Analysis of Retinal Vasculature using a Multiresolution Hermite-Gaussian Model

Li Wang, Student Member, IEEE, Abhir Bhalerao, Member, IEEE, and Roland Wilson

Abstract

This paper presents a vascular representation and segmentation algorithm based on a multi-resolution Hermite-Gaussian model (MHGM). A 2D Hermite-Gaussian polynomial intensity model is developed which models blood vessel profiles in a quad-tree structure over a range of spatial resolutions. The use of a multi-resolution representation allows robust analysis by combining information across scales and to reduces the computational complexity. Moreover, the local image model can accurately represent vessel directions, widths, amplitudes and branch points. A Fourier based modelling and estimation process is used, followed by an EM type of optimization scheme to estimate local model parameters. An information based process is then presented to select the most appropriate scale/model for modelling each region of the image. In the final stage, a stochastic inference approach within a Bayesian framework is employed for linking the local features to obtain a description of the global vascular structure. Experimental results on a number of standard retinal images are shown to demonstrate the application of the proposed algorithms. Some preliminary results on 3D data are also presented showing the possible extension of the methods to 3D branching structures.

Index Terms

Retinal Images, Hermite-Gaussian Modelling, EM, AIC, Kruskal M-Spanning Tree, Stochastic linking algorithm
Analysis of Retinal Vasculature using a Multiresolution Hermite-Gaussian Model

I. INTRODUCTION

A MAJOR task of medical image analysis is the extraction of an appropriate feature description to represent the meaningful content or structure of the image data including organs, anatomical features (e.g., blood vessels, tissues) and pathological features (e.g., lesions, tumours, exudates). A variety of medical image techniques such as magnetic resonance imaging (MR), computed tomography (CT) and fluorescein angiography are capable of obtaining data on human vasculature. In most images, however, signal noise and lack of image contrast pose significant challenges to the extraction of blood vessels. To this end, a number of methods have been proposed for the accurate segmentation of vascular structures (see [1] for a comprehensive review). As with any image segmentation method, its robustness to image noise and pathology, accuracy, and utility will rely greatly on the choice of image feature detection used, the appropriate use of local and global feature models, and the extent to which domain or prior knowledge is incorporated into the image model. Prior knowledge may take the form of contrast thresholds, signal and noise intensity distributions, object shape descriptions (templates, deformable or statistical shape models) or, in the case of vasculature, a tree or graph description of the anatomical features. In this work, we develop an image model for vascular segmentation and target it to the accurate segmentation of retinal vasculature.

In ophthalmic practice, angiography is a common procedure and several different techniques are commonly used, such as fluorescein and indocyanine green angiography and digital fundus photography. In fluorescein angiography a coloured dye injection into a vein in the arm of the patient an image of the retina is taken some 8-10 seconds [2]. Fundus photography without contrast agents is used more commonly for screening for hypertension and diabetes. A colour CCD camera will obtain red-yellow images of the retina, or coloured filters can be applied to produce monochromatic fundus photographs to highlight different pathologies. Nevertheless, the resulting images can pose significant challenges for human experts to interpret and, in the case of screening, manual assessment can be tedious and error prone [3]. The hope is therefore that quantitative, computer-based assessment of the characteristics and functionality of blood vessels for these angiographies may be faster and yield more consistent results than a manual examination process.

Central to such analyses is the detection of the blood vessels in an image [4]. This allows for a quantitative measurement of the geometrical changes of arteries, in their diameters, tortuosity or lengths and can provide the localization of landmark points (such as bifurcations) needed for image registration [5]. Furthermore, it has been shown that a segmentation of the vessel can be effectively used to measure the size of the optic disc and fovea, and estimate the leakage of blood into the retina (exudate) that is a significant indicator of retinal disease and diabetes [6].

Retinal vasculature has several distinct characteristics that can be used by image processing techniques to separate blood vessels of interest from the background. First of all, vessel cross-sectional intensity profiles approximate a Gaussian (or mixture Gaussian) in shape for fundus images. Secondly, the orientation and grey level of a vessel does not change abruptly; they are locally linear and gradually change in intensity along their lengths. Finally, vessels can be expected to be connected and, in the retina, form a binary tree-like structure. However, the shape size and local grey level of blood vessels can vary hugely and some background features may have similar attributes to vessels. Vessels crossing and branchings can further complicate the simplicity of the Gaussian profile model.

Many segmentation algorithms have been presented to provide either automated or semi-automated detection of vascular structure. These methods attempt to exploit one or more of the characteristics of retinal vessel and can be broadly categorised into: tracking-based approaches; neural network approaches; and template and model based approaches. Tracking based approaches work by first locating an initial point and then exploiting local image properties to trace the vasculature recursively [7]. They are essentially region-growers, albeit designed specifically for linear structures. These algorithms require user intervention to determine the thresholds for vessel versus background and the region seed point and the proposal of adaptive thresholds during the tracing process. Another set of approaches use neural networks for classification. In these kinds of methods, pre-labelled angiograms are used as the training set to determine the network’s weight [8]. Both tracking and neural network classifiers do not extract the vascular structure directly and a post-process of line-thinning or skeletonisation has to be applied to estimate the global connectivity. Line-thinning is problematic as small variations in the local geometry in relation to the pixel sampling grid can lead to different estimates of the vessel’s medial axis. Thinning is also notoriously sensitive to noise and can result in ambiguous results at bifurcations. The results will lead to inaccuracies in width, vessel direction and tortuosity measurements, but also provide misleading connectivity inferences.

The third class of methods is template or model based approaches. The ‘matched filter response’ method is a widely used template-based technique, first introduced by Chauduri [9] and further developed by Hoover [10] and others, e.g. [11], [12]. A set of 2D Gaussian kernels with fixed size and orientation are used to enhance the vessels. A
local threshold is then set to differentiate their outputs from retinal background features. The method has some interesting characteristics such as handle bifurcations, to some extent (by detecting the conjunction in peak outputs of more than one filter), and obtaining fairly robust separation of vessels from background. However, it is relatively time consuming as the convolution kernel used may be quite large and needs to be applied repeatedly. In addition, the kernels respond optimally to vessels that have a fixed widths, expressed as the standard deviation $\sigma$ of the underlying Gaussian function of the convolution kernel. Therefore, they may output weak responses to thin vessels as well as very thick vessels. In addition, the vessels with high tortuosity may also be left out from the kernel response. The matched filter response output can thus locate likely vessel pixels if they match the fixed set of input widths and orientation, but they do not explicitly estimate the local vessel width or orientation. Additionally, a post-process is still required to join out the filter responses either by tracking/region-growing followed by a skeletonisation or by a some other labelling processes.

