

Analysing the Structure of Collagen Fibres in SBFSEM Images

Yassar Almutairi¹, Timothy Cootes¹ and Karl Kadler²

¹Imaging Science, The University of Manchester, UK

² Wellcome Trust Centre for Cell-Matrix Research, The University of Manchester, UK

yassar.almutairi@postgrad.manchester.ac.uk

{timothy.f.cootes, karl.kadler}@manchester.ac.uk

Abstract

Collagen fibres form important structures in tissue, and are essential for force transmission, scaffolding and cell adhesion. Each fibre is long and thin, and large numbers group together into complex networks of bundles, which are little studied as yet. Serial block-face scanning electron microscopy (SBFSEM) can be used to image tissues containing the fibres, but analysing the images manually is almost impossible - there can be over 30,000 fibres in each image slice, and many hundreds of individual image slices in a volume. We describe a system for automatically identifying and tracking the individual fibres, allowing analysis of their paths, how they form bundles and how individual fibres weave from one bundle to another.

1. Introduction

We describe a system for tracking large numbers of collagen fibres in Serial Block-Face Scanning Electron Microscopy (SBFSEM) images volumes. Collagen fibres are found in many types of connective tissue, often forming a matrix which helps hold the tissue together [11]. In tendons they are formed into rope-like bundles. To study such structures SBFSEM images of tendons can be collected. However, since each may contain tens of thousands of individual collagen fibres, automatic methods are essential to analyse such images.

To learn more about tendon development the exact positions and trajectories of these fibres need to be examined. During different stages of development significant changes occur (see Figure 1). For example, for mouse-tail tendon at embryonic day 15.5 fibres have a circular shape and are organised in small groups. At birth their diameter and radii have increased and their number increased roughly four times [5]. The analysis of these fibres requires tracing thousands of fibres over hundreds of images, which is hugely time consuming to do manually.

Recent technological developments have led to auto-

matic 3D electron microscopes allowing the acquisition of large voxel volumes. These high resolution images are essential in fields such as connectomics, which aims to reconstruct the structures comprehensively using these images. Serial-Sectioning TEM (ssTEM) can produce large 3D datasets to be reconstructed. This technique has a limitation in terms of volume size since it is time-consuming and laborious to collect serial sections of the sample. Thin sections are difficult to acquire and image deformation could be observed between slices which is challenging because of the ambiguity between coordinate transformation and structure changes.

Serial block-face scanning electron microscopy (SBFSEM) involves repeatedly removing a very thin slice (as thin as 25nm) from a sample block (with a diamond knife) then taking an EM image of the remaining face. The individual slices can be put together into a volume [4] and provide large 3D datasets. This technique allows biologists to study fibre connectivity from the visual inspection of image volume. Figure 2 shows an example of the image volume with reconstructed fibres.

A fully automatic system for locating and tracking fibres is challenging due to the complexity of the networks that the bundles form and the variation of shape of fibres within the image volume.

Particular challenges include:

- The tortuous morphology of fibres.
- Fibres may disappear and appear again due problems in sectioning.
- The large numbers of individual fibres - as many as 20,000 per slice.
- Bundles of fibres may split or merge.
- Some individual fibres may leave one bundle and join another, following significantly different paths to the majority of the fibres in the bundle - these are often

