A *divide et impera* strategy for automatic classification of retinal vessels into arteries and veins

Enrico Grisan and Alfredo Ruggeri

Department of Information Engineering, University of Padova, Italy

Abstract

The first pathologic alterations of the retina are seen in the vessel network. These modifications affect very differently arteries and veins, and the appearance and entity of the modification differ as the retinopathy becomes milder or more severe. In order to develop an automatic procedure for the diagnosis and grading of retinopathy, it is necessary to be able to discriminate arteries from veins. The problem is complicated by the similarity in the descriptive features of these two structures and by the contrast and luminosity variability of the retina. We developed a new algorithm for classifying the vessels, which exploits the peculiarities of retinal images. By applying a *divide et impera* approach that partitioned a concentric zone around the optic disc into quadrants, we were able to perform a more robust local classification analysis. The results obtained by the proposed technique were compared with those provided by a manual classification on a validation set of 443 vessels and reached an overall classification error of 12 %, which reduces to 7 % if only the diagnostically important retinal vessels are considered.

Index Terms

Retinal vessels, arteries, veins, classification, fundus imaging, clustering.

I. INTRODUCTION

In the automatic analysis of retinal fundus, a number of algorithms have been proposed for extracting the vascular structure [1] [2] [3] [4] and for identifying non-vascular lesion (exudates, haemorrhages, ischemic regions) [5] [6] [7]. The first changes in the retina that point out the onset of a retinopathy, e.g. from a systemic disease, appear in the vessels. Changes in vessel structure can affect very differently arteries and veins. Tortuosity has a very different course in arteries and veins while retinopathy progresses, and some lesions are peculiar of only one type of vessels, e.g. focal narrowing for arteries, beading for veins. Moreover, one of the early signs of retinopathy is the so called *generalized arteriolar narrowing*, in which a ratio between arteries and veins diameters decreases.

In order to realize an automatic tool for the diagnosis and grading of retinopathy, it is mandatory to automatically distinguish between arterial and venous vessels (A/V classification). The presence of inter- and intra-image contrast, luminosity and color variability is a first problem, whereas the fading of the differences between the two types of vessels in the periphery of the retina is a second, but of no lesser impact, problem. Even after image contrast and luminosity normalization, A/V may be recognized reasonably well only in an area around the optic disc: inside the optic disc, vessels become so intertwined that it is very difficult even for an expert to track a vessel, whereas in the periphery of the image (far from the optic disc) they become thinner and thinner and almost undistinguishable. Moreover, even around the optic disc only vessels close to each other can be reliably recognized as arteries or veins by direct comparison, whereas vessels far apart from each other can be easily misclassified without any further knowledge other than their image features.

The retinal vessel network has a specific structure: the main vessels emerge at the optic disc and then follow a double-parabolic path by branching and thinning (Fig. 1). This structure applies to both arteries and veins and we can assume that, at a small distance from the optic disc border, both types of vessels are distributed in a balanced way around the optic disc.

These observations led us in our strategy to develop a reliable A/V classification technique. To begin with, we decided to classify the vessels only in a well defined concentric zone around the optic disc. Then, by using the vessel structure reconstructed by tracking techniques, we would propagate this classification outside this zone, where little or no information is available to discriminate arteries from veins. Second, the A/V classification algorithm would not be designed considering all together the vessels in the zone, but rather partitioning the zone in four quadrants and working separately and locally on each of them.

II. METHODOLOGY

A. Available Data

Retinal images have been acquired with a fundus camera, centered on the fovea and with a 45° or 50° field of view. The films were subsequently digitized with a scanner, with a color depth of 24 bits and a resolution of 1360 dpi. Resulting 1944x1296 pixels images were cropped to about 1300x1100 pixels, in order to match the retina image bounding box. 35 fundus images have been analyzed in this study: 11 were initially used to develop the algorithm, the remaining 24 were used as validation set.

