An Informational Approach to Measuring Tortuosity

By

George Walden

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTERS IN SCIENCE

(MATHEMATICS)

at the

CALIFORNIA STATE UNIVERSITY - CHANNEL ISLANDS

2017

APPROVED FOR THE MATHEMATICS PROGRAM

12/20/17

Dr. Geoff Dougherty, Advisor | Date

12/19/17

Dr. Ivona Grzegorczyk | Date

APPROVED FOR THE UNIVERSITY

12/20/17

Dr. Joseph Shapiro | Date

Interim AVP & Dean

Extended University & International Programs

Non-Exclusive Distribution License

In order for California State University Channel Islands (CSUCI) to reproduce, translate and distribute your submission worldwide through the CSUCI Institutional Repository, your agreement to the following terms is necessary. The author(s) retain any copyright currently on the item as well as the ability to submit the item to publishers or other repositories.

By signing and submitting this license, you (the author(s) or copyright owner) grants to CSUCI the nonexclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that CSUCI may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that CSUCI may keep more than one copy of this submission for purposes of security, backup and preservation.

You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright. You also represent and warrant that the submission contains no libelous or other unlawful matter and makes no improper invasion of the privacy of any other person.

If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant CSUCI the rights required by this license, and that such third party owned material is clearly identified and acknowledged within the text or content of the submission. You take full responsibility to obtain permission to use any material that is not your own. This permission must be granted to you before you sign this form.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN CSUCI, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

The CSUCI Institutional Repository will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.

An Informational Approach to Measuring Tortuosity

Title of Item

Image Processing, Tortuosity, Curvature

3 to 5 keywords or phrases to describe the item

George Walden

Author(s) Name (Print)

George Walden — 2017-12-20

Author(s) Signature — Date

Abstract

Tortuosity is an intuitive term used to describe paths on the plane that exhibit multiple twists and turns. It is used in a variety of applications as a measure of how much a path deviates from a straight line, especially to analyze images. However, unlike curvature, tortuosity does not have a good mathematical definition yet. Nor is there a consensus on how to measure tortuosity.

This paper examines the information obtained from curvature and how it applies to the measure of tortuosity. Methods from the literature are examined and a new method is proposed using multiple measures instead of a single value.

Table of Contents

CHAPTER 1

TORTUOSITY

1.1 Introduction

The word tortuosity is an intuitive term used to describe paths that exhibit multiple twists and turns. It is often used to describe very commonplace shapes, like that of a road or a river. It is usually something that is not straight, nor is it a gentle curve. Rather, it is something that is crooked, lumpy, or misshapen. Definitions and descriptions of tortuosity often use phrases like “marked by repeated twists, bends, or turns” [PBG+81], “serpentine” [HGC+99] and “a type of geometric irregularity.” [Jos12]

Figure 1.1: A tortuous river (meander of Nowitna River, Alaska) [riv]

Mathematically, tortuosity should be a measure of how much a path deviates from a straight line. However, there is no commonly used way to measure tortuosity. Many have been proposed, but none have gained widespread acceptance as they do not capture the concept well. The literature on tortuosity shows little consilience towards a single description. This, in spite of the great advantage in having an objective measure of tortuosity, one that does not rely on the subjective opinion of a human observer.

The goal of this paper is to develop an objective measure of tortuosity based on the information that can be extracted from a path, and that describes its twists well.

This paper examines the mathematical definition of curvature of a curve and applying it to define the measurement of tortuosity. We review the literature and examine various algorithms that have been proposed to measure tortuosity of paths. Finally, this paper proposes using multiple measures of tortuosity to improve the ability to distinguish various shapes as a single measure cannot capture all of the information needed to describe tortuosity in a way that is useful to an application.

1.2 Guidelines for Paths

In this paper we use the word "path" for various non-self-intersecting curves in $\mathbb{R}^2$ being measured for tortuosity.

We give some examples here to underline the restrictions on the types of paths that one considers. Note that all paths have a starting point and an end point.

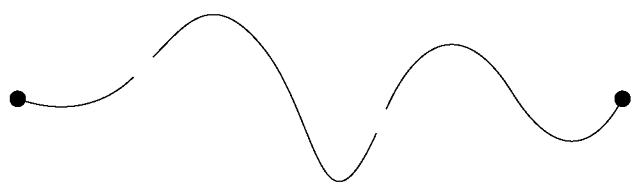


Figure 1.2: Paths cannot have any gaps, i.e. we consider only continuous paths.

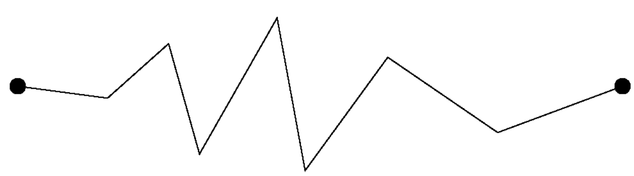


Figure 1.3: Paths cannot have any sharp corners, i.e. we consider only locally differentiable curves.

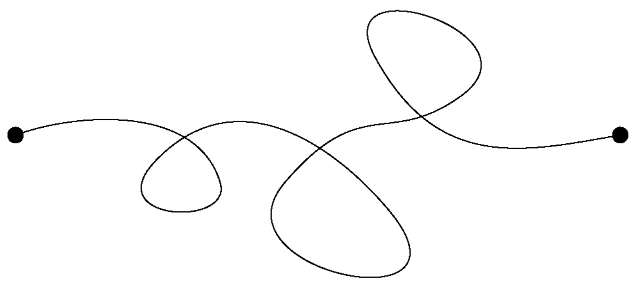


Figure 1.4: Paths cannot have any loops, or self-intersections.

1.3 Formal Definition for Paths

A path is described by a parametric equation $P(t) = \big(x(t), y(t)\big)$ over the non-empty interval $[t_0, t_1]$ and is subject to the following restrictions:

(a) $P(t)$ is continuous and differentiable over the interval $[t_0, t_1]$

(b) If $a \neq b$ then $P(a) \neq P(b)$ for all $a, b \in [t_0, t_1]$

1.4 Examples

1. $P(t) = (t, t)$ defined on $[0, 1]$ gives a segment of the straight line $y = x$ in $\mathbb{R}^2$.

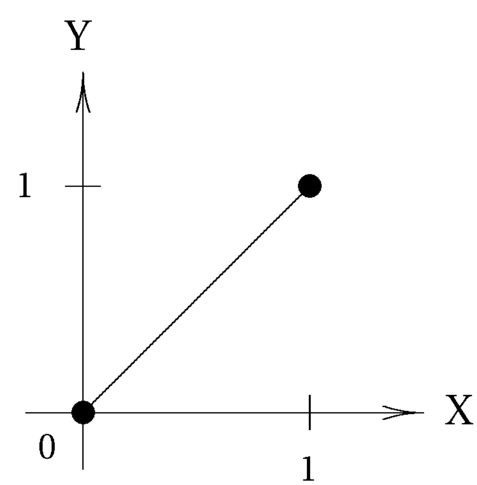


Figure 1.5: Segment of the Straight Line $y = x$

2. $P(t) = (\sin t, \cos t)$ on $[0, \pi]$ describes a half-circle.

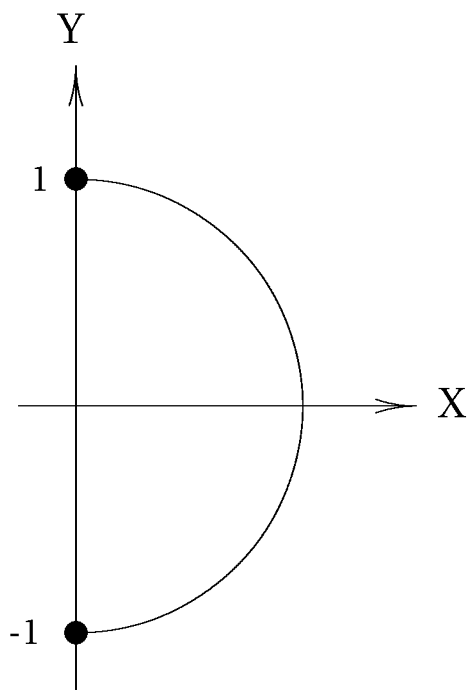


Figure 1.6: Half-Circle

3. $P(t) = (t, \sin t)$ on $[0, 4\pi]$ describes a sinusoidal curve.

