Retinal image analysis: Concepts, applications and potential

Niall Pattona,b,*, Tariq M. Aslamc, Thomas MacGillivrayd, Ian J. Dearye,
Baljean Dhillonb, Robert H. Eikelboomf,g, Kanagasingam Yogesan,a, Ian J. Constablea

aLions Eye Institute, 2, Verdun Street, Nedlands, WA 6009, Australia
bPrincess Alexandra Eye Pavilion, Chalmers Street, Edinburgh EH3 9HA, UK
cManchester Royal Eye Hospital, Oxford Road, Manchester M13 9WH, UK
dWellcome Trust Clinical Research Facility, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
eSchool of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7, George Square, Edinburgh EH8 9JU, UK
fLions Ear and Hearing Institute, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia
gCentre for Ophthalmology and Visual Science, University of Western Australia, Australia

Abstract

As digital imaging and computing power increasingly develop, so too does the potential to use these technologies in ophthalmology. Image processing, analysis and computer vision techniques are increasing in prominence in all fields of medical science, and are especially pertinent to modern ophthalmology, as it is heavily dependent on visually oriented signs. The retinal microvasculature is unique in that it is the only part of the human circulation that can be directly visualised non-invasively in vivo, readily photographed and subject to digital image analysis. Exciting developments in image processing relevant to ophthalmology over the past 15 years includes the progress being made towards developing automated diagnostic systems for conditions, such as diabetic retinopathy, age-related macular degeneration and retinopathy of prematurity. These diagnostic systems offer the potential to be used in large-scale screening programs, with the potential for significant resource savings, as well as being free from observer bias and fatigue. In addition, quantitative measurements of retinal vascular topography using digital image analysis from retinal photography have been used as research tools to better understand the relationship between the retinal microvasculature and cardiovascular disease. Furthermore, advances in electronic media transmission increase the relevance of using image processing in ‘teleophthalmology’ as an aid in clinical decision-making, with particular relevance to large rural-based communities.

In this review, we outline the principles upon which retinal digital image analysis is based. We discuss current techniques used to automatically detect landmark features of the fundus, such as the optic disc, fovea and blood vessels. We review the use of image analysis in the automated diagnosis of pathology (with particular reference to diabetic retinopathy). We also review its role in defining and performing quantitative measurements of vascular topography, how these entities are based on ‘optimisation’ principles and how they have helped to describe the relationship between systemic cardiovascular disease and retinal vascular changes. We also review the potential future use of fundal image analysis in teledmedicine.

© 2005 Elsevier Ltd. All rights reserved.

Contents

1. Introduction ................................................................. 100
2. Principle of digital image capture, processing and analysis .................................................. 101
   2.1. Image capture ......................................................... 101
   2.2. Image processing .................................................... 101
       2.2.1. Image enhancement ............................................. 102

*Corresponding author. Lions Eye Institute, 2, Verdun Street, Nedlands, WA 6009, Australia. Tel.: +61 8 9346 1216.
E-mail address: niallpatton@hotmail.com (N. Patton).
1. Introduction

The retina is the only location where blood vessels can be directly visualised non-invasively in vivo. Increasing technology leading to the development of digital imaging systems over the past two decades has revolutionised fundal imaging. Whilst digital imaging does not still have the resolution of conventional photography, modern digital imaging systems offer very high-resolution images that are sufficient for most clinical scenarios (Facey et al., 2002; Fransen et al., 2002; Hansen et al., 2004b; Klein et al., 2004a; van Leeuwen et al., 2003). In addition, digital imaging has the advantage of easier storage on media that do not deteriorate in quality with time, can be transmitted over short distances throughout a clinic or over large distances via electronic transfer (allowing expert “at-a-distance” opinion in large rural communities), can be processed to improve image quality, and subjected to image analysis to perform objective quantitative analysis of fundal images and the potential for automated diagnosis. In the research or screening setting, large databases of fundal images may be automatically classified and managed more readily than labour-intensive observer-driven techniques. Automated diagnosis may also aid decision-making for optometrists.

In this review, we outline the principles upon which retinal digital image analysis is based. We discuss current techniques used to automatically detect landmark features of the fundus, such as the optic disc, fovea and blood vessels. We review the use of image analysis in the automated diagnosis of pathology (with particular reference to diabetic retinopathy). We also review its role in defining and performing quantitative measurements of vascular topography, how these entities are based on ‘optimisation’ principles and how they have helped to describe the relationship between systemic cardiovascular disease and retinal vascular changes. We also review the potential future use of fundal image analysis in telemedicine. Whilst this technology can be
employed for other image acquisition modalities, such as confocal scanning laser ophthalmoscopes (cSLO) (Deckert et al., 2005), ultrasound (Schachar and Kamangar, 2005) and optical coherence tomography (Yanuzzi et al., 2004), in this article we concentrate solely on image processing based on fundal colour photography and fluorescein angiography.

2. Principle of digital image capture, processing and analysis

Digital images are made up in such a way that makes them accessible to simple and complex mathematical manipulation. For black and white images (grey scale) at any given locus of pixel, typically there is a corresponding intensity on a range from 0 (black) to 255 (white) (28 for 8-bit images) {for 12-bit images, there are 4096 grey levels (212), etc.}. Hence, the image is composed of an array of pixels of varying intensity across the image, the intensity corresponding to the level of “greyness” at any particular point in the image. If we were to express this image as an equation or function, we could say that at any point of spatial co-ordinates \((x,y)\) the image has a set shade or intensity. When \(x,y\) and the amplitude of intensity of points of an image are all described as finite and discrete quantities, the image is termed digital. A simple digital image may consist of many such points, or pixels (derived from ‘picture element’). Each pixel’s intensity for a monochrome grey-scale image is known as its grey value. Thus, a grey-scale digital image may be defined as a two-dimensional function, \(f(x,y)\), where \(x\) and \(y\) are spatial co-ordinates and \(f\) the amplitude at any pair of co-ordinates. Pixels surrounding any given pixel constitute its ‘neighbourhood’. This mathematical means of describing an image is the basis for allowing complex manipulations and calculations that are termed image processing and analysis.

Colour images use three channels (red, green and blue—RGB) to produce an overall composite (Mc Andrew, 2004). RGB images require a three-dimensional array to convey the extra colour information. The first plane in the extra, third dimension represents the red pixel intensities, the second plane represents the green pixel intensities and the third plane represents the blue pixel intensities. Often contrast is greater when the green channel alone is utilised in fundal image analysis as this enhances contrast between the background and features, such as blood vessels and haemorrhages (Hipwell et al., 2000). Many processing and measurement tools, however, are written to operate on grey-scale image and this may need to be extracted from an RGB colour image.

Indexed images use matrices, or colour maps that define a limited set of combinations of RGB values. Then, instead of each point in the digital image defining RGB levels individually, the pixel value simply refers to the closest combination from the colour map, thus saving computational memory for storage of the image.

A binary image is one containing only black and white pixels. The image consists of a binary array, typically of 0’s and 1’s. Images of any type may be converted to this format for processing or analysis.

2.1. Image capture

The first stage in fundal digital image analysis is image capture. This is normally acquired by a fundal camera (mydriatic or non-mydriatic) that has a back-mounted digital camera. The digital camera operates in the same fashion as a conventional camera, but instead of having film, digital cameras use an image sensor. Direct digital sensors are either a charge-coupled device (CCD) or complimentary metal oxide semiconductor active pixel sensor (CMOS-APS) (Gonzalez and Woods, 1992). The CCD is an array of tiny light-sensitive diodes which convert the light signal (photons) into electrical charge (electrons). This then converts the analogue light image into a digital pixelated (pixel = picture element) image. At each element (or pixel) in the array, the electrical current proportional to the analogue light level is converted into a digital level. The spatial resolution of the image depends on the number of pixels that can be created from the analogue image by the CCD array. The CMOS-APSs employ active pixel technology and are less expensive to manufacture. The APS technology reduces by a factor of 100 the system power required to process the image compared with the CCD. In addition, the APS system eliminates the need for charge transfer and may improve the reliability and lifespan of the sensor (Fossum, 1993).

2.2. Image processing

Image-processing operations transform the grey values of the pixels. There are three basic mechanisms by which this is done. In its most simple form, the pixels grey values are changed without any processing of surrounding or ‘neighbourhood’ pixel values. Neighbourhood processing incorporates the values of pixels in a small neighbourhood around each pixel in question. Finally, transforms are more complex and involve manipulation of the entire image so that the pixels values are represented in a different but equivalent form. This may allow for more efficient and powerful processing before the image is reverted to its original mode of representation.

The aims of processing of an image normally falls into one of the three broad categories: enhancement (e.g., improved contrast), restoration (deblurring of an image)
and segmentation (isolating particular areas of interest within the image) (Gonzalez and Woods, 1992).

2.2.1. Image enhancement

One of the difficulties in image capture of the ocular fundus is image quality which is affected by factors, such as medial opacities, defocus or presence of artefact (Kristinsson et al., 1997; Liesenfeld et al., 2000). Image enhancement involves the development or improvement of an image so that the result is more suitable for subsequent use. Improvements may mean the image is more acceptable for viewing, processing or analysis. This might involve processes, such as improving contrast or brightening an image.

The image histogram provides basic information about the appearance of an image. It consists of a graph indicating the number of times each grey level occurs in the image. Across the horizontal axis of this graph is the range of possible pixel intensity values, e.g., 0–255. The vertical axis represents a measure of the frequency of occurrence of each intensity value. In an excessively dark or bright image, the grey level would be clustered to the extremes of the histogram, but in a well-contrasted image these levels would be well spread out over much of the range. Histogram stretching algorithms act to distribute grey levels more equally across the range according to specific user defined equations and thus produce an image with greater contrast than the original. Histogram equalisation works on a similar principle but is an entirely automatic procedure that aims to make the histogram as uniform as possible.

2.2.2. Image restoration

Processes in this class aim to reverse damage by known causes. Algorithms such as deblurring or removal of interference patterns belong to this category. Noise occurs due to errors in pixel values caused by external disturbance. There are many forms of noise, such as salt-and-pepper noise, Gaussian noise or periodic noise.

Salt-and-pepper noise causes the appearance of randomly scattered white or black pixels over the image but it is possible to reduce this by using filters in which the mask evens out aberrations or ignores excessively high or low values. Gaussian noise is caused by random fluctuations in the signal. It can be reduced by using several versions of that same image and averaging values for each pixel. Periodic noise occurs if the imaging equipment is subject to electronic repeating disturbance. This can be reduced by transforming the image to a different structure known as a Fourier transform, then applying noise filters before transforming back to the original image. Deblur functions rely on modelling of the blurring process then using filters to remove the known effects of blur.

2.2.3. Image segmentation

Segmentation involves dividing images into subsections that are of particular interest, such as defining areas of an image that are appropriate to be subsequently analysed, or finding circles, lines or other shapes of interest. Segmentation can stop when such objects of interest have been isolated. Segmentation algorithms for monochrome images are generally based on discontinuity of image intensities such as edges in an image, or on similarities judged by predefined criteria (see below).

2.2.3.1. Thresholding. Thresholding allows the separation of an image into separate components by turning it into a binary image. This involves the image being separated into white or black pixels on the basis of whether their intensity value is greater or less than a certain threshold level. The process of thresholding may be particularly useful to remove unnecessary detail or variations and highlight detail that is of interest. A global threshold value may be chosen automatically or on the basis of clear points in the image histogram that would allow for efficient separation. More complex intensity criteria may be used to allocate whether pixel values become white or black. For some images, adaptive or local thresholding is useful where different thresholds are applied to different sections of the image, e.g., the image has varying levels of background illumination.

