PATCH-BASED FEATURE MAPS FOR PIXEL-LEVEL IMAGE SEGMENTATION

Shuoying Cao, Saadia Iftikhar, Anil Anthony Bharath

Imperial College London

ABSTRACT

In this paper, we describe the use of phase-invariant complex wavelet filters, coupled to a training process involving a small, high-quality training dataset, to build an image segmentation system capable of performing in very low signal-to-noise, and under conditions of strong object-background contrast change. The three main components of our approach are: *i)* a patch-based feature description of local phase-invariant orientation fields; *ii) a priori* ground-truth data; *iii)* a machine learning method, such as Multilayer Perceptron (MLP) or kernel-based Support Vector Machine (SVM), to build an accurate classifier that is customised to the segmentation problem. A key feature of the approach is that it may be easily retrained and is, therefore, more adaptable to different imaging modalities. A representation of phase-invariant local image orientation using geometric algebra is first introduced; this is important to the patch-based approach. The quality of our trained systems is then assessed using Receiver Operating Characteristic (ROC) curves in two different biomedical applications: the human retinal vessel-bed in colour fundus images from the publicly available DRIVE database, and the rabbit endothelial cell boundaries of thoracic aorta microscopy images.

Index Terms— Segmentation, visual features, geometric algebra, Multilayer Perceptron, Support Vector Machine, microscopy, retinal imaging.

1. INTRODUCTION

Segmentation is an important first step in the automated analysis of several types of biomedical image data. Unlike object recognition, the techniques that have been developed for structure segmentation are optimised for particular resolutions, modalities and contrast (staining) mechanisms. Examples of methods that have been employed include active contours [1], maximum likelihood [2], watershed algorithms [3], and various threshold-based morphological approaches [4]. The variety of these techniques is a reflection of the fact that there are few methods that generalise well across the many contrast mechanisms increasingly found in biomedical imaging. The wealth of data has spawned a wealth of methods, some of which claim to be optimal for its designated data source. Yet there is a shortage of a unified segmentation approach that is transferable among large multi-modal data sets. In this paper, we describe one such approach based on capturing phase-invariant orientation and a few intensity-based measures to detect line-like structures. Our motivation is a need to be able to adapt a method quickly to tackle new segmentation problems with minimal code redevelopment. Our requirements are therefore: *i)* to use features that either generalise well to new problems, or are easily described; *ii)* to use a training process that can be applied to a small set of manually segmented data (i.e. ground truth) in order to teach a system to segment new data sets; *iii)* to use a training process and feature set combination that generalises well from small training sets to more data of the *same* type.

The outline of the paper is as follows: Section 2 presents an explanation and justification of the feature sets that we use, and justifies the use of a phase-invariant patch representation to capture spatial structure. In Section 3, we explain the details of how the training of a machine learning system is performed on different image modalities, retinal fundus images (Section 3.1 & 3.2) and cellular microscopy images (Section 3.3 & 3.4). A description of performance evaluation is then given in Section 4 and an analysis of the performance of the proposed segmentation approach is provided in Section 4.1 & 4.2. Finally, in Section 5, we provide some concluding remarks around further work and the generalisability of the approach to other problems of a similar, but not identical type.

2. FEATURE SPACE CONSTRUCTION

A common problem of medical or biomedical image segmentation is that the connectivity of boundaries should be enforced. This is often achieved through a post-processing method known as edge-linking. Incorrect edge linking can be a source of poor segmentation results. Capturing the properties of pixels across larger scales of space provides a method of making edge-linking more reliable. But it is inelegant to extend basic edge-linking algorithms beyond immediate neighbourhood pixels, because the number of decisions and the complexity of the decision process become high. Our aim is to use a sufficiently rich feature vector that incorporates neighbourhood information appropriate to describing the weak, thin nature of the cell boundaries, and implicitly encodes the continuity information in the feature vector itself. Rather than using explicitly coded rules to en-

force continuity, the feature vector and a training process are used instead, making this a more adaptable approach, should imaging statistics change. For the classification of pixels into object or background classes, the high-dimensional nature of the problem might suggest the use of either a Multilayer Perceptron (MLP) or a Support Vector Machine (SVM). The former is known to perform fast classification decisions and achieve good recognition rate of curvilinear shapes [5]. The latter is known to perform well on high-dimensional input vectors [6], partly due to the use of a margin-based training approach. This combination of feature space and classifier allows us to take the following approach: rather than merely using features at a pixel which are computed over an image patch, we use a *patch* of the *feature image* around the pixel for which we are creating the feature vector. This is *not* the same as merely using feature vectors estimated using linear operators over larger neighbourhoods, because of the non-linearities that are involved in either generating feature vectors or in learning classifiers.