The central light reflex which is the specular reflection of the light source during imaging along large surface vessels is an artefact which confounds the Gaussian profile assumption. To address this problem, several authors have suggested the extension of the profile model to use mixtures, Laplacian of Gaussians or Difference of Gaussians [13], [12], [9]. This complicates the matched filtering since it conlates the problem of what size Gaussian to use with the need to estimate a mixing and second size parameter. Our solution is to use a Hermite-Gaussian model that requires a single mixing parameter and generalises our vessel profile model without the need to change the estimation strategy. Another artefactual problem is intensity inhomogeneity across the image – so called, “vignetting” [14] and is generally handled by local intensity normalisation (e.g. [15], [16]). Here, our local feature model incorporates a piece-wise linear background variation which is estimated simultaneously with the Hermite-Gaussian vessel model.

Methods by Martínez-Pérez [17] and Staal and Niemeijer et. al [18], [15] both use a Gaussian scale-space and ridge features. This is followed by a region grower or a grouping process; Staal et. al [15] use a supervised classification step using 20 or so features in the neighbourhoods of the nominal vessel centre lines. By combining ridge features across multiple scales, the local vessel size is decoupled from the model. Our multiresolution approach achieves the same end by local fitting across as set of fixed resolutions on a pyramid of the image and then performing a model selection process which trades off model complexity against residual error to find an optimal selection scale. Unique to our approach is the use of a topological prior that seeks out binary trees in the image [9], [7]. This allows us to extract out a meaningful global description of the vasculature that is robust to noise and obviates the need for a post-labelling skeletonization step.

The main goal of this work has been to develop a branching structure segmentation and detection which can be used for detecting blood vessel structures in retinal images that tries to improve on both the low-level feature estimation and the subsequent vessel tracking/labelling processes by providing a unified, model-based representation. To achieve this goal, we have developed a vessel modelling and estimation technique that operates directly on the image intensities for detection of the blood vessels. Vessel segments with their local directions, widths and amplitudes, and bifurcations are identified explicitly by a local model. To account for changes in scale of the features of interest, a block-based multi-resolution approach is employed followed by an EM type of optimization scheme to fit local model parameters. The local models are combined across scales to produce a data summary by a model selection process. Then a stochastic Bayesian approach is used to link the local models, i.e. vessel segments and bifurcations, and infer the global vascular structure under the assumption of a binary tree structure.

The paper is organised as follows. The vessel model is introduced next and the local-to-global estimation strategy is described in the following section. Two ways to infer the global vascular topology are proposed: a heuristic-linking method which uses a modified maximum spanning tree algorithm; and a stochastic linking method which employs a Bayesian framework and a Metropolis sampling algorithm. We evaluate our methods on a both the standard image sets from the STARE database [10], and those from the DRIVE database from Utrecht’s group [18] for selectivity, specificity and accuracy in the standard way. We further demonstrate width and tortuosity estimation and the detection of bifurcations with our model. We conclude by discussing the merits and limitation of our methods and make suggestions for future work.

II. A MULTiresolution HERmite-GAUssian Model (MHGM) FOR 2D VASCuLATE

Retinal images are modelled as consisting of a forest of binary trees, $\mathcal{F} = \{T_1, T_2, ..., T_n\}$, where each tree $T_i$ is made of the joint set of vertices $X$ and a set of links or edges $L$. Each vertex $x_i \in X$ is restricted to have degree at most 2 i.e. the trees are binary. Furthermore, vertices are made to lie in disjoint, square regions size $2^l$, e.g. $1 \times 1$, $2 \times 2$, $4 \times 4$, etc. For any set of vertices, $X$, in the image plane, the disjoint square regions can be discovered by recursively subdividing the plane into quadrants until only 1 vertex occupies any given quadrant, or is approximated by a pixel (figure 1). Square image regions are modelled as a vessel segments over background variation and noise:

$$f(x) = G(x) + B(x) + n$$

where $G(x)$ and $B(x)$ are the foreground and background intensity variation in the block and the noise $n$ independently distributed Gaussian noise $n \sim N(0, \sigma^2)$ with standard deviation $\sigma$. Each link or edge, $l_{ij} = \{x_i, x_j\}$, is parameterised by the bi-variate Gaussian kernel,

$$G(x) = A \exp(-{(x - \mu)^T R^T \Sigma^{-1} R(x - \mu)}/2),$$

where $A$ is the amplitude of the edge, $R$ is a 2D rotation matrix which rotates the link $l_{ij}$ to the x-axis and $\Sigma$ is a diagonal covariance matrix representing the length and widths
of the kernel. The image background, $B(x)$, is modelled in
dpiecewise linear form across each square block:

$$B(x) = \alpha x + \beta$$  \hspace{1cm} (3)

The model estimation strategy is bottom-up: the image is
divided into overlapping square regions over a range of sizes
from $8 \times 8$ to $64 \times 64$ pixels, and parameters of the com-
bined Gaussian vessel model and the piece-wise linear model
estimated. Disjoint regions are then selected from this over-
complete image representation. Each image region is assigned
a vertex, $x_i$, and a curve-tracing process is initiated by a
neighbourhood linking process. This process is designed to
infer $F$ given the data by maximising the posterior probability
$P(F|f(x))$.

\subsection{A. Local Feature Analysis}

The global shape of the retinal vessel structure (figure 2(a)),
in general, cannot be modelled by a single primitive form,
therefore any model can only represent the local shape of a
vessel in a small region (figure 2(b), (e)). Here, we exploit
the symmetry and translational invariance properties of the Fourier
domain to model linear and branching structures.

The local feature models derived in this section are based
upon the characteristics of local spectra. Any given image
regions of varying sizes are modelled using their local spectra,
and the model parameters are then estimated from the repre-
sentation.

\subsubsection{1) Gaussian Intensity Model}

In the spatial frequency do-
main, a single linear feature (figure 2(b)) can be approximated
by a 2-dimensional Gaussian function $G()$, whereas multiple
linear features within the same region (figure 2(c)) are mod-
elled as a superposition of Gaussian models and the spectrum
is approximated as a sum of component spectra:

$$G(\omega) = \sum_{m=1}^{M} |G_m(\omega)| \exp[-j \phi_m(\omega)]$$  \hspace{1cm} (4)

where $G_m(\cdot)$ is the $m$th Gaussian feature and $\phi_m(\cdot)$ is the
 corresponding phase spectrum.