B. Image preprocessing and vessel tracking

Retinal images suffer quite often from inhomogeneity in luminosity and contrast, both within the same image (intra-image variability) as well as between images (inter-image variability). In order to obtain meaningful brightness information, it is necessary to compensate for this variability. To this end, we employed a previously developed algorithm [8], which analyzes the retinal background area to detect changes of luminosity and contrast, and through an estimation of their local statistical properties derives a compensation for their drifts.

The first task in fundus image analysis is to extract the vessel network, often through a vessel tracking procedure. It provides a set of vessel segments, each usually described by a set S of n points p, representing the x and y Cartesian coordinates of samples of the vessel center line and the vessel diameter d at these locations:

$$p_i = (x_i, y_i, d_i) \tag{1}$$

$$S = \{p_i : i = 1 \ldots n\} \tag{2}$$

In our study, the vessel network has been automatically extracted by a sparse tracking algorithm previously developed [9].

C. Divide

In order to exploit the local nature of the A/V classification procedure and the symmetry of the vessel network layout, we partitioned the retina in regions, in which there is a reasonably similar number of veins and arteries, and in which the two types of vessels hopefully have significant local differences in features. A concentric zone around the optic disc was identified and then partitioned into four regions, each containing one of the main arcs of the A/V network: superior-temporal (ST), inferior-temporal (IT), superior-nasal (SN) and inferior-nasal (IN), as in
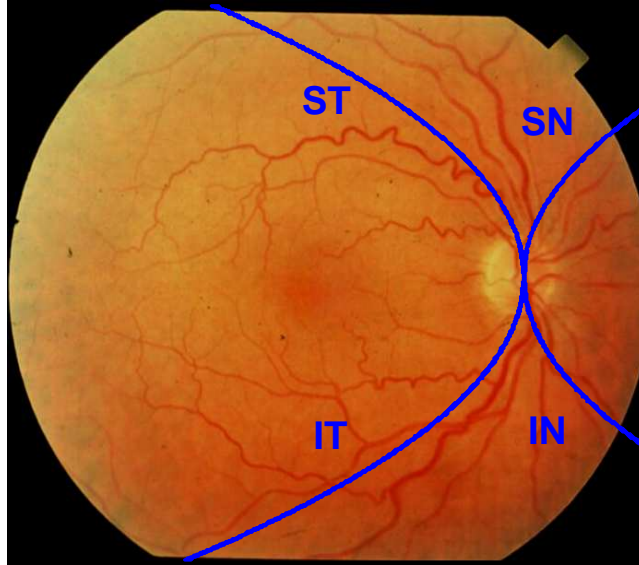
Fig. 1.   Principal arcades of the retinal vessel network

Fig. 1.

To perform this partitioning, we first needed to identify the position of the optic disc and its approximate diameter; this can be done manually or automatically, e.g. as in [10] or [11]. Then we identified the cardinal axes that divide the retina into the four quadrants $Quad_i$ $i = 1 \ldots 4$, that contain the above mentioned main vessel arcs. Given the center of the optic disc c_{OD} and its radius r_{OD}, the concentric zone C delimiting the region of analysis was centered in c_{OD} and had inner radius $w_i \cdot r_{OD}$ and outer radius $w_o \cdot r_{OD}$, with $w_i = 1$ and $w_o = 4$ (see Fig. 2).

For each quadrant $Quad_i$, the 5 largest tracked vessels $S_1, \ldots, S_5$, i.e. the vessels that had the largest mean diameter, were automatically selected by the algorithm. Therefore, for each quadrant we had a set $\mathcal{S}_{Quad_i} = \{S_1, \ldots, S_5\}$ of vessels under analysis. An example for all quadrants is given in Fig. 3.

This selection was performed to restrict the analysis to main vessels and avoid the possibly confusing information coming from small arterioles and venoules. The balanced presence of veins and arteries in each of the four quadrants holds only if main vessels and their branches are considered.

