

Two-Dimensional Plane for Multi-Scale Quantification of Corneal Subbasal Nerve Tortuosity

Roberto Annunziata,¹ Ahmad Kheirkhah,² Shruti Aggarwal,² Bernardo M. Cavalcanti,² Pedram Hamrah,^{2,3} and Emanuele Trucco¹

¹Computer Vision and Image Processing Group, School of Science and Engineering (Computing), University of Dundee, Dundee, United Kingdom

²Ocular Surface Imaging Center and Cornea and Refractive Surgery Service, Massachusetts Eye and Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts, United States

³Boston Image Reading Center and Cornea Service, New England Eye Center, Tufts Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States

Correspondence: Emanuele Trucco, Computer Vision and Image Processing Group, School of Science and Engineering (Computing), University of Dundee, Perth Road, Dundee, United Kingdom; e.trucco@dundee.ac.uk.

Submitted: October 27, 2015

Accepted: December 17, 2015

Citation: Annunziata R, Kheirkhah A, Aggarwal S, Cavalcanti BM, Hamra P, Trucco E. Two-dimensional plane for multi-scale quantification of corneal subbasal nerve tortuosity. *Invest Ophthalmol Vis Sci.* 2016;57:XXX-XXX. DOI:10.1167/iovs.15-18513

PURPOSE. To assess the performance of a novel system for automated tortuosity estimation and interpretation. ?1

METHODS. A supervised strategy (driven by observers' grading) was employed to automatically identify the combination of tortuosity measures (i.e., tortuosity representation) leading to the best agreement with the observers. We investigated 18 tortuosity measures including curvature and density of inflection points, computed at multiple spatial scales. To leverage tortuosity interpretation, we propose the tortuosity plane (TP) onto which each image is mapped. Experiments were carried out on 140 images of subbasal nerve plexus of the central cornea, covering four levels of tortuosity. Three experienced observers graded each image independently. ?2 ?3 ?4

RESULTS. The best tortuosity representation was the combination of mean curvature at spatial scales 2 and 5. These tortuosity measures were the axes of the proposed TP (*interpretation*). The system for tortuosity *estimation* revealed strong agreement with the observers on a *global* and *per-level* basis. The agreement with each observer (Spearman's correlation) was statistically significant ($\alpha_s = 0.05$, $P < 0.0001$) and higher than that of at least one of the other observers in two out of three cases ($\rho_{OUR} = 0.7594$ vs. $\rho_{Obs3} = 0.7225$; $\rho_{OUR} = 0.8880$ vs. $\rho_{Obs1} = 0.8017$, $\rho_{Obs3} = 0.7315$). Based on paired-sample *t*-tests, these improvements were significant ($P < 0.001$).

CONCLUSIONS. Our automated system stratifies images by four tortuosity levels (discrete scale) matching or exceeding the accuracy of experienced observers. Of importance, the TP allows the assessment of tortuosity on a two-dimensional continuous scale, thus leading to a finer discrimination among images. ?5?6

Keywords: tortuosity plane, confocal microscopy, corneal subbasal nerve plexus

Changes in corneal nerves have been observed in various conditions such as ocular allergies and dry eye disease, neurotrophic keratopathy, infectious keratitis, contact lens wear, keratorefractive surgeries, corneal transplantation, and cataract surgery.¹⁻⁷ For clinical evaluation of the corneal nerve structures in patients, *in vivo* confocal microscopy (IVCM) is often used, which visualizes corneal nerves and in particular the subbasal nerve plexus.^{6,8,9} In addition to measuring the corneal nerve density, this imaging modality enables the assessment of morphologic features of the nerves such as width, reflectivity, orientation, branching patterns, bead-like formations and tortuosity.^{10,11}

Increased tortuosity, in particular, has been associated with various conditions, such as dry eye disease, Graves' ophthalmopathy, corneal infections, small-fiber neuropathy, and diabetic retinopathy.¹²⁻¹⁹ However, tortuosity has no widely accepted definition, and it is often estimated qualitatively by comparing patient images with reference images.¹¹ Moreover, several corneal nerve fibers with different tortuosity character-

istics (e.g., amplitude, number of inflection points) might be present in the same image. This inevitably introduces a degree of subjectivity; therefore, quantitative, objective and repeatable methods of tortuosity estimation would advance the field.

Several quantitative measures of tortuosity have been proposed for the retinal vasculature as observed in fundus camera images,²⁰⁻²⁷ and for the brain vasculature²⁸ (see Ref. 29 for a recent comparative study). In contrast, there is very limited published work on corneal nerve fibers; Kallinikos et al.³⁰ were the first to propose an objective, semiautomated method for quantifying the tortuosity of corneal nerve fibers. Scarpa et al.³¹ adapted the algorithm proposed by Grisan et al.²¹ for retinal vessels to work with corneal nerve fibers.

Previously reported approaches have the following limitations, which we address herein: (1) they are single-scale, as they do not differentiate between high- and low-frequency turns often characterizing corneal nerve fibers (Fig. 1); (2) they are based on a single tortuosity index (or coefficient) in which the combination and weight of each tortuosity measure (e.g.,



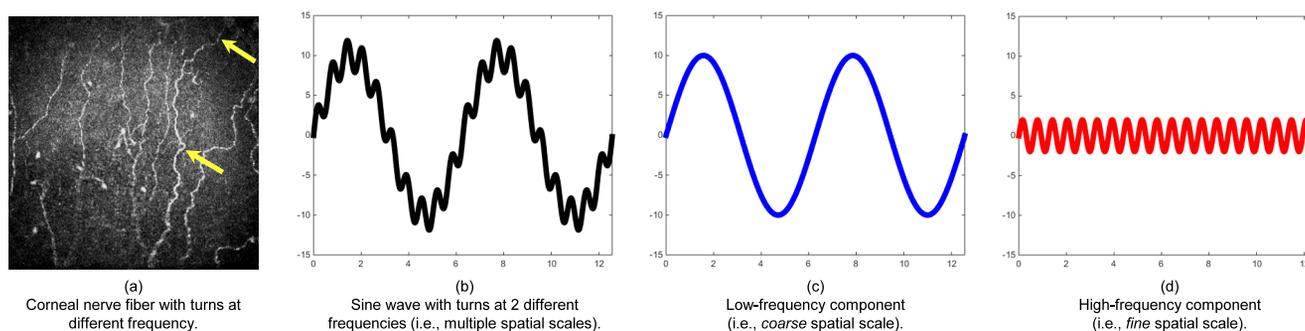


FIGURE 1. Example of multiple spatial scales observed in a corneal nerve fiber (a, *yellow arrows*). Multi-scale decomposition (c, d) of a synthetic corneal nerve fiber (b).