In particular, the phase-spectrum of a component exhibits
a phase variation which is related to the feature centroid. We
assume that this phase variation is independent for each of the
$M$ components of the model [19]:

$$\phi(\omega) = \sum_{m=1}^{M} \phi_m(\omega_m) = -\frac{B_l}{2\pi} \sum_{m=1}^{M} \mu_m \cdot \omega_m,$$  \hspace{1cm} (5)

According to the Fourier shift theorem, the centroid $\mu_m$ can
therefore be estimated within the block, $B_l$, by taking the av-
erage pairwise correlations between neighbouring coefficients
along each axes.

Feature components are estimated by first separating out
the component centroids using K-means clustering. Using the
$M$ spatial frequency coordinate partitions, the orientations
and the widths (as variances) of the Gaussian components can be
estimated by performing PCA on the covariances matrix, $C_m$,
which is calculated from the inertia tensor of the spectral
energy [20]. Figure 3 illustrate single and double feature
estimation on a region taken from a retinal image.

\subsubsection{2) Optimization Algorithm}

The above Fourier based esti-
mation is computationally efficient and accurate in noise free
data. However, because of uncertainties in most real images,
an optimization step is needed to improve the initial estimation
$\Theta_0$. An iterative Minimum Mean Square Error (MMSE) type
of approach is employed to minimise the error between model
$G_m(\cdot)$ and windowed data $f_B(\cdot)$.

$$arg\min_x \sum_x (f_B(x) - G_m(x|\Theta))^2$$  \hspace{1cm} (6)

which can be regarded as an error function.

This is equivalent to maximising the sample statistics $\Theta$
weighted by the inner product between the windowed data,
$f_B(x)$, and the model $G_m(\cdot|\Theta)$. The regression leads to a set of
equations which are similar to the Expectation Maximisation
(EM) algorithm [21] (Appendix I):

$$A^{(t+1)} = \frac{\sum_x W(x|\Theta^{(t)})}{\sum_x G^2(x|\Theta^{(t)})},$$  \hspace{1cm} (7)
\[ \mu^{(t+1)} = \frac{\sum x W(x|\Theta^{(t)}) x}{\sum x W(x|\Theta^{(t)})} \]
\[ \Sigma^{(t+1)} = \frac{\sum x W(x|\Theta^{(t)})(x - \mu_t)(x - \mu_t)^T}{\sum x W(x|\Theta^{(t)})} \]

Convergence is achieved rapidly in 7-10 iterations. In figure 4, the estimation is tested in a sample region using a two component mixture. Note that the parametric description of the model \( \Theta \) has information about feature amplitude and width together with position and orientation. A plot of MSE across all the blocks for 2, 5, and 10 iterations is given in figure 5.

3) **Background Estimation**: Background estimation is carried out as follows...

Explanation of how to get an estimate of \( B(x) \).

4) **Hermite Approximation**: Blood vessel segments are approximately Gaussian in shape after the subtraction of the piecewise linear approximation of the background and applying a Hanning window each image block \( B \times B \). However, as reported by other researchers [22], [23], in some special cases, the intensity profile of some large vessel segments is not exactly Gaussian due to the so called 'central light reflex' caused by specular reflection (see figure 7(top-row)). Note that this happens more frequently in digital fundus images since the light reflex is comparatively larger than in fluorescein images [23].

For the above special cases, Gaussian intensity modelling tends to over-segment the features. Gao et al. [22] proposed an algorithm which uses the difference of two separate Gaussian functions to model each blood vessel segment. The model is able to fit the data well, however, doubles computational complexity since the number of parameters is doubled.

Our approach to solve this problem uses a Hermite approximation applied after the above optimization scheme which introduces only one additional parameter [24]. The 1d Hermite polynomial is defined as;

\[ H_n(x) \equiv e^{x^2/2}(x - \frac{d}{da})^n e^{-x^2/2} \]

By Cartesian separability, its 2D extension can be represented as:

\[ H_{mn}(\Theta, a) = H_n(x)H_m(y) \]

Here, a second order Hermite polynomial \( H_{2,0}(x) \) is used with an adaptive parameter \( a \) as a factor that is multiplied by the optimized Gaussian \( G(x|\Theta) \), i.e.

\[ H_{2,0} = (1 + a(x^2 - 1))G \]

We can see from figure 6, a 1D plot of equation 11 with different \( a \), that when \( a = 0, H \equiv G \) and the two peaks are further apart when \( a \) increases. Note that \( a \) remains as a scalar parameter perpendicular to the orientation of the Gaussian when we perform 2D modelling.

The estimation results on four sample images are shown at figure 7. From the MSE comparison result, we can see that more accurate approximation can be derived by using a Hermite equation rather than a simple Gaussian formula.

5) **Hierarchical Feature Analysis**: As suggested by Box and Jenkins [25], the principle of model selection “...using the smallest possible number of parameters for adequate representation of the data” is needed from a statistical point of view. The principle of model selection is a bias versus variance tradeoff. In general, bias decreases and variance increases as the number of parameters increases. The fit of any model, generally, can be improved by increasing the number of parameters. However, a trade off with the increasing variance must be considered in selecting a model of inference [26].

The technique described in this work, following Calway’s [27] and Bhalerao’s [28] method, uses a hierarchical quad-tree structure to perform feature estimation at different
levels, i.e. different window sizes. Then a Kullback-Liebler type of penalized distance measure, the Akaike information criteria (AIC) is used to bias the likelihood function with the number of parameters $P$.

The image is firstly estimated using multiple feature models $m = 1, 2$ over a predetermined range of levels $B_l$, starting at $B_{l1}$ and descending maximally to level $B_{ln}$. If the block sizes at each scales, $l$, are chosen to increase by a factor of two, $B_{l+1} = 2B_l$, then each four ‘child’ blocks, $c_0 = (2i, 2j, l)$, $c_1 = (2i+1, 2j, l)$, $c_2 = (2i, 2j+1, l)$, $c_3 = (2i+1, 2j+1, l)$, will correspond to one ‘parent’ block $p = (i, j, l + 1)$ at the next level and forming a quad-tree feature representation of the data which is introduced by Klinger [29]. Unlike a traditional top-down fashion of quad-tree structure, however, we use a bottom-up grouping approach:

1) Without making decisions on whether the measures for a local image region are good or not, feature estimation is first carried out at every scale.
2) The process starts at finer scales of the image and merges each four ‘children’ blocks to one ‘parent’ block according to the scale selection criteria.

