Image Processing and Machine Learning for Fibiril Porous Media

Jim Adriazola (NJIT), Kevin Cotton (Claremont Graduate U.),
Ellie Gurvich (U. Maryland Baltimore County), Richard Moore (NJIT),
Qingxia Li (Fisk U.) Vrushaly Shinglot (U. Texas Dallas),
Aprajita Singh (U. Texas Dallas), Yixuan Sun (NJIT), Thomas Tu (UCLA),
Mingkai Yu (U. Maryland Baltimore County), Naren Vohra (Oregan State U.)

June 17-21 2019

Abstract

Characterizing the microstructure of porous media plays a very important role in a variety of industrial applications. Several algorithms have been studied for the image processing of fibrous materials. The Radon transform and Hough transform investigation shows promising preliminary results for real fiber image processing.

1 Introduction

Characterizing the microstructure of porous media plays a very important role in a variety of industrial applications. For example, in filtration, the performance of the filter is impacted by its microstructure properties, such as pore size distribution and pore connectivity. Researchers are also creating virtual structures that mimic fibril porous media using other microstructure properties such as distributions of fiber size and fiber orientation. The pore size distribution can then be obtained for these virtual structures and compared with results from other experimental techniques like mercury intrusion porosimetry.

Scanning electron microscopy (SEM) is commonly used for imaging surface microstructures. The images produced are usually in grey scale and have a resolution of about 10 nanometers. However, some image analysis is required to accurately extract microstructure properties from these images. The two main research questions for this week-long workshop are: (1) Can we identify accurate distributions of both the fiber diameter and fiber orientation from an image of a sample? and (2) Can we combine fiber size distributions from images of different magnification/resolution?

The group has divided into four subgroups. The first group focused on the existing techniques and algorithms that are employed by the ImageJ software. From experiments done with the various algorithms on fiber images that were generated with known distributions, it was seen that Ridge Detection, Edge Detection, Mavi2dfiber, and Local Thickness algorithms produced the distributions for the diameter of the fibers.

We found that the Local Thickness algorithm works well for all points except for the junctions. Ridge Detection seemed to fit the known distribution much better than the others, but was still considerably off. For that reason, ImageJ's Ridge Detection was investigated for the appropriate parameter values.

With respect to the 3 mandatory parameters, it was found that the lower values of 'Lower Threshold' seem to improve the algorithm's performance at junctions and the higher values of 'Higher Threshold' helps to recognize more 'lines' in the image. On the other hand, we also found out that the different σ (variance of the Gaussian kernel) values detected fibers of different widths. In particular, to detect wider fibers we would want to set higher σ values and narrower fibers are better found by choosing lower σ values. Therefore, though the Ridge Detection algorithm works better than the other algorithms by ImageJ, it would be best suited for those fibers with narrower distribution. Nevertheless, this algorithm still does not seem to resolve the fibers well at the junctions.

Another method given by Pourdeyhimi, computes the Euclidean distance transform of the fibers which in turn directly calculates the width. The caveat here, though, is that transform is calculated of a binary image which leads to a loss of information depending on the threshold value. The Local Thickness algorithm works on similar grounds using the Spherical Distance Transform, and seems to work well for calculating the thickness of the fibers at all points but at the junction.

The second group applied Machine Learning to obtain distribution properties of fibers. In order to have a larger data set to train on, a large database of synthetic images was produced. Each image has a known distribution of two parameters, namely, the fiber angle and the fiber width. Using edge detection, one can take real data and produce data that is quite similar to the synthetic data. The synthetic data is imported and pre-processed using edge detection and the Hough transform. The image in Hough transformed space is then fed to a Convolutional Neural Network that is trained to extract distribution features on the parameters of the fibers.

