ELSEVIER

International Journal of Medical Informatics 53 (1999) 239–252

International Journal of Medical Informatics

Measurement and classification of retinal vascular tortuosity

William E. Hart [a,*], Michael Goldbaum [b], Brad Côté [c], Paul Kube [d], Mark R. Nelson [e]

[a] *Department of Applied and Numerical Mathematics, Sandia National Laboratories, Alburquerque, NM 87185, USA*
[b] *Department of Ophthalmology, University of California, San Diego, CA, USA*
[c] *The Registry, Inc., San Diego, CA, USA*
[d] *Department of Computer Science and Engineering, University of California, San Diego, CA, USA*
[e] *Data Vector, San Diego, CA, USA*

Abstract

Automatic measurement of blood vessel tortuosity is a useful capability for automatic ophthalmological diagnostic tools. We describe a suite of automated tortuosity measures for blood vessel segments extracted from RGB retinal images. The tortuosity measures were evaluated in two classification tasks: (1) classifying the tortuosity of blood vessel segments and (2) classifying the tortuosity of blood vessel networks. These tortuosity measures were able to achieve a classification rate of 91% for the first problem and 95% on the second problem, which confirms that they capture much of the ophthalmologists' notion of tortuosity. Finally, we discuss how the accuracy of these measures can be influence by the method used to extract the blood vessel segments.

Keywords: Tortuosity; Retina; Blood vessel; Automated measurement

1. Introduction

Normal retinal blood vessels are straight or gently curved. In some diseases, the blood vessels become tortuous, i.e. they become dilated and take on a serpentine path. The dilation is caused by radial stretching of the blood vessel and the serpentine path occurs because of longitudinal stretching. The tortuosity may be focal, occurring only in a small region of retinal blood vessels, or it may involve the entire retinal vascular tree. Fig. 1 shows images with tortuous and non-tortuous blood vessels.

Many disease classes produce tortuosity, including high blood flow, angiogenesis and blood vessel congestion. High blood flow may occur locally in arteriovenous anastamo-

* Corresponding author. E-mail: wehart@cs.sandia.gov.

1386-5056/99/$ - see front matter
PII: S1386-5056(98)00163-4

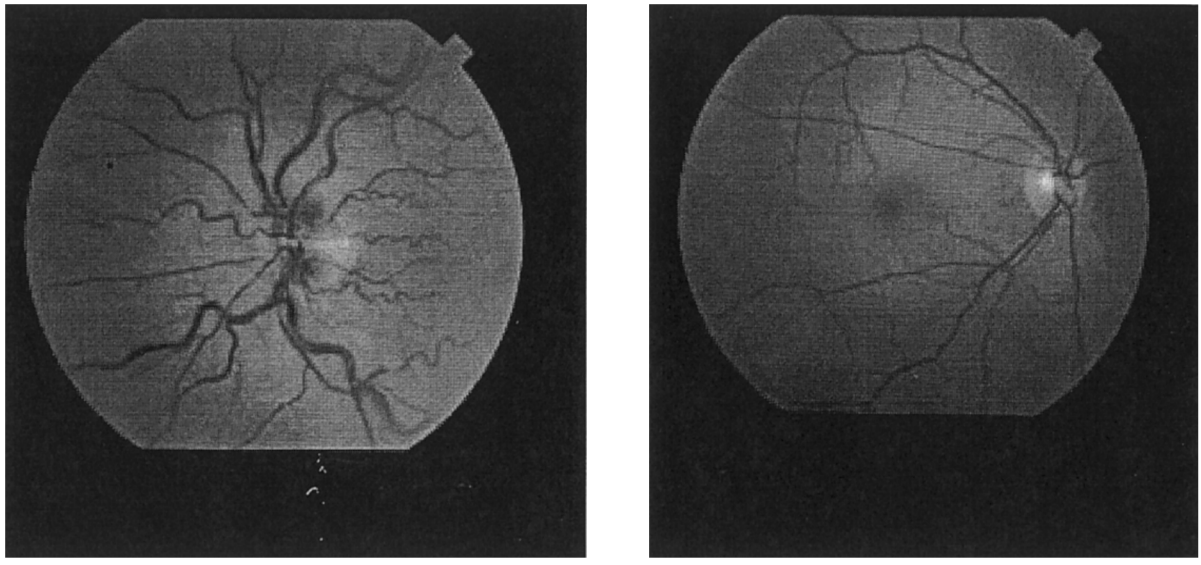

Fig. 1. Images with (a) tortuous and (b) non-tortuous blood vessel segments.

sis or systematically in high cardiac output such as accompanies anemia. In response to ischemia or inflammation, new blood vessels grow, often in a destructive manner. The angiogenic process commonly causes dilation and tortuosity of retinal blood vessels prior to the onset of the neovascularization. Venous congestion can result from the occlusion of the central retinal vein or a branch retinal vein. The resulting elevated intravascular pressure results in tortuosity and dilation in the blocked vein.