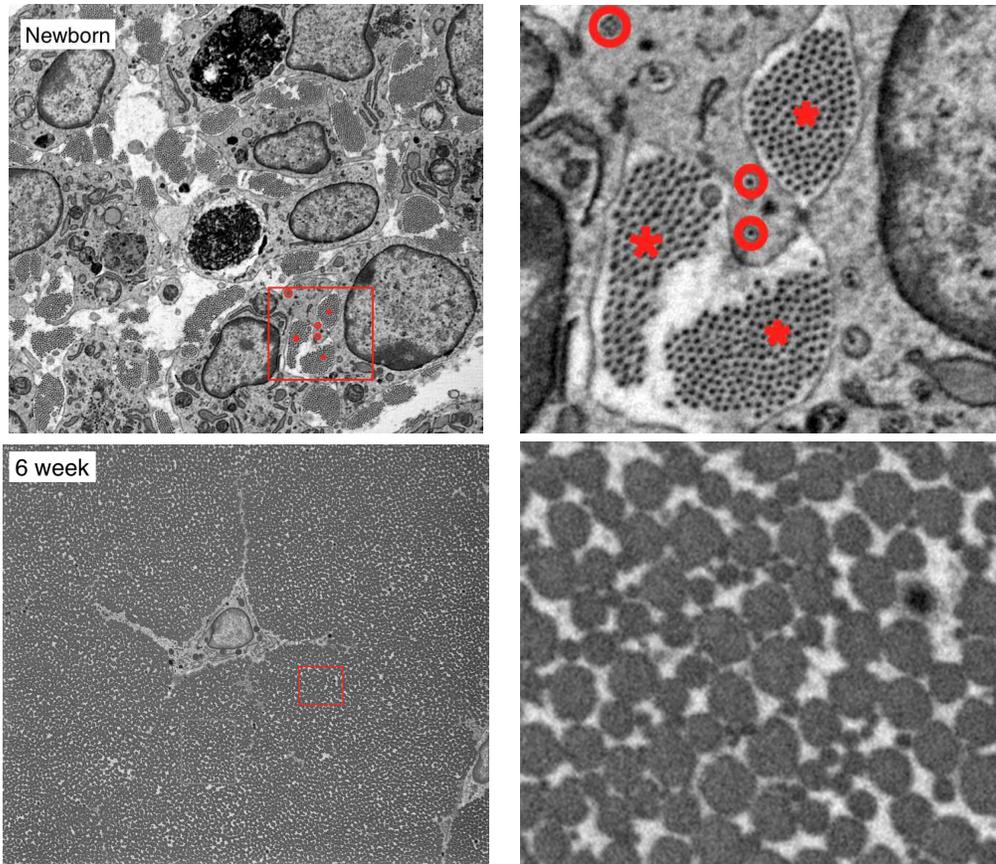


Figure 1. Images of mouse-tail tendon through development. The first row is for newborn where most fibres are grouped into distinct bundles. Stars represent bundles and circles show lone fibres. Second row shows the fibres after 6 weeks where fibres are larger and bundles have merged.

the most interesting to biologists, but are the hardest to track.

The aim of our work is to automate the process of identifying and tracking structures through image volumes, allowing the structures to be measured and visualised in 3D. Fibres appear as small dark blobs as shown in Figure 1. The approach we take is to first detect candidates for fibres in every image, then to link candidates between neighbouring images to form extended fibres. The linking stage can make mistakes, which we attempt to identify and correct in a third stage.

Having tracked all the fibres, we can then identify bundles and lone fibres to study the networks that they form.

2. Related work

Automatic reconstruction methods for ssEM typically focus on obtaining a 2D segmentation of each section. They then match these 2D segments across next slice in the stack [8] [12] [9]. This method relies on the assumption that the initial 2D segments of each slice are good enough for linkage step. Thus the tracking result is dependent on the

detection quality. Some of these methods include indirect penalties by setting a stopping conditions or have rules to ensure that the algorithm converges to the right answers [9]. There are several works are focused on reconstruction neurons from ssEM. These use Markov Random Fields (MRFs) at the pixel level. Such methods are sensitive to the alignment between slices due to anisotropic nature of ssEM [7].

In tracking-by-assignment models, the task is considered as a joint global optimisation problem. It treats every detected object as a potential target by selecting a subset of candidates solution. Each possible solution is given a cost based on how likely that one segment is part of another segment in next frame [6] [2]. Domain knowledge is often used to guide the cost function. Kausler et al. showed that the global optimal assignment can be achieved better over several frames than just two neighbouring frames.

The Random Forest was introduced as a classifier or regressor by Breiman [3] and has been widely adopted because of its simplicity and effectiveness. A Random Forest is an ensemble of decision trees, each trained using random subsets of the full data, and random subsets of the available

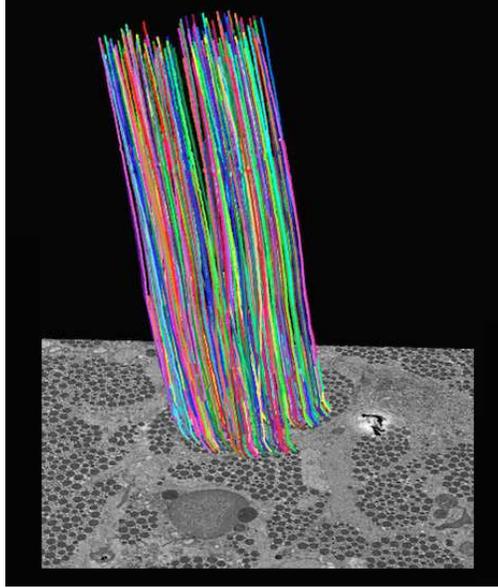


Figure 2. Visualisation of some of the fibres found by the tracking algorithm.

features. This encourages each tree to give an independent estimate of the true answer, so that the uncertainty on the estimates can be reduced by combining results from many such trees.

