultimately be integrated into a scanning microendoscope.[38]

The application of standard machine learning techniques commonly used in computer vision has great potential as image classification algorithms of optical microscopy data. Simple image analysis techniques commonly used in optical microscopy, such as FFT, are highly dependent on having well-aligned structures within the imaging field of view which is rather infrequent in most biological tissues. [39-42] Moreover, FFTs provide a global, rather than local analysis. As a result, they have found their largest use in analyzing SHG images of highly organized structures like tendon. Other transforms, such as curvelets have the capability of local analysis and alleviate limitations of FFT. For example, Keely and coworkers successfully utilized a curvelet transform, a variant of a wavelet transform, in the evaluation of the collagen fiber alignment around tumor boundaries in various stages of breast cancer. [43] However, for the current case, we are interested in analyzing the overall collagen in the image where there is no specific frame of reference such as a tumor boundary.

Texture analyses are superior in this regard, as they locally probe slowly statistically, repetitive features that are present within the image. The local analysis using the integrated Wavelets/PCA/KNN approach affords the development of a tailored classification scheme based on recurring patterns in known images (although the PCAs do not correspond to visually identifiable features) which then can be used classify unknown images. This approach is powerful for the classification of overall sparse, but locally dense collagen fibers found in both normal and IPF diseased lung tissues. We could consider other texture features as well. For example, a collagen-specific morphological filter could have been designed, but this approach is time-intensive and is not flexible in its application, i.e. it must be customized by trial and error for each application. For example, we have used an implementation of texture analysis using "textons"[44] in other work on ovarian cancer.[45] This requires making an extensive library of common features. Still, the textons also do not correlate to specific features. In general, we realize that texture approaches have an inherent abstraction, as they look at slowly varying statistical patterns, rather than tangible features such as fiber length and alignment. The advantage of this approach is that Wavelets are general signal processing tools and are provided in the Matlab toolbox, where a great number are standard and can be tested to get the optimal results.

While the patient numbers here were low, the classification system had high accuracy (94-99%), based on the area under the computed ROC curves. In contrast, simpler metrics of intensity and area covered provided insufficient discrimination. The robustness of this classification algorithm is preserved across all combinations of PCA dimensions and K-Nearest Neighbors and is a major

strength of the approach. In contrast, simpler techniques such as the gray level co-occurrence matrix (GLCM) classification[46] uses only brightness of adjacent pixels rather than patterns and are not always applicable. Moreover, brightness measurements can be misleading due to scattering, change in concentration, and other unknown factors.

2.5 Conclusions

Quantifying and classifying images of biological tissues using optical microscopy remains challenging. Therefore development of standard computer vision techniques such as the current work for nonlinear optical microscopy image classification is highly beneficial to the community and will enhance our own research findings. Specifically in this particular study, it has allowed excellent classification accuracy (>94%) of normal and IPF diseased tissues, providing the initial step toward development of additional studies to probe the disease pathology. Additionally, combining the pattern analysis with the collagen/elastin balance yields a more complete picture of the alterations in the ECM in IPF. Given the robustness of this classification tool, it may be feasible in future studies with larger patient numbers to separate data from large airway and parenchymal areas to determine if classification is possible based on the remodeling of large airways alone potentially allowing for less-invasive imaging studies and diagnostic tools.

2.6 References

1. Tilbury, K., et al., Second harmonic generation microscopy analysis of extracellular matrix changes in human idiopathic pulmonary fibrosis. J Biomed Opt, 2014. **19**(8): p. 086014.

2. Fernandez Perez, E.R., et al., Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. Chest, 2010. **137**(1): p. 129-37.
3. Martinez, F.J., et al., The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med, 2005. **142**(12 Pt 1): p. 963-7.
4. Selman, M., et al., Concentration, biosynthesis and degradation of collagen in idiopathic pulmonary fibrosis. Thorax, 1986. **41**(5): p. 355-9.
5. Raghu, G., et al., Extracellular matrix in normal and fibrotic human lungs. Am Rev Respir Dis, 1985. **131**(2): p. 281-9.
6. Kirk, J.M., et al., Biochemical evidence for an increased and progressive deposition of collagen in lungs of patients with pulmonary fibrosis. Clin Sci (Lond), 1986. **70**(1): p. 39-45.
7. Fulmer, J.D., et al., Collagen concentration and rates of synthesis in idiopathic pulmonary fibrosis. Am Rev Respir Dis, 1980. **122**(2): p. 289-301.
8. Bateman, E.D., et al., Cryptogenic fibrosing alveolitis: prediction of fibrogenic activity from immunohistochemical studies of collagen types in lung biopsy specimens. Thorax, 1983. **38**(2): p. 93-101.
9. Kirk, J.M., et al., Quantitation of types I and III collagen in biopsy lung samples from patients with cryptogenic fibrosing alveolitis. Coll Relat Res, 1984. **4**(3): p. 169-82.
10. Enomoto, N., et al., Amount of elastic fibers predicts prognosis of idiopathic pulmonary fibrosis. Respir Med, 2013. **107**(10): p. 1608-16.
11. Rozin, G.F., et al., Collagen and elastic system in the remodelling process of major types of idiopathic interstitial pneumonias (IIP). Histopathology, 2005. **46**(4): p. 413-21.

12. Raghu, G., et al., An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med, 2011. **183**(6): p. 788-824.

13. Utz, J.P., et al., High short-term mortality following lung biopsy for usual interstitial pneumonia. Eur Respir J, 2001. **17**(2): p. 175-9.

14. Lettieri, C.J., et al., Outcomes and safety of surgical lung biopsy for interstitial lung disease. Chest, 2005. **127**(5): p. 1600-5.

15. Campagnola, P., Second harmonic generation imaging microscopy: applications to diseases diagnostics. Anal Chem, 2011. **83**(9): p. 3224-31.

16. Nadiarnykh, O., et al., Alterations of the extracellular matrix in ovarian cancer studied by Second Harmonic Generation imaging microscopy. BMC Cancer, 2010. **10**: p. 94.

17. Lacomb, R., O. Nadiarnykh, and P.J. Campagnola, Quantitative SHG imaging of the diseased state Osteogenesis Imperfecta: Experiment and Simulation. Biophys J, 2008. **94**: p. 4504-4514.

18. Provenzano, P.P., et al., Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med, 2006. **4**: p. 38.

19. Kapinas, K., et al., Bone matrix osteonectin limits prostate cancer cell growth and survival. Matrix Biol, 2012. **31**(5): p. 299-307.

20. Campagnola, P.J. and C.Y. Dong, Second harmonic generation microscopy: principles and applications to disease diagnosis. Lasers and Photonics Reviews, 2011. **5**: p. 13-26.

21. Cicchi, R., et al., Basal cell carcinoma imaging and characterization by multiple nonlinear microscopy techniques. Biophysical Journal, 2007: p. 157a-157a.

22. Brown, E.B., et al., In vivo measurement of gene expression, angiogenesis and physiological function in tumors using multiphoton laser scanning microscopy. Nat Med, 2001. **7**(7): p. 864-8.

23. Han, X., et al., Second harmonic properties of tumor collagen: determining the structural relationship between reactive stroma and healthy stroma. Optics Express, 2008. **16**(3): p. 1846-1859.

24. Chen, S.Y., et al., In Vivo Virtual Biopsy of Human Skin by Using Noninvasive Higher Harmonic Generation Microscopy. Ieee Journal of Selected Topics in Quantum Electronics, 2010. **16**(3): p. 478-492.

25. Lin, S.J., et al., Discrimination of basal cell carcinoma from normal dermal stroma by quantitative multiphoton imaging. Opt Lett, 2006. **31**(18): p. 2756-8.

26. Adur, J., et al., Recognition of serous ovarian tumors in human samples by multimodal nonlinear optical microscopy. J Biomed Opt, 2011. **16**(9): p. 096017.

27. Strupler, M., et al., Second harmonic imaging and scoring of collagen in fibrotic tissues. Optics Express, 2007. **15**(7): p. 4054-4065.

28. Sun, W., et al., Nonlinear optical microscopy: use of second harmonic generation and two-photon microscopy for automated quantitative liver fibrosis studies. J Biomed Opt, 2008. **13**(6): p. 064010.

29. Kwon, G.P., et al., Contribution of macromolecular structure to the retention of low-density lipoprotein at arterial branch points. Circulation, 2008. **117**(22): p. 2919-27.

30. Le, T.T., et al., Label-free molecular imaging of atherosclerotic lesions using multimodal nonlinear optical microscopy. Journal of Biomedical Optics, 2007. **12**(5): p. -.

31. Lin, S.J., et al., Evaluating cutaneous photoaging by use of multiphoton fluorescence and second-harmonic generation microscopy. Optics Letters, 2005. **30**(17): p. 2275-2277.

32. Abraham, T. and J. Hogg, Extracellular matrix remodeling of lung alveolar walls in three dimensional space identified using second harmonic generation and multiphoton excitation fluorescence. J Struct Biol, 2010. **171**(2): p. 189-96.

33. Pena, A.M., et al., Three-dimensional investigation and scoring of extracellular matrix remodeling during lung fibrosis using multiphoton microscopy. Microsc Res Tech, 2007. **70**(2): p. 162-70.

34. Unser, M., Texture Classification and Segmentation Using Wavelet Frames. Ieee Transactions on Image Processing, 1995. **4**(11): p. 1549-1560.

35. Chang, T. and C.C.J. Kuo, Texture analysis and classification with tree-structured wavelet transform. Ieee Transactions on Image Processing, 1993. **2**(4): p. 429-441.

36. Kim, W.H., et al., Multi-resolutional shape features via non-Euclidean wavelets: Applications to statistical analysis of cortical thickness. Neuroimage, 2014. **93**: p. 107-123.

37. Koehler, M.J., et al., Morphological skin ageing criteria by multiphoton laser scanning tomography: non-invasive in vivo scoring of the dermal fibre network. Exp Dermatol, 2008. **17**(6): p. 519-23.

38. Wu, Y.C., et al., Scanning all-fiber-optic endomicroscopy system for 3D nonlinear optical imaging of biological tissues. Optics Express, 2009. **17**(10): p. 7907-7915.

39. Cicchi, R., et al., Scoring of collagen organization in healthy and diseased human dermis by multiphoton microscopy. J Biophotonics, 2010. **3**(1-2): p. 34-43.

40. Adur, J., et al., Second harmonic generation microscopy as a powerful diagnostic imaging modality for human ovarian cancer. J Biophotonics, 2014. **7**(1-2): p. 37-48.

41. Ambekar, R., et al., Quantifying collagen structure in breast biopsies using second-harmonic generation imaging. Biomed Opt Express, 2012. **3**(9): p. 2021-35.

42. Fung, D.T., et al., Second harmonic generation imaging and Fourier transform spectral analysis reveal damage in fatigue-loaded tendons. Ann Biomed Eng, 2010. **38**(5): p. 1741-51.

43. Conklin, M.W., et al., Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am J Pathol, 2011. **178**(3): p. 1221-32.

44. Varma, M. and A. Zisserman, Unifying statistical texture classification frameworks. Image and Vision Computing, 2004. **22**(14): p. 1175-1183.

45. Wen, B.L., Texture analysis applied to second harmonic generation image data for ovarian cancer classification. J. Biomed Opt, 2014. **19**(9).

46. Zheng, W., et al., Diagnostic value of nonlinear optical signals from collagen matrix in the detection of epithelial precancer. Optics Letters, 2011. **36**(18): p. 3620-3622.

Chapter 3 Texture Analysis Applied to Second Harmonic Generation Image Data for Ovarian Cancer Classification

3.1 Forward:

Remodeling of the extracellular matrix (ECM) has been implicated in ovarian cancer. In this chapter we implement a form of texture analysis to delineate the collagen fibrillar morphology observed in second harmonic generation microscopy images of human normal and high grade malignant ovarian tissues to quantitate the ECM remodeling,. In the learning stage, a dictionary of "textons"—frequently occurring texture features that are identified by measuring the image response to a filter bank of various shapes, sizes, and orientations—is created. By calculating a representative model based on the texton distribution for each tissue type using a training set of respective second harmonic generation images, we then perform classification between images of normal and high grade malignant ovarian tissues. By optimizing the number of textons and nearest neighbors, we achieved classification accuracy up to 97% based on the area under receiver operating characteristic curves (true positives versus false positives). The local analysis algorithm is a more general method to probe rapidly changing fibrillar morphologies than global analyses such as FFT. It is also more versatile than other texture approaches as the filter bank can be highly tailored to specific applications (e.g., different disease states) by creating customized libraries based on common image features.

3.2 Introduction

According to the American Cancer Society, in 2013 about 22,000 new cases of ovarian cancer were diagnosed and about 15,000 women died because of this disease. [1] The survival rate for ovarian cancer has not significantly improved over the last two decades. With current screening and diagnostic abilities, about 70% of women who are detected with ovarian cancer are diagnosed in later stages, [2, 3] leading to a low 5-year survival rate of about 25%. The major problem of current diagnostic methods is the lack of reliable screening/imaging tools to detect early malignancies in the ovary. CA125 is currently the best serum biomarker, however, the sensitivity/specificity are both low. [4] For example, about 20% of women with ovarian cancer do not have elevated CA125. [2] The achievable resolution of clinical modalities (computed tomography, positron emission tomography, ultrasound, magnetic resonance imaging) is limited (only ~0.5 to 3 mm) and is not sufficient for imaging microscopic disease. This is especially important for ovarian cancer as meta- stasis can occur during early stages of tumor growth. [3]

Because of these limitations in detection and the high mortality rate, there is a compelling need for new technologies that can image ovarian cancers with better resolution and specificity and improve the accuracy of diagnosis and prognosis. Although traditional pathology focuses on cellular architecture, many recent studies have demonstrated that there is a close correlation between cancer initiation/progression and remodeling of the extracellular matrix (ECM) in the tumor microenvironment (TME). [5–9] For example, changes in collagen composition and

morphology in the ECM have been documented for many cancers, including those of the ovary, breast, and colon. [7, 10–12] It would then be advantageous to further develop collagen specific microscopic imaging modalities such as second harmonic generation (SHG) imaging microscopy13 for this purpose.

SHG microscopy has already emerged as a highly sensitive/ specific probe of collagen architecture changes in several dis- eases, including many cancers,10,11,14–18 connective tissue disorders, [19, 20] and fibroses. [21, 22] All these diseases are characterized by changes in alterations of collagen density, fibrillar organization, collagen isoform distribution, and combinations thereof. We previously utilized three-dimensional (3-D) imaging in combination with the measurement of bulk optical properties and Monte Carlo simulations to differentiate normal ovarian stroma and high grade serous carcinomas. [10] Collectively, the results indicated an increase in organization in the collagen organization at both the fibril (subresolution) and fiber levels of assembly.

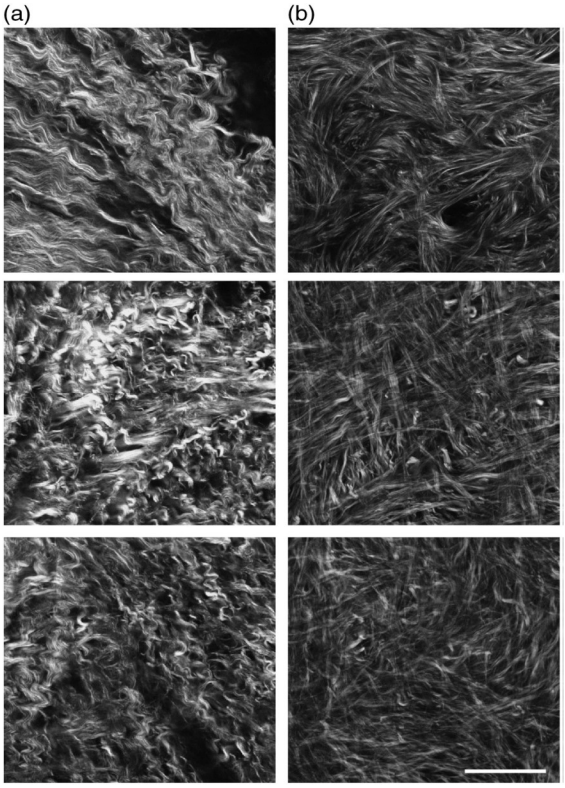


Fig. 3.1 Representative SHG single-optical sections of malignant (a) and normal (b) human ovarian tissues. Scale bar = 50 μm.

Although successful in elucidating detailed structural differences, the method is labor intensive and requires several independent measurements. It would be desirable to have a quantitative, objective measurement that can classify the state (or class) of tissues and is easy to perform. A measure based on fibrillar architecture, i.e., fiber size and organization, is attractive in this regard in general, and particularly for ovarian cancer. This is because a profound remodeling occurs in the stroma. [10, 18] As an example, single-SHG optical sections from ex vivo normal stroma (b) and high grade serous cancer (a) tissues are shown in Fig. 3.1, where these are characterized by shorter mesh-like and longer curvy fibers, respectively. By inspection, we noted that these

respective overall patterns are seen throughout tissues of each type and also between patients in each group. [10] This similarity suggests the development of a reliable image analysis approach toward a system for automated classification of these images. Still, classification is challenging because there are large stochastic variations with no highly defined fiber organization within the image.