Information about disease severity or change of disease with time may be inferred by measuring the tortuosity of the blood vessel network. Consequently, there is a benefit in measuring tortuosity in a consistent, repeatable fashion. Vascular tortuosity measurements are commonly performed by human observers using a qualitative scale (e.g. mild, moderate, severe and extreme). Variability in grading can result because the boundaries between the grades may differ between observers and with the same observer at different times. Furthermore, measurements on such a gross scale make changes in vascular tortuosity difficult to discern.

A quantitative tortuosity measurement was first described by Lotmar, Freiburghaus and Bracher [1] and slightly extended by Bracher [2]. This measurement involves the manual selection of points on fundus photographs that subdivide the tortuous vessel into a series of single arcs. Fig. 2 illustrates the mea-

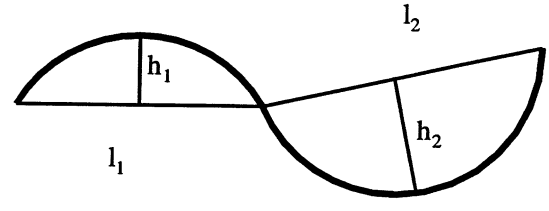

Fig. 2. Illustration of the subdivision used to estimate blood vessel tortuosity with the relative length variation.

surements they use to calculate the vessel's tortuosity. Their method decomposes the vessel into a series of circular arcs, for which the chord lengths l_i and arrow heights h_i are measured. The tortuosity is measured as the *relative length variation*

$$\frac{L-l}{l} \approx \frac{8}{3} \sum_i \left(\frac{h_i}{l_i}\right)^2,$$

where L is the length of the blood vessel and l is the chord length. The approximation is derived using a sinusoidal model of a blood vessel. This method was used by Kylstra et al. [3] to measure the effects of transmural pressure on the tortuosity of latex tubing.

Given the increased availability of digitized fundus photographs, automated tortuosity measurements are now feasible. Kaupp et al. [4] have reported unpublished results of an automated tortuosity measurement that uses a Fourier analysis of the perpendicular along the blood vessel. Smedby et al. [5] describe five tortuosity measures used to measure tortuosity in femoral arteries. Included are several measures of the integral curvature along the blood vessel, the number of inflection points of the vessel and the fraction of the vessel that has high curvature. Their experiments examine properties of these measures like reproducibility and scalability. Capowski, Kylstra and Freedmen [6] describe a measure of blood vessel tortuosity based on spatial frequencies. Zhou et al. [7] have also described a method for distinguishing tortuous and nontortuous blood vessels in angiograms. In a preliminary abstract, we have proposed a tortuosity measure based on the integral curvature along a blood vessel [8].

In this paper we describe tortuosity measures that are used to measure the tortuosity of retinal blood vessels as well as the retinal blood vessel network. We motivate our tortuosity measures by analyzing abstract properties of measures based on medical intuitions of tortuosity. To assess the relative utility of these measures, they were used to classify blood vessel segments and blood vessel networks. The segments used in these classification experiments were extracted manually and automatically and we discuss how the tortuosity measures can be influenced by properties of the method used to extract the vessel segments. The classification rate was as high as 91.5% for blood vessel segments and 95% for blood vessel networks. While no single tortuosity measure was clearly superior, the experiments recommend a total squared curvature measure.

2. Tortuosity measures

2.1. Abstract properties

We begin by discussing several properties of tortuosity measures that are motivated by the ophthalmologist's notion of tortuosity. A tortuosity measure that has many of the properties described below should provide good predictive performance in our classification tasks, thereby matching an ophthalmologist's measure of tortuosity. We will discuss the following properties of tortuosity measures: invariance to translation and rotation, response to scaling and compositionality of vessels and vessel networks. These properties are defined for parametrized differentiable curves $C = (x(t), y(t))$, with t in an interval $[t_0, t_1]$. We model retinal blood vessels with such curves. A tortuosity measure τ takes a curve C as its argument and returns a real number.

2.1.1. Invariance to translation and rotation

Suppose a vessel with tortuosity τ were moved on the retina without changing its shape or size. What should be the tortuosity of the resulting vessel?

While ophthalmologists' judgments of vessel tortuosity do seem to incorporate the relative curvature of other vessels on the fundus, we believe their judgment is largely independent of the location or orientation of the vessels. Our measures assume that vessel tortuosity is not dependent on the location of orientation of the vessel.

2.1.2. Response to scaling

Suppose a vessel with tortuosity τ was enlarged uniformly by some scale factor $\gamma > 1$. What should be the tortuosity of the scaled vessel?

Ophthalmologists do not seem to have unequivocal intuitions about this question, though they believe that a vessel's tortuosity should be roughly invariant to scaling. However, it is clear that if scale does affect the tortuosity then it does so in a multiplicative manner. That is, if curve $C = (x(t), y(t))$ has tortuosity $\tau(C)$, curve $C' = (\gamma x(t), \gamma y(t))$ has tortuosity $\beta(\gamma) \cdot \tau(C)$ for some function β. If tortuosity is invariant to scaling, $\beta(\gamma) \equiv 1$; and if tortuosity is inversely related to scale, $\beta(\gamma) < 1$ when $\gamma > 1$. We examine tortuosity measures that exhibit various tortuosity scale factors $\beta(\gamma)$ and investigate their effect on classification performance.