D. Feature Extraction

For each quadrant, the set $\mathcal{P}$ to be classified is represented by the sample points describing the vessels in $\mathcal{S}_{Quad_i}$ in the concentric zone C.

$$\mathcal{P}_{Quad_i} = \{\boldsymbol{p}_{i,j} = [x_{i,j}, y_{i,j}, d_{i,j}] \,|\, \boldsymbol{p} \,\epsilon\, \mathcal{S}_{Quad_i} \cap C\} \tag{3}$$

For each selected sample point $\boldsymbol{p}_{i,j}$ in $\mathcal{P}_{Quad_i}$, a circular region $CR_{i,j}$ of diameter $d_{CR} = 0.8 * d_{i,j}$ is identified:

$$CR_{i,j} = \{\boldsymbol{p} = [x, y] | \sqrt{(x - x_{i,j})^2 + (y - y_{i,j})^2} \leq d_{CR}\} \tag{4}$$

The triplet of values representing the RGB intensities of all pixels in $CR_{i,j}$ can also be expressed in another color space that more closely relates to a perceptual difference, such as the hue (H), saturation (S) and luminance (L) space. A non-linear mapping is used for this transformation [12]. A great deal of features can be extracted from the RGB and HSL values of the set $CR_{i,j}$. We performed an extensive statistical analysis to identify the features most discriminant for the A/V classification of $p_{i,j}$. The result was that the variance of red values $Var[R(CR_{i,j})] = \overline{R_{i,j}}$ and the mean of hue values $E[H(CR_{i,j})] = \overline{H_{i,j}}$ emerged as the best features to classify $\boldsymbol{p}_{i,j}$ as belonging to an artery or to a vein. The fact that the classes of veins and of arteries are distinguished by looking at their average hue and homogeneity of their red component is also in agreement with clinical experience: when two vessels close to each other are compared for classification, the one with the darker red is classified as vein; if this difference is not significant enough, the one with the highest degree of uniformity is classified as vein.

E. Impera

After features extraction, we have four sets of data F_i, each consisting in a $2 \times N_i$ matrix with variance of red intensities and mean of hue intensities for the N_i points in $\mathcal{P}_{Quad_i}$. The jth position in F_i correspond to the point $\boldsymbol{p}_{i,j}$ and is:

$$F_i(j) = [\overline{R_{i,j}}, \overline{H_{i,j}}] \tag{5}$$

For each set F_i, the centers of the two classes are identified via a fuzzy clustering algorithm [13]:

$$c_{art,i} = [\overline{R_{art,i}}, \overline{H_{art,i}}] \text{ for arteries} \tag{6}$$