The third group focused on the application of the Hough transform. The Hough transform is a feature extraction technique used in image analysis. It helps to solve problem of missing points or pixels in detecting simple shapes like straight lines, circle or ellipses. The simplest case is detecting straight lines. A straight line $y = mx + b$ can be represented as a point (b, m) in the parameter space. However, vertical lines pose a problem as it leads to unbounded value of m. For computational purposes we use Hesse normal form $r = x\cos(\theta) + y\sin(\theta)$. Thus, it is possible to associate with each line of the image a pair (r, θ) in Hough space. Given a single point in the plane, the set of straight lines passing through that point corresponds to a sinusoidal curve in Hough space. Thus, the problem of detecting colinear points (points on a line) can be converted to the problem of finding cocurrent curves. Applying Hough transform to a simple image with three fibers in MATLAB did detect all the edges of the fibers. Once all the edges are detected, edges with equal slope can be paired up. The distance between the two parallel lines will give the width of the fiber.

The fourth group attempted a probabilistic approach to identify a distribution of the fiber diameters. They tested this method on generated images of fibers with known distributions of fiber widths. They considered taking a one-dimensional slice of the image, which intersects a fiber at a normally distributed angle θ. The width of the intersection X is measured by counting the number of consecutive light pixels, since the background is dark. It follows that for a fixed fiber width h, the expected values of X and θ are related by $h = \mathbb{E}[X]/\mathbb{E}[\sec(\theta)]$. The fact that $\sec(\theta)$ is unbounded is adjusted by using the fact that the slice through a fiber can be no longer than the length of the fiber L. Hence, they derived that the expected value of X is given by $\mathbb{E}[X] = \mathbb{E}[\min(h\sec(\theta), L)] \approx 2h(1 + \log(2L/h))/\pi$. However, L is correlated with the probability of sampling a value of h.

Since each measured intersection corresponds to a range of possible width values, the group

multiplied the empirical distribution with a distribution of possible widths. This led to a distribution of fiber widths that is much closer to the known distribution. The changes in the width distribution when the images were magnified by 2X, 4X, and 8X were observed, in which the trend of distributions skewed right accordingly, so that the fibers of smaller width are over represented.

2 Existing Models

We studied the main existing image detection algorithms that have been used. The first of them is Local Thickness Algorithm. As the name suggests, this algorithm detects the thickness of the fiber at each point, locally. The second is the Ridge Detection which has given better results in the experiments and our objective was to understand the working of it, so we could suggest how the plugin in the ImageJ software could be used in the best way possible.

2.1 Local Thickness

The Local thickness algorithm works on the principle of Distance Maps.

2.1.1 Definition

Local Thickness at a point: The diameter of the largest sphere that fits in the fiber containing the point.

2.1.2 Steps of the Method:

Finding Distance Map: The Distance map for a given point is the radius of the largest sphere contained in the fiber that is centered at the given point.
Local Thickness Computation: The local thickness is given in terms of the Distance Map.
$\tau(p) = 2maxD(q)$
$q \in X(p)$
where $X(p)$ is the set of all points whose distance map sphere contains p.

Now since we want to avoid the redundant spheres, i.e., smaller spheres contained in larger ones, the concept of Distance Ridge, $X_R(p)$.

2.1.3 Distance Ridge:

X_R consists of the centers of all those spheres containing p that are not contained in any other sphere containing p. The concept of using this set instead of $X(p)$ increases the efficiency of the algorithm, considerably.

2.1.4 Why Local Thickness doesn't work well at the intersections?

Since the Local Thickness algorithm works on fitting largest spheres at each point, the intersection area can fit multiple spheres and so the local thickness at intersection is vague.

2.2 Ridge Detection Algorithm

The next Image Analysis Algorithm is Ridge Detection Algorithm. The algorithm works by identifying the center line passing through the fiber. This recognition is done by the use of changing gradients. The points on the center line is detected by the Calculus approach. The points are then joined to form the center line. Thereafter, the algorithm goes on to find the edges, using this center line, on both the sides. It then detects the width of the fiber by finding the distance between the edges. This distance will be called the thickness of the fiber.

3 Image Processing

3.1 Radon Transform

- Definition:

$$R(f)(P) = \int_P f,$$

  where P is the hyperplane in $\mathbb{R}^n$, $f \in S(\mathbb{R}^n)$. (Fourier Analysis an introduction, Stein, E., Shakarchi, R.)