2.1.3. Vessel compositionality

Suppose a vessel were composed of two vessel segments C_1 and C_2, one with tortuosity $\tau(C_1)$ and the other with tortuosity $\tau(C_2)$. What should be the tortuosity of the whole vessel?

It appears to be the intuition of ophthalmologists that a vessel that is composed of two segments of different tortuosities would have a degree of tortuosity between the tortuosities of its constituent segments. This assumes that the two segments are smoothly connected, but the 'order' in which segments occur in the vessel is not important. It also appears that if a vessel segment with a given tortuosity were extended with a segment of the same tortuosity, the resulting vessel would have the same tortuosity as either of its segments. That is, a vessel doesn't increase (or decrease) its tortuosity by having the same serpentine pattern extended at greater length.

We formalize these intuitions as follows. We write $C = C_1 \otimes C_2$ and say curve C has constituent segments C_1 and C_2. Whenever $C = (x(t), y(t))$ with $t \in [t_0, t_1]$, $C_1 = (x(t), y(t))$ with $t \in [t_0, t_2]$ and $C_2 = (x(t), y(t))$ with $t \in [t_2, t_1]$, for some $t_2 \in [t_0, t_1]$. Now suppose that $\tau(C_1) \leq \tau(C_2)$. We say that a tortuosity measure has the property of compositionality when, for any such C_1, C_2, C_3,

$$\tau(C_1) \leq \tau(C) \leq \tau(C_2) \tag{1}$$

with the equality holding if and only if $\tau(C_1) = \tau(C_2)$.

The property of compositionality provides constraints on how to compute the tortuosity of a vessel, given the tortuosities of its constituent segments. This computation will be necessary in an automated system that extracts vessel segments and not whole vessels from fundus images. A method of computing the tortuosity of a whole vessel that is consistent with Eq. (1) is to weight the tortuosity of each constituent segment by the fraction of arc length which that segment contributes to the vessel. That is, $\tau(C_1 \otimes C_2) = [s(C_1)\ \tau(C_1) + s(C_2)\ \tau(C_2)]/s(C_1 \otimes C_2)$, where $s(C)$ is the arc length of the curve C (see below). We call this method weighted additivity.

On the other hand, computing the tortuosity of a vessel by averaging the tortuosities of its constituent segments is inconsistent with Eq. (1). It follows from Eq. (1), together with the assumption that the tortuosity of a vessel is independent of how it is segmented, that the tortuosity of a vessel cannot be a function only of the tortuosities of its constituent seg-

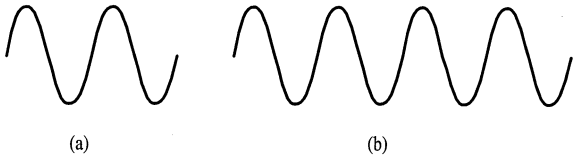

Fig. 3. Two curves with the same tortuosity for a measure with chord-colinear compositionality.

ments. Let $\tau(C_1) = \tau(C_2) < \tau(C_3) = \tau(C_4)$ and consider the curves $C_1 \otimes C_2 \otimes C_3$ and $C_2 \otimes C_3 \otimes C_4$. If we segment these curves as $C_1 \otimes (C_2 \otimes C_3)$ and $(C_2 \otimes C_3) \otimes C_4$, then the tortuosity measures for these two curves are different. However, if these curves are segmented as $(C_1 \otimes C_2) \otimes C_3$ and $C_2 \otimes (C_3 \otimes C_4)$, the tortuosities of the constituent segments of the two vessels are the same, viz. $\tau(C_1)$ and $\tau(C_3)$.

Finally, we note that the property of compositionality implies that extending a vessel segment 'in the same way' does not affect its tortuosity. The vessel segment which constitutes the extension would have the same tortuosity as the vessel segment, so the tortuosity of the extended vessel remains the same. Some of the proposed tortuosity measures will be invariant to some types of extensions without strictly satisfying the compositionality property. We say that a measure has the property of chord-colinear compositionality if a vessel C is segmented such that each segment has the same tortuosity and the chords of the segments are colinear, then the tortuosity of the vessel is the same as the constituent segments. A tortuosity measure with this property is invariant to chord-colinear extensions. Fig. 3 shows two curves that will have the same tortuosity for a measure that has chord-colinear compositionality. This is an interesting property since retinal blood vessels are often roughly periodic along a straight line. Thus they can be partitioned into segments whose chords are colinear and whose tortuosity is roughly equivalent.