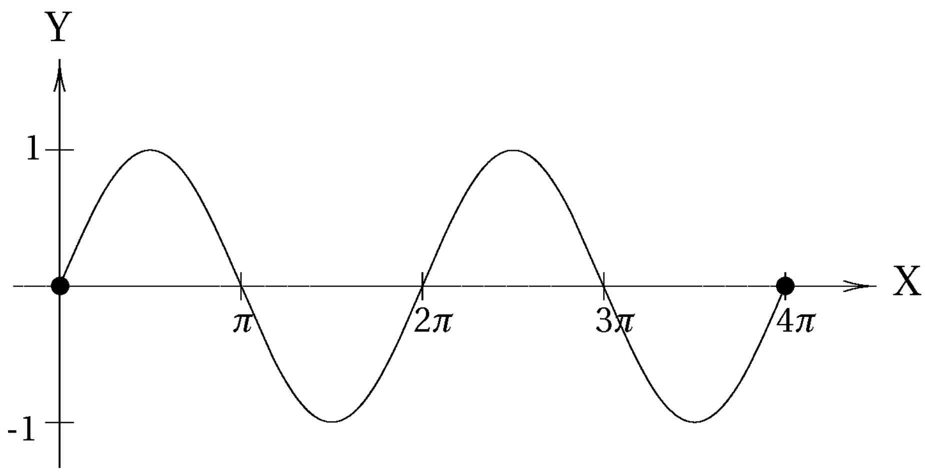


Figure 1.7: Sinusoidal Curve

1.5 Rules for Measuring Tortuosity

The measure of tortuosity should not depend on the orientation of the path. Hence, any measure of tortuosity should not be affected by the following:

1. Translation.
2. Rotating the path around a point.
3. Reflecting the path in a line.
4. Enlarging or shrinking the path (scaling).

Since the concept of tortuosity is often used to analyze images of shapes in various scales, we need to make sure that the measure does not change depending on the scale. Hence, the measure of tortuosity of a curve should not depend on the magnification of the device that obtained the image. It should be scale invariant.

1.6 The Information Contained in Curvature

1.6.1 Introduction

Mathematically, curvature of a path at a point is measured by fitting a circle that approximates the shape of the path the best. Therefore, the radius of the circle is a numerical invariant, and we define geometrical curvature as $\frac{1}{r} = \kappa$.

1.6.2 Definition

The path can be thought of as the trajectory of a moving particle in space over time as it travels on a plane from its starting position to its ending position. As the particle moves, it may be turning in various ways. Curvature is a measure of the size of a turn at any single point along the path.

Consider a path given by a parametric equation $P(t) = \big(x(t), y(t)\big)$ over the non-empty interval $[t_0, t_1]$. As t goes from t_0 to t_1, the particle at point $P(t)$ moves along the path. Since we assume $P(t)$ is twice differentiable, we have a line tangent to the path at any point $\big(x(t), y(t)\big)$. Curvature is the rate of change in the angle of this tangent line with respect to the change in the position of the point.

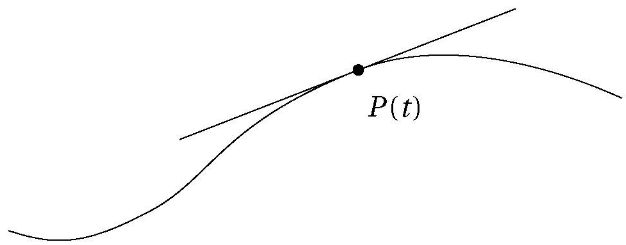


Figure 1.8: Tangent line at point $P(t)$

Let $\Delta\theta$ be the change in the angle, let Δs be the distance the point has moved, and let κ be curvature. Then, curvature is defined as [wil]

$$\kappa = \lim_{\Delta s \to 0} \frac{\Delta\theta}{\Delta s} \tag{1.1}$$

Observe that $\Delta\theta$ can be negative, so κ can be negative as well. We call it signed curvature κ. However, we usually consider $|\kappa|$, i.e. positive curvature only. The precise definition is given here.

Definition: Let $P(t) = \big(x(t), y(t)\big)$. Then, **signed curvature** is given by

$$\kappa(t) = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}} \tag{1.2}$$

and **curvature** by

$$|\kappa(t)| = \frac{|\dot{x}\ddot{y} - \dot{y}\ddot{x}|}{(\dot{x}^2 + \dot{y}^2)^{\frac{3}{2}}} \tag{1.3}$$

1.6.3 Magnitude of the Turn

Geometrically, curvature relates to the circle whose shape matches the path at a point. To illustrate, examine the parabola described by the equation $y = x^2$ at the point $(0,0)$. We have $\dot{x} = 1$, $\ddot{x} = 0$, $\dot{y} = 2t$, and $\ddot{y} = 2$. Hence

$$\kappa(t) = \frac{2}{(1+4t^2)^{\frac{3}{2}}} \tag{1.4}$$

For $t = 0$ we get 2. Hence, a circle with radius $\frac{1}{2}$ approximates the parabola at $(0,0)$.

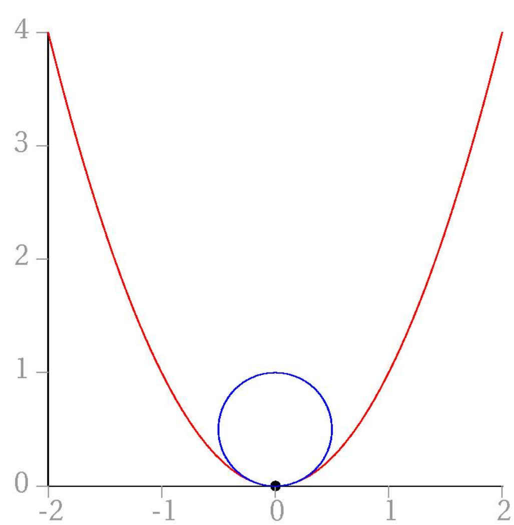


Figure 1.9: Parabola curvature at $(0,0)$

It turns out that the value of the curvature is equal to the inverse of the radius of this circle. [Rau08, p. 64]

By contrast, at the point $(2,4)$, the parabola has a curvature of $\kappa = 0.0285$, which means the inscribed circle is much larger and has a radius of 35.0464.

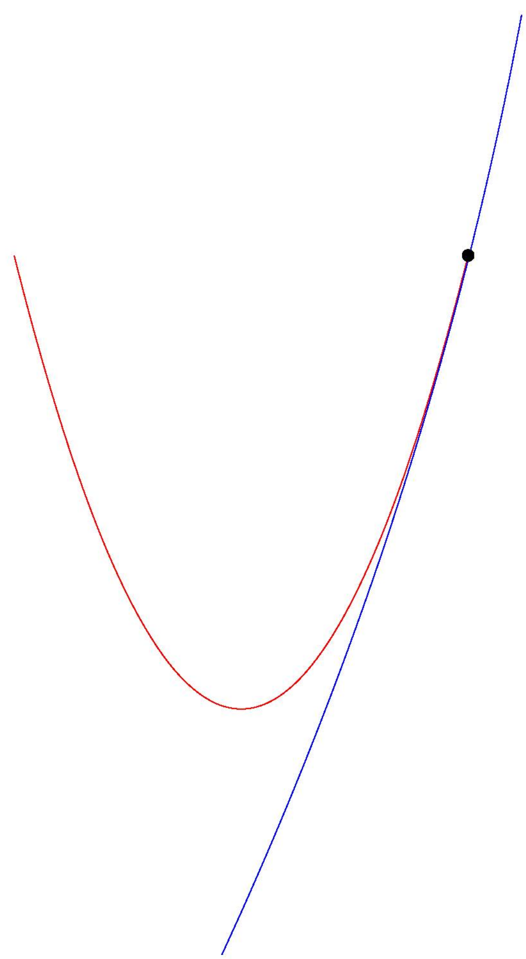


Figure 1.10: Parabola curvature at $(2,4)$

Points on a path with a small curvature approximates the curve by a large circle, i.e. it has a very gradual turn. A path with a large curvature has a very sharp turn. Here, a low $|\kappa|$ indicates a gradual turn while a high $|\kappa|$ indicates a sharp turn. Hence, curvature is an indication of how much the path turns, or the **magnitude** of the turn.

1.6.4 Direction of the Turn

But κ can be negative as well as positive. A positive value means the path is turning in a counter-clockwise direction and a negative value means the path is turning in a clockwise direction. Hence, curvature is also an indication of the **direction** in which the path turns. Here, direction is analogous to rotation: clockwise or counter-clockwise.