Several image processing techniques have been employed for quantitative analysis of the collagen morphology observed in SHG microscopy. The simplest approach is to use segmentation methods. For example, Schanne–Klein used a thresholding process of image segmentation of collagen fibers for scoring fibrosis in a mouse model of kidney disease. [23] Similarly, Tai et al. [24] applied Otsu's segmentation to score liver fibrosis in both mouse and human tissues. However, segmentation is most sensitive to brightness and the collagen area covered in the image and is not as sensitive to fibrillar alignment and organization, which are often more important markers of diseased states. To help alleviate this limitation, several researchers have explored the use of other signal processing concepts. For example, FFT analysis has been used in several studies for analysis of SHG images. [25–28] Although this is simple to implement and has been successful in some cases, it is a global approach, analyzing the frequency components that are present in the entire image. However, perceptually the morphology that often discriminates one type of tissue from another is composed of predominately rapidly changing "local" features. Other transforms, such as wavelets and their variants, are more powerful for local analysis of the fibrillar morphology within such images. For example, we previously used wavelets to examine the length of sarcomeres in normal and

optically cleared skeletal muscle and calculated the entropy as a measure of organization.[29] More recently, we used two-dimensional (2-D) wavelet transforms to delineate normal lung tissue from that diseased with idiopathic pulmonary fibroses. [30] In a different approach, Keely et al. used curvelets, which are highly sensitive to edges, [31] to delineate tumor boundaries in different stages of breast cancer. [32] Although more applicable than FFT these transforms, in their simplest implementation, still lack the ability to analyze more random patterns of collagen that are representative of the stroma of most ECM tissues (nor- mal and diseased). For example, 2-D wavelet transforms were not successful in accurately classifying the ovarian tissues studied here (unpublished results).

To solve this problem for ovarian cancer, we utilized a form of texture analysis of SHG images as a classification tool. In computer vision, texture is an image property based on repetitive patterns with slowly varying local statistical propertues. Texture analysis has the strong advantage of being insensitive to intensity and not requiring long range orientation (e.g., tens to hundreds of microns). Rather, it probes the environment around small individual regions in the image, and using computer vision, extracts common features that are present. Our implementation applies the method developed by Varma and Zisserman. [33, 34] Specifically, we focused on collagen fiber distribution of the image by convolving filter patches in different directions and scales. Instead of extracting visually apparent features like angular distribution, fiber length, or area covered, as has been more commonly done, we trained on large sets of cancer and normal SHG images by clustering the filter responses within small groups of pixels using statistical methods to find common features among each tissue type. This is an important

distinction, as in real tissues it is often difficult to discretize all of the individual fibers, which leads to a loss of information.

Materials and Methods

We conducted an institutional review board-approved study of ex vivo ovarian tissues from 5 normal patients and 5 patients with high-grade serous ovarian cancer from the University of Connecticut Health Center. The diagnoses for all tissues were confirmed by pathological analysis of biopsied tissues. Tissues were fixed in 4% formalin for 24 h, transferred to phosphate buffered saline, and sliced into 100 to 200-μm-thick sections with a vibrating microtome (Vibratome, Buffalo Grove, Illinois).

Tissues were imaged by the SHG microscopy as previously described. [13] The excitation used 890-nm, 100-fs pulses from a commercial Ti:sapphire oscillator (Mira, Coherent, Santa Clara, California). The SHG laser scanning microscope was a modified Fluoview300 (Olympus, Center Valley, Pennsylvania) mounted on a fixed stage upright stand (Olympus BX61). All imaging was performed with a 40× (0.8 NA) water immersion objective lens with an average power of 20 to 50mWat the focal plane. To excite all orientations equally, circularly polarized light was used throughout. This was achieved at the focal plane using the combination of a quarter wave plate and a half wave plate as a compensator. [13] The SHG was collected in the forward direction by a 0.9-NA condenser, isolated with a 20-nm band- width 445-nm bandpass filter (Semrock, Rochester, New York) and detected by a single photon counting photomultiplier tube module

(Hamamatsu 7421, Hamamatsu City, Japan). Images were acquired at three times zoom with a field-of-view of 170 μm by 170 μm and a field size of 512 by 512 pixels to sample.

3.3 Texture Analysis Method

3.3.1 Training Images Selection Machine

Machine learning is required for the texture analysis method to obtain a statistical distribution of repetitive collagen structure patterns. This is acquired from training images of both normal and cancer tissues. For the training image set, we randomly chose single-optical sections that had at least 60% collagen coverage from each image stack. Altogether, there were 1100 selected training images (550 images each from cancer and normal tissues).We normalized the overall image intensity of each optical section to the full 12-bit dynamic range to compensate for any artifacts arising from depth dependent attenuation introduced by scattering within the tissue slice. This also compensates for the increased brightness of the SHG from the tumor specimens relative to the normal tissues. [10]

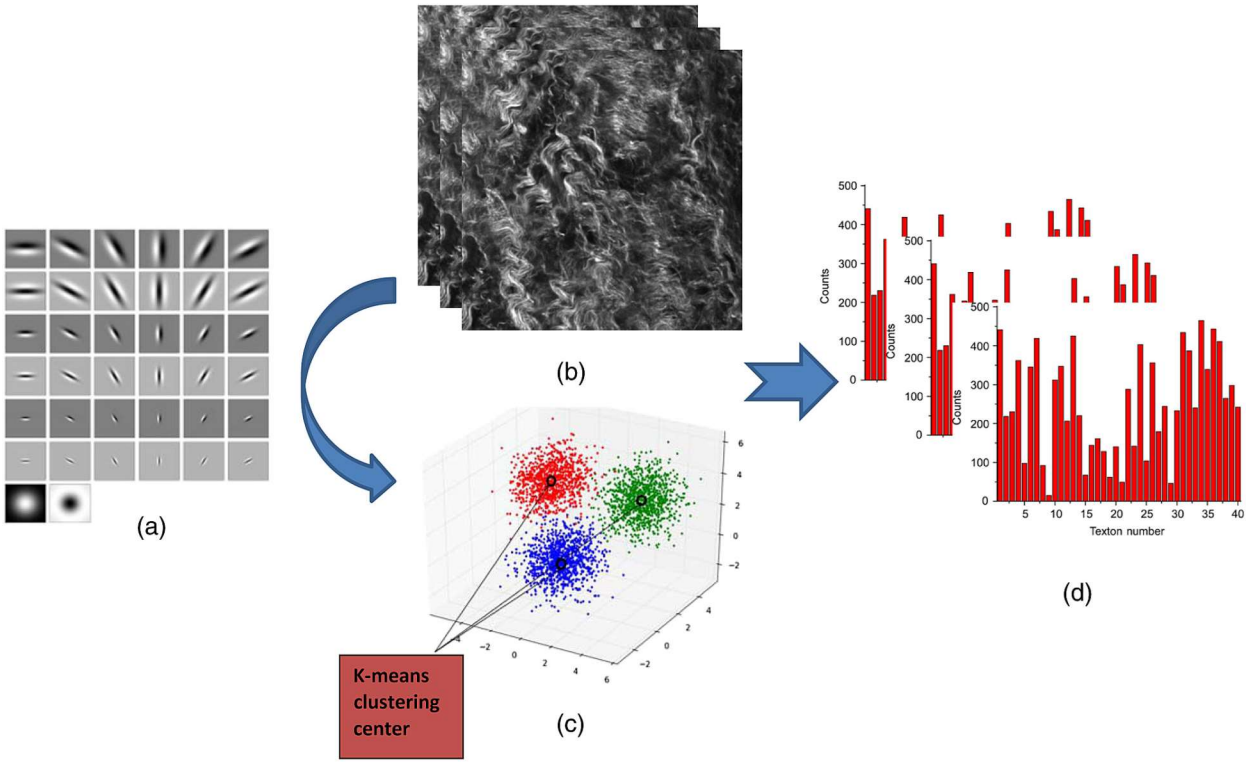


Fig. 3.2 Demonstration of the learning stage, (a) the 38 element MR8 filter bank, (b) representative training images (here cancer); and (c) K-means clustering center-texton demonstration in three dimensional filter response space; and (d) histogram models generated from the training images using 40 texton bases.

3.3.2 Filter Selection and Image Model Construction

In the learning stage, we convolved all training images with the rotationally invariant filter bank MR8 [elements are shown in Fig. 3.2(a)]. This bank has 38 filters and consists of Gaussian and a Laplacian of Gaussian filters, which are rotationally symmetric. It also includes edge filters and bar filters at 3 different scales. Both the edge and bar filters are oriented at 6 orientations at each

scale. Measuring the maximum response only across these orientations reduces the number of responses from 38 (6 orientations at 3 scales for 2 oriented filters, plus 2 isotropic) to 8 (3 scales for 2 filters, plus 2 isotropic). This provides rotation invariant behavior. [34] This is important as we do not know the orientation of the tumor relative to its point of removal and thus have no fiducial markers for placement on the microscope. Each pixel then generates an eight-dimensional vector response after convolution with this subset of the MR8 filter bank. We randomly chose 10,000 pixel responses from each training image [see e.g., Fig. 3.2(b)] to keep the computational cost feasible. These were analyzed in small "patches," with each composed of 49 × 49 pixels. The frequency of occurrence of individual patterns in an image (here histogram of textons) will then provide a so-called topic model for the image. But since the predominant filter responses (i.e., textons) are not known a priori, the standard approach in machine learning is to group the responses via a K-means clustering [Fig. 3.2(c)]. These were then used to build an overall texton dictionary. Finally, we built the classification model using the texton distribution histogram obtained for each training image [as rep- resented in Fig. 3.2(d)]. Through this process, we identified representative structural features in normal and malignant tissues based on prelearned models of the respective SHG images.

3.3.3 Classification

In the classification stage, we built a model for testing images [Figs. 3(c) and 3(d)] based on the statistical distribution of the histogram of the texton distribution for each, as was performed for the training set [Figs. 3(a) and 3(b)]. Then, we adopted a $\chi 2$ nearest neighbor (NN) classification

to determine the identity of the testing image based on the image model. We applied different thresholds for both the cancer and normal images based on the Gaussian weighted [$\exp(-d^2 / 2\sigma^2)$] distribution of NN distances for each case, where σ is the width of the distribution and d is the χ2 distance between training and testing images. We determined σ from the fitted Gaussian distribution of all NN distances from all the training images, which afforded the classification of each test image by comparison with the most similar training images around it. We then used standard 10-fold cross validation, where we randomly divided the total number of images (550 cancers and 550 normal) into 10 groups (each group then had 110 images). In the cross validation procedure, each group serves as the test set once whereas the remaining nine are the training set. This is repeated for each group, i.e., 10 times in all. The summary scores reflect the mean over all folds. In this calculation, we held out a group of images to optimize the number of NNs and achieved the highest accuracy by applying NN = 10.

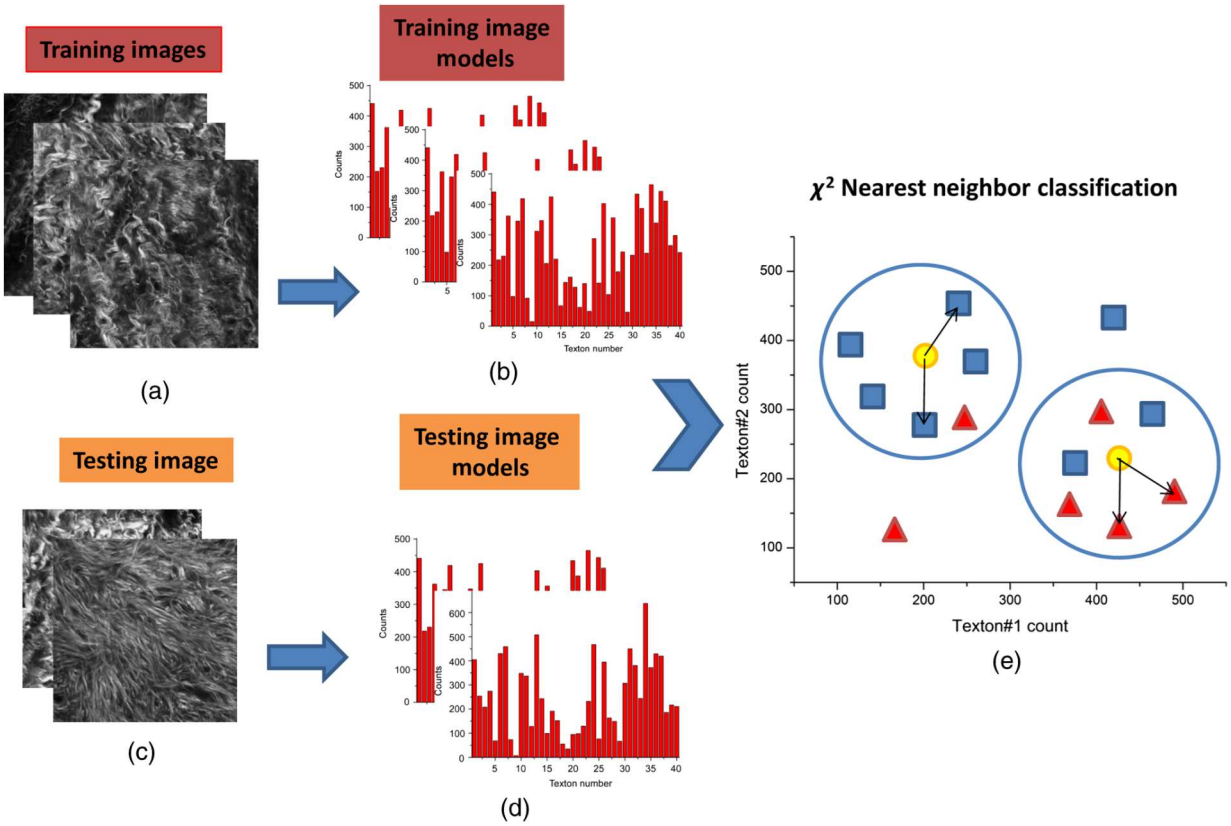


Fig. 3.3 Demonstration of the classification stage (a) representative training images; (b) the resulting histogram model generated from K-means clustering of the training images; (c) representative testing image; and (d) the histogram model generated from testing images; (e) demonstration of the $\chi 2$ nearest neighbor (NN) classification using two textons (x and y axes), where we chose six NNs around the testing images presented as yellow circles, where the normal and cancer training images are presented as blue squares and red triangles, respectively.

We diagram the classification scheme in Fig. 3(e), where purely for demonstration purposes, we assumed that 2 textons were selected to construct the image model. In the demonstration, the yellow circle represents the testing image models; the blue squares and red triangles are the

normal and cancer image models, respectively; the x and y axes are the count numbers of each of the 2 textons in the model. The distances of different image models are evaluated by the $\chi 2$ distance, which reflects the similarity of their respective statistical distributions, i.e., images that are more similar to each other will have a smaller $\chi 2$ distance. In this particular example, we chose six nearest training images away from the testing image. Then, the classification of the testing image is decided by the sum of all six weights around it.

3.4 Results of Classification Accuracy

We use the receiver operating characteristic (ROC) curve formalism [35, 36] of true positives versus false positives (or sensitivity versus 1-specificity) to determine the accuracy of the classification, where the accuracy is defined as the area under the ROC curve (AUROC).We applied the optimized NN = 10 from held out samples (determined in Sec. 3.3), and systematically changed the texton number to best represent the features in each tissue type. Figure 3.4 shows the resulting ROC curves that were generated using a range of 5 to 400 textons. The dis- crimination threshold is crucial for ROC curve generation, and this was chosen by summing up the weighting of the 10 NNs around the testing images. Using 40 textons and NN ¼ 10,we achieved a high accuracy with AUROC = 0.974.We found that if we chose 100 textons, the accuracy of classification decreased slightly to 0.963, since the frequency for each base is then so low that there were insufficient counts for a sufficient statistical distribution to construct the imaging model. As a more extreme example of this result, 400 textons resulted in yet a significantly worse discrimination (0.805). On the other hand, when the texton number is lower

than 20, the accuracy of classification also decreases due to the lack of features differentiating normal and cancer tissues. For example, 5 textons yielded a significantly lower AUROC (0.919) than the optimal value of 40.