Andres et al. [10] proposed a method to segment neural tissue using a hierarchical classifier in three stages. In the first stage, a Random Forest classifier is trained on manually segmented samples to validate each voxel. The watershed algorithm is used to evaluate these classifications, which produces over-segmentation. Finally, the segments produced by the watershed algorithm are then classified as correct or incorrect by another trained Random Forest classifier. This approach is not appropriate for our task, as the fibres are small objects in any individual slice - the main challenge is to find them and link each fibre across slices.

Almutairi et al. [1] describe an approach which involved locating individual fibre candidates in each frame with normalised cross correlation (NCC), then linking them using a tracking algorithm to take account of the drift of bundles between slices. We find that this has limitations, including

- (a) NCC does not work well for the non-circular cross sections that occur where fibres are tightly packed (see Fig. 1)
- (b) The linking method proposed can make incorrect assignments when fibres are close

In this paper we propose solutions to these challenges, using Random Forests to locate fibres, and a Kalman Filter tracking stage to identify and correct fibre connection errors.

3. Method

The aim of this work was to build a robust system to assist biologists to extract 3D structure of collagen fibres. We focused on image volumes from SBFSEM, which were gathered so that the image planes are roughly orthogonal to the direction of the fibres of interest - each fibre then appears as a small disk or ellipsoid of radii in the range 2 to 25 pixels. Biologists are more interested in the paths of the fibres than details of their cross-section. Since many fibres are only a few pixels across, and have little visible internal structure (see Fig. 1) it was found to be sufficient to use template matching to identify the centre and approximate radius of each fibre in each image at early stage of tendon development and to use a trained classifier to locate fibres in the late development. False template matches are eliminated using a Random Forest classifier. Candidates in each slice were then linked to identify extended fibres, which are then checked and corrected in a final filtering stage.

3.1. Fibre Detection and Tracking

For fibre detection and tracking we build on the algorithm described in [1]. We extended the detection algorithm to deal with non-circular shapes and introduce a new step to correct errors in tracking by validating the fibres using a Kalman Filter.

At early stages of embryonic development each fibre has a roughly circular shape. Our method is based on template matching for identifying fibres. We defined set of models each trained to find fibres at a particular radii. To construct a model we annotate set of fibres, as a training examples,

by drawing a circle around each. We construct a set of templates at a range of sizes from the mean of patches extracted around fibres with similar diameters. During search for these circular fibres we use normalised cross-correlation (NCC) to locate candidates at each possible size.

During late development, fibres have a more polygonal shape. We train random forest (RF) classifiers to distinguish non-circular fibres from false positives. Eight RFs have been trained, each designed to deal with a different (narrow) range of radii of fibres, centred on \hat{r} . To obtain training examples we annotate several fibres over a set of images. Thus, we have a set of circles defining the centre and radii of each fibre. To increase the training data we rotated these patches four times. False examples are generated automatically by displacing the circle around the center of fibre’s patch at different scales (see Fig.3). Each RF was then trained on image patches around all the samples which fell within it’s radii range, $[r_{min}, r_{max}]$, where $r_{min} = s^{-1}\hat{r}$, $r_{max} = s\hat{r}$, $s = 1.2$.

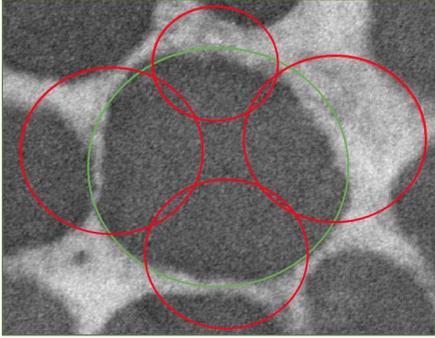


Figure 3. Fibre annotation: Positive example (green circle) and negative examples (red circles)

Image patches of size $2\hat{r} + 3$ were taken, centred on the centre of the circle defining the candidate. The features used to make decisions at each node of each tree in the forest were based on the difference between intensities in two randomly chosen pixels within the patch.

For fibre tracking, the detection algorithm is run on every frame giving a set of candidate disks, each of which is likely to be from a fibre. Each fibre appears as a sequence of candidate disks of similar radii in consecutive frames. The first step is to group the ends of fibres to locate bundles in the previous frame. Then, we estimate the movement of each bundle as the translation of all fibres which minimises the distance of their centres to the centres of candidates in the next slice.

