UNIVERSITA' DEGLI STUDI DI PADOVA

CORSO DI LAUREA IN BIONGEGNERIA

Global tortuosity estimation of retinal vessel network in wide-field images taken from infants with retinopathy of prematurity.

Laureando:
Giovanni OMETTO

Relatore:
Prof. Alfredo RUGGERI

Correlatore:
Dott. Enea POLETTI

13 Marzo 2012

Contents

Introduction

ROP (retinopathy of prematurity) represents one of the leading causes of blindness in infants in developed and emergent countries. It is fundamental being able to detect ROP promptly in order to diagnose and treat it early, in several cases solving the problem or avoiding serious complications. This disease consists in an abnormal development of retinal vasculature which doesn't reach the periphery of the retina and it is characterized by an increased vascular dilatation and abnormal changes of other parameters such as tortuosity and dilatation. These vascular changes are usually only qualitatively assessed by visual inspection of retinal blood vessels in RetCam images. Ophtalmologists and experts have to review a large number of images so the evaluation of these parameters requires frequent, time consuming and subjective intervention by them. 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 a 20-year-old reference photograph (ICROP, [7]) 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,[8]). Moreover a recent study it has shown that ophtamologists' agreement on diagnostic classification is poor, especially when more than two severity levels of classification are involved [5]). For these reasons authors from [5] conclude that the possibility of having an automatic device which provides useful information to the clinician about the actual state of the eye under exam would represent an important aid and that it would improve the quality of care of premature infants. The realization of a software that automatically detects the stage of ROP basing on RetCam images, would standardize and make more objective the way to proceed and would give more information to the ophtalmologists that could easier decide how to treat the pathology before that complex interventions as vitrectomy or scleral buckling were necessary. Several suites of algorithms and their implementation as prototype computer programs [10, 6, 3, 4]have been developed for the tracking of vessels and the following estimation of vascular tortuosity and dilation to classify the images in order of seriousness. Yet none of them has been proved to be a reliable clinical tool. The problem is that there is not yet agreement about

what experts mean by tortuosity and hence no way to decide which of many measures fit best with the evaluations of ophtamologists. In this study we propose a new approach for the tortuosity evaluation of ROP images. We refer to tortuosity as a property of both vessels (vessel-wise level) and images (image-wise level). Global image tortuosity at the image-wise level is defined as combination of vessel-wise measures. A regression approach on the reference ground truth at disposal has been employed to combine vessel-wise measures in order to replicate the evaluations by experts [11].

Chapter 1

Materials

1.1 RetCam Images

In order to treat the baby in the best way its fundamental an attentive monitoring of the retinal condition. The use of RetCam allows to obtain images of the eye that can be collected and compared leading to a more effective and reliable process of screening, diagnosis or follow-up of the patient and a more accurate way to analyze and categorize the images from the experts who also can collect, register and interchange them [1]. The hallmarks of ROP are usually located in the most peripheral area of the eye. In this area retinal blood vessels grow during the last weeks of gestation. The digital technology enables to obtain images that give information about the entire extension of the lesion, because of the wide RetCam field of view. RetCam images have a wide field of view (120°-130°) whereas fundus camera images usually have a field of view which doesnt exceed 60°. These images are frames selected from videos of the retinal fundus. The selection is carried out with the aim of having images with minimum movement artifacts and maximum focus and contrast. Despite this, RetCam images are often affected by a non-uniform illumination, low contrast and they can show fuzzy areas and the presence of vessels not belonging to the retina due to the extreme slightness of the retina in newborns. RetCam images present in general a low level of resolution that makes their automatic analysis challenging. In this work a set of 20 RetCam images 640x480 pixel resolution bitmap file format provided by Clarity Medical Systems Inc., CA, USA has been used. These images are taken from the retina of newborns showing ROP at different stages. All images have a black area which will not be considered in all the process and borders the circular area of interest. See Figure 1.1. This set that has been reviewed by 3 clinical graders of the Department of Ophthalmology, Columbia University, New York, USA, M. Chiang, R. Chan, T. Lee and 3 image experts of the Clarity Medical Systems, B. Linder, X. Mackeen, and D. Fiorin . From now on we will refer to the former ones as GRAD 1, GRAD 2 and GRAD 3, and to the latter ones as

EXP 1, EXP 2, and EXP 3.

The same set of images has been used to create a new set of black and white images which can be processed by the tortuosity and dilatation estimation algorithm.

1.2 Manual tracking of vessels

The algorithm needs black (”BLACK”, ASCII code #000000) and white (”WHITE”, ASCII code #FFFFFF) images to run. Input images must have the same size of the relative RetCam image and must have a white pixel where the same pixel in the RetCam image represent a vessel. Black pixels identify areas where no vessels are clearly shown. Optic disc, where it is really difficult to identify the directions of the vessels due to the many overlays, can be filled with black pixels and will not be considered.For each of the 20 images, a manual segmentation of the vessels was provided by the authors of this work under the supervision of GRAD 1. The manual tracing has been carried out and used for tortuosity estimation, in order to focus on tortuosity formulation alone and avoid problems coming from possible errors in automatic computerized vessel tracing. This is a crucial task because any mistake will affect the quality of the results. To create the new set a graphics editing program has been used. Each one of the 20 RetCam images has been set as the background layer and a new black layer was created over it. Setting a transparency level to the new layer and using a graphic tablet, vessels in the background level have been traced in white, pixel by pixel, on the black layer. The resulting layer has been saved as a TIF image file with no transparency, see Figure 1.2.

1.2.1 Manual tracking of veins and arteries

Similarly to the previous process a new type of images can be created to obtain another tortuosity and dilatation estimation that takes in account also the type of vessel (arteries and veins). From the black and white set of images, arteries and veins can be detected and marked respectively with red (”RED”, ASCII code #FF0000) and blue (”BLUE”, ASCII code #0000FF) pixels instead of white. Pixels corresponding to an overlay of these two type of vessels can be marked in green (”LIME”, ASCII code #00FF00). Optic disc, where it is really difficult to identify the type of the vessels due to the many overlays, can be filled with yellow pixels (”YELLOW”, ASCII code #FFFF00) and will not be considered. In this case the estimation will be the result of a two stage process for each image. In first stage the estimation algorithm will process a B/W image including just the veins, in second stage another B/W image including only arteries. The first-stage image can be immediately obtained with this color conversion: BLACK - BLACK, BLUE - WHITE, LIME - WHITE, RED - BLACK, YELLOW - BLACK. Same conversion with BLUE and RED swapped will result in the second-stage image.

This solution can be considered when a reliable automatic vessel tracing technique will be available.

1.3 Reviews by experts

The 20 image set has been independently ordered by increasing tortuosity by each of the 3 graders and each of the 3 experts. Image files have been sent to the clinicians and experts named in a random order, not to give them any information except from the images themselves. Since the task could have been time-demanding and to avoid mistakes, together with the images experts received a small custom software called TortSort to make the grading easier and quicker. The gradings are shown in Table 1.1. Lists are ordered on the ranking by GRAD 1.

1.3.1 TorTsorT

TorTsorT (available to download at http://bioimlab.dei.unipd.it) is a small program written in Visual Basic programming language. It's based on a comparison sorting algorithm. The software performs a grading based on user's choices. The idea behind TorTsorT is that it is difficult to accurately and efficiently sort a numerous set of image by a certain property,while it is much easier to sort only two images at a time by that property. TorTsorT takes RetCam images from a specified folder and displays pairs of them as shown in Picture . For each pair of images displayed, user will be asked to pick the one showing the higher perceived tortuosity (see Figure 1.3). On background a merge-sort algorithm is selecting the pairs of images and recording choices from the user in order to rank the set. Merge sort is a comparison-based divide and conquer sorting algorithm that ensures average and worst-case performance of $O(n \log(n))$. Having $n = 20$ images to be ordered the average number of pair comparisons required is 60. For this number of comparisons it takes few minutes for the user to complete the task and obtain a graded set.

1.3.2 Ground Truth

Eventually we obtained six different order lists, each expressing, for each image i, the ranking $R(i) \in \{1, ..., 20\}$. In order to have a gold standard reference, a ground truth vector as been defined as $R_{GT}(i) = ranking\Big(mean\big(R_{GRAD1}(i), R_{GRAD2}(i), R_{GRAD3}(i)\big)\Big)$. Please see Table 1.1.

