PyFibre - Python Fibrous Image Analysis Toolkit

Francis G.J. Longford

January 13, 2019

PyFibre (Python Fibrous Image Analysis Toolkit) is a computer program written in Python purposely built to analyse fibrous tissue.

1 Analysis Methods

A variety of image processing techniques have been used to assess the orientation and structure of collagen fibres in the ECM. Most have been applied using the hypothesis that there is a relationship between cancer progression and the alignment of fibres. To this end, the techniques also attempt to define a metric to describe this alignment. These include Fourier[1] and nematic tensor[2] analysis of image intensity as well as more advanced methods to identify individual fibrils employing the curvelet transform[3]. However, generally there is no commonly agreed automated method to measure cancer progression from medical imaging. We outline a range of image analysis techniques below that have either been used to cancer biopsies, or that may prove beneficial for future investigations.

1.1 Fourier Transform

Fourier analysis methods have been used previously to quantify the orientation distribution $P(\phi)$ of fibres in a SHG image[1]. They can be quick and easy to implement using fast Fourier transform (FFT) methods, such as those available in NumPy. The Fourier transform A_{jk} of our 2D SHG images then becomes

$$A_{jk} = \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} I_{uv} \exp\left(2\pi i \left(\frac{uj}{N} + \frac{vk}{M}\right)\right) \tag{1}$$

With the wave amplitudes α_{kj} and angles ϕ_{kj} for each frequency corresponds to equation (2).

$$\alpha_{jk} = \sqrt{A^{*}_{kj} A_{kj}} \tag{2a}$$

$$\phi_{jk} = \mathrm{Imag}\left\{\ln\left(\frac{A_{kj}}{\alpha_j k}\right)\right\} \tag{2b}$$

We calculate the probability distribution $P(\phi)$ of fibril orientations using the average amplitudes $\langle\alpha(\phi_j k)\rangle$

$$P(\phi) = \langle\alpha(\phi_{jk})\rangle \quad where \quad \phi_{jk} = \phi \tag{3}$$

The spectrum of $P(\phi)$ is then able to inform us of the likelihood of finding fibres in a particular orientation. Although the value of ϕ is arbitrary, the distribution of probabilities can inform us of how ordered the system is. For example, a highly ordered system is likely to produce a spectrum dominated by a few values of $P(\phi)$, whereas a relatively disordered system will have a uniform distribution of $P(\phi)$.

Metrics used to describe the Fourier spectrum include the spectral directional index (SDI) and modified directional index (MDI), which have been recently applied to AFM images in order to measure the degradation of cell cytoskeleton fibres under ionising radiation[4].

1.2 Structure Tensor

The structure tensor can be used to estimate the local direction and magnitude of features in an image[5]. It is derived from the Jacobian matrix at each pixel, which can be averaged over a region if desired. Representing an image as a 2D intensity map $f(x, y)$, results in a Jacobian vector $\mathbf{J}(x,y)$ given in equation (4).

$$\mathbf{J}(x,y) = \left[\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right] \tag{4}$$

The structure tensor $\mathbf{K}(x,y)$ is the dot product $\mathbf{J}(x,y) \cdot \mathbf{J}(x,y)^T$. Representing the derivative of $f(x,y)$ with respects each variable as $f'_x(x,y)$ or $f'_y(x,y)$, we can represent the structural tensor as equation (5).

$$\mathbf{K}(x,y) = \mathbf{J}(x,y) \cdot \mathbf{J}(x,y)^T = \begin{pmatrix} f'_x(x,y)^2 & f'_x(x,y)f'_y(x,y) \\ f'_y(x,y)f'_x(x,y) & f'_y(x,y)^2 \end{pmatrix} \tag{5}$$

In practice, the gradients $f'_x(x,y)$ and $f'_y(x,y)$ for a discrete data set such as $f(x,y)$ are often estimated by the central difference method or a spline based approach. The structure tensor of a region centred on $\mathbf{r}(x_0,y_0)$ can be estimated by averaging over a window (typically a Gaussian or uniform filter) $\mathbf{w}(\mathbf{r})$.

$$\mathbf{K}_w(x_0,y_0) = \int\int w(x-x_0, y-y_0)\,\mathbf{K}(x,y)\,dx\,dy \tag{6}$$