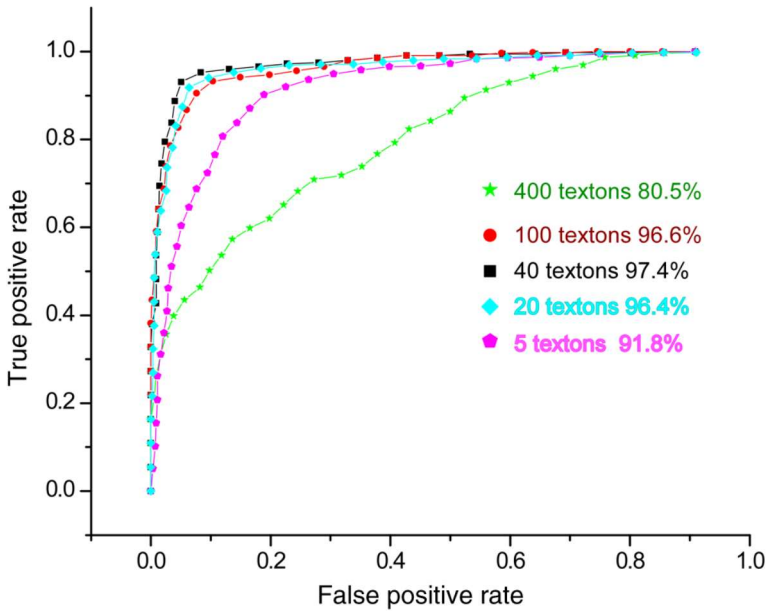


Fig. 3.4 Receiver operating characteristic curves for classification with five textons (pink pentagons) with 91.9% accuracy; 20 textons (turquoise diamonds) with 96.4% accuracy; 40 textons (black squares) with 97.4% accuracy;100 textons (red circles) with 96.6% accuracy; and 400 textons (green stars) with 80.5% accuracy.

3.5 Discussion

The current image analysis methods utilized in SHG microscopy (summarized in Sec. 1) all probe the fiber organization either globally or relative to specific features, and are not completely general in their applicability. Texture based methods may be superior in this regard. The simplest

form of texture, the gray level co-occurrence matrix, compares the brightness of adjacent pixels and has been used for SHG image analysis. [25, 37] However, it is also not a general approach as it is not as sensitive to morphology. As an alternate and more versatile form of texture analysis, Wang et al. [38] applied spectral moments to quantify intervertebral disk damage in a mouse model and successfully developed a linear discriminate classifier to differentiate loaded and sham-loaded tissues. This approach determines the 2-D frequency response and is independent of scale and orientation. As a result, this approach is a fairly general approach used for image classification. Textons, on the other hand, may have some advantages as spectral moments may be "too general;" for instance, in our application, we expect that having a dependence on orientation (as could be probed using the full set of filters in the MR8 filter bank) may actually result in an additional discriminatory capability of the statistical model. For example, one may want to be able to record the orientation of the maxi- mum response when this is relevant, as there is some clear alignment of fibers in the malignant tissues. This will yield higher order co-occurrence statistics on orientation dependent "topics" within the model and such information may be critical in dis- criminating textures that may seem similar in an orientation- independent spectral moment analysis. Further, topic models based on textons are known to yield better discrimination (at least for photographic images of naturally occurring objects) [34] than those based on the formulations where the algorithm trans- forms the data to a 2-D frequency space, where the latter loses potentially meaningful orientation information.

The texture analysis algorithm here successfully recognized the repetitive collagen fiber patterns by convolution with a standard filter bank composed of many shapes, sizes, and orientation. As a

result, this approach also affords the classification and comparison of essentially any morphology present in image data, as long as unique features can be assigned to each class. This criterion was satisfied with normal and high grade malignant ovarian tissues. By optimizing the number of textons and NNs we obtained an excellent classification accuracy of ~97%. We stress that the tumors were all high grade serous malignancies and were not representative of all ovarian cancer types. Still, excellent discrimination was achieved with a small sample set because of the large change in morphology. We note that our previous analysis using 3-D SHG imaging, measurement of optical properties, and Monte Carlo simulations delineated normal stroma and high grade malignancies using a small sample size.[10] Although that study provided insight into sub-resolution structural changes in the latter, the method requires many measurements and simulations. The routine developed here can now be readily implemented on the SHG images of new tissues in a straightforward manner, as the dictionaries from training sets are already created. Although the result of the feature extraction is sensitive to the original filter selection, it is straightforward to change the filter set until common features are located which can differentiate the tissues being compared. The drawback for this method is the large number of images required for the algorithm to extract the common features in each class. It is also not possible to directly visually associate textons with specific visual features such as fiber length and alignment. Still, our approach of comparison to a filter bank affords the specific tailoring of the feature selection to the desired application. For example, even using limited single-optical sections as inputs, the texture analysis employed here showed great potential for ovarian cancer classification.

3.6 Summary

We applied a texture analysis algorithm to evaluate the ECM structural changes in normal ovarian stroma and high grade ovarian serous cancer observed in SHG images. By optimizing the number of textons and NNs, we achieved high accuracy (97%) for classifying high grade cancer tissue and normal ovarian tissue using an ROC curve analysis. The classification algorithm is a relatively general method based on pre-learned SHG images and is well suited for analysis of rapidly changing fibrillar features typical of most tissues. The application here was for the discrete case of normal and high grade serous malignancies, but the approach could be extended to other cases such as low grade and borderline tumors.

3.7 References

1. American Cancer Society, “Cancer Facts and Figures 2013,” (2013). 2. R. C. Bast et al., “New tumor markers: CA125 and beyond,” Int. J. Gynecol. Cancer 15(Suppl 3), 274–281 (2005).
2. C. N. Landen, Jr., M. J. Birrer, and A. K. Sood, “Early events in the pathogenesis of epithelial ovarian cancer,” J. Clin. Oncol. 26(6), 995–1005 (2008).
3. E. L. Moss, J. Hollingworth, and T. M. Reynolds, “The role of CA125 in clinical practice,” J. Clin. Pathol. 58(3), 308–312 (2005).
4. S. Y. Huang et al., “Contributions of stromal metalloproteinase-9 to angiogenesis and growth of human ovarian carcinoma in mice,” J. Natl. Cancer Inst. 94(15), 1134–1142 (2002).

5. H. Hugo et al., “Epithelial-mesenchymal and mesenchymal-epithelial transitions in carcinoma progression,” J. Cell. Physiol. 213(2), 374– 383 (2007).
6. C. Ricciardelli and R. J. Rodgers, “Extracellular matrix of ovarian tumors,” Semin. Reprod. Med. 24(4), 270–282 (2006).
7. S. M. Pupa et al., “New insights into the role of extracellular matrix during tumor onset and progression,” J. Cell. Physiol. 192(3), 259– 267 (2002).
8. N. Theret et al., “Increased extracellular matrix remodeling is associated with tumor progression in human hepatocellular carcinomas,” Hepatology 34(1), 82–88 (2001).
9. O. Nadiarnykh, “Alterations of the extracellular matrix in ovarian cancer studied by second harmonic generation imaging microscopy,” BMC Cancer 10, 94 (2010).
10. P. P. Provenzano et al., “Collagen reorganization at the tumor-stromal interface facilitates local invasion,” BMC Med. 4, 38 (2006).
11. V. Backman and H. K. Roy, “Advances in biophotonics detection of field carcinogenesis for colon cancer risk stratification,” J. Cancer 4(3), 251–261 (2013).
12. X. Chen et al., “Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure,”Nat. Protoc. 7(4), 654–669 (2012).
13. E. Brown et al., “Dynamic imaging of collagen and its modulation in tumors in vivo using second-harmonic generation,” Nat. Med. 9(6), 796–800 (2003).
14. R. Cicchi et al., “Basal cell carcinoma imaging and characterization by multiple nonlinear microscopy techniques,” Biophys. J. 157a (2007).
15. S. Y. Chen et al., “In vivo virtual biopsy of human skin by using non- invasive higher harmonic generation microscopy,” IEEE J. Sel. Top. Quant. 16(3), 478–492 (2010).

16. K. Burke, P. Tang, and E. Brown, “Second harmonic generation reveals matrix alterations during breast tumor progression,” J. Biomed. Opt. 18(3), 031106 (2013).
17. N. D. Kirkpatrick, M. A. Brewer, and U. Utzinger, “Endogenous optical biomarkers of ovarian cancer evaluated with multiphoton microscopy,” Cancer Epidemiol. Biomarkers Prev. 16(10), 2048–2057 (2007).
18. R. Lacomb, O. Nadiarnykh, and P. J. Campagnola, “Quantitative SHG imaging of the diseased state Osteogenesis Imperfecta: experiment and simulation,” Biophys. J. 94(11), 4504–4514 (2008).
19. K. Schenke-Layland et al., “Increased degradation of extracellular matrix structures of lacrimal glands implicated in the pathogenesis of Sjogren’s syndrome,” Matrix Biol. 27(1), 53–66 (2008).
20. A. M. Pena et al., “Three-dimensional investigation and scoring of extracellular matrix remodeling during lung fibrosis using multiphoton microscopy,” Microsc. Res. Tech. 70(2), 162–170 (2007).
21. W. Sun et al., “Nonlinear optical microscopy: use of second harmonic generation and two-photon microscopy for automated quantitative liver fibrosis studies,” J. Biomed. Opt. 13(6), 064010 (2008).
22. M. Strupler et al., “Second harmonic imaging and scoring of collagen in fibrotic tissues,” Opt. Express 15(7), 4054–4065 (2007).
23. D. C. Tai et al., “Fibro-C-Index: comprehensive, morphology-based quantification of liver fibrosis using second harmonic generation and two-photon microscopy,” J. Biomed. Opt. 14(4), 044013 (2009).

24. R. Cicchi et al., “Scoring of collagen organization in healthy and dis- eased human dermis by multiphoton microscopy,” J. Biophotonics 3(1–2), 34–43 (2010).
25. J. Adur et al., “Second harmonic generation microscopy as a powerful diagnostic imaging modality for human ovarian cancer,” J. Biophotonics 7(1–2), 37–48 (2014).
26. R. Ambekar et al., “Quantifying collagen structure in breast biopsies using second-harmonic generation imaging,” Biomed. Opt. Express 3(9), 2021–2035 (2012).
27. D. T. Fung et al., “Second harmonic generation imaging and Fourier transform spectral analysis reveal damage in fatigue-loaded tendons,” Ann. Biomed. Eng. 38(5), 1741–1751 (2010).
28. S. Plotnikov et al., “Optical clearing for improved contrast in second harmonic generation imaging of skeletal muscle,” Biophys. J. 90(1), 328–339 (2006).
29. K. Tilbury et al., “Second harmonic generation microscopy analysis of extracellular matrix changes in human idiopathic pulmonary fibrosis,” J. Biomed. Opt. 19(8), 086014 (2014).
30. E. Candes et al., “Fast discrete curvelet transforms,” Multiscale Model. Simul. 5(3), 861–899 (2006).
31. M. W. Conklin et al., “Aligned collagen is a prognostic signature for survival in human breast carcinoma,” Am. J. Pathol. 178(3), 1221–1232 (2011).
32. M. Varma and A. Zisserman, “Unifying statistical texture classification frameworks,” Image Vision Comput. 22(14), 1175–1183 (2004).
33. M. Varma and A. Zisserman, “A statistical approach to texture classifi- cation fromsingle images,” Int. J. Comput. Vision 62(1–2), 61–81 (2005).

34. M. H. Zweig and G. Campbell, “Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine,” Clin. Chem. 39(4), 561–577 (1993).
35. N. A. Obuchowski, “Receiver operating characteristic curves and their use in radiology,” Radiology 229(1), 3–8 (2003).
36. W. Zheng et al., “Diagnostic value of nonlinear optical signals from collagen matrix in the detection of epithelial precancer,” Opt. Lett. 36(18), 3620–3622 (2011).
37. M. Wang, K. M. Reiser, and A. Knoesen, “Spectral moment invariant analysis of disorder in polarization-modulated second-harmonic-generation images obtained from collagen assemblies,” J. Opt. Soc. Am. A 24(11), 3573–3586 (2007).

Chapter 4: 3-D Texture Analysis For Classification of Second Harmonic Generation Images of Human Ovarian Cancer

4.1 Forward

To follow up with the texture analysis method in chapter 3 we further developed 3D texture analysis algorithm to quantify these alterations remodeling of the extracellular matrix (ECM), which has been implicated in ovarian cancer specifically in terms of collagen architecture. In this chapter we implemented the technique to delineate the fibrillar morphology observed in 3-D second harmonic generation (SHG) microscopy image stacks of normal (1) and high risk (2) ovarian stroma, benign ovarian tumors (3), low grade (4) and high grade (5) serous tumors, and endometrioid tumors (6).

We designed a set of 3-D filters which extract textural features in the 3-D image sets to build (or learn) statistical models of each tissue class. By applying k-nearest neighbor classification using these learned models, we achieved 83-91% accuracies for the six classes. The 3-D method outperformed the analogous 2-D classification on the same tissues. This classification based on ECM structural changes will complement conventional classification based on genetic profiles and can serve as an additional biomarker. The texture analysis algorithm is quite general, as it does not rely on single morphological metrics such as fiber alignment, length, width, etc. but instead relies on combined convolution with a customizable basis. Custom filters could then be designed for other cancers and other pathologies which involve collagen alterations.

4.2 Introduction

Ovarian cancer accounts for 5% of cancer deaths among women and is the most deadly gynecologic cancer. In 2016, an estimated 22,280 new cases and 14,240 deaths of ovarian cancer are expected in the United States. Current screening and imaging techniques are insufficient for early detection as the majority of cases (61%) are diagnosed with widespread metastatic disease, for which the 5-year survival rate is 27%. [1] CA125 is currently the best serum biomarker, but it lacks required sensitivity and specificity.[2,3] A CA125 screening combined with transvaginal ultrasound is provided for women who are at high risk; however, the resulting sensitivity and specificity of this combined approach is still insufficient for screening early tumor growth.[4] With current diagnostic imaging techniques including computed tomography, positron emission tomography, ultrasound, and magnetic resonance imaging, [5-9] only 15% of cases are diagnosed while localized to the ovary (stage 1) for which 5-year survival is 92%. [1] With these current limitations in screening and imaging modalities,[2,5] there remains a compelling need for new technologies that can image early ovarian cancers with better resolution and specificity to improve the accuracy of diagnosis and prognosis and provide new insight into the disease etiology.

Recent studies have shown that ovarian carcinomas are not homogeneous and can be broadly divided into two different types by their respective genetic mutations and epidemiological risk factors, delineated as type I and II.[10,11] The former grouping includes low-grade, borderline, and endometrioid tumors, whereas the latter refers to high grade serous carcinomas, which are the most prevalent (~70%) and have the poorest 5-year survival rates (~27%).[1] This dualistic

classification is primarily based on genetics, but as most epithelial cancers have associated ECM remodeling, it is also important to classify the respective alterations in the tumors as this will provide further diagnostic/prognostic information. Probing changes in the ECM, primarily cellular characteristics and collagen architecture, requires microscopic resolution not achievable by conventional clinical modalities. Microscopic imaging provides an option for observing early tumor initiation and progression. While H&E histology has been the gold standard for pathologic analysis (the low grade and high grade serous micro description is described below in table 4.1), it is limited in terms of numbers of sections and, more importantly, is not highly sensitive to collagen fiber organization. A more complete classification based on cellular and ECM features could improve the current standards of care.

Table 4.1 Micro Description of high grade and low grade serous ovarian cancer tissue [30]

	Pathology description
Low Grade	Extensively papillary with many psammoma bodies
	May have papillae, glands, cribriform glands, cysts, or irregular nests of cells with uniform round to oval nuclei with evenly distributed chromatin, variable nucleolus
	Usually 10 mitoses/HPF
	Variable amounts of fibrous stroma
High Grade	Branching papillary fronds, slit-like fenestrations, glandular complexity, moderate to marked nuclear atypia with marked pleomorphism, prominent nucleoli, stratification, frequent mitoses, stromal invasion (irregular or destructive infiltration by small glands or sheets of cells)
	Variable psammoma bodies (calcium concretions with concentric laminations, may be intracellular due to autophagocytosis)
	Stroma may be fibrous, edematous, myxoid, or desmoplastic

Enabling a more complete classification is important as recent studies have demonstrated a close correlation between cancer stages with remodeling of the ECM, in the tumor microenvironment (TME) in several carcinomas.[12-16] For example, using the collagen specific modality of second harmonic generation (SHG) microscopy, Keely and co-workers identified tumor associated characteristic signatures ("TACS") of collagen alignment in breast cancer. [17,18] We previously used SHG microscopy to characterize structural aspects in normal and high grade ovarian cancer using an integrated approach combining SHG creation physics and optical scattering properties. [19] While successful, this approach is both experimentally and computationally very intensive. Other reports have analyzed histologic ovarian cancer sections using techniques such as whole image Fast Fourier Transforms (FFTs). [20] However, such conventional analyses are limited in ovarian cancers. This is because while altered from normal tissue, ovarian cancers do not display the easily identifiable TACS observed in breast cancer. Moreover, as we will demonstrate here and elsewhere, the modifications are substantially different for type I (and within type I group) and type II. Thus, a general and more versatile approach is needed to classify collagen architecture alterations across the spectrum of ovarian cancers. For example, our attempts at classification using FFTs or even 2-D wavelet transforms did not achieve sufficient accuracy (<70%) in this setting (unpublished results).