2.2.3.2. Edge detection. Edges contain some of the most useful information in an image. They can be used, e.g., to measure the size of objects or to recognise and isolate objects. An edge in a digital image consists of an observable difference in pixel values within a certain area. Most edge detection algorithms assess this change by finding the magnitude of the gradient of the pixel intensity values. This can be done by the application of specialised filters of varying complexity and utility. A threshold can be applied to the resultant image to create a binary image of the edges. Examples of edge detection masks include Sobel (Gonzalez and Woods, 2002) and Canny (1986) edge detection programs. The Sobel edge detector uses a pair of 3 × 3 convolution masks, one estimating the gradient in the x-direction (columns) and the other estimating the gradient in the y-direction (rows). However, in a comparison of three automated techniques of edge detection to identify the boundaries and corresponding widths of retinal blood vessels, Sobel was found to be the most inconsistent, possibly related to the program identifying the central light reflex from the blood vessel as an edge (Chapman et al., 2001). The Canny edge detection program has been used in neural networks to automatically localise retinal blood vessels in fundal RGB images (Sinthanayothin et al., 1999).
2.2.3.3. Filters. Neighbourhood processing extends the power of processing algorithms by incorporating values of adjacent pixels in calculations. A user defined matrix, or mask is defined with enough elements to cover not only a single pixel but also some of its adjacent pixels. Each pixel covered by the elements of the mask is subject to a corresponding function. The combination of mask and function is called a filter. Thus, the result of applying a mask to a particular location is that the final resultant value is a function not only of the central pixel’s values but also of its neighbouring pixel values.

2.2.3.4. Morphological processing. Mathematical morphology in image processing is particularly suitable for analysing shapes in images. The two main processes are those of dilation and erosion. These processes involve a special mechanism of combining two sets of pixels. Usually, one set consists of the image being processed and the other a smaller set of pixels known as a structuring element or kernel. In dilation, every point in the image is superimposed onto by the kernel, with its surrounding pixels. The resultant effect of dilation is of increasing the size of the original object. Erosion is an inverse procedure in which an image is thinned through subtraction via a structuring element or kernel. The kernel is superimposed onto the original image and only at locations where it fits entirely within its boundaries will a resultant central pixel be accepted. The algorithms of opening and closing are based upon these processes. Opening consists of erosion followed by dilation, and tends to smooth an image, breaking narrow joints and removing thin protrusions. Closing consists of dilation followed by erosion and also smoothes images, but by fusing narrow breaks and gulfs and eliminating small holes. Algorithms combining the above processes are used to creating mechanisms of edge detection, noise removal and background removal as well as for finding specific shapes in images (see below).

2.3. Image registration

Image registration is a process of aligning two or more images of the same scene. One image (the base image) is compared to the other input images (Maintz and Viergever, 1998). The aim of registration is to apply spatial transformations to the input image to bring the input image into alignment with the base image. Commonly, the input images may be misaligned due to different camera angles or different imaging modalities. The details of the exact alignment algorithm are calculated after the user identifies pairs of points that should exactly correspond in the two images. A spatial mapping is inferred from the positions of these control points. Registration can be based on identified landmarks (e.g., retinal vessel bifurcation points), on the alignment of segmented binary structures (segmentation based), or directly onto measures computed from the image grey values (Maintz and Viergever, 1998). The input image is transformed then to correspond with the base image so that direct comparisons can be made. Registration of images in this manner is commonly used as a preliminary step in processing applications. Alternatively, two images taken at different time intervals may need to be registered (temporal registration) using an automated process in order for time-dependent changes to be identified. This has been performed for sequential fluorescein angiography (Spencer et al., 1996) and colour fundal images. The available registration methods are divided into intensity based (Matsopoulos et al., 1999; Ritter et al., 1999) and feature based (Zana and Klein, 1999). Intensity-based methods have the drawback of poor performance under varying illumination, whilst feature-based methods rely heavily on accurate and repeatable extraction of the features (Tsai et al., 2004). Image mosaicing is the act of combining two or more images and is used to combine images such that no obstructive boundaries exist around overlapped regions and to create a mosaic image that exhibits as little distortion as possible from the original images, and may be used to create wide-field retinal montages. It also has applications in video frame capturing (Can et al., 2000).

3. Automated localisation (segmentation) of retinal landmarks

A potential use of fundal digital image analysis is the ability to analyse a large database of fundal images in a short period of time. The identification of fundal landmark features such as the optic disc, fovea and the retinal vessels as reference co-ordinates is a prerequisite before systems can achieve more complex tasks identifying pathological entities. Reliable techniques exist for identification of these structures in retinal photographs.

3.1. Optic nerve head localisation

The location of the optic disc is important in retinal image analysis, to locate anatomical components in retinal images, for vessel tracking, as a reference length for measuring distances in retinal images, and for registering changes within the optic disc region due to disease. The optic disc is usually the brightest component on the fundus, and therefore a cluster of high intensity pixels with a high grey-scale value will identify the optic disc location (Chaudhuri et al., 1989a; Lee et al., 1999). This works well, unless there are other potential fundal features such as the presence of exudates, which may mimic similar high grey-scale values (Goldbaum et al., 1990). Techniques such as principal component analysis (PCA) (a way of
identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences—often used in face recognition and other computer vision applications) can help differentiate the true optic disc from other sources (Li and Chutatape, 2004). Li and Chutatape (2004) produced a training set using the brightest pixels that were firstly clustered as candidate optic disc regions. Principle component analysis was then applied to project a new image to the ‘disc space’. Then, the location of the optic disc centre was found by calculating the minimum distance between the original retinal image and its projection. Other features that help differentiate the optic nerve are the confluence of blood vessels at the optic disc, which results in a large variance in intensity of adjacent pixels (Akita and Kuga, 1982; Hoover and Goldbaum, 1998). Sinthanayothin et al. (1999) correctly identified the location of the optic disc employing the variance of intensity between the optic disc and adjacent blood vessels in 111 of 112 colour fundal images, giving both a sensitivity and specificity of 99.1%. However, others have found that this algorithm often fails for fundus images with a large number of white lesions, light artefacts or strongly visible choroidal vessels (Lowell et al., 2004a). Others have exploited the Hough transform (a general technique for identifying the locations and orientations of certain types of shapes within a digital image; Kalviainen et al., 1995) to locate the optic disc (Kochner et al., 1998; Tamura et al., 1988; Yulong and Dingru, 1990). However, Hough spaces tend to be sensitive to the chosen image resolution (Hoover and Goldbaum, 2003). Foracchia et al. (2004) recently report on a new technique for locating the optic disc using a geometrical parametric model (retinal vessels originating from the optic disc and their path follows a similar directional pattern (parabolic course) in all images) to describe the typical direction of retinal vessels as they converge on the optic disc. Hoover and Goldbaum (2003) correctly identify optic disc location in 89% of 81 images, 50 of which were diseased retinas using a “fuzzy convergence” algorithm (finds the strongest vessel network convergence as the primary feature for detection using blood vessel binary segmentation, the disc being located at the point of vessel convergence. Brightness of the optic disc was used as a secondary feature). All of the healthy retinas (n = 31) had successful optic disc localisation, and 41 from the 50 diseased retinas.

Optic disc boundary identification has been used by Mendels et al. (1999), and Osareh (2004) and Osareh et al. (2002) report an accuracy of 90% (n = 75) in locating the optic disc boundary, using active contours (model-based methods for localisation and tracking of image structures) compared to the reference standard of a clinical ophthalmologist. Lowell et al. (2004a) also report identifying the optic disc in 89 of 90 randomly chosen low-resolution diabetic fundal images using a contour-model-based approach (using a parametric approach, encoding a specific shape to fit the simple global model of the optic disc, but allowing for significant variability related to “distractors”, such as variation in blood vessel patterns).

3.2. Foveal localisation

The fovea can be detected exploiting the avascularity of the fovea, thus having different grey levels at its border (Ibanez and Simo, 1999). Sinthanayothin et al. (1999) report a sensitivity and specificity for correct identification of the fovea as 80.4% and 99.1%, respectively, in 112 images, when compared with an experienced ophthalmologist. The location of the fovea was chosen as the position of maximum correlation between a model template and the intensity image, obtained from the intensity-hue-saturation transformation, provided it was appropriately placed temporal to the optic disc and in the region of minimum intensity. Foveal localisation was particularly affected if there was poor centration of the fovea in the image. Goldbaum et al. (1996) fixed the position of the fovea relative to the optic disc. Li and Chutatape (2004) report a 100% detection of the foveal region, using model-based methods. They estimate the position of the fovea by extracting the points on the main blood vessels by a modified active model, and fitting a parabola curve with the optic disc as the centre. The fovea is then located at 2 disc diameters (DDs) from the optic disc on the main axis of the parabola.

3.3. Vascular segmentation

Retinal vascular segmentation techniques utilise the contrast existing between the retinal blood vessel and surrounding background, the cross-sectional grey-level profile of a typical vessel conforms to a Gaussian shape, the vasculature is piecewise linear and may be represented by a series of connected line segments and that the vasculature originates from the same point (the optic disc) and all vessels are connected (Heneghan et al., 2002).

Four main techniques are used to segment the vasculature form retinal images:

3.3.1. Matched filters

This usually employs a two-dimensional linear “structural element” (kernel) that has a Gaussian cross-profile section, extruded or rotated into three dimensions to identify the cross-profile of the blood vessel, which typically has a Gaussian or a Gaussian derivative profile (Chaudhuri et al., 1989b; Hoover et al., 2000; Lowell et al., 2004b) (Fig. 1). The kernel is rotated into many different orientations (usually eight
or 12) to fit into vessels of different configuration. The image is then thresholded (an arbitrary chosen grey level divides all features into a binary classification, depending on whether they have a greater or lesser intensity level than the ‘brightness threshold’) to extract the vessel silhouette from the background. This works reasonably well on images of healthy retinae, but in diseased states such as diabetic retinopathy, there are problems associated with detecting very fine neovascularisation, partly due to image resolution and also smaller vessels are more prone to changes in background intensity and there is a reduced contrast-to-noise ratio. To overcome this, non-linear “tram-line” filters have been used, utilising the contrast between a central line oriented along the vessel and satellite tram-lines at either side (Hunter et al., 2002). Also, using too long a structuring element may have difficulty in fitting into highly tortuous vessels. Matched filters do not operate in isolation, but as part of an algorithmic chain, requiring thresholding into a binary vessel/non-vessel image (Teng et al., 2002).

3.3.2. Vessel tracking

Another technique for vessel segmentation include “vessel-tracking” (Kochner et al., 1998; Tamura et al., 1988), whereby vessel centre locations are automatically sought over each cross-section of a vessel along the vessels longitudinal axis, having been given a starting and end point (Fig. 2). They tend to work on single retinal vessels and require starting and ending points to be identified by the user. The selection of vascular points is normally accomplished by matched filters (Teng et al., 2002). In addition, vessel-tracking techniques may be confused by vessel crossings and bifurcations (Frame et al., 1996; Tamura et al., 1988). However, vessel tracking can provide very accurate measurements of vessel widths and tortuosity.