$\kappa > 0$	turns counter-clockwise	(turns to the left)
$\kappa < 0$	turns clockwise	(turns to the right)
$\kappa = 0$		(does not turn at all)

Table 1.1: Table of Curvature Directions

1.6.5 Two Kinds of Information

Therefore, curvature captures two aspects of how the path turns at any specific point. The **magnitude** of the turn as well as the **direction** of the turn. A parabola may have one sharp turn but never changes direction while a sine wave may never have sharp turns but changes direction frequently.

1.6.6 Curvature Along The Entire Path

Curvature is a measure of tortuosity of the path at a single point. To characterize the entire path, we examine curvature at every point along the length of the path and normalize the result to obtain a numerical invariant of the path.

To illustrate, we present examples of curvature along various paths. For simplicity, every path starts with a point on the left and ends at a point on the right at the same vertical position.

Example 1: Consider a semi-circle of radius $\frac{1}{2}$

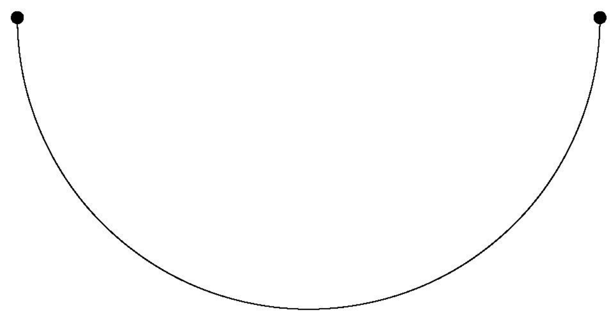


Figure 1.11: Semi-circle of radius $\frac{1}{2}$

The plot below is of the curvature function along the arc length of the path.

Arc length is on the X-axis and curvature is on the Y-axis.

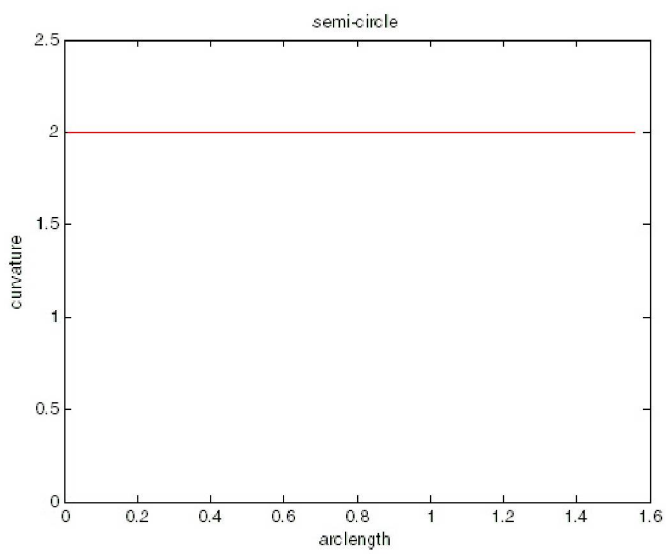


Figure 1.12: Curvature of Semi-circle

We conclude that circles have a constant curvature. The curvature is positive because the path is turning counter-clockwise.

Example 2: Consider a parabola

$$y = 2x(x-1) \text{ over } [0,1]$$

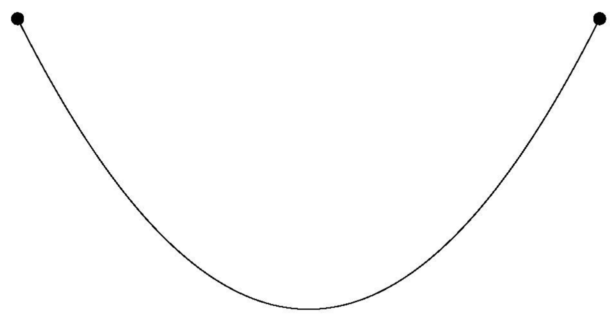


Figure 1.13: Parabola

Hence the curvature function:

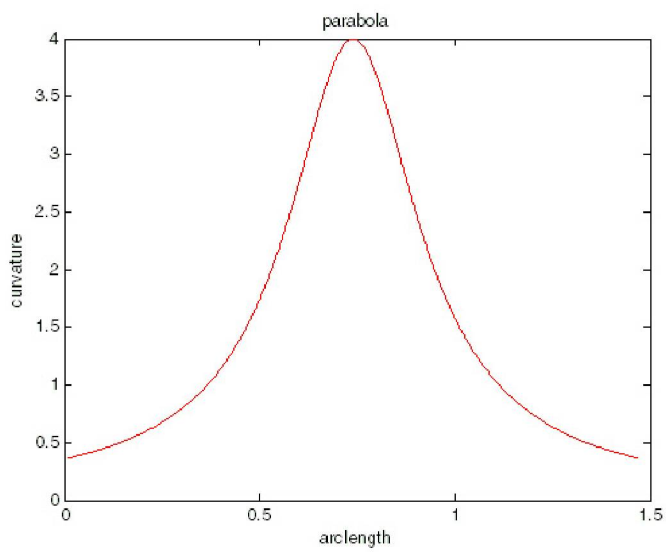


Figure 1.14: Curvature of Parabola

On a parabola, the curvature varies along the length of the path, but it is always positive because the path is always turning counter-clockwise.

Example 3: Consider a cubic polynomial

$$y = 5x(x - \tfrac{1}{2})(x - 1) \text{ over } [0, 1]$$

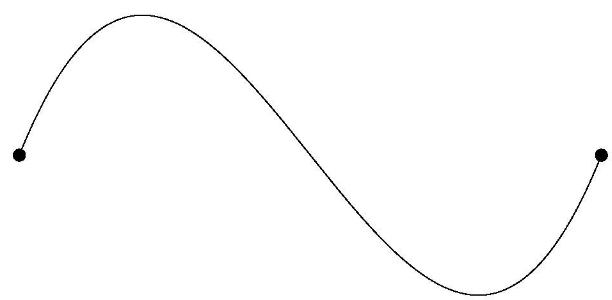


Figure 1.15: Cubic Polynomial

We have the following curvature function:

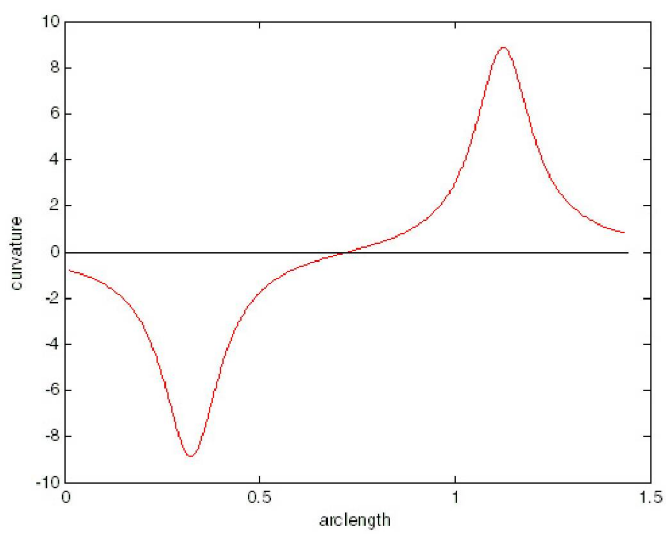


Figure 1.16: Curvature of Cubic Polynomial

The curvature is negative for the first part because the path is turning clockwise. Then, the path turns counter-clockwise and the rest of the curvature is positive.

To characterize the average curvature along the entire path, we calculate the average, which is the area under the curve divided by the length of the interval.

$$\frac{1}{b-a}\int_a^b \kappa(s)\,ds \tag{1.5}$$

The problem with this method is apparent in example 3. Curvature can be negative, so the area of the first part is negative which cancels out the positive area of the second part. The average curvature in example 3 is zero.

This problem is usually solved by examining the **absolute value** of the curvature along the arc length of the path

$$\frac{1}{b-a}\int_a^b |\kappa(s)|\,ds \tag{1.6}$$

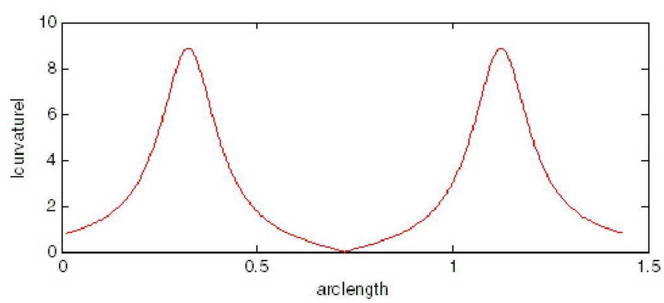


Figure 1.17: Absolute Value of Curvature of Cubic Polynomial

Now, the average $|\kappa|$ is an indication of the magnitude of turns along the path.