curvature, number of inflection points) is defined a priori; (3) the estimated tortuosity consists of a single number, therefore they are lacking in interpretation as the contribution of each tortuosity measure in the final estimate is lost; (4) they do not provide the confidence level of the estimated tortuosity, thus ignoring that some tortuosity levels might be more difficult to distinguish from others.

In our preliminary work on corneal nerve fiber tortuosity estimation,³² we addressed points 1 and 2, investigating which role is played by different spatial scales (i.e., different frequency of turns along the fiber) in tortuosity estimation (Fig. 1). Experimental results suggested that the notion of tortuosity can be captured quantitatively in a richer and more informative way by computing the tortuosity measures at multiple spatial scales, and we proposed an automated methodology to compute it. The need to measure tortuosity at multiple spatial scales is further strengthened by a recent study³³ highlighting the roles played by “short- and long-range tortuosity” in tortuosity grading, which we (a) formalized mathematically with multi-scale analysis and (b) discovered through a machine learning algorithm.³²

Herein, we assess the performance of our automated methodology in greater depth and compare it with the tortuosity assessments provided by expert observers. We measure performance on a global and per-level basis, to evaluate the accuracy of our system in predicting correctly all the tortuosity levels and each tortuosity level, respectively. Moreover, we address points 3 and 4 by proposing the tortuosity plane (TP), a two-dimensional (2-D) plane in which the coordinate axes are the weighted mean curvatures at the most significant spatial scales identified automatically. In this framework, each IVCM image is represented as a point on the TP, thus allowing a better interpretation of the estimated

tortuosity. Of importance, the TP represents images on a continuous 2-D tortuosity scale, thus leading to a finer discrimination among images.

METHODS

Image Acquisition

One hundred forty de-identified images of the subbasal nerve plexus of the central cornea were selected from an existing database. Images were acquired from 76 subjects (49 women and 27 men, 54.9 ± 19.3 years old) with laser scanning IVCM (Heidelberg Retina Tomograph 3 with the Rostock Cornea Module; Heidelberg Engineering GmbH, Heidelberg, Germany). The diode laser has a 670-nm red wavelength, and the microscope is equipped with a $\times 63$ objective lens with a numerical aperture of 0.9 (Olympus, Tokyo, Japan). The images obtained represent a coronal section of the cornea of $400 \times 400 \mu\text{m}^2$, from the subbasal layer. For imaging, a disposable sterile polymethylmethacrylate cap (Tomo-Cap; Heidelberg Engineering) was filled with a layer of hypromellose 0.3% gel (GenTeal; Alcon, Fort Worth, TX, USA) and mounted on the Cornea Module. After topical anesthesia with 0.5% proparacaine hydrochloride (Alcon), a drop of hypromellose 0.3% gel was applied to both eyes. One drop of this gel was also put on the outside tip of the cap to improve optical coupling. Then, the Cornea Module of the confocal microscope was advanced until the gel contacted the central surface of the cornea. Digital images were then recorded at a rate of 3 frames per second in sequence mode, including 100 images per sequence, from all layers of the cornea.

For this pilot-level study, 140 images showing a wide range of tortuosity characteristics were chosen from patients with a

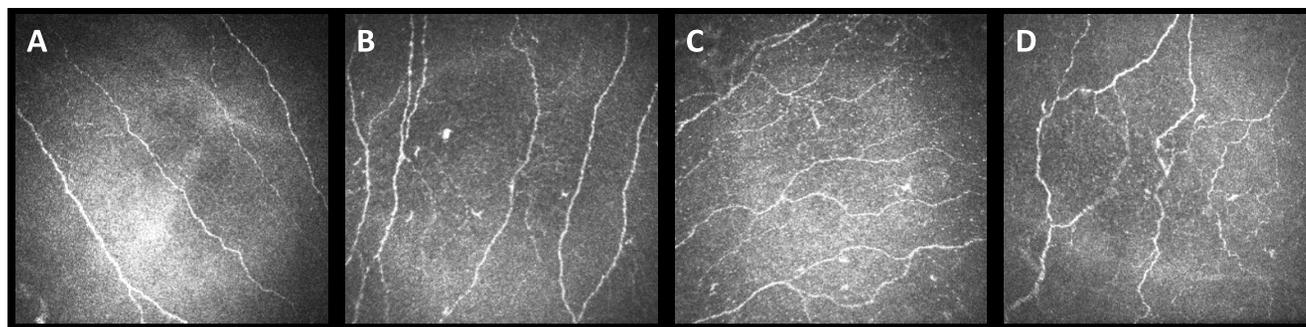


FIGURE 2. Corneal nerve fibers in the subbasal layer in IVCM. Original images from tortuosity level 1 (A) to level 4 (D).