A bottom-up type of approach like this is better for exploring fine features but requires more computation than a top-down technique that can prune out subtrees during selection.

Once the quad-tree structure is set up and the parameters of each model at each scale have been estimated, the most fit model and ‘natural’ scale for the features is determined for a given image region by using a penalised distance measure, the Akaike information criteria (AIC). This biases the residual fit error which is expected to fall with the increasing number of parameters $P$:

$$AIC = N^2 \log(RSS/N^2) + 2P \quad (13)$$

where $N^2$ is number of data and RSS stands for the residual sum of squares $\sum \hat{\epsilon}^2$ (Appendix II).

From a heuristic point of view, equations 43 and 44 can be interpreted as a measure of lack of model fit plus a penalty term for increasing the number of parameters, shown as a tradeoff between bias and variance. However, the penalty term in AIC is not arbitrary but derived from an asymptotic estimator of relative, expected K-L information or distance between
model pairs. Minimising the AIC provides an estimated best approximating model for that particular data set.

The recursive algorithm of this multi-resolution representation can be summarized as follows:
1) Estimate the initial parameters for each model \( m = 1 \) at some starting block size.
2) Use equations 6–8 to improve the first estimate for each model \( m \).
3) Compute \( AIC \) from \( RSS \) using equation 43 or 44 (which depends on the \( N^2/P \) ratio).
4) Repeat steps (1) – (3) for \( m + 1 \leq M \).
5) \( m_{opt} = \arg_m \min(AIC) \).
6) Repeat steps (1) – (5), for \( l + 1 \) until \( m_{opt} \) is calculated at all levels.

According to the quad-tree structure, the ‘natural’ scales of the features can also be determined by taking the smaller value of the AIC summations of child blocks across scales and comparing the information number with their parents;

\[
\sum_{k=0}^{3} AIC(c_k) \rightarrow AIC_m(p) \quad (14)
\]

where \( c_k \) represents four ‘child’ blocks \( c_0 = (2i, 2j, l), c_1 = (2i+1, 2j, l), c_2 = (2i, 2j+1, l), c_3 = (2i+1, 2j+1, l) \) and \( p \) is the corresponding ‘parent’ block \( p = (i, j, l+1) \). The block(s) with the minimum AIC value is then picked to represent the data in a given region. Figure 8 is a flow chart that shows an overview of this multi-resolution estimation algorithm.

To verify the algorithm’s performance, we first test the method on a noise free image. Figure 10 is generated by an original retina fundus image multiplied by the hand labelled image to eliminate the background noise. Figure 10 (b) & (c) shows the reconstruction result of blocks using single and multiple Hermite-Gaussian models respectively. Different block sizes are highlighted that indicate the region is modelled at different levels. As we can see from the result, the main vasculature, small blood vessels, bifurcations and crossing points are modelled accurately using the above multi-scale Hermite-Gaussian representation. The mean square errors of the two reconstruction images are also calculated. The experiments were also carried out on the entire dataset which contains both healthy retinal images and those with pathology. Figure 9 shows the plot of MSE of the whole data set using Hermite-Gaussian approaches.

![Fig. 9. Mean Square Error of reconstruction result using H-G modelling on the whole data set.](image)

To assess the robustness of the modelling algorithm, Gaussian additive noise is added to figures 10(a) at 0dB. The SNR is calculated as [28] [30],

\[
SNR = 20\log_{10}\left(\frac{\mu_f - \mu_b}{\sigma_n}\right) dB \quad (15)
\]

where \( \mu_f \) and \( \mu_b \) are the gray levels of the vessel and background and \( \sigma_n \) is the standard deviation of the additive Gaussian white noise.

Figures 11, 12 shows the reconstruction results under the noisy data. The results illustrate the immunity of the algorithm against such noise. Overall, the Hermite-Gaussian modelling followed by a ML type of estimation and AIC model/scale selection scheme works well. The orientation, position and width of the features both along blood vessels and near bifurcations are accurately modelled on both noise free and noisy data. Table I summarises the mean square error of the reconstruction results using both approaches for 10 images with and without noise. We can see that by using the Hermite approximation on average reduce the reconstruction error by 23%.

### B. Global Structure Inference

For a complete segmentation of the image, an analysis of geometrical and topological property of the global structure are needed. In this work, two algorithms are reported for measuring the global topology of vasculature. Firstly, a heuristic linking algorithm using a graph representation is proposed. Using a modified Kruskal type of maximum-cost spanning tree algorithm, the linking process is reduced to finding the optimum path in a graph representation of the image. The deterministic linking approach is able to successfully track and characterise much of the vessel topology. However, it is prone to becoming trapped in local maxima and unable to explore less certain, alternative explanations of the data.

The second approach discussed in this chapter, includes the prior knowledge of anatomical structure. A Markov chain is employed to sample from the posterior distribution, given the local feature estimates. The sample distribution is an approximate equilibrium of a random process configured in the space of a tree-like structure. As well as gaining information
about the global structure, variation in the posterior samples indicates the uncertainties about the image interpretation.

1) Heuristic Linking Algorithm: If we regard vessel segments at each block as vertices and all linking possibilities among neighbouring blocks as edges, the linking process can be reduced to finding the optimum path in a graph representation.

The cost of arcs between each adjacent blocks are calculated using the Gaussian Product Theorem. The theorem states that the product of two arbitrary angular momentum Gaussian functions on different centres can be written as a form of third Gaussian $G_1 \cdot G_2 = G_3$. (The proof can be found in Appendix B). Using this theory, the probability that two adjacent blocks $i$ and $j$ are part of the same vessel is estimated by calculating a link weight $W_{ij}$ in an n-neighbour system by integrating the product of the two Gaussian feature models along a line.
y(t) = \mu_i + (\mu_j - \mu_i)t, \ 0 \leq t \leq 1, \text{ joining their centroids. The number of neighbours } n \text{ vary according to the block sizes.}

W_{ij} = e^{-(A_i - A_j)^2/(2\sigma_A^2)} \int_0^1 G\{A_i, \mu_i, y(t)\} \cdot G\{A_j, \mu_j, y(t)\} dt (16)

where \( e^{-(A_i - A_j)^2/(2\sigma_A^2)} \) is a coefficient which is formed by a normal distribution of the difference of the amplitude and models the likelihood of changes in amplitude between features. \( \sigma_A^2 \) is estimated from the data.

