Individual Yarn Fibre Extraction from Micro CT: Multilevel Machine Learning Approach

Petr Henyš[a], Lukáš Čapek[a]

[a]*Technical University of Liberec - Textile Faculty, Department of Technologies and Structures, Studentská 1402/2, 461 17 Liberec, Czech Republic*

Abstract

The internal structure and mechanics of the fibre materials, such as yarn or woven textile, are highly complex. Exploring the fibre structure is an essential step in material engineering either from the experimental or computational point of view. In this study, a new method to extract geometrical and morphological parameters of fibre structures is proposed. The method benefits from standard image analysis and machine learning technique to efficiently extract fibre segments from microcomputer tomography data. The proposed algorithm is tested on the yarn and woven textile materials with different resolution and quality. The developed method can extract the individual fibres with varying accuracy from 73–100% with processing time 2–5s on the tested samples.

Keywords: fibre extraction, multilevel tracing, micro CT, fibre morphology, Gaussian Mixture, computer model

1. Introduction

Extracting the geometrical and topological features of complex fibrous structures, e.g. yarn, from micro-structural tomographic data is challenging. Hence, understanding the internal structure of yarn is a crucial step for determining how macro-mechanical parameters are influenced by inter-fibre mechanics and other physical effects. Using an adequately defined micro-structural model of yarn, a computational model can be established to find out the yarn micro-mechanics concerning full-resolution models [1, 2] or more efficient multi-scale models [3, 4, 5]. Nevertheless, prior to any computational modelling, an accurate geometry representation of the yarn structure must be provided. A few studies have focused on the tracing the individual fibres. Geisselmann et al. used a combination of the distance transform and skeletonisation methods to separate fibres individually. However, owing to noise and irregularity in the three-dimensional (3D) data, the fibres might not be continuous and the skeletonisation algorithm may generate the so-called 'H' connectivity (Figure 1) [6, 7]. The challenging step is to properly connect the disconnected part to obtain the original fibres. Hence, a suitable criterion must be met to connect the appropriate segments. Geiselmann et al. proposed the use of stochastic models wherein the angle and length of the interconnecting segment were utilised to control the actual connection probability [6, 7]. Hu et al. used a multilevel tracing algorithm that simplified the decision criteria of pair selection [8]. Another method, which enjoys the benefits of popular geometrical structures (e.g. cylindrical structures), employed the Hessian eigenvalues to extract fibres [9]. Jerome et al. proposed another skeleton-based method suitable for low-density materials [10]. Viguié et al. used a local orientation map combined with dilatation operations to extract fibres and their contacts [11]. Sencu et al proposed an extraction algorithm, which used a combination of standard image methods to filter raw CT data and applied an enhanced Bayesian filter and Bayeisan inference to track fibres in unidirectional composite [12]. Another methods are based on searching fibre centres according to a certain template with suitable correlation metric [13, 14, 15]. Although fibre-like structures are specific structures from the perspective of segmentation algorithms, geometrical extraction can be significantly simplified by reusing several specific properties provided by a certain fibre model. The simplest extraction can be observed in long and almost homogeneously oriented fibres. This is the case of a wide range of composite fibres

*Corresponding author
Email address: lukas.capek@tul.cz (Lukáš Čapek)

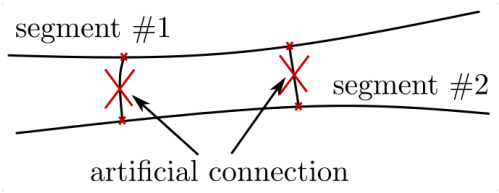

Figure 1: H-type connection occurring during skeletonisation of a noisy structure.

and algorithms used the regularity and straight fibre assumption were proposed (a more enhanced approach has been presented in previous research [16, 17]). The straight fibres allow us to extract centres from individual computer tomography (CT) stack cross sections by employing regular Hough-like methods. However, even in these simple cases, the computational cost can be minimised and more effective methods should be explored. Most algorithms use at least the basic morphological methods to extract useful information from 3D images. A very effective method can be observed when combining the distance map and skeletonisation algorithms, which are used to extract the fibre centres [7]. Enhanced methods, e.g. Hessian eigenvalues-based method, can be used for further extraction of structure-based information [9]. In recent years, deep neural network variants have been designed to extract fibres from CT [18, 19]. Nevertheless, deep learning methods require effective training, which is difficult to achieve for high dimensional CT data in a reasonable time. Herein, we propose a new method based on an efficient combination of well established 3D image analysis and a machine learning technique to efficiently extract individual fibres from micro-CT data. Our method is able to:

- handle spatially complex fibres with many contacts (not possible or untested with methods in [16, 17, 12])
- extract fibres in a reasonable time (compared to [7, 6]),
- use more robust pairing criteria (than used in [8]), and
- extract fibres without time consuming training (compared to deep learning models in [18, 19]).

Proposed method is a fusion of methods introduced in [6, 8], but with enhanced speed and robustness thanks to effective spatial partitioning of space and machine learning method to check the fibre pairing validity.