Observation: There is more than one way to characterize the average curvature along the length of the path. Mean and standard deviation are the two most commonly used calculations.

1.6.7 Second Aspect of Curvature Along The Entire Path

Since the average $|\kappa|$ ignores any changes in direction the path may have taken, it only captures one aspect of the path: the magnitude of the turns.

The other aspect of tortuosity is related to the number of times the path changes direction, which can be obtained by finding the inflection points of the path. Both the first and second examples have zero changes in direction. The third example has one inflection point, hence it changes direction once.

1.6.8 Two Types of Information Obtained from Curvature

Both aspects of curvature can be extracted from any given path. The magnitude of turns can be captured by examining the absolute value of curvature along the length of the path and changes in direction can be captured by the number of inflection points. To fully characterize a path, it seems that all available information should be utilized.

CHAPTER 2

METHODOLOGY

2.1 Measures of Tortuosity in Literature

In this chapter we discuss the algorithms that have been proposed to measure tortuosity in the past.

2.1.1 Arc Length

The most commonly used measure of tortuosity is also the easiest to understand. The arc length of the path.

The path with the shortest length between two points is a straight line, sometimes called the **chord length**. Any other path must be longer. The more a path deviates from a straight line, the longer its arc length.

Most applications that use arc length measure tortousity as the ratio between the arc length of the path (L_c) and the chord length (L_χ).[SHN+93]

Let's call this "arc ratio". Often, this ratio is reduced by one so a straight line will have a tortuosity of zero. [HGC+99]

$$\frac{L_c}{L_\chi} - 1 \tag{2.1}$$

However, arc ratio completely ignores the shape of the path.

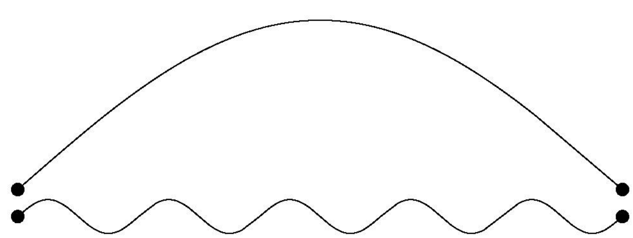


Figure 2.1: Two Paths with the Same Arc Ratio

Both of these paths have the same arc ratio.

While arc ratio is a good measure of how much a path deviates from a straight line, it doesn't capture any other information about the path.

2.1.2 Total Curvature

Because curvature varies along the length of the path, another commonly used measure involves "the use of the integral of the absolute curvature (tc) or the squared curvature (tsc)."[GFR08]

$$tc = \int |\kappa(s)|\, ds \tag{2.2}$$

$$tsc = \int \kappa(s)^2, ds \tag{2.3}$$

where s is the distance along the path.

Interestingly, both tc and tsc are used in the calculation of the mean (μ) and standard deviation (σ) of the absolute curvature $|\kappa|$.

$$\mu = \frac{tc}{arc\ length} \tag{2.4}$$

$$\sigma = \frac{tsc}{arc\ length} - \mu^2 \tag{2.5}$$

However, these measures ignore any changes in direction of the path.

2.1.3 Inflection Points

As demonstrated above (see §1.6.7) these measures ignore changes in the direction of the path. One 1993 paper, [SHN+93], simply counts the number of inflection points as an invariant of tortuosity. Of course this completely ignores the magnitudes of the turns along the path, hence many different curves have the same measurement.

Other researchers have proposed methods that combine different information into a single measure. A 2003 paper, [BGP+03], multiplies inflection points by the arc ratio into what they call an "Inflection Count Metric".

$$ICM = (IPs + 1)\left(\frac{arc\ length}{chord\ length}\right) \tag{2.6}$$

where IPs is the number of inflection points.

While ICM incorporates both number of inflection points and arc ratio by multiplying them together, this has the effect of losing information about the influence of each factor. Given some ICM value, it is impossible to tell if the value was dominated by the number of inflection points or the arc ratio. For example, the following two paths have the same ICM value.

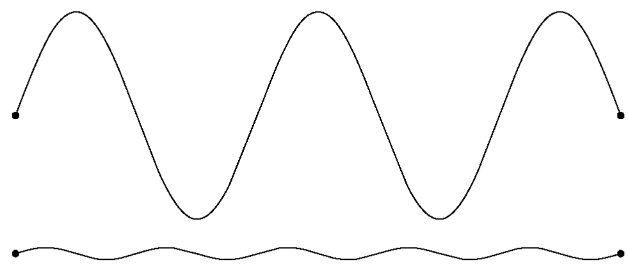


Figure 2.2: Two Paths with Same ICM Value

2.1.4 More Complicated Methods

Several methods have been proposed combining several measurements together. They also have the same limitation as ICM because they multiply all the factors together into a single value.

We examine three methods from the literature. They begin by breaking the path at every inflection point into a sequence of single arcs such that each arc has curvature with a constant sign.

$$[\kappa(s) \geq 0, \forall s \in D_i] \vee [\kappa(s) \leq 0, \forall s \in D_i] \tag{2.7}$$

where $D_i, i = 1, \dots, n$ refers to the set of arcs in the path. [GFR08]

Here are three methods proposed by [GFR08], [TOUC13], and [Jos12].

$$\left(\frac{n-1}{n}\right)\left(\frac{1}{L_c}\right)\sum_{i=1}^{n}\left[\frac{L_{c_i}}{L_{\chi_i}}-1\right] \tag{2.8}$$

$$\left(\frac{n-1}{n}\right)\left(\frac{1}{L_c}\right)\sum \kappa_i \tag{2.9}$$

$$\left(\frac{n}{L_c}\right)\left(\frac{\sum \theta_i}{m}\right)\left(\frac{\sum \frac{L_{c_i}}{L_{\chi_i}}}{m}\right) \tag{2.10}$$

n is the number of arcs with curvature of a constant sign.

L_c and L_χ are the arc length and chord length of the path.

L_{c_i} and L_{χ_i} are the arc length and chord length of a single arc within the path.

m is the number of segments the path is divided into.

θ_i is the change in angle with each segment.

The first equation (2.8) involves calculating the average arc ratio, the second (2.9) involves calculating the average curvature of the path, and the third involves calculating both since $\sum \theta_i$ is just another way to calculate curvature (see §1.6.2). They all use measures described above, but they are all multiplied together, which has the effect of losing information about the tortuosity of the path.

The first (2.8) and second (2.9) methods both use the coefficient $\frac{n-1}{n}$ which approaches one as the number single arcs increases, but it collapses to zero when the path consists of one arc, hence losing all information about that path.

2.1.5 Sum of Angles

Paths are usually described by a sequence of points, which in turn describe a chain of line segments. Each adjacent pair of segments form an angle.

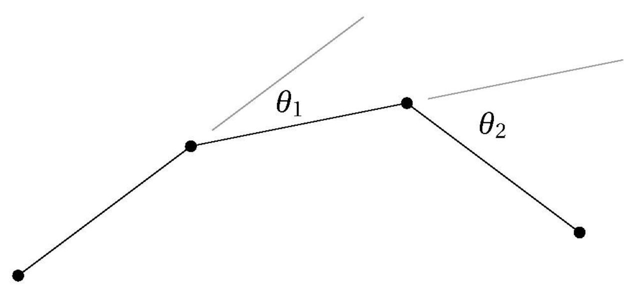


Figure 2.3: Angles between Path Segments

Another method of calculating tortuosity that appears in the literature, as mentioned above in equation (2.10), is the sum of these angles. This is actually a discrete way of calculating total curvature. From the definition of curvature (see equation 1.1 in §1.6.2), the measure of curvature at each point in the chain can be approximated by θ_i. The total curvature is simply the sum of these angles.

2.1.6 Conclusion

All of the methods described above either ignore information about the path or lose information by multiplying various factors together. Hence, distinguishing certain paths from each other using these measurements may be impossible.

2.2 Proposed Measurement of Tortuosity

Our goal is to measure tortuosity more efficiently than it was done in the past. It has been demonstrated that there are several kinds of information that describe the shape of a path. All the methods proposed so far have either ignored some information, or combined different combinations of this information into a single value, losing information in the process.

