34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Image-level Tortuosity Estimation in Wide-field Retinal Images from Infants with Retinopathy of Prematurity

Enea Poletti, Enrico Grisan, and Alfredo Ruggeri

Abstract— **Tortuosity and dilation of retinal vessels are considered of primary importance for the diagnosis and follow-up of the Retinopathy of Prematurity (ROP) disease. We developed an algorithm to estimate vessel tortuosity in images acquired with a wide-field fundus camera in ROP subjects, offering clinicians a quantitative, objective, and reproducible diagnostic parameter.**

Vessels were manually traced in 20 images to provide error-free input data for the tortuosity estimation. At first we investigated different vessel-level measures, some including also caliber information. Then we used them to obtain different image-level tortuosity measures, which were eventually combined in a supervised approach to provide a tortuosity index capable to reproduce the clinical experts assessment.

To provide manual assessment, the 20 images were independently ordered by increasing tortuosity by three clinical graders and three retinal imaging experts. The proposed tortuosity index obtains a Spearman's correlation coefficient of 0.95 with ground truth, a performance comparable to the clinical graders' one and better than the retinal imaging experts' one.

I. Introduction

Retinopathy of Prematurity (ROP) [1] is an eye disease that affects prematurely born infants. It can be mild and resolve spontaneously, but in more serious cases it becomes very aggressive: new blood vessel formation progresses to scarring, retinal detachment and blindness.

ROP is categorized by zone, stage, and presence of plus disease. Plus disease is an indicator of ROP severity and may be characterized by different signs: arterial tortuosity and venous dilation at the posterior pole, vitreous haze, and iris rigidity. Plus disease is difficult to reproducibly quantify: diagnosis is based on 20-year-old reference photographs (ICROP, [1]) Pre-plus disease, described as vascular abnormalities at the posterior pole that are insufficient to diagnose plus disease but demonstrate more arterial tortuosity and venous dilation than normal, was added to the classification of ROP in 2005 (revisited ICROP, [2]).

A recent study [3] shows that inter-reader agreement on central vascular changes is poor, especially when based on more than two severity levels. Authors conclude that further refinement of the revisited ICROP guidelines or/and availability of a reliable and valid computer-based image analysis system to further standardize quantification of central vascular changes are urgently needed. To overcome the inherent inaccuracies in qualitative evaluation, several groups have explored the use of automated techniques [4-7] to analyze infants' retinal fundus. Yet, none of them has advanced to the stage of being widely used as a clinical tool.

All authors are with the Department of Information Engineering, University of Padova, Padova, 35131 PD Italy (phone: (+39)049-827-7758; fax: (+39)049-827-7699; e-mail: enea.poletti@dei.unipd.it).

Several mathematical definitions of tortuosity have been proposed in literature, yielding however different results in various standard situations of tortuosity (e.g., sinusoidal, helical, etc.) [8]. The problem is further compounded by the lack of agreement about what clinical experts mean by tortuosity and thus which of the many measures best fit with the experts' evaluations [9].

In this study we propose a new approach for the tortuosity evaluation of ROP images. We refer to tortuosity as a property of both single vessels (vessel-level) and whole images (image-level). At the first level, we propose the use of different measures that can be integrated with information about vessel calibers. At image-level, the global image tortuosity is defined as a combination of vessel-level measures. Since there are many ways to combine these measures, we employed a regression approach on the clinical graders' ground truth, in order to replicate their evaluation.

II. Materials

Twenty videos of retinal fundus have been acquired in premature infants with the RetCam fundus camera (Clarity Medical Systems Inc., CA, USA) with a 130° field of view and 640×480 pixels image size. From each video a single image was manually selected for analysis, with the aim of having images with minimal movement artifacts and maximal focus and contrast (see sample image in Fig. 1-a).

The expert manual evaluation of tortuosity was provided by six domain experts: three ROP clinical graders (GRADs), and 3 non-clinical ROP imaging experts (EXPs). The set of 20 images was independently ordered by increasing tortuosity by each of them.

In order to make the manual sorting of the images easier, we provided all experts with a custom software we developed, TorTsorT (available for download at http://bioimlab.dei.unipd.it). TorTsorT is based on a merge-sort algorithm and presents pairs of images to the user, asking him/her to select the more tortuous image. Merge-sort is a comparison-based divide and conquer sorting algorithm that ensures average- and worst-case performance of $O(n \log n)$. As in our study we have $n = 20$ images to be ordered, the average number of required pair comparisons is 60. Eventually, we obtained six different ordered lists, each expressing for each image i its ranking $R(i) \in \{1, \dots, 20\}$.

To obtain the ordering to be used as ground truth reference (Table I) we computed the average ordering by the three clinical graders, defined as:

$$R_{GT}(i) = rank\left(mean\left(R_{GRAD_1}(i), R_{GRAD_2}(i), R_{GRAD_3}(i)\right)\right) \quad (1)$$

For each of the 20 images, a careful manual segmentation of the vessels was provided by one of the authors (EP). Manual tracing was used to focus on tortuosity formulation alone and

TABLE I. MANUAL EXPERT ORDERING

image name	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*
$R_{GRAD\ 1}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$R_{GRAD\ 2}$	6	3	4	1	2	7	5	9	8	10	11	12	13	17	14	18	15	16	19	20
$R_{GRAD\ 3}$	4	3	2	1	6	5	7	10	11	9	8	12	13	14	15	16	17	18	19	20
$R_{EXP\ 1}$	9	4	6	2	1	10	3	11	8	7	5	12	14	15	13	18	17	16	19	20
$R_{EXP\ 2}$	8	4	1	3	2	7	9	6	11	10	5	12	16	14	13	19	15	20	18	17
$R_{EXP\ 3}$	4	5	1	6	2	9	3	10	12	7	8	17	11	16	15	13	14	19	20	18
R_{GT}	4	2	3	1	5	6	7	8	9	10	11	12	13	15	14	17	16	18	19	20

The 20 image were independently ordered by increasing tortuosity by each of the 3 clinical graders (GRAD1, GRAD2, GRAD3) and each of the 3 experts (EXP1, EXP2, EXP3). $R \in \{1, \dots, 20\}$ expresses the rank of an image in the list (i.e., the 1st image according to GRAD1 is the 6th according to GRAD2). Images are named in accordance with the ordering of GRAD1.

avoid possible problems that may come from errors in automatic vessel tracing.

III. METHODS

The rationale of our approach is to find the best combination of vessel-based measures (tortuosity, diameters, length) that provides an image-level tortuosity measure capable to reproduce the clinical evaluation of image tortuosity. First, we selected and modified, among the previously proposed vessel-level tortuosity measures, the most promising ones (Sec. III-A). Then we tried several combinations of vessel-level measures in order to obtain image-level measures (Sec. III-B). Lastly, we evaluated a linear weighted combination of the image-level measures with the aim of replicating the clinicians' assessment (Sec. III-C).

A. Vessel-level Tortuosity Measures

Various vessel-level tortuosity indexes have been proposed in the literature [8], each based on different geometrical aspect of the vessels. Since the tortuosity formula proposed in this work employ geometrical information about vessels, we introduce some definitions.

Let the vessel v be described by an ordered set of triplets $v = \{(x_n, y_n, d_n) : n \in \{1, \dots, N\}\}$, where (x_n, y_n) are the coordinates of the n^{th} vessel sample, d_n is the diameter of the vessel at the n^{th} vessel sample and N is the number of the equidistant samples. We assume here that the diameter of a vessel is the average of its diameters at all the sample points

$$D_v = \frac{1}{N}\sum_{n=1}^{N} d_n \tag{2}$$

while its *length* L_v and its *chord length* $L_{\chi v}$ are respectively defined as:

$$L_v = \sum_{n=2}^{N}\sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2} \tag{3}$$

$$L_{\chi v} = \sqrt{(x_N - x_1)^2 + (y_N - y_1)^2} \tag{4}$$

Using these definitions about vessel's geometry, we propose the following vessel-level measures.

1) *Angle-based Tortuosity*

A measure of tortuosity based on the local direction variation of the vessel has been proposed [10], in which the average of the angles between consecutive vessel samples is computed. However, in that formulation angles can only vary between $-\pi/2$ and $\pi/2$, and no information about vessel width is taken into account. We propose here a modified version of this angle-based tortuosity measure.

Being $(x_n,\ y_n)$ the coordinates of the n^{th} sample point of the vessel, the vector from the n^{th} sample to the $(n-1)^{th}$ sample center points is $\vec{v}_n = (x_n - x_{n-1},\ y_n - y_{n-1})$ and describes the local vessel direction. The local angle variation is then defined as

$$\vartheta_n = \ arctg2(\vec{v}_n) - \ arctg2(\vec{v}_{n-1}) \tag{5}$$

where $arctg2$ is a function with values between $-\pi$ and π, representing the counterclockwise angle, measured in radians, between the positive x axis, and the point (x, y) (Fig. 1-b). We define the tortuosity of the vessel v as:

$$\Phi_v = \frac{1}{L_v}\sum_{n=3}^{N} \vartheta_n^2 \tag{6}$$

This tortuosity measure has a dimension of $rad^2/length$ and thus may be interpreted as a *tortuosity density,* allowing its comparison on vessels of different length.

2) *Caliber-weighted Angle-based Tortuosity*

At variance with Eq. 6, we defined also an expression that takes into account the vessel diameter of each sample:

$$\Phi d_v = \frac{1}{L_v}\sum_{n=2}^{N} \vartheta_n^2 d_n \tag{7}$$

3) *Twist-based Tortuosity*

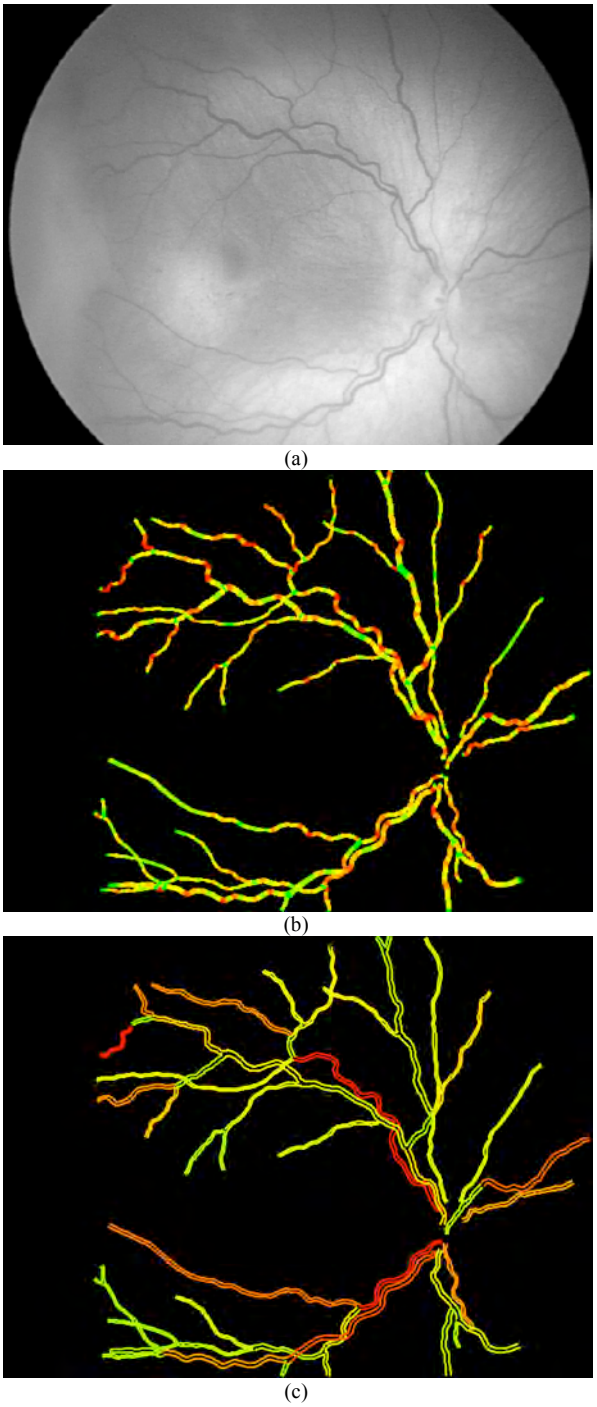
Authors of [8] proposed a vessel tortuosity measure based on twists (or inflections), located at points where the curvature sign changes. The "turn curve" v_t is defined as the portion of a vessel v located between two consecutive twists. This measure requires at first a partitioning of the vessel v into T turn curves: $v = \{v_1,\ v_2, \dots, v_T\}$. The twist-based tortuosity τ_v of the vessel v (Fig. 1-c) is defined as:

$$\tau_v = \frac{T-1}{T}\ \frac{1}{L_v}\sum_{t=1}^{T}\left[\frac{L_{v_t}}{L_{\chi v_t}} - 1\right] \tag{8}$$

where the *length* $L_{v_t}$the *chord length* $L_{\chi v_t}$ refers to each turn curve $v_t \in v$. Also this measure can be interpreted as *tortuosity density,* having dimension of $1/length$. It is worth noting that when $T = 1$ then $\tau = 0$ and thus vessels with constant convexity have zero tortuosity. Eq. 8 satisfies two assumptions: 1) the greater the number of twist, the more tortuous the vessel, and 2) the greater the amplitude of a turn curve, the greater the tortuosity associated with it.

B. *Image-level Tortuosity Measures*

In order to describe the tortuosity of an entire image, a method for combining vessel-level tortuosity measures needs to be defined. We analyzed 8 different image-level tortuosity measures, obtained by combining the measures proposed in III.A and computing the average on all the vessels in the image:



(a) (b) (c)

Figure 1. (a) Original image. (b) Value, expressed for each vessel sample n, of the local angle variation $\vartheta_n = arctg2(\vec{v}_n) - arctg2(\vec{v}_{n-1})$, used in the angle-based tortuosity formulation described in Sec. III-A. (c) Value, expressed for each vessel v, of the twist-base tortuosity measure τ_v, described in Sec. III-A. False colors in (b) and (c) from green to yellow to red represents respectively low to medium to high values.

$$IT_1 = \frac{1}{V}\sum_{v=1}^{V}\tau_v \qquad IT_2 = \frac{1}{V}\sum_{v=1}^{V}\tau_v D_v$$
$$IT_3 = \frac{1}{V}\sum_{v=1}^{V}\tau_v L_v \qquad IT_4 = \frac{1}{V}\sum_{v=1}^{V}\tau_v D_v L_v$$
$$IT_5 = \frac{1}{V}\sum_{v=1}^{V}\Phi_v \qquad IT_6 = \frac{1}{V}\sum_{v=1}^{V}\Phi d_v$$
$$IT_7 = \frac{1}{V}\sum_{v=1}^{V}\Phi_v L_v \qquad IT_8 = \frac{1}{V}\sum_{v=1}^{V}\Phi d_v L_v \tag{9}$$

where V is the number of vessels in the image. It is worth noting that all IT measures are linear with respect to tortuosity components (Φ and τ do not appear together), diameter components (same for D and d) and length. It is also worth mentioning that in the IT expressions in which the length of the vessel L_v appears (i.e., IT_3, IT_4, IT_7, IT_8), the vessel-level tortuosity measures cannot be considered tortuosity densities. In fact, in Eq. 6, 7, and 8 the term L_v always appear at denominator.

C. *Combination of Image-level Measures*

Lastly, we evaluated a linear weighted combination of the previously defined IT measures, with the aim of replicating the clinicians' assessment.

Linear regression analysis is used to understand which among the independent variables $x_{i,j}$ are linearly related to the dependent variable y_i by parameters β, save for the error ε:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,j} + \varepsilon_i$$
$$y = X\beta + \varepsilon \tag{10}$$

In our case the dependent variable is the ground truth ordering R_{GT} associated to image i, as defined in Sec. II, while the independent variables are the measures IT_j *(j=1,8)* on image i:

$$y_i = R_{GT}(i) \qquad x_{ij} = IT_j(i) \tag{11}$$

Ordinary least squares estimator, which minimizes the sum of squared residuals, have been employed to estimate the value of the unknown parameter vector β:

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{12}$$

Once $\hat{\beta}$ is obtained, from the automatic measures IT_j we can compute for any image the estimate $\hat{R}_{GT}$, which we propose as global image-level tortuosity index:

$$\hat{R}_{GT} = [1,\ IT_1,\ \dots,\ IT_8] \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_8 \end{bmatrix} \tag{13}$$

In order to assess the method's performance on unknown data, we used a leave-one-out technique: for each of the 20 images, when estimating the image tortuosity of the image i, $\hat{R}_{GT}(i)$, parameters $\hat{\beta}$ were estimated using only the remaining 19 images j, with $j \neq i$. With this approach we assure that the assessment of the method's prediction performance is correct, since each estimate is performed on data not used to determine the parameters.

We performed a Durbin-Watson statistical test [11] in order to detect the possible presence of autocorrelation in the residuals: d resulted to be equal to 1.97 ($0 \leq d \leq 4$, and $d = 2$ indicates no autocorrelation), and the p-value for the

TABLE II. SPEARMAN CORRELATION COEFFICIENTS FOR DIFFERENT TORTUOSITY EVALUATIONS

	Clinical graders			*ROP imaging experts*			*Automatic approach*		GT
	GRAD 1	GRAD 2	GRAD 3	EXP 1	EXP 2	EXP 3	best IT	$\hat{R}_{GT}$	
GRAD 1	1	0.94	0.97	0.85	0.88	0.89	0.89	0.97	0.98
GRAD 2		1	0.95	0.93	0.90	0.88	0.79	0.90	0.98
GRAD 3			1	0.89	0.92	0.89	0.88	0.93	0.98
EXP 1				1	0.89	0.86	0.71	0.82	0.89
EXP 2					1	0.85	0.84	0.88	0.91
EXP 3						1	0.73	0.86	0.88
best IT							1	0.92	0.87
$\hat{R}_{GT}$								1	**0.95**

"Best *IT*" is the best-performing single image-level tortuosity index, whereas $\hat{R}_{GT}$ is the combination of tortuosity indices. In bold the correlation of our method with GT..

test is 0.62, strongly rejecting the hypothesis of correlation in the residuals.

IV. RESULTS

Results are summarized in Table II, which shows the Spearman's rank correlation coefficient computed for manual and automatic tortuosity ordering. It assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman's correlation of +1 or −1 occurs when each of the variables is a perfectly monotone function of the other.

Average correlation among clinical graders (GRADs) is 0.96 and among experts (EXPs) is 0.87. Average correlation between GRADs and EXPs is 0.89. Correlation with ground truth (GT) is on average 0.98 for GRADs and 0.89 for EXPs.

The best single *IT* has a correlation with GT of 0.87 and an average correlation of 0.85 with GRADs and 0.76 with EXPs. Our proposed combination measure $\hat{R}_{GT}$ has a correlation of 0.95 with GT and an average correlation of 0.93 and 0.85 with GRADs and EXPs respectively.

V. DISCUSSION AND CONCLUSION

The problem of defining clinical tortuosity is not trivial. Tortuosity definitions that have been proposed in literature are mainly vessel-level, as well as the experiments that test their consistency in controlled examples of tortuosity (e.g., sinusoidal, helical, etc.) [8]. There is yet no agreement about what graders/experts mean when evaluating tortuosity, both at vessel and at image level.

In this work we addressed the problem of reproducing the clinical graders' perception of tortuosity. To provide a more sound assessment of the proposed method, we involved several experts at different levels: clinical graders, whose combined assessment was also used as the reference ground truth, and non clinical ROP imaging experts. In this way, we were able to appreciate both the inter- and intra-reproducibility of the two groups of evaluators, as well as the level of performance achieved by the proposed method.

Results highlights that the proposed measure $\widehat{\mathrm{R}}_{\mathrm{GT}}$ correlates very well with the GT, at the same level as clinical graders do and better than ROP imaging experts. At the same time, $\hat{R}_{GT}$ correlates with the clinical graders' ordering as well as they do among themselves, and better than how the ROP imaging experts do (with respect to both themselves and clinical graders).

It is finally worth noting that although $\hat{R}_{GT}$ is an estimation of the R_{GT} ground truth reference (which is, in turn, a set of ranking values between 1 and 20), it can be viewed as a sheer tortuosity measurement value, capable to assess the tortuosity of any image in a quantitative and objective way. The robustness of the system has been evaluated by assessing its performance only analyzing unknown images, by means of the leave-one-out validation approach.

When paired with an automatic vessel tracing technique (under development), it may provide a completely automated tool for the reliable quantitative estimation of vascular tortuosity in ROP images.

ACKNOWLEDGEMENTS

We would like to thank M. Chiang (Oregon Health and Science University, USA), R.V.P. Chan (Weill Cornell Medical College, USA), T.C. Lee (Children's Hospital of Los Angeles, USA), B. Linder and L. MacKeen (Clarity Medical Systems, USA), and D. Fiorin (University of Padova) for having provided images and manual assessment.

REFERENCES

[1] The Committee for the Classification of Retinopathy of Prematurity. "An international classification of retinopathy of prematurity". *Arch Ophthalmol*, Vol. 102(8), pp. 1130–1134, 1984.

[2] Committee for the Classification of Retinopathy of Prematurity, "The International Classification of Retinopathy of Prematurity revisited". *Arch Ophthalmol*, Vol. 123 (7), pp. 991–999, 2005.

[3] Slidsborg C et al., "Experts do not agree when to treat retinopathy of prematurity based on plus disease", *Br J Ophthalmol*, Vol 96, pp. 549-553, 2012.

[4] Wallace DK, Zhao Z, Freedman SF. "A pilot study using ROPtool to quantify plus disease in retinopathy of prematurity". *J AAPOS*, Vol. 11(4), pp. 381–387, 2007.

[5] Swanson C et al. "Semiautomated computer analysis of vessel growth in preterm infants without and with ROP". *Br J Ophthalmol*, Vol. 87(12), pp. 1474–1477, 2003.

[6] Gelman R et al. "Diagnosis of plus disease in retinopathy of prematurity using Retinal Image multi Scale Analysis". *Invest Ophthalmol Vis Sci*, Vol. 46(12), pp. 4734–4738, 2005.

[7] Poletti E et al., "Automatic vessel segmentation in wide-field retina images of infants with Retinopathy of Prematurity". in Proc. IEEE EMBC 2011, pp. 3954-3957, 2011.

[8] Grisan E, Foracchia M, Ruggeri A, "A Novel Method for the Automatic Grading of Retinal Vessel Tortuosity". *IEEE Trans Med Imag*, Vol. 27(3), pp. 310-319, 2008.

[9] Wilson, CM et al., "Computerized Analysis of Retinal Vessel Width and Tortuosity in Premature Infants". *Invest Ophthalmol Vis Sci*,Vol. 49(8), pp. 3577-3585, 2008.

[10] Chandrinos KV et al., "Image processing techniques for the quantification of atherosclerotic changes", in Proc. MEDICON98, 1998.

[11] Durbin J, and Watson GS, "Testing for Serial Correlation in Least Squares Regression, I". *Biometrika*, Vol. 37, pp. 409–428, 1950.