2.1.4. Network compositionality

Another type of compositionality concerns the way in which the tortuosity measures for vessel segments are combined to determine a tortuosity measure for an entire vessel network. This type of compositionality might differ from vessel compositionality because the method for combining the tortuosity values from vessel segments may depend on whether they are part of the same vessel. We have yet to develop a method for extracting a complete vessel network, so we calculate the tortuosity of a blood vessel network using the weighted additivity of all of the blood vessel segments in the image.

2.2. Definitions

Using our previous definition of a curve segment C, we define a suite of tortuosity measures. We begin by defining the components used to construct these measures. These measures are defined for a curve $C = (x(t), y(t))$ on the interval $[t_0, t_1]$.

2.2.1. Arc length

The arc length of C is

$$s(C) = \int_{t_0}^{t_1} \sqrt{x'(t)^2 + y'(t)^2}\ \mathrm{d}t.$$

2.2.2. Chord length

The chord length of C is

$$\mathrm{chord}(C) = \sqrt{(x(t_1) - x(t_0))^2 + (y(t_0) - y(t_1))^2}$$

2.2.3. Total curvature

The curvature of C at t is

$$\kappa(t) = \frac{x'(t)y''(t) - x''(t)y'(t)}{[y'(t)^2 + x'(t)^2]^{3/2}}, \tag{2}$$

and the total curvature of a curve segment C is

Table 1
Summary of the tortuosity measures and their properties

Tortuosity measure	Response to scale	Chord-colinear compositionality	Compositionality	
τ_1	$s(C)/\text{chord}(C)-1$	1	✓	
τ_2	$\text{tc}(C)$	$1/\gamma$		
τ_3	$\text{tsc}(C)$	$1/\gamma^2$		
τ_4	$\text{tc}(C)/s(C)$	$1/\gamma^2$	✓	✓
τ_5	$\text{tsc}(C)/s(C)$	$1/\gamma^3$	✓	✓
τ_6	$\text{tc}(C)/\text{chord}(C)$	$1/\gamma^2$	✓	
τ_7	$\text{tsc}(C)/\text{chord}(C)$	$1/\gamma^3$	✓	

$$\text{tc}(C) = \int_{t_0}^{t_1} |\kappa(t)| \, \mathrm{d}t.$$

2.2.4. Total squared curvature

The total squared curvature of C is

$$\text{tsc}(C) = \int_{t_0}^{t_1} \kappa(t)^2 \, \mathrm{d}t.$$

Table 1 defines the measures that we examined in our experiments and describes their properties. These measures have zero measure for straight vessel segments and increasing positive measure for segments as they become tortuous. The variety of tortuosity measures allows us to compare measures to determine the relative importance of the tortuosity properties described previously.

The measure τ_1 simply computes the tortuosity of the segment by examining how long the curve is relative to its chord length. This measure is the same as the *distance factor* tortuosity measure described by Smedby et al. [5] and it is very similar to the relative length variation proposed by Lotmar et al. [1]. If the curve C is a single arc, or if C is chord-colinear, then $l = \text{chord}(C)$ and τ_1 is equivalent to relative length variation.

Measures τ_2 and τ_3 directly calculate the curvature of the curve. Unfortunately, neither of these measures have either of the compositionality properties. The τ_3 measure differs from τ_2 in that it places a greater emphasis on the parts of the curve that have high curvature and de-emphasizes the parts of the curve that have low curvature. Since the curvature is greater for small vessels, τ_3 will emphasize the tortuosity of smaller vessels more than τ_2. The τ_2 measure is the same as the total curvature measure described by Smedby et al. [5].[1]

The remaining measures are 'length-normalized'. Measures τ_4 and τ_5 average the total curvature measures by the arclength, while τ_6 and τ_7 average by the chord length. The measures τ_4 and τ_5 have the advantage that they have the property of compositionality. We believe that these two measures a priori come closest to the ophthalmologist's notion of tortuosity.

2.3. Tortuosity calculation

The definitions for our tortuosity measures apply to abstract curves. The skeletonized blood vessels in our data sets consist of sequences of pixel locations that approximate the center line of the blood vessels in the retinal images. The arc length of a skeletonized blood vessel with n pixel locations

[1] We disagree with the claim by Smedby et al. that the total curvature measure is invariant to linear scaling. If $x(t)$ and $y(t)$ are multiplied by a constant factor, A, then their derivatives are multiplied by A. Thus the curvature measure in Eq. (2) is scaled by $1/A$.

(x_i, y_i) can be calculated by walking along the curve and summing the distance between neighboring points:

$$\sum_{i=1}^{n-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2}.$$

To calculate the total curvature and total squared curvature, recall [9] that the curvature of a parameterized curve can also be defined as

$$\kappa(t) = \frac{d\alpha}{ds}(t),$$

where $\alpha(t)$ is the angle of the tangent line at t. To compute $(d\alpha/ds)(t)$, we measured $\alpha(t)$ and $s(t)$ and computed the derivative of $\alpha(t)$ with respect to $s(t)$. Here, $s(t)$ is the arc length between (x_1, y_1) and (x_t, y_t) We determined $\alpha(t)$ by calculating the angle of the line connecting (x_t, y_t) and (x_{t+1}, y_{t+1}). To compute the derivative $(d\alpha/ds)(t)$, we used a window of 10 values of $s(t)$ and $\alpha(t)$: $(s(t-4), \alpha(t-4)),\ldots, (s(t+5), \alpha(t+5))$. A line, $f(x) = mx + b$, was fitted to these points with a linear regression method to minimize

$$\sum_i (\alpha(i) - f(s(i)))^2.$$

The derivative $(d\alpha/ds)$ was equated with m, the slope of this line. The $n-9$ values for $(d\alpha/ds)$ were used to numerically estimate the integrals needed to calculate the total curvature and total squared curvature.