3.2. Filling Gaps in Fibres

Occasionally a candidate disk for a fibre is not detected in an image, either due to the failure of the detector, or some image slices are corrupted by ‘tearing’ the surface of the

block when the diamond knife cuts the slice. Such missing disks cause a long fibre to be split into two (or more) shorter fibres. To detect and correct small gaps caused by such detection failures we use a linear prediction of the fibre’s location at frame z . We identify the end of every fibre at frame $z-1$ and estimate their center projected onto frame z . Similarly, we identify every fibre starting at frame $z+1$ and estimate their center projected onto frame z , then we link fibres that satisfy constraint according to the following procedure:

Let disk i in plane z have centre $\vec{p}_{i,z}$ and radius $r_{i,z}$, $i = 1..n_z$, n_z is the number of candidates in plane z . For every fibre i ending at frame $z-1$ we estimate the centre of the fibre at frame z as

$$\mathbf{p}_{i,z} = \mathbf{p}_{i,z-1} + (\mathbf{p}_{i,z-1} - \mathbf{p}_{i,z-2}) \quad (1)$$

For every fibre j starting at frame $z+1$ we estimate

$$\mathbf{p}_{j,z} = \mathbf{p}_{j,z+1} - (\mathbf{p}_{j,z+2} - \mathbf{p}_{j,z+1}) \quad (2)$$

Then, we link fibre i to j if $|\mathbf{p}_{i,z} - \mathbf{p}_{j,z}| < 1/2(r_{i,z} + r_{j,z})$ and $0.7 \leq |r_{i,z}/r_{j,z}| \leq 1.5$.

3.3. Correcting Mis-connections

When two fibres approach closely or a fibre changes direction, the connection algorithm above can create mis-matches or lose track. This is particularly common for lone fibres (which leave one bundle and join another) or when two bundles split or merge. Because of their stiffness, fibres tend to be fairly straight and thus the position of their centre point in a slice varies smoothly from one slice to another. We use a Kalman Filter (KF) framework to predict the position of the fibre centre in sequential frames, and thus identify when a mismatch may have occurred.

Each fibre will be associated with its own KF. The filter is initialised by the first five slices using fibre’s centre location. In subsequent slices we use the Kalman equations to predict fibre’s location in the next slice and compare it against that produced by the original linking algorithm. Where it finds a discrepancy, that position is recorded and the filter is re-initialised on the next 5 slices.

Each discrepancy is then re-visited, and corrected where necessary using **Algorithm 1**. There are several cases to deal with, including gaps caused by missing detections, two fibres which have been incorrectly swapped when they are close and cases where one fibre has terminated.

The algorithm is run forward and then backward through the volume to locate as many such cases as possible.

3.4. Lone Fibres

Most fibres are grouped together forming bundles. However, some of the fibres leave their bundle and join others, or enter other structures. Biologists are interested in the path

Algorithm 1: Correcting Mis-connections

Input: Fibre $fibres_{nm}$; // list of fibres
where n is number of fibres and m
is the length of fibre; each
fibre consist of set of disks.
Fibre $cfibre_m$; // fibre to be checked.
 $z = \text{FindDiscrepancyUsingKalman}(cfibre_m)$;
disk $d \leftarrow cfibre_z$;
for $i = 1 \leftarrow$ **to** n **do**
 if ($isDiskMatchWithFibre(fibres_{im}, d)$) **then**
 if ($isFibreStart(fibres_{im}, z)$) **then**
 // break fibre $cfibre_m$ at
 frame z then add its tail
 to fibres list
 $head_{cf} \leftarrow \text{BreakFibreH}(cfibre_m, z)$;
 $tail_{cf} \leftarrow \text{BreakFibreT}(cfibre_m, z)$;
 $\text{addFibre}(fibres_{nm}, tail_{cf})$;
 $\text{joinFibres}(head_{cf}, fibres_{im})$;
 else
 // break the two fibres
 $head_{cf} \leftarrow \text{BreakFibreH}(cfibre_m, z)$;
 $tail_{cf} \leftarrow \text{BreakFibreT}(cfibre_m, z)$;
 $head_{fi} \leftarrow \text{BreakFibreH}(fibres_{im}, z)$;
 $tail_{fi} \leftarrow \text{BreakFibreT}(fibres_{im}, z)$;
 $\text{addFibre}(fibres_{nm}, tail_{cf})$;
 $\text{addFibre}(fibres_{nm}, head_{fi})$;
 $\text{joinFibres}(head_{cf}, tail_{fi})$;

of such lone fibres. Lone fibres leave a bundle as a single fibre or as a small group of fibres. They might leave their bundle and merge with another or they might return to the original bundle.