Fig. 1 summarises the principle, contrasting it with edge-linking post-processing.

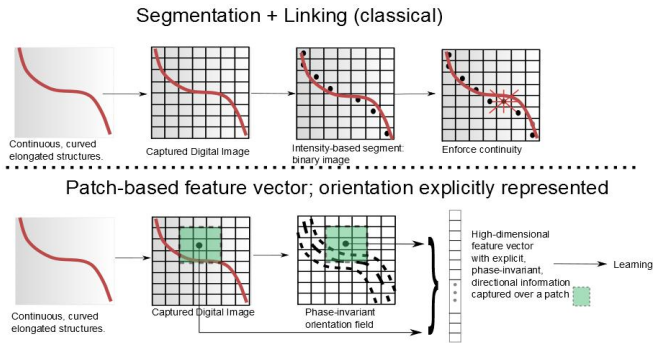

Fig. 1. *Two contrasting approaches. Top: Segmentation is followed by explicit edge linking; Bottom: The decision process at individual pixels utilises explicit information on the local phase-invariant orientation* ***field****.*

We use an intensity patch around the pixel to be classified; a good classifier can learn the distribution of pixel intensities that would otherwise be used by a thresholding approach. In addition, to capture information that relates to thin, elongated structures, we use an eighteen-dimensional feature subspace extracted from the 3×3 nearest neighbours of symmetric and antisymmetric responses of filter pairs that are in complex quadrature [7]. A sample of the estimated phase-invariant orientation field imposed over a patch from a retinal fundus image is shown in Fig. 2, and over a sub-region in thoracic aortic microscopy image is shown in Fig. 3. Note, in Fig. 3, the contrast in the sub-region has been manually enhanced for the purpose of visibility in print. Contrast enhancement is not used prior to calculating the orientation field itself.

The orientation field $\mathbf{O}^{(\ell)}$ of the image, $I(x, y)$, can be described at decomposition level, ℓ, for a multi-rate, quadrature-pair steerable pyramid by:

$$\mathbf{O}^{(\ell)} = \frac{u^k \psi_k^{(\ell)}}{\varepsilon + \sqrt{\sum_{k \in \{0,1,2,3\}} (\psi_k^{(\ell)})^2}} \tag{1}$$

where a term ab signifies the geometric product between multivectors in a $\mathcal{C}\ell_2$ algebra; Einstein's summation convention is used in the numerator of Eq. (1). ε is a constant, described in [7], and the variables $\psi_k^{(\ell)},\ k = 0, 1, 2, 3$ refer to the *magnitude* outputs of four, directed, quadrature filter pairs which are applied by spatial convolution to $I(x, y)$ at multiple scales. The orientation field $\mathbf{O}^{(\ell)}$ is represented as a field in $\mathcal{C}\ell_2$ by weighting unit-length grade-1 vectors by the magnitudes of phase-quadrature filter outputs. The unit-length vectors, $u^k, k = 0, 1, 2, 3$, are formed by $u^k = v^k v^0 v^k$ (in the non-commutative $\mathcal{C}\ell_2$ algebra), where $v^k, k = 0, 1, 2, 3$ represent the unit-length grade-1 vectors describing the k^{th} quadrature filter axis of spatial phase symmetry/antisymmetry. The RHS of Eq. (1) takes into account the double-coverage of the directed quadrature filter, and introduces a phase-invariant nature to the spatial orientation field $\mathbf{O}^{(\ell)}$. The magnitude of this field ranges from 0 to 1 and can be used as an indication of local anisotropy, in which strongly isotropic neighbourhoods produce values near to 0 and strongly anisotropic neighbourhoods produce values near to 1.