Our approach to the addressing the heterogeneity of this disease is to implement a computer vision scheme based on texture analysis by creating so-called "topic models" for each type of

ovarian cancers and benign tissues. In computer vision, texture analysis attempts to quantify perceptual qualities (e.g., roughness, smoothness, or orientation of signal gradient) and repetitive patterns as a function of the spatial variations in pixel intensities around small individual regions in the image. In previous work, we had successfully implemented 2-D texture analysis for SHG images from two classes (normal stroma and high grade serous cancer tissue) using textons.[21] Textons are repeating features determined by convolution with a filter set that fundamentally comprise of a series of edge, bar, Gaussian, and Laplacian filters. [22] Models for training set images are created by identifying common responses of the collagen morphology, and test images are classified by k-nearest neighbor (K-NN) analysis, a non-parametric classification algorithm based on a number of training samples, k, with the shortest χ^2 statistical distance from testing sample. This was highly successful in classifying high grade serous ovarian cancers and normal ovarian stroma (95% accuracy). [21] Importantly, this approach is general as it does not rely on direct visuals such as fiber alignment and size.

While successful, we utilized this approach only on a series of 2-D optical sections. As SHG imaging has intrinsic 3-D imaging capabilities, and the stroma has 3-D architecture in terms of fiber organization, we are likely missing valuable information. In this paper, we extend the texton-based analysis to 3-D and apply the process to different ovarian tissues types. Although 3-D texture analysis (e.g., using grey level co-occurrence matrix and wavelet transforms) has been successfully adopted in radiomics to extract image features in CT/MRI for classifying tumors, to the best of our knowledge, there are no analogous reports using 3-D SHG images of tumor ECM. This is an important goal for classification of ovarian cancer given its

heterogeneity. Here we demonstrate the 3-D texton approach to differentiate six classes of ovarian tissues across a spectrum of diseases, and using the *one-vs.-rest* approach, we found the receiver operator characteristic (ROC) with better than 83-91% accuracy. By comparison, we achieved much lower accuracy and unsatisfactory performance using our previous 2-D approach when applied to the same six classes.

4. 3 Material & Methods

(1) Tissue Acquisition

We conducted an institutional review board-approved study of ex vivo ovarian tissues from 5 normal patients, 5 patients with high-grade serous (HGS) ovarian cancer, 5 patients with benign tumors, 5 high risk patients with BRCAI and/or BRCAII gene mutations, 3 patients with low-grade serous (LGS) tumors, and 5 patients with endometrioid tumors from the University of Connecticut Health Center and University of Wisconsin-Madison. The diagnoses for all tissues were confirmed by pathological analysis. For SHG imaging, tissues were fixed in 4% formalin for 24 h, transferred to phosphate-buffered saline, and sliced into 100 to 200 μm thick sections using a Leica Vibratome 1200S (Leica Biosystems, Buffalo Grove, IL).

(2) SHG Imaging

Tissues were imaged by SHG microscopy as previously described. [19, 23] The excitation used 890 nm, 100 fs pulses from a Ti: sapphire oscillator (Mira, Coherent, Santa Clara, California). The SHG laser scanning microscope was a modified Fluoview300 (Olympus, Center Valley, Pennsylvania) mounted on a fixed stage upright stand (Olympus BX61). All imaging was

performed with a 40× (0.8 NA) water immersion objective lens with an average power of 20 to 50 mW at the focal plane. To excite all dipole orientations equally, circularly polarized light was used throughout. This was achieved at the focal plane using the combination of a quarter wave plate and a half wave plate as a compensator. The SHG emission was collected in the forward direction by a 0.9 NA condenser, isolated with a 20 nm bandwidth 445 nm bandpass filter (Semrock, Rochester, New York), and detected by a single photon counting photomultiplier tube module (Hamamatsu 7421, Hamamatsu City, Japan). Images were acquired at two times zoom with a field-of-view of 170 μm by 170 μm corresponding to a field size of 512 by 512 pixels to sample at the Nyquist frequency. Similarly, axial sectioning was acquired in one micron steps. Considering the reduction of the signal with increasing depth into the tissue, we only employed the top 30-40 μm of the tissue stack to maintain approximately the same level of contrast between images. 3-D image reconstructions were performed in Bitplane Imaris (Bitplane AG, Zurich, Switzerland).

(3) Texture Analysis Method

We previously applied 2-D texton analysis, utilizing k-means clustering to train 2-D texture features and adapted K-NN classification to classify normal and high grade serous ovarian cancer tissues. [21] Here we further develop and adapt this algorithm for 3-D texture analysis targeting classification with six different types of ovarian tissues.

First, we redesigned the texture filter set for 3-D application. The 3-D filter set is a multi-scale, multi-orientation filter bank with 110 filters. The representative filter bank is demonstrated in Figure 4.1. It consists of first (Type I, Fig. 4.1 (a)(b)) and second (Type II, Fig. 4.1 (d)(e)) derivatives of 3-D Gaussians at six orientations about x, y, and z-axis in three different scales

(2×6×3×3=108), one Laplacian of Gaussian (LOG) filter, and one Gaussian filter. The Gaussian (Fig. 4.1 (c)) and Laplacian of Gaussian (Fig. 4.1 (f)) filters were generated with σ =10 pixels; the first and second derivatives of 3-D Gaussians were generated at three different scales with $(\sigma_x, \sigma_y, \sigma_z) = \{(1, 1, 3), (2, 2, 6), (4, 4, 12)\}$. Measuring the maximum response only across orientations at each scale about one axis reduced the number of responses from 110 to 20, which provided rotationally invariant behavior along each axis.

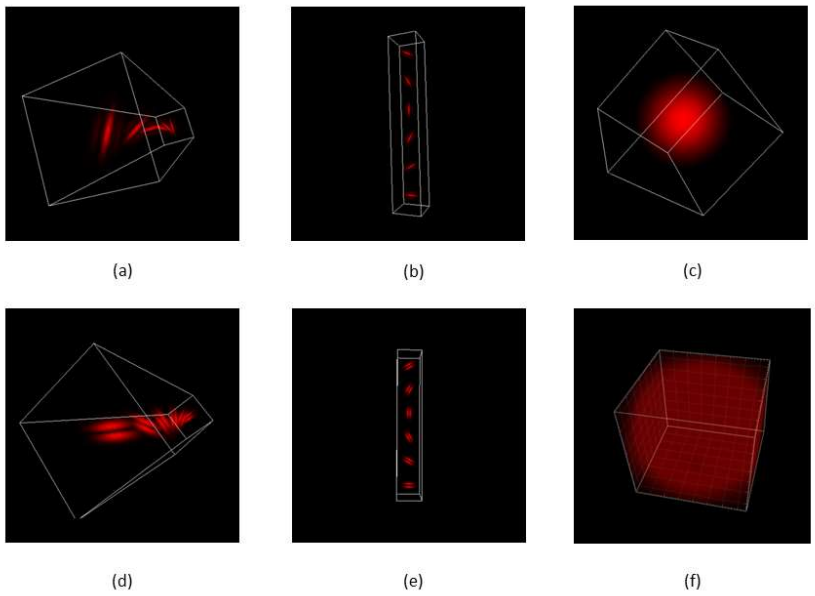


Figure 4.1. Representative 3-D image filters: 6 type-I filters at mid-scale around z-axis (a), 6 type-I filters at mid-scale around y-axis (b), 3-D Gaussian filter (c), 6 type-II filters at mid-scale around z-axis (d), 6 type-II filters at mid-scale around x-axis (e), 3-D Laplacian filter(f)

Secondly, we applied convolution to the 3-D filter bank and the 3-D SHG image data. To satisfy Nyquist criterion, the voxel sizes in the acquired data were 1.0 and 0.37 μm in the axial and lateral directions, respectively. For the convolution, we interpolated the 3-D image reconstruction so that the voxel sizes were equivalent in x, y, and z-axis. Here we employed GPU computing within the Matlab parallel computing toolbox to greatly decrease the computation time. Further, we randomly chose 12,500 voxels from the original 512 by 512 by 60 voxels in the interpolated image stack (170 μm by 170 μm by 20 μm) to keep the computational cost feasible. The filter responses of chosen voxels were grouped by k-means clustering which aims to partition the full set of patch-wise image convolution responses, where each observation belonged to the cluster with the nearest mean, serving as a prototype of the cluster or image feature. Here we clustered 20 dimensional filter responses of the 12,500 randomly chosen voxels into 40 response centers or textons. Each filter response vector is a corresponding filter descriptor for the 3-D stack patch around the chosen voxel. Therefore, the cluster centers would be filter descriptors of the 3-D stack patches representing different image features. Similar to the 2-D analysis, we found that 40 textons gave the highest accuracy. Next, we built a model for each training image stack as a histogram of the determined 3-D texton statistical distribution. The image training workflow is displayed in Figure 4.2.

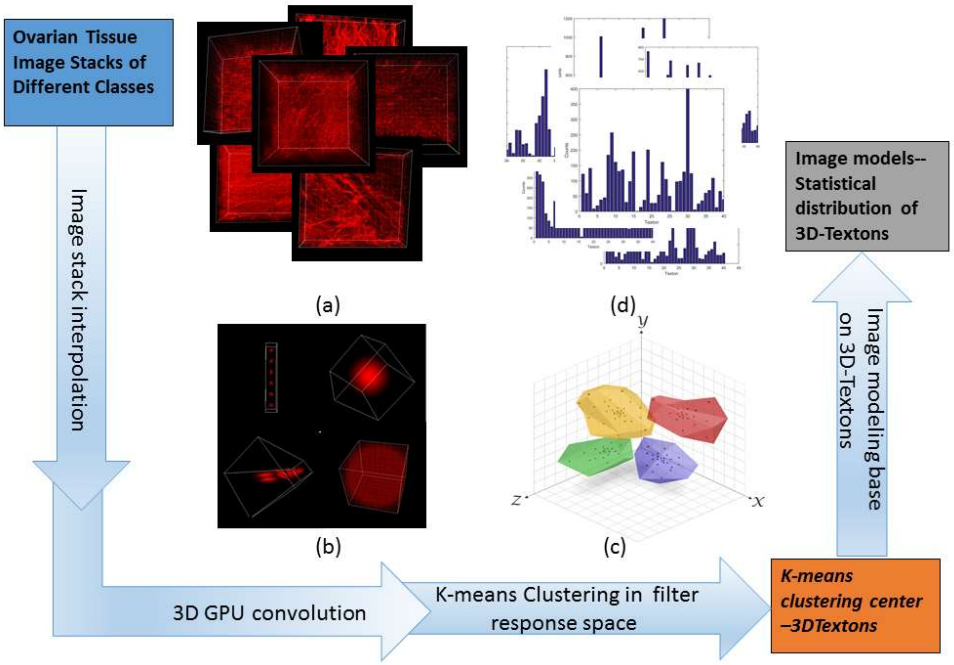


Figure 4.2. Training stage flowchart (a) 3-D renderings of SHG 3-D images of the six different types of ovarian tissues (b) representative 3-D filters (c) K-means clustering (d) histogram models generated from the training images using 40 texton bases.

In classification stage, each testing image was also built into a histogram model based on the trained 3-D textons. We used the χ^2 statistical distance, d, to evaluate differences between the statistical distributions of the textons in the training and testing stacks. Then we performed 10-fold cross validation, where we randomly divided the total number of image stacks into 10 groups. In cross validation procedure, each group rotated as the test set with the remaining nine

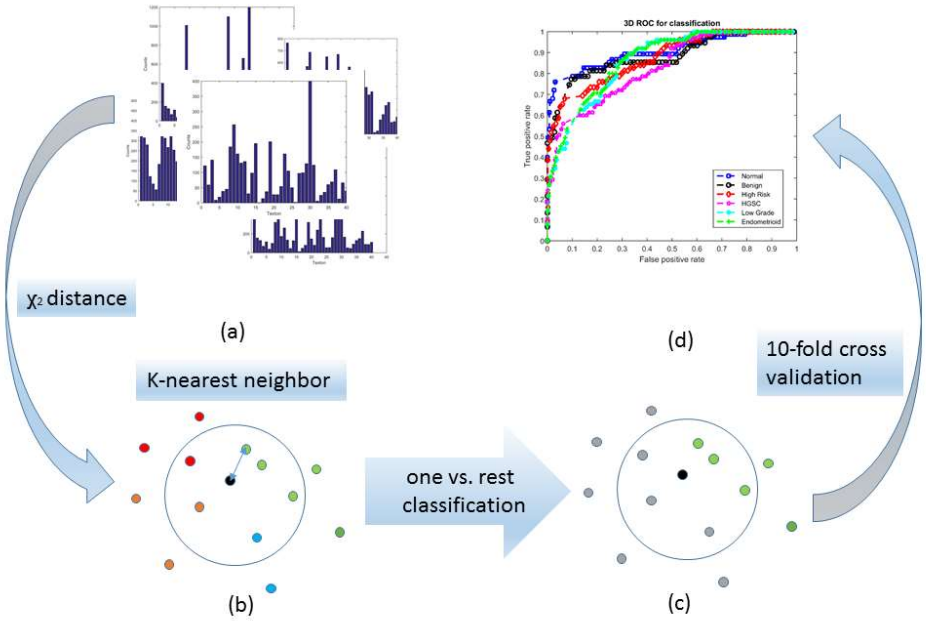


Figure 4.3. Classification stage flowchart showing (a) representative training image models; (b) demonstration of the $\chi 2$ nearest neighbor (NN) classification with four different classes (circle in four different colors); (c) demonstration of one vs. rest classification, where the black circles represent a model of the test image stacks, the green circles represent training models of one of the tissue classes, and the grey circles represent models of the remainder of the classes; (d) Resulting ROC curve.

serving as the training set. This process was then repeated and the accuracy corresponds to the average performance of the model over all ten folds. Lastly, we performed *one-vs.-rest* strategy for multiclass classification. Specifically, the *one-vs.-rest* strategy involves training a single classifier per class, with the samples of that class as positive samples and all other image stacks as negatives. This strategy requires the base classifiers to produce a real-valued confidence score

for its decision, rather than just a class label. In this case, we applied the sum of Gaussian weighting $q(\exp(-d^2/\sigma^2))$ for all nearest neighbors models relative to the testing 3-D image to evaluate thresholds for ROC curve. We determined σ by fitting a Gaussian distribution with variance σ^2 for all χ^2 distances from each training image pairs. Additionally, we optimized q (5~10) as the weighting factor for all the positive cases and we added q = -1 for all the negative cases, considering our *one-vs.-rest* classification is a semi-balanced classification method. Figure 3 demonstrates the workflow for the classification process.

In this study, we analyzed 75 image stacks for each of the six types of ovarian tissues in order to provide a comparison of the new 3-D texture analysis and our previously reported 2-D method (now expanded to multiclass classification (6)). For the latter study, we extracted five randomly chosen single optical sections from the full volume stacks to serve as training set images.

4.4. Results

Our 3-D texture classification algorithm is utilized for analyzing SHG image stacks from six different types of ovarian tissues whose class was initially identified by pathology: (1) normal ovarian stroma with no evidence of disease; (2) high risk ovarian tissue from patients with BRCAI/II gene mutations but without cancer initiation/progression as determined by pathology; (3) benign ovarian tumors which lack the ability to invade neighboring tissue or metastasize; (4) LGS carcinomas which often have a non-invasive serous borderline component; (5) endometrioid tumors of the ovary that closely mimic their uterine counterparts; and (6) HGS carcinomas. In the dualistic classification scheme,[10] endometrioid and LGS tumors are

classified as type I tumors and HGS are classified as type II, where the latter accounts for 70% of ovarian carcinomas and are usually associated with BRCAI/II and p53 gene mutations.

In Fig. 4, we first present a representative single SHG optical section and 3-D rendering from the six different types of ovarian tissues, where (a) ex vivo normal stromal tissues are characterized by shorter collagen fibers arranged in a mesh-like pattern; (b) high risk tissues are more heterogeneous and have a mixture of curvy and straight fibers; (c) benign tumor tissues are recognized by thicker, short wavy fibers relative to normal; (d) HGS tumors have characteristic long wavy fibers; (e) endometrioid tumors have a high degree of alignment but have sparser fibers than HGS, and (f) LGS tumors appear fibrotic with shorter fibers than HGS. These overall appearances are common in all the respective tissues and form the basis of using machine learning for classification. However, because of the different typical morphologies, metrics such as fiber size and alignment were not sufficient and we use the 3-D texture analysis described in the Methods and the flow charts in Figures 4.2 and 4.3.

To determine the accuracy of classification, we use the receiver operator characteristic (ROC) curves of true positives versus false positives (or sensitivity versus 1-specificity) to determine the accuracy of the classification, defined as the area under the ROC curve (AUROC). We optimized the nearest neighbor number, k, and the texton number systematically in the K-NN classification for each tissue type. We found 40 textons and k = 10 to provide the best sensitivity. Using these parameters, we achieved good accuracy with AUROC for the 6 different classes of tissues: normal stroma 90.6%, benign tumors 88.3%, high risk stroma 87.5%, HGS tumors 83.0%, LGS

tumors 86.3%, and endometrioid tumors 86.8%. Using a similar method for 2-D texture analysis as a comparison, we built a 2-D texture imaging library with five single optical sections randomly chosen from the 3-D image stack, and found lower accuracies of 89.8% for normal; 81.0% for benign tumors; 84.6% for high risk tissue; 76.5% for HGS; 82.1% for LGS; and 76.8% for endometrioid tissue. These values are summarized in Table 1 and the respective AUROC curves are shown in Figure 5 and 6. In sum, the 3-D texture analysis performs much better than the analogous 2-D approach for multi-class classification.