The standard graph-theoretic adjacency matrix for a graph has elements that indicate whether the corresponding vertices are connected. In the absence of loops and multiple edges: if the linking state \( A_{ij} = 0 \) then \( V_i \) and \( V_j \) are not connected, if \( A_{ij} = 1 \) then \( V_i \) and \( V_j \) are connected. Using function (15), the nodes can then be connected using a modified Kruskal (mKruskal) method for finding a MST of a weighted graph, \( W_{ij} \), as the cost along the arcs [31]. The algorithm creates a forest of trees. Initially the forest consists of \( n \) single node trees and no edges. At each step, each entry to the graph indicates the smallest vertex number to which vertices are connected; eg, if \( K[j] = [i] \), then \( i \leq j \) and \( i \) is the vertex of smallest numeric label to which vertex \( V_j \) is connected. Edges are progressively added to the graph. At each step, we add the edge with the highest cost (\( W_{max} \)) without creating circuits.

Based on the biological nature of the data, we further constrain the resulting tree to be binary in this application, i.e vertices can have degree 3 or less. The algorithm proceeds as follows:

1) Pre-sort the edge costs, \( E \), into descending order of weight.
2) Examine each edge and add to the tree \( T \) if it satisfies:
   a) It connects two vertices from different components.
   b) The degree of vertices at each end is currently less than 3.
3) Terminate when the weight of the current edge is lower than some user defined threshold.

The experimental results presented in this section illustrate the operation of the heuristic linking process using a modified Kruskal MST algorithm. To test the robustness of the method, noise was added to the clean image at $SNR = 1$ and original noisy images (see figures 13 and 14). The results after the linking process are shown on the right hand side of figure 13, different colours indicate the separate tree structures. The results demonstrate the ability of the greedy algorithm to successfully track and characterise much of the vessel topology, however, the linking result degrades as the uncertainty of the data increases, figure 14. To overcome this drawback, prior knowledge can be incorporated into the simulation and locally uncertain solutions in the data can be explored, which leads to a Markov chain type of algorithm.

2) Stochastic Linking Algorithm: As shown in the previous section, the Kruskal deterministic algorithm can produce a fast and reliable neighbourhood linking result in low noise images. However, since the heuristic linking algorithm is a greedy operation, at no stage does it try to look ahead more than one edge and is prone to becoming trapped in a local maximum. A Bayesian, stochastic type of approach can be used instead of the deterministic algorithm to explore less certain, alternative explanations of the data.

The probability of the link state given the data in the neighbourhood can be written as the probability of the link state given the parameters of an edge between a neighbouring block

$$P(l_{ij}|f(\Theta))$$

Using Bayes’ law of conditional probability equation 16 can be rewritten as;

$$P(l_{ij}|f(\Theta)) = \frac{P(l_{ij}P(f(\Theta)|l_{ij})}{P(f(\Theta))}$$

$$\propto P(l_{ij})P(f(\Theta)|l_{ij})$$

where $P(l(\zeta, \xi))$ is a predefined prior distribution which, in the simplest case, works by setting $P(l(1) = 1) = 2/N$ for blocks represented by a single Hermite-Gaussian model and $P(l(1) = 1) = 3/N$ for multiple H-G models, where $N$ is the number of neighbouring blocks. The sample distribution $P(f(\Theta)|l(\zeta, \xi))$ can be defined by;

$$P(f(\Theta)|l_{ij}) = \phi(\gamma_\nu|\gamma_\eta, \eta \in A(\nu)) \times L(l|f)$$

in which $\phi(\gamma_\nu|\gamma_\eta)$ is the distribution of the parameters of the vertex, $\gamma_\nu$ [32]. It forms a first order auto-regressive process $AR(1)$. That is, the location of the root vertices $\gamma_\eta$ are selected by choosing the centroid of the Hermite-Gaussian model which has biggest width $w_{\text{max}}$ among the sampling windows. The location of the children $\phi(\gamma_\nu)$ are selected by a directional Gaussian distribution centred at the root. The $AR(l)$ process is there to ensure the tree models do not get tangled in a small area and to govern the branch linking/growing process. By transforming the original image pixels into a parametric H-G representation, the likelihood function $L(., f)$ can be estimated by integrating the product of the two Gaussian feature models along a line $y(t) = \mu_i + (\mu_j - \mu_i)t$, $0 \leq t \leq 1$, joining their centroids.

$$L_{ij}(l = 1|f) = e^{-(A_i - A_j)^2/(2\sigma_i^2)} \int_0^1 G_i \cdot G_j dt$$

$$L_{ij}(l = 0|f) = (1 - e^{-(A_i - A_j)^2/(2\sigma_i^2)}) \int_0^1 (1 - G_i) \cdot (1 - G_j) dt$$

where $e^{-(A_i - A_j)^2/(2\sigma_i^2)}$ is a coefficient which is formed by a normal distribution of the difference of the amplitudes, and models the likelihood of changes in amplitude between features ($\sigma_i^2$ is estimated from the data).

In order to arrive at a maximum posterior (MAP) estimate from a given image, it is necessary to sample from the p.d.f. A Metropolis algorithm is used to generate a sequence of selections from the distributions as follows [33]:

1) Start with any initial value $\theta_0$ satisfying $p(x) > 0$

2) Using current $\theta$ value, sample a candidate point $\theta^*$ from some jumping distribution $q(\theta_1, \theta_2)$, which is the probability of returning a value of $\theta_2$ given a previous value of $\theta_1$. The distribution is also referred to as the proposal or candidate-generating distribution. The only restriction on the jump density in the Metropolis algorithm is that it is symmetric, i.e. $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$.

3) Given the candidate point $\theta^*$, calculate the ratio of the density at the candidate ($\theta^*$) and current ($\theta_{t-1}$) points,

$$r = \frac{p(\theta^*)}{p(\theta_{t-1})}$$

4) If the jump increases the density ($r > 1$), accept the candidate point (set $\theta_t = \theta^*$) and return to step 2. If the jump decreases the density ($r < 1$), then with probability $r$ accept the candidate, else reject it and return to step 2.