TABLE 1. Tortuosity Measures Investigated With Our Automated Feature Selection Approach*

Tortuosity Measure	Description
$K_{\text{mean}}(1)$	Weighted average of mean curvatures estimated on each fiber at scale 1
$D_{\text{ip}}(1)$	Weighted average of density of inflection points at scale 1
$K_{\text{max}}(1)$	Weighted average of maximum curvature estimated on each fiber at scale 1
$K_{\text{mean}}(2)$	Weighted average of mean curvatures estimated on each fiber at scale 2
$D_{\text{ip}}(2)$	Weighted average of density of inflection points at scale 2
$K_{\text{max}}(2)$	Weighted average of maximum curvature estimated on each fiber at scale 2
$K_{\text{mean}}(3)$	Weighted average of mean curvatures estimated on each fiber at scale 3
$D_{\text{ip}}(3)$	Weighted average of density of inflection points at scale 3
$K_{\text{max}}(3)$	Weighted average of maximum curvature estimated on each fiber at scale 3
$K_{\text{mean}}(4)$	Weighted average of mean curvatures estimated on each fiber at scale 4
$D_{\text{ip}}(4)$	Weighted average of density of inflection points at scale 4
$K_{\text{max}}(4)$	Weighted average of maximum curvature estimated on each fiber at scale 4
$K_{\text{mean}}(5)$	Weighted average of mean curvatures estimated on each fiber at scale 5
$D_{\text{ip}}(5)$	Weighted average of density of inflection points at scale 5
$K_{\text{max}}(5)$	Weighted average of maximum curvature estimated on each fiber at scale 5
$K_{\text{mean}}(6)$	Weighted average of mean curvatures estimated on each fiber at scale 6
$D_{\text{ip}}(6)$	Weighted average of density of inflection points at scale 6
$K_{\text{max}}(6)$	Weighted average of maximum curvature estimated on each fiber at scale 6

* Each corneal nerve image shows a variable number of fibers. To obtain an image-level value for each tortuosity measure, we apply weighted averaging in which each fiber contribution is weighted by its length. These measures are computed at multiple spatial scales, from scale 1 (high-frequency turns) to scale 6 (low-frequency turns).

variety of eye conditions, typically at a depth of 50 to 80 μm (Fig. 2). Specifically, 20 images were taken from healthy subjects, 100 images from patients with dry eye disease, and 20 images from subjects with herpes simplex keratitis. Images were graded independently by three expert observers (PH, AK, and SA) using four tortuosity levels according to a previously published protocol.¹¹ Nerves were traced using NeuronJ (in the public domain at <http://www.imagescience.org/meijering/software/neuronj/>), a plug-in for the ImageJ software (in the public domain at <http://imagej.nih.gov/ij/>).

The study was performed in accordance with the tenets of the Declaration of Helsinki.

Proposed Tortuosity Estimation and Interpretation System

Our methodology can be summarized in four steps: (1) to compute multiple-scale tortuosity measures (or features) from fiber centerlines; (2) to use a feature selection approach to

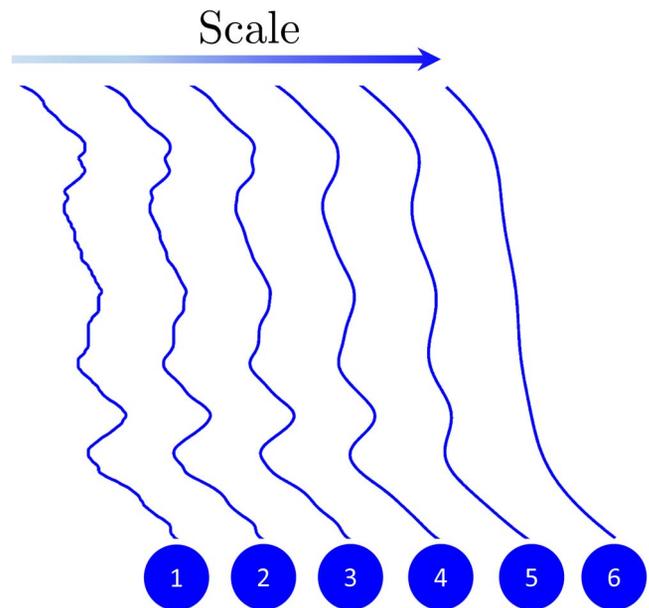


FIGURE 3. Multi-scale analysis of a corneal nerve fiber. From left to right: the original fiber including turns at all frequencies (spatial scale 1), and its smoothed versions at spatial scales from 2 to 6 to take into account turns at intermediate and high frequencies.

identify automatically the combination of tortuosity measures and scales leading to the best agreement with experienced observers; (3) to assign a new image to one of the four tortuosity levels according to the value of its tortuosity measures; and (4) to map the new image as a point on the TP and show the level of confidence (i.e., reliability) for the tortuosity level assigned by the system.

MATLAB (Release 2014b, The MathWorks, Inc., Natick, MA, USA) was used to implement the proposed system.

Computing Multiple-Scale Shape Features

The initial feature pool is formed by 18 measurements taken on the fibers (Table 1). We use maximum curvature and inflection points, the basis of the best-performing tortuosity indices in recent reports (see introduction), and average curvature, which led to promising performance in several methods.^{22,24,25,34} These three measures are computed at six spatial scales to take into account the contribution of turns at six different frequencies. As Figure 3 shows, we start by measuring curvatures and inflections points on the original fiber (including all the frequencies, scale 1). Then, we smooth out its shape to remove high-frequency turns and concentrate on lower frequency turns (scales 2–6).

Curvature and inflection points are measured with a highly accurate algorithm specifically designed to work on anatomic structures imaged at low resolution (such as corneal nerve fibers in our data set).³² Notice, we use the density of inflection points (i.e., number of inflection points per unit length) rather than their number, to yield a consistent comparison among fibers of different length.

Since each IVCM image includes a variable number of corneal nerve fibers, we average the 18 measures over all fibers in the image to obtain 18 global (i.e., image-level) measures. Assuming that longer fibers are more informative than shorter ones, we apply weighted averaging in which each fiber contribution to the global measure is weighted by its length.