Table 4.1. The accuracy of classification for 2-D and 3-D texton analysis with 1 (left column) and 5 (middle) random optical section(s) from the middle axial region of each stack and the corresponding 3-D texton analysis.

Accuracy	2-D texton 1 section	2-D texton 5 sections	3-D texton
Normal	87.4%	89.8%	90.6%
Benign	77.1%	81.0%	88.3%
High Risk	80.7%	84.6%	87.5%
High Grade	70.3%	76.5%	83.0%
Low Grade	75.7%	82.1%	86.3%
Endometrioid	75.1%	76.8%	86.8%

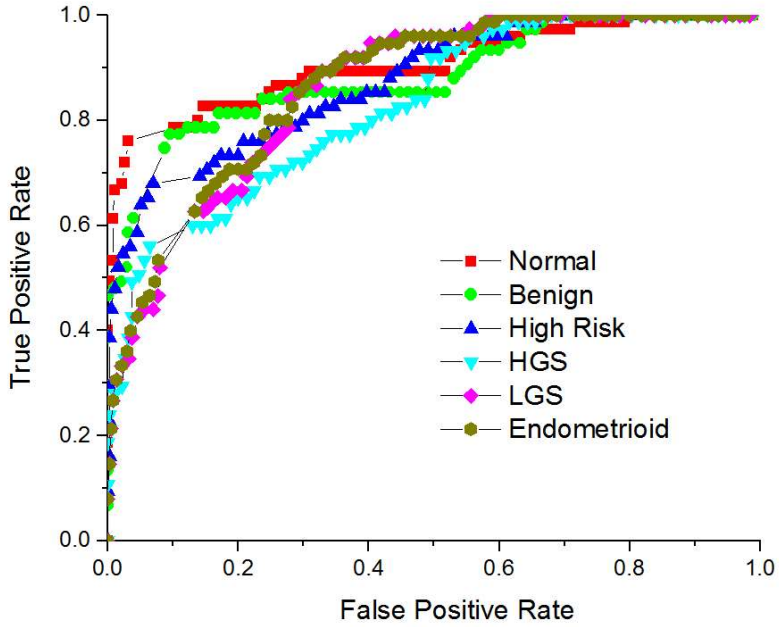


Figure 4.5. Receiver operating characteristic curves for 3-D classification with normal ovarian tissue 90.6% accuracy (red squares), benign tumor ovarian tissue 88.3% accuracy (light green circles), high risk ovarian tissue 87.5% (blue triangles), high grade ovarian tissue 83.0% (turquoise triangles), low grade ovarian tissue 86.3% (pink diamonds) and endometrioid tissue 86.8% (dark green circles).

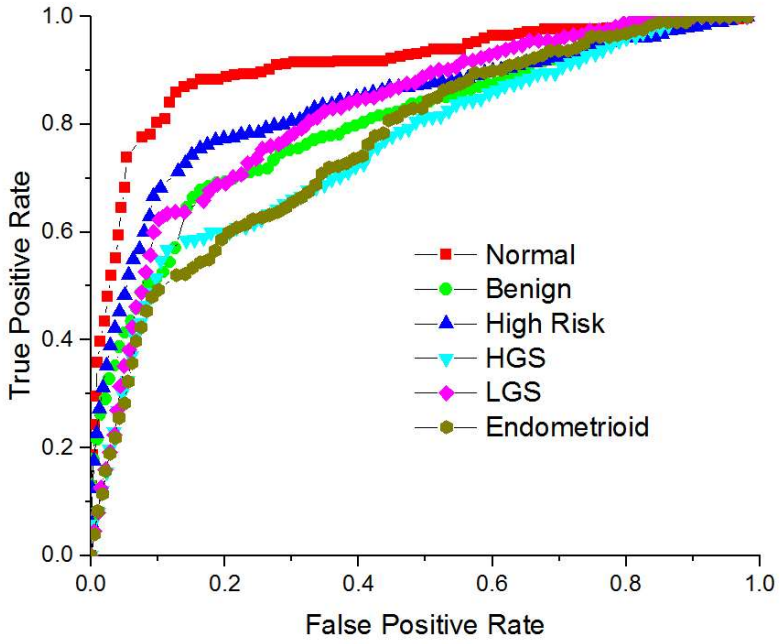


Figure 4.6. Receiver operating characteristic curves for 2-D classification with normal ovarian tissue 89.8% accuracy (red squares), benign tumor ovarian tissue 81.0 % accuracy (light green circles), high risk ovarian tissue 84.6% (blue triangles), high grade ovarian tissue 76.5% (turquoise triangles), low grade ovarian tissue 82.1% (pink diamonds), and endometrioid tissue 76.8% (dark green circles).

4.5. Discussion

Structural Significance

The overall relatively poor performance of standard screening/imaging tools [5,24] suggests that early detection of ovarian cancer demands new technologies that have the resolution and the specificity to detect and monitor small lesions. We have investigated using 3-D texture analysis

of SHG images of the ECM across a spectrum of ovarian tissues as a step in this direction. Instead of extracting visually apparent features such as angular distribution, fiber length, or area covered, as has been more commonly done, we convolved 3-D image filters with 3-D image patches. This is an important distinction, as in real tissues it is often impossible to fully discretize all of the individual fibers. Moreover, not all ECM modifications have easily apparent changes in fiber architecture, and subtle variations require a sensitive computer vision algorithm for feature extraction. Such analysis can be, however, less sensitive by only considering 2-D morphology. We previously performed an analogous 2-D analysis where we limited the application to HGS and normal ovarian tissues, which have the most visually noticeable changes in fiber alignment. [21] The task is more difficult for the other tissue classes, where the changes are not as pronounced or have more heterogeneity. For example, in the case of high risk tissues, there is visible heterogeneity not only between patients, but even within the same tissue specimen. By contrast, the respective overall morphologies of HGS tumors and normal stroma are conserved between patients. Comparisons of the 2-D vs. 3-D results are summarized in Table 1, where we found significant improvement with increased dimension, especially for the high risk patients. From a diagnostic/prognostic perspective, this is the most important class, given the high probability of developing HGS cancer in the lifetime of women with BRCA mutations. The increase in accuracy with 2-D vs. 3-D texture analysis likely arises from two factors. First, the 3-D texture analysis includes 3-D information from the z-axis and provides a more complete picture of the ECM structure. Second, the 3-D stack is a combination of tens of 2-D images. Therefore, the image library itself is much more sensitive and tolerant to diversity and heterogeneity of the ECM over the imaged depth profile.

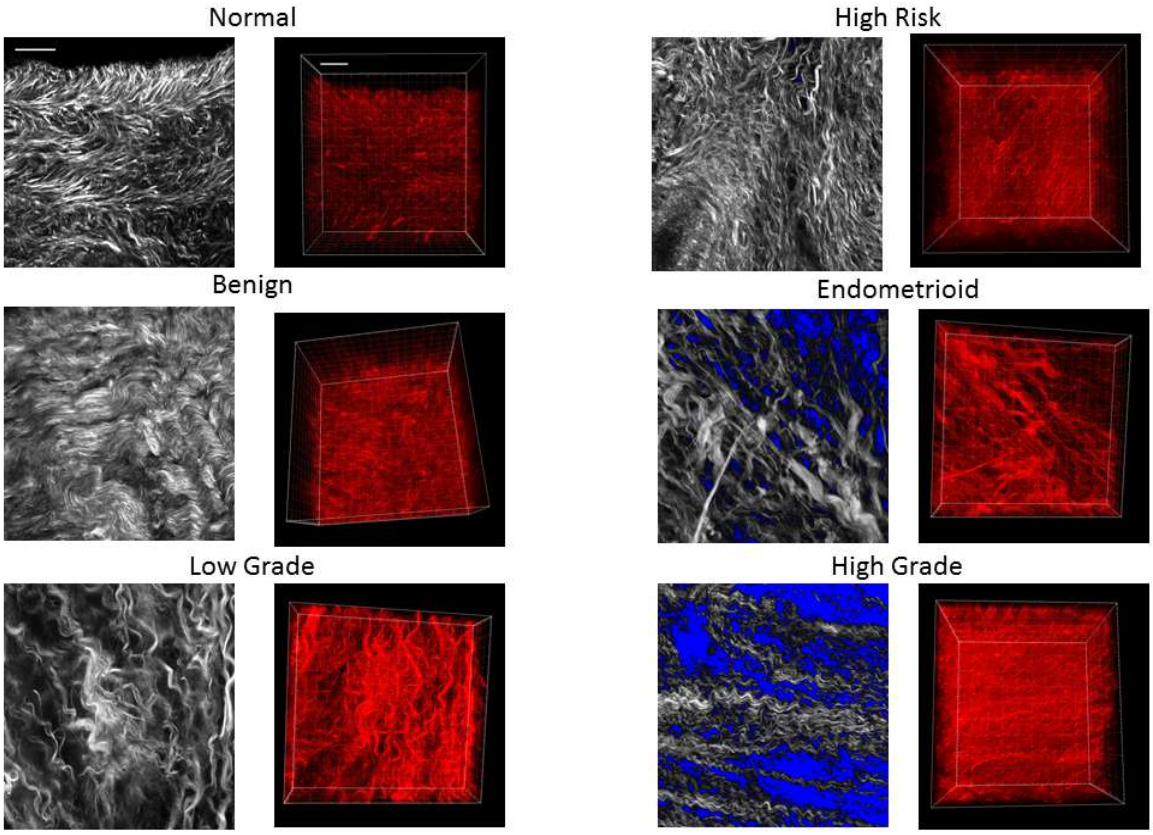


Figure 4.4. Representative SHG single optical sections (left) and 3-D renderings (right) of normal (a), high risk (b), benign tumor (c), endometrioid tumor (d), low grade serous (e) and high grade serous (f) human ovarian tissues. Scale bar = 30 μm.

We note that the overall accuracies achieved here for 2-D and 3-D are slightly lower than for the 2-D HGS vs normal in our previous work. This is a natural consequence of increasing the number of classes from 2 to 6, and because those tissues displayed the most substantial and uniform ECM changes. Still, our classification results are consistent with HGS, LGS and

endometrioid carcinomas being essentially distinct diseases based on genetic profiles, growth patterns, and chemotherapy responses. However, in the dualistic type I and II ovarian cancer classification scheme, LGS and endometrioid tumors are often grouped together,[25] whereas our classification clearly delineates them. This is also consistent with work from our lab using SHG creation physics and optical scattering to characterize the ECM changes in normal vs. HGS tissues. Importantly, none of our other metrics were able to differentiate high risk tissues due to the heterogeneity between and within specimens, whereas the 3-D texture analysis demonstrated here was successful in this task.

Texture Analysis Considerations

The 3-D textons used to build the histogram model for each image stack represent image features described by the whole filter bank. The number of image features is optimized based on that which provides the best representation of the image library. However, as we previously showed, using more textons may not provide higher accuracy as this can over-describe the class. [21] In other words, the histogram model gave inadequate statistical distributions when applying too many features to model the image stacks. On the other hand, too few textons will not provide sufficient discrimination. For these tissues, with our customized filter bank, we have found 40 textons to be optimal. Here we randomly selected 12,500 patches out of 512*512*60 voxels stack for the analysis so that the computational time is reasonable for image feature clustering. Then using 40 textons, 12,500 filter response vectors are needed to construct an effective statistical distribution model of the 3-D image set. To avoid artifacts, similar SHG image intensities are required in the analysis. As the SHG image intensity rapidly decreases even within

100 microns of stromal thickness, we restricted the analysis to the first 40-50 microns. This is not limiting as we have shown that stromal modifications mostly occur within the first 100-200 microns of thickness and are most pronounced near the tissue surface. [19]

Our distribution with the multi-class classification method is semi-balanced within the training set. We use *one-vs.-rest* multi-class classification, which is a strategy that involves training a single classifier per class, with the samples of that class deemed positive, and all other samples are negative. This strategy requires the base classifiers to produce a real-valued confidence score as a basis for the decision. Although our training set is semi-balanced across all six classes, the *one-vs.-rest* binary classification learning algorithm results in unbalanced distributions because the designated set of negative samples is typically much larger than the set of positives. Therefore, before decision making, we add a parameter weight and optimize it for each test image to acquire the best accuracy. The number of nearest neighbors in the classification algorithm was also optimized for the accuracy of the ROC curve.

Diagnostic Potential

ECM alterations are thought to be a critical step in the intiation and progression of many epithelial carcinomas [12, 27-29], and these are increasingly suggested as potential biomarkers.[18-20] In ovarian carcinogenesis and progression, changes in the reactive stroma can occur in the form of increased collagen concentration (i.e., desmoplasia), more ordered alignment of fibers, and changes in collagen isoform expression.[12] We have previously studied these changes across different size-scales using 3-D imaging in combination with the

measurement of bulk optical properties and Monte-Carlo simulations.[26] These studies revealed differences in collagen sub-resolution fibril and fiber architecture and best differentiate HGS tumors from the other classes, where the classification was moderately successful. While the texture analysis here does not provide the same detailed structural information, it results in equal if not better classification. Importantly, the analysis here only requires standard 3-D image data, whereas the other studies need directionally resolved SHG data and optical scattering measurements. Perhaps most importantly, the required SHG images could be acquired minimally invasively in conjunction with a standard ovarian laparoscope, where backward acquired image data would be sufficient. While this approach may not currently be feasibly used widely as a screening tool, it could be applied periodically to screen high risk patients with known BRCA or p53 mutations. Currently, the standard of care is to remove suspicious ovaries but this comes with considering the cost of quality of life as well as increased risk factors for other cancers.

4.6. Conclusions

We applied a new 3-D texture analysis algorithm to evaluate the ECM structural changes in normal ovarian stroma, high risk ovarian stroma, benign ovarian tumor, as well as high and low grade ovarian serous cancer tissues observed in 3-D SHG imaging stacks. By optimizing the number of textons, training image stack weighting, and nearest neighbor number, we achieved high accuracies between ~83-91% between classes, which greatly outperformed the analogous 2-D version. This successful application demonstrates the power of quantitative computer vision evaluation of 3-D SHG image features as a potential biomarker for cancer stage/type evaluation.

Importantly, it does not rely on extracting simple fiber characteristics such as size and alignment. This classification algorithm is a general method based on pre-trained SHG images and is well suited for analysis of rapidly changing fibrillar features for different kinds of tissues**.** It must be noted, the hand crafted filters used here may not necessarily be the best ones for image feature extraction for other cancer types or other disease states. Also, the effectiveness of the algorithm depends on the diversity of the acquired image library. With the coming era of big data and personalized medicine, this image analysis may assist pathologists in diagnoses as well as surgeons and oncologists with treatment decisions.

4.7 References

1 American Cancer Society, "Cancer Facts and Figures 2013," (2016).

2 Skates, S., Troiano, R. & Knapp, R. C. Longitudinal CA125 detection of sporadic papillary serous carcinoma of the peritoneum. *Int J Gynecol Cancer* **13**, 693-696, doi:13050 [pii] (2003).

3 Skates, S. J. *et al.* Calculation of the risk of ovarian cancer from serial CA-125 values for preclinical detection in postmenopausal women. *J Clin Oncol* **21**, 206s-210s, doi:10.1200/JCO.2003.02.955 [pii] (2003).

4 van Nagell, J. R., Jr. *et al.* The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. *Gynecol Oncol* **77**, 350-356, doi:10.1006/gyno.2000.5816S0090-8258(00)95816-5 [pii] (2000).

5 Iyer, V. R. & Lee, S. I. MRI, CT, and PET/CT for ovarian cancer detection and adnexal lesion characterization. *AJR Am J Roentgenol* **194**, 311-321, doi:10.2214/AJR.09.3522 194/2/311 [pii] (2010).

6 Bristow, R. E. *et al.* Combined PET/CT for detecting recurrent ovarian cancer limited to retroperitoneal lymph nodes. *Gynecol Oncol* **99**, 294-300 (2005).

7 Nakamoto, Y., Saga, T. & Fujii, S. Positron emission tomography application for gynecologic tumors. *Int J Gynecol Cancer* **15**, 701-709 (2005).

8 Qayyum, A. *et al.* Role of CT and MR imaging in predicting optimal cytoreduction of newly diagnosed primary epithelial ovarian cancer. *Gynecol Oncol* **96**, 301-306 (2005).

9 Pandit-Taskar, N. Oncologic imaging in gynecologic malignancies. *J Nucl Med* **46**, 1842-1850 (2005).

10 Bowtell, D. D. The genesis and evolution of high-grade serous ovarian cancer. *Nat Rev Cancer* **10**, 803-808, doi:10.1038/nrc2946 [pii] (2010).

11 Prat, J. New insights into ovarian cancer pathology. *Ann Oncol* **23 Suppl 10**, x111-117, doi:mds300 [pii]10.1093/annonc/mds300 (2012).