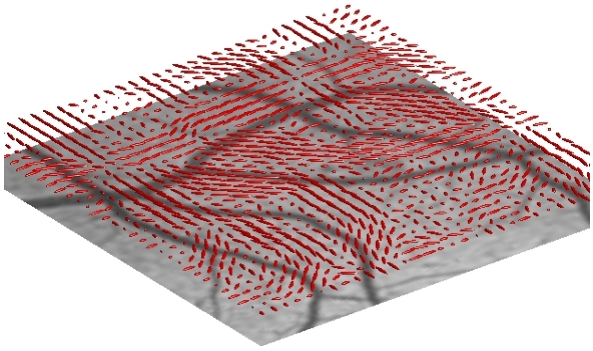

Fig. 2. *Estimated phase-invariant orientation field (in red) of a* 128×128 *patch from a retinal fundus image.*

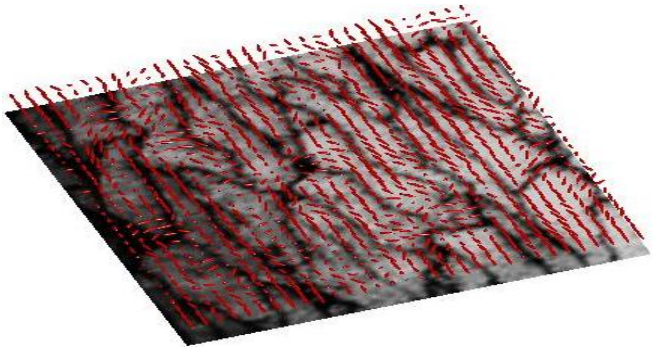

Fig. 3. *Estimated phase-invariant orientation field (in red) for a high-quality sub-region from a microscopy image. Regions are typically lower contrast and with greater degrees of noise.*

For some segmentation problems, it is also advantageous to use 6 statistical features from 3×3 intensity neighborhoods of the candidate pixel. These feature vectors are z-scale transformed, i.e. normalised in standard deviation and zero-mean transformed using: $\nu(i,m) = \frac{\nu(i,m)-\mu_m}{\sigma_m}, \forall i, m$, where $\{\mu_m\}_{m\in\{1,2,\ldots,M\}}$ and $\{\sigma_m\}_{m\in\{1,2,\ldots,M\}}$ are the mean and standard deviation in the training population of the m^{th} feature. M is the total number of features. $\{\nu(i,m)\}_{(i,m)\in\{1,2,\ldots,N\}\times\{1,2,\ldots M\}}$ denote the feature vector components before normalisation. One has to exercise caution *not* to normalise the orientation field information from $\mathbf{O}^{(\ell)}$ – which is already well-behaved and bounded – in this way, as the information encoded by the relation between vector components of individual observation vectors will be destroyed.

3. MACHINE LEARNING IN DIFFERENT IMAGE MODALITIES (VASCULAR AND CELLULAR)

3.1. Retinal Data

The photographs for the DRIVE database [8] were obtained from a diabetic retinopathy screening program in The Netherlands. The 40 images were acquired using a Canon CR5 nonmydriatic 3CCD camera with a 45 degree field of view (FOV). Each image was captured using 8 bits per color plane at 768 by 584 pixels. The FOV of each image is circular with a diameter of approximately 540 pixels.

Niemeijer [8] proposed a supervised pixel classification method such that for each pixel in the image, a 31-dimensional feature vector is constructed and a k-Nearest Neighbor (kNN) classifier is trained with these feature vectors. Their study also showed that the kNN classifier performed better than a linear or a quadratic classifier. This architecture first appeared as *convolutional neural nets* [9] used for recognizing handwritten characters. We suggest instead a smaller feature vector that relies on the phase-invariant orientation field over a local patch, combined with pixel intensities over the spatial patch.

