

NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

# **TFY4500 biophysics specialization project**

by

Arve Seljebu

Fakultet for informasjonsteknologi, matematikk og elektroteknikk  
Institutt for fysikk

December 2014

# Summary

In the search for curing cancer researchers have had a clear focus on understanding tumor cells. Though still outnumbered by the number of published articles on tumor cells, studies focusing on tumor cell environment is getting traction. Some studies have shown that the tumor environment, also known as the tumor stroma, may have significant impact on development of a tumor. It's suspected that connective and supporting tissue in the tumor stroma have a significant role in the metastasis, the state when tumor cells spreads to new parts of the body.

In fact research suggest that there is a correlation between structure of collagen tissue and survival of the patient. This can be used in diagnosis, but also suggest that there might be possibilities to avoid cancer cells spreading by altering the tumor stroma. This report will focus on the diagnosis part, in specific on collagen tissue from breast cancer.

In example, one property of tissue structure is alignment and curliness of collagen fibers. An earlier study at NTNU have measured curliness and alignment properties in a manual qualitative manner on a data set of 37 samples. The purpose of this report is documenting project work that has been done on automating the process such that a larger data set can be obtained and studied. To be specific, the project have explored possibilities for obtaining large number of high quality images of breast tissue samples with automated nonlinear microscope scanning. The project have also created some quantitative data from the images.

The goal for this report is informing about the process and the results. The details are:

- Find good parameters for obtaining high quality images
- Find an effective way to scan whole glass slides of 135 samples with little human intervention
- Store the images in a structured manner so that
  - Correlation to patient data is possible

- Quantitative data can be subtracted
- Try some algorithms for creating quantitative data

**The results are:**

- Nonlinear scanning microscope with femto-second laser and non-descanned sensors give images with visible fiber structures. Photonmultiplier tubes (PMT) are subject to more noise than hybrid detectors (HyD), but PMT are more reliable when sample contain areas of very high intensity.
- Having a way to adjust plane of sample making it horizontal is crucial for scanning an area of  $\approx 3 \text{ cm}^2$  without loosing focus. Microscope software should also have a way to define plane of focus before starting the scan, as auto focusing all coordinates is not a viable option due to time constraints, adding 7+ hours to the scan. Using auto focus at regular intervals are neither an option, as it may hit a coordinate with no signal which results in selecting focus at random. It is hard to find a reliable process to detect if little signal is due to focus problems after the scan have been done and it's also cumbersome to rescan selective parts of the glass slide. Tweaking settings of microscope, making xyz-coordinates easily available and storing spatial positions in memory, makes aligning of the sample before scan more effective.
- Selecting and grouping specific part of the glass slide already at scanning is preferred, as diving the scan images into groups representing one sample is non-trivial for a data set containing 3-4000 images. In contrast, grouping samples when scanning makes image processing and keeping record of spatial placement to specific tissue samples a trivial task.
- Quantifying fiber direction with Fourier transform gives a feature that possibly can be used in classification to separate anisotropic samples from samples having fibers in a few specific directions. Using gradients for creating features have not been successful.

# Contents

<b>Summary</b>	<b>i</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Theory</b>	<b>3</b>
2.1 Image processing . . . . .	3
2.1.1 Representation . . . . .	3
2.1.2 Contrast enhancement . . . . .	3
2.2 Scanning microscope . . . . .	4
2.3 Nonlinear light interaction . . . . .	5
2.4 Fourier transform . . . . .	6
<b>3 Method</b>	<b>8</b>
3.1 Sample slide . . . . .	8
3.2 Instrumentation . . . . .	9
3.2.1 Microscope . . . . .	9
3.3 Semi-automatic image collection . . . . .	9
3.3.1 Tile scan . . . . .	10
3.3.2 Matrix scan . . . . .	10
3.4 Algorithms . . . . .	11
3.4.1 Histogram of oriented gradients . . . . .	11
3.4.2 Frequency domain . . . . .	11
<b>4 Result</b>	<b>13</b>
4.1 Scan settings . . . . .	13
4.2 Structured scan . . . . .	14
4.3 Quantitative data . . . . .	14
4.3.1 Histogram of oriented gradients . . . . .	15
4.3.2 Fourier transform . . . . .	17
4.3.2.1 Line fit . . . . .	17
4.3.2.2 Sum of angles . . . . .	19
<b>5 Discussion</b>	<b>20</b>
5.1 Scan settings . . . . .	20
5.2 Automation of the process . . . . .	21
5.3 Quantification . . . . .	21
5.3.1 Gradient . . . . .	21
5.3.2 Frequency domain . . . . .	22

<b>6 Conclusion</b>	<b>24</b>
---------------------	-----------

<b>Bibliography</b>	<b>25</b>
---------------------	-----------

# Chapter 1

## Introduction

Over three million published articles on pubmed with keyword *cancer* shows the huge amount of research effort for understanding, diagnosing and treating cancer diseases. The research focus is mainly on tumor cells, but a segment of interest which is increasing is research on tumor stroma as seen in figure 1.1. Tumor stroma is basically the environment of cells that is suppressing or supporting the function of the tumor cells. It is suggested that under the development of a tumor, the stroma is changing from being suppressive to supportive of the tumor cells.

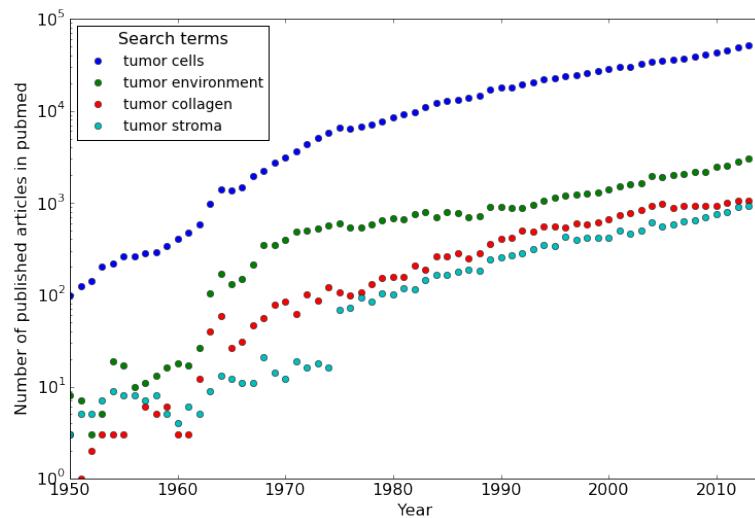


FIGURE 1.1: Amount of published articles by year on different search terms. The search term *tumor cells* outnumber the others by two orders of magnitude. Note the logarithmic scale on y-axis.

In particular, collagen fiber is known to be altered in the surroundings of tumor cells under the development towards metastasis. One bio-marker for collagen fibers, is their alignment at the vicinity of the tumor, which may predict if a tumor is malignant. The fiber alignment can be used as a diagnosis tool for malignant tumor, and an article

written at NTNU have studied collagen fiber alignment in a manual qualitative manner [1].

St. Olav hospital have taken breast tissue samples from 900 patients. In total three samples per patient, one sample inside, one sample at the boundary and one sample outside the tumor. The samples are laid in a matrix on a glass slide, each glass slide having about 130 samples. As microscope scanning and analysis of such a large data set is not straightforward, this project have explored possibilities for automating the process.