Following a suggestion of Flynn and Jain [10], we smoothed the pixel representation of a blood vessel segment before making our tortuosity calculations. We independently smoothed the x and y coordinate sequences, $\{x_1,\ldots, x_n\}$ and $\{y_1,\ldots, y_n\}$ using the smooft method described by Press et al. [11], which applies a low-pass filter to the data. This method eliminated undesirable noise that is due to the discrete nature of the pixel representation. For example, a blood vessel at a 45° angle in a image can have a pixel representation that is a zigzag line of pixels along the blood vessel. The smoothed coordinates are not restricted to the discrete pixel grid, so their tortuosity measures more closely reflect the tortuosity of the actual blood vessel. Preliminary experiments indicated that two applications of this smoothing method were sufficient to reduce this noise.

The smoothing operation also eliminated noise that was introduced when independent segments were linked together to form a longer segment. Adjacent segments extracted from a blood vessel can form a non-smooth sequence of pixels. Smoothing combined segments produced a sequence of points that more closely resembled the contour of the original blood vessel.

3. Methods

We examined the utility of our tortuosity measures by using them as features in two different classification problems. In the first problem, we classified blood vessel segments as tortuous or non-tortuous. In the second problem, we classified the tortuosity of the entire blood vessel network.

3.1. Data

Blood vessel segments were extracted from a set of 20 retinal images in two different ways: automatically and manually. To extract blood vessel segments, we applied the blood vessel filter described in Chaudhuri et al. [12] to the green plane of an RGB image; the green plane was selected because it typically exhibits the greatest contrast. The filter was applied at 12 orientations over 180°. The final response map was computed by taking the maximum response of the 12 filters at each location. We thresholded and thinned the

response map of the blood vessel filter to produce an image containing binary edge segments. The edge segments were primarily blood vessels, but also included some edges of large objects like the optic nerve.

To create the set of *automatically* extracted blood vessel segments, the edge segments were classified as blood vessels or non-blood vessels using the linear classifier described in Côté et al. [13]. There were 981 automatically extracted blood vessel segments, of which 252 were tortuous and 729 were non-tortuous.

The unclassified edge segments were also used to extract blood vessel segments *manually*. The edge segments were manually identified as blood vessels and linked together to form the final blood vessel segments. Care was taken to link vessel segments only if the smoothness of the link reflected the curvature of the underlying blood vessel. This is reasonable since blood vessel segments are smoothed before the tortuosity measures are calculated. There were 284 blood vessels extracted manually, of which 133 were tortuous and 151 were non-tortuous.

We are primarily interested in the automatically extracted blood vessels, since they can be used as part of automated diagnostic tools. However, the blood vessel filter tends to break up longer blood vessels. This usually occurs when two vessels cross in an image, where a vessel branches or when a vessel bends very rapidly. Since tortuous blood vessels can bend very rapidly, they are broken up more often than non-tortuous vessels. The manually extracted blood vessel segments provide a benchmark against which the performance of the automatically extracted segments can be compared.

After extraction, the tortuosity measures were calculated for each segment and the segments were labeled by one of the authors (MG), a retinal specialist who has experience with diseases causing tortuosity in retinal blood vessels. These two sets of extracted blood vessels were used for the first classification problem.

For the second classification problem, the extracted blood vessels were used to calculate the network tortuosity measures for each image. The vessel network tortuosity of the 20 retinal images were labeled; there were ten tortuous and ten non-tortuous images.

3.2. Classification

3.2.1. Logit models

We used a logit model [14] to classify the data for both classification problems. For problems with two classes, a logit model computes a weighted sum of the input features passed into a logistic function. Let $\{f_1,\ldots, f_n\}$ be n input features and $\{w_0,\ldots, w_n\}$ be the $n+1$ weights. The two-class logit model is

$$f(x, w) = g\left(w_0 + \sum_{i=1}^{n} w_i\ f_i\right),$$

where $g(x)$ is the logistic function $g(x) = 1/(1+\mathrm{e}^{-x})$. The output of the logit model is between zero and one. To perform classification, the output is thresholded to zero or one, depending on whether the output is greater or less than 0.5.

