Tracking Collagen Fibres through Image Volumes from SBFSEM

Yassar Almutairi[1]
yassar.almutairi@postgrad.manchester.ac.uk
Timoothy Cootes [1]
timothy.f.cootes@manchester.ac.uk
Karl Kadler [2]
karl.kadler@manchester.ac.uk

[1] Imaging Science,
The University of Manchester, UK
[2] Welcome Trust Centre for Cell-Matrix Research, Faculty of Life Sciences,
The University of Manchester, UK

Abstract

Serial block-face scanning electron microscopy (SBF-SEM) has become an important modality for examining small structures in biology, and new instruments are capable of gathering large volumes of data. Analysing these datasets manually is difficult because of the number of individual images involved. Objects of interest often span many slices, and must be tracked by the user through these slices to understand their full 3D structure. In this paper we focus on analysing the shape and structure of collagen fibres. There can be over 10,000 fibres in a single image, and ideally each would be tracked through the volume in order to estimate their lengths and how they form into bundles. Manual annotation of such data has proved impractical. Here we describe an automated system which detects the fibres and tracks them through large image volumes. We show that multi-scale normalised cross correlation is effective for finding candidates in a single image, and that false positive matches can be eliminated using a random forest classifier. Tracking is performed by linking candidates from one image to the next. We describe the system in detail, including experiments assessing the accuracy and reliability of the approach when tracking through hundreds of image slices in a volume.

1 Introduction

Collagen is the main structural protein found in connective tissues. It has many essential functions including force transmission, scaffolding, cell adhesion and cell migration. Collagen fibrils appear in ordered bundles in the extracellular matrix where they are the major tensile element in vertebrate tissues [7]. They can be found in skin, tendon, bone and hollow organs and vary in diameter ranging between 12 to 500 nanometres. These collagen fibrils are closely packed together in curvilinear bundles.

Recent technological developments have led to automatic 3D electron microscopes allowing the acquisition of large voxel volumes. These high resolution images are essential to several fields such as connectomics, which aims to reconstruct the structures comprehensively using these images. Serial-Sectioning TEM (ssTEM) could 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. Trelstad and Hayashi [8]

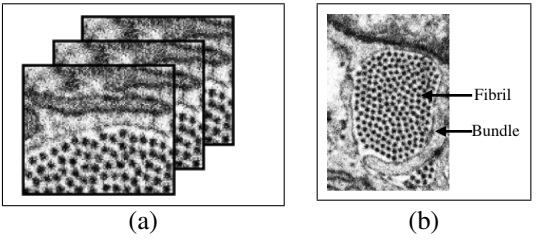


Figure 1: (a) Stack of 3 slices; (b) Bundle and fibril.

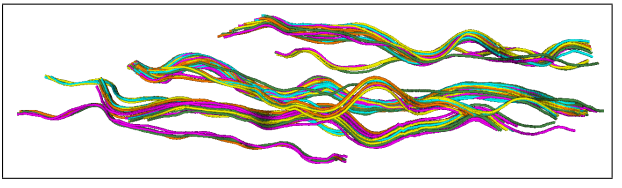


Figure 2: 3D paths of some of the detected fibres.

were the first to show the structure of collagen fibrils by ssTEM. Serial block-face scanning electron microscopy (SBFSEM) can automate the process of ssTEM [2] and provide large 3D datasets. Figure 1 (a) shows an example of stack of three slices.

Background We are not aware of any previous publications on tracking collagen fibres on EM images. The closest work is that of Jurrus et al. [5] [3] on electron microscopy (EM) images to track membranes and [4] to track multiple axons across large image volumes.

The methods rely on the assumption that the initial 2D segmentation of each section is good enough for the posterior grouping, or that every object has been over segmented [6] [9]. In addition, some of these methods need indirect penalties to prevent trivial yet incorrect clusterings [9], or require setting stopping conditions or manually-designed rules to converge to the right solution. Finally, a number of methods rely on greedy segment-merging strategies and cannot guarantee global clustering optimality.

2 Methods

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. False template matches were eliminated using a Random Forest classifier (RF) [1]. Candidates in each slice were then linked to identify extended fibres.

2.1 Fibre Detection and Tracking

Each fibre has a roughly circular shape which may vary from slice to slice due to sectioning. We developed our method based on template matching. We defined multiple models each trained to locate fibres at a particular radii.

We use normalised cross-correlation (NCC) to search each image with the given template. As can be seen from Figure 3 (left), there are also many false positives, non-fibre patches.

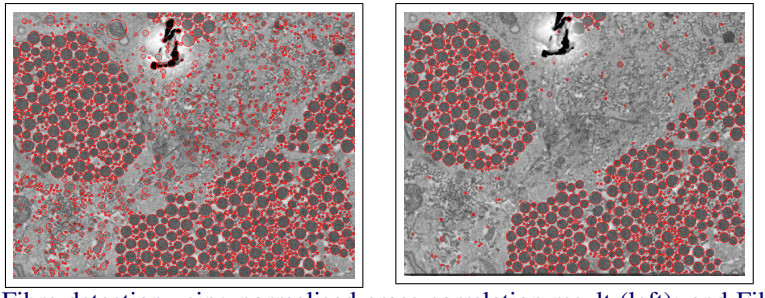


Figure 3: Fibre detection using normalised cross correlation result (left); and Fibre candidates after removing false positives using RF classifiers (right) (Best viewed in colour).