- In 2D, it coincides with X-ray transform:

$$X(\rho)(L) = \int_L \rho,$$

  where L is a line in $\mathbb{R}^2$, $\rho \in S(\mathbb{R}^2)$ is typically a density function.

- Connection with Hough transform (Hough, P. V. C. 1962 Patent No. 3,069,654).

$$\begin{cases} y = y_i x + x_i, & \text{Rosenfeld, 1969} \\ \rho = x_i \cos\theta + y_i \sin\theta, & \text{Duda and Hart, 1972} \end{cases}$$

$$R(f)(\theta, \rho) = \int_{-\infty}^{\infty} f(z\sin\theta + \rho\cos\theta, -z\cos\theta + \rho\sin\theta)dz$$

3.2 Hough Transofrm

The Hough transform is a feature extraction technique used in image analysis. In image processing, a problem often arises of detecting simple shapes, such as straight lines, circles or ellipses. The Hough transform helps to solve problem of missing points or pixels in detecting such simple shapes.

The simplest case is detecting straight lines. A straight line $y = mx + b$ can be represented as a point (b, m) in the parameter space. However, vertical lines pose a problem. For computational purposes we use $r = x\cos(\theta) + y\sin(\theta)$.

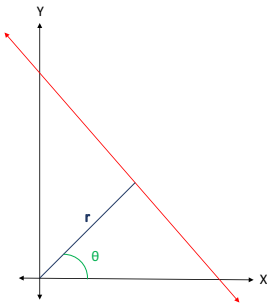

Figure 1: Parametrization of straight line.

Thus, it is possible to associate with each line of the image a pair (r, θ) in Hough space. Given a single point in the plane, the set of all straight lines going through that point corresponds to a sinusoidal curve in (r, θ), which is unique to that point. A set of two or more points that form a straight line will produce sinusoidals which cross at (r, θ) for that line. Thus, the problem of detecting colinear points can be converted to the problem of finding intersecting curves.

Let's consider the example given in 2. The figure on the left shows two intersecting straight lines. The figure on the right shows its Hough transform. The two distinctly bright spots are the Hough parameters of the two lines. From these spots' positions, angle and distance from image center of the two lines in the input image can be determined.

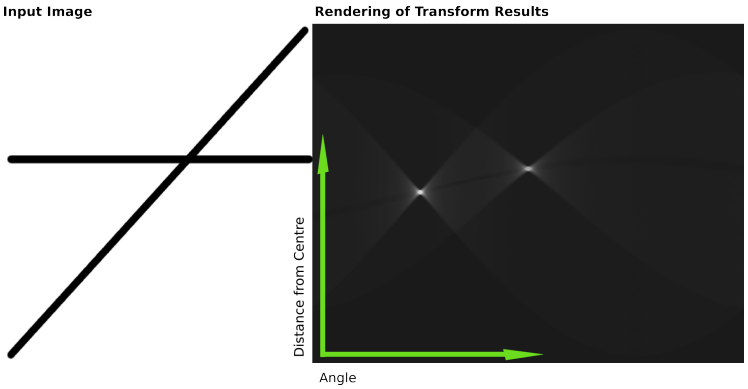

Figure 2: Hough transorm of two intersecting straight lines.

3.3 Image processing procedure

- change image to grey scale (rgb2gray)
- find the edge (edge)
- radon transform the image with just edges (radon)
- locate peaks (tricky part)
- peak location gives information about size and orientation (couple peaks)