The maximum and minimum eigenvalues of the structure tensor, λ_{max}, λ_{min} and their corresponding eigenvectors $\mathbf{e}_{max}$, $\mathbf{e}_{min}$, can be used to measure the local anisotropy n, and orientation θ:

$$n = \lambda_{max} - \lambda_{min} \tag{7a}$$

$$\theta = \frac{1}{2}\arctan\left(\frac{2\left\langle f'_x(x,y)f'_y(x,y,)\right\rangle}{\left\langle f'_y(x,y)^2\right\rangle - \left\langle f'_x(x,y)^2\right\rangle}\right) \tag{7b}$$

1.3 Nematic Tensor

Similar to the structure tensor, in order to measure the alignment of the collagen fibril network we adopt the same methodology as Garcia *et al.*[6], who employed the FibrilTool plugin[2] of the ImageJ software package[7]. This tool calculates the tangent to the structure unit vector $\mathbf{t}(x,y)$:

$$\mathbf{t}(x,y) = \left[\frac{-f'_y(x,y)}{\sqrt{f'_x(x,y)^2 + f'_y(x,y)^2}}, \frac{f'_x(x,y)}{\sqrt{f'_x(x,y)^2 + f'_y(x,y)^2}}, \right] \tag{8}$$

Which is transformed into components of the 2x2 nematic tensor $\mathbf{n}(x,y)$ for each pixel.

$$\mathbf{n}(x,y) = \mathbf{t}(x,y) \cdot \mathbf{t}(x,y)^T = \begin{pmatrix} f'_y(x,y)^2 & -f'_x(x,y)f'_y(x,y) \\ -f'_y(x,y)f'_x(x,y) & f'_x(x,y)^2 \end{pmatrix} \frac{1}{f'_x(x,y)^2 + f'_y(x,y)^2} \tag{9}$$

The nematic tensor for a selected area then becomes the average local tensor $\langle \mathbf{n} \rangle$ of the constituent pixels. Once again, the eigenvalues and corresponding eigenvectors of $\langle \mathbf{n} \rangle$ are used to measure the anisotropy of collagen fibrils in the sampled image region.

1.4 Graph Theory

Fibril networks may be considered graphs, with the fibres themselves as edges between n nodes at the regions they interconnect. A graph cant be described by an adjoint matrix, an nxn matrix that records whether an edge is present between each node in a system. represents a collection of nodes Before conversion to a graph

1.5 Curvelet Transform

One of the most powerful and widely used tool to identify collagen fibres from medical images is CT-FIRE (Curvelet Transform–FIbeR Extraction software), found in the ImageJ[] suite. CT-FIRE uses the curvelet transform[8] (CT) (a wavelet transform variant) to filter out background image noise from individual fibres, making use of the CT's ability to resolve curved, cylindrical features. The software also contains a fibre extraction tool in order to quantify the structure of the collagen network[9]. [3]

1.6 Hough Transform

The Hough transform is a feature extraction tool that was originally developed to identify lines and curves in images[10]. It was later extended for arbitrary shapes[11], and became a popular technique in the computer vision community. The Hough transform has been recently been implemented to measure the degradation of the cell cytoskeleton under ionising radiation[]. Considering the similarities between the appearance of a stiff collagen fibre network and ...., it is likely that the same technique could also be applied to measure the anisotropy of the ECM.

1.7 Hierarchical Clustering

Until this point we have considered image analysis techniques that attempt to define a metric based on the rationale of fibre alignment. However, it is also possible to measure the similarity of images without the need for quantification of some known or commonly agreed features. This approach may be used to investigate the appropriateness of the classification of biopsies based on features outlined in the Gleason score (figure **??**). Simply put, it is beneficial to assess whether a machine would classify a set of images in the same way that a trained human would.

1.8 Non-Negative Matrix Factorisation (NMF)

We consider a matrix solely comprised of non-negative elements $\mathbf{V}$ to be able to be factorised into two smaller matrices $\mathbf{W}$ and $\mathbf{H}$.

$$\mathbf{V} = \mathbf{WH} \tag{10}$$

The features matrix $\mathbf{W}$, represents a set of key (hidden) features, and the coefficients matrix $\mathbf{H}$, provides the weighting of these features in the original $\mathbf{V}$. The factorisation relies on the property that all matrices do not possess any negative elements, making it particularly applicable to image analysis. Solutions to the matrices $\mathbf{H}$ and $\mathbf{W}$ are estimated by the following criteria, where $||A||_F$ signifies the Frobenius norm (equation (12)).