To identify such fibres we scan each fibre and count the number of neighbouring fibres within a range of r_L within each slice. Any fibre which has more than two slices with no neighbours (excluding those close to the image boundaries) is labelled as a lone fibre.

Examples of such fibres are shown in Figure 4 below.

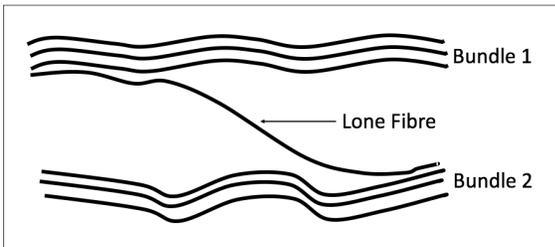


Figure 4. Lone fibre moving from bundle 1 to bundle 2.

4. Experiments

4.1. Data

We use images from three different datasets; (i) an embryonic 16.5 day wild type mouse tail sample used as a control for an MT1 knock out protease that cleaves collagen molecules (among other things), (ii) an embryonic 17.5 day wild type mouse tail sample used as a control for a collagen mutation that protects the fibrils from cleavage, (iii) an embryonic wild type mouse close to 17.5 day used as a control for a collagen receptor knock-down mouse.

The datasets were collected by SBFSEM system [4] under the brand name 3View. The images were created by collecting the back-scattered electrons before an in-chamber ultramicrotome removes a section [11] with assist of Gatan DigitalMicrograph software.

Annotated Data To quantitatively assess the performance of the tracking algorithm. We manually annotated 209 fibres across 102 slices. The location of the centre of each fibre was recorded on each of 102 images in a sequence - a total of 21019 points.

To compare our detection algorithm using RF with the work of [1] which based on NCC. we manually annotated 531 fibres across four images from 6 week dataset. Fibres at this stage have a more polygonal shape rather than a circular shape.

4.2. Evaluation Measure

To evaluate the performance of the algorithm we defined the following metric. Let $m_j(z)$ define the position of the marker for j th fibre at the z th slice.

The tracking for fibre i at slice z is defined as correct if $m_i(z)$ falls inside the circle that is identified by the detection and tracking algorithm for the outline of that fibre. We also report mismatched errors that occur when two fibres are swapped as they pass close to each other. Then we consider whether the fibre is correctly tracked after the error correction algorithm (see Figure 5).

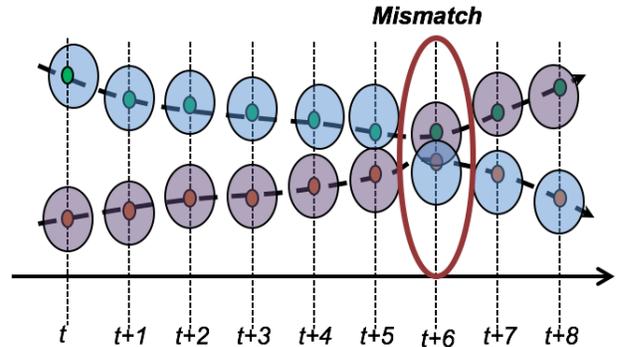


Figure 5. Evaluation measure

	Method	
	Gap Filling	Kalman Filter
Correctly Tracked	20949	20951
Miss	70	68
Mismatched	289	82

Table 1. Improvements given by Kalman Filter based correction

Table 1 shows the tracking algorithm result against the ground truth. Here we excluded the false positives since there are over 5000 fibres in the image. The Kalman Filter-based method corrects over 70% of the matching errors.

We have performed three experiments on three datasets to show the tracking performance with gap filling and Kalman correction. Figure 6 shows a histogram of the number of fibres with particular lengths (number of consecutive frames in which they are located) when analysing a block of images. As can be seen from all the figures that many more longer fibres are tracked when Kalman is applied. The first experiment shows the number of fibres with a length 183 were about 5000. However, after applying the Kalman based error correction the number increased to 6600.

We performed experiments to evaluate how well the random forest classifiers could discriminate between fibres and non-fibre. Manual annotation on a set of images gave 952 true fibre candidates and 3808 non-fibre candidates, which were used for training and testing the classifier. The testing data consist of 363 true fibres and 1452 non-fibres. To increase the training and testing data we performed rotation and scaling on the data.