3.3.3. Neural networks

Others have proposed the use of neural networks to segment retinal vasculature (Akita and Kuga, 1982). Artificial neural networks employ mathematical “weights” to decide the probability of input data belonging to a particular output. This “weighting” system can be adjusted by training the network with data of known output, often with a “feedback” mechanism allowing retraining. Unlike conventional computer programs that employ serial processing, neural networks use parallel processing. Neural networks have been used in association with edge detection programs (Sinthanayothin et al., 1999) having initially pre-processed the image with PCA to reduce background noise. Sinthanayothin et al. (1999) report a success rate (as compared with an experienced ophthalmologist manually mapping out the location of the blood vessels in a random sample of 73 20 × 20 pixel window and requiring an exact match between pixels in both images) of 99.56% for training data and 96.88% for validation data, respectively, with an overall sensitivity and specificity of 83.3% (standard deviation 16.8%) and 91% (standard deviation 5.2%), respectively.

3.3.4. Morphological processing

Morphological image processing exploits features of the vasculature shape that are known a priori, such as it being piecewise linear and connected. Algorithms that extract linear shapes can be very useful for vessel segmentation. Structuring elements of a certain intensity can be added (dilation) or subtracted (erosion) to the underlying image. Opening (erosion followed by dilation) with a structuring element of a certain shape can separate objects in an image, by preserving image structures that can contain the structural element and

---

Fig. 1. Three-dimensional image of an intensity profile from a retinal vessel of “double-Gaussian” construct.

Fig. 2. Grey-scale image of a retinal vessel-tracking process.
removing those that cannot (Heneghan et al., 2002). Closing (dilatation followed by erosion) can be used to ‘fill-in’ small holes within an image. Gregson et al. (1995) utilise morphological closing to help identify veins in the automated grading of venous beading by filling in any “holes” in the silhouette of the vein created during the processing procedure.

4. Automated detection of pathology using retinal digital image analysis

Automated diagnosis of retinal fundal images using digital image analysis offers huge potential benefits. In a research setting, it offers the potential to examine a large number of images with time and cost savings and offer more objective measurements than current observer-driven techniques. Advantages in a clinical context include the potential to perform large numbers of automated screening for conditions such as diabetic retinopathy, and hence to reduce the workload required from manual trained graders. Image management systems such as STARE offer an approach that is designed to measure key aspects of fundal images, diagnose images, compare images and search for images similar in content, using statistical neural network learning modules (Goldbaum et al., 1989; Hoover et al., 2000). The greatest emphasis in automated diagnosis has unsurprisingly been given to the detection of diabetic retinopathy.

4.1. Automated detection of diabetic retinopathy (ADDR) using retinal digital image analysis

Diabetic retinopathy is the leading cause of blindness in people of working age in the developed world, affecting more than 2% of the UK population (Evans et al., 1996). In England and Wales, approximately 1000 diabetic patients are registered as blind or partially sighted each year (Watkins, 2003) and blindness due to diabetes costs the US Government and general public $500 million annually (Klein and Klein, 1995). A WHO collaborative study projected that the global diabetic burden is expected to increase to 221 million people by 2010 (Amos et al., 1997). However, treatment can prevent visual loss from sight-threatening retinopathy if detected early (Diabetic Retinopathy Study, 1978). In order to address the impact of diabetes, screening schemes are currently being put into place around the world, many based on digital fundal photography. In England and Wales, a national screening program, based on digital photography, has been recommended by the National Screening Committee. However, there are concerns regarding the cost of any screening scheme used for detecting sight-threatening diabetic retinopathy in the population. Whilst screening schemes could reduce the risk of blindness to less than half compared with unscreened controls and reduce the economic burden of blindness, significant resources are required to put these screening schemes into place (Mason, 2003). One of the greatest sources of expenditure in setting up any diabetic retinopathy screening program is the cost of financing trained manual graders. As a means to reduce this cost, the potential ability of ADDR to reduce this workload by using computerised algorithms on digitalised retinal images to define an image as showing presence or absence of diabetic retinopathy has been extensively investigated. If automated detection programs are able to exclude a large number of those patients who have no diabetic retinopathy, it will reduce the workload of the trained graders and thus reduce costs. Other benefits of an automated detection program include improved repeatability and immunity from fatigue. Great interest over the past 10 years has centred on developing algorithms that detect diabetic retinopathy with sufficient sensitivity to be able to implement them into screening programs as an adjunct to current strategies. Current technology offers no potential for a completely independent automated diabetic detection program, as systems that have been tested show unacceptable specificity. Some of the reasons include a high number of false-positives confusing drusen with exudates (Usher et al., 2003) and potential problems with different pigmentation form different ethnic races (Usher et al., 2003). However, sensitivity is more important than specificity, to err on the side of caution and not miss a case of sight-threatening retinopathy. Systems have concentrated on highly sensitive systems to be used in conjunction with manual graders. ADDR systems have concentrated largely on the automated detection of microaneurysms, haemorrhages and exudates. Detection of clinically significant macular oedema has proved more problematic (Olson et al., 2003) and no system to our knowledge has explored the automatic identification of diabetic retinal neovascularisation form colour images.

4.1.1. Detection of microaneurysms/haemorrhages

Some of the first automated techniques utilised fluorescein angiography and relied on global (involving the whole image) image-processing procedures to segment microaneurysms from the background fundus image by thresholding at a grey level which was a compromise between that which was low enough to detect all the microaneurysms and that which did not detect any other features (Lay et al., 1983; Spencer et al., 1992). Binary morphological processing and structuring elements in the thresholded images allows further discrimination between microaneurysms and other features, such as small vessels sections. Fluorescein angiography has also been utilised to provide measurements of overall fluorescein intensity variation over the
fluorescein transit (Hipwell et al., 1998), and the use of parameters such as time to maximum intensity of fluorescein can produce quantifiable values that relate to the state of health of the retinal circulation, and hence offer a potential diagnostic index of retinopathy severity. However, the drawbacks related to fluorescein angiography as an intervention prohibit its use in large-scale screening analysis. Image processing of colour fundal photographs is more challenging, due to the different ‘distractors’ within the image that may be confused with diabetic lesions, e.g., small vessels, choroidal vessels and reflection artefacts.

Spencer et al. (1996) employed a morphological transformation to segment microaneurysms from fluorescein angiograms. After ‘opening’ (erosion followed by a dilatation) the shade-corrected image with an 11-pixel linear structural element in eight rotational orientations that when combined included all of the vessel sections, and excluded all the ‘circular’ microaneurysms, this opened image was extracted from the original shade-corrected image (a ‘top-hat transformation’) producing an image that only contains microaneurysms. A two-dimensional Gaussian matched filter model of a microaneurysm was then applied, before thresholding resulted in a binary image containing candidate microaneurysms. A region-growing algorithm (based on choosing a single-pixel ‘seed’ with the highest grey level as the origin of each ‘candidate’ microaneurysm and assessing neighbouring pixels’ grey levels against this to determine inclusion or exclusion of the pixel into the growing object) delineated each marked object and subsequent inclusion or exclusion of the pixel into the growing object. The classification algorithm was based on 13 different calculations (based on a training set of 102 images of variable degrees of retinopathy) including shape, intensity, circularity, perimeter length and length ratio, and if all criteria were met, the area under scrutiny is nominated as a microaneurysm. Based on four 10” x 8” printed images containing microaneurysms, they report good agreement with five clinicians’ opinions, with 82% sensitivity and 86% specificity. However, this was at the cost of approximately 100 false positives per image.

Cree et al. (1997) developed Spencer et al.’s technique but redesigned the region-growing and classification algorithms and included a process to remove the need for operator intervention in selecting regions-of-interest (ROI) (found by locating the fovea and centring a 512 x 512 ROI at this point) within which microaneurysms are counted, and to include an automated process for image registration to allow sequential comparisons of microaneurysm turnover, based on a cross-correlation algorithm (Cideciyan et al., 1992). However, the authors state that the automated registration process for sequential studies often failed for poor-quality images, and those with prior laser photocoagulation. Automated selection of the macular ROI was reported as being accurate in 93 of 95 images. The classifier was trained on a set of 68 fluorescein images to detect features that may be helpful to discriminate microaneurysms and other microvascular abnormalities, and was further tested in 20 angiogram images (all individual patients with varying degrees of retinopathy, containing 297 true microaneurysms as referenced by the joint agreement of an ophthalmologist and a medial physicist) and achieved a sensitivity of 82%, with 2.0 false positives per image. Against five clinicians looking at 20 digital angiograms on a computer screen, sensitivity was 82% with 5.7 false positives per image, a great improvement over Spencer et al. (1996).

Further development by Hipwell et al. (2000) led to a microaneurysm detection program on digitalised 50° digital red-free images. The images were initially processed by the same technique described by Spencer et al. (1996), viz. shade correcting the image, followed by removal of vessels and other distractors by the top-hat transformation, and use of the Gaussian filter to retain candidate microaneurysms for subsequent classification. The classification algorithm was based on 13 different calculations (based on a training set of 102 images of variable degrees of retinopathy) including shape, intensity, circularity, perimeter length and length ratio, and if all criteria were met, the area under scrutiny is nominated as a microaneurysm. Based on a total of 3783 images from 589 patients on 977 visits (graded for presence/absence of microaneurysms and/or haemorrhages against the reference standard of an experienced clinical research fellow according to the EURODIAB HMA protocol; Aldington et al., 1995) produced a sensitivity of 81%, with 93% specificity. However, this was when EURODIAB HMA grade 2 images (questionable HMA present) were excluded. When based on individual patients (two images per eye), sensitivity of the computer program increased to 85%, while specificity reduced to 76%. Whilst this protocol for diabetic screening requires two 50° images per eye, and hence increases the sensitivity of the screening process, it decreases the specificity as well as having cost implications for widespread screening programs. However, the authors point out that 15% of the patients only had retinopathy identified on the nasal image, and hence would have been missed if a single macular field only had been used.

Goatman et al. (2003) have developed an ADDR program to detect microaneurysms on fluorescein angiography, with the ability to follow-up temporal changes in microaneurysm turnover. Using the same microaneurysm detection algorithm on fluorescein images and image registration as Cree et al. (1997), the turnover of microaneurysms (static, new or regressed) was compared with a reference standard of an ophthalmologist experienced in identifying microaneurysms and grading of retinopathy in angiographic images. Compared with manual measurements of nine manual graders, the automated system was fast and reliable with similar sensitivity and specificity to manual graders. It also worked with red-free images.
Ege et al. (2000) detected microaneurysms and haemorrhages with a sensitivity of 69% and 83%, respectively (based on 30 images, 14 of which had retinopathy and 16 were no-retinopathy), using a Mahalanobis classifier (a statistical measure of similarity of an object based on ‘nearest-neighbour’ classification, often used in pattern recognition), after employing a training set of 30 images (15 with and 15 without retinopathy). Reference standard was an experienced clinical ophthalmologist.

Lee et al. (2001) employed colour fundus photographs and image processing (image enhancement, noise removal and image normalisation) in conjunction with pattern recognition to test for particular features of early diabetic retinopathy. In comparison with two expert human graders, they report sensitivity of 77% and specificity of 94% (when compared with a general ophthalmologist) and 87% and 98%, respectively (when compared with a grader from the University of Wisconsin Fundus Photograph Reading Center), for haemorrhages/microaneurysms from 428 images. However, no details regarding the nature of the processing or pattern recognition were provided.