$$\min_{W,H} ||V - WH||\,|_F^2 \qquad \text{where} \qquad W \geq 0, H \geq 0 \tag{11}$$

$$||A||_F^2 = \sum_{i,j} A_{ij}^2 \tag{12}$$

In practice, there are several algorithms available to find a solution to equation (11). One of the most popular and straight forward methods is the multiplicative update rule, developed by Lee and Seung[12]. Using an initial (non-negative) guess of $\mathbf{W}$ and $\mathbf{H}$, each index i, j is iteratively updated at step n using the following scheme, until both matrices are stable.

$$H_{ij}^{n+1} \leftarrow \mathbf{H}_{ij}^{n} \frac{((W^n)^T V)_{ij}}{((W^n)^T W^n H^n)_{ij}} \tag{13}$$

$$W_{ij}^{n+1} \leftarrow \mathbf{W}_{ij}^{n} \frac{(V(H^{n+1})^T)_{ij}}{((W^{n+1} H^{n+1} (H^{n+1})_{ij}^T} \tag{14}$$

1.9 Neural Networks

2 Routines

1. Preprocessing of input images, including high and low-pass filtering followed by adaptive equalisation and Gaussian convolution.
2. Form nematic tensor using Gaussian convolution with a standard deviation of $\sigma = 0.5$ px
3. Form structure tensor using Gaussian convolution with a standard deviation of $\sigma = 0.5$ px
4. Form Hessian matrix using Gaussian convolution with a standard deviation of $\sigma = 0.5$ px
5. Calculate “tubeness“ T at each pixel
6. Perform a modified version of the FIRE algorithm to extract

References

[1] A. Ghazaryan, H. F. Tsai, G. Hayrapetyan, W.-L. Chen, Y.-F. Chen, M. Y. Jeong, C.-S. Kim, C. S-J., and D. C-Y., “Analysis of collagen fiber domain organization by fourier second harmonic generation microscopy,” *J. Biomed. Opt.*, vol. 18, p. 031105, 2012.

[2] A. Boudaoud, A. Burian, D. Borowska-Wykret, M. Uyttewaal, R. Wrzalik, D. Kwiatkowska, and O. Hamant, “Fibriltool, an imagej plug-in to quantify fibrillar structures in raw microscopy images,” *Nat. Protocols*, vol. 9, p. 457–463, 2014.

[3] J. S. Bredfeldt, Y. Liu, C. A. Pehlke, M. W. Conklin, J. M. Szulczewski, D. R. Inman, P. J. Keely, R. D. Nowak, T. R.Mackie, and K. W. Eliceiri, “Computational

segmentation of collagen fibers from second-harmonic generation images of breast cancer," *J. Biomed. Opt.*, vol. 19, p. 16007, 2014.

[4] M. Manghi, L. Bruni, and S. Croci, "MDI: Integrity Index of Cytoskeletal Fibers Observed by AFM," *The European Physical Journal Plus*, vol. 131, no. 6, p. 213, 2016.

[5] B. J ahne, *Digital Image Processing.* Springer-Verlag Berlin Heidelberg, 2005.

[6] A. M. Garcia, F. L. Magalhes, J. S. Soares, E. Paulino-Jr, M. F. de Lima, M. Mamede, and A. M. de Paula, "Second harmonic generation imaging of the collagen architecture in prostate cancer tissue," *Biomed. Phys. Eng. Express*, vol. 4, p. 025026, 2018.

[7] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "Nihimage to imagej: 25 years of image analysis," *Nat. Methods*, vol. 9, p. 671–675, 2012.

[8] J. L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans Image Process*, vol. 11, p. 670–684, 2002.

[9] A. M. Stein, D. A. Vader, L. M. Jawerth, D. A. Weitz, and L. M. Sander, "An algorithm for extracting the network geometry of three- dimensional collagen gels.," *J. Microsc.*, vol. 232, p. 463–475, 2008.

[10] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Comm. ACM.*, vol. 15, pp. 11–15, 1972.

[11] D. H. Ballard, "Generalising the hough transform ro detect arbitrary shape," *Pattern Recognition*, vol. 13, pp. 111–122, 1981.

[12] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *In NIPS*, pp. 556–562, MIT Press, 2000.