It seems that the best approach is to keep the various kinds of information separate to control the differences between the paths. For example, in statistics when mean and standard deviation are used to characterize a data set, they are kept separate. They are not combined into a single value as this obscures the information about the underlying data set.

The tortuosity of a path is also a way to characterize the path. Two similar paths should have similar measures.

The proposed technique is to treat tortuosity as a problem of classification. To sort paths into classes based on a small set of independent features. [Dou12, p. 3] The features would be the various kinds of information that can be extracted from paths. Then, the measure, instead of being a single value along the real number line, becomes a location in n-dimensional space where n is the number of features. Paths are characterized, or “measured”, by the position they hold in this “feature space.”

2.2.1 Selection of Features

The first step in defining this classifying space is to decide upon a set of features. This set should be small but still capture the available information. The technique is to generate several features, analyze them, and remove features that are redundant. If two features are highly correlated, then one can be eliminated without appreciable loss of information. The result is a small set of the most significant features that will be used to characterize paths.

2.2.2 Test Suite of Paths

In order to evaluate features, a test suite of paths was generated. Each path $P(t)$ was formed from two sine waves. The second wave was added to add some irregularity. The function used was

$$P(t) = \left(t, M_1 \sin\left(t\frac{2\pi}{T_1}\right) + M_2 \sin\left(t\frac{2\pi}{T_2}\right)\right) \tag{2.11}$$

where t ranged from zero to 10.

To ensure a good distribution of data, the number of inflection points was controlled so that each path had a fixed number of inflection points. This number varied from zero up to and including 32. This was controlled by the T_1 parameter. For each number of inflection points, 10 paths were generated by randomly assigning values to the other parameters. This resulted in a test suite of 330 paths with a range of inflection points.

	Range of values		Range of values
T_1	0.6 to 20	M_1	0.1 to 0.3
T_2	10 to 20	M_2	0.0 to 0.3

Table 2.1: Table of Parameters

2.2.3 Available Features

Based on the discussion above, and a review of the literature, the following seven features were extracted from each path $P(t)$ (with fixed parameters):

TC	total curvature	$\int \lvert\kappa\rvert$
TSC	total squared curvature	$\int \kappa^2$
ARCLEN	arc length of entire path	
RATIO	arc ratio	$\frac{arc\ length}{chord\ length} - 1$
MEAN	mean positive curvature	$\frac{TC}{ARCLEN}$
STDEV	standard deviation of positive curvature	$\frac{TSC}{ARCLEN} - MEAN^2$
IPS	number of inflection points	

Table 2.2: Table of Features

These features were calculated from the entire test suite of 330 paths, using the parameters as described above, and were imported into JMP [jmp] to produce the correlations and to graph a scatter plot between each pair of features.

Correlations

	TC	TSC	ARCLEN	RATIO	MEAN	STDEV	IPS
TSC	0.9420	1.0000	0.9459	0.9459	0.8370	0.9476	0.8049
ARCLEN	0.9502	0.9459	1.0000	1.0000	0.8618	0.9669	0.8012
RATIO	0.9503	0.9459	1.0000	1.0000	0.8618	0.9669	0.8012
MEAN	0.9692	0.8370	0.8618	0.8618	1.0000	0.9538	0.9876
STDEV	0.9959	0.9476	0.9669	0.9669	0.9538	1.0000	0.9169
IPS	0.9413	0.8049	0.8012	0.8012	0.9876	0.9169	1.0000

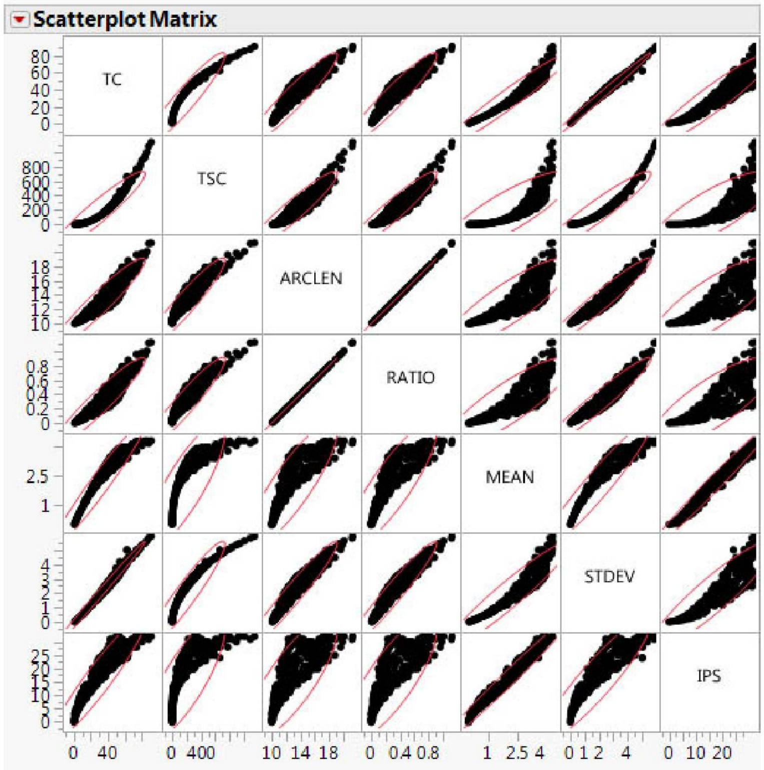


Figure 2.4: Scatterplots

The numbers on the scales of the scatter plot reflect the values that were generated. There are several highly correlated features. ARCLEN and RATIO have a direct correlation. While not linear, the plot of TC versus TSC shows a distinct relationship. And TC has a large correlation with both MEAN and STDEV.

The question is which ones to eliminate.

The answer comes from an earlier discussion (see §1.5). Features should be scale invariant.

To determine the scale invariance of each feature, a small set of paths were generated. For each path $P(t)$ the following function was used:

$$P(t) = \left(t, M_1 \sin\left(t\frac{2\pi}{T_1}\right)\right) \tag{2.12}$$

where T_1 varied from 0.6 to 20 in order to generate paths with inflections points ranging from zero up to and including 32. M_1 was set to one of the three values: 0.1, 0.2, and 0.3, while t ranged from zero to 10. Each path was scaled by a factor of 1,3,5,7, and 9. Then, for each path and each of its scaled paths, the features were calculated and the standard deviation was calculated for each feature. The features with the lowest standard deviation would have the smallest scale invariance.

The following figure shows data from two of the paths to illustrate the results of the entire set. All other paths had similar results. The standard deviations are outlined in bold in the last row of each set of data. The first set shows the data generated by $T_1 = 0.606061$, which results in 32 inflection points, and the second set shows the data generated by $T_1 = 2.22222$, which results in 8 inflection points. Both sets used $M_1 = 0.1$.

SCALE	TC	TSC	ARCLEN	RATIO	MEAN	STDEV	IPS
1	41.8387	212.967	12.1255	0.212552	3.45047	2.37861	32
3	41.8387	70.989	36.3766	0.212552	1.15016	0.79287	32
5	41.8387	42.5934	60.6276	0.212552	0.690094	0.475722	32
7	41.8387	30.4238	84.8787	0.212552	0.492924	0.339802	32
9	41.8387	23.663	109.13	0.212552	0.383386	0.26429	32
	0	70.3098	34.29625	0	1.139152	0.785284	0
1	4.84471	2.89393	10.1957	0.019567	0.475173	0.240934	8
3	4.84471	0.964642	30.587	0.019567	0.158391	0.080311	8
5	4.84471	0.578785	50.9783	0.019567	0.095035	0.048187	8
7	4.84471	0.413418	71.3697	0.019567	0.067882	0.034419	8
9	4.84471	0.321547	91.761	0.019567	0.052797	0.02677	8
	0	0.955414	28.8377	0	0.156876	0.079543	0

Figure 2.5: Results of Two Paths

TC, RATIO, and IPS have a standard deviation of zero, meaning they are also scale invariant.

In the test suite of data, TC, RATIO, and IPS have high correlations, but when the same analysis was performed on two sets of retinal images, this correlation disappears.