2. Materials and Methods

The overall working-flow of the image analysis on the CT data is shown in Figure 2. The analysis is composed of optimal de-noising of the raw data CT up to final step consisting of re-sampling and proper tracking of the fibre segments. The steps are discussed in detail in the following subsections.

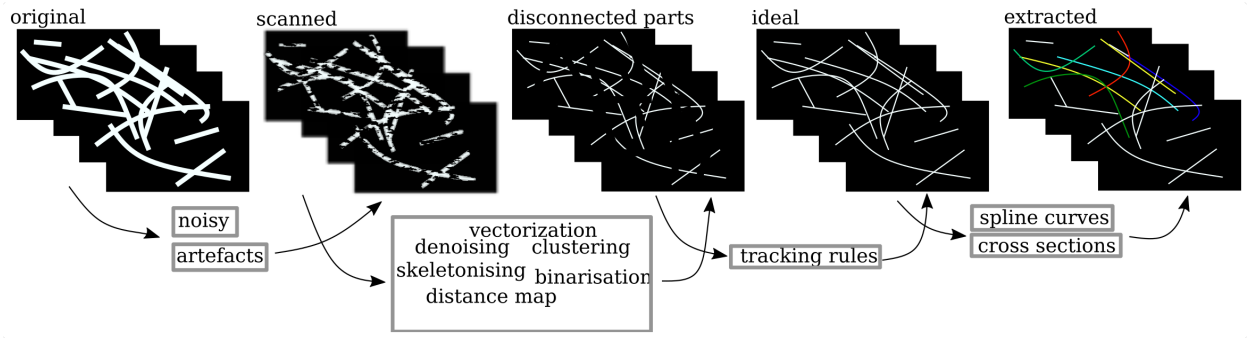

Figure 2: The work-flow of extraction of fibres from microCT data.

2.1. Tested Textile Materials

The tested materials represent planar textile named **TEX-2DW** and yarn structure named **TEX-Y**. The planar textile is made of polyamide multifilament yarns with fineness 167 [dtex], yarn twist 420 [1/m], and warp/weft sets 16/22 [1/cm]. The second structure of yarn is made of polyester with a fibre diameter of ~12μm, yarn diameter of ~131μm and fineness of 25 [tex]. Both structures are shown in Figure 3. The series of images are represented by an

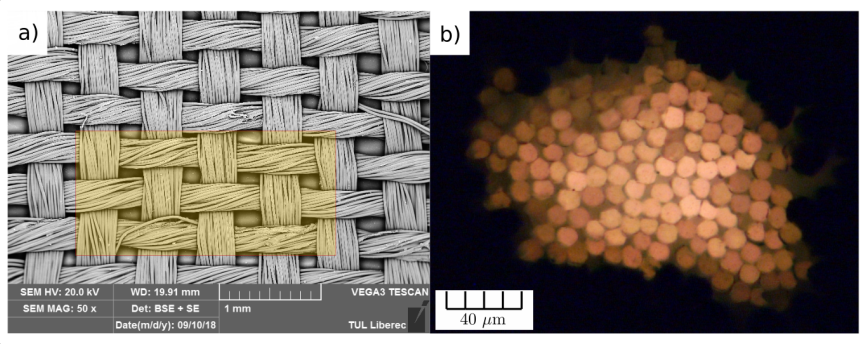

Figure 3: Electron microscopy slice of woven structure **TEX-2DW** with region of interest(yellow window) used for 3D scanning (a). A slice cut of yarn **TEX-Y** with visible individual fibre cross-sections made by light microscopy (b).

array with orthogonal indices, where each member represents a grey value in range 0–255. An example of raw grey structure slices is shown in Figure 4.

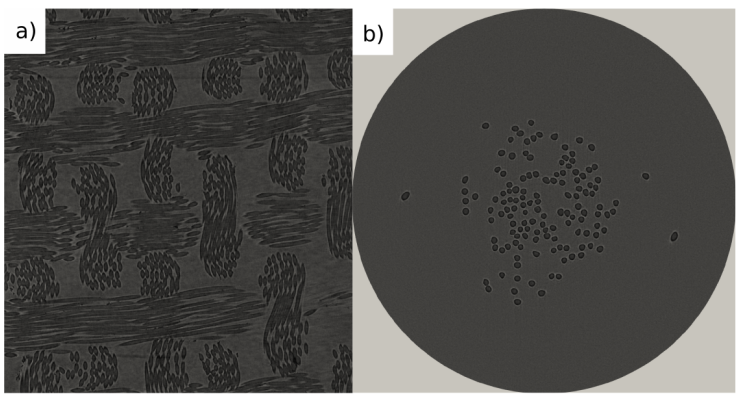

Figure 4: Gray slice (z axis) example of **TEX-2DW** (a) and **TEX-Y** (b).