Sinhthanayothin et al. (2002) employed a recursive region-growing technique (works by localising similar pixels within a certain boundary by comparison of their respective intensities) and adaptive intensity thresholding in conjunction with a ‘moat operator’ (creates a trough around lesions, which aids segmentation of the image to enhance edge detection). After pre-processing and segmentation of major fundal landmarks, detection of diabetic features was performed. To detect microaneurysms and haemorrhages, they utilised a Moat Operator (creates a trough around lesions, which aids segmentation of the image) to enhance edge detection and then used the recursive region-growing segmentation algorithm, using neural networks to extract the similar retinal blood vessels. For a total of 30 images (14 of which contained haemorrhages/microaneurysms), they report a sensitivity and specificity for haemorrhage/microaneurysm detection as 77.5% and 88.7%, respectively (clinical ophthalmologist as the reference standard).

Gardner et al. (1996) employed a back propagation neural network to detect diabetic retinopathy features. Training a neural network refers to adjusting the weights in response to incorrect results. The degree to which a connection has contributed to a particular error determines the degree to which the weight associated with the connection will be adjusted. This weight update method, known as the delta rule, is the basis of back propagation. Initial training was performed on 147 diabetic and 32 normal images, analysing the green channel from 60° colour images, and dividing images into 20 × 20 pixel or 30 × 30 pixel windows, which were each individually graded manually by a trained observer (normal without vessel, normal vessel, exudate and haemorrhage/microaneurysm). This information was used to teach the neural network, prior to employing a testing set. They report detection rates for haemorrhages as 73.8% for both sensitivity and specificity, compared with the reference standard of a clinical ophthalmologist, based upon 200 diabetic and 101 normal images. When classifying images into normal, diabetic requiring referral and diabetic not requiring referral, based on the reference standard of the clinical ophthalmologist, they report a sensitivity of 88.4% and a specificity of 83.5%. Increasing the sensitivity to 99% results in a fall of specificity to 69%. A drawback of the neural network approach is the length of time that it takes to ‘teach’ the network form the training data. In this study, it required 79 h to train the haemorrhage/microaneurysm protocol, although this is based on a 60 MHz computer processor, which would now be obsolete.

Usher et al. (2003) also employed an artificial neural network, based on a single 45° macular centred colour image. After pre-processing and segmentation of normal structures, haemorrhages/microaneurysms were extracted using recursive region growing and adaptive intensity thresholding in conjunction with a ‘moat operator’ similarly to Sinthanayothin et al. (2002). Training was performed on 500 patients before analysis of performance in comparison with a trained clinical diabetologist (audited by a consultant ophthalmologist) in 773 patients. On a per patient basis, sensitivity for detection of any exudates and/or haemorrhages/microaneurysms was 95.1% (95% confidence interval (CI) 92.3 to 97.7%) and a specificity of 46.3% (95% CI 41.6 to 51%). For detection of diabetic retinopathy, maximum sensitivity and specificity was 70.8% and 78.9%, respectively. However, 70 of 773 images were ungradable and a significant number of false positives were related to drusen. If the system was implemented at 94.8% sensitivity and 52.8% specificity, the authors conclude that the workload for manual trained graders in screening for diabetic retinopathy in a population could be reduced by a third and the number of patients examined without diabetic retinopathy reduced by a half by utilising this automated method in conjunction with established diabetic screening procedures, with considerable cost benefits.

4.1.2. Detection of retinal exudates and cotton wool spots

Important objects within fundal images that act as distractors in the identification of retinal exudates include the optic disc, drusen, light artefacts, choroidal vessels and laser photoagulation scars.

Sinhthanayothin et al. (2002) identified exudates in colour images based on the same recursive region-growing technique described above to define an “exudate” and “non-exudate” image. After thresholding to produce a binary image, the regions containing the
exudates were overlaid onto the original image. They report a sensitivity and specificity (with reference to a clinical ophthalmologist grading images manually) of 88.5% and 99.7%, respectively, for 30 images (21 of which contained exudates).

Employing the neural network as before, Gardner et al. (1996) report a sensitivity of 93.1%. Osareh et al. (2003) also employ neural networks to locate exudates. After pre-processing of images, the image was segmented depending on colour using ‘Fuzzy C-means clustering’ (Bezdek et al., 1999). For object classification, 18 different features that discriminate exudates were then inputted to a three layer neural network. They report detection \( n = 67 \) with 93.0% sensitivity and 94.1% specificity in terms of lesion based classification (employing a back propagation neural network), and 95.0% sensitivity and 88.9% specificity for the identification of patients with evidence of retinopathy. The whole processing time using a 700 MHz computer processor was only 11 min per image.

Ege et al. (2000) detected exudates and cotton wool spots with sensitivity of 99% and 80%, respectively. In comparison with a general clinical ophthalmologist, Lee et al. (2001) report sensitivities of 96% and 80% and specificities of 93% and 93% for hard exudates and cotton wool spots, respectively (when compared with a general ophthalmologist).

It must be borne in mind that the above studies may not be directly applicable to the clinical situation as the reference standard was a clinician examining an image, rather than using indirect slit-lamp biomicroscopy. However, this may more appropriately reflect the screening scenario. In contrast, Olson et al. (2003) compared automated digital analysis with direct slit-lamp biomicroscopy by ophthalmologists as the reference standard. In 586 patients, they report a sensitivity of 83% (95% CIs 77–89%) and a specificity of 85% (82–88%) for diagnosis based on detection of haemorrhages/microaneurysms. Technical failure rates were lower with digital imaging, compared with conventional photography and automated digital imaging had the same sensitivity as manual digital grading, and was even superior to screening by optometrist examination. This is very encouraging for the use of automated digital imaging as a screening tool for diabetic retinopathy, as this is the first study to find it of superior sensitivity to optometric examination. As expected however, specificity was lower for automated digital imaging than other modalities (71% vs. 89%, 79% and 82% for manual examination using colour slides, manual examination using digital images and optometric examination, respectively).

4.1.3. Detection of clinically significant macular oedema

Previous studies quantifying macular oedema have employed fluorescein angiography, comparing early transit (15–30 s) with late transit (250–300 s) images, and using a threshold gradient below which were considered representative of leakage (Philips et al., 1991). Whilst high-resolution digital non-simultaneous stereoscopic fundal photographs have been successfully employed to detect clinically significant macular oedema, this was based on manual grader examination using liquid crystal shutter goggles (Rudnisky et al., 2002). Olson et al. (2003) are the first to report the use of ADDR for clinically significant macular oedema from fundal photographs using an automated system. Based on findings of haemorrhages/microaneurysms or exudates within 1 DD of the fovea, the automated analysis detected 16 of 21 cases (76% sensitivity, 85% specificity) of macular oedema (with reference to slit-lamp biomicroscopy by an ophthalmologist). Whilst this is encouraging (the automated system was more markedly more sensitive than optometric examination), automated systems are currently limited in their detection of macular oedema, due to the need for stereoscopic macular examination to make this diagnosis.

As well as identification of sight-threatening diabetic retinopathy, computerised digital fundal analysis has the potential to quantitatively analyse diabetic fundi in terms of features, such as haemorrhages and exudates. Evidence suggests quantitative analysis may help identify patients with early diabetic retinopathy that may later develop vision-threatening maculopathy (Hove et al., 2004).

4.1.4. Other issues concerning ADDR

One of the issues arising from the use of digital images for diabetic retinopathy screening is the time and space involved in capture and storage of the files. Currently, the use of image compression using utilities such as Joint Photographic Experts Group (JPEG) have not been recommended, although there is some evidence that while large file compression significantly reduces the ability of automated detection programs, a compression ratio of 1:12 or 1:20 would produce little reduction in sensitivity (Basu et al., 2003). Another consideration for diabetic screening is the use of routine mydriasis. Hansen et al. (2004a) address the impact of pharmacologically dilated pupils on ADDR. They report a change in sensitivity before and after pupil dilatation of 90% and 97%, respectively, for detection of ‘red lesions’ (haemorrhages/microaneurysms) and specificity before and after pupil dilatation was reported as 86% and 75%, respectively \( n = 165 \) eyes of 83 patients. The use of routine mydriasis for diabetic screening is controversial. Currently, the National Screening Committee in England and Wales have recommended routine mydriasis for all screened patients, whereas the Health Technology Assessment Board for Scotland (Facey et al., 2002) only recommend mydriasis under certain defined circumstances.
No current ADDR systems can identify neovascularisation. Whilst in the context of screening to identify presence/absence of retinopathy in conjunction with manual assessment, this may not be important, but if ADDR systems are ever going to have the potential to identify and classify diabetic retinopathy, this will need to be addressed. Fractal geometrical analysis may hold more promise in this context than conventional vessel identification (Daxer, 1993a, b).

Whilst the detection of sight-threatening diabetic retinopathy has received the most attention with respect to automated digital image analysis, other pathologies offer potential to use this tool as well, including morphological evaluations of the optic nerve in glaucoma (Corona et al., 2002; Wolfs et al., 1999) and the macular region in age-related macular degeneration (Barrhes et al., 2001; Shin et al., 1999; Smith et al., 2003, 2005a, b; Soliz et al., 2000) and retinopathy of prematurity (ROP) (Heneghan et al., 2002; Swanson et al., 2003). Table 1 summarises sensitivities and specificities of selected studies of ADDR.

5. Quantitative measurements from Fundal images

An important role of retinal digital image analysis is the ability to perform quantitative objective measurements from retinal colour photographs. However, the effect of image magnification resultant from fundal photography has to be overcome, either incorporating an adjusted measurement to take the magnification into account, or to use dimensionless measurements so that results between patients can be compared.

5.1. Magnification effect of fundal photography

Magnification is defined as the image height divided by the actual object height. For images that are close to the ocular optical axis, the “actual” retinal size (t) is related to the image size (s) by the formula

\[ t = \frac{psq}{s} \]

where (p) is a camera factor and (q) an ocular factor.

Therefore, both (a) camera factors and (b) ocular factors will have a bearing on the degree of magnification obtained from fundal photography. Other factors that may need to be taken into consideration include the degree of eccentricity of the measured object from the optical axis (Bennett et al., 1994; Holden and Fitzke, 1988) and camera–eye distance (Arnold et al., 1993; Behrendt and Doyle, 1965; Bengtsson and Krakau, 1977, 1992; Lotmar, 1984; Pach et al., 1989). The following discussion of camera and ocular magnification factors are based on the Gullstrand schematic eye.