We train eight RF classifiers, each consisting of 10 trees, with mean radii of 4, 5, 7, 9, 11, 13, 15 and 17 pixels. Figure 7 shows the Random Forest performance for each radii.

We performed a pilot experiment to compare Random Forest classifier for locating fibres with the work of [1] which based on NCC. We compute the precision and recall, where we define recall = $TP/(TP+FN)$ and precision = $TP/(TP+FP)$, where TP is the number of true positives, FN is the number of false negatives and FP is the number of false positives. We summarise the precision-recall in single number using the F1 score:

$$F = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

The F1 score for finding fibres using template matching achieved 62% where the Random Forest classifier achieved 96%.

5. Discussion

We have demonstrated a system to detect fibres and track them across image volumes which is fully automatic. The detection system involves finding candidates using template matching for circular shape and Random Forest classifier

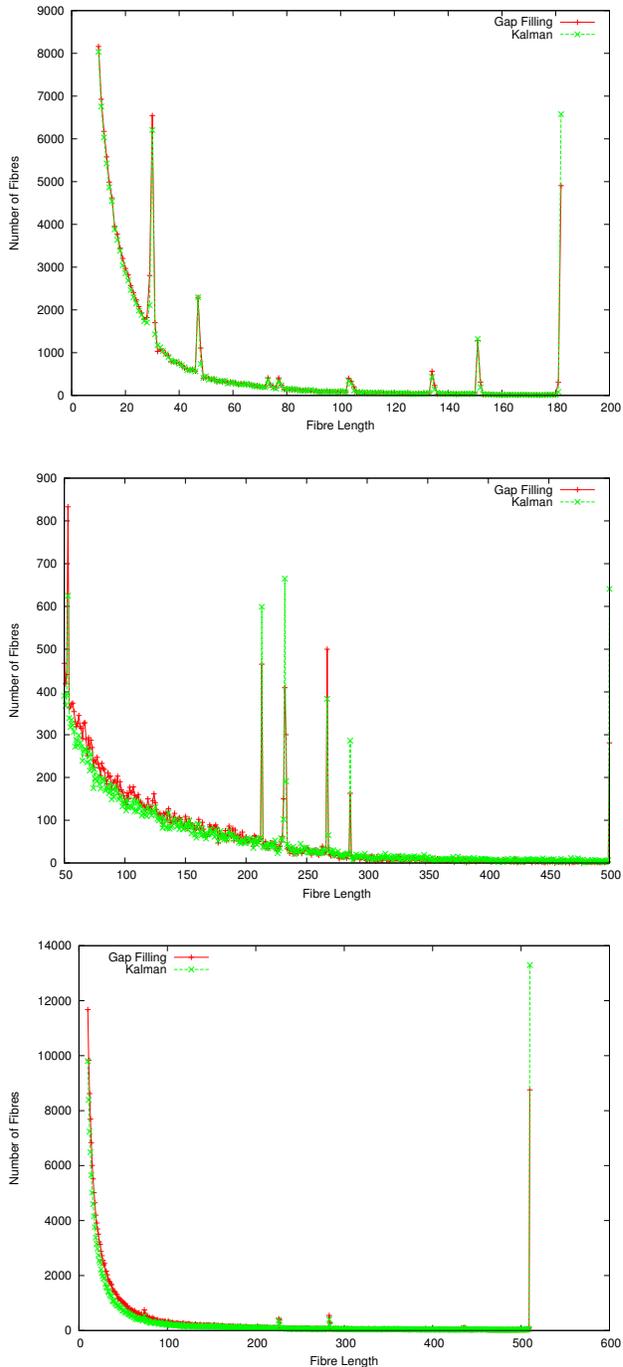


Figure 6. Histograms of the number of fibres with particular lengths on three data sets, with gap-filling and with Kalman correction, performed on dataset (i) (above), dataset (ii) (middle) and dataset (iii)(bottom). It demonstrates that Kalman significantly increases the number of longer fibres detected.

for polygonal shape, then discarding false matches using a Random Forest classifier. Locating fibres that have polygonal shape using template matching produced a large number