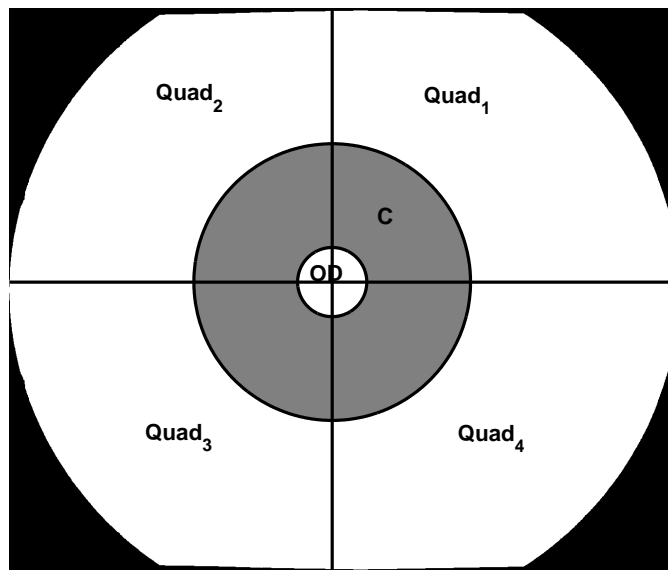$$c_{vein,i} = [\overline{R_{vein,i}}, \overline{H_{vein,i}}] \text{ for veins} \tag{7}$$
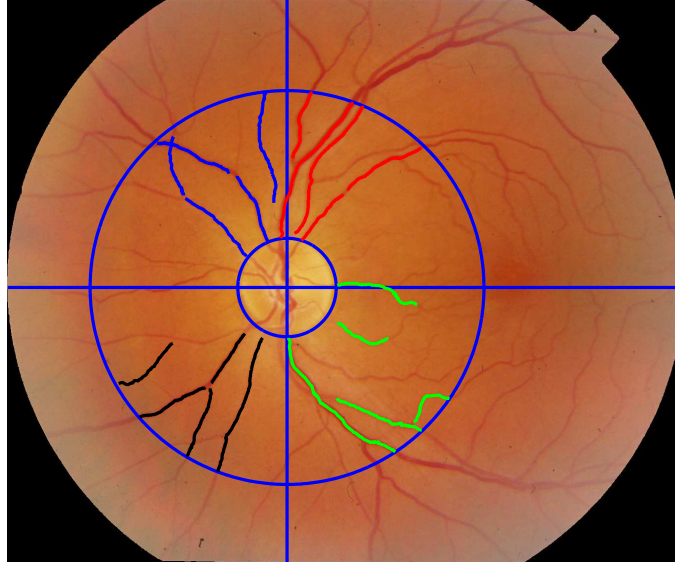

Fig. 2. Division of the fundus

Fig. 3. Selected vessels $\bigcup_{i=1}^{4} \mathcal{S}_{Quad_i} \cap C$ in a sample image

with $\overline{R_{art,i}} > \overline{R_{vein,i}}$.

The Euclidean distance of each point from the centers is the criterion used for the classification:

$$\boldsymbol{p}_{i,j} \epsilon Artery \Leftrightarrow \|F_i(j) - c_{art,i}\| \leq \|F_i(j) - c_{vein,i}\| \tag{8}$$

After all points are assigned to either of the two classes, an empirical probability P for each vessel to be an artery (or vein) can be determined. Being p_v the number of points $\boldsymbol{p}_{i,j}$ of vessel S assigned to the vein class and p_a the number of points assigned to the artery class, these probabilities are:

$$P[S \epsilon \, Artery] = \frac{p_a}{p_v + p_a} \tag{9}$$

$$P[S \epsilon \, Vein] = 1 - \frac{p_a}{p_v + p_a} \tag{10}$$

These values are then used to classify the whole vessel, according to the class having the higher probability.

III. RESULTS AND DISCUSSION

In Fig. 4 features for all the points $\boldsymbol{p}_{i,j}$ inside the concentric zone are plotted; in Fig. 5 features for the points of one quadrant only are plotted. It is quite evident that the two clusters of Fig. 5 are much more well defined than those of Fig. 4 and thus the strategy of partitioning the retina in four quadrants greatly improves the classification.