<table>
<thead>
<tr>
<th>Study</th>
<th>Any retinopathy Sensitivity%</th>
<th>Any retinopathy Specificity%</th>
<th>HA Sensitivity%</th>
<th>HA Specificity%</th>
<th>MA Sensitivity%</th>
<th>MA Specificity%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olson et al. (2003) (reference slit-lamp biomicroscopy by ophthalmologist)</td>
<td>586a 83</td>
<td>71</td>
<td>75</td>
<td>93</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Hipwell et al. (2000) (reference clinical research fellow)</td>
<td>3783a 85 (per-patient basis)</td>
<td>76 (per-patient basis)</td>
<td>83.5</td>
<td>93 (MA only)</td>
<td>73.8 (HMA only)</td>
<td>93</td>
</tr>
<tr>
<td>Gardner et al. (1996) (referenced to clinical ophthalmologist assessing images)</td>
<td>301a 88.4</td>
<td>83.5</td>
<td>83</td>
<td>93.1</td>
<td>93.1</td>
<td></td>
</tr>
<tr>
<td>Lee et al. (2001) (reference general ophthalmologist)</td>
<td>428b —</td>
<td>—</td>
<td>77</td>
<td>94</td>
<td>96</td>
<td>93</td>
</tr>
<tr>
<td>Sinthanayothin et al. (2002) (reference ophthalmologist looking at slides)</td>
<td>30</td>
<td>—</td>
<td>77.5</td>
<td>88.7</td>
<td>88.5</td>
<td>99.7</td>
</tr>
<tr>
<td>Cree et al. (1997) (reference ophthalmologist)</td>
<td>20c</td>
<td>—</td>
<td>82</td>
<td>5.7 false positives per image</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Spencer et al. (1996) (reference clinician examining image negatives)</td>
<td>4c</td>
<td>—</td>
<td>82 (MA only)</td>
<td>86 (MA only)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Osareh et al. (2003)</td>
<td>67</td>
<td>95</td>
<td>88.9</td>
<td>94.8</td>
<td>94.8</td>
<td></td>
</tr>
<tr>
<td>Hansen et al. (2004a) (reference manual graders)</td>
<td>165</td>
<td>97</td>
<td>95</td>
<td>73.8</td>
<td>73.8</td>
<td></td>
</tr>
</tbody>
</table>

MA, microaneurysm; HMA, haemorrhage; n refers to number of images. 
*Red-free images. 
†Sensitivity and specificity for detection of cotton wool spots 80% and 90%, respectively.
‡Values shown are with mydriatic photography, using red lesion detection and image quality control. For non-mydriatic photography, sensitivity and specificity were 89.9% and 85.7%, respectively.
5.1.1. Camera factors

The magnification effect of the camera relates the angle emergent from the first principal point of Gullstrand’s schematic eye to the image size ($s$) of the retinal feature, expressed as a quotient (Garway-Heath et al., 1998). For any particular fundal camera, this ratio will be a constant, and therefore if attempting to make between-patient comparisons of exact measurements from fundal photographs correcting for magnification, the camera constant of each camera used needs to be known.

5.1.2. Ocular factors

Ocular magnification is solely related to the vergence of the internal axis of the eye (Bengtsson and Krakau, 1977) (Fig. 3). Thus, ocular magnification ($q$) is directly proportional to the distance between the second principal point and the fovea. Several strategies exist to calculate $q$ from ocular biometric factors. The most accurate technique is to use ray tracing to calculate $q$, knowing the axial length of the eye, the anterior and posterior radii of curvature of both the cornea and the lens, the asphericity of these curvatures, corneal and lenticular thickness, anterior chamber depth, the refractive indices of the all ocular elements involved in light transmission, and the eccentricity of the retinal feature being measured (Garway-Heath et al., 1998). Because of the impracticality of gathering all of the above information, summarising formulae that make certain assumptions of the eye can be used to obtain an accurate estimate of the ocular effect of magnification. Techniques used include those based solely on spectacle refraction (Bengtsson and Krakau, 1992), ametropia and keratometry (Bengtsson and Krakau, 1992; Littman, 1982), axial length only (Bengtsson and Krakau, 1992; Bennett et al., 1994), axial length and ametropia (Littman, 1988; Wilms, 1986), and those utilising all of axial length, anterior chamber depth, lens thickness, keratometry and ametropia (Bennett et al., 1994). Garway-Heath et al. (1998) found the abbreviated axial length method employed by Bennett et al. (1994) differs little from the more detailed calculations using keratometry, ametropia, anterior chamber depth and lens thickness. They found that Littman’s (1982) technique based on keratometry and ametropia to be the least accurate.

5.2. Dimensionless measures of retinal topography

Whilst all the above techniques make assumptions about the optics of the eye, they serve as reasonable estimates for calculating true retinal features from retinal photographic images. However, in studies collecting large numbers of patients, it may be difficult to acquire such information. Hence, studies have sought dimensionless measures, thus nullifying any magnification effect and allowing measurements between subjects to be compared. Such dimensionless entities that have been used include the arteriovenous ratio (AVR), junctional exponents, angles at vessel bifurcations, measures of vascular tortuosity, length:diameter ratios and fractal dimensions.

5.2.1. Measuring retinal vessel widths

Attempts at quantifying retinal arteriolar calibres were first considered by Wagener et al. (1947). The introduction of retinal photography in the 1960s allowed semi-objective methods of performing measurements on retinal vasculature using enlarged projected images (micrometric methods) (Arzabe et al., 1990; Bracher et al., 1979; Burgess, 1967; Cunha-Vaz and Lima, 1978; Hodge et al., 1969; Hubbard et al., 1992; Parr and Spears, 1974a, b). The introduction of digital image analysis in the mid-1980s provided more objective measurements of retinal vascular widths (Brinchmann-Hansen, 1986; Delori et al., 1988; Eaton and Hatchell, 1988; Gao et al., 2000; Newsom et al., 1992; Penn and Gay, 1992; Rassam et al., 1994; Stromland et al., 1995; Wu et al., 1995). Digitalised image analysis techniques are more reliable than previous micrometric techniques (Delori et al., 1988; Newsom et al., 1992; Sherry et al., 2002). Densitometric techniques employ intensity profiles of a grey-scale image of the fundus (microdensitometry). The location of each pixel can be identified with spatial co-ordinates and each has a defined intensity, known as its grey value. An intensity profile of a line crossing perpendicular to the blood vessel will tend to produce a distinct Gaussian distribution curve against the background intensity of the surrounding retina. The
Gaussian (or double-Gaussian) model can then be analysed using image processing, and an estimate of the width of the blood vessel can be obtained.

The single Gaussian model is given by the equation

\[ f(x) = a_1 e^{-\frac{(x-a_2)^2}{2a_3^2}} + a_4, \]

where \( a_1 \) is the amplitude of the peak of the profile, \( a_2 \) the position of the peak, \( a_3 \) a specific parameter of Gaussian function that controls the width of profile and \( a_4 \) the background retinal intensity.

The most common technique for acquiring the vessel width is to estimate the width of the vessel at half the height of the peak of the intensity profile of the Gaussian curve (half-height method). This strategy minimises any effect of defocusing at image acquisition (Brinchmann-Hansen, 1986), which may be caused by medial opacities. Poor-quality captured images can be enhanced using grey-level transformation functions (e.g., linear contrast stretch, histogram equalisation or contrast-limited adaptive histogram equalisation; Gonzalez and Woods, 1992) to improve the contrast of the retinal vessels. Other potential problems with retinal vessel width measurements include width variation due to the cardiac cycle (Chen et al., 1994; Dumskýj et al., 1996; Knudtson et al., 2004), degree of systemic autonomic nerve stimulation (Baer and Hill, 1990; Lanigan et al., 1988) and degree of fundus pigmentation (Hubbard et al., 1992). Because retinal arterioles are small (approximately 50–200 µm in width), very high-resolution digital images must be obtained to perform accurate measurements from vessels that may be as small as 15–20 pixels in width. Also, measurements from retinal photography are based on the width of the blood column, rather than the actual blood vessel width, as it does not take into account the plasma component of blood in the peripheral vessels.

Other techniques of automated vessel width measurement have included the use of edge detection masks (Gonzalez and Woods, 1992) and sliding linear regression filters (Chapman et al., 2001; Gang et al., 2002). Rassam et al. (1994) have used “kick-points” on the image histogram which, although appearing to be more accurate in determining vessel width for good-quality images, are more prone to errors due to defocus.

Chapman et al. (2001) compared three different automated strategies to measure retinal vessel widths (Gaussian intensity profiles, edge detector, sliding regression linear filter) with manual measurement in red-free images. They found the most reliable of the three techniques was the sliding linear regression filter. The edge detector program frequently misinterpreted the central bright light reflex from the arteriole as the vessel edge.

Whilst performing individual retinal vessel measurements can provide some information regarding an individual’s retinal vasculature, it would be more advantageous to obtain a measure of overall retinal arteriolar and venular calibre. The most commonly performed dimensionless measurement that has been used as a measure of the width of the retinal vessels is known as the arteriolar–venular ratio (AVR) (Hubbard et al., 1992, 1999; Stanton et al., 1995a).

### 5.2.2. The arteriovenous ratio

The AVR was first suggested as a good parameter to investigate retinal vascular geometry by Stokoe and Turner (1966). It was developed as a general measure of the ratio between the average diameters of the arterioles with respect to the venules. It is comprised of two components, the central retinal artery equivalent (CRAE) and the central retinal vein equivalent (CRVE), expressed as a quotient. The CRAE was first devised by Parr and Spears (1974a, b), who developed an estimation from arteriolar trunk and branch vessels around a predefined zone concentric with the optic disc. Each individual vessel was measured, and paired vessels were combined to estimate the trunk vessels, and then paired trunk vessels were combined, and this iterative process was continued until all vessels had been combined into a summary measure of the mean CRAE. The formula that Parr et al. devised to calculate the calibre of the trunk vessel from the two branch vessels is detailed below:

**For arterioles**

\[
W_c = e(1.01 W_a^2 - 0.22 W_a W_b - 10.76),
\]

where \( W_c \) is the calibre of trunk arteriole, \( W_a \) the calibre of the smaller branch arteriole and \( W_b \) the calibre of the larger branch arteriole.

The Parr approach to calculate the CRAE was dependent on carefully tracing out the individual paired vessels, and was labour intensive and time consuming.

Hubbard et al. (1992) developed a similar measure to calculate the CRVE, again using a selection of young normotensive individuals and calculating a formula that would best describe the relationship between the trunk retinal venule and its branches.

**For venules**

\[
W_c = e(0.72 W_a^2 + 0.91 W_b^2 + 450.05),
\]

where \( W_c \) is the calibre of trunk venule, \( W_a \) the calibre of the smaller branch venule and \( W_b \) the calibre of the larger branch venule.

A further development by Hubbard et al. (1999) was to allow vessels to be paired according to an arbitrary pattern, where the largest vessel was combined with the smallest vessel and the second largest with the second smallest, etc. This was continued until all vessels had been combined. If there were an odd number of vessels, the residual vessel was carried over to the next iteration. This technique offered clear advantages by being less time-consuming and in an analysis of 10 eyes correlated.
well with the original Parr technique, with no evidence of fixed or proportional systematic bias. Thus, the AVR was calculated based on the calibres of all arterioles and venules passing through a concentric ring, which was defined as between 0.5 and 1 DD from the optic disc margin. This was chosen as it was felt that retinal blood vessels at the margins of the disc may be of an arterial configuration, whereas they are unambiguously arteriolar approximately 0.5–1 DD from the disc margin (Hubbard et al., 1999; Parr, 1974). Other amendments were made based on the individual calibre of vessels (if vessel calibre was > 80 μm, then the branches were considered, rather than the vessel itself and if vessels were < 25 μm, then they were not included in the calculations). The atherosclerosis risks in communities (ARIC) study was the first to utilise an objective, semi-automated AVR as a measure of generalised retinal arteriolar narrowing in response to systemic disease (Hubbard et al., 1999). The AVR was felt to be a good measure of generalised arteriolar attenuation, as there was evidence that arterioles would be much more affected by narrowing in response to cardiovascular disease than corresponding venules (Hubbard et al., 1999; Leung et al., 2003b).

The AVR has been used in a large number of epidemiological studies, such as the ARIC study, the Blue Mountains Eye Study, the Wisconsin Epidemiologic Study of Diabetic Retinopathy, the Cardiovascular Health Study, the Beaver Dam Eye Study and Rotterdam Study. It has proved to be a useful measure of generalised arteriolar attenuation (Table 2). In addition, there is good evidence that the AVR correlates well between right and left eyes (Leung et al., 2003a; Wong et al., 2004b).

