International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:4, No:7, 2010

A Similarity Function for Global Quality Assessment of Retinal Vessel Segmentations

Arturo Aquino, Manuel Emilio Gegúndez, José Manuel Bravo, Diego Marín

Abstract—Retinal vascularity assessment plays an important role in diagnosis of ophthalmic pathologies. The employment of digital images for this purpose makes possible a computerized approach and has motivated development of many methods for automated vascular tree segmentation. Metrics based on contingency tables for binary classification have been widely used for evaluating performance of these algorithms and, concretely, the accuracy has been mostly used as measure of global performance in this topic. However, this metric shows very poor matching with human perception as well as other notable deficiencies. Here, a new similarity function for measuring quality of retinal vessel segmentations is proposed. This similarity function is based on characterizing the vascular tree as a connected structure with a measurable area and length. Tests made indicate that this new approach shows better behaviour than the current one does. Generalizing, this concept of measuring descriptive properties may be used for designing functions for measuring more successfully segmentation quality of other complex structures.

Keywords—Retinal vessel segmentation, quality assessment, performance evaluation, similarity function.

I. Introduction

DIGITAL retinal images are employed in diagnosis of ophthalmic pathologies. Vessels assessment is an important diagnosis key to detect and evaluate many of them that produce vascular anomalies, such as diabetic retinopathy. As previous step, it requires segmentation of the vascular tree from the background. However, when number of vessels in an image is large, or when a large number of images is acquired, manual delineation of vessels becomes tedious or even impossible. That is why many methods for segmenting vessels automatically have been proposed over the last years from many different approaches. *Mathematical morphology* [1], [2], [3], [4], *matched filtering* [5], [6], [7], [8], [9], [10], *supervised methods* [11], [12], [13], [14], [15], [16] or *model-based locally adaptive thresholding* [17] among other, are techniques used in this topic.

Metrics based on contingency tables for binary classification [18] have been widely used over the last years for objectively evaluate quality of automated retinal vessel segmentations [4], [6], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. Specifically, the *accuracy* (Acc) [18] has been accepted as the metric for quantifying global quality of these segmentations and therefore, it has been accepted as the objective measure for comparing segmentation algorithms and for providing a global idea of their behaviour. It is calculated as the ratio of pixels well-classified into the two classes, vessel and background. For that, segmentations are compared pixel by pixel with hand labelled images.

The difficulties of quality evaluation of retinal vessel segmentations have been previously pointed out and discussed by Niemeijer *et al.* in [13]. In the authors' opinion, the exclusive use of this metric is not suitable for assessing global quality of these segmentations and presents some outstanding difficulties.

In this paper, a new similarity function for evaluating global quality of retinal vascular tree segmentations as a complement of the metrics based on contingency tables is proposed. This similarity function is based on the authors' perception that the three main measurable features that describe the vascularity are connectivity, area and length of all vessels. Since the new function enhances the numeric assessment of the global quality of this kind of segmentations, it can be considered potentially useful for researchers of this field.

A. Aquino, J.M. Bravo and D. Marín are with the Department of Electronic, Computer Science and Automatic Engineering, "La Rábida" Polytechnic School, University of Huelva, Spain (e-mail: {arturo.aquino, caro, diego.marin}@diesia.uhu.es).

M.E. Gegúndez is with the Department of Mathematics, "La Rábida" Polytechnic School, University of Huelva, Spain (e-mail: gegundez@uhu.es).

II. New Similarity Function for Quality Assessment of Retinal Vessel Segmentation

Here, in the first sub-section, the development of a new similarity function for assessing quality of retinal vessel segmentations is motivated. For that, the main identified weaknesses of the Acc in its use for this purpose are described. The second sub-section presents and describes the new proposal.

A. Motivation: Weaknesses of the Accuracy

Subjective distortion measure, obtained by subjective testing, quantifies the dissatisfaction of the viewer in observing the distorted image in place of the original. On the other hand, objective distortion measure gives the distortion between the original and the distorted image mathematically. For an appropriate and comprehensive evaluation of quality, it is desirable to obtain a good matching between both subjective and objective distortion measure.