2.2. Scanner Setting

The micro CT scanner Rigaku nano3DX was used with tube voltage in range from 20 to 50 kV and current up to 30 mA. The dimension of the **TEX-Y** is $[1427 \times 1442 \times 1250]$ with voxel resolution $[0.528 \times 0.528 \times 0.528]$ μm. The **TEX-2DW** has dimension $[204 \times 1201 \times 551]$ with voxel resolution $[1 \times 1 \times 1]$ μm.

2.3. Image Preprocessing

The resultant CT stacks were converted to 3D array with metadata regarding to resolution and origins. The noise was removed by median filter with spherical structure element. Further, the binarisation based on the Otsu's method was performed. Once the binarised images were obtained, the Euclidean distance map was computed. It benefits from the circular cross-section of individual fibres. At fibre centre, a distance is minimal with respect to the background and peak occurs. Optimally, the peak has the half value of fibre diameter, but due to a noise, the peak filter works in

range 4.7–6.3 micro meters for **TEX-Y** and 4.5–7 for the woven structure **TEX-2DW**. The fibre diameter is known either from manufacturer (see subsection 2.1) or can be estimated from distance map histogram. Moreover, the crossing fibres can be separated by the distance peaks and thus no additional step is required (Figure 5). The new binarised data is obtained from filter the distance peaks defined by range above. Due to small gaps in the peaks, the morphological dilatation with small structuring element was used to close such gaps. The final step is to apply the skeletonisation on binary array and get one-pixel centres of fibre parts. In order to remove the 'H' like connections and another crossings, the pixel were tested for the neighbourhoods by 26-connectivity and those with more than two neighbourhoods were removed. Finally, the disconnected parts were labelled and sorted ascendantly according to fibre segment length.

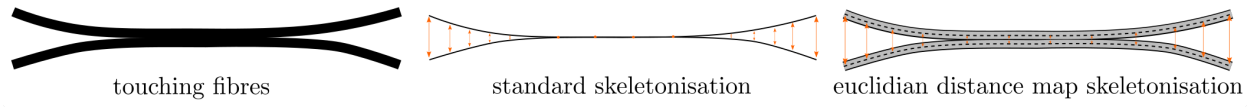

Figure 5: A regular skeletonisation issue (middle) with two fibres (left). The distance map disconnecting two fibre segments

2.4. Segment Vectorisation

For further work with disconnected discrete parts of the fibres, the transformation into continuous representation is useful. Let call the disconnected parts of fibres as 'segments'. Each segment represented a structure containing the unsorted number of pixels, but with known ends, is approximated by the b-spline curve (Figure 6). Once two candidates for the merging are found, the b-spline curve is also used to connect them by approximating the missing coordinates between the candidate segments. The spline approximation quality was carefully investigated in dependence on the smoothing factor s (for details, see the Scipy library manual [20]). The smoothing factor influences a compromise between the accuracy of interpolation and smoothness. The smoothing parameter in the range 0.01-0.05 was usually given good b-spline approximations. The spline is of cubic order and also the first derivative is used to get a segment direction. The more details of how the spline approximation is constructed can be found in Scipy library documentation [20].

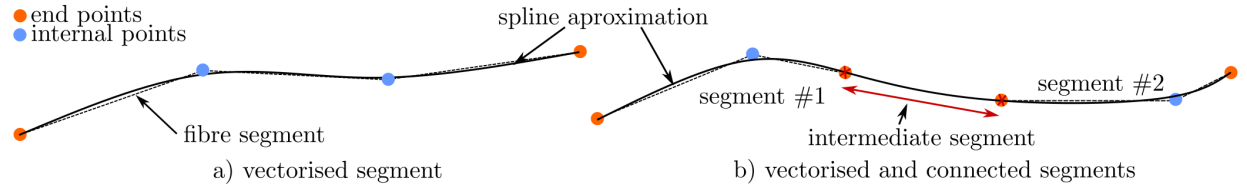

Figure 6: Fibre segment with cubic spline approximation (a). Merging of two segments by spline curve with smooth transition (b).

2.5. Fibre Reconstruction

Each segment contains two end-points (start/end). The initial connection of two segments is defined by a straight line, which is called intermediate segment. Segments can be connected only by their end-points. A potential pair of two segments that belong to a fibre is searched based on geometrical relations inspired by work of Geiselmann [7, 6]. Consider two segments i and j in a pair. We can define Euclidean length l_{ij} of intermediate segment and angles α_{ij}, α_{ji} (see Figure 7). Introducing the weighting factor w_r and maximal values l_{max}, α_{max}, we can write down the criterion function as weighted sum of normalized partial members:

$$c_{ij} = (1 - w_r)\frac{l_{ij}}{l_{max}} + w_r\frac{\alpha_{ij} + \alpha_{ji}}{\alpha_{max}} < 1 \tag{1}$$

Every pair of segments satisfying the above criteria is taken into account as a potential candidate. The final pair is selected according the smallest value of c_{ij} and should also pass the back checking defined later. Increasing the maximal distance between fibre ends leads to decreasing of the maximal angles, so there is a clear geometrical relation [7]. But we found that separating both criterion (distance, angles) can provide more flexibility in some cases.