However, there are some conflicting results regarding AVR, particularly in its association with atherosclerosis (Ikram et al., 2004; Klein et al., 2000, 2004b; Wong et al., 2001b, 2002a, 2003b), which may reflect different populations between the various studies. In an elderly population, after controlling for age, gender, race, mean arterial blood pressure and antihypertensive medication, the AVR was not associated with prevalence of coronary heart disease, stroke, myocardial infarction or presence of carotid disease (Wong et al., 2003b). The ARIC study did find an association between AVR and carotid plaque, but not with any other markers of atherosclerosis, either clinical (cardiovascular disease or stroke) or subclinical (carotid artery or popliteal thickness, lower limb peripheral vascular disease), serum cholesterol (Klein et al., 2000) or incidence of congestive cardiac failure (Wong et al., 2005). Furthermore, it is unclear whether using measures such as the AVR from retinal image analysis provides additional information regarding future risk of these systemic disease, over and above current standardised methods of clinical assessment (Wong, 2004). For a review of retinal micro-

<table>
<thead>
<tr>
<th>Report associations between arteriovenous ratio (AVR) and systemic/ocular factors</th>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atherosclerosis risk in communities</td>
<td>Cardiovascular health study (Wong et al., 2003b)</td>
</tr>
<tr>
<td>Blue mountains eye study</td>
<td>Retina study (Ikram et al., 2004)</td>
</tr>
<tr>
<td>Blood pressure (Wong et al., 2004); south and blood pressure (Wong et al., 2004); familial tendency (Lee et al., 2004); retinal pigment epithelial degeneration (Klein et al., 2000); cardiovascular mortality (Klein et al., 2000); cardiovascular mortality in age range 44–74 years old (Wong et al., 2003a)</td>
<td></td>
</tr>
<tr>
<td>Age (Wong et al., 2003b)</td>
<td></td>
</tr>
<tr>
<td>Blood pressure (Klein et al., 2003); age (Wong et al., 2004b); body mass index (Klein et al., 2003); myocardial infarction in both sexes (Klein et al., 2004b) and severity of diabetic retinopathy (Klein et al., 2004d)</td>
<td></td>
</tr>
<tr>
<td>Blood pressure (current, past and incident) (Hubbard et al., 1999; Sharett et al., 1999; Klein et al., 2000); metabolic syndrome (Wong et al., 2004); no association was found between AVR and prevalence of coronary heart disease, myocardial infarction, stroke, or presence of carotid disease (Wong et al., 2003b)</td>
<td></td>
</tr>
<tr>
<td>Blood pressure (current, past and incident) (Hubbard et al., 1999; Sharett et al., 1999; Klein et al., 2000); metabolic syndrome (Wong et al., 2004); no association was found between AVR and prevalence of coronary heart disease, myocardial infarction, stroke, or presence of carotid disease (Wong et al., 2003b)</td>
<td></td>
</tr>
<tr>
<td>Age (Wong et al., 2003b)</td>
<td></td>
</tr>
<tr>
<td>Blood pressure (Klein et al., 2003); age (Wong et al., 2004b); body mass index (Klein et al., 2003); myocardial infarction in both sexes (Klein et al., 2004b) and severity of diabetic retinopathy (Klein et al., 2004d)</td>
<td></td>
</tr>
<tr>
<td>Blood pressure (current, past and incident) (Hubbard et al., 1999; Sharett et al., 1999; Klein et al., 2000); metabolic syndrome (Wong et al., 2004); no association was found between AVR and prevalence of coronary heart disease, myocardial infarction, stroke, or presence of carotid disease (Wong et al., 2003b)</td>
<td></td>
</tr>
</tbody>
</table>
vascular changes in cardiovascular disease, see Wong et al. (2001b).

A limitation of the AVR is that venules and arterioles may have a different response to different pathologies, e.g., the venules may be dilated in response to an inflammatory condition, whereas the arterioles are attenuated due to underlying hypertension. Hence, the independent use of the CRAE and CRVE may provide information regarding vessel changes in certain pathological states that would otherwise be undetected by using the combined AVR. Whilst the CRAE and CRVE are not dimensionless measurements, studies have reported these measurements of retinal vascular calibre in association with systemic disease (Ikram et al., 2004; Klein et al., 2004c, d; Lee et al., 2004; Leung et al., 2003b; Wang et al., 2003; Wong et al., 2004a). A few of these studies had refractive data in order to partially adjust for magnification effect from retinal photography (Wong et al., 2004b, c). The Beaver Dam Eye Study (Wong et al., 2004b) found that myopic refraction was associated with smaller retinal vessel diameters, but there was no data on axial length in this study, and the authors speculate as to whether this is purely a magnification effect, or whether it represents a biological or pathological process in eyes with different refractions. They also highlight the need for future studies with axial length data to explore more precisely its impact on retinal vascular diameters and their association with systemic cardiovascular disease. The Blue Mountains Eye Study (Wong et al., 2004c) found that smaller arterioles and venules (as determined by the CRAE and CRVE) were associated with myopic refraction. After correction for magnification using the Bengtsson (1976) formula, there was no association between retinal vessel diameters and refraction, but it did not alter the association between AVR and hypertension. No data on axial length was included in this study, and using the Bengtsson method for image magnification is known to be associated with a systematic increase in method error in long eyes, as only part of the variation in axial length is manifest by ametropia (Garway-Heath et al., 1998). However, correcting for magnification error did appear to increase the statistical power to detect associations with retinal vessel diameters, and the authors recommend that refraction be taken into account for the detection of retinal vessel changes in association with cardiovascular disease for future studies. More recently, we found axial length to have no bearing on retinal vascular network geometry, including AVR, junctional exponents and angles at vessel bifurcations (Patton et al., 2005b).

5.2.3. "Revised" AVR

A limitation of the Parr–Hubbard formula is the measurements are converted from pixels to micrometres, and therefore direct pixel calculations cannot be performed. An estimate of the pixel-to-micrometre ratio is calculated based on an average optic DD of 1850 μm (Hubbard et al., 1999). Another limitation is that the number of vessels measured has a significant impact on the overall AVR calculation (Knudtson et al., 2003). Knudtson et al. (2003) developed a revised measure of AVR formula based on the six largest arterioles and venules passing through the previously defined zone B (concentric area between 0.5 and 1 DD, centred on the disc), which is independent of the units of scale, and less dependent of the number of vessels measured. This revised Parr–Hubbard formula correlated strongly with the previous formula, but was found to be independent of the number of vessels measured, unlike the previous Parr–Hubbard formula (p < 0.05). They arbitrarily chose to measure the six largest arterioles and venules and calculate a “branching coefficient” based on vessel widths between the trunk vessel and the two branch vessels:

\[
\text{branching coefficient} = \frac{w_1^2 + w_2^2}{W^2},
\]

where \( W \) is the width of the trunk vessel, and \( w_1 \) and \( w_2 \) are the two branch vessels.

In a sample of 44 healthy young normotensive subjects, measuring a total of 187 arteriolar junctions, the branching coefficient was found to be 1.28 (95% CIs 1.25–1.32). This compared well with a theoretical value of 1.26 (based on a dichotomous symmetrical vessel bifurcation—see below) (Fig. 4) (Sherman, 1981). From 151 venular junctions, the branching coefficient was calculated as 1.11 (95% CIs 1.08–1.14). Thus, by placing the calculated values into the above formula, they calculated that:

\[
P = \text{Power} \quad D = \text{Drag} \quad V = \text{Volume} \quad S = \text{Surface}
\]

Fig. 4. Graph of power losses, drag, volume and surface area costs for junctional exponents (X) and angles at bifurcations. Note, as Murray predicted, power losses and volume are minimised when X approximates to 3, and the angle at bifurcations is approximately 75°.
Formula and their calculated branching coefficient of difference in root mean square deviation between their healthy, young, normotensive population. Parr found a (Knudtson et al., 2003) and both groups employed a junctions, compared to Knudtson’s calculated value of 1.2 for vascular vessels (Parr and Spears, 1974b). For the vessel radius (Sherman, 1981). The original Parr study undertaken reanalysis of some of the previously published analyses using the revised formulae, and noted overall associations were still detected but with tighter CIs. Based on these findings, the revised Parr–Hubbard formula should be regarded as the new reference standard for the measurement of AVR. Studies are now employing the revised AVR to determine retinal vessel changes in cardiovascular disease (Ikram et al., 2004; Wong et al., 2004a, d).

The AVR is a useful device for obtaining an estimate of generalised arteriolar width. However, it has limitations, other than its dependence on the number of retinal vessels measured and the presence of formulaic constants requiring measurements to be performed in micrometres. The AVR was constructed by producing a formula that minimised the observed spread of values for retinal vascular branching points using a least-squares strategy (Parr and Spears, 1974b). For the CRAE, this was done using micrometric methods, which have been shown to be less reliable than modern microdensitometric techniques (Newsom et al., 1992). The theoretical optimum for the branching coefficient of a dichotomous, symmetrical junction is 1.26 [(2)\(^{1/3}\)] (Sherman, 1981; Young, 1809). The original Parr study found a branching coefficient of 1.2 for vascular junctions, compared to Knudtson’s calculated value of 1.28 which is much closer to the theoretical value (Knudtson et al., 2003) and both groups employed a healthy, young, normotensive population. Parr found a difference in root mean square deviation between their formula and their calculated branching coefficient of 0.45 μm (mean parent widths both 83 μm), which may be considered marginal. It is unclear from Parr’s original paper how much difference would have existed if the Parr formula had been compared with theoretically optimum values of branching coefficients. In addition, all images were considered to have a magnification of 2.5, with no correction for magnification considered, although the subjects’ refractions ranged from −3.5 to +2 DS. Whilst the Parr–Hubbard formula has served well as a calculation of the AVR, the ‘new’ revised formula of Knudtson et al. (2003) based on branching coefficients may have greater power to detect smaller associations between the AVR and systemic factors.

### 5.2.4. Optimality at vascular junctions

Vascular topographical geometry, far from being a totally random network, has a tendency to conform to some ‘optimal’ principals, in order to minimise physical properties such as shear stress and work across the vascular network (Murray, 1926a,b; Sherman, 1981; Zamir, 1976a; Zamir and Medeiros, 1982; Zamir et al., 1979). In 1926, Murray calculated the most efficient circulation across a vascular network can be achieved if blood flow is proportional to the cubed power of the vessel’s radius (known eponymously as Murray’s law). This was deduced from the assumption of blood acting as a Newtonian fluid (flow rate is proportional to the pressure difference across the vessel, and excluding any effect of gravity and kinetic energy) and Poiseuille’s law (resistance to fluid in a vessel is proportional to the fourth power of the vessel radius, and inversely proportional to the vessel length) and assuming that the viscosity of blood is constant, and metabolism of the blood and vessel tissue remain constant throughout the vascular system. The power required to maintain flow is greatly reduced by small increments in vessel radius (proportional to the fourth power of the vessel radius), but the power to maintain metabolism is increased by small increments in the vessel radius (proportional to the square of the vessel radius). By differential calculus, it can be shown that for flow across a vascular network to be constant requires it to be proportional to the cube of the vessel radius (Sherman, 1981).