3.2. MLP Training

We apply MLP with the Levenberg-Marquardt training algorithm [10] over a small sample of hand-labelled pixels. The network uses a mean squared error performance function, hyperbolic tangent sigmoid transfer functions for both hidden and output layers, and scaled conjugate gradient descent for iteration. The set of 40 images has been divided into a training and a test set, each with 20 images. The separability of the training and test data was encouraged [8] to avoid ambiguities and discrepancies when benchmarking classification algorithms. For training, we generated 5 different batches of random patch from each training image: 4000 pixels (0.89% of total image pixels), 8000 pixels (1.78%), 12000 pixels (2.68%), 16000 pixels (3.57%) and 24000 pixels (5.35%). Each patch contains approximately equal numbers of vessel and background pixels. We compared network models constructed from training based on 8×10^4 pixels, 1.6×10^5 pixels, 2.4×10^5 pixels, 3.2×10^5 pixels and 4.8×10^5 pixels. Two test fundus images (number 9 and number 16) with their segmentation results are shown in Fig. 4.

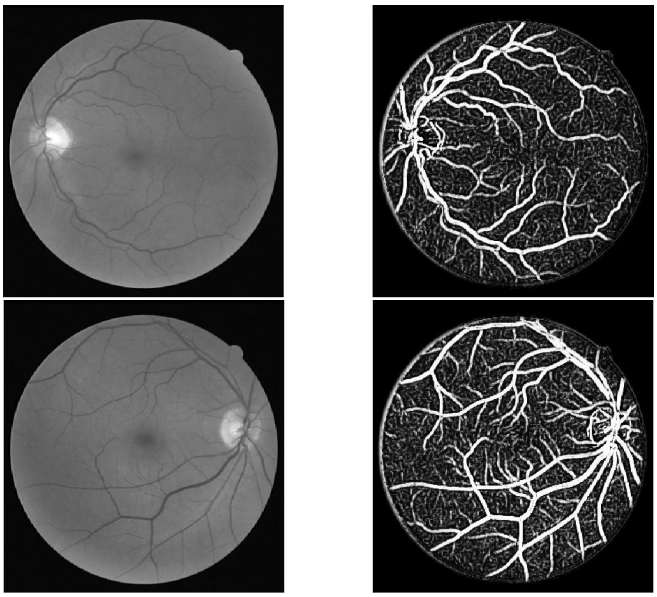

Fig. 4. *The posterior probability map ("soft decision") of our global vessel-bed segmentation result using test image 9 (top) and test image 16 (bottom).*

3.3. Cell Microscopy Data

The detailed experimental procedure to image endothelial cells in intercostal arteries in rabbit aortas is described in [11]. The 33 features used to classify each pixel included the image intensity patches, phase-invariant orientation field, and standard, single-pixel features including median, range and moment values. Full details are provided in [12].

3.4. SVM Training

An SVM [13] can be used to construct a separating hyperplane which maximizes the distance to the nearest training data inputs. If the data is linearly separable in the feature space implied by the kernel function then the decision rule (in standard algebra) will be:

$$f(\boldsymbol{\nu}) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\boldsymbol{\nu}, \boldsymbol{\nu}_i) + \beta\right) \qquad (2)$$

in which $\boldsymbol{\nu}$ is the normalised feature vector. N is the number of training examples. $y_i \in \{-1, 1\}$ are the scalar labels. K is the training kernel. $\{\alpha_i\}_{i=1,2,\ldots,N}$, and β are scalar parameters selected to maximize the margin that partitions the positively and negatively labelled data. $\{\boldsymbol{\nu}_i\}_{i=1,2,\ldots,N}$ are M-dimensional training vectors. The training examples ($\boldsymbol{\nu}_i$) are called *support vectors*, such that $\alpha_i \neq 0$. For this work, we

used `libsvm` [14] via its *Matlab* interface, and restricted our tests to the in-built kernels.