The Metropolis sampling can be summarised as

$$r = \min\left(\frac{f(\theta^*)}{f(\theta_{t-1})}, 1\right)$$

The proposing move will be accepted with probability $r$. The equation 22 leads to the generation of a Markov chain as the transition probabilities from $\theta_t$ to $\theta_{t+1}$ depend only on $\theta_t$ and not any previous stages $(\theta_0, ..., \theta_{t-1})$. To eliminate the symmetric requirement of the Metropolis test, i.e. $q(\theta_1, \theta_2) = q(\theta_2, \theta_1)$. Hastings generalised the Metropolis algorithm by using an arbitrary transition probability function $q(\theta_1, \theta_2) = Pr(\theta > \theta_2)$, and the testing equation defined as

$$r = \min\left(\frac{f(\theta^*)q(\theta^*, \theta_{t-1})}{f(\theta_{t-1})q(\theta_{t-1}, \theta^*)}, 1\right)$$

An iterative simulation is then designed based on the above formulae (17-20), in order to maximise a posteriori probability. At each iteration, the Metropolis-Hastings test [34] is used to determine whether to accept or reject the linking between the visiting nodes and each of their neighbours using equation 23;

1) Calculate the proposed posteriori density $Pr(\theta^*|f)$ at time $t$.

2) Calculate the ratio of the densities

$$r = \frac{Pr(\theta^*|f)}{Pr(\theta_{t-1}|f)}$$
3) Set
\[
\theta^t = \begin{cases} 
\theta^* & \text{with probability } \min(r, 1) \\
\theta^{t-1} & \text{otherwise.}
\end{cases}
\]  
(26)

This is effectively a simplified form of a Monte Carlo simulation using just one move at each iteration. As a result of this, the simulation rapidly converges to a MAP estimate. Figure 15 shows the simulation converging to its stationary distribution as the iteration increases. The first experiment was carried out on a set of images with different signal to noise ratios (SNR). Figures 13 and 14 show the performance of the stochastic linking algorithm on a noise free image, an image...
with 0dB white noise and the image with original background. As we can see from the result, the simulation is able to explore the majority of the vessel structures despite the increase in noise.

The results shown above illustrate the ability of a Markov Chain type of stochastic linking strategy on exploring the vessel topology on the data with high uncertainty. The algorithm needs approximately 10 million iterations to converge to its equilibrium distribution which takes 75 mins to process the example data (#IM0077) on a Sun Ultra-10 machine. The results demonstrate the robustness of the method when the uncertainty of the data increases compared with the deterministic algorithm. Since the random sampling is restricted to 1 move, the convergence of the simulation is faster than a full MCMC algorithm. However, it requires a good estimation of the parameters of the modelling, including orientation, centroid and width of any blood vessel segments.

### III. Evaluation and Comparison

#### A. Materials and Methods

This section gives a quantitative analysis to verify the performance of the presented MHGM algorithm. For this purpose, two validation studies against manual measurements for diameters and branching angles were undertaken. Automatic measurements of individual bifurcations were compared with manual measurements for randomly chosen bifurcations from fundus retina images. The final classification results were also measured against hand-labelled results. Specificity/sensitivity plots were produced in the standard way to demonstrate the algorithm’s performance. By means of comparison, two other commonly used methods are also discussed in length. Results from both methods are presented and a comparative analysis is given towards the end of the section.

Comparative assessments were carried out two retinal image databases: on 20 fundus image data sets provided by Hoover et al. [10] (the STARE database); and 36 fundus images provided by Staal et al. [15] (the DRIVE database). The STARE images were digitised slides captured by a TopCon TRV-20 fundus camera at 35 degree field of view. The slides were digitised to $700 \times 605$ pixels, 8 bits per colour channel. There were hand labelled images in the database by two nominal experts.

Evaluation results are presented below for both databases against the best results of Hoover and Staal’s methods and the GIMM method against the manually labelled ground truth. Since the GIMM results are yield a parametric representation of the original image, we binarise the image into vessel and background classes by setting different threshold values for the standard deviation along the principal axes of each vessel feature. The performance of the GIMM system is then measured with a receiver operating characteristic (ROC) curve [35]. An ROC curve plots the sensitivity (SE) against specificity (SP) of the true/false positive (TP/FP) and true/false negative (TN/FN) rates estimated from the pixel classification data:

$$\text{Specificity(}	ext{SP}) = \frac{FN}{FP + FN}$$

$$\text{Sensitivity(}	ext{SE}) = \frac{TP}{TP + TN}$$

The closer a curve approaches the top left corner, the better the performance of the system. A single measure to quantify this behaviour is the area under the curve, $A_{\text{roc}}$, which is 1 for a perfect system. A system that makes random classifications has an ROC curve that is straight line through the origin with slope 1 and $A_{\text{roc}} = 0.5$.

#### B. Preprocessing

As noted in [40], the optic disc and exudate appear as bright patterns in colour fundus retinal images. Most of them contrast well against the background because the grey level variation caused by these patterns are often similar to that caused by the vessels, and the appearance of optic disc and lesions becomes one of the main reasons for the failure of vascular structure detection [39] [41]. Since our model does not cater for these regions, a preprocessing step has to be carried out prior to the vessel detection, to remove the optic disc and exudate in order to reduce its interference on vessel segmentation.

A simple and effective morphological filtering technique is used to localize the candidate region of both the optic disc and exudates [40]. The vessels are eliminated by using a closing operator $\mathcal{I}_1 = (f \circ b)$ with a circle structuring element $b$ bigger than the maximum width of vessels (figure 16). The candidate region can then be found by thresholding and reconstructing the derivative of $\mathcal{I}_1$ (figure 16(a)). In order to find the contours.

![Fig. 15. The posterior value of the first million iterations shows that the distribution converges when the iteration increases (image #IM0077).](image-url)
of the exudates, all candidate regions are set to 0 in the original image and the morphological reconstruction is then calculated. This operation propagates the values of pixels \( x \) next to the candidate regions into the candidate regions by successive geodesic dilation, with the original image \( f \) acting as the mask. The contours of exudates and the main part of the optic disc are also obtained by applying a simple threshold operation to the difference between the original image and the reconstructed image (figure 16(b)).