R_{GRAD1}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
R_{GRAD2}	6	3	4	1	2	7	5	9	8	10	11	12	13	17	14	18	15	16	19	20
R_{GRAD3}	4	3	2	1	6	5	7	10	11	9	8	12	13	14	15	16	17	18	19	20
R_{EXP1}	9	4	6	2	1	10	3	11	8	7	5	12	14	15	13	18	17	16	19	20
R_{EXP2}	8	4	1	3	2	7	9	6	11	10	5	12	16	14	13	19	15	20	18	17
R_{EXP3}	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

Table 1.1: The 20 image set were independently ordered by increasing tortuosity by each of the 3 graders (GRAD 1, GRAD 2, GRAD 3) and each of the 3 experts (EXP 1, EXP 2, EXP 3). $R(i) \in \{1, ..., 20\}$ expresses the rank of an image in the list (i.e., the 1st image according to GRAD 1 is the 6th according to GRAD 2). Lists are ordered on the ranking by GRAD 1.

1.3.3 Second re-evaluation and Disease Severity Grading

A second re-evaluation has been asked them to assess also the intra-expert reproducibility of tortuosity assessment. To ensure as much as possible a blind re-evaluation the names of image files have been randomized once again. GRAD 1 has been able to provide also this second grading ($R_{GRAD1}*$, see 3rd row in Table 1.2).

1.3.4 Diagnostic classification

An automatic estimation of the image tortuosity and dilatation will be expected to return a severity condition of the disease. Therefore output gradings must be consistent with the diagnostic classification. In a diagnostic classification each of the 20 images are graded into one of the three severity of ROP: plus disease, pre-plus disease, and neither. To make sure results fulfill this condition a diagnostic classification is required. GRAD 1 provided this classification grading 4 images as plus disease, 6 as pre-plus disease and the remaining 10 as neither (DIAG 1, see last row in Table 1.2).

R_{GT}	4	2	3	1	5	6	7	8	9	10	11	12	13	15	14	17	16	18	19	20
R_{GRAD1}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$R_{GRAD1}*$	3	2	4	1	6	5	7	9	10	8	11	12	14	13	16	15	17	19	18	20
DIAG 1	n	n	n	n	n	n	n	n	n	n	pp	pp	pp	pp	pop	p	p	pp	p	p

Table 1.2: Ground truth in first row, 1st and 2nd grading (R_{GRAD1} and $R_{GRAD1}*$) and diagnostic classification (DIAG 1) by GRAD 1 (p=plus, pp=pre-plus, n=neither).

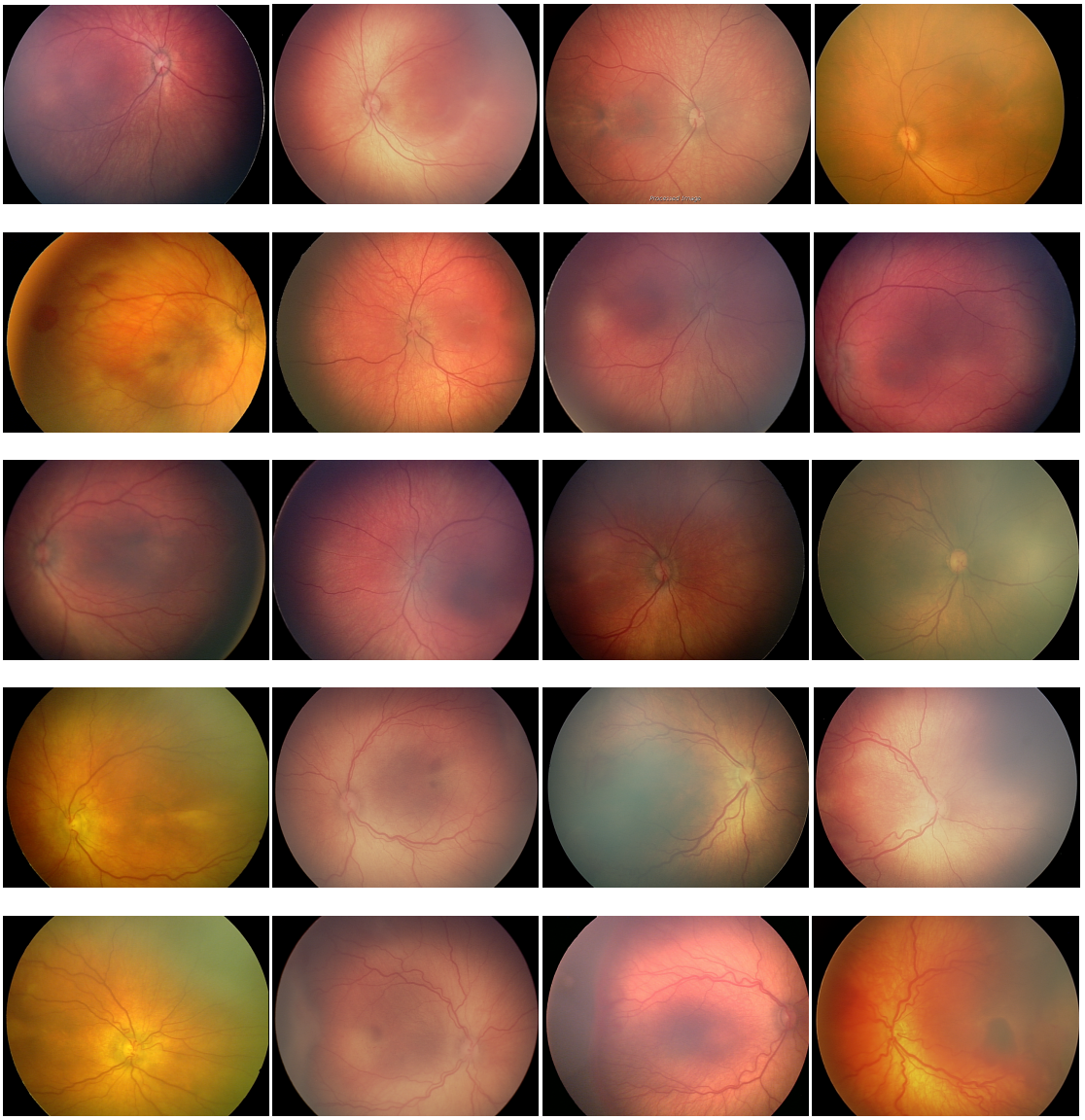

Figure 1.1: The set of 20 RetCam Images.