To be specific, the goals of the project were:

- Find good parameters for obtaining high quality images
- Find an effective way to scan whole glass slides of 135 samples with little human intervention
- Store the images in a structured manner so that
  - Correlation to patient data is possible
  - Quantitative data can be subtracted
- Try some algorithms for creating quantitative data

# Chapter 2

## Theory

### 2.1 Image processing

#### 2.1.1 Representation

All images in this report are two dimensional arrays of integer values. The image is denoted  $f(x, y)$ , where  $f(0, 0)$  is the intensity value of the top left pixel. Positive direction for x and y-axis is respectively right and down as seen in figure 2.1 (a). Pixel axis will be omitted unless useful for description of image.

An image intensity value of zero equals no signal and representation is dependt on colormap chosen. As microscope images in this report tend to have a lot of near zero pixel values, an inverted colormap is used for better visibility. In other words zero are represented as white, in contrast to black that one might assume. As long as possible, the same colormap is used on same type of images. It is suggested that the reader note the colorbar beside the image which indicate correspondence between numerical value and color.

#### 2.1.2 Contrast enhancement

Images where structures are more important than the numerical values, *histogram equalization* is used to enhance contrast. Histogram equalization tries to make each intesity likely probable, which often result to a higher contrast and therefor better visibility. An example is shown in figure 2.1, where (a) is the original image and (c) is the equalized image. The algorithm used to equalize figure 2.1 (c) is called *adaptive histogram equalization* and are found in the python library scikit-image. As this are just for visibility

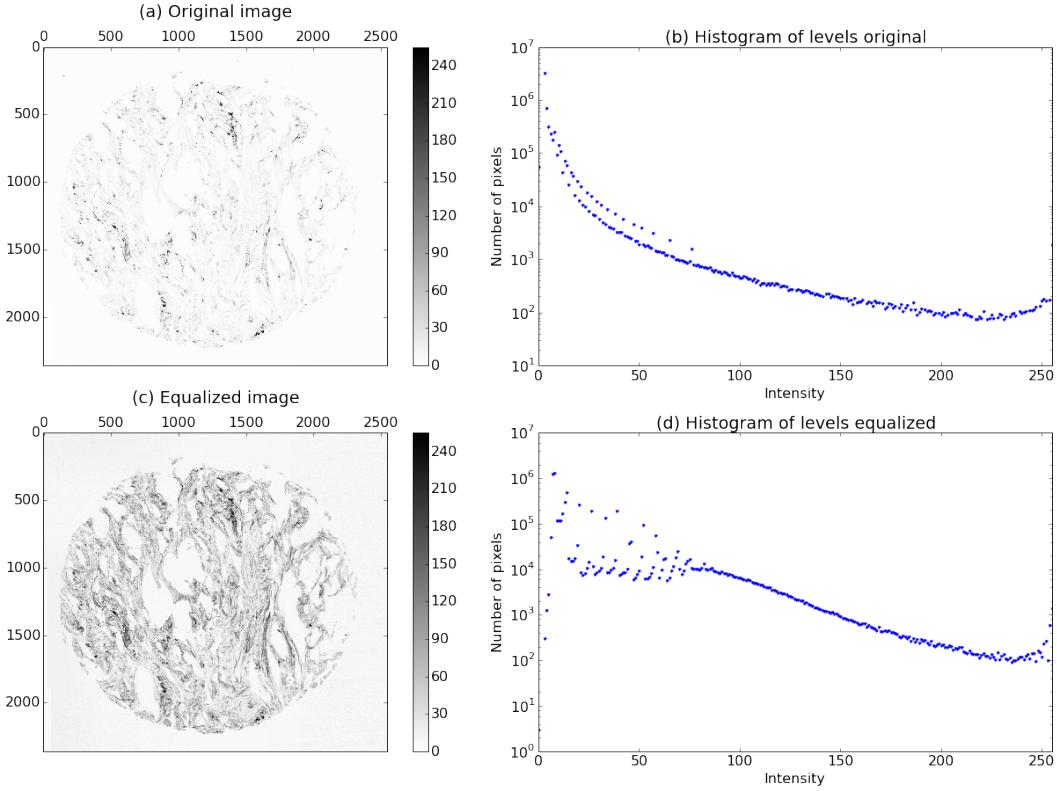


FIGURE 2.1: (a) Original image from microscope. (b) Histogram of intensity values in image (a). (c) Adaptive histogram equalization of image (a). (d) Histogram of intensity values in image (c). Note that the scale of the y-axis in the histograms are logarithmic.

in the report, its suggested to read scikit-image documentation or source code if more detail is needed. Figure caption will always note if images has been processed with any kind of contrast enhancement method.

## 2.2 Scanning microscope

The microscope in use is a scanning microscope. Its called scanning due to the moving mirror which directs the laser to different spatial coordinates on the sample. The laser will move in a raster pattern and a sensor detect light in samples. Each sample represent one pixel and computer software assign the sensed values to a digital image. By scanning speed, we mean the count of lines scanned per second. In example a scanning speed of 600 Hz will use 853 milliseconds to scan a whole image of 512x512 pixels.

As its possible to control the amplitude of the mirror oscillation, choosing a smaller amplitude of the oscillation will effectively zoom. This gives the user of a scanning microscope control over spatial resolution without changing objective. Though the zooming is not without limitation, as too high oscillation frequency could damage the mirror. In

---

example, a high mirror oscillation frequency would limit the possible amplitude which effectively limits how low zoom one can use.

Even if it's possible in software to select resolution below the theoretical resolution defined by the Rayleigh criterion, we are still limited by the diffraction limit. Resolution is therefor given by

$$d = 0.61 \cdot \frac{\lambda}{NA} \quad (2.1)$$

where  $\lambda$  is wavelenght of light and  $NA$  is numerical aperture of objective.

### 2.3 Nonlinear light interaction

Second harmonic generation (SHG) is a interaction between photons and material which generate light of twice the frequency as the incomming light. The generation is due to the material's nonlinear susceptibility, which can be explained by the polarized density

$$\vec{P} = \chi_1 \vec{E} + \chi_2 \vec{E} \cdot \vec{E} + \chi_3 \vec{E} \cdot \vec{E} \cdot \vec{E} + \dots \quad (2.2)$$

When a material have  $\chi_2 \neq 0$ , two photons may interact with the material to create a polarized field. The polarized field then again generates an  $\vec{E}$ -field with twice the frequency as the incoming light (two photons interact with the material, creating a single photon with twice the energy). The interaction only happens upon a high intensity beam where two photons are at the same place at the same time. Practically this is achieved by using a pulsed laser and focusing the beam with an objective, as a steady beam of this intensity magnitude would damage the sample. Also since the probability for interaction is proportional to intensity squared, one get the convenient benefit of essentially no SHG outside the focal point.

$\chi_2$  is also dependent on wavelength of incoming light and properties of the material, which results in material needs to be non-centrocymetric. This happens to be the case for collagen fibers as some of few biological materials, making it simple to take images of the collagen fibers only with little sample preparation.