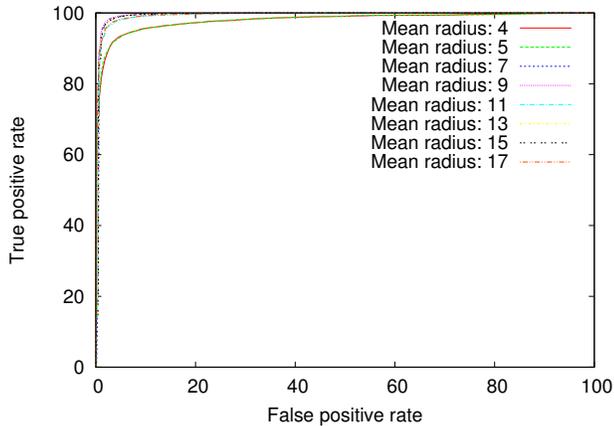


Figure 7. ROC for Random Forest classification performance for models of different radii.

of false positives.

A relatively simple tracking algorithm, which takes account of the movement of fibre bundles, is found to be effective for linking the detected disks together into extended fibres (see Fig.8). However, there are some errors when two fibres almost touch. To identify such errors we use a Kalman Filter to look for inconsistencies. Each discrepancy is then re-visited, and corrected where necessary.

The algorithm is able to track thousands of fibres across hundreds of slices, showing that the fibres follow complex paths through the tissue (see Fig. 9 and Fig.10).

References

- [1] Y. Almutairi, T. F. Cootes, and K. Kadler. Tracking collagen fibres through image volumes from sbfsem. In T. Lambrou and X. Ye, editors, *MIUA*, pages 40–45. BMVA, 2015. **3, 5, 6**
- [2] R. Bise, Z. Yin, and T. Kanade. Reliable cell tracking by global data association. In *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, pages 1004–1010, March 2011. **2**
- [3] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. **2**
- [4] W. Denk and H. Horstmann. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biol*, 2(11):e329, 2004. **1, 5**
- [5] N. S. Kalson, Y. Lu, S. H. Taylor, T. Starborg, D. F. Holmes, and K. E. Kadler. A structure-based extracellular matrix expansion mechanism of fibrous tissue growth. *Elife*, 4:e05958, 2015. **1**
- [6] B. X. Kausler, M. Schiegg, B. Andres, M. Lindner, U. Koethe, H. Leitte, J. Wittbrodt, L. Hufnagel, and F. A. Hamprecht. *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III*, chapter A Discrete Chain Graph Model for 3d+t Cell Tracking with High Misdetec-

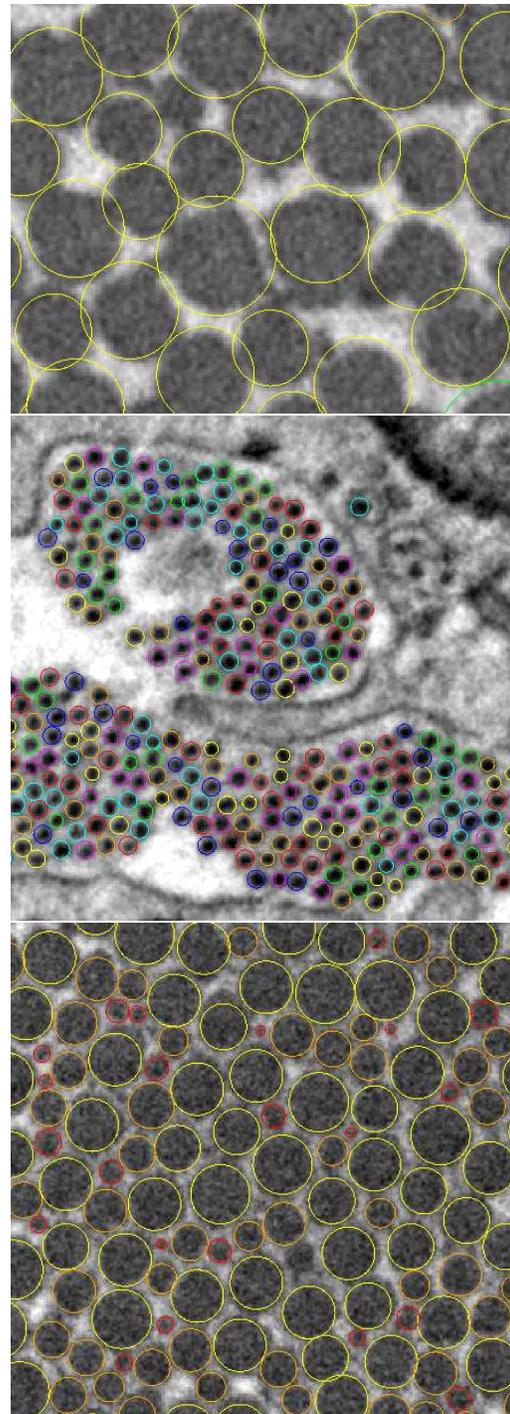


Figure 8. Result of fibre tracking on three different data sets.