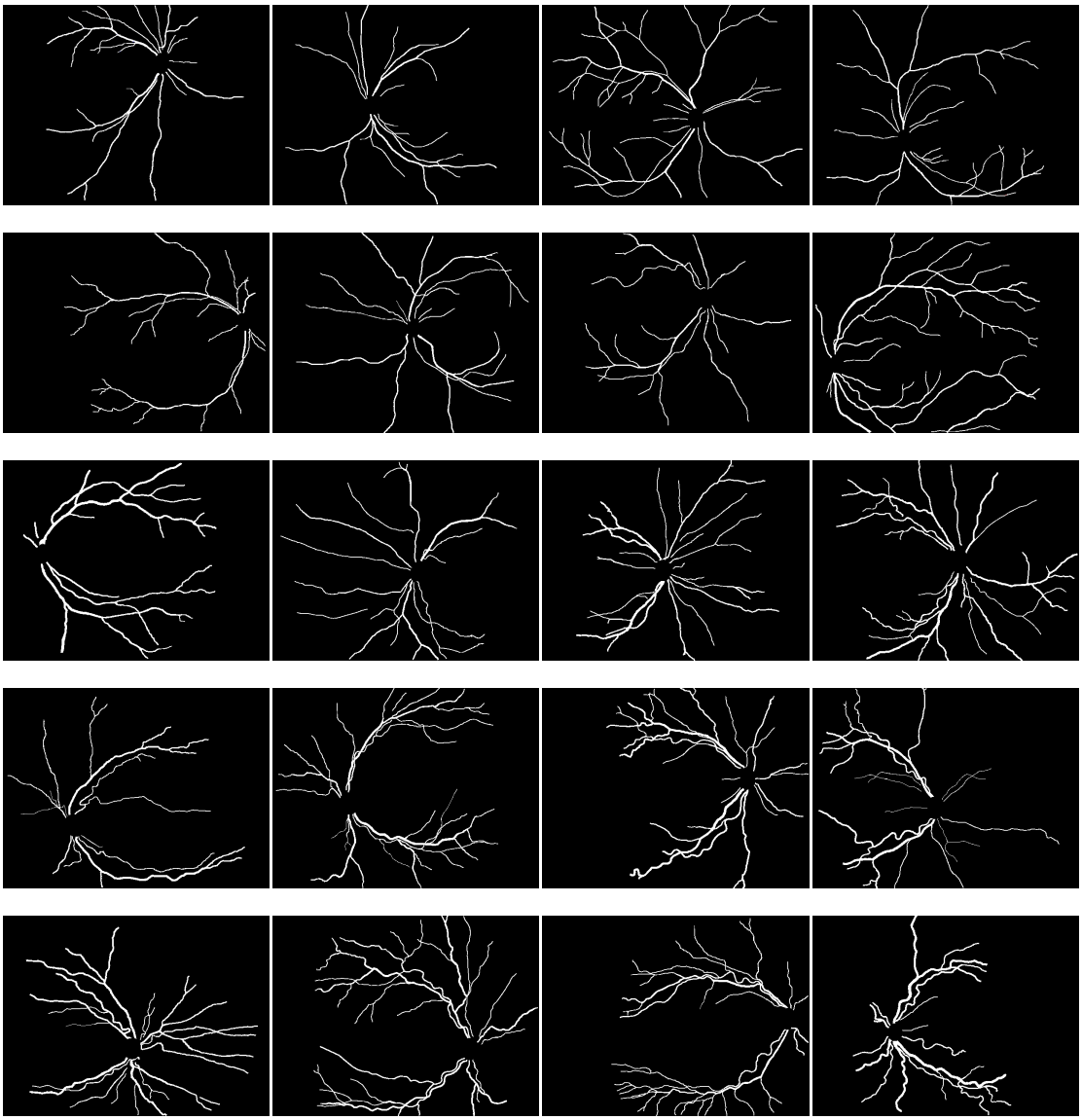

Figure 1.2: Manually tracked vessels for the set of RetCam Images.

Figure 1.3: TorTsorT software running, images are displayed pair by pair.

Chapter 2

Methods

The rationale of our approach is to find the best combination of vessel-related measures (tortuosity, diameters, length) that provides an image-A linear weighted combination of the defined global measures has been defined with the aim of both determining the ones that are most informative and replicating the clinical measurements.

2.1 Automatic grading algorithm

To test if an automatic process can give reliable results for diagnostic purposes measures of tortuosity and dilatation have been proposed and developed in Matlab and combinations of them have been tested to reveal a grading function that is consistent with the gradings from the experts. This means assessing whether or not output indexes of a grading resulting from the algorithm could replace the experts providing a compatible diagnosis in the ideal setting of 'exactly' traced vessels. This is a necessary step to provide a final, totally automatic tool. Satisfying results would mean that this algorithm, paired with an automatic vessel tracing technique still under development, can provide that tool.

2.1.1 Vessel-wise Measures

The various vessel-wise tortuosity indexes have been proposed in the literature (see e.g., [9]). They are estimated by using geometric quantities, but not the width of the vessel. Let the vessel v be an ordered set of triplets $v = \{(x_n, y_n, d_n) : n \in \{1, ..., N\}\}$ 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 samples. Starting with that definition of vessel we can devise the following vessel-wise measures.

Vessel Diameter

We assume here that the diameter of a vessel is the average of all its diameters at the sample points.

$$D_v = \frac{1}{N}\sum_{n=1}^{N} d_n \tag{2.1}$$

It is worth noticing that, even if this measure is not tortuosity-related at the vessel-wise level, it could potentially turn out to be tortuosity-significant at the global image-wise level.

Vessel Length

In order to have a measure that is invariant to the number of samples and to the orientation of the vessel, vessel length is defined as:

$$L_v = \sum_{n=2}^{N} \sqrt{(x_n - x_{n-1})^2 + (y_n - y_{n-1})^2} \tag{2.2}$$

Angle-based Tortuosity

A measure of tortuosity based on the local direction variation of the vessel has been proposed [2]), in which the average of the angles between consecutive vessel samples is computed. However, in that formulation angles can vary only between $-\pi/2$ and $\pi/2$, and no information about vessel width has been taken into account. We propose here a new, improved version of the angle-based tortuosity measure. Being (x_n, y_n) the coordinates of the n^{th} sample point of the vessel, the vector going from the n^{th} sample to the $n^{th} - 1$ sample center points is:

$$\vec{v}_n = (x_n - x_{n-1}, y_n - y_{n-1}) \tag{2.3}$$

and describes the local vessel direction. The local angle variation is then defined as

$$\vartheta_n = arctg2(\vec{v}_n) \tag{2.4}$$

$$arctg2(x, y) = \begin{cases} cos^{-1}\frac{x}{\sqrt{(x^2+y^2)}}, & y \geq 0 \\ -cos^{-1}\frac{x}{\sqrt{(x^2+y^2)}}, & y < 0 \end{cases} \tag{2.5}$$

where $arctg2$ is a function that takes values between $-\pi$ and π, representing the counterclockwise angle, measured in radians, between the positive x axis, and the point

(x, y). We define the vessel-wise angle-based tortuosity of the vessel v as:

$$\phi_v = \frac{1}{L_v} \sum_{n=2}^{N} \vartheta_n^2 \tag{2.6}$$

This tortuosity measure has a dimension of $1/length$ and thus may be interpreted as a tortuosity density, allowing its comparison on vessels of different length.

Caliber-weighted Angle-based Tortuosity

At variance with Eq. 2.6, a formulation that takes into account the vessel diameter of each sample has also been defined:

$$\phi d_v = \frac{1}{L_v} \sum_{n=2}^{N} \vartheta_n^2 d_n \tag{2.7}$$

Twist-based Tortuosity

Work [9] propose a vessel tortuosity measure based on twists (or inflections), which are located at samples where the curvature sign changes. The turn curve v_t is defined as that portion of a vessel v located between two consecutive twists. The twist-based tortuosity definition fulfills two assumptions: 1) the greater the number of twist, the more tortuous the vessel and 2) the greater the amplitude (maximum distance of the curve from the underlying chord) of a turn curve, the greater the tortuosity associated with it. The measure requires at first a partitioning of the vessel v into T turn curves:

$$v = v_1 \oplus v_2 \oplus ... \oplus v_T \tag{2.8}$$

The twist-based tortuosity τ_v of the vessel v is defined as:

$$\tau_v = \frac{T-1}{T} \frac{1}{L_v} \sum_{t=1}^{I} \left[\frac{L_{v_t}}{L_{X_{v_t}}} - 1 \right] \tag{2.9}$$

where the chord length L_{X_v} is defined as the distance between the curve end points $L_{X_v} = \|(x_1, y_1) - (x_N, y_N)\|$. Also this tortuosity measure has a dimension of $1/length$. It is worth noting that when $T = 1$ then $\tau = 0$ and thus vessels with constant convexity have zero tortuosity.

2.1.2 Image-wise Measure

In order to catch the tortuosity of an entire image, measures proposes in 2.1.1 have been randomly aggregated and averaged on all the vessels in the image to obtain a number of

image-wise tortuosity measures. We will call them IT (image tortuosity) and the most revealing results are:

$$IT_1 = \frac{1}{V}\sum_{V=1}^{V}\tau_v \tag{2.10}$$

$$IT_2 = \frac{1}{V}\sum_{V=1}^{V}\tau_v D_v \tag{2.11}$$

$$IT_3 = \frac{1}{V}\sum_{V=1}^{V}\tau_v L_v \tag{2.12}$$

$$IT_4 = \frac{1}{V}\sum_{V=1}^{V}\tau_v D_v L_v \tag{2.13}$$

$$IT_5 = \frac{1}{V}\sum_{V=1}^{V}\phi_v \tag{2.14}$$

$$IT_6 = \frac{1}{V}\sum_{V=1}^{V}\phi d_v \tag{2.15}$$

$$IT_7 = \frac{1}{V}\sum_{V=1}^{V}\phi_v L_v \tag{2.16}$$

$$IT_8 = \frac{1}{V}\sum_{V=1}^{V}\phi d_v L_v \tag{2.17}$$

It is worth noticing that the IT measures are linear with respect to the tortuosity components (ϕ and τ do not appear together), to the diameter component (same for D and d) and to the length. It is also worth mentioning that in the IT formulations in which the length of the vessel L_v appears (i.e., Eq. 2.12, 2.13, 2.16, 2.17), the vessel-wise tortuosity measures cannot be considered tortuosity densities. In fact, in Eq. 2.6, 2.7, and 2.9 the term L_v always appeared at denominator.