The Radial Basis Function (RBF) kernel $K(\boldsymbol{\nu}, \boldsymbol{\nu}_i) = \exp(-\gamma||\boldsymbol{\nu} - \boldsymbol{\nu}_i||^2)$ is used, where γ determines the area of influence $\boldsymbol{\nu}_i$ has over the data space. We find the RBF kernel effective for pixel-level segmentation of the cell data. The SVM is first trained using the input feature space with its *a priori* determined ground-truth on a small training image of size 128×128. The accuracy of the SVM model is strongly dependent on the selection of the model parameters, and our search over these parameters is computationally expensive, trying values of each parameter across a wide range. We search over the two "standard" parameters (the cost parameter C and the aforementioned γ), using a range of -4 to +4 with an interval of 0.1 for C and a γ range from 0.1 to 2 with an interval of 0.001. 5-fold cross-validation is done on the training image patch. This validation step is important to ensure the learned model is stable and generalises well to new data. Once the model with the smallest error is generated, segmentation of new data is achieved by casting a new image into 33-dimensional feature space, normalising the vectors and applying the SVM on a pixel-by-pixel basis. A sample (original) test image and binary segmentation training result is shown in Fig. 5 using the SVM-RBF approach. Once trained, this model was applied without further modification to the rest of the data. In the analysis that we report in the Section 4.2, samples were drawn from 56 microscopy slides, each of dimension around 1100×660 pixels.

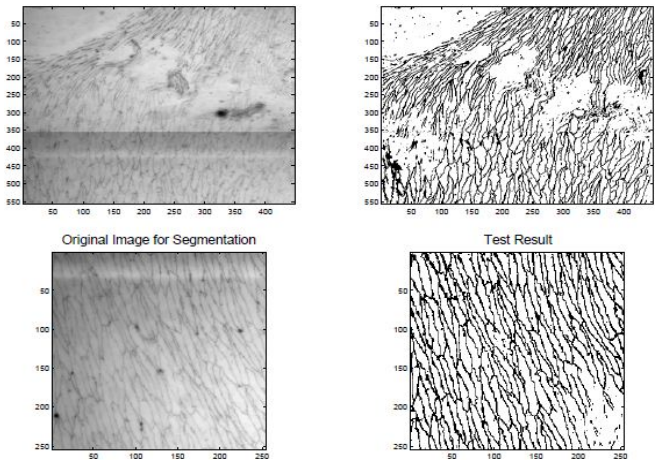

Fig. 5. *Left column: Original images; Right column: The segmentation results using SVM with RBF kernel. No post-processing is applied.*

4. PERFORMANCE EVALUATION

4.1. Retinal Vascular Segmentation

Since poor performance on a single image may be masked in an ROC curve for an image set, we apply the same neural network model on each of the 20 test images in the DRIVE database individually. The ROC for the entire test data superimposed on the ROC lower and upper bounds (per image) is depicted in Fig. 6. Our algorithm achieved 0.9779 for the area under the ROC curve (A_z), better than 0.93 reported by Niemeijer [8].

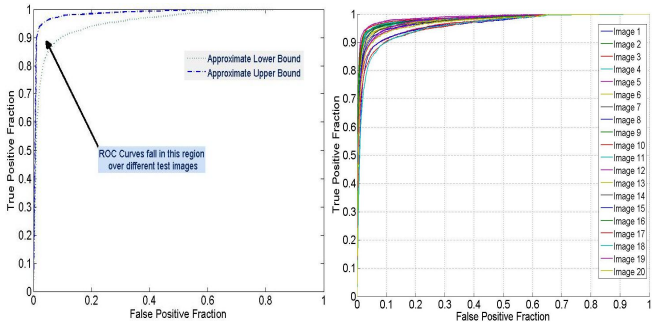

Fig. 6. *Left: The ROC envelope curves across DRIVE database; Right: The ROC curves for 20 test images using one neural network model. Individual results on each of the test images contributes to the lower and upper bound.*

4.2. Endothelial Cell Boundaries

The performance of the SVM approach to segment cell boundaries was evaluated on six images containing *manually traced* boundaries. The manually traced boundaries provide ground-truth data that allows the accuracy of the segmentation to be evaluated on a pixel-by-pixel basis. These are distinct to the manually traced boundaries used for training. Performance of an RBF kernel was compared against sigmoid and linear kernels (see Fig. 7).