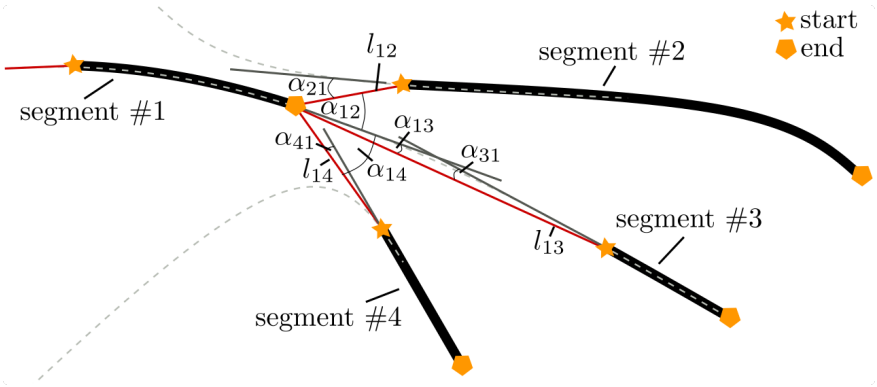

Figure 7: Geometrical quantities defined between end-points of segments. The red line represents an intermediate segment. Dashed line represents true fibre trajectory.

2.6. Back Checking

Once the two segments are merged, the newly constructed segment should be evaluated whether it corresponds with original CT data. The spline curve of the intermediate segment is discretised onto the original CT data space and dilated with small structuring element sphere. The mean of grey values in CT data corresponding to newly occurred intermediate segment (see Figure 6) was evaluated with help of clustering algorithm [21]. Firstly, we let a Gaussian clustering algorithm detects two density clusters in CT data. The clusters separated the CT data on grey density belongs to fibre structure and background noise. Secondly, we computed a probability that a newly occurred intermediate segments belongs to a one of the cluster (see the details in clustering algorithm implementation in [21]). We found a that a probability level 85 % was optimal for tested examples in this study.

2.7. Algorithms and Implementation Details

The image algorithms work usually with large data collections and in 3D it is even more demanding issue. In the proposed algorithms in Figure 8 main steps required for the successfully extraction of fibres are introduced. The input required by algorithm composed of distance map field and raw CT data. Both arrays are encoded in hdf5 format [22] with uint8/float64 data type. The vectorisation is one of the most computationally intensive task, as it requires to compute filtering (if required), binarisation and skeletonisation. The initial control parameters are set by the user and are defined by following python dictionary:

- "l_{max}": [...,...,...] a list of maximal distance parameters
- "α_{max}": [...,...,...] a list of alpha angles
- "w_r": [...,...,...] a list of weighting constants

From a programming language point of view, a segment is represented by the object allowing recursive merging of segments and respecting the parent-child relation. As the segment connectivity is defined by spatial coordinates of the segment end-points, neighbourhood lookup (Kd tree [20]) is constructed before main loop begins in order to significantly speed up searching of segment pairs. At current point, the pairing of segments runs on the two queues named as **finished** and **unfinished**. Once the pair of two matching segments is found and matched back-checking criteria, they are merged and partial segments are drop out from queue **unfinished** and the new segment is pushed to queue named **finished**. If necessary the matching criteria can be adapted during each level. The main code is developed in python language with critical parts (graph construction, searching, merging and b-spline interpolation) written in C++. The code is written in objective manner.