2.1.3 Image-wise Measure Combination

From the first attempts to compute IT measures on the images to obtain a disease severity grading the research has moved to a mathematical approach where the proposed measures have been combined and weighted together according to the solution of a

regression problem aiming for the most informative and most clinically consistent combination. A linear regression analysis is used here 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:

$$y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_p x_{i,j} + \varepsilon_i \quad \Rightarrow \quad y = X\beta + \varepsilon \tag{2.18}$$

Here the dependant variable are the ground truth R_{GT} associated to the image i, while the independent one are replaced by the j^{th} measure IT_j on the image i:

$$y_i = R_{GT}(i) \qquad x_{ij} = IT_j(i) \tag{2.19}$$

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

$$\hat{\beta} = (X^T X))^{-1} X^T y \tag{2.20}$$

Once $\hat{\beta}$ are estimated, it is also possible to estimate the clinical graders tortuosity $\hat{R}_{GT}(i)$ on the basis of the automatic measures $IT_j(i)$:

$$\hat{R}_{GT}(i) = [1, IT_1(i)), ..., IT_8(i))] \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_8 \end{bmatrix} \tag{2.21}$$

In order to assess the methods performance on unknown data, a leave-one-out technique has been used: when estimating the image tortuosity of the image i, $\hat{IT}(i)$, parameters $\hat{\beta}$ were estimated using only the others 19 images j, with $j \neq i$. With this approach we assure that the parameters are not over-fitted to the data, since each estimate is done on data not used to determine the parameters. In this case it is safe to compute the Spearmans correlation coefficient between $\hat{IT}$ and R_{GT}. A Durbin- Watson statistic test has been also performed in order to detect the possible presence of autocorrelation in the residuals: $d = 1.97$ ($0 \leq d \leq 4$, and $d = 2$ indicates no autocorrelation), and the p-value for the test is 0.6154, strongly rejecting the hypothesis of correlation in the residuals.

2.2 Comparing gradings

To compare the gradings, a simple mismatch count would not be a good index of performances. What is most important for the diagnostic aim is not having the images ranked exactly as in the gradings by expert, but having a grading that is consistent with them. This means that an image very tortuous on experts' opinion should have a very high ranking position and should not necessarily have the same ranking. The more this is true the better the relationship between the two rankings can be described using a monotonic function. The Spearmans rank correlation coefficient can assess this and has been used to compare manual and automatic orderings.

2.2.1 Spearmans rank correlation coefficient

Spearman's rho (ρ) measures the statistical dependence between two variables. 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 correlation of $+1$ or -1 occurs when each of the variables is a perfect monotone function of the other. Two identical rankings would have $\rho = +1$, two opposite rankings would have a $\rho = -1$, while a $\rho \approx 0$ means no correlation between the rankings. The Spearman correlation coefficient is defined as the Pearson correlation coefficient between two ranked variables. For a sample of size n, the n raw scores R_a, R_b are converted to ranks r_{a_i}, r_{b_i}, and ρ is computed from these:

$$\rho = \frac{\sum_i (r_{a_i} - \bar{r_a})(r_{b_i} - \bar{r_b})}{\sqrt{\sum_i (r_{a_i} - \bar{r_a})^2 \sum_i (r_{b_i} - \bar{r_b})^2}} \tag{2.22}$$

Tied values are assigned a rank equal to the average of their positions in the ascending order of the values. With these gradings ties are known to be absent and a simpler procedure can be used to calculate rho. Differences $d_i = r_{a_i} - r_{b_i}$ between the ranks of each observation on the two variables are calculated, and ρ is given by:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}. \tag{2.23}$$

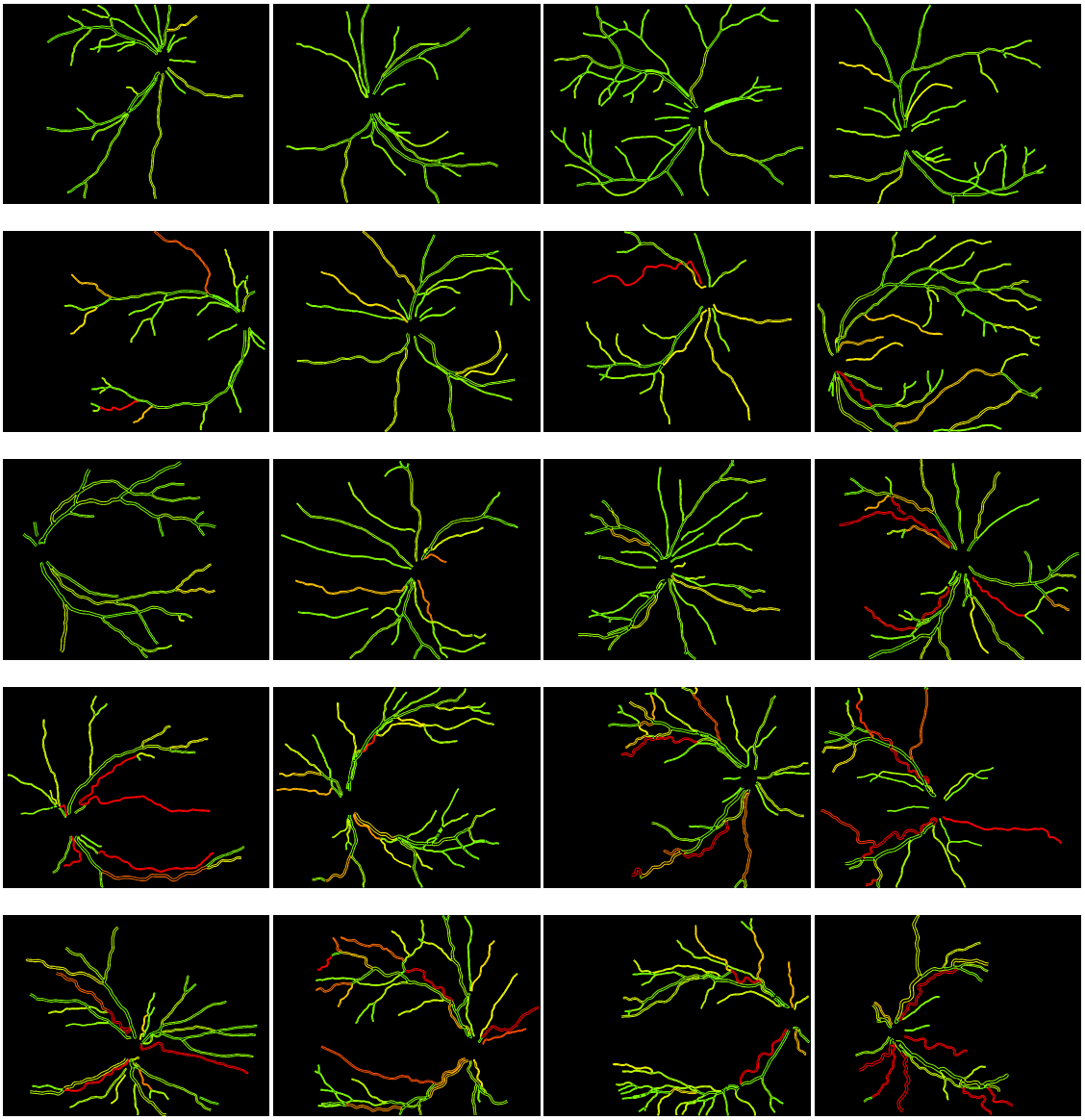

Figure 2.1: τ_v: Semantic Tortuosity for the set of RetCam Images. The value is shown in a color scale going from the lowest value in green to the highest in red.