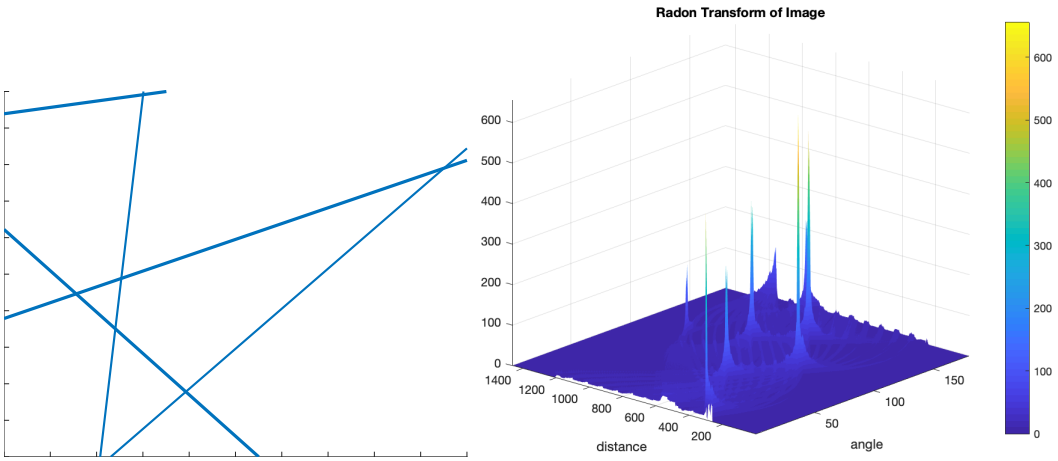

Figure 3: Fiber image processing using Radon transform.

3.4 Preliminary results

Figs 3 shows image processing using Radon transform. Figs 4 shows image processing using Hough transform. Figs 5 shows image processing of real fiber image with Radon transform.

4 Machine Learning

5 1D Slicing

We also considered another image processing approach to calculating the distribution of fiber diameters. By taking a single 1D slice of the image (a row or a column), we can measure the width of the intersection of a fiber with the slice simply by counting the number of consecutive bright pixels. However, the fiber diameter (h) will be less than the width of the intersection (X), since the slice intersects the fiber at an angle θ, as shown in Figure 6.

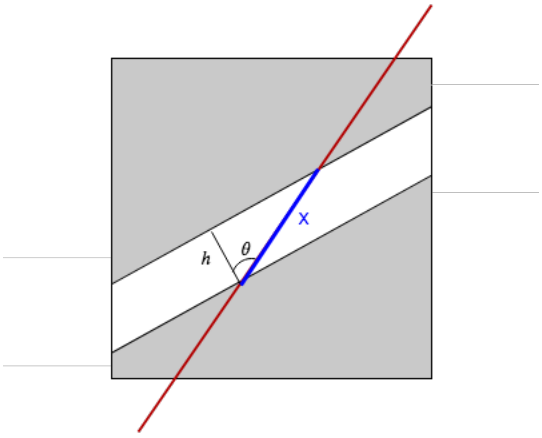

Figure 6: Diagram of the intersection of a 1D slice with a single fiber of diameter h at an angle θ

Figure 4: Fiber image processing using Hough transform. Code from Davin.

Here, we assume that the distribution of θ is uniform in the range $\left[0, \frac{\pi}{2}\right)$; however, the general approach can be used for any known distribution of θ, so long as it is independent of X. We also assume that $h \ll$ the length of the fiber in the frame, so that the probability of taking a slice which intersects the ends of the fiber is low.

Because $X = h\sec(\theta)$, we considered taking an approximation of the distribution of h as the distribution of X divided by the expectation of $\sec(\theta)$. However, this expectation is unbounded when θ ranges from 0 to $\frac{\pi}{2}$. We can make this bounded by enforcing a maximum length L on the measurement, due to the bounds of the frame of the image. The expectation of X is then:

$$\mathbb{E}\left(\min(h\sec(\theta), L)\right) = \frac{2}{\pi}\left[\int_0^{\cos^{-1}\frac{h}{L}} h\sec(\theta)d\theta + \left(\frac{\pi}{2} - \cos^{-1}\frac{h}{L}\right)L\right] \tag{1}$$

$$= \frac{2}{\pi}\left(h\log\left(\frac{L+\sqrt{L^2-h^2}}{h}\right) + L\sin^{-1}\frac{h}{L}\right) \tag{2}$$

$$\approx \frac{2h}{\pi}\left(1 + \log\left(\frac{2L}{h}\right)\right) \tag{3}$$

Unfortunately, this results in an equation that has multiple solutions for h for a given X. Additionally, we cannot measure the value of L using a 1D slice.

Instead, let the probability density functions of h, X, and $\cos(\theta)$ be f_h, f_X, and f_Y respectively. f_X can be estimated through measurement as:

$$f_X(x) \approx \hat{f}_X(x) := h_j, x_j < x < x_{j+1} \tag{4}$$

where h_j is the (normalized) number of intersections measured with width between x_j and x_{j+1}.