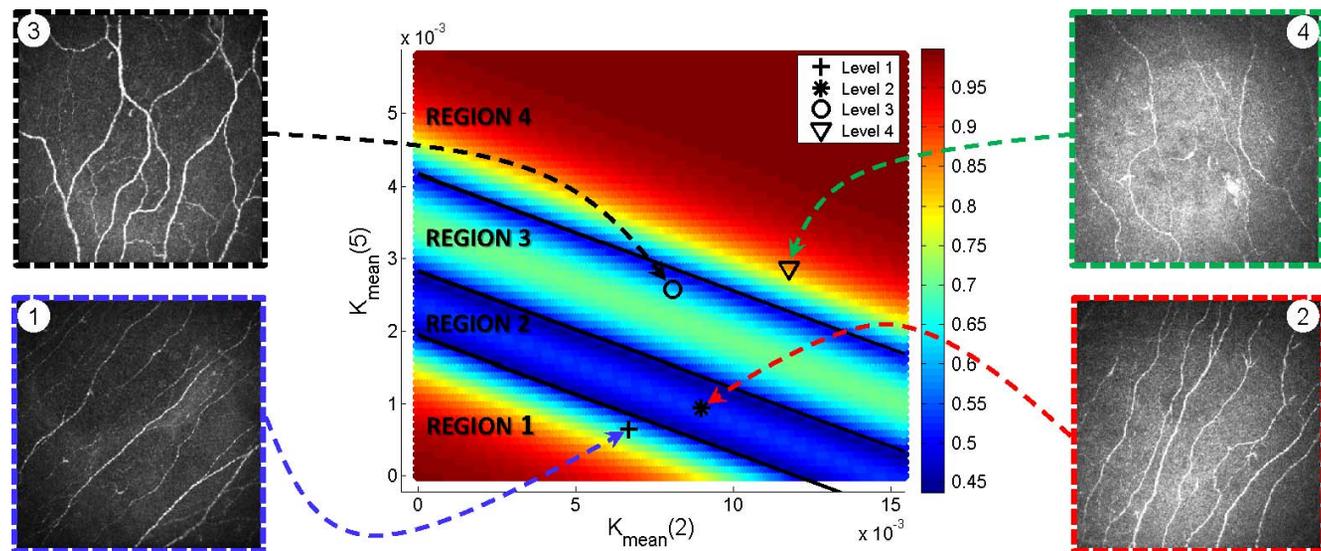


FIGURE 4. Tortuosity estimated for four corneal nerve images by the proposed method. Images are projected onto the TP as points (markers indicate the tortuosity level assigned by the expert observers), whose coordinates are the estimated mean curvature at scale 2 (high-frequency turns) and scale 5 (low-frequency turns). The level of confidence for each tortuosity estimate is encoded with color (red and blue mean high and low confidence, respectively). Intuitively, the confidence is related to the distance from the decision boundaries (indicated in black) separating each tortuosity region.

Identifying the Best Combination of Tortuosity Measures

Feature selection³⁵⁻³⁷ is a class of machine learning algorithms identifying the most discriminative features (or measures) for a task, here for tortuosity assessment. As our pool of measures is computed at multiple spatial scales, it also selects the most discriminative scales.

Here, we use a wrapper strategy,³⁷ in which the classification is based on a multinomial logistic ordinal regressor (MLOR). The MLOR models the log-cumulative odds³⁸; that is, the logarithm of the ratio of two probabilities: the probability that an image y has level of tortuosity c_j or lower, $P(y \leq c_j)$ and the probability that its tortuosity level is greater than c_j , $P(y > c_j)$, using a linear combination of the P tortuosity measures (features) X_j :

$$\ln\left(\frac{P(y \leq c_j)}{P(y > c_j)}\right) = \alpha_j + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p, \quad (1)$$

where, $j \in \{1,2,3\}$ and α and β are the weights of each tortuosity measure into the final assignment.

An MLOR model is built for each combination of tortuosity measures (an optimization step is performed to estimate the weights α and β), and it is used to assign the tortuosity level to a subset of training images. In this framework, the best combination of tortuosity measures and their scale is the one minimizing the number of misclassified images. Notice, we used cross-validation (i.e., randomly splitting the data set in training and test sets) to provide a fair assessment.

Assigning New Images to a Tortuosity Level

Once the best combination of tortuosity measures is found, we build the *final* MLOR model (i.e., using all the training images). The best weights α and β are estimated for each case (i.e., level 1 vs. level 2, 3, or 4; level 1 or 2 vs. level 3 or 4; level 1, 2, or 3 vs. level 4). Then, the selected tortuosity measures are computed on the new image, and its tortuosity level is

obtained using the maximum a posteriori criterion: the tortuosity level that the MLOR predicts as the most likely among the four considered is assumed as the *estimated* tortuosity level.

Tortuosity Plane and Confidence of the Estimated Tortuosity Level

Previously proposed methods typically provide a *single* tortuosity level (or coefficient value), obfuscating the different effects of the factors considered, and limiting interpretations that could be important to the ophthalmologist for diagnosis. Using the geometric interpretation of the MLOR model, we map each IVCM image onto a plane whose axes are the best tortuosity measures identified automatically by the feature selection procedure (in our case, weighted mean curvatures at spatial scales 2 and 5 corresponding to the fourth and 13th tortuosity measures in Table 1, respectively). Geometrically, the best weights α and β of the MLOR model define the best (in the MLOR sense) linear *decision boundaries* separating the TP into four regions corresponding to the four tortuosity levels (black lines in Fig. 4). Any new corneal nerve image can be plotted as a point on the TP by computing the values of the two best tortuosity measures (Fig. 4 shows four examples). The region containing the new point gives the estimated tortuosity level for the corneal nerve image.

Of importance, the TP also provides a *level of confidence* for the estimated tortuosity level, quantifying the reliability of the system. This confidence is given by the probability of belonging to one of the four regions (i.e., tortuosity levels) estimated automatically by the MLOR model and is intuitively proportional to the distance of the point (i.e., image) from the linear decision boundaries. In fact, the closer the point to the boundary between two adjacent regions, the less reliable is the estimated tortuosity level. The level of confidence for *all* the points on the TP can be estimated once the system is trained (i.e., before analyzing the target images) and color coded for immediate, intuitive visualization (Fig. 4).