Eight Random Forest classifiers have been trained to remove false positives, each designed to deal with a different (narrow) range of radii of fibres. To obtain training examples we ran the NCC template model described above over a set of images, using a low threshhold to minimise the number of false negatives (missed fibres). This produced a set of circles defining the centre and radii of each candidate. These were manually annotated as either true fibres or false matches. Each random forest 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 (right) shows the result after removing these false positives using the RF classifier.

After running the detection stage on every frame we have a set of candidate disks, each of which is likely to be from a fibre. 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. Each fibre appears as a sequence of candidate disks of similar radii in consecutive frames.

The simplest approach to tracking would be to extend each fibre with the nearest candidate disk of the correct radius in the next slice, if it was within a suitable threshold distance. In practise this does not work well for fibres in bundles, as the 'drift' of the bundle between frames can be larger than the separation between nearby fibres. This leads to ambiguity in matching the candidates to the fibres and thus incorrect linking.

To reduce the chances of this we take account of the drift (the movement of the fibre bundle from one frame to the next) and assume each fibre only moves small amounts relative to its neighbours in the bundle when locating suitable candidates in the next frame.

We assume all candidate disks in the first frame are the start of a fibre, creating n_1 fibres each containing a single disk. We then process the subsequent frames one at a time, using the candidate disks to either extend an existing fibre, or to create new fibres if there is no match in the previous frame.

When processing a frame, the first step is to group the ends of fibres in the previous frame in order to identify bundles. This is done by a clustering algorithm, in which each fibre is added to an existing bundle if its disk at slice $(z-1)$ has a centre within a radius r_c of any other fibre centre in the cluster.

We then estimate the "drift" of each bundle as the translation of all fibres which minimises the distance of their centres to the centres of candidates on the next slice. In particular, let

$D_z(\vec{x})$ be the distance transform of the centres of disks in image z

$$D_z(\vec{x}) = min_i|\vec{p}_{i,z} - \vec{x}| \tag{1}$$

Let $\{\vec{x}_{b,j}\}$ be the centres of the n fibres identified in a bundle b on frame $(z-1)$. The movement of the bundle is then the translation $\hat{\vec{t}}_b$ which minimises

$$S_b(\vec{t}) = \sum_j D_z(\vec{x}_{b,j} + \vec{t}) \tag{2}$$

Let $\vec{x'}_{b,j} = \vec{x}_{b,j} + \hat{\vec{t}}_b$ be the estimated centre of each fibre from frame $z-1$ projected onto frame z, and $r_{b,j}$ be the radius of the disk for that fibre in frame $(z-1)$.

We now consider every candidate disk $\{\vec{p}_{i,z}, r_{i,z}\}$ in turn. If $|\vec{p}_{i,z} - \vec{x'}_{b,j}| < d_t$ and $|\log(r_{i,z}/r_{b,j})| < \log(1.25)$, then the disk (i,z) is used to extend fibre j in bundle b, otherwise the disk is used to start a new fibre. Fibres are assumed to have ended if no match is found in the next frame.

Occasionally a candidate disk for a fibre is not detected in an image, either due to the failure of the detector, or where 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 to predict fibre's location at frame z. We identified the end of every fibre at frame $z-1$ and estimate their center projected onto frame z. Similarly, we identified every fibre starting at frame $z+1$ and estimate their center projected onto frame z, then link those fibres.

3 Experiments and Results

Our data sets consist of; (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.

We performed experiments to evaluate how well the random forest classifiers could discriminate between fibres and non-fibre candidates found by the NCC-based models. Manual annotation of the output of the NCC models on a set of images gave 53480 true fibre candidates and 47022 non-fibre candidates, which were used for training and testing the classifier. We randomly split the examples into two sets where 70% of the examples are used for training and the remainder for testing.

We train eight RF classifiers, each consisting of 20 trees, with mean radii of 4, 5, 7, 11, 13, 15, 17 and 19 pixels. Figure 4 shows classifiers performance for each radii.

To evaluate the tracking algorithm we compared results with manual annotations. We annotated 208 fibres on each image of 10 images in a sequence - a total of 2080 points. To quantitatively assess the performance of the algorithm we defined the following metric. Let $m_j(z)$ define the position of the marker for jth fibre at the zth 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 that fibre.