### 5.2.5. Junctional exponent

If we consider the relationship between the diameter of the parent vessel \(D_0\) and the diameter of the two daughter vessels \(D_1\) and \(D_2\) (Fig. 5), then the following relationship exists in vascular junctions:

\[
D_0^X = D_1^X + D_2^X,
\]

where \(X\) is the junctional exponent.

Thus, according to Murray, theoretical values for the value of \(X\) (junctional exponent) approximate to the value of 3 in healthy vascular networks in order to minimise power losses and intra-vascular volume (Fig. 4).

The branching coefficient used to calculate the ‘revised’ AVR is derived from Murray’s law, in the situation where the two daughter vessels are equal in diameter \((D_1 = D_2)\) (Knudtson et al., 2003).

Consider \(D_0^1 = D_1^1 + D_2^1\).
In a symmetrical, dichotomous junction, \( D_1 = D_2 \), and thus this can be rewritten as \( D_3^0 = 2D_3^1 \). Thus, \( D_0 = (2)^{1/3}D_1 \).

The branching coefficient detailed in Section 5.2.3 relates the area of daughter to parent vessels as a ratio, such that daughter:parent ratio (area) \( = 2^{1/3} : 1 \). Hence, the theoretical value of 1.26 \((2^{1/3})\) for daughter:parent area ratio (Fig. 6). This compared favourably with Knudson’s et al. calculated value of 1.28. This can also be expressed in relation to the width of parent to daughter vessels as a ratio, such that daughter:parent ratio (width) = \(1:2^{1/3}\).

A variety of animal and human tissue circulations conform to an approximation of Murray’s law (Horsfield, 1978; Mall, 1888; Miller, 1893; Weibel, 1963). In addition, changes in this optimal geometrical topography are known to occur with increasing age (Stanton et al., 1995b) and in diseased coronary arteries (Hutchings et al., 1976).

### 5.2.6. Optimality parameter

Chapman et al. (2002) further developed the concept to try to overcome some of the problems associated with calculating junctional exponents. Occasionally, a peripheral branch vessel width may be greater than the parent vessel width, particularly in vascular junctions that do not conform to optimal junctional bifurcation.

In this situation where \( D_1 \) or \( D_2 > D_0 \), no such real value of \( X \) can exist. In addition, junctional exponents are sensitive to even small changes in vessel measurement. This is significant when dealing with what may be vessels of no more than 10–15 pixels diameter, even in high-resolution images. Hence, Chapman et al. developed a new “optimality parameter”, to be able to get a measure of how much the pattern of vessel widths at any junction deviate from the optimum junctional exponent of 3.

This is given by the equation:

\[
\rho = \frac{D_0^3 - (D_1^3 + D_2^3)^{1/3}}{D_0},
\]

where \( \rho \) is the optimality parameter, \( D_0 \) the diameter of the parent vessel, and \( D_1 \) and \( D_2 \) are the diameters of the two daughter vessels.

This new calculation was found to be less prone to small errors in vessel measurement than an iterative procedure designed to calculate the junctional exponent. In addition for circumstances where \( D_1 \) or \( D_2 > D_0 \), a value for \( \rho \) can still be calculated as it is possible to calculate a cube root of a negative number. Using this new optimality parameter, Chapman et al. (2002) found that there was a significant difference in the ‘optimality’ of retinal vascular junctions between healthy individuals and those with peripheral vascular disease. Griffith et al. (1991) suggest a possible role for endothelium in maintaining optimal junctional exponents, possibly via nitric oxide and endothelin-1.

Despite the evidence that junctional exponents and optimality parameters are affected by systemic factors,
they have been relatively few studies using these as markers of vascular topography, when compared with the AVR. The concept of ‘optimalisation’ of vascular systems is of fundamental importance in the future construction of artificial organs. If vascular supply to organs can be described by theoretical biological optimisation principals, then new organs can be designed with blood supplies that minimise shear stress, fatigue, volume and energy across the vascular network (Schreiner et al., 1996, 2003; Zhou et al., 1999). By being easily visualised and photographed in vivo and readily analysed using digital image analysis, the retinal blood vasculature is uniquely placed in helping to elucidate and define optimality for vascular networks.

5.2.7. Vascular bifurcation angles

In addition to junctional exponents fitting theoretical values in an ‘optimised’ vascular network, the angle subtended between two daughter vessels at a vascular junction has also been found to be associated with an optimal value, approximately 75° depending on which costs (surface, volume, drag or power) (Woldenberg, 1986; Zamir, 1976b) are considered and the degree of asymmetry between the two daughter vessels (Griffith and Edwards, 1990) (Fig. 6). Retinal arteriolar bifurcation angles are known to be reduced in hypertension (Stanton et al., 1995b), increasing age (Stanton et al., 1995a) and low birth weight males (Chapman et al., 1997). Reduced angles at vascular junctions are associated with less dense vascular networks (Kiani and Hudetz, 1991). In addition, vascular responsiveness to high oxygen saturation leads to a reduced angle at retinal vascular junctions, but this responsiveness is known to be reduced in hypertensives (Chapman et al., 2000). No relationship was reported between vascular bifurcation angles and peripheral vascular disease, compared with normal healthy controls (Chapman et al., 2002). Using X-ray microangiography in an animal model, Griffith et al. (1991) found branching angles to be unaffected by blood flow rate or change in vasomotor tone. Associations with angles between daughter vessels at vascular junctions and systemic factors have not been extensively investigated and it is unclear how angles at vascular junctions may serve as independent predictors of systemic disease. Further studies exploring these relationships are needed.

5.2.8. Vascular tortuosity

The degree of tortuosity of a vessel could be summarised as the ratio between the distances a vessel travels from A to B, and the shortest distance between points A and B drawn by a straight line. Conditions such as ROP have utilised indices of vascular tortuosity as a measure of disease severity (Capowski et al., 1995; Heneghan et al., 2002; Swanson et al., 2003). In 1995, Capowski et al. (1995) reported using an arterial tortuosity index from fundal photographs as a useful measure of ROP disease state. Freedman et al. (1996) used computer-aided analysis of fundus photographs from eyes with a wide range of ROP severity, and traced posterior pole blood vessels diameter and tortuosity.

There is a need to identify and quantify signs of plus disease as early as possible before ROP has progressed to the point where outcome is compromised. The earliest signs of plus disease are venous engorgement and increased arteriolar tortuosity around the optic disc. Hence, great potential lies in the use of digital image analysis in providing quantitative objective measurements of retinal tortuosity in these patients, which can be compared between different examiners and sequentially over time. In this regard, Swanson et al. (2003) describe a semi-automated retinal vascular image analysis system (retinal image scale-space analysis) to measure retinal diameters and tortuosity in preterm infants. Whilst the difficulties in acquiring sharp, focused images from these patients may be considerable, the authors were able to obtain enough measurements on relatively low-resolution images to find a significant association between arteriolar tortuosity and ROP severity. The authors point out some potential advantages of using such a system rather than regular clinical examination in the screening of ROP. Visualising the posterior pole alone with the aim of diagnosing plus disease would considerably reduce the duration and trauma of examination and may allow healthcare professionals other than ophthalmologists to undertake screening. They point out the potential of tele-screening to further improve cost effectiveness of the scheme, and the future potential of automated diagnosis (or grading) based on the captured fundal images.

Venous beading, as a feature of diabetic retinopathy has also been examined using Fourier analysis (Kozusek et al., 1992) (image transformation based on the fact that a periodic function may be written as the sum of sine’s and cosine’s of varying amplitudes and frequencies) or using thresholding to extract the vein, partitioning the vein segment into 32 pixel segments and generating diameter data along the length of these vessel segments (Gregson et al., 1995). Hart et al. (1999) were able to measure and classify retinal vascular tortuosity from RGB colour images using an automated approach.

5.2.9. Length:diameter ratio

King et al. (1996) developed the length:diameter ratio as another dimensionless measure of network topography, reflecting retinal arteriolar attenuation. This is calculated as the length from the midpoint of a particular vascular bifurcation to the midpoint of the preceding bifurcation, expressed as a ratio to the diameter of the parent vessel at the bifurcation. They found this to be increased in hypertension (King et al.,
1996), but Chapman et al. (2002) found no association with peripheral vascular disease. Quigley and Cohen (1999) developed a measure of retinal topography that is derived from Poiseuille’s law, Ohm’s law and Murray’s law. This “pressure attenuation index” also reduces to the length:diameter ratio of a retinal arteriole segment. This index predicts that the longer and/or thinner the retinal arterioles, the greater will be the pressure attenuation. This may explain the observed “protective” effect of conditions such as myopia for diabetic retinopathy (Pierro et al., 1999). Further studies to explore systemic influences on length:diameter ratios in retinal vessels are needed.

5.2.10. Fractal geometrical analysis

Fractal geometry is commonly encountered in nature, e.g., branching patterns in trees, snowflake patterns, etc. The concept of fractals as mathematical entities to describe complex natural branching patterns, such as that present in biological systems was first considered by Mandelbrot (1967, 1982). Fractals are based on the concept of self-similarity of spatial geometrical patterns despite a change in scale or magnification so that small parts of the pattern exhibit the pattern’s overall structure. The fractal dimension \( D \) (in the context of vascular branching patterns) describes how thoroughly the pattern fills two-dimensional spaces. Unlike Euclidean dimensions such as length, area or volume which are normally described by integer values (1, 2 or 3), fractal dimensions are usually non-integers, and lie somewhere between 1 and 2. Different models of formation of fractals have been developed, but the one most commonly used to describe vascular branching patterns is the diffusion limited aggregation (DLA) model, developed by Witten and Sander (1981). The basic principal involves a particle that moves in a random fashion until it gets close to part of the existing structure, at which point it becomes an adherent component of the structure. The process is started with a seeding structure, normally a single point, and continues until the structure reaches a desired size.

Just as Murray (1926a) predicted that junctional exponents should be approximate to the value \( x = 3 \), Mandelbrot suggests that this value would also generate a vascular network in which the most distal vessels would exactly fill the available space (i.e., fills less of the available space) within the macular region using a region-based fractal analysis of retinal fluorescein angiograms, although no such difference was observed outside the macular region (Avakian et al., 2002). However, as the authors point out, use of fractal analysis in clinical practice requires more comprehensive studies to elucidate what additional information over and above conventional assessment is gathered in pathological vascular states. Fractal dimensions may elucidate significant biological change during the early stages of disease, before other features of retinopathy appear. However, the global analysis of the retinal circulation may miss these changes and not be sensitive to early disease (Masters, 2004). Zamir (1999) has pointed out that fractal analysis of vascular systems is not useful unless the variability within the arterial system (such as the junctional exponents and vascular junction angles) is permitted within the analysis. This variability may be masked by a simple fractal dimension that is unable to differentiate two vasculatures that have the same space-filling properties, but widely different structures in their fluid dynamic design and function (Zamir, 2001).

It is currently unclear what role fractal analysis may have, but potential knowledge of an optimised framework whereby vascular branching structures are formed may have future implications in the design of optimal artificial organs. If large numbers of branching structures are described by optimised fractal dimension models such as DLA, can we infer the human vascular system is no different?