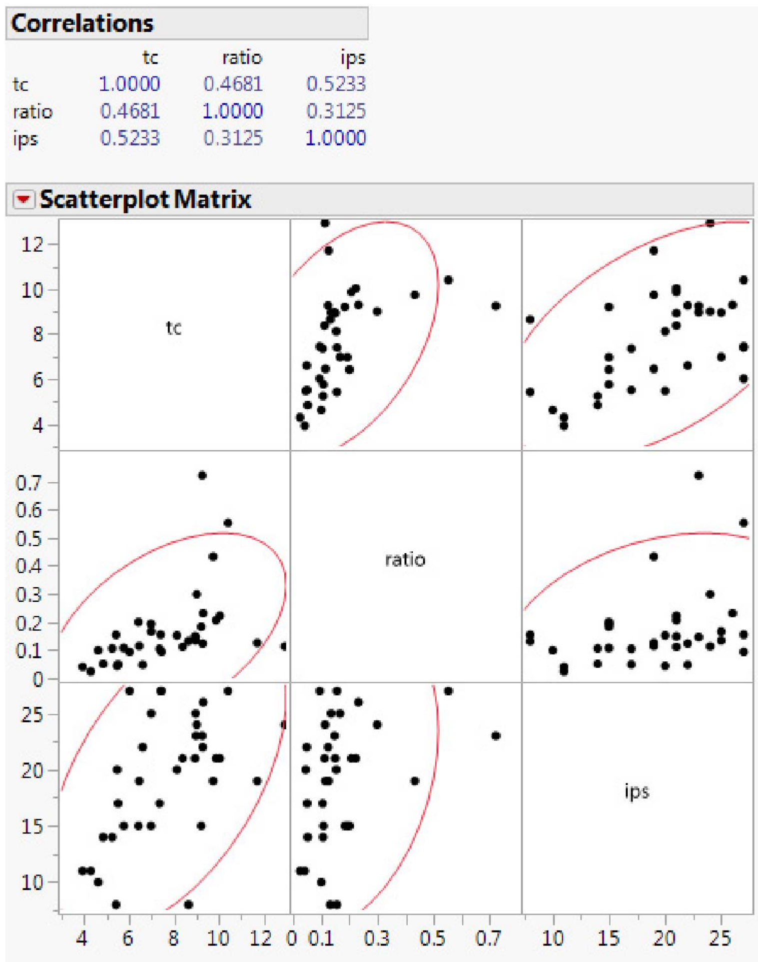


Figure 2.6: Scatterplots of Retinal Data

This plot is from the STructured Analysis of the Retina data set [sta]

Correlations

	tc	ratio	ips
tc	1.0000	0.2850	0.2914
ratio	0.2850	1.0000	0.5884
ips	0.2914	0.5884	1.0000

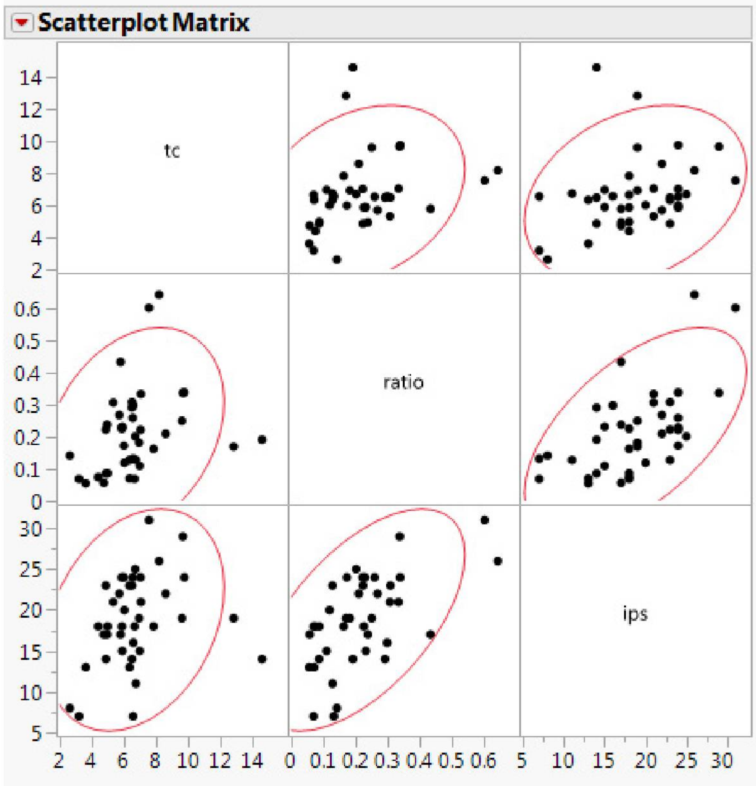


Figure 2.7: More Scatterplots of Retinal Data

This plot is from the Digital Retinal Images for Vessel Extraction data set [dri]

Hence we conclude that there is no reason to eliminate any more features.

Therefore, TC, RATIO, and IPS are the best features to use.

2.2.4 Selected Features

TC, RATIO, and IPS are the set of features selected for analysis. There are several advantages to this selection.

They capture all of the information available from a path. TC captures information about the magnitudes of the turns in a path, IPS captures the number of changes in direction, and RATIO captures the total length of the path, which is a very intuitive measure. As stated above (see §1.1) tortuosity is a measure of how much a path deviates from a straight line. Arc length captures this nicely, and RATIO is calculated from arc length.

They are all scale invariant. As stated above (see §1.5), a measure of tortuosity should not depend on the scale of the data. All three of these features remain the same as the scale of the path changes.

There are only three of them. This means it is possible to plot out the data onto a graph and examine it visually. This is an enormous advantage over having a larger set of features. To illustrate, TC, MEAN, and IPS from the test suite of data are plotted on a 3D graph for visual examination.

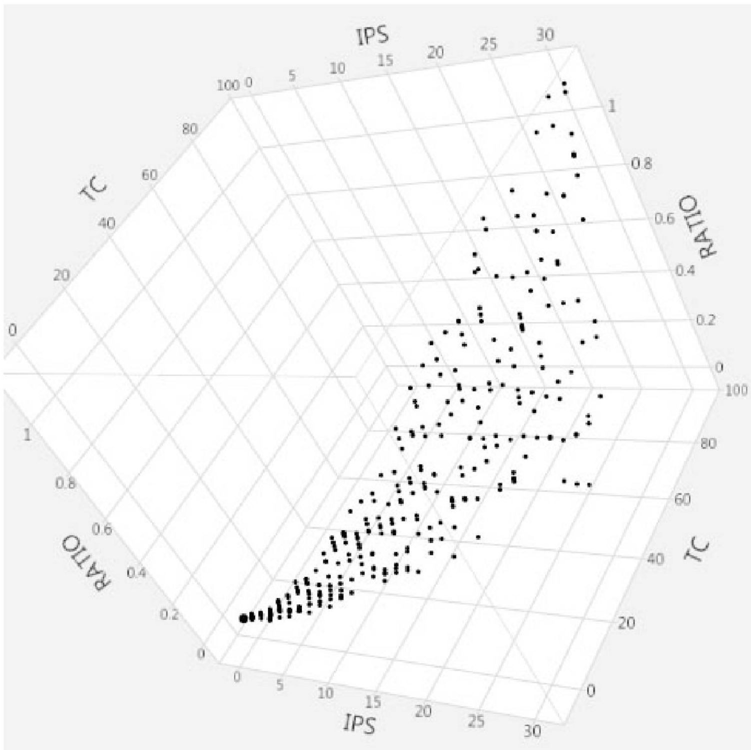


Figure 2.8: Scatterplot of Feature Data

The resulting graph, when rotated in 3D to display the maximum variance, shows a sail-like shape where the point at the bottom is the point corresponding to a straight line. The points are spread out fairly evenly so that every path is distinct from every other path. This eliminates the ambiguity of previous measures where dissimilar paths had the same measured value. To illustrate how the results are distributed throughout this graph, seven different paths are identified and shown on the next graph.