The A/V classification results obtained by the proposed algorithm on the validation set were compared to those provided by a manual classification on the same vessels performed by an expert. The classification performances are shown in Table I, where they are reported both for the whole data set and separately for the main vessels, i.e. those belonging to the main vascular arcades, and the secondary vessels. The origin of the latter ones is that it is possible that in some quadrant fewer than five segments belonging to the main vessels are present, in which case small arterioles and venoules are also taken into account. Since they are quite similar to each other and their
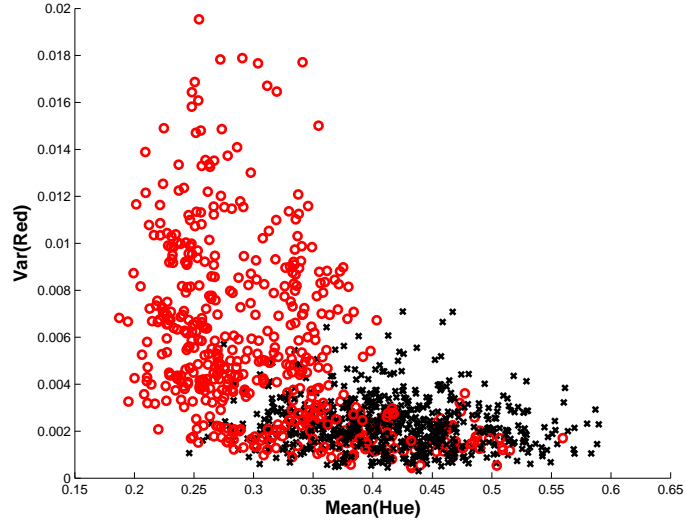
Fig. 4.   Features for the points from the vessels in $\bigcup_{i=1}^{4} \mathcal{S}_{Quad_i} \cap C$. Circles are artery sample points, crosses are vein sample points.

discrimination is not very significant from a clinical point of view, we can accept a higher error rate for them. A correct classification of the main vessels, on the contrary, has a far greater impact on the further diagnostic analysis of the retinal fundus.

Finally, it is worth noting that it was not always possible to identify five vessels in each quadrant. This is the reason why the total number of vessels is 443 and not 480. As long as there are at least one artery and one vein in each quadrant, the algorithm is applicable.

Going through the classification of single points first, and then of whole vessels, a more robust classification
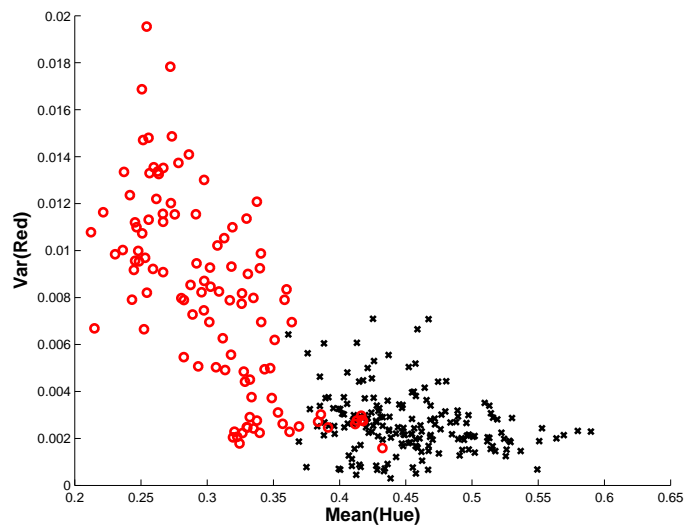

Fig. 5.   Features for the points from the vessels in $\mathcal{S}_{Quad_1} \cap C$. Circles are artery sample points, crosses are vein sample points.

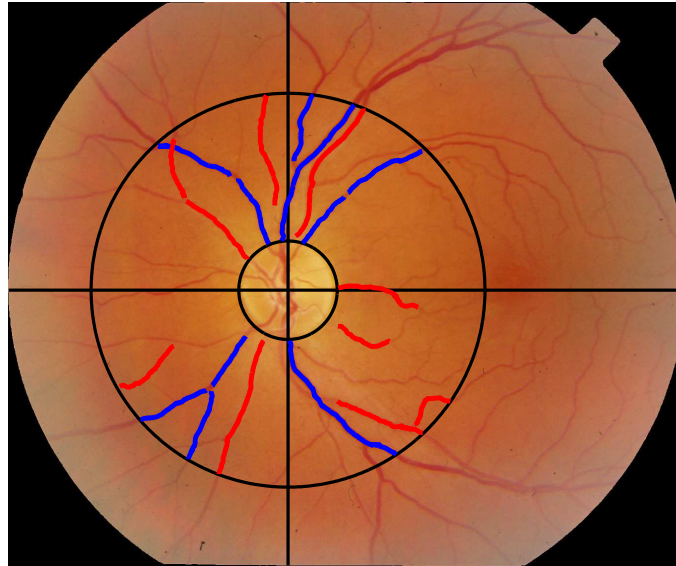|	Number of vessels	Misclassifications	Error
Total	443	55	12.4 %
Main Vessels	264	18	6.7 %
Secondary Vessels	178	37	20.7 %