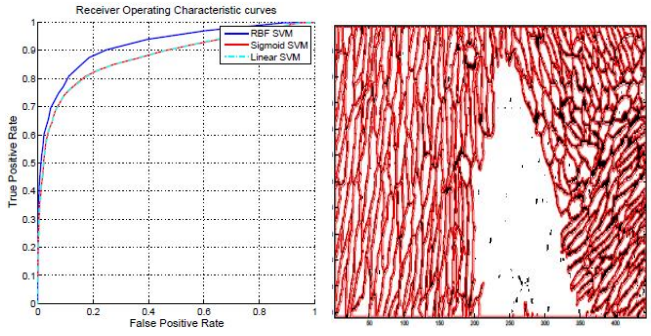

Fig. 7. *Left: ROC curves using different SVM kernels; Right: Segmented cell boundary (in red) superimposed on the original microscopic image using SVM+RBF kernel as the training model.*

4.3. Computational Effort

The computational effort of training the classifier and the process of then applying it to an image is of the order of minutes (see numbers below). Whilst this is not sufficiently fast for real-time performance, we envisage its usage to be in the laboratory or research setting, where accurate segmentation at pixel level is required, and the (easy) training of a system by a human expert is necessary. Classifying an image of size 512×512 takes under 1 minute on sub-optimal *Matlab* code, includ-

ing feature generation. We estimate that *Matlab* code optimisation could yield a factor of 4 improvement. The greatest limitation in the current implementation is memory size, because of the high dimensions of the feature vector. With code optimisation, a good representation choice for the feature vector (e.g. data requantisation) and re-implementation in C, this approach could be turned into a tool that would easily run on a laptop or tablet device.

5. DISCUSSION

In this paper, we proposed a patch-based approach to segmenting elongated structures, including retinal vasculature in standardized clinical colour fundus images and thoracic aortic endothelial cell boundaries in stained microscopy images. A key feature of the approach is its extensibility to other difficult segmentation problems. We have tested the approach on the DRIVE database and reported 0.9779 for the Area Under the (ROC) Curve (AUC, A_z) when evaluated across the entire test data set (a total of 20 test images) [15]. We also obtained good results on the microscopy data and managed to achieve 94% accuracy using best performing kernel on a total of 3.9 million pixels [11].

One key observation is that the performance of *patches* of features is more robust than a simple combination of selected features. This is not reflected in standard feature selection tools, as one cannot generalize the contribution of each individual element of a patch without prior information. In our experience, a powerful machine learning method, such as MLP or SVM, tends to interpret all the elements in the patch to reach a joint classification decision. Patches of non-linear feature spaces offer the potential for much better performance than single features, even if such features might be produced from operators of wide spatial support.

In future work, we plan to investigate feature selection methods that explicitly group fields of features, and further our pursuit of encoding such fields in a patch-based approach.

6. REFERENCES

[1] F. Leymarie and M. D. Levine, "Tracking deformable objects in the plane using an active contour model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 617 – 634, 2002.

[2] T. Mouroutis, S. J. Roberts, and A. A. Bharath, "Robust cell nuclei segmentation using statistical modelling," *Bioimaging*, vol. 6, pp. 7991, 2001.

[3] L. Vincent and P. Soille, "Watersheds in digital spaces: an efficient algorithm based on immersion simulations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583–598, 1991.

[4] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal of Electronic Imaging*, vol. 13, pp. 146–165, 2004.

[5] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63–84, 2000.

[6] M. E. Mavroforakis and S. Theodoridis, "A geometric approach to support vector machine (SVM) classification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 671 –682, 2006.

[7] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 948–959, 2005.

[8] M. Niemeijer, J. J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," *SPIE Medical Imaging*, vol. 5370, pp. 648–656, 2004.

[9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[10] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Trans. on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[11] A. R. Bond, S. Iftikhar, A. A. Bharath, and P. D. Weinberg, "Morphological evidence for a change in the pattern of aortic wall shear stress with age.," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 31, no. 3, pp. 543–550, 2011.

[12] S. Iftikhar, A. R. Bond, A. I. Wagan, P. D. Weinberg, and A. A. Bharath, "Segmentation of endothelial cell boundaries of rabbit aortic images using a machine learning approach," *International Journal of Biomedical Imaging*, vol. 2011, no. ID 270247.

[13] V. Vapnik, "The nature of statistical learning theory," *Springer-Verlag*, 1995.

[14] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," 2001.

[15] S. Cao, A. A. Bharath, K. Parker, J. Ng, J. Arnold, A. McGregor, and A. Hill, "Microvasculature segmentation of co-registered retinal angiogram sequences," *Annals of British Machine Vision Association*, vol. In Press, 2012.