TABLE 2. Per-Level Performance Measures of the Proposed Tortuosity Estimation Approach When the CGT Is Taken as Reference

Performance Measure	CGT Tortuosity Level			
	1	2	3	4
Accuracy, %	88.24	75.63	76.47	89.08
Sensitivity, %	78.13	63.16	51.85	63.64
Specificity, %	91.95	81.48	83.70	94.85
Positive predictive value, %	78.13	61.54	48.28	73.68
Negative predictive value, %	91.95	82.50	85.56	92.00

Performance Criteria and Statistical Analysis

Discriminating between four tortuosity levels is a multi-class classification problem. Following the standard performance assessment methodology for these problems,³⁹ we use accuracy (Acc), sensitivity (Se), specificity (Sp), positive predicted value (Ppv), and negative predictive value (Npv), as defined below.

$$\text{Acc} = \frac{1}{N} \sum_{i=1}^N \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{FN}_i + \text{FP}_i + \text{TN}_i} \quad (2)$$

$$\text{Se} = \frac{1}{N} \sum_{i=1}^N \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \quad (3)$$

$$\text{Sp} = \frac{1}{N} \sum_{i=1}^N \frac{\text{TN}_i}{\text{FP}_i + \text{TN}_i} \quad (4)$$

$$\text{Ppv} = \frac{1}{N} \sum_{i=1}^N \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i} \quad (5)$$

$$\text{Npv} = \frac{1}{N} \sum_{i=1}^N \frac{\text{TN}_i}{\text{FN}_i + \text{TN}_i} \quad (6)$$

where TP_i indicates true positives; TN_i , true negatives; FP_i , false positives; and FN_i , false negatives for the i -th tortuosity level ($N = 4$). We use weighted averaging as our data set is unbalanced, as suggested by Sokolova and Lapalme.³⁹ The above performance measures do not take into account the distance from the actual value when considering errors. To this aim, we also include the mean squared error (MSE) defined as

$$\text{MSE} = \frac{1}{N_{\text{imgs}}} \sum_{i=1}^{N_{\text{imgs}}} (y_i - l_i)^2 \quad (7)$$

where N_{imgs} is the total number of test images, y_i is the predicted tortuosity level, and l_i the true one, for the i -th image.

In addition to the aforementioned performance measures, we use Spearman's correlation as done elsewhere (see Ref. 27 for review). To assess the statistical significance of improvements among tortuosity estimation methods we use pair-sample t -tests.

RESULTS

System Agreement With Consensus Ground Truth

Assessing tortuosity is a difficult and subjective task. As such, the level of agreement among expert observers is arguably

TABLE 3. Spearman's Correlation and P Values Computed When an Individual Observer (Reported as Obs1, Obs2, and Obs3) Is Taken as Reference or Ground Truth (GT)*

GT	Spearman's Correlation ($P < 0.0001$)			
	Obs1	Obs2	Obs3	Proposed
Obs1	-	0.8017	0.7225	0.7594
Obs2	0.8017	-	0.7315	0.8880
Obs3	0.7225	0.7315	-	0.6916

* Performance of the proposed method is compared with the other two expert observers.

smaller compared with other tasks such as density estimation (see, for example, the repeatability study by Petropoulos et al.⁴⁰) To address this problem, we create a consensus ground truth (CGT): the tortuosity level of each image is given by the majority voting (i.e., the CGT tortuosity level for an image is the one assigned independently by at least two expert observers). Consistent with the difficulty of the task, 11 images were discarded owing to complete disagreement (all assigned levels different). We trained the system using the CGT and achieved promising global-level performance, reported previously (Annunziata et al.,³² Table 1). However, some tortuosity levels might be easier to estimate compared with others, or a specific level could be predicted with poor performance. To better investigate this point, we compute each performance measure on a per-level basis and show the results in Table 2. Our system achieves almost 90% accuracy for the tortuosity levels 1 and 4. Performance decreases for the middle levels owing to smaller differences with neighboring regions. Quantitatively, sensitivity and positive-predictive values decrease considerably owing to the increase in false negatives. Qualitative considerations on the TP are made below.

System Agreement With Individual Expert Observers

To explore agreement beyond consensus, we take the assessment made by one expert observer as ground truth and compare the performance of our framework (MLOR) with the other two expert observers using Spearman's correlation. This procedure is repeated three times, taking each expert observer as reference, in turn (Table 3). The correlations of our method with expert observers are significant ($\alpha_s = 0.05$, $P < 0.0001$). Moreover, the first and second rows of Table 3 show that the correlation of our system with the reference observer is higher than that of at least one of the other observers. Computed P values for the pair-sample t -tests on the MSE show that these improvements are significant for both hypotheses " $\text{MSE}_{\text{ours}} < \text{MSE}_{\text{Obs3}}$ " (row 1, $P < 0.001$) and " $\text{MSE}_{\text{ours}} < \text{MSE}_{\text{Obs1}}$ " (row 2, $P < 0.001$).