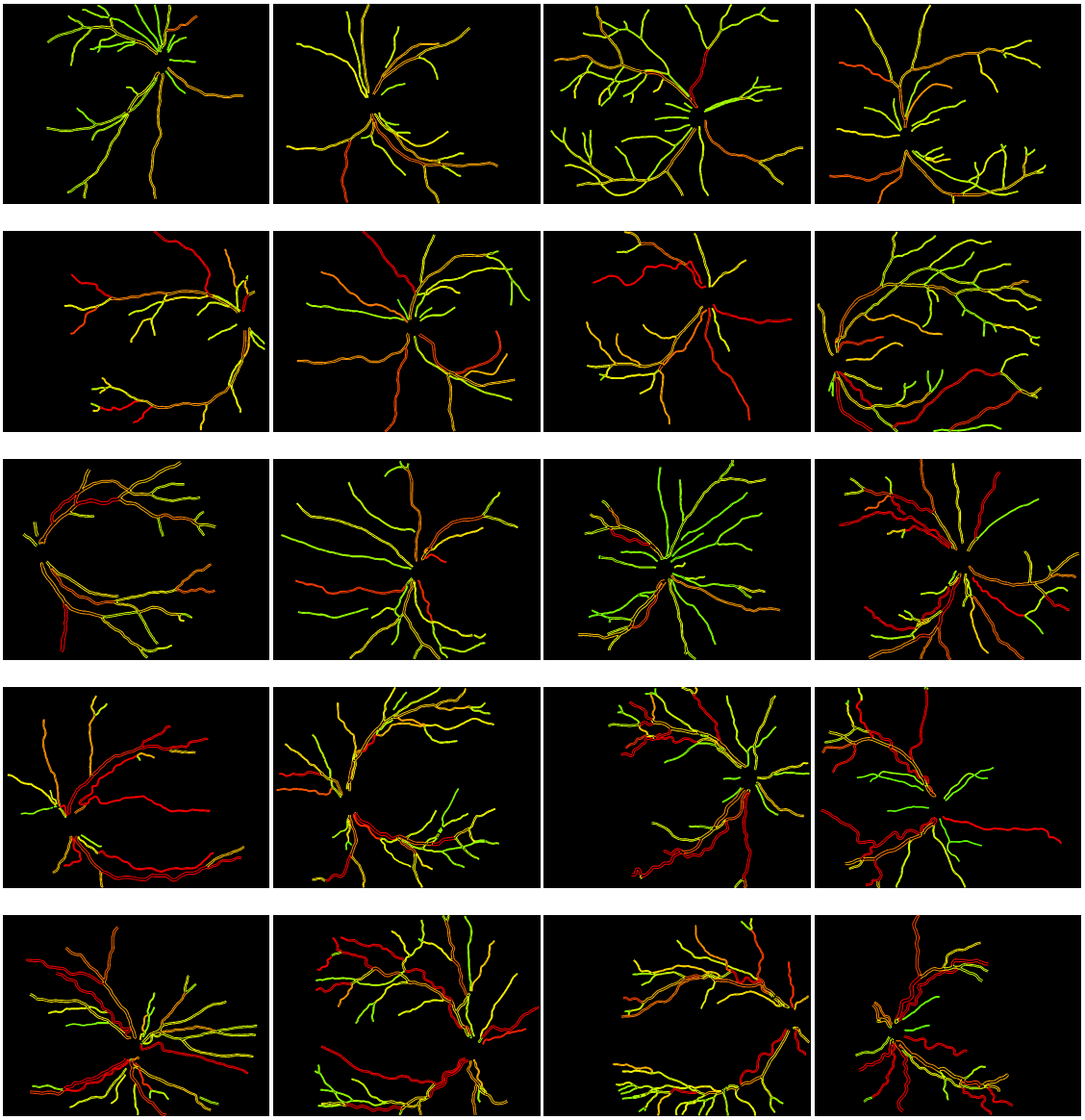

Figure 2.2: $\tau_v D_v$: Diameter-weighted Semantic Tortuosity for the set of RetCam Images. The value is shown in a color scale going from the lowest value in green to the highest in red.

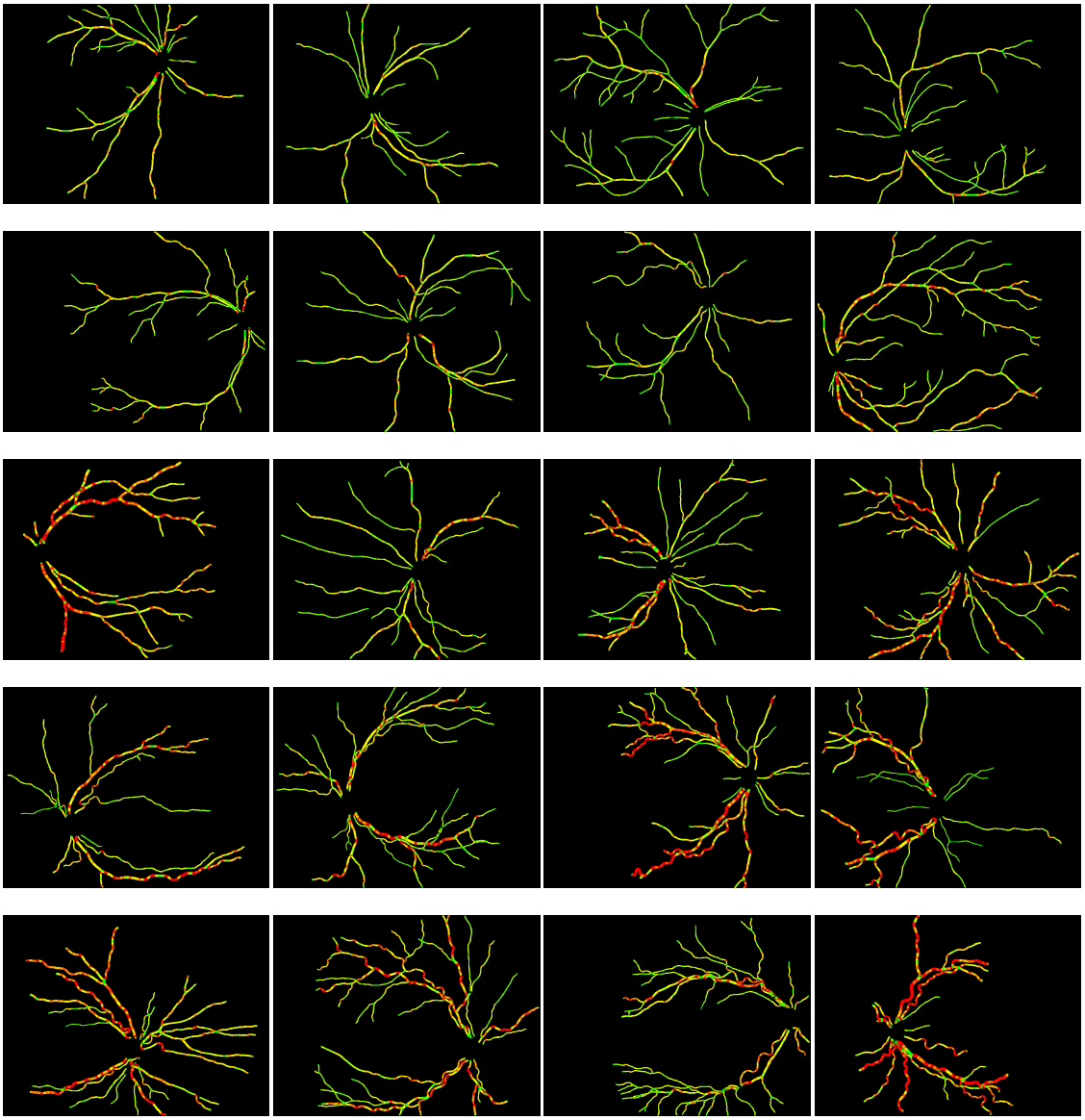

Figure 2.3: ϕ: Angle-based Tortuosity value for the set of RetCam Images. The value is shown in a color scale going from the lowest value in green to the highest in red.