For a numerical assessment of the quality of a segmentation generated by means of an automated process, a binary image representing the "perfect" case to be compared with is needed. This last, typically called ground truth or gold standard image, is made by a medical expert by manually labeling the original one.

To assess the degree of correspondence between subjective testing and objective measurement using the Acc, an experiment was performed. Firstly, five different retinal vessel segmentations from the same image having all of them an Acc value of 0.9798 were generated. Then, a set of observers was asked for evaluating them, obtaining for the lowest scored segmentation an average of 0.35 and for the highest one 0.95. This outstanding mismatching can be explained by identifying objective deficiencies derived directly from the nature of the metric and the vascularity. This last is a tree-like connected structure composed of some main gross vessels that progressively split in other thinner. Anatomically, thin vessels constitute in average more than 50% of the components of the vascular tree and, despite their small size, they are relevant from a medical point of view, since they may suffer damages and contribute diagnose information. Nevertheless, in a digital image they only suppose about 21% of the pixels of the whole vasculature. Furthermore, the whole vascular tree is approximately only the 12% of the pixels of the retina.

Besides all mentioned above, since automated segmentations are compared with ground truth images made by human observers, another difficulty is therefore the subjectivity introduced by these observers. Differences in vessels tracing, width, length or even in painting some vessels or not can be observed in gold standards made by different observers for the same image. It is important to highlight that any function for evaluating segmentation quality would be sensitive to differences in vessels presence and absence in different gold standards, as this is a difficulty derived of the use of human-made gold standards. However, impact of slight differences in vessels tracing or width should be minimized by any metric for this purpose, and the Acc does not do it.

Summarizing, the following are the main identified deficiencies of the Acc:

1) High insensitivity to absence of thin vessels detection.
2) Low sensitivity to poor vascular tree segmentations.
3) Strong dependence with the gold standard used for measuring.
4) Very poor matching with human perception.

B. New Similarity Function Description: The CAL Similarity Function

The similarity function based on connectivity, area and length (in the following CAL), is a product of three factors that evaluate each of them. So, mathematically it can be defined as

$$CAL = C * A * L \tag{1}$$

This function takes values from the interval $[0, 1]$, where values 0 and 1 denote the worst and perfect cases of segmentation respectively. Description of each factor is as follows:

- Connectivity (C): This factor assesses the fragmentation degree of the segmentation with respect to the gold standard. As the vascular tree is a connected structure, a good vascular segmentation is expected to have only a few connected components (ideally one). So, this factor penalizes fragmented segmentations according to:

$$C = 1 - min\left(1, \frac{|\#_C(S_G) - \#_C(S)|}{\#(S_G)}\right) \tag{2}$$

where min is the minimum function, $\#_C(S_G)$ and $\#_C(S)$ are the number of connected components in gold standard S_G and segmentation S respectively, and $\#(S_G)$ denotes the cardinal of S_G. Note that here, for simplifying, the word segmentation reffers the set of vessel pixels exclusively, excluding therefore the set of background pixels from this term. This is held in the rest of this section.

- Area: This factor evaluates the degree of overlapping areas between the segmentation and the gold standard:

$$A = \frac{\#((\delta(S) \cap S_G) \cup (S \cap \delta(S_G)))}{\#(S \cup S_G)} \tag{3}$$

where S_G is the gold standard and S the segmentation to be evaluated. Function δ is a morphological dilation using a small disc of three pixels in diameter. Introduction of this operator provides tolerance with slight differences in vessel width.

- Length: Here the degree of coincidence between the segmentation and the gold standard in terms total length is measured. Formally it is expressed as,

$$L = \frac{\#((\varphi(S) \cap \delta(S_G)) \cup (\delta(S) \cap \varphi(S_G)))}{\#(\varphi(S) \cup \varphi(S_G))} \tag{4}$$

where φ is an homotopic skeletonization [19] and δ is again a morphological dilation (with the same defined disc) for decreasing impact of slight differences in vessel tracing.

III. