An Experimental Assessment of Five Indices of Retinal Vessel Tortuosity with the RET-TORT Public Dataset

Aneta Lisowska[1], Roberto Annunziata[1], Graeme Kenneth Loh[2], David Karl[2] and Emanuele Trucco[1]

***Abstract*— We compare the performance of five indices of retinal vessel tortuosity against sampling rates of vessel centerlines. We consider distance measure, tortuosity density, two curvature-based measures, and a recently introduced slope-chain coding for general curves, never before assessed comparatively with retinal vessels. To enable replication of our results, we use the public dataset for retinal tortuosity, RET-TORT. We find that (1) the tortuosity density index offers good performance overall, but is not always the best performer; (2) curvature-based methods are the best if high-frequency resampling is possible, but (3) are the most sensitive to variations of the number of samples; (4) slope-chain coding performs best at low sampling rates, but the length of the linear elements must be chosen carefully. In general, performance may vary considerably with resampling, suggesting that the choice of a tortuosity index for clinical inference requires attention to numerical details, and ideally standardization thereof.**

I. INTRODUCTION

The retinal microvasculature is one of the few portions of the human vascular system which can be non-invasively visualized in vivo. Among various retinal imaging techniques, fundus photography allows clinicians to observe, measure and monitor properties of the retinal vasculature related to eye diseases [1], screening programmes [2], and biomarker discovery [3]. Retinal vessel tortuosity is one of the clinical parameters used by ophthalmologist to assess the potential risk, severity and progression of various diseases [4]: e.g., increased tortuosity has been found to correlate with plus disease in retinopathy of prematurity [5], type-1 diabetes [6] and chronic anemia [7].

Clinicians grade tortuosity subjectively, usually on a 3- to 5-point scale. Grading results have been shown to vary with ophthalmologist, e.g., [8]. An objective, repeatable and quantitative assessment of vessel tortuosity has been long deemed desirable, as it sometimes informs important treatment decisions (e.g., ROP-plus disease [18]), and for biomarker research [19]. Several quantitative tortuosity indices and associated automatic algorithms have been proposed [9], [10], [11], [12], [17] but, due to the lack of public, large annotated datasets, ideally disease-specific, it remains very difficult to compare algorithms comprehensively and fairly [13]. The aim of this study is to compare and assess five quantitative tortuosity indices using the publicly available RET-TORT dataset [14], including a recent one [15] not yet evaluated comparatively. The indices, described in Section II, are (a) the distance measure, probably the earliest index, (b) the tortuosity density [14], a top performer in recent papers, (c) the two best-performing curvature-based measures from the well-known paper by Hart et al. [12], and (d) the index recently introduced by Bribiesca, assessed here comparatively for the first time with retinal vessels [15].

The indices were selected considering together their performance (as reported in the original papers), recency and impact on the field (citations, take-up). Following the established methodology [10], [14], [18], we estimate the tortuosity values using all indices for all vessels in both sets (artery and vein), rank them from least to most tortuous, and applied to retinal images [12], compute the Spearman correlation, ρ, with the ranking obtained from clinicians.

In the following, we summarize the tortuosity indices (Section II) and the data set used (Section III), describe the experimental protocol (Section IV), report and discuss results (Section V) and offer conclusions (Section VI).

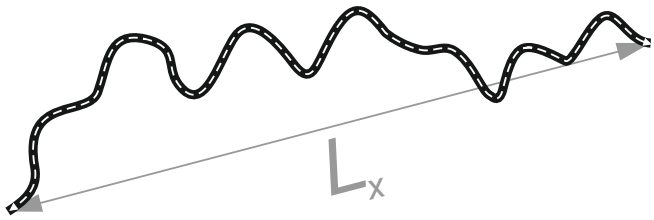

Fig. 1. Key quantities for the DM.

II. TORTUOSITY ALGORITHMS

A. Distance Measure (DM)

The DM is one of the earliest tortuosity indexes [12]. Its popularity may depend on its simple and intuitive definition:

$$DM = \frac{L_c}{L_x} \tag{1}$$

where L_c is the vessel centreline length (dashed white line in Figure 1) and L_x the chord length (line joining the vessel's endpoints, Figure 1. DM is 1 when a vessel is perfectly straight and increases with tortuosity. The limits of the DM have been pointed out previously [5], [12], crucially its inability to distinguish vessels with multiple bends (very tortuous) from vessels with a single arc (less tortuous) that have the same average deviation from the chord. This problem arises because DM is a global index which fails to capture local changes.

[1]Lisowska, Annunziata, Trucco are with the VAMPIRE project at the School of Computing, University of Dundee, Dundee, DD1 4HN. E-mail: {*e.trucco,r.annunziata,a.lisowska*}*@dundee.ac.uk*.

[2]Dr Loh and Dr Karl are with the Dept of Ophthalmology, Ninewells NHS and University Hospital, Dundee. E-mail: {*graemeloh,david.karl*}*@nhs.net*.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

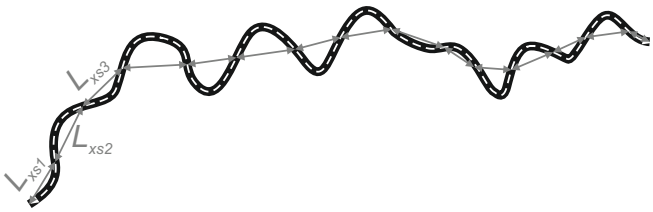

Fig. 2. The vessel partitioning used by the TD algorithm.

B. Tortuosity Density (TD)

To address the problem above, Grisan et al. proposed the TD index [10] (Figure 2). TD assesses vessel tortuosity by summing the contributions to tortuosity of uniformly convex or concave arcs, as follows:

$$TD = \frac{n-1}{n}\frac{1}{L_c}\sum_{i=1}^{n}\left[\frac{L_{csi}}{L_{xsi}} - 1\right] \tag{2}$$

Here, n is the number of "turns" (curvature sign changes, i.e., zero-crossings of the second derivative of the centreline), L_{csi} is the arc length of segment i, L_{xsi} is the chord length of segment i. L_c is the length of the whole vessel centreline. A vessel centreline with only one turn has $TD = 0$; with more than one turn, the tortuosity is greater than 0 (avoiding the problem of DM). The TD index is also normalized to vessel length ($1/L_c$), which allows comparison of vessels of various lengths and scale invariance. The authors found TD to be the most accurate index to model clinical scores of retinal vessel tortuosity with hypertensive retinopathy images [10], [14].

C. Slope Chain Coding (SCC)

The SCC index was suggested recently as a general index for planar curves by Bribiesca [15]. The paper includes a qualitative demonstration on a single ROP image but no comparative tests against other indices. To calculate SCC, a vessel centreline is approximated by a linear piecewise curve formed by line segments of fixed length, and the slope change between segments is computed. The SCC index is defined as the integral of the absolute values of the slope changes along the centreline:

$$SCC = \sum_{i=1}^{n} |a_i| \tag{3}$$

Here, n is the number of slope changes (the number of segments minus one) and a_i is the slope change after the i-th segment. The influence of n on the tortuosity is addressed in [15] by an example showing invariance of SCC with two values of n, but no proof is offered. To our best knowledge, we are the first to compare the SCC index with other tortuosity indices from the literature in terms of association with ground truth tortuosity scores.

D. Curvature-integral measures

The curvature-based measures selected are the two best performing ones from the well-known paper by Hart et al. [12]. These are τ_3 (overall best) and τ_5, defined as

$$\tau_3 = t_{sc} \tag{4}$$

$$\tau_5 = \frac{t_{sc}}{L_c} \tag{5}$$

Where t_{sc} is the total squared curvature and L_c is the vessel centreline length, as above. To improve digital curvature estimation, we use the highly accurate, noise-resilient algorithm developed by Annunziata and Trucco [16], based on multi-window ellipse and line fitting to curve segments.

III. MATERIALS

Most public datasets available concentrate on retinal signs of diabetes (e.g., microaneurysms, exudates) and a few other eye diseases, e.g., glaucoma. We refer the reader to the list of 11 public datasets with URLs published recently in [13], Appendix. To enable replication of our results, we use the publicly available dataset for retinal tortuosity, RET-TORT, created by the University of Padova (bioimlab.dei.unipd.it) [14]. To our best knowledge, this is the only public data set with clinical annotations for retinal vessels. RET-TORT consists of 30 arteries and 30 veins of similar length and calibre, extracted from 60 1300×1100 fundus images of retinal vessels from normal and hypertensive patients.

All vessel centrelines were sampled manually at regular intervals and with the same frequency. The resulting number of samples varies between 19 and 50 depending on vessel length. Images were acquired with a 50^o-FOV Topcon TR50 fundus camera in a clinical context. The authors chose major arteries and veins with minimal overlap with surrounding vessels.

Three clinical specialists, denoted C_1, C_2, C_3, ranked the vessels independently by tortuosity. C_1 is the specialist whose rankings are included in the RET-TORT data. C_2 and C_3 are our clinical co-authors (Loh, Karl). The agreement (Spearman correlation) between specialists is shown in Table I. We notice that the correlation for veins is lower than for arteries.

TABLE I
RANK CORRELATION BETWEEN ANNOTATORS.

Arteries	C_2	C_3	**Veins**	C_2	C_3
C_1	0.93	0.93	C_1	0.88	0.88
C_2	–	0.97	C_2	–	0.94

IV. METHODS

Our aim was to compare the performance of the five algorithms selected, using the Spearman rank correlation of automatic and clinical rankings, as done elsewhere [14], [18]. Values obtained by each index for arteries and veins were obtained separately. All algorithms were implemented by ourselves in MATLAB. We first rotated the centreline samples to guarantee a one-to-one explicit representation, $y = y(x)$. Following [14], we then used spline interpolation on the rotated centrelines.

Test set 1: high sampling rates. We first tested performance with high sampling rates on the spline, as for centrelines obtained by automatic vessel segmentation. "High" means

that the number of samples is much larger than the minimum needed to reconstruct the digital curve for the purpose of tortuosity estimation; as this is a qualitative process, we do not use quantitative criteria, e.g., Nyquist. The original sampling rate in RET-TORT is approximately 1/10, yielded by the manual sampling (between 19 and 54 samples per artery, and between 19 and 48 samples per vein). Here we used 100, 200 and 300 samples from the interpolated spline, i.e., one order of magnitude larger than the original.

For the TD index, we divided the vessel into segments of constant curvature sign, i.e., convex or concave, by locating inflection points as zero-crossings of the second derivative of the centrelines. If the zero-crossing occurred on a locally straight segment, we placed the inflection point in its middle. Following [10], we limited the impact of numerical noise on the location of inflection points using an hysteresis algorithm for reliable zero-thresholding (threshold $t = \pm 0.02$).

Test set 2: medium sampling rates. We tested performance with more modest sampling rates of the spline, using the original sampling rate, and twice and thrice that rate. This yielded a smaller number of samples than in test set 1, respectively in the intervals [19,54], [38,108], [57,162] (over both arteries and veins), partially overlapping the lowest end of the range of test set 1.

Test 3: SCC. This index is based on a uniform linear piecewise approximation of the centrelines. We tested the influence of the number of linear segments, n, (or equivalently, of their length, as it is a constant) on the Spearman rank correlation. We used n values in [15,70], step 5; 15 and 70 yield, respectively, longer and shorter elements than the original sampling rate.

V. RESULTS AND DISCUSSION

TABLE II
SPEARMAN CORRELATION, HIGH SAMPLING RATE (TEST SET 1). NES = NOT ENOUGH SAMPLES; C_1, C_2, C_3 = CLINICAL ANNOTATOR; ORIGINAL NUMBER OF SAMPLES BETWEEN 19 AND 54. BOLDED = BEST ALGORITHM GIVEN OBSERVER.

		Original		**100 samples**		**200 samples**		**300 samples**	
		A	**V**	**A**	**V**	**A**	**V**	**A**	**V**
	TD	**0.89**	0.76	**0.91**	**0.82**	**0.89**	**0.83**	0.87	**0.81**
	SCC	0.85	**0.77**	0.83	0.66	0.83	0.66	0.83	0.66
	DM	0.8	0.63	0.84	0.64	0.84	0.64	0.84	0.64
C_1	τ_5	NES		0.76	0.56	**0.89**	0.65	**0.92**	0.72
	τ_3	NES		0.76	0.57	**0.89**	0.67	**0.92**	0.74
	TD	0.85	**0.81**	**0.9**	**0.92**	**0.91**	**0.92**	0.87	**0.86**
	SCC	0.83	**0.81**	0.76	0.75	0.76	0.75	0.76	0.75
	DM	0.71	0.74	0.74	0.74	0.75	0.75	0.74	0.75
C_2	τ_5	NES		0.71	0.7	0.88	0.88	**0.91**	0.85
	τ_3	NES		0.71	0.71	0.88	0.88	**0.91**	**0.86**
	TD	0.85	0.73	**0.93**	**0.87**	**0.94**	**0.85**	0.89	**0.81**
	SCC	**0.89**	**0.79**	0.82	0.7	0.82	0.71	0.82	0.71
	DM	0.79	0.69	0.82	0.69	0.82	0.69	0.82	0.69
C_3	τ_5	NES		0.79	0.63	0.93	0.71	0.96	0.76
	τ_3	NES		0.79	0.63	**0.94**	0.71	**0.96**	0.76

Table II shows the Spearman correlation for test set 1 (high sampling rate) and highlights the best performer for arteries and veins at each sampling rate and for each observer. In line with the inter-observers' correlations (Table I), performance is in general better for arteries. All correlations were significant ($p < 10^-7$, significance level $\alpha = 0.05$).

For test set 1 (high sampling rates, Table II), TD is the most frequent winner and is never beaten for veins. As the sampling rate increases, Hart's curvature-based algorithms become the best for arteries (300 samples), but not always for veins. This is in line with the the fact that high sampling rates improve the accuracy of curvature estimation, which in turn may lead to tortuosity estimates better in line with clinical judgement. Interestingly, SCC is the best for arteries at the original sampling rate, suggesting that, contrary to curvature-based ones, this index is appropriate for low sampling rates.

For test set 2 (medium sampling rate; Table III), SCC and TD are the best for the lowest $2\times$ sampling rate. This is in line with the behaviour observed in test set 1, reinforcing the practical recommendation stated above. Again curvature-based measures become the best at higher sampling rates, in line with what found in test set 1. TD remains a good performer even at lower sampling rates than in test set 1, but SCC is a good alternative at the original and two times the original sampling rate. The abnormally low figures (0.34 to 0.58) for τ_3, τ_5 at the original and $2\times$ sampling rates are due to not enough samples on several short vessels for those sampling rates.

TABLE III
SPEARMAN CORRELATION, MEDIUM SAMPLING RATE (TEST SET 2). SYMBOLS AND NOTATION AS IN TABLE II

		Original		**2×**		**3×**	
		A	**V**	**A**	**V**	**A**	**V**
	TD	**0.89**	0.76	**0.9**	**0.84**	**0.91**	**0.79**
	SCC	0.85	**0.77**	0.87	0.77	0.85	0.76
	DM	0.8	0.63	0.82	0.64	0.84	0.62
C_1	τ_5	NES		0.51	0.36	0.87	**0.79**
	τ_3	NES		0.51	0.34	0.88	0.77
	TD	**0.85**	**0.81**	**0.88**	**0.86**	**0.9**	**0.89**
	SCC	0.83	**0.81**	0.85	0.81	0.83	0.82
	DM	0.71	0.74	0.74	0.73	0.75	0.73
C_2	τ_5	NES		0.53	0.43	0.89	0.83
	τ_3	NES		0.53	0.41	0.89	0.83
	TD	0.85	0.73	0.9	0.77	0.92	**0.8**
	SCC	**0.89**	**0.79**	**0.91**	**0.79**	0.91	0.79
	DM	0.79	0.69	0.81	0.69	0.81	0.69
C_3	τ_5	NES		0.58	0.36	**0.94**	**0.8**
	τ_3	NES		0.58	0.36	**0.94**	0.79

As the number of samples increases, the performance change for the curvature-based indices is generally larger than that of the others, although we used the multi-window curvature estimation algorithm developed by Annunziata and Trucco [16], which proved resilient to sampling noise and sampling rate. This suggests that a high-frequency resampling should be used whenever curvature-based indices are adopted.

Test 3: SCC. With the protocol described, the ρ values varied from 0.68 and 0.83, a substantial variation given the range of values in Tables II and III, and one order of magnitude larger than the maximum variation among annotators (Table I: .97-.88 = .09). We conclude that n is a critical parameter of SCC.

VI. CONCLUSIONS

We have presented an experimental comparative study of the performance of five retinal tortuosity indices with varying sampling rates of vessel centrelines. To facilitate reproducibility, we used a public dataset, RET-TORT, and the commonly adopted Spearman correlation coefficient as performance measure. We assessed, for the first time, the performance of Bribiesca's index. Our results indicate that the number of samples does influence performance and its choice must be considered carefully whenever tortuosity estimates are used for clinical associational studies. We also arrived at the following practical guidelines: (1) the tortuosity density index offers good performance overall, but not the best one at all sampling rates; (2) curvature-based methods are the best if high-frequency sampling is possible (always the case when reliable vessel centrelines are estimated automatically), but (3) are the most sensitive to variations of the number of samples; (4) slope-chain coding, the new measure tested quantitatively for retinal images, performs as well as the tortuosity density at low sampling rates, but the length of the linear elements must be chosen carefully.

ACKNOWLEDGMENT

This research was made possible in part by the EU Marie Curie Initial Training Network (ITN) REtinal VAscular Modelling, Measurement And Diagnosis" (REVAMMAD), project number 316990. The authors are indebted with the VAMPIRE group members for useful discussions.

REFERENCES

[1] G Liew, JJ Wang, P Mitchell, TY Wong. "Retinal vascular imaging: a new tool in microvascular disease research". Circulation: Cardiovascular Imaging, vol 1, pp. 156-161, 2008.

[2] Harding SP, Broadbent DM, Neoh C, White MC, Vora J. "Sensitivity and specificity of photography and direct ophthalmoscopy in screening for sight threatening eye disease: the Liverpool diabetic eye study". BMJ; vol.311 pp. 1131-1135, 1995.

[3] Nunes S, Pires I, Rosa A, Duarte L, Bernardes R, Cunha-Vaz J. "Microaneurysm turnover is a biomarker for diabetic retinopathy progression to clinically significant macular edema: Findings for type 2 diabetics with nonproliferative retinopathy". Ophthalmologica., vol 223 pp. 292-297, 2009.

[4] Angelos A, Kalitzeos, Gregory Y.H. Lip, Rebekka Heitmar, "Retinal vessel tortuosity measures and their applications". Experimental Eye Research, vol. 106, pp: 40-46, 2013.

[5] Capowski JJ, Kylstra JA, Freedman SF. "A numeric index based on spatial frequency for the tortuosity of retinal vessels and its application to plus disease in retinopathy of prematurity". Retina, vol. 15, pp. 490-500, 1995

[6] Sasongko, M.B., Wang, J.J., Donaghue, K.C., Cheung, N., Benitez-Aguirre, P., Jenkins, A., Hsu, W., Lee, M.L., Wong, T.Y., "Alterations in retinal microvascular geometry in young type 1 diabetes". Diabetes Care vol.33, pp. 1331-1336, 2010a.

[7] Incorvaia, C., Parmeggiani, F., Costagliola, C., Perri, P., D'Angelo, S., Sebastiani, A. "Quantitative evaluation of the retinal venous tortuosity in chronic anaemic patients affected by [beta]-thalassaemia major". Eye vol.17, pp. 324-329, 2003.

[8] Kagan, A., Aureli, E., Dobree, J. "A note on signs in the fundus oculi and arterial hypertension: conventional assessment and significance". Bulletin of the World Health Organization vol.34, pp. 955-960, 1966.

[9] Joshi V, Reinhardt JM, Abramoff MD. "Automated measurement of retinal blood vessel tortuosity". SPIE Medical Imaging. vol. 7624, February 2010.

[10] Grisan E, Foracchia M, Ruggeri A. "A novel method for the automatic evaluation of retinal vessel tortuosity". Proceedings 25th IEEE Intern Conf on Engineering in Medicine and Biology (EMBS), Cancun, Mexico; September 17-21, 2003. New York: IEEE; 2003:866-869.

[11] E. Bullitt, G. Gerig, S. Pizer, W. Lin and S. Aylward. "Measuring Tortuosity of the Intracerebral Vasculature from MRA Images". IEEE Transactions on Medical Imaging, vol.22, pp. 1163-1171, 2003.

[12] W. Hart, M. Goldbaum, B. Cote, P. Kube, M. Nelson. "Measurement and classification of retinal vascular tortuosity". Intern Journal of Medical Informatics, vol. 53, no. 2-3, pp. 239-252 ,February 1999.

[13] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J.P. Hubschman, B. al-Diri, C. Cheung, D. Wong, M. Abrmoff, G. Lim, D. Kumar, P. Burlina, N. Bressler, H. Jelinek, F. Meriaudeau, T. MacGillivray, B. Dhillon. "Validating retinal fundus image analysis algorithms: issues and a proposal". Investigative Ophthalmology and Visual Science, vol. 54, pp. 35463559, 2013.

[14] Grisan E, Foracchia M, Ruggeri A. "A novel method for the automatic grading of retinal vessel tortuosity". IEEE Transactions on Medical Imaging, vol. 27, pp.310319, 2008 .

[15] Bribiesca E. "A measure of tortuosity based on chain coding". Pattern Recognition, vol.46, pp. 716-724, 2013.

[16] Annunziata R., Trucco E. "A Multiple-Window Approach for Digital Curvature Estimation". Pattern Recognition, submitted for publication, 2014.

[17] Azegrouz H., Trucco E., Dhillon B. "Modeling the tortuosity of retinal vessels: does caliber play a role?". IEEE Transactions on Biomedical Engineering, vol 34 no. 9, pp. 2239-47, 2010.

[18] Wilson C, Wong K, Ng J, Cocker KD, Ells AL, Fielder AR. "Digital image analysis in retinopathy of prematurity: A comparison of vessel selection methods". Journal of the American Association for Pediatric Ophthalmology and Strabismus, vol 16, no 13, pp. 223-8, 2012.

[19] Doubal F, Hokke P, Wardlaw J. "Retinal microvascular abnormalities and stroke: a systematic review". Journ Neurol Neurosurg Psychiatry, vol 80, pp. 158-165, 2008.