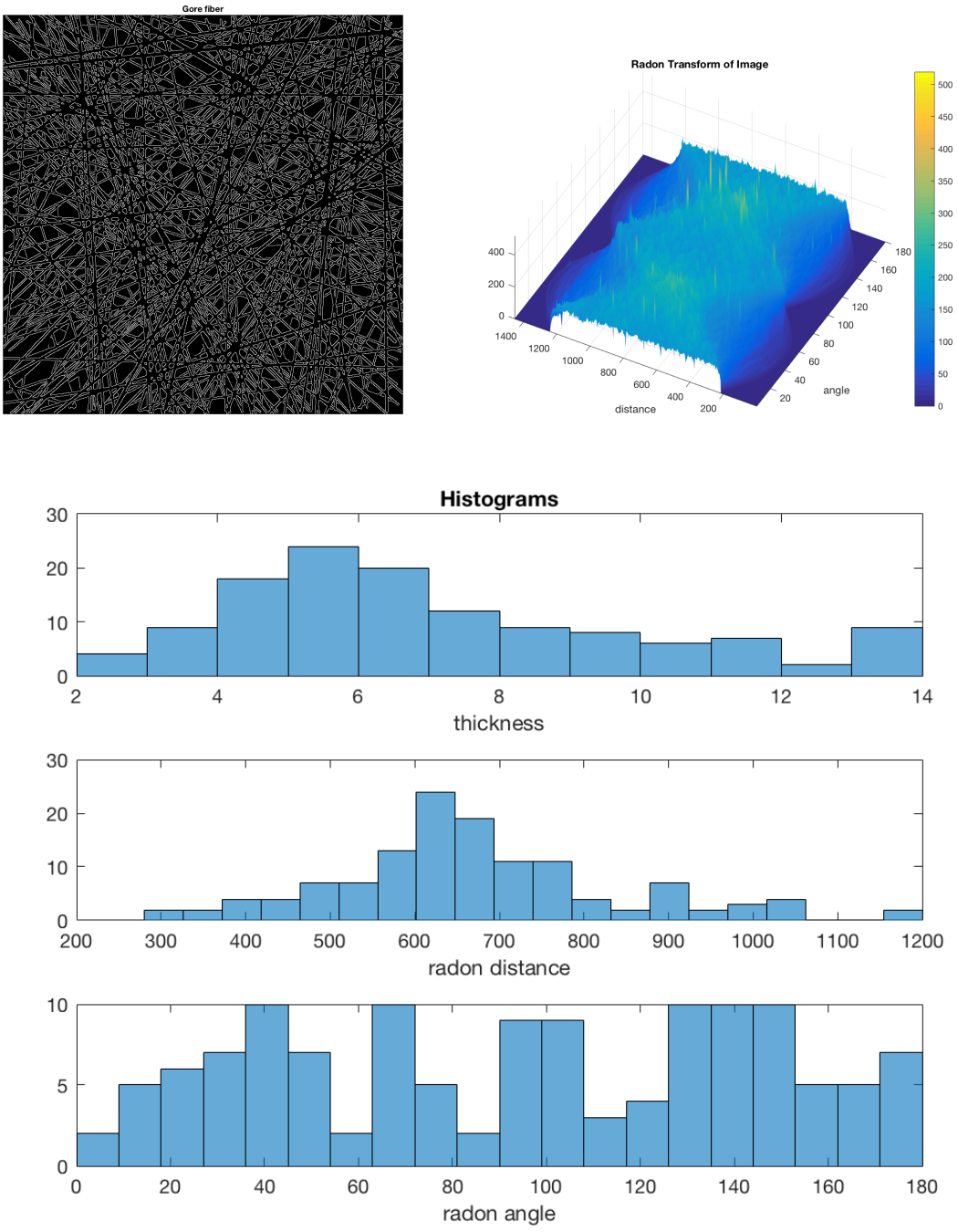

Figure 5: Actual fiber image processing using Radon transform. With distribution of fiber size.