The pulsed laser is normally scanned over the sample, generating SHG-signal and detected by sensors. As compared to a confocal microscope, descanning of the light is not necessary to take away background signal.

## 2.4 Fourier transform

A common way to view data is in its frequency domain. To achieve this, one does a Fourier transform defined by

$$F\{f(t)\} = \int_{-\infty}^{\infty} f(t)e^{-i2\pi\mu t} dt. \quad (2.3)$$

The transform has the property that it "picks out" amplitude and phase of oscillating signals in  $f(t)$ . This is based on the Fourier series which states that any periodic function can be represented by a sum of finite or infinite set of sines and cosines.

As we are working with images in the discrete domain, one defines  $f(t)$  to be a sampled function using the Dirac delta function and gets the discrete Fourier transform

$$F(u) = \sum_{x=0}^{M-1} f(x)e^{-i2\pi ux/M}. \quad (2.4)$$

Where  $u$  and  $x$  are discrete values for frequency and position, and  $M$  is number of samples in the function  $f(x)$ . Adding a dimension yields the 2D discrete Fourier transform given by

$$F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)e^{-i2\pi(\frac{ux}{M} + \frac{vy}{N})}. \quad (2.5)$$

Here  $u$  and  $v$  are frequencies for accordingly  $x$  and  $y$  dimensions in the real image  $f(x, y)$ .  $M$  and  $N$  are the size of the image. The transform yields an image with same dimensions as the real image, where  $F(u, v)$  gives amplitude and phase of given frequencies  $u$  and  $v$ . All Fourier images in this report are logarithms of power spectrum which are shifted. The power spectrum of the Fourier transform is given by

$$|F(u, v)| = \sqrt{F(u, v) \cdot \overline{F(u, v)}}, \quad (2.6)$$

where  $\overline{F(u, v)}$  is the complex conjugate of  $F(u, v)$ . Values below 1 are set to 1, to avoid division by zero and  $-\infty$  values. Because the domain of values may range  $10^6$  in the power spectrum, logarithmic scaling is done. In example the Fourier spectrum to figure 2.2 (a), have a maximum value of  $8 \cdot 10^6$ . Shifting means that  $u = 0$  and  $v = 0$  are

set to the center of the image, as it's easier to grasp what frequencies are involved and direction of features in the image.

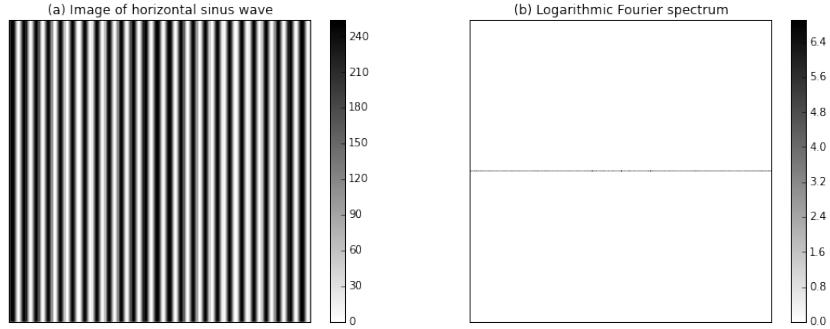


FIGURE 2.2: (a) An 8 bit grayscale image with sine wave running in x-direction. (b) The logarithmic scaled power spectrum of the Fourier transform. The Fourier transform has been shifted for better visibility, as we have no repeating signal in y-direction and the line actually lays on the border of the image ( $v = 0$ ).

As mentioned, the transform picks out amplitude, phase and frequency, which are properties we can exploit to find orientation of fibers. Figure 2.2 shows this concept.

# Chapter 3

## Method

### 3.1 Sample slide

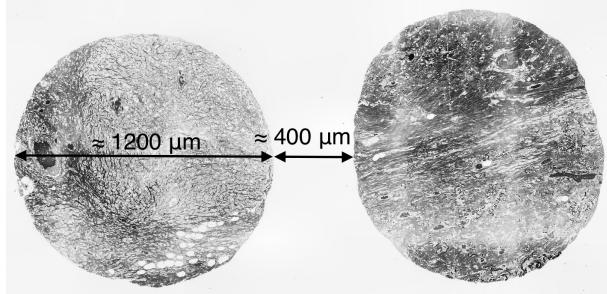


FIGURE 3.1: Two spots side by side with approximate sizes. The left spot have more circular shape than the right one.

Each glass slide contains 9x15 sample spots of breast tissue with diameter  $\approx 1200 \mu\text{m}$ ,  $\approx 0.4 \text{ mm}$  spacing between and  $5 \mu\text{m}$  thickness. An illustration taken with fluorescence microscopy of a stained sample is shown in figure 3.1. All other images in this report is unstained samples. Some of the sample spots seem transparent to human eye and give little or no SHG-signal. It has not been confirmed if this is due to failure in sample preparation or if the tissue is transparent.

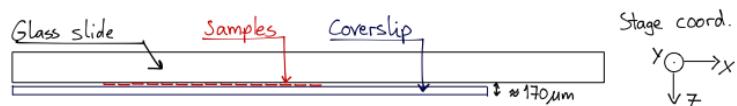


FIGURE 3.2: Illustration of glass slide with coverslip mounted too far to the left, leaving no room for resting the glass slide on the microscope stage.

When glass slide is lying on the microscope stage, sample plane is measured to have tilt by  $\approx 170 \mu\text{m}$  in x-direction due to misaligned coverslip (illustrated in figure 3.2) and  $\approx 70 \mu\text{m}$  in y-direction (unknown cause, might be microscope sample holder).

## 3.2 Instrumentation

### 3.2.1 Microscope

The project used a Leica SP8 microscope with Leica LAS AF software version 3.3.0.10134. Objectives in microscope are shown in table 3.2.1. Laser of microscope is a Chameleon Vision-S from Coherent with 80Mhz pulsing frequency and  $\approx$  75 picoseconds pulse width rated to 2.5 Watt. In all images 890 nanometer was used and output power was shown in the software to be  $\approx$  1950 mW.

Mag/NA	Immersion	Name
20x/0.75	None	HC PL APO
25x/0.95	Water	HCX IRAPO
63x/1.20	Water	HC PL APO
63x/1.40	Oil	HC PL APO

## 3.3 Semi-automatic image collection

Two types of semi-automatic scan jobs were used in the Leica LAS AF software; *tile* and *matric* scan. Before scanning, sample was aligned by using the controlling touch screen seen in figure 3.3.

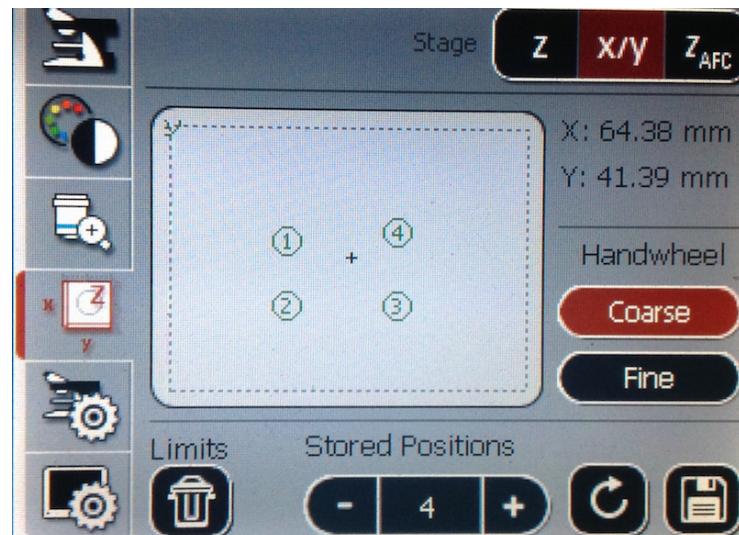


FIGURE 3.3: Touch screen with four stored positions.