5.3. Reliability of quantitative measurements from retinal image analysis

Newsom et al. (1992) compared the retinal vessel width measurement techniques of observer-driven
micrometric techniques (making manual measurements from a projected image), and objective computer-driven microdensitometry, based on a vessel’s profile “grey-level” intensity level, using the previously described “half-height” technique, that has been shown to be the most accurate in the presence of focusing errors (Brinchmann-Hansen, 1986). The coefficient of variation for computer-driven microdensitometry was calculated as 1.5–7.5%, compared to 6–34% for the observer-driven technique. Delori et al. (1988) also found a greater variability for micrometric than microdensitometric techniques. Brinchmann-Hansen (1986) reports a coefficient of variation using microdensitometry of 1.5%. George et al. (1990) report a coefficient of variation ranging from 1.2 to 3.4% with an average of 2.2% for microdensitometry.

Sherry et al. (2002) report intra-observer reliability $\kappa$ values ranging from 0.8 (for trunk AVR ratios) to 0.93 (for CRVE measurements). $R^2$ correlation analysis showed agreement ranging from $R^2 = 0.79$ to 0.92. For inter-observer reliability, $\kappa$ ranged from 0.71 (for branch AVR measurements) to 0.9 (for CRVE measurements), and correlation statistics showed $R^2$ ranging from 0.78 to 0.9. As one would expect, there was better agreement for larger vessels (CRVE) and better intra-observer than inter-observer agreement. In the ARIC study (Hubbard et al., 1999), inter-observer agreement ($n = 151$ eyes) was $R = 0.74, 0.77$ and 0.79, for CRAE, CRVE and AVR, respectively. For intra-observer agreement, $R = 0.69, 0.89$ and 0.84 for CRAE, CRVE and AVR, respectively.

Suzuki (1995) reports on a direct method of vessel measurement using automated detection of vessel edges by processing the one-dimensional retinal image obtained by a linear image sensor set in a fundus camera, producing results in real-time. The coefficient of variation and the inter-observer variation of the direct method for all measurements were $1.71 \pm 1.13\%$ and $2.25 \pm 1.92\%$, respectively. The inter-observer variation of the direct method was smaller than those of the microdensitometric methods.

As stated earlier, Chapman et al. (2002) found the junctional exponent to be sensitive to small changes in vessel width measurement. The introduction of the ‘optimality parameter’ significantly reduced the consequences of small errors measurement, using a Monte-Carlo simulation (Chapman et al., 2002). They also report intra-observer repeatability for retinal vessel widths (used to calculate junctional exponents and optimality parameters) on the basis of within-subject standard deviation as 0.92 pixels (mean width 21.1 pixels; coefficient of variation 4.36%) and for angles 2.66° (mean angle 68.9°; coefficient of variation 3.86%). For calculating length:diameter ratio, within-subject standard deviation was 5.09 pixels (mean length 406.9 pixels; coefficient of variation 1.25%).

Little data are available regarding the reliability of fractal analysis, although studies show little variation of the fractal dimension within the population being studied (Avakian et al., 2002; Landini et al., 1995; Masters et al., 1992).

5.4. Measurement of retinal vessels in real-time

The quantitative measures of retinal vascular topography described above are all based on digital image analysis of retinal images from retinal photography. However, conventional or digital photography cannot be used to obtain real-time continuous recording of vessel changes, which may be useful to illicit short-term retinal vascular changes to different pharmacological effects, for example. The retinal vessel analyser (RVA®) (Imedos, Weimar, Germany) consists of a retinal fundus camera, a CCD video camera, a real-time monitor for electronic online image acquisition, and a PC for overall system control, image analysis and result archiving (Seifert and Vilser, 2002; Vilser et al., 2002). It allows real-time assessment of retinal vascular diameters at a maximum frequency of 50 Hz (allowing 25 vessel diameters readings per second) and has demonstrated reproducible results (Pache et al., 2002; Polak et al., 2000). Adaptive algorithms allow for measurement of retinal vessel widths, utilising the absorbing properties of haemoglobin in each blood vessel. The system is able to automatically correct for slight adjustments in luminance that may occur due to slight eye movement, and thus vessel diameter can be recorded as a function of time, as well as position along the vessel. A major limitation of the RVA® is that it assumes that the eye under measurement has no refractive error (emmetropia) and uses standardised units to measure vessel diameters. Therefore, the RVA® is unable to give actual measurements of vessel wall widths if a significant number of subjects do not conform to the assumptions of emmetropia. However, attempts at finding a value for the diameter of the central retinal artery in vivo using the RVA® have been performed, utilising the diameters of all retinal veins entering the optic disc and laser Doppler velocimetry as a measure of the total retinal blood flow, and combining this with the velocity of blood flow in the central retinal artery (Dorner et al., 2002).

6. Digital retinal vascular image analysis and telemedicine

Because ophthalmology is largely dependent on visual information, it is an ideal specialty for telemedicine (Constable et al., 2000; Lamminen et al., 2003; Murdoch, 1999; Yogesan et al., 1998). Digital capture of images and the potential for transmission of these
images via electronic transfer across large distances with subsequent image analysis offers the potential for more efficacious use of medical resources in large, rural communities that may otherwise have difficulty obtaining expert opinion (Yogesan et al., 2000). The most common system utilised in ‘tele-ophthalmology’ is ‘store-and-forward’, where images are captured, and later transmitted electronically to be analysed at a later date. This contrasts with live video-conferencing, which is currently limited by electronic transmission rates (Murdoch, 1999). “Tele-ophthalmology” could be utilised between primary health care practitioners, optometrists and ophthalmic specialists, or between different ophthalmic units (Bowman et al., 2003; Murdoch, 1999). Telemedicine has even been used to aid prison medical officers in diagnosing ophthalmic complaints, and thus reducing costs and potential complications of prisoners attending specialist medical centres (Yogesan et al., 2001). In a collaborative international project, telemedicine has been found to be cost effective in reducing the burden of eye-disease, and that richer countries may aid capacity building in health care systems of poorer countries (Johnstone et al., 2004). Countries with large areas of sparsely populated communities such as Canada (Burnier, 2003), Australia (Constable et al., 2000) and India (Kumar et al., 2003) may greatly benefit in terms of health care delivery to these areas.

Telemedicine has a potential role in diabetic screening (Choremis and Chow, 2003; Constable et al., 2000). Kawasaki et al. (2003) report that 1076 of 1170 eyes’ fundal images (92%) were successfully evaluated by a consultant ophthalmologist, when images were transferred via electronic mail. Lin et al. (2002) report single non-mydriatic monochromatic wide-field digital photography of the disc and macula to be more sensitive for diabetic retinopathy screening than mydriatic ophthalmoscopy, when transmitted electronically to a reading site. When adjudicated by standard seven-field colour photographs, the higher sensitivity of digital photography primarily reflected the reduced sensitivity of ophthalmoscopy in detecting early retinopathy. TOSCA (Tele-Ophthalmological Services Citizen-Centred Application) was developed in Europe as a project to reduce the incidence of blindness caused by diabetic retinopathy (Luzio et al., 2004). Telemedicine has been explored in screening for ROP. Yen et al. (2002) found RetCam (Massie research Laboratories, Inc., Dublin, CA) images captured by a neonatal nurse compared well with examinations performed by an experienced ophthalmologist with good sensitivity, but only moderate specificity. In addition, teleophthalmology has been utilised in macular diseases (Berger and Shin, 1999. Ophthalmology 106, 1935–1941).

Eikelboom et al. (2000) report on the effect of JPEG and wavelet digital image compression on the quality of images for telemedicine. JPEG image compression breaks the image into blocks of \(8 \times 8\) pixels and converts these blocks into spatial frequency components. Sampling of this frequency domain information by closely preserving the low-frequency components and approximating the high-frequency components is performed and the amount of information discarded determines the amount of compression. Wavelet employs band filters and low pass filters to the pixel rows and columns of an image. This produces information on the low-frequency components of the image and the horizontal, vertical and diagonal detail in the image (and is more computationally intensive). Eikelboom et al. found that wavelet compression to 15 KB for digital image transmission was ideal when time and costs are to be minimised. For computational time to be minimised, the use of JPEG compression to 29 KB was a good alternative.

All studies to date using telemedicine in ophthalmology have not extended the digital process to digital image analytical techniques. This may be difficult due to the need for relatively high-resolution images in order to perform quantitative digital image analysis. Transmission of such high-resolution images is currently impractical for telemedicine. However, image compression algorithms are currently still evolving and with improved technology, it may be possible to transmit sufficiently high-resolution images to enable digital image analysis.

7. Future directions

Retinal digital image analysis is able to exploit the ease with which the retinal circulation can be visualised, photographed, and analysed non-invasively in vivo. Using objective, quantitative measures from retinal vasculature which are based on principals of optimisation of a branching vasculature, studies have been able to improve our understanding of the effect of systemic factors on the microvasculature. The most commonly performed quantitative measurement from digital retinal vascular image analysis has been the AVR. Whilst this has proved to be a very useful research tool to measure generalised arteriolar narrowing, very large epidemiological studies have been required to have sufficient statistical power to be able to detect associations of this entity with systemic factors. It is also unclear from current studies whether the detection of retinal microvascular changes has additional predictive value above current standardised methods (Wong, 2004). Recently, ‘revised’ formulae for the AVR (Knudtson et al., 2003) may hold greater promise for future studies to find weaker associations with greater statistical power. Furthermore, in contrast to the widespread use of the AVR, other measures of retinal
vascular topography have been under utilised. Future studies to explore junctional exponents and optimality parameters at vascular junctions, angles at vessel junctions and fractal geometrical fundal patterns may provide more information regarding the retinal vasculature in systemic disease. Large epidemiological studies have also found value in the relationship between features of focal microvascular changes and systemic factors. In future, development of automated detection systems may perform these tasks without recourse to observer-driven methods which are more prone to fatigue, bias and are more time-consuming.

Technology is ever developing and improving the way that we can analyse fundal images (Stefansson, 2004). The concept of using digital retinal images from scanning laser ophthalmoscopes for biometric identification is already a reality (www.retinaltech.com). The promise of greater image resolution (modern digital cameras exceed 16 megapixels) and computer processing power in the future may allow more sensitive detection of retinal microvascular changes and lead to automated diagnosis from fundal images being a practical and efficient adjunct to ophthalmic diagnostics. Digital retinal vascular image analysis may also permit an assessment of a particular individual’s specific risk stratification for a variety of cardiovascular conditions, and may have particular relevance to cerebrovascular risk (Patton et al., 2005a). The advent of non-mydriatic wide-field cSLO providing over 200° images of the retina may allow the detection of further information regarding peripheral retinal vascular junctions, and allow a more complete examination of the retinal vascular tree. Such technology with further improvements may provide screening opportunities for those considered at high risk of such conditions affecting the far retinal periphery, e.g., retinal tears in high myopes based on automated computer algorithms. New imaging modalities such as fundus autofluorescence offer further opportunities to employ digital image analysis for the objective, quantitative measurements of fundal characteristics (Bellmann et al., 2003).

With an increasingly aged population and increased strain on medical resources, the use of strategies such as telemedicine and widespread screening of individuals at risk of certain diseases will increase. Retinal vascular digital image analysis will play an ever greater role in clinical ophthalmology.

References


Klein, B., Klein, R., McBride, P., Cruickshanks, K., Palta, M., Knudtson, M., Moss, S., Reineke, J., 2004b. Cardiovascular disease,


Miller, W., 1893. The structure of the lung. J. Morphol. 8, 165–188.


Murray, C., 1926b. The physiological principle of minimum work applied to the angle of branching arteries. J. Gen. Physiol. 9, 835–841.


Young, T., 1809. On the function of the heart and arteries. Philos. Trans. R. Soc. Lond.