12 Ricciardelli, C. & Rodgers, R. J. Extracellular matrix of ovarian tumors. *Semin Reprod Med* **24**, 270-282 (2006).

13 Sobel, G. *et al.* Changes of cell adhesion and extracellular matrix (ECM) components in cervical intraepithelial neoplasia. *Pathol Oncol Res* **11**, 26-31 (2005).

14 Alowami, S., Troup, S., Al-Haddad, S., Kirkpatrick, I. & Watson, P. H. Mammographic density is related to stroma and stromal proteoglycan expression. *Breast Cancer Res* **5**, R129-135 (2003).

15 Cunha, G. R., Hayward, S. W. & Wang, Y. Z. Role of stroma in carcinogenesis of the prostate. *Differentiation* **70**, 473-485 (2002).

16 Blumenthal, R. D., Hansen, H. J. & Goldenberg, D. M. Inhibition of adhesion, invasion, and metastasis by antibodies targeting CEACAM6 (NCA-90) and CEACAM5 (Carcinoembryonic Antigen). *Cancer Res* **65**, 8809-8817 (2005).

17 Provenzano, P. P. *et al.* Collagen density promotes mammary tumor initiation and progression. *BMC Med* **6**, 11 (2008).

18 Conklin, M. W. *et al.* Aligned collagen is a prognostic signature for survival in human breast carcinoma. *Am J Pathol* **178**, 1221-1232, doi:S0002-9440(10)00233-6 [pii] 10.1016/j.ajpath.2010.11.076 (2011).

19 Nadiarnykh, O., Lacomb, R. B., Brewer, M. A. & Campagnola, P. J. Alterations of the extracellular matrix in ovarian cancer studied by Second Harmonic Generation imaging microscopy. *BMC Cancer* **10**, 94 (2010).

20 Adur, J. *et al.* Second harmonic generation microscopy as a powerful diagnostic imaging modality for human ovarian cancer. *J Biophotonics* **7**, 37-48, doi:10.1002/jbio.201200108 (2014).

21 Wen, B. L. *et al.* Texture analysis applied to second harmonic generation image data for ovarian cancer classification. *J Biomed Opt* **19**, 096007, doi:10.1117/1.JBO.19.9.0960071906043 [pii] (2014).

22 Varma, M. & Zisserman, A. A statistical approach to texture classification from single images. *Int J Comput Vision* **62**, 61-81, doi:DOI 10.1007/s11263-005-4635-4 (2005).

23 Chen, X., Nadiarynkh, O., Plotnikov, S. & Campagnola, P. J. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nat Protoc* **7**, 654-669, doi:10.1038/nprot.2012.009 nprot.2012.009 [pii] (2012).

24 Vikram, R., Yeung, H. D., Macapinlac, H. A. & Iyer, R. B. Utility of PET/CT in differentiating benign from malignant adrenal nodules in patients with cancer. *AJR Am J Roentgenol* **191**, 1545-1551, doi:10.2214/AJR.07.3447191/5/1545 [pii] (2008).

25 Prat, J. Ovarian carcinomas, including secondary tumors: diagnostically challenging areas. *Mod Pathol* **18 Suppl 2**, S99-111 (2005).

26 Tilbury, K. B., Campbell, K. R., Eliceiri, K. W., Patankar, M. & Campagnola, P. J. Analysis of stromal alterations in ovarian cancers via wavelength dependent Second Harmonic Generation microscopy and optical scattering measurements. *PNAS* (submitted).

27 Pupa, S. M., Menard, S., Forti, S. & Tagliabue, E. New insights into the role of extracellular matrix during tumor onset and progression. *Journal of Cellular Physiology* **192**, 259-267 (2002).

28 Landen, C. N., Jr., Birrer, M. J. & Sood, A. K. Early events in the pathogenesis of epithelial ovarian cancer. *J Clin Oncol* **26**, 995-1005 (2008).

29 Theret, N. *et al.* Increased extracellular matrix remodeling is associated with tumor progression in human hepatocellular carcinomas. *Hepatology* **34**, 82-88 (2001).

30 http://www.pathologyoutlines.com/topic/ovarytumorserouscarcinoma.html

Chapter 5: Multi-view Second Harmonic Generation Imaging

5.1 Forward

In this chapter we discussed on the development of a multi-view second-harmonic generation imaging system and the computational challenges for the later imaging processing such as registration and reconstruction. We first experimentally show that SHG imaging is not sensitive to collagen fibers oriented parallel to the direction of laser propagation and, as a consequence, can potentially miss important structural information. As an alternative approach, we demonstrate the use of reflective micro-prisms to enable multi- view SHG imaging of mouse tail tendon by redirecting the focused excitation and collection of subsequent emission. Our approach data corroborates the theoretical treatment on vanishing and non-vanishing orientations, where fibers along the laser direction are largely transparent by SHG. In strong contrast, the two-photon excited fluorescence of dye-labeled collagen fibers is isotropic and is not subject to this constraint. We utilized Pearson correlation to quantify differences in fluorescent and backward detected SHG images of the tendon fiber structure, where the SHG and TPEF were highly statistically correlated (0.6--0.8) for perpendicular excitation but were uncorrelated for excitation parallel to the fiber axis. The results suggest that improved imaging of 3D collagen structure is possible with multi-view SHG microscopy. Further we conceptualized and constructed several iterations of a commercial system compatible multi-view SHG imaging platform. We acquired detailed structures of mouse tail tendon and ovarian tissues using the

multi-view SHG platform. Further we discussed some of the imaging analysis framework for data registration and reconstruction. In the end we covered the further challenge and direction for future research.

5.2 Introduction

Second-harmonic generation (SHG) imaging has emerged as a powerful modality for visualizing the collagen assembly in a wide range of normal and diseased tissue types [1, 2]. Applications for imaging structural changes in many pathologic conditions, including cancers [3–5], fibroses [6,7], and connective tissue disorders [8] have received considerable attention, as changes in the collagen rich ECM are often revealed by SHG imaging via changes in fibrillar morphology, intensity, and polarization properties. A limiting aspect of SHG imaging is that it is not a true 3D technique. Specifically, while 3D data is built up from stacking a series of 2D *en face* images, due to the electric dipole interaction, fully axially oriented fibers (i.e., along the laser direction) are transparent. This phenomenon is not commonly seen in fluorescence imaging as probe molecules (either dyes or fluorescent proteins) typically have rotational freedom, and absorb at all angles. The most notable exception is imaging of membrane staining dyes (e.g., DiI or ANEPPS), where a “ring stain” is often observed due to the rotational constraints of the dye molecules being bound in the membrane. The endogenous SHG contrast from collagen molecules within fibrils has these same constraints. This situation results in a loss of information in determining the structure of 3D ECM.

A solution to this problem is to acquire SHG images from different directions of the excitation laser relative to the fixed specimen. We note that this is distinct from probing different structural

aspects via performing polarization analysis from the same direction of laser propagation [9]. Here we demonstrate SHG microscopy with different iteration of multi-view SHG imaging system to image the collagen fiber structure in mouse tail tendon and bio-tissue to tailor true 3D visualization of the collagen fibers within a matrix. The micro prisms have previously been used for several other microscopy applications [10–13]. Here we utilize the micro-prisms to excite and collect the backward directed SHG from different views. Backward detected SHG is comprised of a mixture of the emitted signal and subsequent scattering at the SHG wavelength. In tendon, the emitted directionality, which we have denoted F_{SHG}/B_{SHG}, is ~7:1 [14, 15], but given the strength of the absolute intensity, this is more than sufficient for imaging.

5.3 Theoretical Calculation

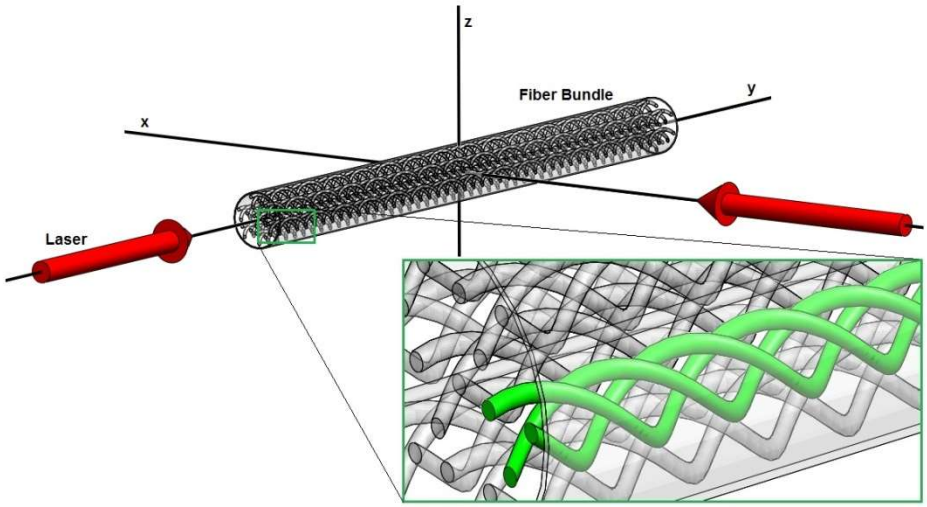


Fig. 5.1. End view (y axis) and side view (x axis) of laser propagation direction relative to the fiber axis (y axis).

Collagen I fibers consist of a complex hierarchal assembly, as shown in Fig. 5.1. First, individual triple helical collagen molecules are covalently linked to into fibrils with diameters ranging from 20 to 200 nm, depending on the tissue. Fibrils are further organized into fibers, where the latter are the quantity visualized in the SHG microscope. Additionally, fibers can crimp, causing the components of the bundle to appear offset from the long axis. The induced polarization of a medium subjected to an in- tense electromagnetic field can be related in a power series of the field strength Ei (i; j; k are Cartesian components) by the expression

$$P_i = \varepsilon_0 \chi_{ij}^{(1)} E_j + \varepsilon_0 \chi_{ijk}^{(2)} E_j E_k + \varepsilon_0 \chi_{ijkl}^{(3)} E_j E_k E_{l+\cdots} \quad (1)$$

where P_i is the *i*th component of the induced polarization, and ε_0 is the vacuum permittivity, $\chi_{ij}^{(n)}$ denotes the *n*th order susceptibility and is a tensor of rank corresponding to the number of subscripts. For example, $\chi_{ijk}^{(2)}$ can be expressed by the third-rank d-tensor given by $d_{ijk}=\chi_{ijk}^{(2)}/2$, and the effective d-value is written as $d_{eff} = \hat{\mathbf{e}}d$, where ê is a unit vector describing the electric field or polarization field of the light wave. The tensor related to SHG, $\chi^{(2)}$, reflects the symmetry and nonlinear optical properties of the material and is the quantity visualized in the microscope.

For the cylindrical structure of collagen fibrils (C_∞ symmetry), the most general vector expression for the polarization dependence of second harmonic generation is: [16]

$$d_{eff} = [\hat{e}_1\hat{e}_2\hat{e}_3]\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & d_{16} \\ d_{21} & d_{22} & d_{23} & 0 & 0 & 0 \\ 0 & 0 & 0 & d_{34} & 0 & 0 \end{bmatrix}\begin{bmatrix} \hat{e}_1^2 \\ \hat{e}_2^2 \\ \hat{e}_3^2 \\ 2\hat{e}_2\hat{e}_3 \\ 2\hat{e}_3\hat{e}_1 \\ 2\hat{e}_1\hat{e}_2 \end{bmatrix} \quad (2)$$

where the coordinate system of the laser electric field or the polarization field of the light wave is related to the collagen fiber by the unit vectors $\hat{e}_1$, $\hat{e}_2$ and $\hat{e}_3$.

In accordance with previous approaches, collagen is assumed to have $C_\infty mm$ symmetry (x=z) along the fiber axis (the y-axis in Fig. 1). [17,18] Cylindrical symmetry (x=z) implies that $d_{16}=d_{34}$ and $d_{21}=d_{23}$, and Kleinman symmetry gives $d_{16}=d_{21}$. There for we can assume $d=d_{16}=d_{34}= d_{21}=d_{23}$. For circular polarization, the electric field of the laser along the collagen fiber can be described by:

$$(\hat{e}_1,\ \hat{e}_2, \hat{e}_3)\ =\ (sin(\omega t), 0, cos(\omega t)) \quad (3)$$

After inserting **ê,** equation (2) becomes:

$$d_{eff} = 2d_{16}\, sin^2(\omega t)\ \hat{e}_2 \quad (4)$$

since here $\hat{e}_2 = \emptyset$, $d_{eff}=\emptyset$,it is seen that there will not be any SHG emission in this case. This is rigorously true only if all of the collagen molecule dipole moments align along exactly along the direction of the fiber axis. However, previous studies suggested that the collagen molecules may have a small tilt angle with respect to the fibril axis,[19] where in this more realistic case, the SHG will be greatly diminished but not completely extinguished. By contrast, if the excitation laser propagates along the z axis as (sin(ω*t), cos(ω*t), 0), the following expression would apply:

$$d_{eff} = [d_{22} \cos^2(\omega t)\, \hat{e}_2 + 2d \sin^2(\omega t)]\, \cos(\omega t)\, \mathbf{y} \qquad (5)$$

(**y** here is unit vector in y direction) in which case only y axis polarization components will result in non-vanishing SHG emission.

We can experimentally verify these suppositions and take steps toward 3D imaging by imaging mouse tail tendon from different orthogonal views. Tendon is ideal for this purpose due to the regularity of the fibril/fiber structure. In this experiment, we arrange two 1 mm micro-prisms (Precision Optical, Costa, Mesa, CA,) at orthogonal vantage points of the tendon. One micro-prism is placed facing the end (endview) of the mouse tail tendon and one is placed on the side (sideview) as shown in figure 2. The excitation beam path changes direction on the mirrored face of the prism and propagates into different sides of the tendon. Using a 40× 0.8 NA water immersion lens (Olympus), we achieved a penetration depth of about 160 um on both the side and end views. Images are collected in the backward SHG geometry.

5.4 Multi-View SHG Imaging of Mouse Tail Tendon via Reflective Micro-Prisms

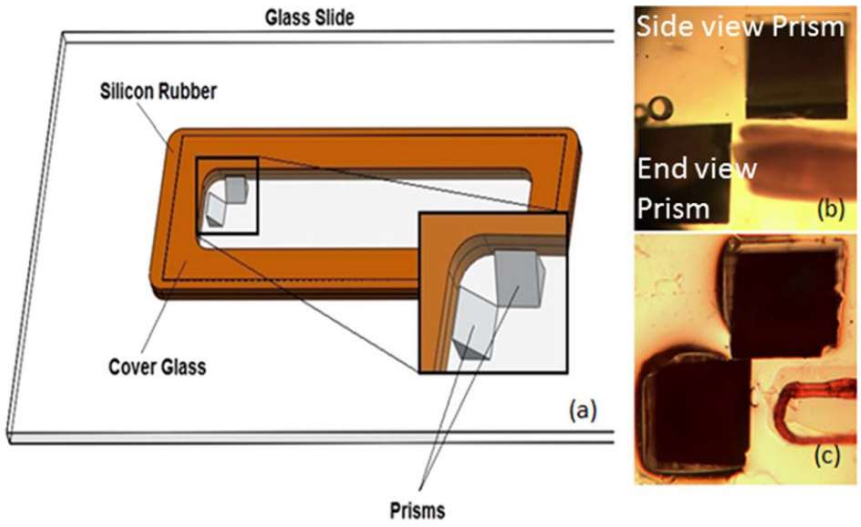


Fig. 5.2. (a) Experiment setup showing the layout of the micro-prisms in the chamber. (b) Straight tendon in bright field with a 10× objec-tive. (c) Bent u-shaped tendon in bright field with the same objective.

Correlation analysis was used to statistically evaluate the similarity of TPEF and SHG images on a pixel by pixel basis for different views of the tendon fibers. The Pearson coefficient was determined over the entire 100-160 μm stacks in 1 μm step sizes of the two imaging modalities for both the side and endviews. When the two image stacks are perfectly similar, the Pearson coefficient is 1.0 and becomes 0.0 when no correlation exists.