Given a set of training examples, $\{(x_1, y_1),\ldots, (x_n, y_n)\}$, the optimal weight vector is found by minimizing

$$J(w) = \sum_i E(y_i h, f(x_i, w))$$

where

$$E(a, b) = \begin{cases} -\log(1-b)\ , & a = 0 \\ -\log(b)\ , & a = 1 \end{cases}$$

This is the 'cross-entropy' error, which is motivated by an analysis based on the relative entropy of the distribution of outputs

with the distribution of binary data [15]. We found the optimal weight vector by minimizing $J(w)$ with the conjugate gradient method [11].

Logit models are well suited for modeling the probability distribution of binary data. If the data are linearly separable, logit models will simply perform a linear classification of the data. However, for data that are not linearly separable, logit models perform differently from linear models. Aldrich and Nelson [16] describe how nonlinear transformations like the logistic function modify the behavior of the linear model. Further, they note that the there are a variety of reasons why the assumption that a probability model is linear is unrealistic in most cases. In preliminary experiments, we found that logit models had better classification performance than linear models.

3.2.2. Performance measures

Two measures were used to compare the performance of classifiers using the different tortuosity measures: classification rate and the integrated relative operating characteristic (ROC). The classification rate is simply the proportion of test samples that are correctly classified. The ROC measures the proportion of test samples that are correctly classified as positive instances (true-positives) as a function of the proportion of negative test samples that are classified as positive instances (false-positives) [17]. The integrated ROC measures how a classifier's performance varies for different classification thresholds. The integrated ROC, A, typically assumes values between 0.5 and 1.0. When A is near 0.5, the classifier performs no better than chance alone. When A is near 1.0, the classifier has near perfect discrimination for most classification thresholds.

The ROC was measured by calculating the output of the logit model for all of the test samples. The outputs were split into 20 groups based on the rank of their values. The true-positive and false positive proportions were measured for each group. The trapezoidal rule was used to integrate the ROC values. This method of calculating the ROC is similar to the method described by Swets [17].

The performance on the samples used to train a classifier is typically an optimistic estimate of the classifier's performance on a new set of data [18]. To estimate the true error rate, we used cross-validation to partition the data into two subsets that are used to train and test the data. In the first classification problem, we used 50 randomly selected partitions for which 70% of the data were used to train the classifiers. A different classifier was trained and tested on each partition. The mean classification rate and integrated ROC on the testing subsets were used to evaluate the expected performance of a classifier.

The second classification problem has a much smaller set of data, so we performed 10-fold cross-validation. The data set was randomly partitioned into ten sets, each of which contained one positive and negative example that were used for testing the classifier. Again, classifiers were trained and tested on each partition. The mean classification rate was used to evaluate the expected performance of a classifier. Because of the small size of the data set, the integrated ROC was calculated once using the values of the logit model for every test sample.

3.3. Experiments

In both classification problems, we performed experiments that examine the classification rate using each tortuosity by itself. We also performed experiments using all of the

Table 2
Results for the problem of classifying blood vessel segments and blood vessel networks. The mean of the cross-validation classification rate and integrated ROC

Measure	Segments				Networks			
	Classification rate		Integrated ROC		Classification rate		Integrated ROC	
	Manual	Auto	Manual	Auto	Manual	Auto	Manual	Auto
τ_1	79.5	91.3	0.897	0.955	65	90	0.61	0.91
τ_2	79.3	82.7	0.875	0.915	70	85	0.79	0.85
τ_3	82.9	89.5	0.927	0.960	90	90	0.86	0.86
τ_4	80.4	88.5	0.905	0.951	80	95	0.83	0.99
τ_5	81.2	87.7	0.906	0.941	85	90	0.84	0.83
τ_6	82.1	89.1	0.921	0.956	80	90	0.81	0.91
τ_7	82.0	88.1	0.914	0.944	85	90	0.88	0.91
All	85.6	91.5	0.935	0.970	95	90	0.95	0.86

tortuosity measures together. We applied Tukey's method for multiple comparisons [19] to see if there were significant differences in the classification rate of the classifiers in the two classification problems ($p < 0.05$).

4. Results

Table 2 summarizes the experimental results for the two classification problems. The experiment name indicates whether the data were manually extracted or automatically extracted and indicates which tortuosity measures were used in the experiment. For the segment classification problem, the manually extracted data set exhibited significant differences between the following pairs of tortuosity measures: (τ_1, τ_3), (τ_1, τ_6), (τ_1, τ_7), (τ_2, τ_3), (τ_2, τ_6) and (τ_2, τ_7). All of the interactions in the automatically extracted data set were significantly different, except the following: (τ_3, τ_6), (τ_4, τ_5), (τ_4, τ_6), (τ_4, τ_7) and (τ_5, τ_7). For the network classification problem, no significant differences were found for either the manually or automatically extracted data sets.

For most of the classifiers, the relative integrated ROC values mirrored the relative classification rates. An interesting exception is the integrated ROC for τ_1. For the segment classification problem, the classification rate of the classifier using τ_1 to classify automatically extracted blood vessels was significantly higher than the rest of the classifiers. However, the integrated ROC for τ_1 was lower than the integrated ROC for τ_3 and τ_6. This indicates that the classification threshold used in these experiments favors τ_1, but the τ_3 and τ_6 measures are robust for a wider range of classification thresholds.