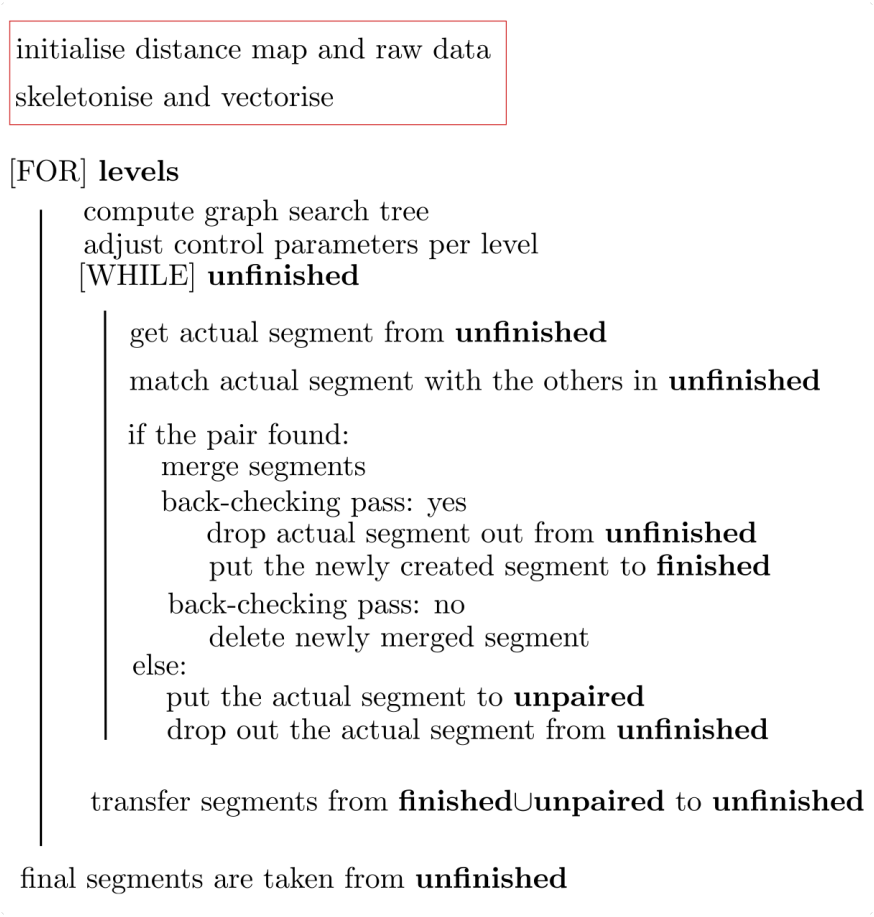

Figure 8: Algorithm scheme of fibre segmentation. The **unfinished**, **finished**, **unpaired** are queries.

3. Results & Discussion

The model parameters $l_{max}, \alpha_{max}, w_r$ are experimentally estimated based on the parametric approach and careful investigating the initial and level-based potentially matched pair of segments. The mean values of grey data for both models were estimated by Gaussian mixed clustering (see the Figure 9). The clustering algorithm is used to decide whether the newly created segment corresponds to grey values. The pairs which do not pass the quality criteria are excluded. It must be pointed out that proposed algorithm could be more improved in control parameter evaluation. The current model relies on the simple deterministic criteria, which need usually a user initial set-up and due to multilevel approach can be adapted during the level iteration. Nevertheless, the unique decision on the best suited segment pair is not always detectable (especially in case of noisy CT data) and mismatches occur. Although the mismatches are detected by evaluation of the newly created segment which leads to delete the newly created segment, still some mismatched segments propagate up to end of level iteration or some potentially connectible segments with long proximity are not correctly detected. Using more advanced criteria can help to improve the decision sensitivity and should be supported by probabilistic approach either proposed in [7] or some another approach such machine learning classifier or Fuzzy logic [23]. The optimal model parameters are defined in the Table 1. To decide whether to start new level, the algorithm checks the number of new pairs for the previous set of control parameters and if it is non-zero, it continues to another level with the same parameters or if the number of new pairs is zero, the algorithm increases the control parameters and starts new pairing level. The control parameters can be also controlled interactively, but it is not efficient. Nevertheless in some cases the manual correction may be required or the weighting parameter w_r can be corrected to get more fine control. Optimal values of parameters $\alpha_{ij}, \alpha_{ji}, l_{ij}$ should be as smallest

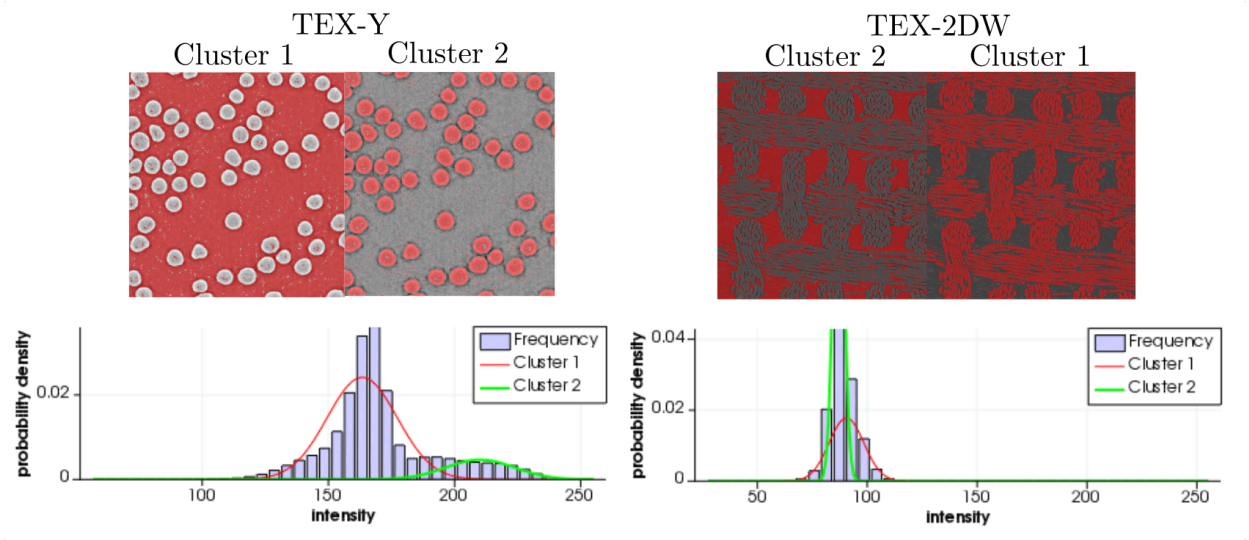

Figure 9: Gaussian mixture cluster classification [21, 24] for woven textile and yarn CT data. Two clusters of grey values were identified for each materials. The red colour corresponds to a mask of clustered data. The following density clusters corresponding to fibre materials and background noise were identified: **TEX-Y**[Cluster 1 (165±12.5); Cluster 2 (210±13.8)]; **TEX-2DW**:[Cluster 1 (90±10.5); Cluster 2 (85±5.0)]