### 3.3.1 Tile scan

Tile scan is a feature in the standard view of Leica LAS AF. In tile scan one defines two points which makes a rectangle. Several scan jobs can be made, but is no way to control different jobs to different tiles, which means any job defined will run on all tiles. Auto focusing can be either all tiles or at regular intervals.

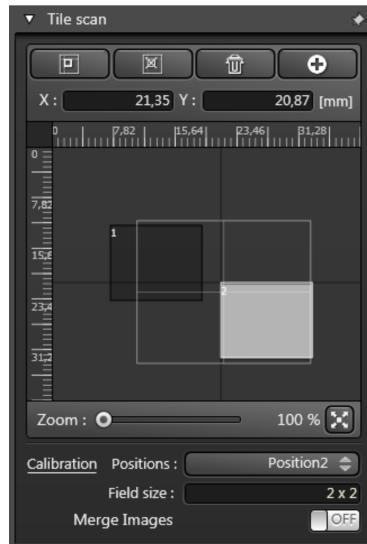


FIGURE 3.4: Leica LAS AF tile scan feature. Two points define an area of tile images. Images are stored in sequence with numbers and software may also merge images.

### 3.3.2 Matrix scan

Matrix scan is a feature that splits a scan into so called *wells* and *scan fields*. This is illustrated in figure 3.5 which have 3x2 wells with 5x5 scan fields. Each well will represent a sample spot.

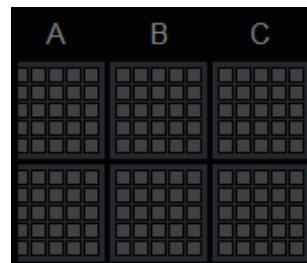


FIGURE 3.5: 3x2 wells, each well containing 5x5 scan fields.

The matrix scan allows for individual adjusting placement of wells and focus of scan fields, making it very flexible. After the matrix scan, images are stitched with ImageJ to form one image per well.

## 3.4 Algorithms

Two types of processing have been tried out for quantifying the images, namely gradient and Fourier transform. The product of the algorithms is a histogram of angles.

### 3.4.1 Histogram of oriented gradients

As the name implies, this algorithm calculates gradient of an image and then use the direction of that gradient to create a histogram of angles. The algorithm is straight forward:

---

```
calculate gradient
calculate angle of gradient
add every angle to histogram weighted by the gradient magnitude
```

---

### 3.4.2 Frequency domain

Two types of strategies was tried out in the frequency domain. The first technique described in psuedo code:

---

```
divide image into sub images of 100x100 pixels
for every sub image:
    if mean intensity < 1: do next sub image
    do Fourier spectrum
    threshold Fourier spectrum
    do line fit
    find angle of fitted line
    add angle to histogram
```

---

An example of the line fitting is illustrated in figure 3.6.

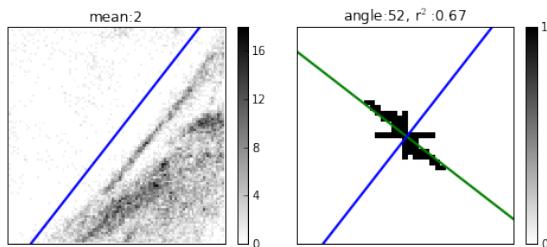


FIGURE 3.6: Example of line fitting in Fourier spectrum. Fourier spectrum have been zoomed, omitting a border of 30 empty pixels.

The second technique is similar:

---

```
do Fourier spectrum
threshold Fourier spectrum
```

```
for every pixel in thresholded Fourier spectrum:  
    find angle  
    add angle to histogram
```

---

# Chapter 4

## Result

### 4.1 Scan settings

Figure 4.1 show a comparison of a 20x dry objective to a 63x water objective. Note that (a) have stronger signal.

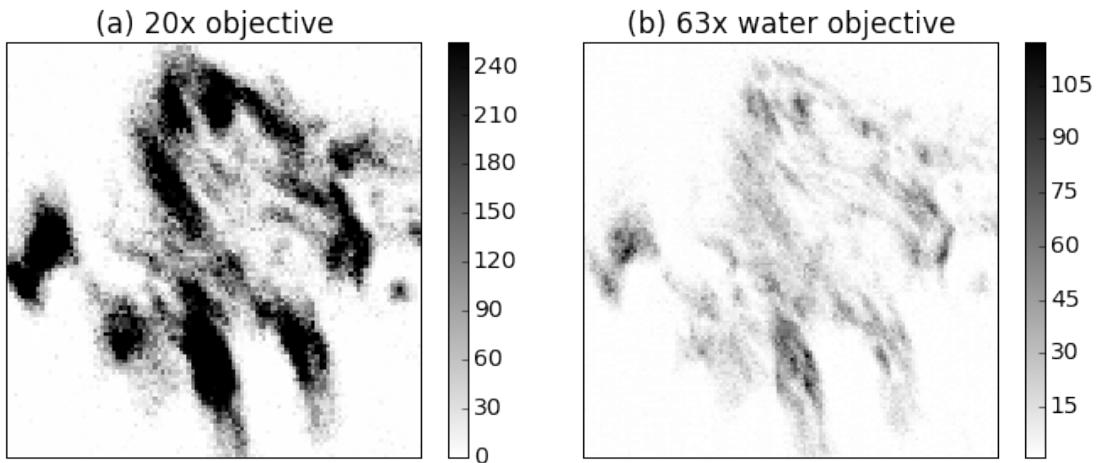


FIGURE 4.1: Comparison of 20x dry objective to 63x water objective. Note that picture (a) have several pixels with maximum value (255). Settings for image (a) are PMT-gain of 700 Volt and scanning speed 400 Hz. Settings for image (b) are PMT-gain of 650 Volt and scanning speed 200 Hz.

Power equal 10%, gain 60% and offset 50% for laser at 890 nm was found to create images taking use of whole 8 bit intensity range when PMT detector had gain of  $\approx$  600 Volt, but not the whole range of the HyD detector. Also some bright spots will make HyD detector shut down, giving large areas with zero value seen in figure 4.2.