- tion Robustness, pages 144–157. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. **2**
- [7] V. Kaynig, T. Fuchs, and J. M. Buhmann. Neuron geometry extraction by perceptual grouping in sstem images. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE*



Figure 9. Visualisation of lone fibres found by the tracking algorithm.

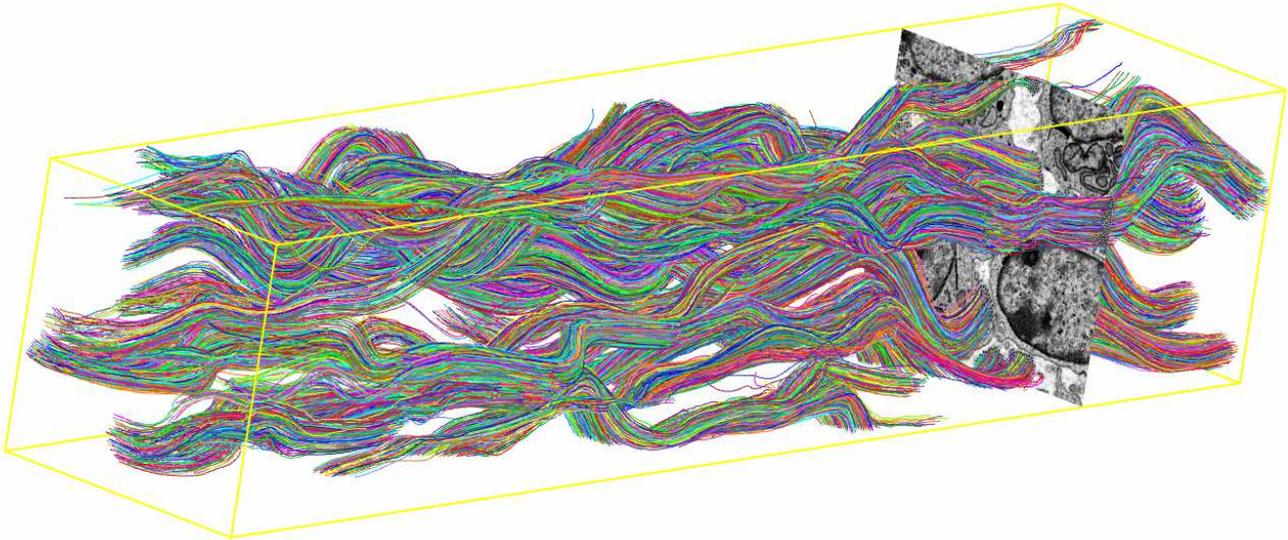


Figure 10. Visualisation of all fibres found by the tracking algorithm.

- Conference on*, pages 2902–2909. IEEE, 2010. 2
- [8] V. Kaynig, T. J. Fuchs, and J. M. Buhmann. Geometrical consistent 3d tracing of neuronal processes in sstem data. In *Proceedings of the 13th International Conference on Medical Image Computing and Computer-assisted Intervention: Part II*, MICCAI’10, pages 209–216, Berlin, Heidelberg, 2010. Springer-Verlag. 2
- [9] A. Lucchi, K. Smith, R. Achanta, V. Lepetit, and P. Fua. A fully automated approach to segmentation of irregularly shaped cellular structures in em images. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010*, pages 463–471. Springer, 2010. 2
- [10] G. Rigoll, B. Andres, U. Köthe, M. Helmstaedter, W. Denk, and F. Hamprecht. Segmentation of SBFSEM Volume Data of Neural Tissue by Hierarchical Classification - *Pattern Recognition - Lecture Notes in Computer Science. Pattern Recognition*, 5096:142–152–152, 2008. 3
- [11] T. Starborg, N. S. Kalson, Y. Lu, A. Mironov, T. F. Cootes, D. F. Holmes, and K. E. Kadler. Using transmission electron microscopy and 3View to determine collagen fibril size and three-dimensional organization. *Nature protocols*, 8(7):1433–48, 2013. 1, 5
- [12] S. N. Vitaladevuni and R. Basri. Co-clustering of image segments using convex optimization applied to em neuronal reconstruction. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2203–2210. IEEE, 2010. 2