In preliminary experiments, we examined tortuosity measures that multiplied the length-normalized measures by the average vessel diameter. However, these measures performed slightly worse than their counterparts. We thought these would improve the performance of our tortuosity measures since this increases the measure's scale relation without changing its tortuosity properties.

5. Discussion

Our experimental results demonstrate that the proposed tortuosity measures can be used to effectively classify the tortuosity of blood vessel segments and blood vessel networks. In particular, these results show that the tortuosity measures can be used with automatically extracted blood vessel segments, which is important for the development of retinal image analysis tools.

5.1. Blood vessel segment tortuosity

The better classification performance on the automatically extracted data set can be attributed to the short, simple blood vessel segments in this data set. While an entire vessel segment may be labeled tortuous, subsegments of the vessel may not be particularly tortuous. These non-tortuous subsegments will influence the overall tortuosity of the vessel segment, making it appear less tortuous and thereby more difficult to classify. The automatically extracted blood vessels are usually short enough to be entirely tortuous or non-tortuous.

This also explains why the τ_1 tortuosity measure performed much better on the automatically extracted data set. This measure does not accurately discriminate between long segments because it does not analyze the tortuosity along the segment. Since the tortuosity is more constant along shorter segments, this global measure performs well.

We can use the statistical analysis of the classification rates to analyze the relative importance of the abstract tortuosity measures. The only clear pattern for both of the data sets is that the τ_2 measure had consistently poor performance. In fact, τ_2 was significantly worse than τ_3 for both data sets. The only property not shared by τ_2 and τ_3 is response to scale, which suggests that the response to scale is an important property. The fact that the difference between τ_2 and τ_3 was greater on the manually extracted data set supports this conclusion.

5.2. Blood vessel network tortuosity

The results for the vessel network classification problem follow the same pattern as the results for the blood vessel classification problem. The classification performance was better on the automatically extracted data set and the τ_1 tortuosity measure performed much better on the automatically extracted data set.

It is surprising that the performance also increased for τ_4 and τ_5. These measures have the compositionality property, so the blood vessel networks should have roughly the same tortuosity measure for both data sets. An analysis of the data set indicated that the performance difference was due to a difference in the extraction process on one of the images. This particular image was tortuous, but only slightly so. The automatically extracted segments included some exudate edges that were misclassified as blood vessel edges during extraction. These edges had a high curvature and thereby made the image appear more tortuous. The manually extracted segments did not include these segments and thereby reduced its tortuosity measure to a point where the image appeared non-tortuous.

One other image was responsible for most of the remaining error for both data sets. This image is clearly tortuous, but it has very small blood vessels. The segmentation algorithm tends to smooth out very small blood vessels, so the tortuosity measures for this image were low enough to make it appear non-tortuous.

5.3. Blood vessel extraction

The experimental results indicate that the utility of the tortuosity measures can be affected by the manner in which the blood vessel segments are extracted. We have observed three different ways in which the extraction process influenced the performance in the classification tasks. While these influences are specific to our data, they highlight factors that will influence the performance of tortuosity measures on any data set.

First, the segmentation method can affect the tortuosity of the extracted vessels. Our segmentation algorithm breaks up vessels at points of very high curvature and tends to smooth out very thin blood vessels. These issues are less problematic when calculating the tortuosity of blood vessel networks. The broken blood vessels can be reconnected and factors that cause tortuosity will usually affect both large and small blood vessels.

The length of the extracted segments can also affect the utility of the tortuosity measure for classification. The shorter segments in the automatically extracted data sets is the principle reason for the better classification results for these data sets.

Finally, extracting blood vessel segments automatically can effect the tortuosity measures of blood vessel networks since they can include misclassified segments. Non-blood vessel segments are often more tortuous than blood vessel segments, so the tortuosity of a blood vessel network can be biased when misclassified segments are included. However, misclassified segments occur almost only in images that contain a large number of abnormal objects such as lesions. This problem did not seriously impact our results because the blood vessel classifier used to automatically extract the blood vessel segments is quite accurate. It has a 89.6% classification rate for non-blood vessels on a data set containing images with lesions [13].

6. Conclusions

Our experimental results demonstrate that the proposed tortuosity measures can be used to classify the tortuosity of blood vessel segments and blood vessel networks. We believe that automated tortuosity measures like these will play an important part of retinal image analysis tools. Automated measures are important because they provide consistent measurements that are directly comparable and can be used to detect changes in vascular tortuosity accurately.

Both classification problems used a coarse, two-class scale to evaluate the tortuosity measures. A coarse scale was chosen for the experiments to minimize the possible error of the hand labeling. One advantage of the proposed tortuosity measures is their ability to make tortuosity measurements in a continuous fashion. However, additional experiments are needed to evaluate whether the proposed tortuosity measures correctly predict tortuosity on a finer scale.