The sideview images for the straight tendon remain similar (column c). However, this configuration shows highly significant differences in the SHG (row 2) and TPEF (row 1) endview images (column d) Fiber and fiber bundles sticking out of the focus plane and also the tip of the tendon where it was initially cut can be clearly observed in the fluorescent images. As

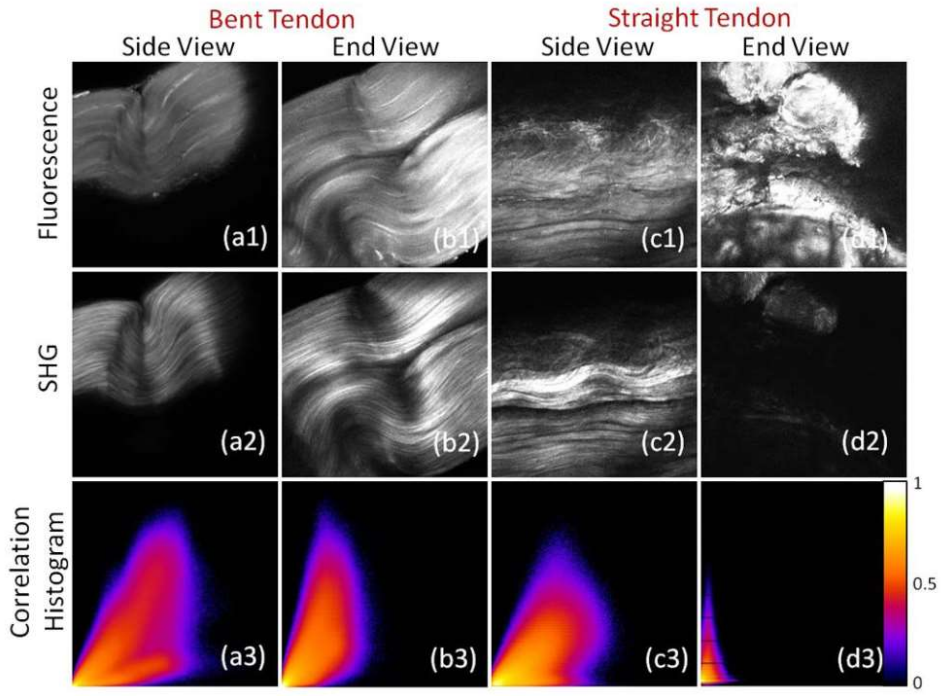


Figure 5.3. Fluorescence image (a1) SHG image (a2) and 2D intensity histogram (a3) for fluorescence and SHG image from side view for bent tendon; fluorescence image (b1) SHG image (b2) and 2D intensity histogram (b3) for fluorescence and SHG image from the endview for bent tendon; fluorescence image (c1) SHG image (c2) and 2D intensity histogram (c3) for fluorescence and SHG image from side view for straight mouse tendon; fluorescence image (d1) SHG image (d2) and 2D intensity histogram (d3) for fluorescence and SHG image from end view for straight mouse tendon. The x and y axis for 2D intensity histograms correspond to normalized SHG and TPEF image pixel intensities, respectively. The heatmap is normalized from highest

pixel frequency to lowest in 256 color bins. Scale bar=40 microns. Predicted by Eq. 4, there is essentially no SHG contrast from these fibers as they are aligned parallel to the excitation laser beam (see Fig. 5.1) where only the boundary of each fiber bundle can be discerned with much lower comparative intensity. Furthermore, when examining axial sections through the fiber, this low signal quickly vanishes away from the boundary and may be the result of the cutting process. These imaging results qualitatively agree with our theoretical predictions of the diminished SHG emission when the collagen fibers align with the propagation direction of the excitation laser.

For the u-shaped mouse tail tendon orientation, the obtained Pearson coefficients (after thresholding) for the front and side views were 0.77 and 0.81, respectively and are statistically correlated. Note that complete correlation is not expected due to the crimped regions, where fluorescence will occur but SHG will be largely extinguished as these fibers will be parallel to the laser propagation. For the straight tendon arrangement, the Pearson coefficient is 0.60 from the side view but decreased significantly to -0.08 indicating very little correlation between the TPEF and SHG image stacks. To view these correlations (and lack thereof) graphically, Row 3 of Figure 3 shows the overlap of the greyscale intensity histograms of the respective SHG and TPEF distributions, where for perfect correlation, all points would fall on the diagonal. The observed greatly diminished SHG intensity and lack of correlation with the TPEF for the endview of the straight tendon configuration agree with the theoretical prediction that little SHG emission will result from collagen fibers that lie parallel to the direction of laser propagation.

While the emission angles for SHG from tissues are not known as they are not perfectly phasematched based on phasematching arguments, little SHG is expected in the orthogonal direction to the excitation.[21] Moreover, scattering from the orthogonal direction will be minimal, as the high scattering anisotropy of tendon (~0.95) leads to essentially all forward directed scatter.[14] Thus the SHG collected in the backward direction from the reflective prims is highly similar to that of *en face* imaging. We note that it is possible in principle to achieve this result through tilting the specimen relative to the laser direction, obtaining multiple views through two micro-prisms is superior due to design considerations.

In summary, by comparing fluorescence and SHG images, we demonstrate some of the limitations of the SHG method for true 3D imaging of collagen. While we have not presented new theory, validation of these visualization constraints of SHG imaging require the appropriate arrangement for multi-view imaging. This approach can be implemented for a range of SHG investigations. This is important as there is great promise in understanding 3D collagen structure, as the multi-view approach here coupled with directional SHG (forward– backward), and polarization-resolved measurements could re- veal new insight in the role of collagen in normal and diseased biological processes.

5.5 3D Imaging Platform Development

While the reflective micro prisms demonstrated the ability for imaging bio-tissue in different perspective, we further conceptualize a Multiview SHG platform using micro prism for imaging bio-tissue from different directions. In the following part, the chapter will go over some of the

prototype iterations for Multiview SHG imaging platform and further applications with mouse tail tendon and ovarian tissue. The main motivations for the design relative to current imaging method are to, firstly, allows anisotropic multi-photon imaging of biological tissues by acquiring 3D images of the same sample from multiple views. Further, the advantage for the platform is to match the axial resolution with the lateral resolution, since the current imaging technique the lateral resolution is much lower than the axial resolution due to laser focus profile. Secondly, collagen fibers perfectly aligned with the excitation laser cannot generate any SHG emission signal because of the dipole interaction physics in multiphoton imaging.

5.5.1 Device Development

The development of the 3D imaging platform so far includes three main iterations. The earliest version of the Multiview SHG platform is demonstrated in Fig 5.5, in which we applied the reflective micro-prisms for imaging tissue from different perspective. The initial idea is that we mount the reflective micro-prim on the rotational stage. The excitation laser would be redirected by the micro-prim and sent into the tissue. The sample would be mounted in the middle, and the reflective micro prim would redirect the excitation laser into tissue from different perspectives. Therefore, we could acquire images the ECM structure of biomedical tissue from multiple angles.

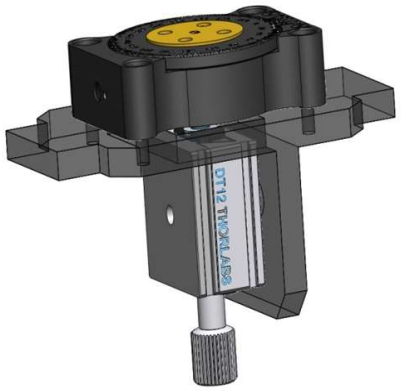


Figure 5.5 Multi-view SHG imaging platform first iteration with rotational stage

The platform was setup with rotation stage from THORLAB, in which can rotate the micro reflective prisms. However, this method will introduce unnecessary motion by hand in the measurement, which will be significant in the registration with image stacks of several hundreds of microns in size.

Further, we applied the motorized rotational stage to reduce the hand motion from stage in Fig 5.6. We applied rotational stage for sample mounting and rotated the sample besides the micro-prism so that we can sequentially image the 3D stack of sample with redirected scanning laser from different directions. The whole sample is in a water emersion chamber which is much like the sample preparation chamber in Fig 5.2. The whole platform was designed so that it is compatible for the PRIOR transitional stage which can provide transitional adjustment.

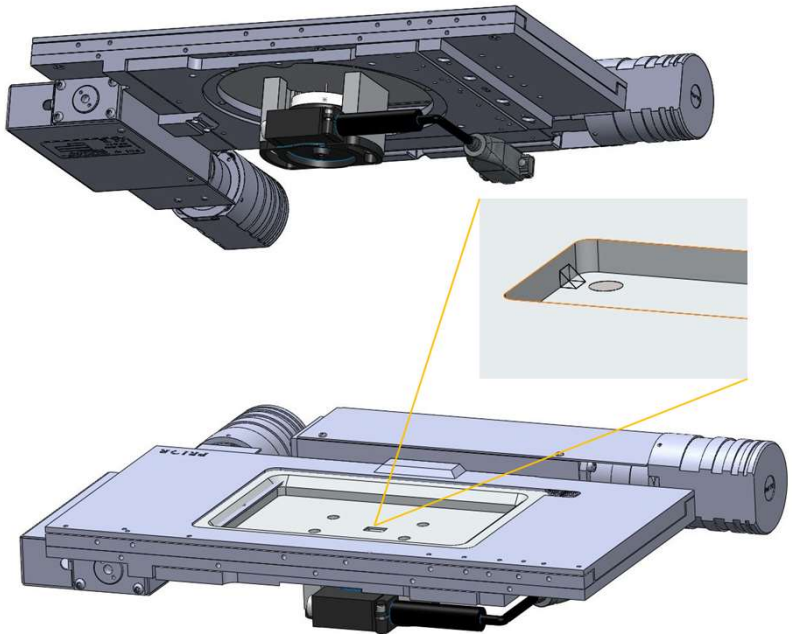


Figure 5.6 Multi-view second harmonic generation imaging platform second iteration

There are still a couple of issues for this iteration. Firstly, the sample preparation is difficult since it would be hard to keep the sample in place on mounting stage. Considering that the size of the prism is 1 mm, the preparation for the sample would be limited in a very small range. Also this platform would only provide limited numerical aperture and penetration depth with different objectives considering the geometry of the prism. We did a detailed analysis with relationship between tissue imaging penetration depth of different objectives and Numerical Aperture (N.A.) conservation in Table 5.1 in experimental setup from Fig. 5.7.

Table 5.1 Working distance, numerical aperture and penetration depth with different objectives

Lens (magnification/ numerical aperture/ media)	Working Distance L(μm)	Penetration Depth D(μm)	Focus Angle Θ(degree)
10X/0.5/Air	1900	366	30
20X/0.5/Water	3300	737	22
40X/0.8/Water	3300	160	37

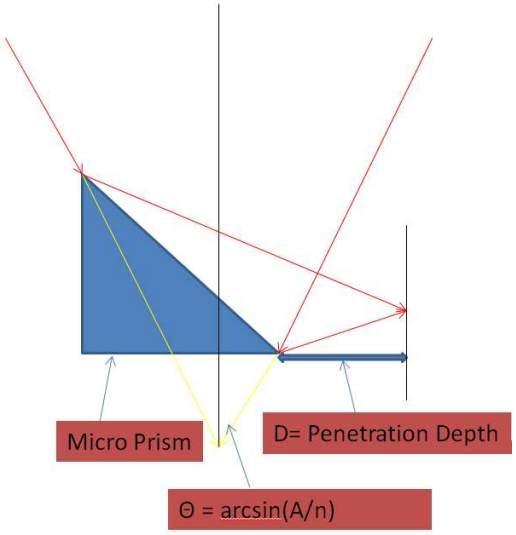
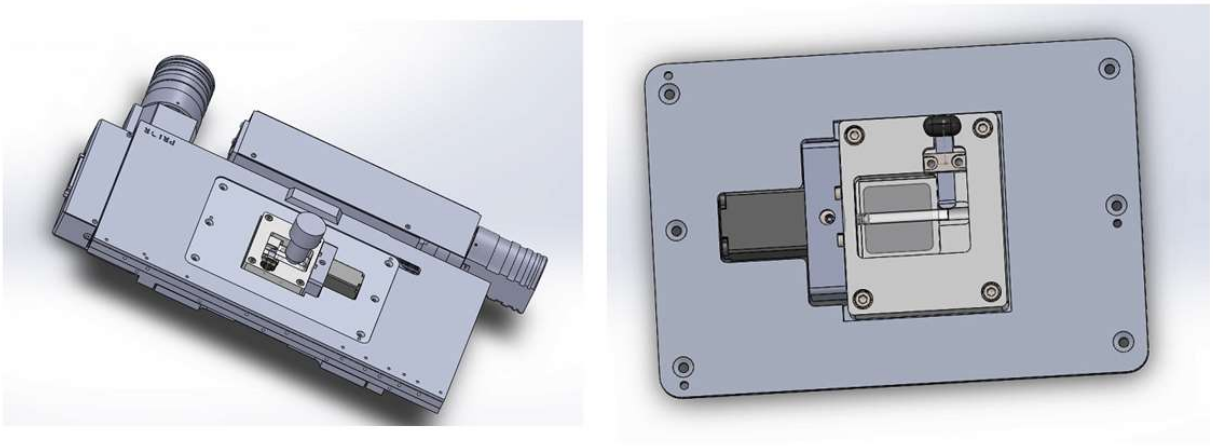


Figure 5.7 Multi-view second harmonic generation imaging platform second iteration

The limitations of tissue penetration depth and imaging NA make it impossible for acquiring high fidelity 3D image of the thick biological tissues. Therefore, we further move on for developing a Selective Plan Illumination Microscopy (SPIM) [] inspired 3D imaging platform. The normal SPIM system is an open platform microscopy setup and is not very effective for

SHG imaging due to its low signal-noise-ratio. We came up with the similar idea and designed a SPIM inspired inexpensive and commercially compatible imaging platform system. As the imaging platform setup shown in Fig 5.8, we embed the sample in agarose in the FEP tubing; the motor will drive the rotation of FEP tubing with sample mounted; therefore, the scanning system would be able to image the sample from different perspectives. We will discuss about the detailed calibration and sample preparation later.



(a) (b)

Figure 5.8 Multi-view SHG imaging platform third iteration with the transitional stage (a) and without the transitional stage (b)

5.5.2 Platform Calibration and Sample Preparation

Calibration: The motor we were using for FEP tubing rotation is the nema8 stepper motor controlled by the LEADERSHINE software. Before imaging it is important to calibrate the rotation angle to ensure that we are rotating sample with good accuracy for the latter registration. After the calibration, the smallest accurate stepping size is 36 degree (0.1 rounds) is. However,

the step size set smaller than 36 degrees can cause error more than 5 degrees and the registration would be refined by such rotation angle. Further development will require a better motor. Also when doing imaging, the center of the sample should be aligned with the objective.

Sample preparation: We prepare the sample in the FEP tubing since the refractive index is the same as the water. First we heat the agarose in the water bath and then fill the FEP tubing with sample. Also in this imaging application it is better to use water immersion objective since it would provide long working distance for thick tissue imaging and refractive index matching.

Since the 3D imaging system will suffer from optical scattering and absorption in tissue; the image quality will decrease as the optical path length as the sample thickness increases. Therefore, we need to optical clear the sample before imaging. Here we have utilized 50% glycerol as clearing agent, as previously done by the Campagnola lab.

Lastly, we must consider the appropriate size of the sample. The optimal size depends on the objective magnification or field of view since we need overlap the different 3D stacks from different perspectives to reconstruct as a 3D image. Therefore different objects should be matched with different FEP tubing size. Within the field of view, the more perspective overlap with each other for the sample would result in better visualization after construction.

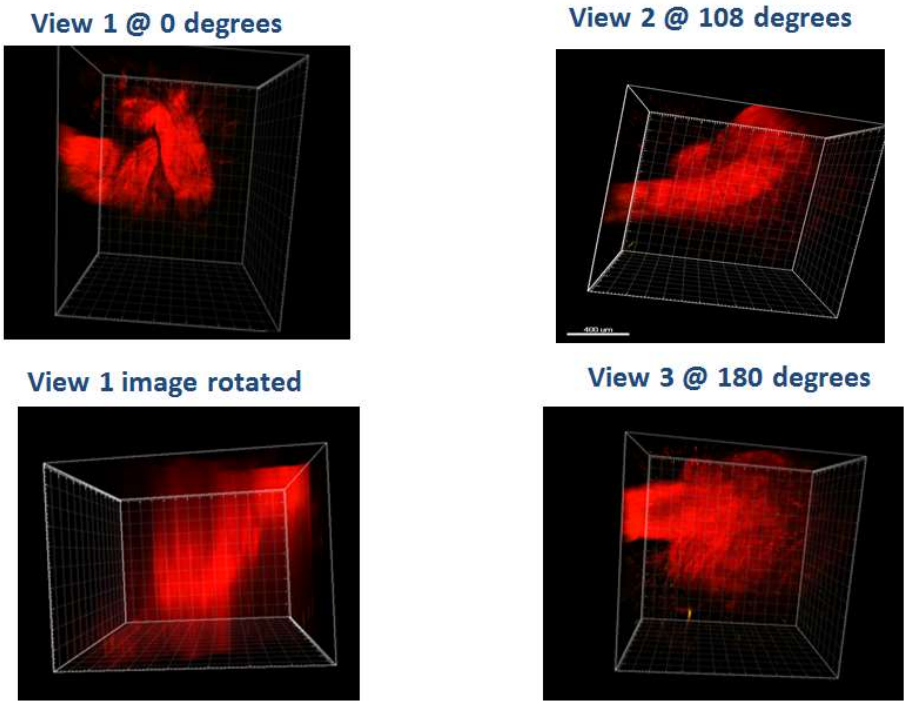


Figure 5.9 Multi-view second harmonic generation imaging platform third iteration

5.5.3 Image Registration and Reconstruction

To realize the full potential of multi-view imaging, it is necessary to reconstruct a single 3D image from the individual views to image collagen structure with high fidelity by SHG imaging. Although we have not fully completed the development of the algorithm, we will go through some of the basics of the image registration and reconstruction.