Table 4.1 show the minimum time for scanning an area of  $\approx$  1.2x1.2 mm<sup>2</sup> with pixel resolution below 200 nm. All times are given with scanning speed 600 Hz, which is

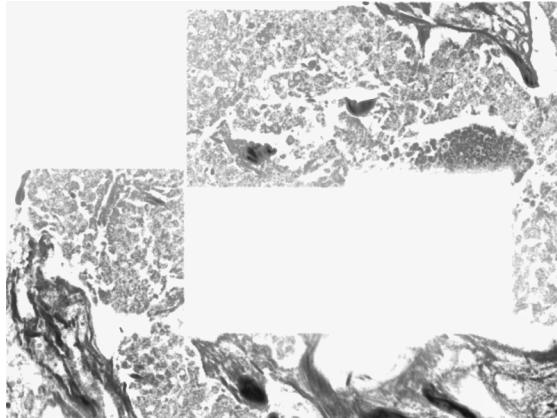


FIGURE 4.2: HyD detector have been shut off due to too strong signal. Image contrast is enhanced to improve visibility.

the highest possible mirror oscillation frequency when using zoom factor 0.75 (minimum zoom factor possible).

Objective	Zoom	Image size	Pixels	Number of images	Total time
20x	0.97	600 $\mu\text{m}$	4192	2x2 = 4	28 s
25x	1.09	400 $\mu\text{m}$	2048	3x3 = 9	31 s
63x	0.92	200 $\mu\text{m}$	1024	6x6 = 36	61 s

TABLE 4.1

## 4.2 Structured scan

Figure 4.3 show a merged tilescan. Note that several spots are uneven and that spots on left side are clipped off. Another defect is lower intensities at merging boarder, which is not easy to spot in the zoomed out figure 4.3, but is clearly seen in figure 3.1.

Figure 4.4 shows an image taken with matrix scan. The image was merged from 5x5 scan fields, saving 11 images per well by omitting some of the sample.

## 4.3 Quantitative data

Three samples have been picked out to illustrate the quantitative results. One image with fibers aligned in mostly one direction, one image with fibers in several directions and one image that seem anisotropic to human eye.

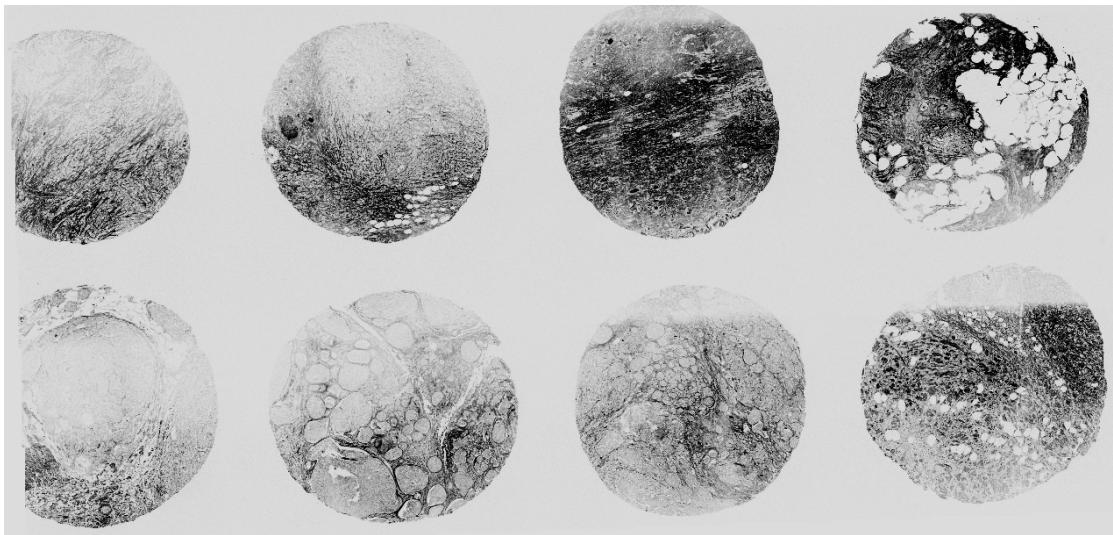


FIGURE 4.3: Fluorescence tile scan with autofocus at regular intervals. Image is merged with Leica LAS AF.

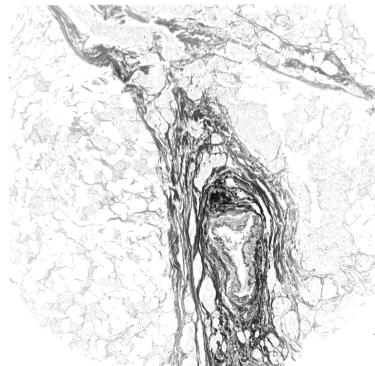


FIGURE 4.4: SHG image taken with matrix scan. Image have been merged with ImageJ and contrast enhanced with scikit-image. Area of image is  $1136 \times 1136 \mu\text{m}^2$ .

#### 4.3.1 Histogram of oriented gradients

Figure 4.5, 4.6 and 4.7 show histogram of angles calculated by the histogram of angles algorithm. Histogram of angles to all images have very strong response for angles 0, 45, 90 and 135, showing the weakness of calculating gradient on two dimensional data.

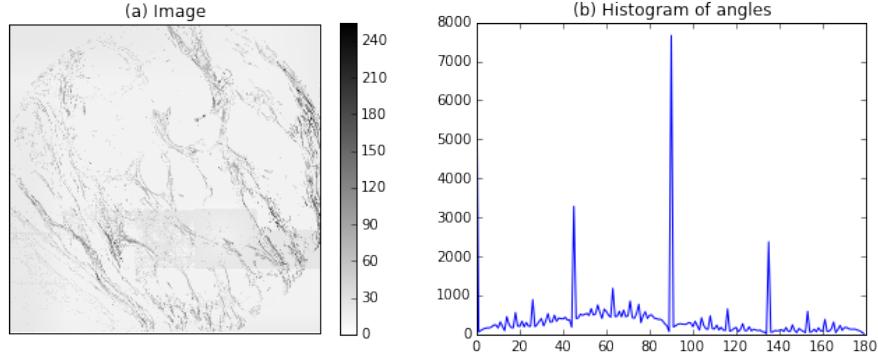


FIGURE 4.5: (a) Image of fibers in mostly one direction. (b) Histogram of angles calculated from gradient to image (a). A response between 40 and 90 is seen if one compare this histogram with figure 4.6 and 4.7.

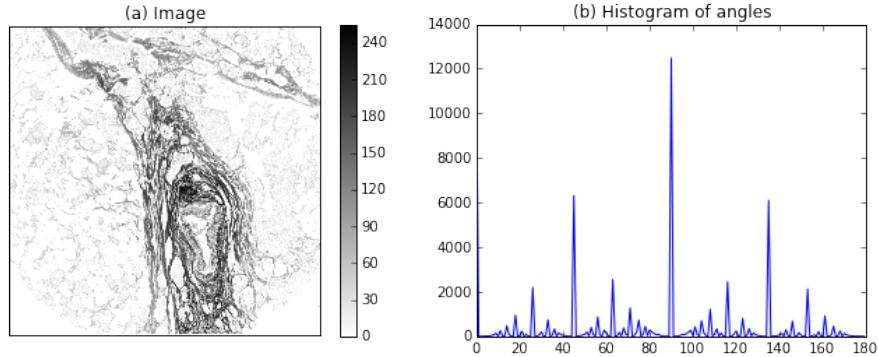


FIGURE 4.6: (a) Image with fibers in several directions. (b) Histogram of angles calculated from gradient to image (a). Histogram doesn't seem to contain any information about the direction of fibers in the image.