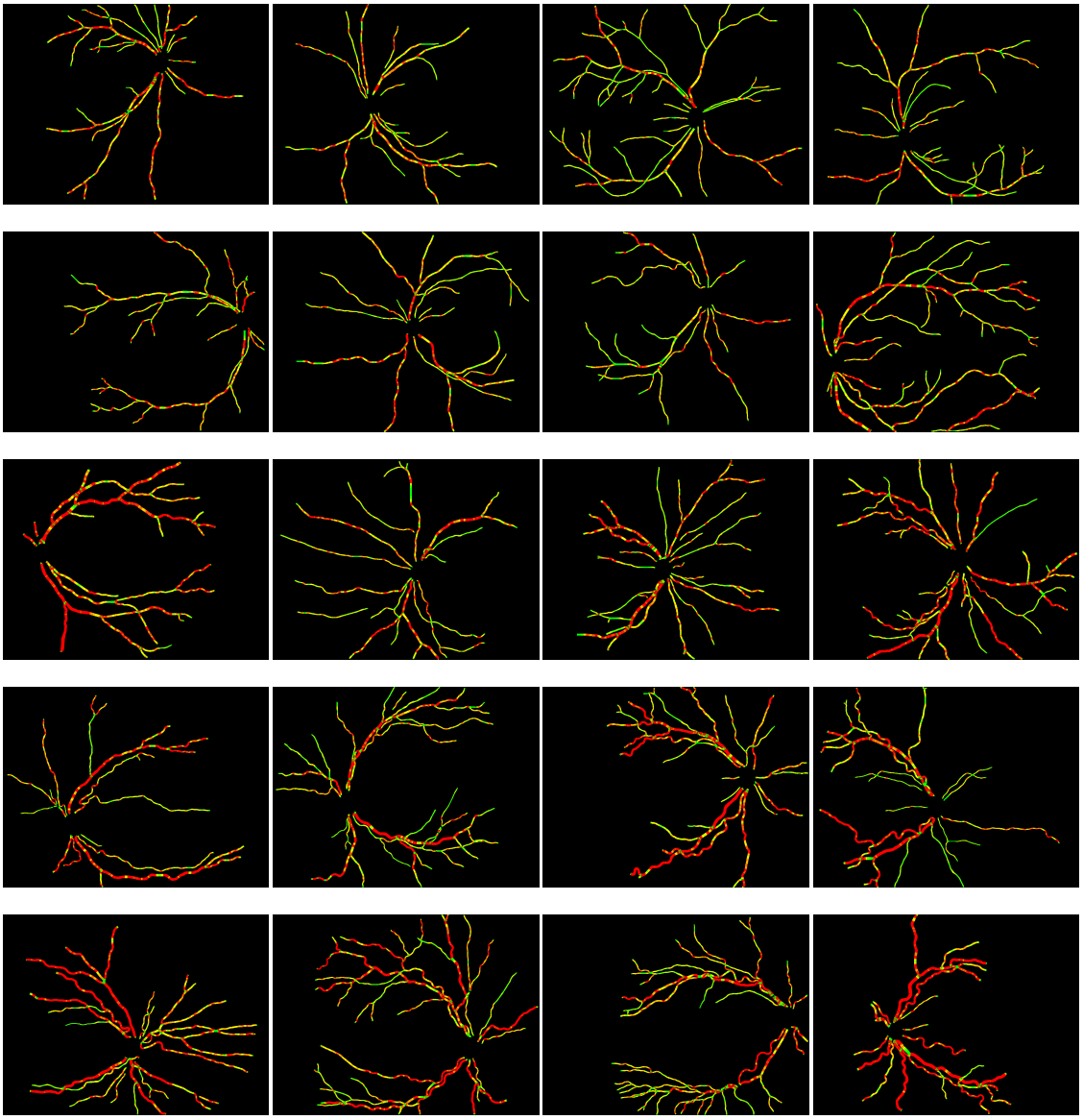

Figure 2.4: $\phi d_v L_v$: Caliber-weighted Angle-based Tortuosity value multiplied by vessel Length for the set of RetCam Images. The value is shown in a color scale going from the lowest value in green to the highest in red.

Chapter 3

Results

Results are summarized in Table 3.1 and Fig. 3.1. Table 3.1 shows the Spearmans 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 Spearmans correlation of $+1$ or -1 occurs when each of the variables is a perfect monotone function of the other. Average inter-correlation amongst clinical graders (GRADs) is 0.96; amongst experts (EXPs) is 0.87. Average correlation between GRADs and EXPs is 0.89. $\hat{R}_{GT}$ correlation is respectively 0.93 with GRADs and 0.85 with EXPs. As the ground truth ordering is concerned (the ranking of the mean of the GRADs rankings), correlation with GRADs is in average 0.98, with EXP is in average 0.89, and with $\hat{R}_{GT}$ is 0.95. In Fig. 3.1, we plot the $\hat{R}_{GT}$ estimated value of every image (y-axis) against the classification of the images in healthy, pre-plus and plus disease provided by GRAD 1 (x-axis). We can see that if the value of $\hat{R}_{GT}$ were used to classify images, only 1 image would be misclassified as false positive. In Fig 3.2, 3.3 and 3.4 the same comparisons expressed with the Spearmans rank correlation coefficient value in Table 3.1 are shown in plots to highlight the monotonic relationship between the most correlated rankings. Each point's X-coordinate is the index from one grading while Y-coordinate is the index from the compared grading. Each point is displayed with a green square when the indexed image shows a healthy condition according to GRAD1 diagnostic classification, with black circle if it is a pre-plus condition, and with a red cross in case of plus condition.

	Clinical Graders				*ROP image experts*				
	GRAD 1	GRAD 1*	GRAD 2	GRAD 3	EXP 1	EXP 2	EXP 3	$\hat{R}_{GT}$	GT
GRAD 1	1	0.98	0.94	0.97	0.85	0.88	0.89	0.97	0.98
GRAD 1*		1	0.94	0.98	0.86	0.88	0.87	0.96	0.98
GRAD 2			1	0.95	0.93	0.90	0.88	0.90	0.98
GRAD 3				1	0.89	0.92	0.89	0.93	0.98
EXP 1					1	0.89	0.86	0.82	0.89
EXP 2						1	0.85	0.88	0.91
EXP 3							1	0.86	0.88
$\hat{R}_{GT}$								1	**0.95**
GT									1

Table 3.1: The table reports the Spearman Correlation Coefficient for each pair of ordering. $\hat{R}_{GT}$ is the combination of tortuosity indices. In bold the final result: the correlation of our proposed method with the ground truth reference.

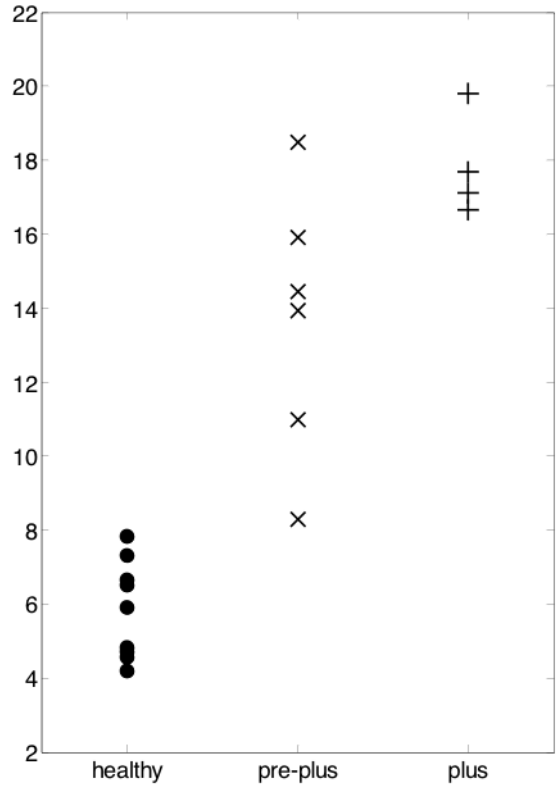

Figure 3.1: X-axis: classification of the images in healthy, pre-plus and plus disease as provided by GRAD 1. Y-axis: the corresponding estimated R_{GT} index.