Registration:

Multi-view registration is inspired by work from Stephan Preibisch's work [22]. Multi-view SHG imaging is complicated by degradation of the signal along the illumination axes, limited

overlap between the views, different orientations of the optical sections and anisotropic SHG signal. We are currently developing a registration method which enables efficient, sample-independent registration. Our 3D imaging platform has two channels of output. We can switch from SHG imaging to multi-photon fluorescence imaging by changing the wavelength filter. Fluorescent beads are embedded in rigid mounting medium around the biological tissue, and they serve as fiduciary markers multi-photon fluorescence acquisitions mode for SHG imaging. Therefore, the registration problem for multi-view SHG imaging is reduced to the matching of point clouds.

We first detected the beads using a difference of Gaussian 3D filter. To efficiently identify corresponding beads in different views, we utilize a translation and rotation invariant local geometric descriptor that identifies each bead by the unique constellation of its neighboring beads. This constellation is preserved across views transformed by rotation and translation in three dimensions. For efficient matching, we define an orthogonal local coordinate system in each descriptor, expressing the 3D constellation of four beads by a vector of six scalar values, achieving translation and rotation invariance. Similar descriptors in different views have a small Euclidean distance in the six-dimensional descriptor space, and for efficient identification of nearest neighbors we presorted the six-dimensional scalar vectors using a hierarchical tree-based algorithm to reduce the matching problem to logarithmic complexity. Constellations of four beads that accidentally look similar are rejected using the random sample consensus with an affine transformation model. [22]

Reconstruction: In the final step of the registration framework, we globally minimize the displacement of all true correspondences identified in all pairs of views using an iterative optimization scheme resulting in an affine transformation model for each view. Typically, we will identify thousands of corresponding bead descriptors equally distributed around the imaged sample. The global optimization converge within to a final average displacement of a number of pixels below threshold. The average bead displacement and the ratio between correspondence candidates and true correspondences is a quantitative measure of the reconstruction success.

We combine content-based fusion with nonlinear blending to compensate for brightness differences at boundaries between views. The reconstructed multi-view acquisition in contrast to the single view would be comparable lateral and axial resolution. In the middle of the specimen, the resolution is lower since the scattering of the sample limits light penetration. [23]

There are still work to be done for the 3D imaging in the algorithm considering the polarization response and phase-matching for the tissue with different alignment between collagen fiber and excitation laser. There is strong correspondence between excitation laser direction and collagen fiber alignment. For a better morphology visualization, not only need we understand how is whether there is any collagen fiber at the space but also how the contrast is introduced. We will discuss more about it in Chapter 6.

5.6 References

1. P. J. Campagnola and C. Y. Dong, Laser Photon. Rev. 5, 13 (2011).

2. P. Campagnola, Anal. Chem. 83, 3224 (2011).

3. P. P. Provenzano, K. W. Eliceiri, J. M. Campbell, D. R. Inman, J. G. White, and P. J. Keely, BMC Med. 4, 38 (2006).

4. R. Cicchi, D. Massi, S. Sestini, P. Carli, V. De Giorgi, T. Lotti, and F. S. Pavone, Opt. Express 15, 10135 (2007).

5. E. Brown, T. McKee, E. diTomaso, A. Pluen, B. Seed, Y. Boucher, and R. K. Jain, Nat. Med. 9, 796 (2003).

6. A. M. Pena, A. Fabre, D. Debarre, J. Marchal-Somme, B. Crestani, J. L. Martin, E. Beaurepaire, and M. C. Schanne-Klein, Microsc. Res. Tech. 70, 162 (2007).

7. W. Sun, S. Chang, D. C. Tai, N. Tan, G. Xiao, H. Tang, and H. Yu, J. Biomed. Opt. 13, 064010 (2008).

8. R. Lacomb, O. Nadiarnykh, and P. J. Campagnola, Biophys. J. 94, 4504 (2008).

9. X. Chen, O. Nadiarynkh, S. Plotnikov, and P. J. Campagnola, Nat. Protocol 7, 654 (2012).

10. H. Hajjoul, J. Mathon, Y. Viero, and A. Bancaud, Appl. Phys. Lett. 98, 243701 (2011).

11. S. Lindek, T. Stefany, and E. H. K. Stelzer, J. Microsc. 188, 280 (1997).

12. K. T. Seale, R. S. Reiserer, D. A. Markov, I. A. Ges, C. Wright, C. Janetopoulos, and J. P. Wikswo, J. Microsc. 232, 1 (2008).

13. J. Yuan, R. J. Melder, R. K. Jain, and L. L. Munn, Biotechniques 30, 388 (2001).

14. R. LaComb, O. Nadiarnykh, S. Carey, and P. J. Campagnola, J. Biomed. Opt. 13, 021109 (2008).

15. R. Lacomb, O. Nadiarnykh, S. S. Townsend, and P. J. Campagnola, Opt. Commun. 281, 1823 (2008).

16. A. Erikson, J. Ortegren, T. Hompland, C. de Lange Davies, and M. Lindgren, J. Biomed. Opt. 12, 044002 (2007).

17. P. Stoller, B.-M. Kim, A. M. Rubinchik, K. M. Reiser, and L. B. Da Silva, J. Biomed. Opt. 7, 205 (2002).

18. A. E. Tuer, M. K. Akens, S. Krouglov, D. Sandkuijl, B. C. Wilson, C. M. Whyne, and V. Barzda, Biophys. J. 103, 2093 (2012).

19. O. Nadiarnykh, S. Plotnikov, W. A. Mohler, I. Kalajzic, D. Redford- Badwal, and P. J. Campagnola, J. Biomed. Opt. 12, 051805 (2007).

20. J. Mertz and L. Moreaux, Opt. Commun. 196, 325 (2001).

21. J. Huisken, J. Swoger, FD Bene, J. Wittbrodt, EK Stelzer, Science, Vol305, Issue 568613 August (2004)

22. Stephan Preibisch and others, *Nature Methods*, 7.6 (2010), 418–19

23. Stephan Preibisch, Fernando Amat, and others, *Nature Methods*, 11.6 (2014), 645–48.

Chapter 6: Conclusion and Future Directions

From previous chapters, we discussed how SHG imaging microscopy has great potential for imaging ECM morphology structure of biological tissues which can be used for disease classification. Moreover we conceptualized and developed 3D multi-view second harmonic generation microscopy system to improve the fidelity of current 3D imaging.

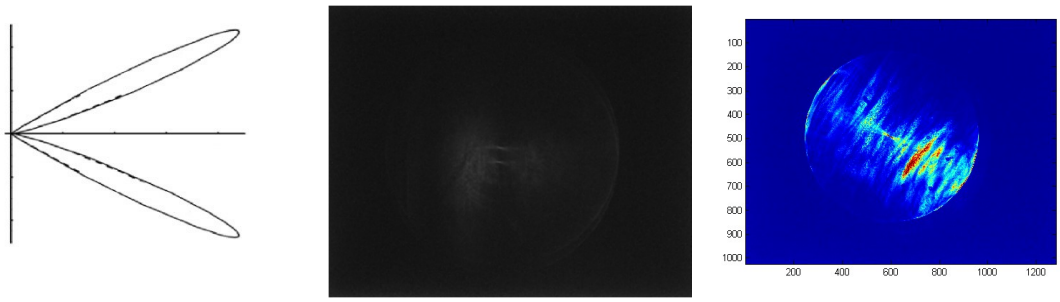
6.1 Future Directions for Multiview 3D SHG Platform

In addition to the morphology study, there are many other applications for the 3D SHG platform. Here I will discuss about future research directions for with respect to phase-matching, polarization measurement and open source SHG microscopy.

6.1.1 Spatial Phase-Matching

Researchers have generalized models describing the SHG by highly focused excitation laser excitation. The spatial frequency of the sample inhomogeneity is very important for the SHG radiation spatial patterns. Most of the work in our lab was has simplified such spatial pattern as forward and backward emission.[1,2] However, this is too simplified theory. Recent studies have model the thin membrane as double lobe shaping for the emission pattern. We have also used pupil plane imaging to acquire the same pattern for emission as demonstrated in Fig. 6.1. However, the phase-matching pattern is still unknown for a lot of more complicated biological

tissues. The difficulty lies in, first, how to model such physical emission patterns for more complicated system; and, second, how to measure 3D spatial emission pattern correctly. The rotational tubing setup of the imaging system makes the 3D emission measurement possible by excite the tissue sample from different perspective. In another aspect the direction of the detector might need also be able to change from different directions, which might further requires open platform SHG microscopy with customized configurations.



(a) (b) (c)

Figure 6.1 (a) double lobe structure (b) pupil plane imaging for membrane (c) pupil plane imaging for mouse tail tendon

6.1.2 Polarization Studies

In another aspect, current assumption for the phase matching studies have focused on mostly linear polarized with one direction for the membrane model. As we mentioned in introduction, a profound remodeling in ECM structure occurs in many epithelial chancers, and the minor isoform of collagen III and V up- or down-regulated during the cancer initiation and progression, which also contribute to the morphology alternation in ECM structure. Colleagues in our lab

have used a pixel-based polarization resolved approach to probe the net collagen α-helix pitch angle within the gel mixtures, where this based on the previous single axis molecular model also developed by our lab. The pitch angle and polarization anisotropy have sufficient sensitivity to differentiate Col I from the Col I/Col III mixtures. Therefore the polarization probing was proved to be an efficient method for measuring collagen isoform distributions in the ECM. [3-5] The 3D multi-view SHG imaging platform coupled with polarization analysis would provide a more advanced measurement technique to yield a better metric of collagen composition and disease progression.

6.1.3 Enhancement of Second-Order Nonlinear-Optical Signals

One of the biggest challenges for SHG imaging is that the inefficient nonlinear optical emission conversion leads to poor signal-to-noise ration especially significant for the 3D thick tissue imaging. The photon scattering and absorption even make this situation more demanding for the signal acquisition. Such issue raises the limitation for the detection method, time and type of the tissue applicable for the SHG imaging. Fortunately, recent researchers have developed ways using optical stimulation for the enhancement of SHG signal. Those experiment demonstrated stimulated enhancement of SHG and difference frequency generation (DFG) in a configuration that is suitable for a wide variety of samples. A quantitatively description was observed for power and phase dependences using a coupled-wave formalism and achieve signal amplification of $>10^4$ in the biologically relevant sample collagen I. Such optical stimulation would be extremely advantageous in systems with the biological tissue spontaneous signals. [6] This

method can accelerate the current signal acquisition by providing more options for detector and enable better imaging qualify for current SHG imaging modality.

6.2 Future Directions for Computer Vision Algorithms Applied for Biological Tissue Classification

Computer vision and machine learning has been establish as a more and more important tool in disease diagnosis and prognosis in recent researches and clinical application. In this thesis we discussed texton analysis and wavelet transform for IPF and ovarian cancer classification. There are many more directions of future work that might be worth exploring for other kinds of biological tissues. In our previous chapters, we have analyze the lung fibrosis and also the ovarian tissue, in which the tissues are assumed to be homogenous and image features were generated without too much consideration of the position of the pixel/voxels. However this method has its limitations for the choice of tissues. For example, the breast cancer was diagnosed considering the relative relationship between collagen and collection of cells, where the collagen alignment related to the cell boundary is extremely important. In that case, we could further develop more sophisticated algorithm for tissue classification with segmentation and recognition of relationship between collagen fiber and cell aggregation boundaries. First we can apply the segmentation so that we can choose the part with collagen and find out the common image feature for the collagen, in which area we are still consider the tissue with collagen is homogenous. At the same time, we can also find out what is the alignment distribution between collagen fiber and collection of cell boundary using the curvelet transform or CT-fire developed

by the colleagues from LOCI [7, 8]. The alignment will also be considered as the new feature in the spectral of all image features.

Another important application will be the tissue micro-array analysis where there are not only information about types of disease but also the patient information such as age, disease stage, body type, prognostic information (how long the patient survived and when they were treated with chemotherapy, what kind of the chemo therapy they were treated, how long it takes for the cancer reoccurrence. The logistical regression can be a useful algorithm relating the background information and the patient personal information. And therefore, the image features and the patient background information would be both taken into account for the further diagnosis or prognosis purpose. The introduced logistical weighting for different features would indicate which information would be more important in the diagnosis and prognosis process.

Another direction worth exploring is called spatial point process, which enable us to study more information about cell aggregation. A spatial point process is a random pattern of points in d-dimensional space. Spatial point processes are useful as statistical models in the analysis of observed patterns of points, where the points represent the locations of some object of study. Point processes play a special role in stochastic geometry, as the building blocks of more complicated random set models (such as the Boolean model), and as instructive simple examples of random sets. This method would be extremely helpful for microscopy image processing

including the SHG and wild field microscopy imaging, where we focus more about the cell and cell nuclear grouping.

At last, the most recent imaging recognition process utilized deep learning algorithms. Deep learning [10] (also known as deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers, with complex structures or otherwise, composed of multiple non-linear transformations. This is a method that is more close to artificial intelligence and requires high volume of the data for the analysis. But instead of focusing on some supervised features, it is more semi-supervised or non-supervised method that could relate the abstract features for prognosis and diagnosis purpose. This method is extremely successful for high volume data. In the coming age of big data and personal medicine where picture archiving and communication system (PACS) system is becoming mature and widely applied, the machine learning and computer vision is going to play a significant role in diagnosis and prognosis.

6.3 SHG Imaging and Clinical Impact

SHG's superb ability of probing ECM structure changes promise us a great potential in clinical application for disease diagnosis. SHG imaging has already demonstrated its capability in cancer

detection during initiation and progression. Such changes have been shown in both ex vivo and in vivo with proper experimental setup.

Recently, researchers have demonstrated the integration of SHG imaging with combinations of endoscope and laparoscope [11], which provide more clinical standard solution. And in terms of ovarian cancer with SHG and texture analysis, the implementation of endoscope and laparoscope provides possible screening solution for women who are at high risk of developing ovarian cancer due to BRCA1/BRCA2 gene mutations with high sensitivity and specificity. This would be huge improvement from current CA125 [12, 13] method since women with BRCA1/BRCA2 gene mutations have 30-40% probability of developing ovarian cancer or breast cancer during their lifetime. The yearly screening and monitor the ECM structure would help them for diagnose ovarian cancer at early stages. Our lab has been also developing multi-scale imaging modality to combine ultra-sound imaging and more localized SHG imaging and scattering. In this case the imaging modality could cover different imaging scale and first using ultrasound detecting suspicious part of the diseased tissue, then zoom into the area and implement SHG imaging for ECM structure measure. This is very promising technique for evolving current CA125 detection and integrate SHG imaging for clinical practice.

6.4 References

1. Moreaux, L.; Sandre, O.; Mertz, J., Membrane imaging by second-harmonic generation microscopy. *J. Opt. Soc. Am. B* **2000,** 17, 1685-1694.
2. Mertz, J.; Moreaux, L., Second-harmonic generation by focused excitation of inhomogeneously distributed scatterers. *Opt Commun* **2001,** 196, (1-6), 325-330.
3. Chi-Hsiang Lien and others, 'Precise, Motion-Free Polarization Control in Second Harmonic Generation Microscopy Using a Liquid Crystal Modulator in the Infinity Space.', *Biomedical Optics Express*, 4.10 (2013), 1991–2002.
4. Karissa Tilbury and others, 'Differentiation of Col I and Col III Isoforms in Stromal Models of Ovarian Cancer by Analysis of Second Harmonic Generation Polarization and Emission Directionality.', *Biophysical Journal*, 106.2 (2014), 354–65.
5. Oleg Nadiarnykh and Paul J Campagnola, 'Retention of Polarization Signatures in SHG Microscopy of Scattering Tissues through Optical Clearing.', *Optics Express*, 17.7 (2009), 5794–5806.
6. a. J. Goodman and W. a. Tisdale, 'Enhancement of Second-Order Nonlinear-Optical Signals by Optical Stimulation', *Physical Review Letters*, 114.18 (2015), 1–5.
7. Provenzano PP, Eliceiri KW, Campbell JM, Inman DR, White JG, Keely PJ. Collagen reorganization at the tumor-stromal interface facilitates local invasion. BMC Med. 2006;4:38.
8. Stein AM, Vader DA, Jawerth LM, Weitz DA, Sander LM. An algorithm for extracting the network geometry of three-dimensional collagen gels. J Microsc. 2008;232(3):463–75

9. Gatrell AC, Bailey C, Diggle PJ, Rowligsont, BS. ‘Spatial point pattern analysis and its application in geographical epidemiology’, Transactions of the Insstitute of British Geographers. Vol. 21, No. 1(1996), pp. 256-274.
10. Arel I, Rose DC, Karnowski TP, ‘Deep machine learning-a new frontier in artificial intelligence research’, IEEE Computational Intelligence Magazine. Vol. 5, Issue:4(2010), 13-18
11. Yuying Zhang and others, ‘A Compact Fiber-Optic SHG Scanning Endomicroscope and Its Application to Visualize Cervical Remodeling during Pregnancy.’, *Proceedings of the National Academy of Sciences of the United States of America*, 109.32 (2012), 12878–83.
12. Bast RC, Xu FJ, Yu YH, Barnhill S, Zhang Z, Mills GB, ‘CA 125: the past and the future’, The International Journal of Biological Markers, (1998), 13(4): 179-187