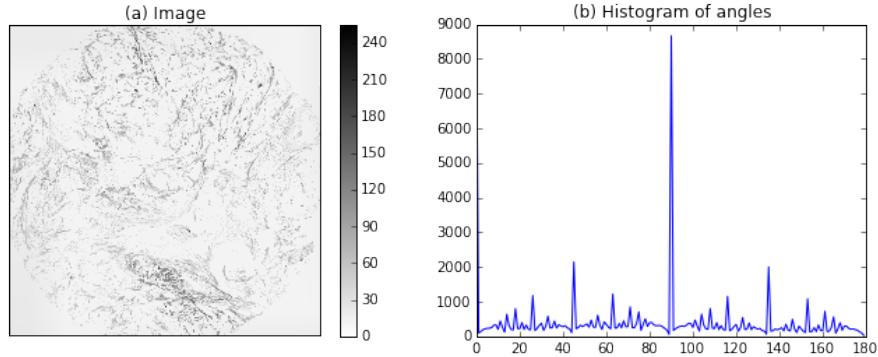


FIGURE 4.7: (a) Image which seems to be anisotropic to human eye. (b) Histogram of angles calculated from gradient of image (a). Histogram is very similar to figure 4.6 (b).

### 4.3.2 Fourier transform

#### 4.3.2.1 Line fit

Figure 4.8, 4.9 and 4.10 show histogram of angles calculated by fitting a line in the Fourier spectrum. All images have similar response with high count of angles around 0 and 90 degrees.

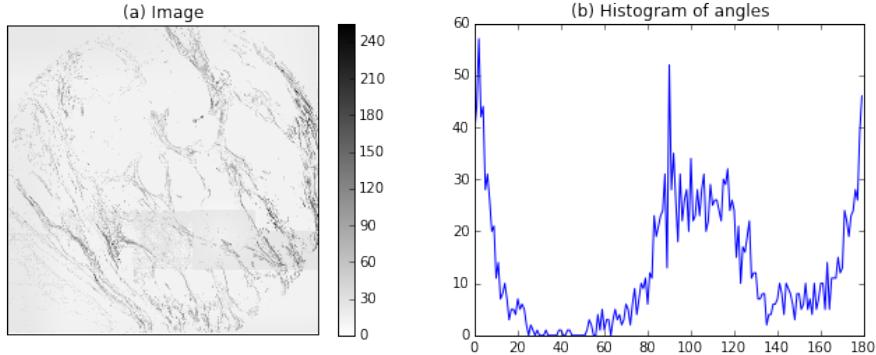


FIGURE 4.8: (a) Image of fibers in mostly one direction. (b) Histogram of angles calculated by using line fit in Fourier spectrum.

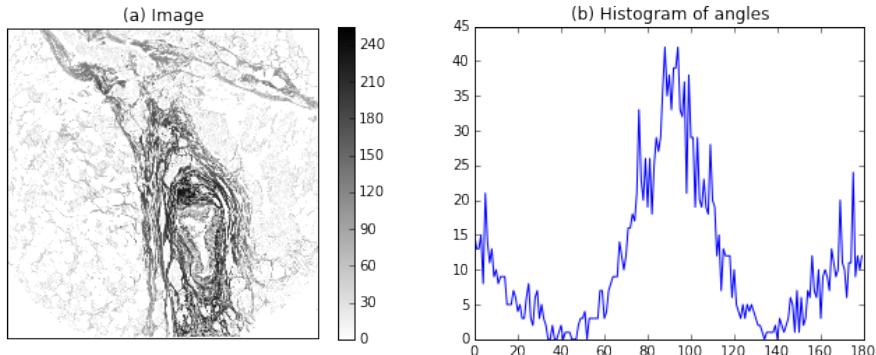


FIGURE 4.9: (a) Image with fibers in several directions. (b) Histogram of angles calculated by using line fit in Fourier spectrum.

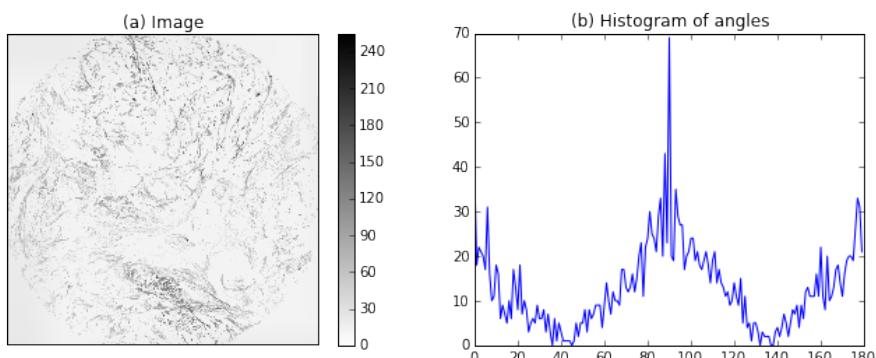


FIGURE 4.10: (a) Image which seems to be anisotropic to human eye. (b) Histogram of angles calculated by using line fit in Fourier spectrum.

As figures 4.8, 4.9 and 4.10 show similar response, lines with a bad fit where discarded making figure 4.11, 4.12 and 4.13. Lines with correlation coefficient  $r^2$  below 0.3 are discarded.

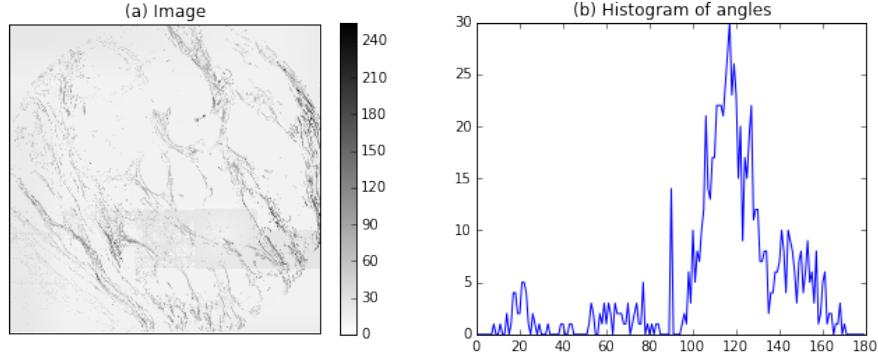


FIGURE 4.11: (a) Image of fibers in mostly one direction. (b) Histogram of angles calculated by using line fit in Fourier spectrum. Lines with  $r^2 \leq 0.3$  are discarded. A clear response at about 120 degree is shown.

