FiberFind: Machine learning-based segmentation and identification of individual fibers in µCT images of fibrous media

Andreas Grießer*[1], Rolf Westerteiger *[2], Christian Wagner*[3], and Andreas Wiegmann*[4]

*Math2Market GmbH, Germany

Keywords: x-ray tomography, segmentation, machine learning

Summary: A machine learning-based approach to segment curved fibers in a binarized µCT image is presented. A neural network has been trained to label the center curve of each fiber using synthetic training data generated with the GeoDict software. The analysis of these centerlines provides an analytical representation of individual fibers, allowing for analysis of diameter, orientation distribution, curliness, and more.

1. INTRODUCTION[1]

In recent years, machine learning techniques have been successfully applied to the segmentation of objects in 2D images. A major challenge in adapting these approaches to 3D µCT images is the difficulty of obtaining sufficient training data for the neural network. In most 2D applications, these ground truth segmentations are produced by human operators. However, this is prohibitively time-consuming in the 3D domain,

In this work, we present an approach to segmenting individual fibers in µCT images of fibrous material samples. For this purpose, we generate synthetic training data for which the desired segmentation outcome is known. This is a fully automated process in which no user interaction is needed to generate an arbitrary number of training structures. A neural network is trained on millions of data points and subsequently applied to real µCT images in order to identify the individual fibers.

We apply this method to a nonwoven material sample containing curved fibers and use the analytical fiber representations so obtained to analyze local orientation distribution and curliness.

2. ALGORITHM

Fibers in µCT images are often difficult to segment because, after thresholding, they are frequently fused together at contact points. In this situation, simply analyzing the connected components results in a severe under-segmentation of the fibers.

Our approach is based on the idea of removing these contact points by shrinking each fiber down to its centerline, represented as a thin, contiguous curve of voxels. For isolated fibers, the resulting image resembles a morphological skeleton. In the presence of contact points, however, our approach is topology preserving and can separate neighboring fibers which the skeleton fails to do.

A neural network is trained to identify the voxels belonging to the centerline curve. Given a binarized (thresholded) µCT image, it classifies each voxel as belonging to a centerline (output 1) or not (output 0). By keeping only those voxels classified as binary 1, a new binary image is obtained where the fibers have been shrunk to a thin centerline, removing most of the contact points. The connected component analysis of this image ideally results in a single component per fiber.

The network architecture is fully convolutional, producing a dense binary classification of each individual input voxel [1]. On the resulting centerline image, two different postprocessing steps are performed.
First, a watershed algorithm is applied on the original binary image, using unique labels for the found

[1] e-mail: andreas.griesser@math2market.de

[2] e-mail: rolf.westerteiger@math2market.de

[3] e-mail: christian.wagner@math2market.de

[4] e-mail: andreas.wiegmann@math2market.de

centerline components as seeds. This results in a labelled image in which each fiber voxel is assigned a label corresponding to the original fiber.

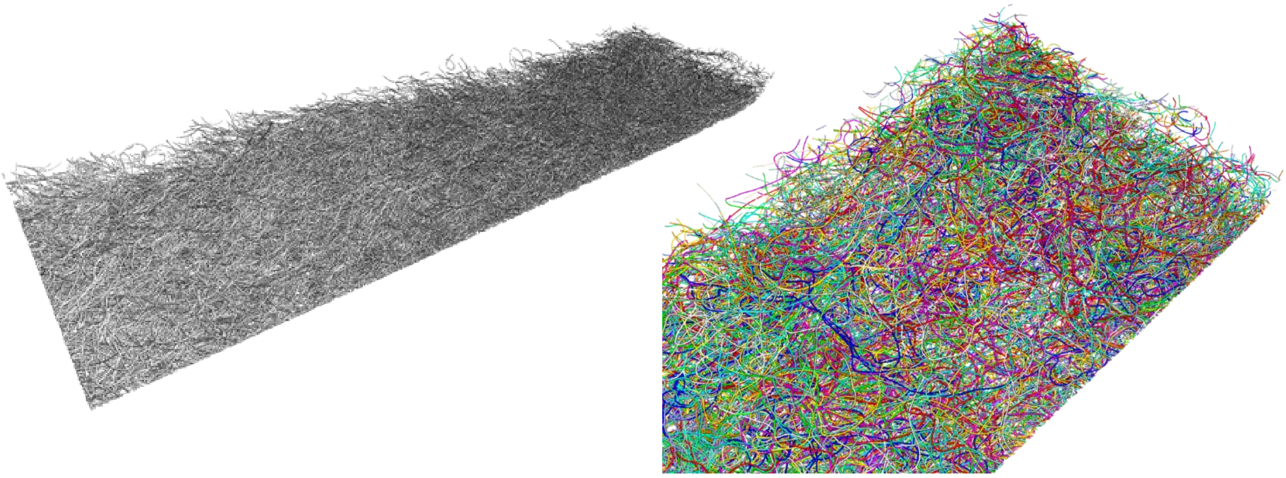
Second, the centerline voxel curves are traced to obtain a piecewise linear analytical representation of the fiber's trajectory in the sample space. Then, the Euclidean Distance Map is sampled along the found trajectory in order to estimate the fiber diameter. At this point, a vectorized representation of all fibers in the sample has been obtained. Using this representation, statistical properties such as the distributions of fiber length (for fibers fully contained in the scan), diameters, orientations, and curliness can be computed with ease.

3. TRAINING DATA GENERATION & TRAINING PROCESS

The structure generator in FiberGeo is used to generate the synthetic training data. FiberGeo is a module of the GeoDict software package. This module can generate 3D models of a large variety of fibrous materials in a stochastic process. Given a set of statistical geometric properties such as porosity, fiber length, diameter and orientation distributions, FiberGeo generates closely matching synthetic material samples in the form of binary voxel images. For these images, the centerlines and positions of all fibers are known and can therefore be used to train the neural network. The training process takes approximately 7 days on a single consumer-grade Nvidia GeForce GTX 1080 GPU.

4. APPLICATION & RESULTS

The algorithm was applied to multiple CT scans. For example, one scan of a nonwoven material has a size of 15,619 x 4,032 x 1,796 voxels. The size of the scan required a fast and memory efficient implementation of the analysis. The final version took one 1 day to extract all fibers and create analytic representations. The trajectories of the resulting identified fibers were visually inspected to verify if any fibers were wrongly connected. Also checked was if the fibers where connected to the boundaries of subdomains smaller than the expected fiber length. These inspections determined that the fiber identification results where correct. From the individual fibers it was then possible to obtain advanced properties, such as length, orientation, curvature and curliness. The same algorithm was successfully applied to a scan of glass-fiber reinforced plastic.



(a) **(b)**

Figure 1: (a) Complete nonwoven sample. (b) Portion of the sample showing individual fibers (color coded).

References

[1] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).