Tortuosity Plane

Figure 4 shows the TP obtained with the proposed method applied to a subset of images during a random 20-fold cross-validation procedure. The four regions related to the four tortuosity levels are identified automatically, and the best separating lines (visualized in black) are traced. Then, the two best tortuosity measures are computed on each of the testing images, and they are mapped onto the TP. Since each image belongs to a different tortuosity level, they are (correctly) mapped to four different regions. Visualizing the images as points on the TP allows an immediate and consistent interpretation. First, image 1 is close to the boundary 1-2, so the confidence of tortuosity level 1 (instead of 2) is limited; this

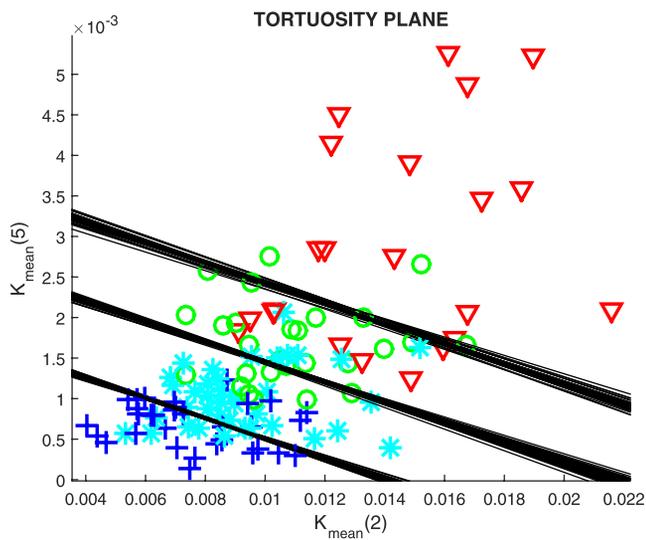


FIGURE 5. All IVCM images are projected onto the TP after a leave-one-out cross-validation on unseen images. *Markers* indicate the CGT tortuosity level. The estimated tortuosity level corresponds to the region within which an IVCM image is mapped. During each cross-validation, the best decision boundaries (shown in *black*) are computed based on the training set.

is consistent, as some fibers appear rather tortuous for level 1. The same considerations apply to image 3. Second, although image 2 belongs to the tortuosity level with the lowest overall accuracy (level 2), it is mapped to the center of the corresponding TP region, and therefore, the confidence of the prediction is relatively high. Image 4 is classified with the highest confidence as it clearly shows an abrupt large change (scale 5) in the direction of the second fiber (counting from left to right), and also high-frequency changes in other fibers (scale 2). Previous, single-scale methods would not tease out this difference. Third, the contribution of high- and low-frequency turns (i.e., curvatures at scales 2 and 5) for the predicted tortuosity level can be easily deduced. For instance, images 1 and 2 show a relatively low mean curvature at scale 5, but this is higher for images 3 and 4. Then, although image 2 is correctly assigned to tortuosity level 2, a higher mean curvature at scale 2 than image 3 is present: this suggests that, in our data set, the mean curvature at scale 2 (i.e., high-frequency turns) is slightly less discriminative than the one at scale 5.

We ran a leave-one-out cross-validation simulating the scenario in which the tortuosity of a new unseen image is assessed. For each validation, the best boundaries were estimated automatically, and the tortuosity level was assigned. Figure 5 shows the result of this process. We use different markers to indicate the true tortuosity level (i.e., l in Equation 7) of each image as per CGT. The four clusters reflecting the tortuosity levels are clearly visible and separated automatically by our system, although some misclassifications are present. We notice that the maximum error made by the system is always within one tortuosity level, showing very low MSE errors. Moreover, different regions show different spreading on the TP, especially in region 2 (very low) and region 4 (very high), suggesting that *intra*-level tortuosity characteristics should be better investigated in future studies. Of interest, tortuosity does not vary much in the Y axis, but does vary in the X axis, for levels 1 and 2. At levels 3 and 4, tortuosity varies in both the Y and X axes.

DISCUSSION

We have presented a novel methodology, the tortuosity plane, to represent and quantify tortuosity in the context of corneal nerve fibers in IVCM images of the subbasal nerve plexus. Unlike other methods using a single index in which the combination and weight of each tortuosity measure are defined a priori, we capture tortuosity with two measures selected by an automatic procedure on the basis of their agreement with grading from experienced observers.

Our approach reveals that tortuosity is best represented with curvature-related measures taken at multiple spatial scales (specifically, scales 2 and 5 accounting for high- and low-frequency turns), a computational finding supported by a recent, general study on the perception of tortuosity of corneal nerve fibers by expert observers.³⁵ Unlike this previous study, the proposed method attempts to discover the best tortuosity representation and the most discriminative spatial scales for a specific data set, using a fully automated procedure (feature selection). This novel approach to tortuosity estimation has the potential to answer interesting research questions related to the role played by each tortuosity measure and spatial scale for specific disease conditions.

Visualizing the predictions made by the proposed tortuosity estimation method on a 2-D plane, the TP, offers various advantages compared with traditional approaches providing a single number (i.e., tortuosity level or index).

First, it allows a finer (i.e., continuous) tortuosity scale compared with methods estimating tortuosity using a few levels. Compared with approaches based on indices (i.e., providing a continuous number), the TP provides a 2-D continuous tortuosity scale, thus allowing a better interpretation of the estimated tortuosity. Moreover, the contribution of each tortuosity measure into the final estimate becomes trivial (i.e., it is readily available on the axes of the plane). Overall, the TP allows a better tortuosity stratification and the possibility to identify subcategories clustering in specific regions of the plane.

Second, it visualizes the level of confidence (i.e., reliability) of the estimated tortuosity simply and intuitively, using color and also by the distance of a point from its closest boundaries (Fig. 4), an aspect ignored so far. As Figure 4 shows, the proposed method takes into account that the tortuosity of some images can be estimated with less confidence compared with others. A visual investigation suggests that a decreased level of confidence is due to the presence of fibers with different tortuosity characteristics within the same image. This is the result of training our system with image-level information (expert observers were asked to provide the level of tortuosity for the whole image). As expected, tortuosity levels 1 and 4 are estimated with higher confidence on average (e.g., regions 1 and 4 are mainly colored in red), compared with levels 2 and 3.

We showed in a pilot experiment that images with four tortuosity levels are grouped in distinct clusters on the TP. This immediately suggests that the TP could provide a quantitative, intuitive tool to stratify patients by condition, or by stage of a specific disease, a research avenue that we are pursuing.