C. Results
D. Measuring the Width and tortuosity

The width, orientation and connectivity measurements that GIMM provides could be usefully used in clinical practice since there are multiple eye diseases that affect the geometry and topology of the retinal vasculature. For instance, hyper-tension increases retina artery dilation by as much as 35% [36]. Age and hyper-tension are thought to cause changes in the bifurcation geometry of retinal vessels [37]. Additionally, retinal arteriolar narrowing is an indication of the onset of diabetes and may be related to the risk of coronary heart disease for female patients [38]. Many of the diseases such as hyper-tension, angiogenesis and blood vessel congestion can also increase the tortuosity of the blood vessels[39]. Typical vein diameter is about 100\( \mu m \) near the optic disc and gradually reduces to approximately 20\( \mu m \) towards the end of the vessel. It translates into 4-5 pixel diameter and can be as thin as 1 pixel depending on the resolution [42].

From the parametric Hermite-Gaussian representation of the image, further experimentation has been carried out to produce the binary output of the gray-level reconstruction image. We use a simple iterative minimum mean square error (MMSE) fitting method to determine a threshold on the standard deviation of the H-G model on the minor axes which are perpendicular to the principal orientation of the vessel segments. A vessel/non-vessel classification can be produced based on the chosen threshold at each block of the H-G model. (see figure 17). The performance is examined by comparing the sensitivity vs specificity. Table II shows the sensitivity and specificity result against the hand label images across the data set comprising 20 images. An optimum choice of threshold using the MMSE method yields a sensitivity of 82% and a specificity of 93%.

Classification results can be verified by comparing the sensitivity and specificity. However, this calculation is pixel based and does not take into account the object size. To verify how well the algorithm is detecting thick vessels and small vessels, we measured the width at every point in the vessel segment skeleton. The width at a skeleton point \( S_p \) is defined as the largest line segment passing through the point \( S_p \) that is perpendicular to the principal orientation of the vessel segment. The figures 17(c) show the width histograms of the H-G classification results and the corresponding hand labelled images. The similarity between the two histograms shows that H-G segmentation and the classification method are able to detect the majority of main blood vessels as well as small vessel segments. Based on the measurement of arc length, the tortuosity of the vascular structure can be quantified. Information about disease severity or change of disease with time may be inferred by measuring the tortuosity of the blood vessel network [43].

The length and curvature of the blood vessels are calculated based on a skeletonized binary image representing the centre line of vasculature (figure 21(b)). The arc length of a vessel segment \( C \) is defined as

\[
s(C) = \int_{t_0}^{t_1} \sqrt{x'(t)^2 + y'(t)^2} dt \quad (28)
\]

The chord length (straight length) of \( C \) is

\[
ch(C) = \sqrt{(x(t_0) - x(t_1))^2 + (y(t_0) - y(t_1))^2}. \quad (29)
\]

where \( t_0 \) and \( t_1 \) are the start and end points of the vessel.
segment. The curvature at point $t$ is also defined as

$$\kappa(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{[y'(t)^2 + x'(t)^2]^{3/2}}$$  \hspace{1cm} (30)$$

The total curvature of a curve segment is $tc(C) = \int_{t_0}^{t_1} |\kappa(t)|^2 dt$.

In this work, two types of measurements are used to calculate the vessel tortuosity.

$$\tau_1 = \frac{s(C)}{ch(C)}$$ \hspace{1cm} (31)

$$\tau_2 = \frac{tc(C)}{ch(C)}$$ \hspace{1cm} (32)

The measure $\tau_1$ simply measures the tortuosity of the vessel segment by examining how long the curve is relative to its straight length. It is called distance factor tortuosity measure, described by Smedby [44]. The drawback of this measurement is that it cannot distinguish the two types of vessels which have different pathological meaning (see figure 18). An alternative way of measuring tortuosity is using $\tau_2$ defined above, which is named the length normalised total curvature measure [44]. Using this type of measurement, figures 18 (a) and (b) can be very distinguished well.

**E. Detection of Vascular Intersections and Bifurcations**

Vascular intersections and bifurcations are important landmark points for registration and segmentation processes [45] [17]. Corner and branch point detection algorithms can be broadly classified into those that first estimate image boundaries or curves using image gradient operators of one sort or another and then infer the position of the junction, to those that directly apply a curvature measure or template to the grey level data. Methods that use the intersections of boundaries to label corners necessitate the thinning of curves to a single pixel width e.g. skeletonisation. Because of ambiguities caused by the line thinning, the subsequent labelling of branch points can be problematic, particularly in 3D. In general, methods that use second or greater order differentials of the image intensity are sensitive to noise, whereas template approaches are limited in the range and type of branch points that can be described [46]. Scale-space curvature provides some trade-off between these approaches because of the noise immunity gained by repeated smoothing, plus the ability to naturally model the size of features and by tracking curvature through scale, allowing labelling decisions to be confirmed by comparing estimates at different scales in the feature space. The disadvantages are that branch points are not modelled explicitly and the implementation does not readily extend to 3D [47].

Based on the Multiresolution Hermit-Gaussian modelling and feature selection scheme, potential areas which contain branch points or crossovers can be identified by highlighting the multiple Gaussian $M > 1$ regions. A skeleton of the classification result after the application of the Bayesian linking algorithm is also used to verify the type of intersections. The detection process contain two steps:
• Highlight all the regions represented by multiple Gaussian models $M > 1$ from the AIC model selection result, (see figure 21(a)). Potential bifurcations are identified by labelling the intersection of the Gaussian functions at each region.
• High curvature and crossover points from figure 21 can be eliminated by examination of the skeleton results, (figure 21(b)), i.e. a point is classified as a bifurcation if only 3 of its neighbours are vessel points. Figure 21(c) shows the results of branching points detection after elimination of cross-over points using (b).

![Fig. 19. The binary and skeleton of three types of feature.](image)

The accuracy of the algorithm are also assessed by comparing the automatic detection against human measurement and yield 95% specificity and 92% sensitivity.

IV. Conclusions

The work described in this paper has been concerned with image segmentation within a multi-resolution framework. A new Hermite-Gaussian modelling algorithm is proposed together with an EM type of optimization scheme and statistical linking algorithm for the modelling and analysis of vascular structure from retinal fundus images. A number of interesting features of the proposed algorithm have been described and shown to be effective and robust on vessel segmentation from noisy data.

• The vessel profile is accurately modelled using a multi-resolution Hermite-Gaussian model followed by an EM type of optimization scheme.
• Bifurcations and branches are handled by the superposition of Hermite-Gaussian modelling.
• Global topology and complete segmentation of the vascular structure is achieved by using a stochastic linking algorithm.