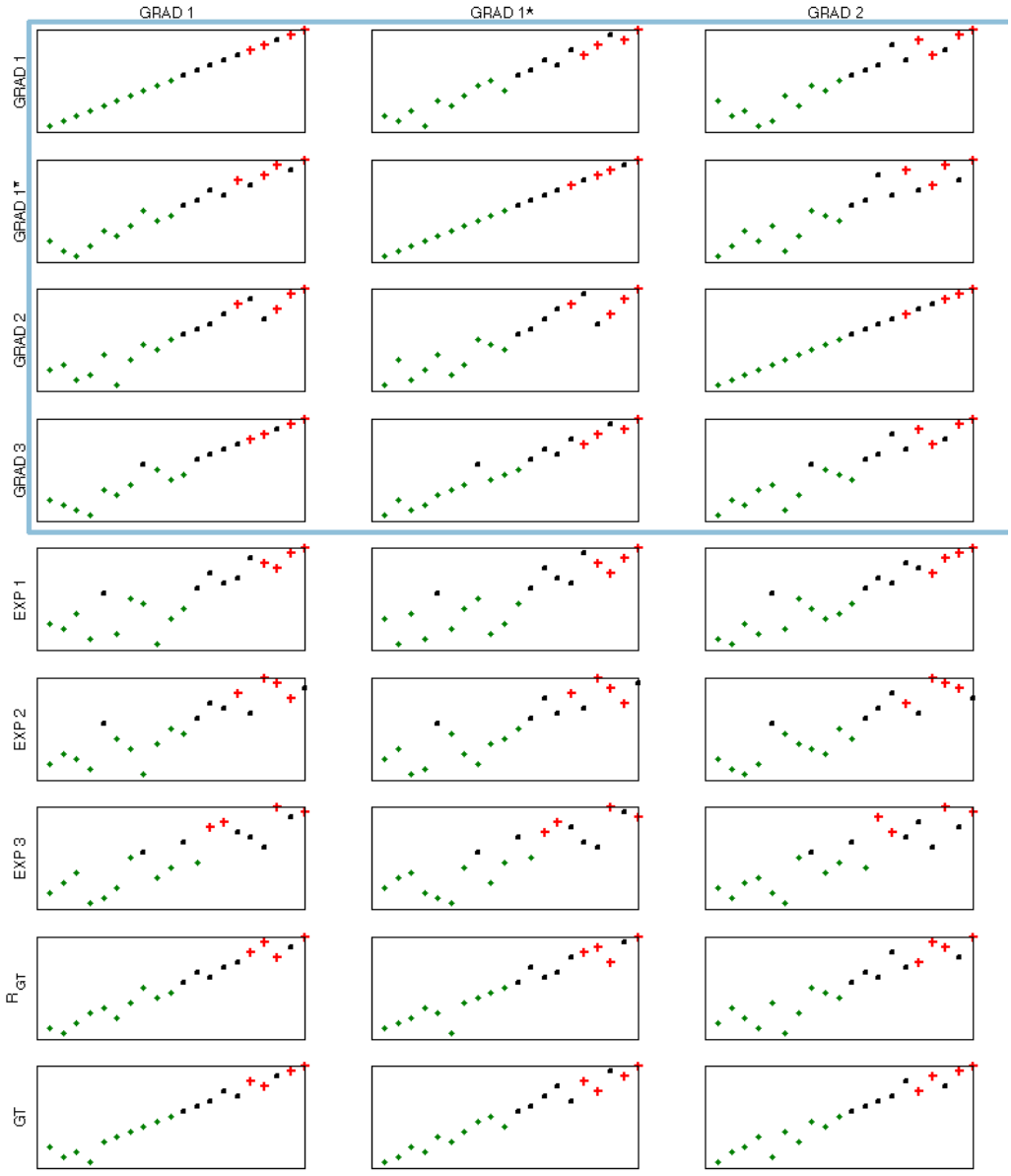

Figure 3.2: X-axis and Y-axis: rank of the images from the corresponding grading. Green square: "healthy" classification by GRAD1. Black circle: pre-plus classification by GRAD1. Red cross: plus classification by GRAD1.

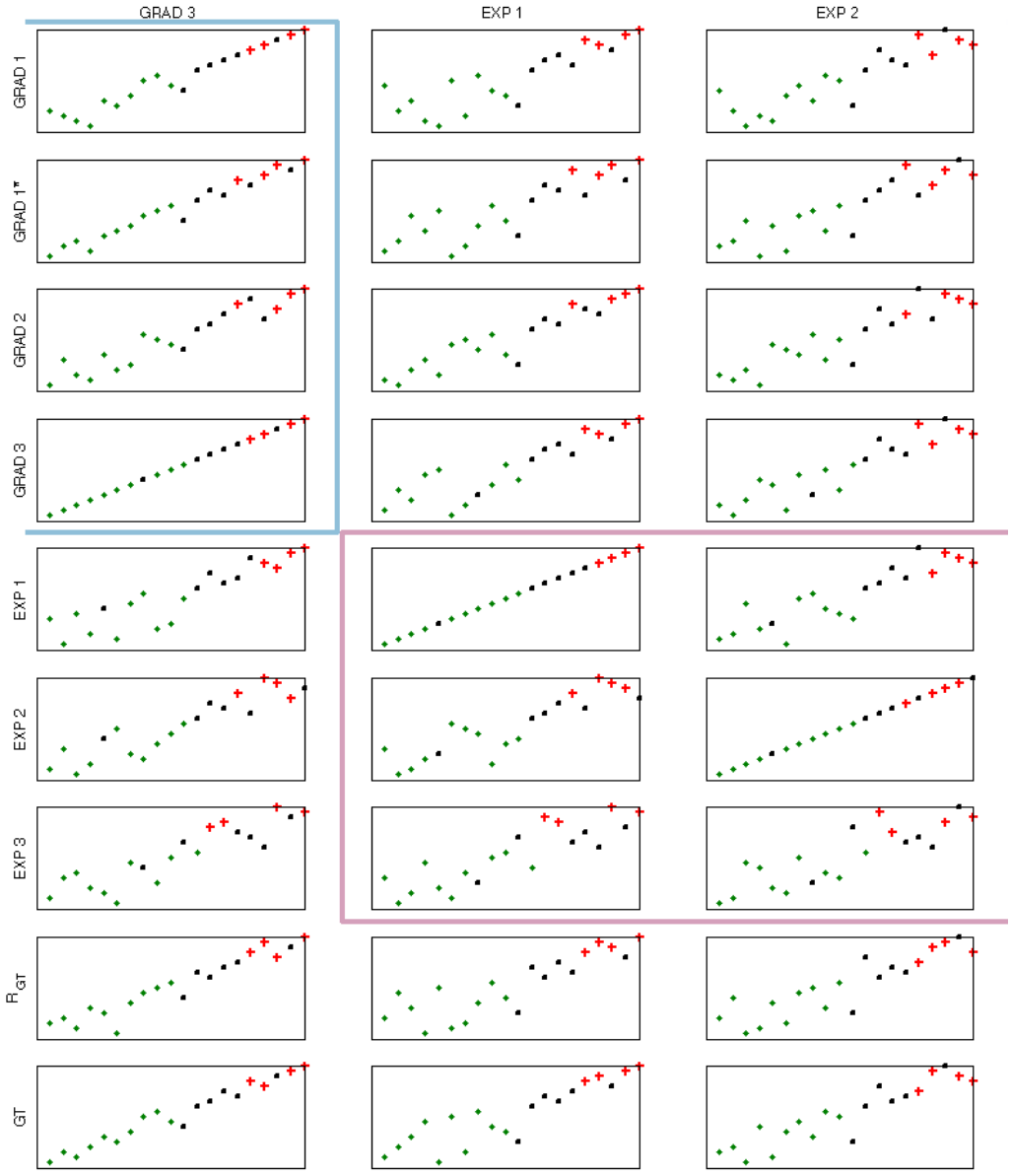

Figure 3.3: X-axis and Y-axis: rank of the images from the corresponding grading. Green square: ”healthy” classification by GRAD1. Black circle: pre-plus classification by GRAD1. Red cross: plus classification by GRAD1.