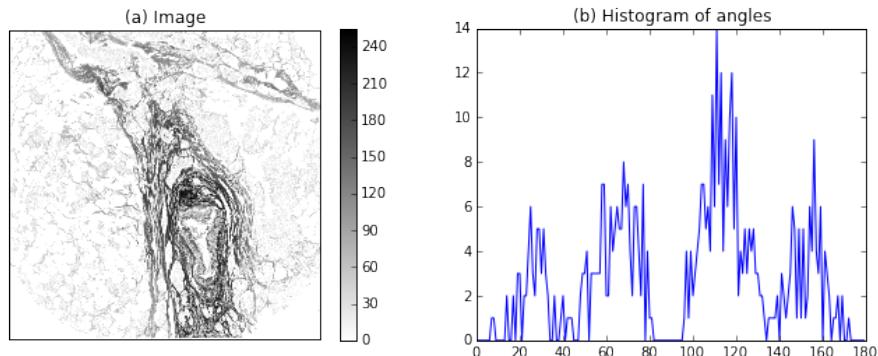


FIGURE 4.12: (a) Image with fibers in several directions. (b) Histogram of angles calculated by using line fit in Fourier spectrum. Lines with  $r^2 \leq 0.3$  are discarded. Four clusters of angles are shown,  $\approx 100\text{-}120$  degree being the strongest response.

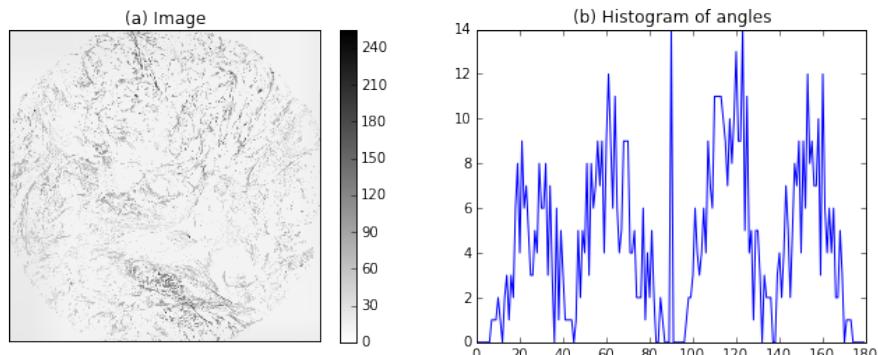


FIGURE 4.13: (a) Image which seems to be anisotropic to human eye. (b) Histogram of angles calculated by using line fit in Fourier spectrum. Lines with  $r^2 \leq 0.3$  are discarded. Four roughly equal cluster of angles are shown.

#### 4.3.2.2 Sum of angles

Figure 4.14, 4.15 and 4.16 show histogram of angles calculated by summing up angles of top intensity pixels in Fourier spectrum.

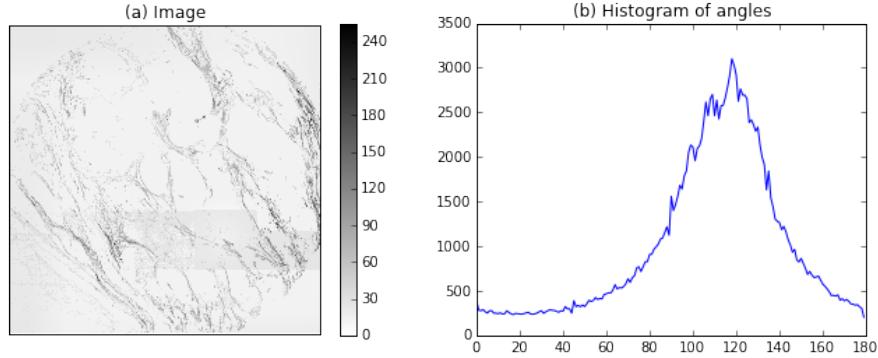


FIGURE 4.14: (a) Image of fibers in mostly one direction. (b) Histogram of angles calculated by summing up the angle of top intensity pixels in Fourier spectrum. Histogram have a peak at  $\approx 125$  degree with  $\approx 40$  degrees width.

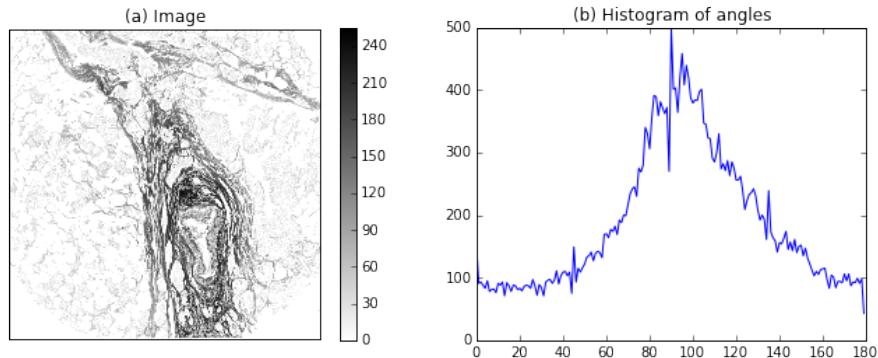


FIGURE 4.15: (a) Image with fibers in several directions. (b) Histogram of angles calculated by summing up the angle of top intensity pixels in Fourier spectrum. Histogram have a peak at  $\approx 95$  degree with  $\approx 50$  degrees width.

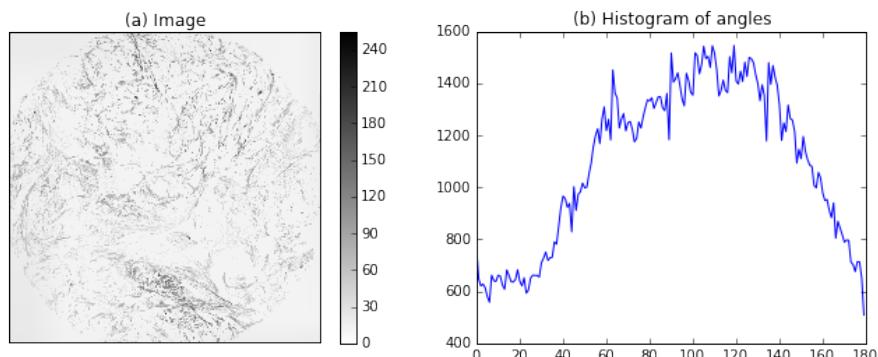


FIGURE 4.16: (a) Image which seems to be anisotropic to human eye. (b) Histogram of angles calculated by summing up the angle of top intensity pixels in Fourier spectrum. Histogram have a high count of angles from  $\approx 40$  to  $160$  degrees, making a broad peak.

# Chapter 5

## Discussion

### 5.1 Scan settings

Selecting acceptable settings is a compromise between resolution, signal, noise and time. As water evaporates rather quick (maximum evaporation time was found to be an hour), scanning a whole slide in one job is not possible with any of the water objectives. This leaves us left with the 20x dry and 63x oil objective. As seen in table 4.1 using 63x oil objective a scan job will take about twice as long as with a 20x dry objective. This was found to be an acceptable trade off between time and resolution, with a whole scan taking minimum two hours.

As we can see in figure 4.1, we have weaker signal with the 63x water objective compared to 20x dry. Although the gain of the detector is 50 Volt higher in image (a), this should not give difference of such magnitude. As a lot of things will effect the intensity, ranging all the way from laser power to mirror oscillation frequency, one should confirm the intensity range by taking a small test scan before scanning a whole slide.

One factor that is limiting to laser power, is HyD detector shutting down due getting several photons at the same time. A good solution to this problem has not been found. Limiting laser power will decrease amount of bright spots which is able to shut down HyD detector, but the problem is not completely excluded until setting laser power too low to generate any significant signal. It is suggested that one concentrate on forward PMT detectors until one can eliminate bright spots.