TABLE I

VESSEL CLASSIFICATION PERFORMANCE

Fig. 6. Result of the A/V classification achieved by the proposed algorithm on a sample image; red vessels were classified as artery, blue ones as vein.

strategy is achieved, since small local variations in chromaticity/luminosity or vessel tracking errors have a reduced impact on the overall classification. An example of resulting A/V classification is shown in Fig. 6

IV. CONCLUSION

We have developed an algorithm for vessel classification in retinal images. The *divide et impera* procedure we proposed is justified by the balanced layout of arteries and veins in retinal fundus and by the local nature of the classification process.

The result obtained on the 24 images of the validation set are satisfactory, since main vessels have been correctly classified 93% of the times. An extensive evaluation of the proposed algorithm will be performed on a larger set of images. An automatic algorithm for diffusing the classification obtained in the concentric zone around the optic disc along the two vascular networks towards the periphery of retina will also be developed.

ACKNOWLEDGMENT

REFERENCES

[1] L. Gang, O. Chutatape, and S. M. Krishnan, "Detection and measurement of retinal vessels in fundus images using amplitude modified second-order Gaussian filter," *IEEE Transactions on Biomedical Engineeing*, vol. 49, no. 2, pp. 168–172, February 2002.

[2] A. Hoover, V. Kouznestova, and M. Goldbaum, "Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, March 2000.

[3] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Transactions on Medical Imaging*, vol. 8, no. 3, pp. 263–269, September 1989.

[4] Y. A. Tolias and S. M. Panas, "A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering," *IEEE Transactions on Medical Imaging*, vol. 17, no. 2, pp. 263–273, March 1998.

[5] B. M. Ege, O. K. Hejlesen, O. V. Larsen, K. Möller, B. Jennings, D. Kerr, and D. A. Cavan, "Screening for diabetic retinopathy using computer based image analysis and statistical classification," *Computer Methods and Programs in Biomedicine*, vol. 62, pp. 165–175, 2000.

[6] M. Goldbaum, S. Moezzi, A. Taylor, S. Chatterjee, J. Boyd, E. Hunter, and R. Jain, "Automated diagnosis and image understanding with object extraction, object classification, and inferencing in retinal images," in *1996 IEEE International Conference on Image Processing*, 1996, pp. 695–698.

[7] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee, "An effective approach to detect lesions in color retinal images," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2000.

[8] E. Grisan, M. Foracchia, and A. Ruggeri, "Color fundus images luminance and contrast normalization," in *Abstracts of the 3rd CAFIA workshop, in European Journal of Ophthalmology*, March 2003, pp. 228–229.

[9] M. Foracchia, E. Grisan, and A. Ruggeri, "Quantitative description of vessel features in hypertensive retinopathy fundus images," in *Abstracts of the 2nd CAFIA workshop*, A. online at http://www.cafia.org/cafia2001b/index.html, Ed., 2001, p. 15.

[10] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *unpublished, available on-line at: http://www.parl.clemson.edu/stare/nerve/*.

[11] A. Ruggeri, M. Foracchia, and E. Grisan, "Detection of optic disk based on a directional model of vessel network," in *Abstracts of the 3rd CAFIA workshop, in European Journal of Ophthalmology*, March 2003, p. 229.

[12] R. C. Gonzalez, *Digital Image Processing*. Addison Wesley Publishing Company, 1992.

[13] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.