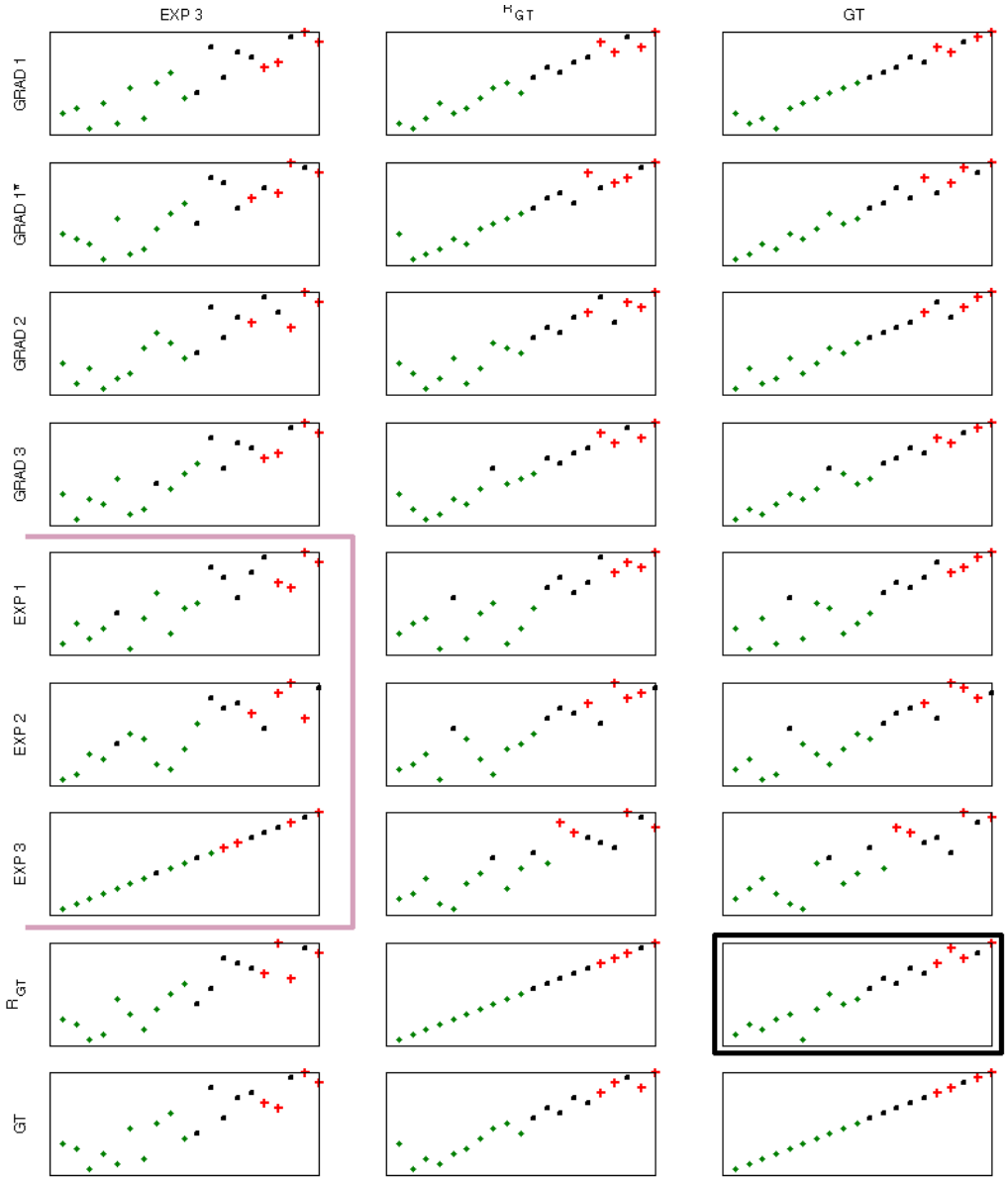

Figure 3.4: X-axis and Y-axis: rank of the images from the corresponding grading. Green square: ”healthy” classification by GRAD1. Black circle: pre-plus classification by GRAD1. Red cross: plus classification by GRAD1.

Chapter 4

Discussion and conclusion

The problem of defining clinical tortuosity is not trivial. Tortuosity definition that have been proposed in literature are vessel-wise, as well as the experiments that test their consistency when assessing controlled examples of tortuosity (e.g., sinusoidal, helical, and random) [8]. There is not yet agreement about what graders/experts mean when evaluating tortuosity, both at the vessel-wise level and at the image wise level. In this work the problem of interpreting the graders' perception of tortuosity has been tackled by defining the following framework:

- a set of lists ordered by tortuosity by a number of ROP experts has been gathered; they were provided with a custom software that ease the task;
- vessel-wise tortuosity measures able to include vessel caliber information have been defined;
- several formulations for aggregating vessel-wise measures into image-wise measures have been defined;
- the best linear weighted combination of image-wise measures has been evaluated in order to replicate the ROP experts gradings; leave-one-out approach permits to avoid over-fitting.

Results highlights that the defined measure correlates well with graders and coherently as grader do (i.e., better than how non-clinical imaging experts do with respect to graders and to themselves). Proposed measure has a stronger correlation with GRADs than with EXPs which is a remarkable result since opinions by clinical are shown to be more coherent and are the point of reference for the diagnose. Images were also classified in terms of severity. Both clinical graders tortuosity ordering (Table 1.1) and proposed measure fail to correctly classify only 1 image (the same image) out of 20, misclassifying it as false positive (Fig. 3.1). It is worth noticing that a false alarm in a clinical contest is a better mistake than a false negative: in this situation an intervention would

be required when not strictly necessary, while a misclassification of a Plus in a PrePlus condition might deny essential treatments. The only re-evaluation available shows a strong correlation (0.98) with the earlier one, denoting high reproducibility for GRAD1. Reproducibility is one of the hallmarks of an automatic evaluation: the proposed algorithm applied on an image will always output the same ranking. It is also worth noticing that in this work ties in rankings have been supposed to be absent. Graders have been asked to always choose the most tortuous between two images, even when showing a very similar condition. Good performances from the algorithm are even more satisfying under this assumption since the algorithm can be asked to rank very close images as clinical graders do. The proposed algorithm has been shown to behave in this dataset at the same level as clinical experts. When paired with an automatic vessel tracing technique (under development), it will provide a completely automated tool for the reliable quantitative estimation of vascular tortuosity in ROP images.

Bibliography

[1] D.K.Vanderveen C. Wu, R.A. Petersen. Retcam imaging for retinopathy of prematurity screening.

[2] Chandrinos KV et al. Image processing techniques for the quantification of atherosclerotic changes. *MEDICON98*, 1998.

[3] Gelman R et al. Diagnosis of plus disease in retinopathy of prematurity using retinal image multi scale analysis. *Invest Ophthalmol Vis Sci*, 2005.

[4] Poletti E et al. Automatic vessel segmentation in wide-field retina images of infants with retinopathy of prematurity. *in proc. IEEE EMBC*, 2011.

[5] Slidsborg C et al. Experts do not agree when to treat retinopathy of prematurity based on plus disease. *British Journal of Ophthalmology*, 2011.

[6] Swanson C et al. Semiautomated computer analysis of vessel growth in preterm infants without and with rop. *British Journal of Ophthalmology*, 2003.

[7] The Committee for the Classification of Retinopathy of Prematurity. *An international classification of retinopathy of prematurity*. Arch Ophthalmol, 1984.

[8] The Committee for the Classification of Retinopathy of Prematurity. *An international classification of retinopathy of prematurity*. Arch Ophthalmol, 2005.

[9] Ruggeri A Grisan E, Foracchia M. A novel method for the automatic grading of retinal vessel tortuosity. *Medical Imaging, IEEE Transactions on*, 2008.

[10] Freedman SF. Wallace DK, Zhao Z. A pilot study using roptool to quantify plus disease in retinopathy of prematurity. *J AAPOS*, 2007.

[11] Clare M et al. Wilson. *Computerized Analysis of Retinal Vessel Width and Tortuosity in Premature Infants*. Investigative Ophthalmology & Visual Science, 2008.