## 5.2 Automation of the process

Tile scan give limited options to adjusting a scan and also makes it hard to separate samples. Using matrix scan is therefore more appropriate for scanning slides where samples are spread over a large area. Also, a feature in the software called *computer assisted microscopy* may be used for automatic adjusting of placement to wells.

## 5.3 Quantification

### 5.3.1 Gradient

Using gradient for quantifying fiber direction does not seem to be a good match. This is illustrated in figure 5.1 and 5.2, where a lot of random directions get generated. Especially angles which are multiples of 45 degrees get a high response, which is due to the nature of the data (being a two dimensional image). As one might imagine, a single pixel of high intensity (as noise often is), would generate gradients with magnitude of n·45 degrees in the pixels around. This is not easily remedied with filtering, as one would simply create a "mountain" where the strong pixel is (and get the same effect further away from the pixel).

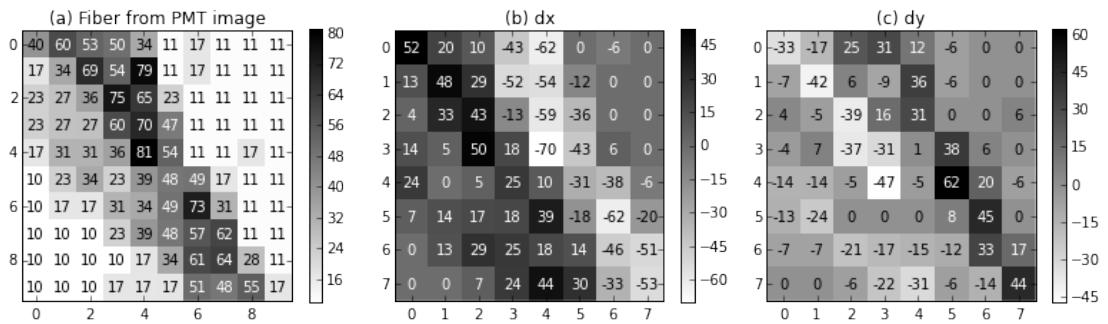


FIGURE 5.1: (a) Section of image with strong signal from a fiber. Numbers in cells are intensity values of image. (b) X-component of gradient. (c) Y-component of gradient.

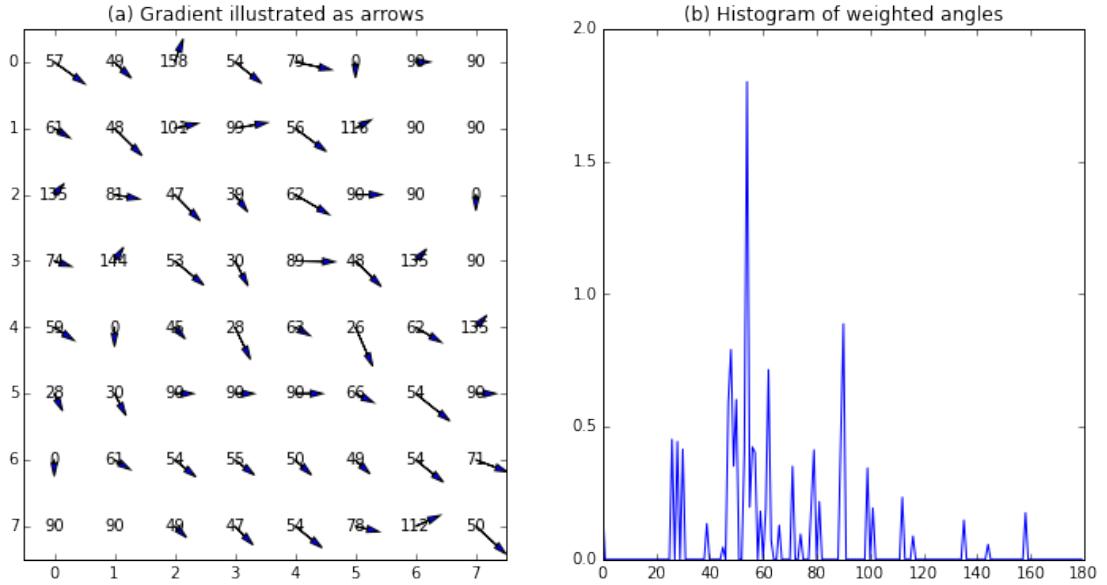


FIGURE 5.2: (a) Gradient to image in figure 5.1 (a) drawn as arrows. An angle of 90 degrees have been added to gradient to align the gradient in the same direction as the fiber. (b) The weighed histogram of angles to gradients in (a).

A possible improvement could be to discard all but the strongest or averaged direction to gradients of sub images. Also, one could try to preprocess the image with some kind of edge detection and then using similar technique on a binary image.

### 5.3.2 Frequency domain

Line fitting of sub images in the Fourier image give good indication of strong fiber orientations. Lines with low correlation coefficient, as in figure 5.3, should be discarded to not include angles which are merely selected at random. A possibility to improve this technique is summing up angles instead of using line fit.

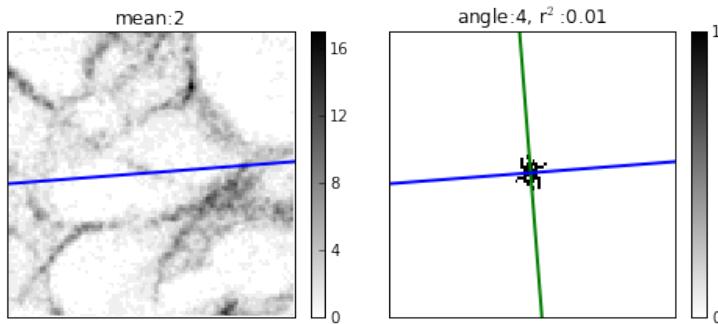


FIGURE 5.3: (a) Sub image of a sample. (b) Fourier spectrum and line fit. Note the low correlation coefficient  $r^2 = 0.01$ .

Summing up angles of highest intensities in Fourier spectrum seem to be good approach in detecting direction of fibers in image. In the images selected, the broadness of the peak in the histogram is a good measure of fiber direction anisotropy. This should be verified on a larger data set. One problem might be directions with similar angles, which may overlap in the histogram and broadening the peak. A possible solution to this is dividing the image into sub images as the previous Fourier algorithm.

## **Chapter 6**

# **Conclusion**

High quality images of collagen fiber has been collected in a structured fashion which may be used for further collection of data. Techniques for analysing the fiber orientation have been tested out. For further work, it is suggested that one should improve collection of images by using Computer Assisted Scanning, to make the collection more time efficient. For analysis it is suggested that one do machine learning on a large data set and try to predict outcome of patients.

# Bibliography

- [1] MONICA J. ENGSTRM OLAV A. HAUGEN LINDA A. DYRNES BJORN O. ASVOLD MAGNUS B. LILLEDAHL ANDERS BRABRAND, IAN I. KARIUKI and ANNA M. BOFIN. Alterations in collagen fibre patterns in breast cancer. A premise for tumour invasiveness? *APMIS*, 12298, 2014. doi: <http://dx.doi.org/10.1111/apm.12298>.