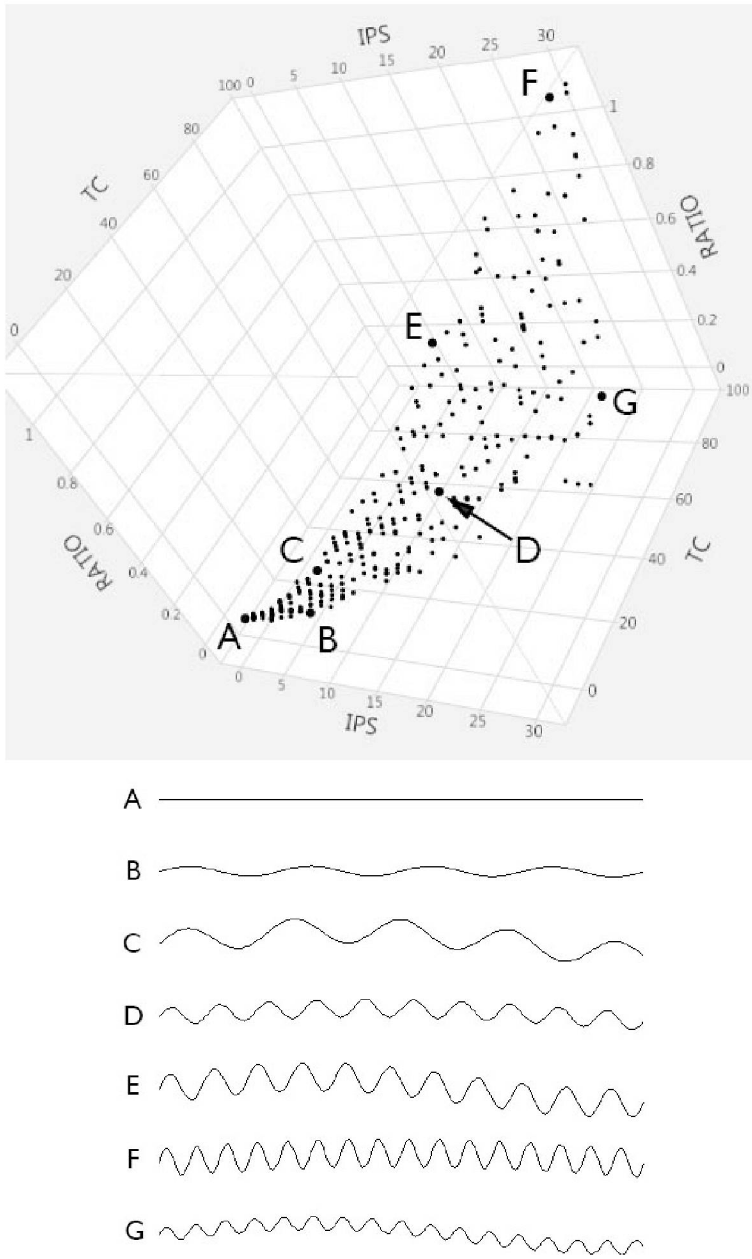


Figure 2.9: Scatterplot of Feature Data with Labeled Points

2.2.5 Tortuosity in Feature Space

All elements of the proposed technique are now in place. Tortuosity of a path is measured, not by one number, but by three: the total positive curvature, the ratio of arc length to chord length, and the number of inflections points. This set of measures forms a location in the feature space for paths. The "location" in feature space is used to evaluate and classify the path.

It is up to the specific application to decide how to characterize, or classify, the path. This technique is meant to work for any application. For example, some applications might want to classify paths as being either "tortuous" or "non-tortuous" [GFR08, p. 310]. Such applications would have to determine what regions of feature space constitute one class or the other.

2.2.6 Previous Work

We want to comment that this is not the first time a set of features has been proposed as a measure of tortuosity. A 2013 paper, [TCU13], examined a set of features and concluded the best results came from a combination of two, (see equations (2.9) and (2.6)).

$$\tau = \left(\frac{n-1}{n}\right)\left(\frac{1}{L_c}\right)\sum \kappa_i \tag{2.13}$$

$$ICM = (IPs + 1)\left(\frac{arc\ length}{chord\ length}\right) \tag{2.14}$$

Note that ICM makes use of the ratio of arc length to chord length and the number of inflection points while τ makes use of total positive curvature, the three features selected for use in this paper. The difference is that they multiply the features together while this paper keeps them separate, obtaining a larger classifying space that can differentiate better curves of different tortuosity.

CHAPTER 3

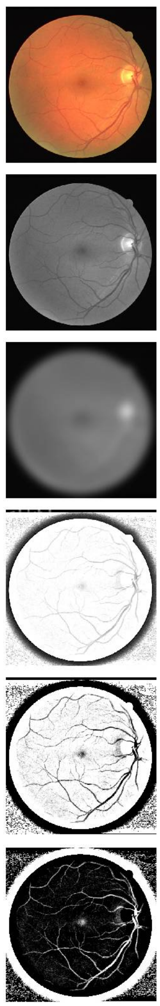
Retinal Data and Results

3.1 Retinal Data

To illustrate how this technique can be used, a small set of retinal scans was analyzed.

Each retinal image was preprocessed using Adobe Photoshop CS6 and converted into a black and white image that the software could read.

The following steps were used in the conversion of images.



Original image

Green color extracted and converted to grayscale

Image duplicated and gaussian blur with radius 15 pixels

Greyscale and blurred images divided and layers flattened

Greyscale range changed from 0-255 to 200-255

Image inverted

Figure 3.1: Preprocessing of Retinal Images

The path of a major vessel from each black and white image was extracted using the image processing program ImageJ [ima] with the plugin software package NeuronJ [neu].

A test suite of 35 images was selected from the STructured Analysis of the Retina website [sta]. The retinal images were labeled with various diagnoses. Finally, for the path extracted from each image, the three features were calculated and the results plotted in 3D.

On the following graphs, two groups of retinals are highlighted based on their diagnoses.

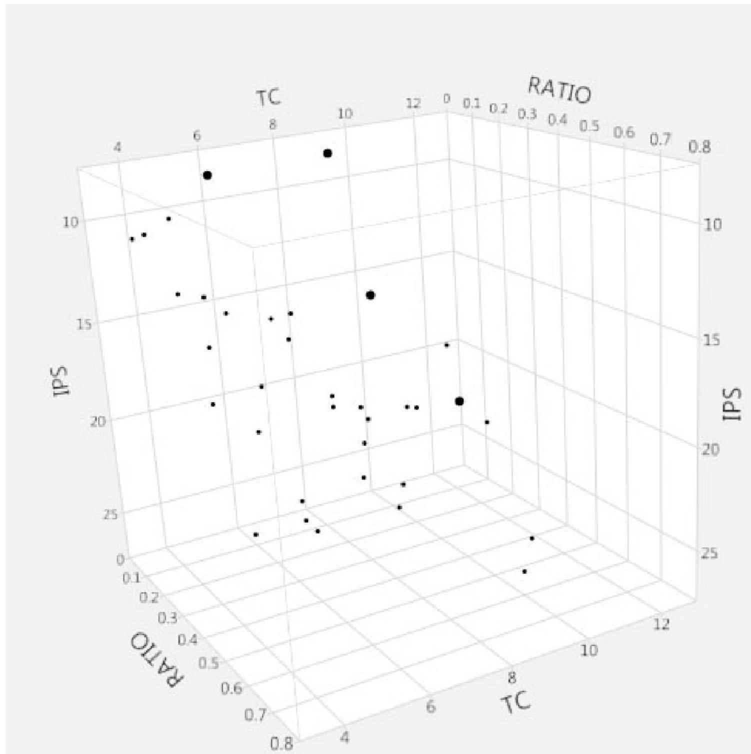


Figure 3.2: Central Retinal Vein Occlusion

This plot shows the retinals diagnosed with Central Retinal Vein Occlusion.

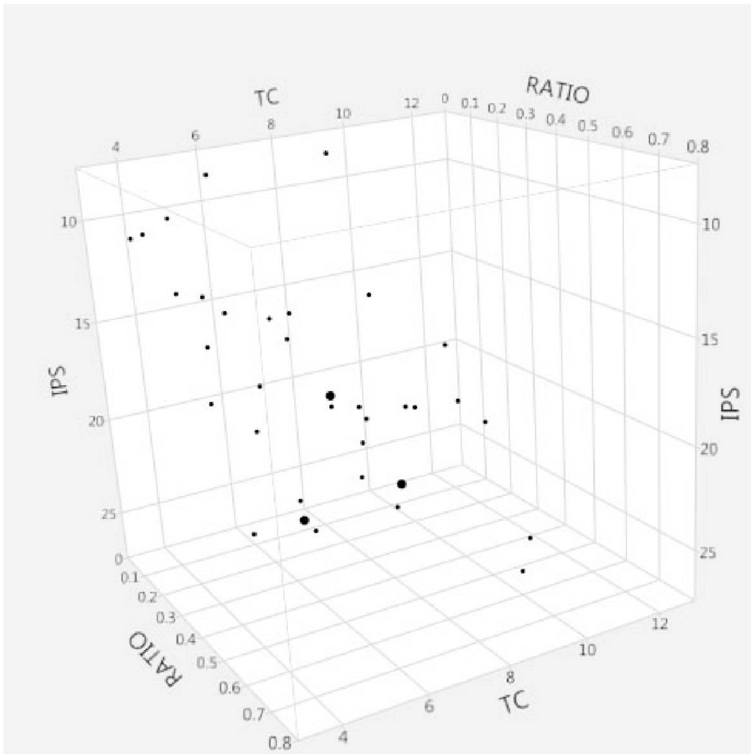


Figure 3.3: Choroidal Neovascularization

This plot shows the retinals diagnosed with Choroidal Neovascularization.