Table 1: Optimal control parameters and number of levels for both models. The weighting parameter w_r was set to 0.5. There is a couple of numbers: the first means the l_{max} [mm] and second means α_{max}. [°]

	number of levels						
TEX-2DW	0.01; 10	0.01; 10	0.03; 30	0.03; 30	0.05; 40	0.08; 40	0.1; 40
TEX-Y	0.01; 50	0.01; 50	0.01; 50	0.015; 50	0.015; 50	–	–

as possible, which indicates that two fibre segments could potentially belong to a fibre. Nevertheless, in some cases the fibre segment ends are close and excentered a bit. The angles can be close to 90°, while distance can be still small. In such case, the user's careful intervention can be required. The control parameters are yet adjusted by a user, which allows us to interactively control segment pairing each level, but on the other hand it requires a certain manual tuning of those parameters by a user to get valid results. The automatic optimal selection of control parameters and their levels is intensively developed by authors and will be a part of future study.

The segmentation results of the yarn model **TEX-Y** are shown in Figure 10. With given level-controlled parameters, the resultant fibres are correctly identified with 100 % success. There are 136 fibres at total (counted from slice cuts). The mean radius of fibres is 0.0051 mm. The initial number of fibre segments is shown in the most left in Figure 10. The results of fibre extraction for high noise and corrupted CT data for model **TEX-2DW** are shown in Figure 11. The only 73% of fibres were correctly identified (at total there are 255 fibres counted by careful looking at the CT data). Lower accuracy is caused by corrupted segments, which cannot be uniquely recognised by the deterministic approach used. The fibre mean radius of **TEX-2DW** is 0.0057 mm. The multilevel strategy allows us to adaptively change pairing criteria (one of the main difference from [8]), which may lead to enhanced segmentation results. The extracting accuracy can be further improved by clever using of morphological operations and more advanced filtering techniques. Nevertheless, the most critical part is to have sufficient quality of CT data, which is seen in our results.

The computational sources needed for the fibre segmentation are relative low (136 fibres identified in 2s and 186 fibres identified in 5s) in comparison with stochastic algorithm (83 minutes for 903 fibres) proposed in [7]. Nevertheless, it can depend on fibre-structure complexity and segment-ends proximity. The critical part is the quick nearest-neighbour lookup, which needs to be implemented efficiently by building optimal graph such KD-tree algorithm [25]. Euclidean distance and skeletonisation algorithms are already implemented efficiently in library [20]. They are not usually the performance bottleneck for moderate size CT.