Quantitative results suggest that our system achieves a level of performance that in two out of three cases is significantly better than that of experienced observers when compared against each other. Notice that we assessed the performance in a more challenging experimental setting compared with previous approaches (i.e., using four tortuosity levels instead of, at most, three in previous work).

The aforementioned considerations demonstrate the importance of using a 2-D plane and a multi-scale tortuosity estimation over single tortuosity indices, as proposed previously. Notice that the best scale for each tortuosity measure is

selected automatically by our feature selection strategy. This is important as it avoids taking into account tortuosity measures computed at very high scales (i.e., very low frequency). In fact, they could be affected by the presence of branching points breaking the continuity of the main fiber. We found no significant correlation between the tortuosity assigned by our system and the number of branches per image ($\rho = 0.0174$, $P = 0.85$). Likewise, averaging each tortuosity measure at fiber level is important to avoid the effect of higher-order aberration, which could favor longer fibers, for instance. No significant correlation between total nerve length per image and the tortuosity assigned by the system was found ($\rho = -0.1693$, $P = 0.0657$).

However, our method has some limitations: (1) Measurements are taken on manually traced fibers, thus introducing a certain level of subjectivity. Combining the method proposed herein with a reliable algorithm for automated fiber tracing would eliminate this residual subjectivity and make the whole system fully automated. (2) Since corneal nerve images were annotated as a whole (i.e., each image is associated with a single tortuosity level by the experienced observers), the contribution of each fiber is lost. We assumed that longer fibers should weight more in the global tortuosity measure, but other possibilities can be explored. (3) The number of tortuosity measures leading to the best agreement with experienced observers depends on the data set used, and more than two measures could lead to better performance on other data sets. This poses a visualization question: three tortuosity measures can still be visualized in a *tortuosity volume*, but four or more cannot. Although further tests are needed, the need for a higher dimensional representation including more than three tortuosity measures seems unlikely, given that we used a representative set of clinical IVCM images including different diseases.

Our immediate future work will explore the combination of this method for automated tortuosity estimation and interpretation with automated corneal nerve fiber detection.⁴¹⁻⁴³ This will lead to a fully automated system, which would contribute to disease stratification by tortuosity and better diagnosis based on objective, highly informative measurements.

Acknowledgments

Supported by EU Marie Skłodowska-Curie Initial Training Network Retinal Vascular Modelling, Measurement And Diagnosis (REVAM-MAD), Project Number 316990 (RA, ET) and by the National Institutes of Health Grant NIH R01-EY022695 (PH), Falk Medical Research Foundation (PH), and Massachusetts Eye and Ear Infirmary Foundation (PH). The authors alone are responsible for the content and writing of this paper.

Disclosure: **R. Annunziata**, None; **A. Kheirkhah**, None; **S. Aggarwal**, None; **B.M. Cavalcanti**, None; **P. Hamra**, None; **E. Trucco**, None

References

- Cruzat A, Pavan-Langston D, Hamrah P. In vivo confocal microscopy of corneal nerves: analysis and clinical correlation. *Semin Ophthalmol*. 2010;25:171-177.
- Cruzat A, Witkin D, Baniyasi N, et al. Inflammation and the nervous system: the connection in the cornea in patients with infectious keratitis. *Invest Ophthalmol Vis Sci*. 2011;52:5136-5143.
- Kurbanyan K, Sejjal KD, Aldave AJ, Deng SX. In vivo confocal microscopic findings in Lisch corneal dystrophy. *Cornea*. 2012;31:437-441.
- Muller RT, Abedi F, Cruzat A, et al. Degeneration and regeneration of subbasal corneal nerves after infectious keratitis: a longitudinal in vivo confocal microscopy study. *Ophthalmology*. 2015;XX:xxx-xxx. ?10
- Guthoff RF, Zhivov A, Stachs O. In vivo confocal microscopy, an inner vision of the cornea: a major review. *Clin Experiment Ophthalmol*. 2009;37:100-117.
- Patel DV, McGhee CN. Quantitative analysis of in vivo confocal microscopy images: a review. *Surv Ophthalmol*. 2013;58:466-475.
- Villani E, Mantelli F, Nucci P. In-vivo confocal microscopy of the ocular surface: ocular allergy and dry eye. *Curr Opin Allergy Clin Immunol*. 2013;13:569-576.
- Patel DV, McGhee CN. In vivo confocal microscopy of human corneal nerves in health, in ocular and systemic disease, and following corneal surgery: a review. *Br J Ophthalmol*. 2009;93:853-860.
- Villani E, Baudouin C, Efron N, et al. In vivo confocal microscopy of the ocular surface: from bench to bedside. *Curr Eye Res*. 2014;39:213-231.
- Aggarwal S, Kheirkhah A, Cavalcanti BM, et al. Autologous serum tears for treatment of photoallodynia in patients with corneal neuropathy: efficacy and evaluation with in vivo confocal microscopy. *Ocul Surf*. 2015;13:250-262.
- Oliveira-Soto L, Efron N. Morphology of corneal nerves using confocal microscopy. *Cornea*. 2001;20:374-384.
- Bucher F, Schneider C, Blau T, et al. Small-fiber neuropathy is associated with corneal nerve and dendritic cell alterations: an in vivo confocal microscopy study. *Cornea*. 2015;XX:xxx-xxx. ?11
- Chang PY, Carrel H, Huang JS, et al. Decreased density of corneal basal epithelium and subbasal corneal nerve bundle changes in patients with diabetic retinopathy. *Am J Ophthalmol*. 2006;142:488-490.
- De Cilla S, Ranno S, Carini E, et al. Corneal subbasal nerves changes in patients with diabetic retinopathy: an in vivo confocal study. *Invest Ophthalmol Vis Sci*. 2009;50:5155-5158.
- Hamrah P, Cruzat A, Dastjerdi MH, et al. Unilateral herpes zoster ophthalmicus results in bilateral corneal nerve alteration: an in vivo confocal microscopy study. *Ophthalmology*. 2013;120:40-47.
- Hamrah P, Cruzat A, Dastjerdi MH, et al. Corneal sensation and subbasal nerve alterations in patients with herpes simplex keratitis: an in vivo confocal microscopy study. *Ophthalmology*. 2010;117:1930-1936.
- Kocabeyoglu S, Mocan MC, Cevik Y, Irkec M. Ocular surface alterations and in vivo confocal microscopic features of corneas in patients with newly diagnosed Graves' disease. *Cornea*. 2015;34:745-749.
- Kurbanyan K, Hoegl LM, Schrems WA, Hamrah P. Corneal nerve alterations in acute Acanthamoeba and fungal keratitis: an in vivo confocal microscopy study. *Eye*. 2012;26:126-132.
- Labbe RM, Irimia M, Currie KW, et al. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells*. 2012;30:1734-1745.
- Ghadiri F, Pourreza H, Banaee T, Delgir M. Retinal vessel tortuosity evaluation via Circular Hough Transform. *Proceedings of the 2011 18th Iranian Conference of Biomedical Engineering (ICBME)*. City of Publisher: IEEE; 2011:181-184. ?12
- Grisan E, Foracchia M, Ruggeri A. A novel method for the automatic grading of retinal vessel tortuosity. *IEEE Trans Med Imaging*. 2008;27:310-319.
- Hart WE, Goldbaum M, Côté B, Kube P, Nelson MR. Measurement and classification of retinal vascular tortuosity. *Int J Med Inform*. 1999;53:239-252.
- Heneghan C, Flynn J, O'Keefe M, Cahill M. Characterization of changes in blood vessel width and tortuosity in retinopathy of