In each graph, the retinals with a particular diagnosis are gathered together in one area of the feature space. The retinals with Central Retinal Vein Occlusion are in a different area from the retinals with Choroidal Neovascularization.

If an application wanted to identify retinals with Central Retinal Vein Occlusion or Choroidal Neovascularization, it could use this technique to see what retinals fall in the two regions identified above.

While this used a very small set of data, it does illustrate how this technique can be used to classify retinal images.

CHAPTER 4

Discussion and Conclusion

4.1 Future Work

4.1.1 Extraction of data

One issue that emerged during this research was how the data for paths was obtained. It turns out that the calculation of curvature from a set of points is very sensitive to noise. [JD07]

In digitial images, positions are restricted to the discrete locations of pixels, introducing noise into the data. Noise can also come from rounding errors when numerical data are recorded. The problem is one common to signal processing: how to extract the signal from noisy data. [Orf95, p. 382]

One technique is to smooth the data to eliminate the noise but not remove significant movement of the path. There are a variety of ways to smooth data. Some attempts were made using moving average but the results were unsatisfactory. Another technique is to fit a polynomial spline through the raw data. [JD07] Here, a spline is fit through a sequence of circles, each centered on a data point, all with some radius. The radius determines the smoothness of the resulting spline. In both moving average and splines, it was unclear how much smoothing was required and when too much information was removed about the path. How to remove noise remains an open question.

The analysis performed for this paper did not remove any noise from the data.

4.1.2 Principal Component Analysis

There are various techniques that can enhance the distribution of the data. Principal Component Analysis (PCA) can translate the features onto a set of axes that maximizes the variation in the data. The test suite of data was subjected to PCA and the resulting three components were plotted on a 3D graph.

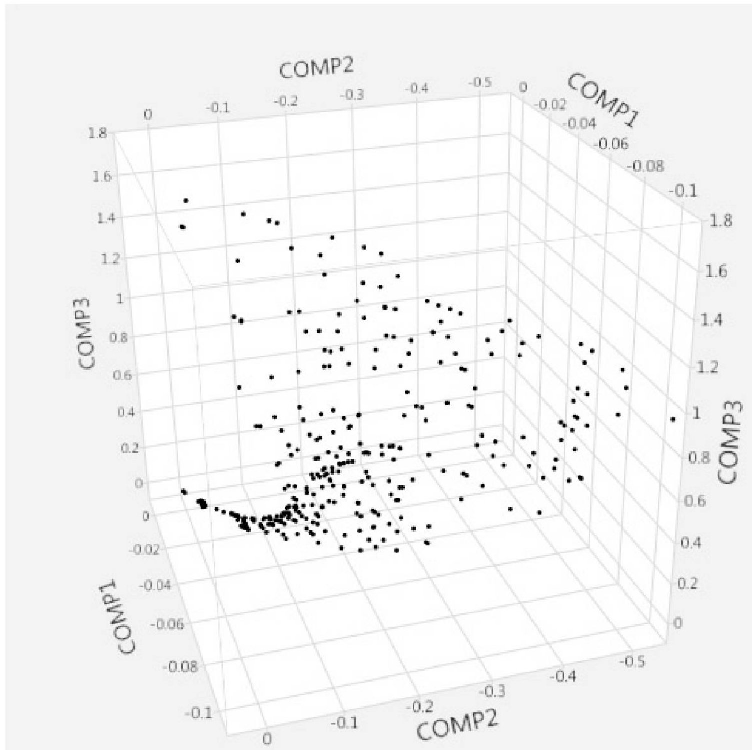


Figure 4.1: Principal Component Analysis

This is a graph of the test suite of data showing the results after PCA. The result is also a fairly even spread of the data points. It remains to be seen if this enhances the analysis of the data.

While it can't be seen in the above graph, the data points form a curved surface, something like a saddle curve. Another open question is whether a nonlinear analysis would give better results.

4.2 Conclusion

Measuring tortuosity with a single value is not a good way to characterize a path. Too much information is lost and dissimilar paths can have the same measured value.

A much better approach is to use more than one value. This set of features captures more of the information that can be extracted from a path and can be used to more accurately characterize it.

This paper proposes using a set of three measures. The total positive curvature, the ratio of arc length to chord length, and the number of inflection points. They are independent of each other and they capture most of the available information. They are also scale invariant, which is a valuable property. In most applications, the data is extracted from digital images. Being scale invariant, the feature measurements do not depend on the size of the images, nor do they depend on the magnification of the camera.

This approach is also general enough for use by any application.

List of Figures

LIST OF TABLES

BIBLIOGRAPHY

[BGP+03] Elizabeth Bullitt, Guido Gerig, Stephen M Pizer, Weili Lin, and Stephen R Aylward. Measuring tortuosity of the intracerebral vasculature from mra images. *IEEE transactions on medical imaging*, 22(9):1163–1171, 2003.

[Dou12] Geoff Dougherty. *Pattern recognition and classification: an introduction*. Springer Science & Business Media, 2012.

[dri] Digital retinal images for vessel extraction. https://computervisiononline.com/dataset/1105138662. Accessed: 2017-11-27.

[GFR08] Enrico Grisan, Marco Foracchia, and Alfredo Ruggeri. A novel method for the automatic grading of retinal vessel tortuosity. *IEEE transactions on medical imaging*, 27(3):310–319, Mar 2008.

[HGC+99] William E Hart, Michael Goldbaum, Brad Côté, Paul Kube, and Mark R Nelson. Measurement and classification of retinal vascular tortuosity. *International journal of medical informatics*, 53(2):239–252, 1999.

[ima] Imagej image processing and analysis in java. https://imagej.nih.gov/ij/. Accessed: 2017-11-25.

[JD07] Michael J Johnson and Geoff Dougherty. Robust measures of three-dimensional vascular tortuosity based on the minimum curvature of approximating polynomial spline fits to the vessel midline. *Medical engineering & physics*, 29(6):677–690, 2007.

[jmp] Statistical discovery from sas. https://www.jmp.com/en_us/home.html. Accessed: 2017-11-25.

[Jos12] Vinayak Shivkumar Joshi. Analysis of retinal vessel networks using quantitative descriptors of vascular morphology. 2012.

[neu] Neuronj an imagej plugin for neurite tracing and analysis. https://imagescience.org/meijering/software/neuronj/. Accessed: 2017-11-25.

[Orf95] Sophocles J Orfanidis. *Introduction to signal processing*. Prentice-Hall, Inc., 1995.

[PBG+81] Ph.D. Philip Babcock Gove et al., editors. *Webster's third new dictionary of the English language unabridged*. Merriam-Webster Inc., 1981.

[Rau08] Martin Raussen. Elementary differential geometry: Curves and surfaces. *Department of mathematical sciences*, 2008.

[riv] By oliver kurmis, cc by-sa 2.0 de. `https://commons.wikimedia.org/w/index.php?curid=241148`. Accessed: 2017-03-12.

[SHN+93] Örjan Smedby, Nils Högman, Sven Nilsson, Uno Erikson, Anders G Olsson, and Goran Walldius. Two-dimensional tortuosity of the superficial femoral artery in early atherosclerosis. *Journal of vascular research*, 30(4):181–191, 1993.

[sta] Structured analysis of the retina. `http://cecas.clemson.edu/~ahoover/stare/`. Accessed: 2017-11-25.

[TCU13] Rashmi Turior, Pornthep Chutinantvarodom, and Bunyarit Uyyanonvara. Automatic tortuosity classification using machine learning approach. In *Applied Mechanics and Materials*, volume 241, pages 3143–3147. Trans Tech Publ, 2013.

[TOUC13] Rashmi Turior, Danu Onkaew, Bunyarit Uyyanonvara, and Pornthep Chutinantvarodom. Quantification and classification of retinal vessel tortuosity. *Science Asia*, 39:265–277, 2013.

[wil] The formula for curvature. https://www.ima.umn.edu/~miller/1372curvature.pdf. Accessed: 2017-11-27.