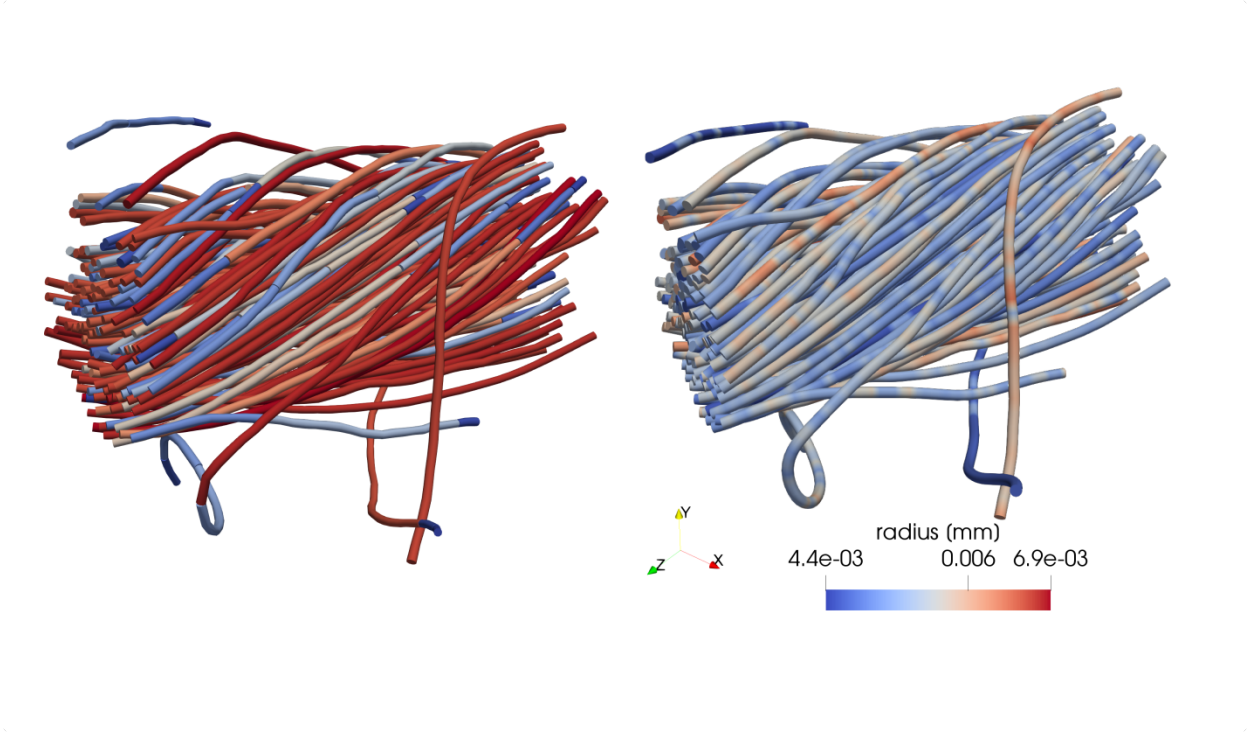

Figure 10: The initial fibre segments before algorithm applying(left). Segmented yarn with 100% success in all available fibre segments(right).

The proposed algorithm has not been tested on large CT data set containing thousands of fibres. This fact can be considered as a shortcoming of this study because Euclidean distance and skeletonisation algorithms are computer memory-intensive. Hence they may become a performance bottleneck for a large data set.

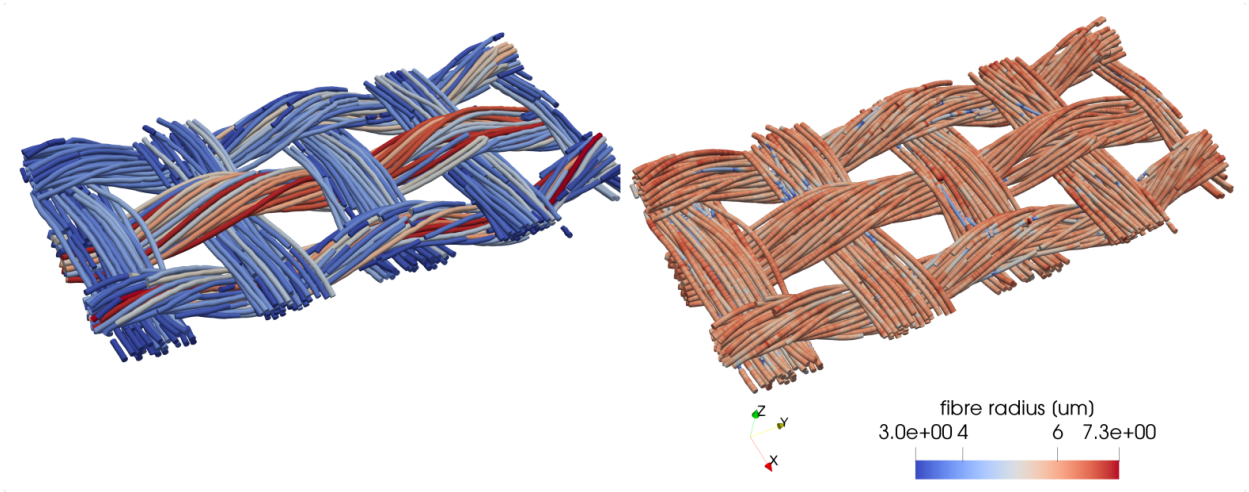

Figure 11: Extracted woven fibres and their properties from highly noisy and corrupted micro-CT data: initial fibre segments (left); final fibre segments with fibre radius is shown (right).

4. Conclusion

Building a complex and physically relevant model is of great interest in the scientific community. The fibre-like structures can be seen in many materials. The inspection of their physics is an important task. Although the computer tomography can provide an important view of the structure of fibres, often more than is required, especially in the computational models, where the geometry is of great interest. Due to the complexity of fibre structures with many contact bonds and overlapping, the proper individual fibre identification is a challenging task. Although a few studies are dealing with fibre extraction, a consistent family of methods is not well developed. The method relies on the multilevel strategy, where at each level pairs of segments are searched so it matches the given control criteria. The method is based on the common morphological image operations enriched by machine learning classifier. The method was tested on the two fibre structure samples and shown high efficiency and good identification accuracy.

5. Conflict of interest statement

The authors declare no conflict of interest.

References

[1] D. Durville, Finite element simulation of textile materials at the fiber scale, arXiv preprint arXiv:0912.1268 (2009).

[2] N. Naouar, E. Vidal-Sallé, J. Schneider, E. Maire, P. Boisse, Meso-scale fe analyses of textile composite reinforcement deformation based on x-ray computed tomography, Composite Structures 116 (2014) 165 – 176.