We have motivated our choice of tortuosity measures with abstract properties that formalize ophthalmological intuitions concerning tortuosity measures. Abstract properties provide a formal basis for comparing tortuosity measures, which enables us to analyze the prospective performance of new tortuosity measures. Our experimental results suggest that response to scaling is an important property for tortuosity measures, while the compositionality properties did not exhibited discernible influence on the classification performance. The importance of the response to scaling is particularly interesting because the results indicate that inverse scaling is preferred, which is contrary to our ophthalmological intuitions.

While our classification results do not strongly support the use of one measure over the others, the τ_3 measure appears the best

choice for a tortuosity measure. Both the τ_3 and τ_4 measures were closest to the ophthalmologist's notion of tortuosity, though the τ_3 measure performed slightly better on the classification tasks. Classification with all of the tortuosity measures was better than τ_3, but not in a statistically significant manner.

Because tortuosity measures are used to infer information about disease severity or change of disease over time, the reproducibility of tortuosity measures is an important property. We have not examined this issue in this study, but the analysis in Smedby et al. [5] suggests that our measures will be very reproducible. Smedby et al. [5] observe that the total curvature (τ_2) and distance factor (τ_1) measures are very reproducible across a series of images of the same blood vessel. All of the tortuosity measures that we have examined are very similarly (if not identical) to τ_1 and τ_2.

Acknowledgements

This work was supported by NIH Grant No. RO1LM05759. This work was performed in part at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

References

[1] W. Lotmar, A. Freiburghaus, D. Bracher, Measurement of vessel tortuosity on fundus photographs, Graefe's Arch. Clin. Exp. Ophthalmol. 211 (1979) 49–57.

[2] D. Bracher, Changes in peripapillary tortuosity of the central retinal arteries in newborns, Graefe's Arch. Clin. Exp. Ophthalmol. 218 (1986) 211–217.

[3] J.A. Kylstra, T. Wierzbicki, M.L. Wolbarsht, M.B. Landers III, E. Stefansson, The relationship between retinal vessel tortuosity, diameter and transmural pressure, Graefe's Arch. Clin. Exp. Ophthalmol. 224 (1986) 477–480.

[4] A. Kaupp, H. Toonen, S. Wolf, K. Schulte, R. Effert, D. Meyer-Ebrecht, M. Reim, Automatic evaluation of retinal vessel width and tortuosity in digital fluorescein angiograms, Invest. Ophthalmol. Vis. Sci. 32 (1991) 952.

[5] Ö. Smedby, N. Högman, S. Nilsson, U. Erikson, A.G. Olsson, G. Walldius, Two-dimensional tortuosity of the superficial femoral artery in early atherosclerosis, J. Vasc. Res. 30 (1993) 181–191.

[6] J.J. Capowski, J.A. Kylstra, S.F. Freedman, A numeric index based on spatial-frequency for the tortuosity of retinal-vessels and its application to plus disease in retinopathy of prematurity, Retina 15 (6) (1995) 490–500.

[7] L.A. Zhou, M.S. Rzeszotarski, L.J. Singerman, J.M. Chokreff, The detection and quantification of retinopathy using digital angiograms, IEEE Trans. Med. Imag. 13 (4) (1994) 619–626.

[8] N.P. Katz, M.H. Goldbaum, S. Chaudhuri, M.R. Nelson, Automated measurements of blood vessels in digitized images of the ocular fundus, Invest. Ophthalmol. Vis. Sci. 31 (1990) 1185.

[9] R. Courant, J. Fritz, Introduction to Calculus and Analysis, Wiley-Interscience, New York, 1965.

[10] P.J. Flynn, A.K. Jain, On reliable curvature estimation, in: IEEE Proceedings on Computer Vision and Pattern Recognition, 1989, pp. 110–116.

[11] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipies in C—The Art of Scientific Computing, Cambridge University Press, Cambridge, MA, 1990.

[12] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, Detection of blood vessels in retinal images using two dimensional blood vessel filters, IEEE Trans. Med. Imag. 8 (3), September, 1989.

[13] B. Côté, W.E. Hart, M. Goldbaum, P. Kube, M.R. Nelson, Classification of blood vessels in images of the ocular fundus. Technical Report CS94-350, University of California, San Diego, CA, 1994.

[14] J.S. Cramer, The Logit Model: An Introduction for Economists, E. Arnold, London, 1991.

[15] J.S. Bridle, Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters, in: D. Touretzky (Ed), Advances in Neural Information Processing Systems, vol. 2, Morgan Kaufman, San Mateo, CA, 1990, pp. 211–217.
[16] J.H. Aldrich, F.D. Nelson, Linear Probability, Logit and Probit Models. Number 07-045 in Quantitative Applications in the Social Sciences. Sage, 1984.
[17] J.A. Swets, Measuring the accuracy of diagnostic systems, Science 240 (1988) 1285–1293.
[18] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.
[19] J. Rice, Mathematical Statistics and Data Analysis. Statistics and Probability Series. Wadsworth and Brooks/Cole, 1988.