The vessel classification results using the MHGM model were compared on two publicly available databases and were shown to have excellent selectivity and specificity.

Extensions of this work could look at using the local feature modelling for exudates and micro-aneurysms. We have already demonstrated the Hermite-Gaussian model to be well suited to blob-like features on other imaging modalities [21]. The tortuosity estimation is preliminary and needs further study with clinical validation and the bifurcation detection may be useful for retinal image mosaicing (as in [5]). There is some scope for extending the vascular model to 3D vessel extraction and using a more elaborate global prior and stochastic estimation scheme. Some of these ideas already being currently investigated e.g. [?].

APPENDIX I

EM Type Optimization for Gaussian Model

The EM algorithm is an elaborate technique for finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data are incomplete or have missing values. It is often used when optimizing the likelihood function is analytically intractable, but the likelihood function can be simplified by assuming the existence values of additional but hidden parameters [48]. In Section II-A.2 we need to minimize

$$\arg\min_x \sum (f_B(x) - G_m(x|\Theta))^2.$$  \hspace{1cm} (33)

Unlike the EM however, this estimation implicitly takes into account the spatial arrangement of the data $f_B(x)$ relative to the intensity model $G_m(x|\Theta)$ whereas EM estimates the underlying distribution from which $f_B(x)$ are drawn [21]. The algorithm first calculates the expected value,

$$Q(\Theta, \Theta^{(t)}) = \mathbb{E}[\sum_x G(x|\Theta)W(x|\Theta^{(t)})]$$  \hspace{1cm} (34)

and

$$G(x|\Theta) = A \exp(-(x - \mu)^T \Sigma^{-1}(x - \mu)/2)$$  \hspace{1cm} (35)

$$\sum_x W(x) = f_B(x)G_m(x|\Theta^{(t)});$$  \hspace{1cm} (36)

where $\Theta^{(t)}$ are the current parameter estimates that we used to evaluate the expectation and $\Theta$ are the new parameters that we should optimize to increase $Q$. However, instead of of maximising $Q(\Theta, \Theta^{(t)})$, to find some $\Theta^{(t+1)}$ such that, $Q(\Theta^{(t+1)}, \Theta^{(t)}) > Q(\Theta, \Theta^{(t)});$, a modified form is used.

The second step of the algorithm is to maximise the expectation we computed from equation 33, i.e.

$$\Theta^{(t+1)} = \arg\max\Theta Q(\Theta, \Theta^{(t+1)}).$$  \hspace{1cm} (37)

Taking the log of Equation 33, we obtain;

$$\log(Q(\Theta^{(t+1)}, \Theta^{(t)})) = \sum_x (-ln \frac{1}{\sqrt{2\pi}} - \frac{1}{2} ln(|\Sigma|))$$

$$- \frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu))W(x|\Theta^{(t)})$$  \hspace{1cm} (38)

By taking the derivative of the with respect to each parameter in set $\Theta$ and setting the results equal to zero, we derive a set of iterative equations;

$$A^{(t+1)} = \sum_x W(x|\Theta^{(t)}) \frac{\sum_x G^2(x|\Theta^{(t)})}{\sum_x W(x|\Theta^{(t)})},$$  \hspace{1cm} (39)

$$\mu^{(t+1)} = \sum_x W(x|\Theta^{(t)}) x \frac{\sum_x W(x|\Theta^{(t)})}{\sum_x W(x|\Theta^{(t)})},$$  \hspace{1cm} (40)

$$\Sigma^{(t+1)} = \sum_x W(x|\Theta^{(t)}) (x - \mu_t)(x - \mu_t)^T \frac{\sum_x W(x|\Theta^{(t)})}{\sum_x W(x|\Theta^{(t)})}. $$  \hspace{1cm} (41)
APPENDIX II

DERIVATION OF THE AKAIKE INFORMATION CRITERION

AIC is derived from Kullback-Leibler Information which is defined as

\[ K(f, g) = \int f(x) \log \left( \frac{f(x)}{g(x|\theta)} \right) dx, \quad (42) \]

where \( f, g \) are notations for full reality of truth and approximating models in terms of probability distributions respectively. \( K(f, g) \) denotes the information lost when \( g \) is used to approximate \( f \), or the distance from \( g \) to \( f \) as a heuristic interpretation [26]. The K-L information provides a distance measure between two models or probability distributions, however, both \( f \) and \( g \) must be known to calculate the distance using equation 41.

Akaike [49] proposed a model selection criterion to estimate the K-L distance based on the empirical log-likelihood function at its maximum point. Since in practice, the model parameters are estimated \( \hat{\theta} \), rather than the true parameters \( \theta \), we have to change our model selection criterion to that of minimizing the expected estimated K-L distance rather than minimising the known K-L distance over the set of models considered. Akaike [49], showed that the key issue for getting an applied K-L model selection criterion was to estimate \( E_g E_x[\log(g(x|\theta(y)))] \) where \( x \) and \( y \) are independent random samples from the sample distribution and both statistical expectations are taken with respect to truth \( (f) \). He also found that under certain conditions, it can be estimated by the maximised \( \log(L(\hat{\theta}|y)) \) minus a bias factor which is approximately equal to the number of estimable parameters \( P \) in the approximating model. A model selection criteria, AIC, based on the log-likelihood is then defined as

\[ AIC = -2\log(L(\hat{\theta}|y)) + 2P. \quad (43) \]

In the special case of least squares estimation with normally distributed errors, AIC can be expressed as a function of residual sum of squares:

\[ AIC = N^2 \log(RSS/N^2) + 2P, \quad (44) \]

where \( N^2 \) is number of data and RSS stands for the residual sum of squares \( \sum \epsilon^2 \).

One of the conditions of using equation 43 is that the sample size is relatively large with respect to the number of estimated parameters \( N^2/P > 40 \). For small data sets, Hurvich and Tsai [50], reported a second-order bias adjustment which led to a criterion called \( AIC_c \),

\[
AIC_c = -2\log(L(\hat{\theta})) + 2P \left( \frac{n}{P-K-1} \right),
\]

\[
= -2\log(L(\hat{\theta})) + 2P + \frac{2P(P+1)}{n-P-1},
\]

\[
= AIC + \frac{2P(P+1)}{n-P-1},
\]

when \( n \) is large with respect to \( P \) then the second-order correction is negligible and \( AIC_c \approx AIC \). The full derivation of AIC can be found in [26].

REFERENCES