In these experiments we only report the true positive rate since there are over 3000 fibres in the image - we cannot identify false positives (other positives may correspond to correct

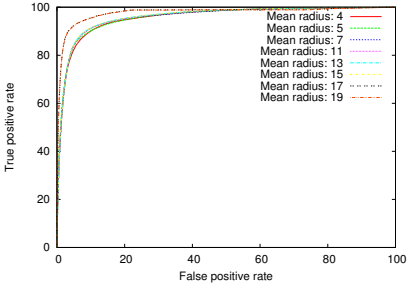
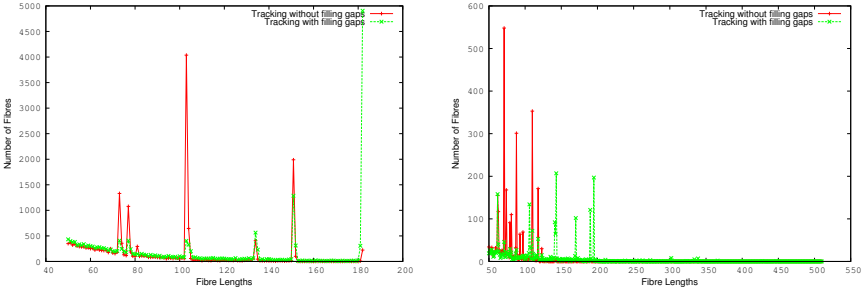


Figure 4: ROC for Random Forest classification performance for models of different radii.

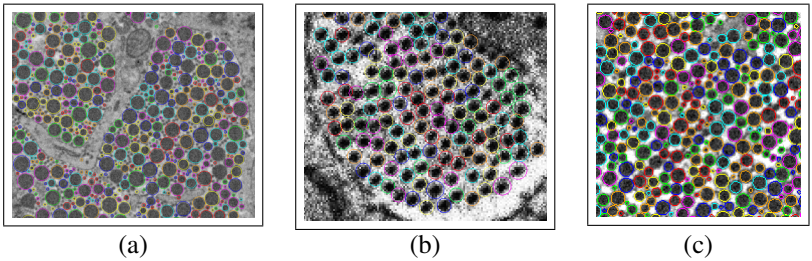
but unannotated fibres). When tested against the ground truth in this way 94% of the fibre slices are correctly identified.



(a) from dataset (i) (b) from dataset (ii)

Figure 5: Histograms of the number of fibres with particular lengths on two data sets, with and without gap-filling. It demonstrates that the gap filling significantly increases the number of longer fibres detected.

We have performed two experiments on two data sets to show the tracking performance with and without gap filling. Figure 5 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 the figures that many more longer fibres are tracked when gap filling is applied. The first experiment shows the number of fibres with a length 183 were about 200. However, after applying gap filling the number increased to 4800.



(a) (b) (c)

Figure 6: Result of fibre tracking on three different data sets.

4 Discussion and Conclusions

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, then discarding false matches using a Random Forest classifier. 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.6). 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. 2).

One limitation of the current approach is that by assuming all fibres in a bundle move in roughly the same way we can fail to track those few fibres which pass through the bundle at an angle, often exiting to join other bundles - these are often of particular interest to biologists, so we will design algorithms to explicitly identify such fibres. Similarly we can get miss-matches when one bundle splits in two - again the assumption of all fibres moving the same way breaks down. Potentially this can be corrected by a multi-pass approach.

References

[1] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[2] Winfried Denk and Heinz Horstmann. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biol*, (11):e329. doi: 10.1371/journal.pbio.0020329.

[3] Elizabeth Jurrus, Ross Whitaker, Bryan W. Jones, Robert Marc, and Tolga Tasdizen. An optimal-path approach for neural circuit reconstruction. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI*, pages 1609–1612, 2008.

[4] Elizabeth Jurrus, Melissa Hardy, Tolga Tasdizen, P Thomas Fletcher, Pavel Koshevoy, Chi-Bin Chien, Winfried Denk, and Ross Whitaker. Axon tracking in serial block-face scanning electron microscopy. *Medical Image Analysis*, 13(1):180–188, 2009.

[5] Elizabeth Jurrus, Antonio R C Paiva, Shigeki Watanabe, James R Anderson, Bryan W Jones, Ross T Whitaker, Erik M Jorgensen, Robert E Marc, and Tolga Tasdizen. Detection of neuron membranes in electron microscopy images using a serial neural network architecture. *Medical Image Analysis*, 14(6):770–783, 2010.

[6] Verena Kaynig, Bernd Fischer, Elisabeth Müller, and Joachim M. Buhmann. Fully automatic stitching and distortion correction of transmission electron microscope images. *Journal of Structural Biology*, 171(2):163–173, 2010.

[7] Tobias Starborg, Nicholas S Kalson, Yinhui Lu, Aleksandr Mironov, Timothy F Cootes, David F Holmes, and Karl E Kadler. Using transmission electron microscopy and 3View to determine collagen fibril size and three-dimensional organization. *Nature protocols*, (7):1433–48. ISSN 1750-2799.

[8] Robert L Trelstad and Kimiko Hayashi. Tendon collagen fibrillogenesis: intracellular subassemblies and cell surface changes associated with fibril growth. *Developmental biology*, 71(2):228–242, 1979.

[9] Shiv N. Vitaladevuni and Ronen Basri. Co-clustering of image segments using convex optimization applied to EM neuronal reconstruction. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2203–2210, 2010.