78

79

- prematurity using image analysis. *Med Image Anal.* 2002;6:407-429.
24. Kalitzeos AA, Lip GY, Heitmar R. Retinal vessel tortuosity measures and their applications. *Exp Eye Res.* 2013;106:40-46.
25. Onkaew D, Turior R, Uyyanonvara B, Akinori N, Sinthanayothin C. Automatic retinal vessel tortuosity measurement using curvature of improved chain code. In: *Proceedings of the 2011 International Conference on Electrical, Control and Computer Engineering (INECCE)*. City of Publisher: IEEE; 2011:183-186.
26. Trucco E, Azegrouz H, Dhillon B. Modeling the tortuosity of retinal vessels: does caliber play a role? *IEEE Trans Biomed Eng.* 2010;57:2239-2247.
27. Wilson CM, Cocker KD, Moseley MJ, et al. Computerized analysis of retinal vessel width and tortuosity in premature infants. *Invest Ophthalmol Vis Sci.* 2008;49:3577-3585.
28. Bullitt E, Gerig G, Pizer SM, Lin W, Aylward SR. Measuring tortuosity of the intracerebral vasculature from MRA images. *IEEE Trans Med Imaging.* 2003;22:1163-1171.
29. Lisowska A, Annunziata R, Loh GK, Karl D, Trucco E. An experimental assessment of five indices of retinal vessel tortuosity with the RET-TORT public dataset. *Conf Proc IEEE Eng Med Biol Soc.* 2014;XX:5414-5417.
30. Kallinikos P, Berhanu M, O'Donnell C, Boulton AJ, Efron N, Malik RA. Corneal nerve tortuosity in diabetic patients with neuropathy. *Invest Ophthalmol Vis Sci.* 2004;45:418-422.
31. Scarpa F, Zheng X, Ohashi Y, Ruggeri A. Automatic evaluation of corneal nerve tortuosity in images from in vivo confocal microscopy. *Invest Ophthalmol Vis Sci.* 2011;52:6404-6408.
32. Annunziata R, Kheirkhah A, Aggarwal S, Cavalcanti BM, Hamrah P, Trucco E. Tortuosity classification of corneal nerves images using a multiple-scale-multiple-window approach. In: Chen X, Garvin MK, Liu JJ, eds. *Proceedings of the Ophthalmic Medical Image Analysis (OMIA) First International Workshop, MICCAI 2014*. Publisher city: Publisher name; 2014:113-120.
33. Lagali N, Poletti E, Patel DV, et al. Focused tortuosity definitions based on expert clinical assessment of corneal subbasal nerves. *Invest Ophthalmol Vis Sci.* 2015;56:5102-5109.
34. Wilson CM, Cocker KD, Moseley MJ, et al. Computerized analysis of retinal vessel width and tortuosity in premature infants. *Invest Ophthalmol Vis Sci.* 2008;49:3577-3585.
35. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell.* 2005;27:1226-1238.
36. Lazar C, Taminau J, Meganck S, et al. A survey on filter techniques for feature selection in gene expression microarray analysis. *IEEE/ACM Trans Comput Biol Bioinform.* 2012;9:1106-1119.
37. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res.* 2003;3:1157-1182.
38. Bishop CM. *Pattern Recognition and Machine Learning*. City: Springer; 2006.
39. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag.* 2009;45:427-437.
40. Petropoulos IN, Manzoor T, Morgan P, et al. Repeatability of in vivo corneal confocal microscopy to quantify corneal nerve morphology. *Cornea.* 2013;32:e83-e89.
41. Annunziata R, Kheirkhah A, Hamrah P, Trucco E. Scale and curvature invariant ridge detector for tortuous and fragmented structures. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. City: Springer International Publishing; 2015:588-595.
42. Annunziata R, Kheirkhah A, Hamrah P, Trucco E. Boosting hand-crafted features for curvilinear structure segmentation by learning context filters. In: Navab N, Hornegger J, Wells W, Frangi A, eds. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. City: Springer International Publishing; 2015:596-603.
43. Annunziata R, Kheirkhah A, Hamrah P, Trucco E. Combining efficient hand-crafted features with learned filters for fast and accurate corneal nerve fibre centreline detection. In: *Engineering in Medicine and Biology Society - EMBS*. City: IEEE; 2015:XX-XX.