Then, since $f_Y(y)$ is known, we can calculate the product distribution as:

$$f_h(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(\frac{z}{x}) \frac{1}{|x|} dx \tag{5}$$

$$f_Y(y) = \frac{2}{\pi} \frac{1}{\sqrt{1-y^2}}, -1 < y < 1, \quad 0 \text{ elsewhere} \tag{6}$$

$$f_h(z) \approx \hat{f}_h(z) = \frac{2}{\pi} \left(\int_z^{x_i} h_i \frac{1}{\sqrt{x^2 - z^2}} dx + \sum_{\substack{j \\ z<x_j}} \int_{x_j}^{x_{j+1}} h_j \frac{1}{\sqrt{x^2 - z^2}} dx \right) \tag{7}$$

where i is the smallest index such that $x_i > z$. Evaluating the integral, we get

$$\hat{f}_h(z) = \frac{2}{\pi} \left(h_i \left(\log(\sqrt{x_i^2 - z^2} + x_i) - \log(z) \right) + \sum_{\substack{j \\ z<x_j}} h_j \left(\log \left(\frac{\sqrt{x_{j+1}^2 - z^2} + x_{j+1}}{\sqrt{x_j^2 - z^2} + x_j} \right) \right) \right) \tag{8}$$

For the following experiments, we used a dataset of fiber images generated via simulation by Gore with a known distribution. Each image was taken at four magnification levels, 1x, 2x, 4x, and 8x, zooming in on the same simulated patch of material. Intersection widths were sampled over every row in the image to form a combined distribution. Due to a high risk of error in measurement, intersections with a width of less than three pixels were not considered. Below, we compare the results of our method to the known ground truth distribution.

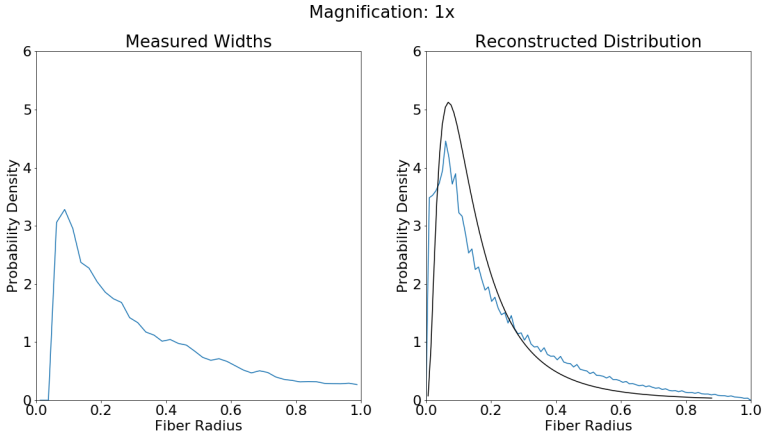

Figure 7: Distribution at 1x magnification. In black is shown the ground truth distribution the dataset was sampled from.

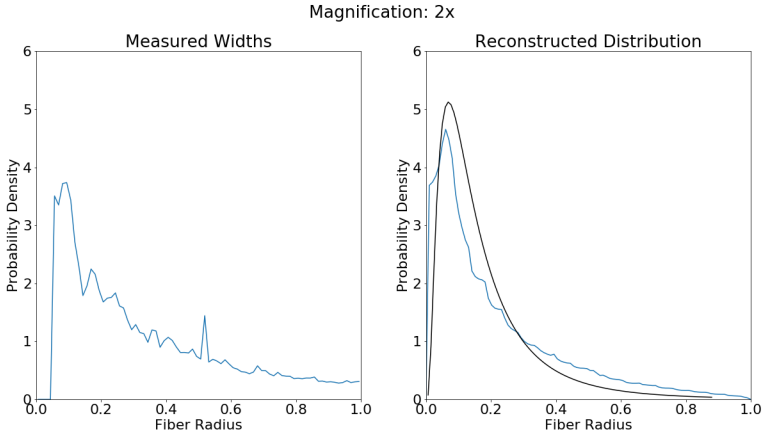

Figure 8: Distribution at 2x magnification. In black is shown the ground truth distribution the dataset was sampled from.