[3] Z. Ullah, L. Kaczmarczyk, S. Grammatikos, M. Evernden, C. Pearce, Multi-scale computational homogenisation to predict the long-term durability of composite structures, Computers and Structures 181 (2017) 21 – 31. UK Association of Computational Mechanics.

[4] X.-Y. Zhou, P. Gosling, C. Pearce, Z. Ullah, L. Kaczmarczyk, Perturbation-based stochastic multi-scale computational homogenization method for woven textile composites, International Journal of Solids and Structures 80 (2016) 368 – 380.

[5] A. Vanaerschot, B. N. Cox, S. V. Lomov, D. Vandepitte, Multi-scale modelling strategy for textile composites based on stochastic reinforcement geometry, Computer Methods in Applied Mechanics and Engineering 310 (2016) 906 – 934.

[6] G. Gaiselmann, D. Froning, C. Tötzke, C. Quick, I. Manke, W. Lehnert, V. Schmidt, Stochastic 3d modeling of non-woven materials with wet-proofing agent, International Journal of Hydrogen Energy 38 (2013) 8448–8460.

[7] G. Gaiselmann, I. Manke, W. Lehnert, V. Schmidt, Extraction of curved fibers from 3d data, in: Forum Bildverarbeitung 2012, KIT Scientific Publishing, p. 119.

[8] X. Huang, D. Wen, Y. Zhao, Q. Wang, W. Zhou, D. Deng, Skeleton-based tracing of curved fibers from 3d x-ray microtomographic imaging, Results in physics 6 (2016) 170–177.

[9] M. Teßmann, S. Mohr, S. Gayetskyy, U. Haßler, R. Hanke, G. Greiner, Automatic determination of fiber-length distribution in composite material using 3d ct data, EURASIP journal on advances in signal processing 2010 (2010) 545030.

[10] J. Lux, Automatic segmentation and structural characterization of low density fibreboards, Image Analysis & Stereology 32 (2013) 13–25.

[11] J. Viguié, P. Latil, L. Orgéas, P. Dumont, S. R. du Roscoat, J.-F. Bloch, C. Marulier, O. Guiraud, Finding fibres and their contacts within 3d images of disordered fibrous media, Composites Science and Technology 89 (2013) 202–210.

[12] R. Sencu, Z. Yang, Y. Wang, P. Withers, C. Rau, A. Parson, C. Soutis, Generation of micro-scale finite element models from synchrotron x-ray ct images for multidirectional carbon fibre reinforced composites, Composites Part A: Applied Science and Manufacturing 91 (2016) 85 – 95.

[13] M. W. Czabaj, M. L. Riccio, W. W. Whitacre, Numerical reconstruction of graphite/epoxy composite microstructure based on sub-micron resolution x-ray computed tomography, Composites Science and Technology 105 (2014) 174–182.

[14] W. Whitacre, M. Czabaj, Automated 3d digital reconstruction of fiber reinforced polymer composites, in: AIAA Guidance, Navigation, and Control Conference, p. 0342.

[15] G. Fang, C. Chen, S. Yuan, S. Meng, J. Liang, Micro-tomography based geometry modeling of three-dimensional braided composites, Applied Composite Materials 25 (2018) 469–483.

[16] M. J. Emerson, K. M. Jespersen, A. B. Dahl, K. Conradsen, L. P. Mikkelsen, Individual fibre segmentation from 3d x-ray computed tomography for characterising the fibre orientation in unidirectional composite materials, Composites Part A: Applied Science and Manufacturing 97 (2017) 83 – 92.

[17] M. J. Emerson, V. A. Dahl, K. Conradsen, L. P. Mikkelsen, A. B. Dahl, Statistical validation of individual fibre segmentation from tomograms and microscopy, Composites Science and Technology 160 (2018) 208 – 215.

[18] T. Konopczyński, T. Kröger, L. Zheng, J. Hesser, Instance segmentation of fibers from low resolution ct scans via 3d deep embedding learning, arXiv preprint arXiv:1901.01034 (2019).

[19] V. Wegmayr, A. Sahin, B. Saemundsson, J. Buhmann, Instance segmentation for the quantification of microplastic fiber images, in: The IEEE Winter Conference on Applications of Computer Vision, pp. 2210–2217.

[20] E. Jones, T. Oliphant, P. Peterson, et al., SciPy: Open source scientific tools for Python, 2001–. [Online; accessed ¡today¿].

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[22] M. Folk, G. Heber, Q. Koziol, E. Pourmal, D. Robinson, An overview of the hdf5 technology suite and its applications, in: Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases, ACM, pp. 36–47.
[23] J. Yen, R. Langari, Fuzzy logic: intelligence, control, and information, volume 1, Prentice Hall Upper Saddle River, NJ, 1999.
[24] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, G. Gerig, User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability, Neuroimage 31 (2006) 1116–1128.
[25] S. Maneewongvatana, D. M. Mount, Analysis of approximate nearest neighbor searching with clustered point sets, Data Structures, Near Neighbor Searches, and Methodology 59 (2002) 105–123.