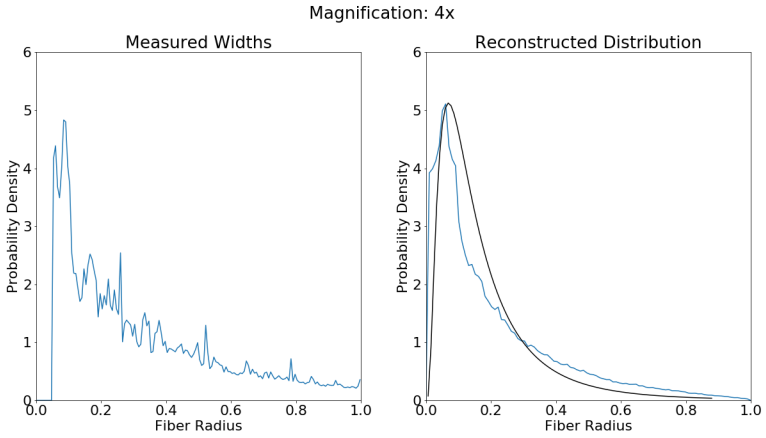

Figure 9: Distribution at 4x magnification. In black is shown the ground truth distribution the dataset was sampled from.

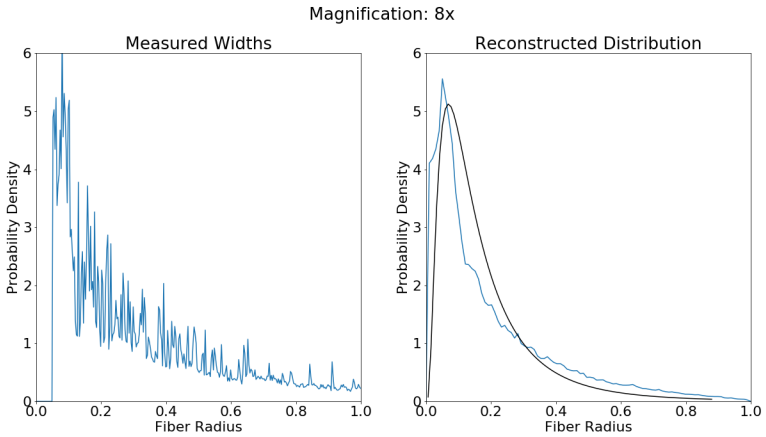

Figure 10: Distribution at 8x magnification. In black is shown the ground truth distribution the dataset was sampled from.

We observe that the distributions closely approximate the ground truth. However, there is still some bias in the distribution, as it does not skew right heavily enough. One hypothesis for why this might be the case is that, due to the sampling of every row in the image, fibers at a larger angle are sampled more times, leading to a bias in measurement. Another is that low diameter fibers are difficult to resolve clearly, and are thus underrepresented.

6 Conclusion

Several algorithms have been studied for the image processing of fibrous materials. The Radon transform and Hough transform investigation shows promising results for real fiber image processing. The machine learning method is promising too and it's certainly worth looking into to generate some real results after sufficient data being generated and processed.

Acknowledgements

The research team thanks Zhenyu He, Marina Chugunova, Sunil Dhar, David Edwards, and Richard Moore for helpful discussions, and the W.L. Gore and Associates, MPI ,NSF, NJIT for their support.

References

[1] B. Pourdeyhimi and R. Dent (1999) Measuring Fiber Diameter Distribution in Nonwovens *Textile Res.* 69(4), 233–236

[2] Carsten Steger (1998) An Unbiased Detector of Curvilinear Structures *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(2), 113–125

[3] Richard O. Duda, Peter E. Hart (1971) Use of the Hough Transformation to Detect Lines and Curves in Pictures *Artificial Intelligence Center*

[4] D.H. Ballard (1981) Generalizing the Hough Transform to Detect Arbitrary Shapes *Pattern Recognition*, 13(2), 111–122

[5] Elias M. Stein, Rami Shakarchi (2003) Fourier Analysis: An Introduction *Princeton University Press*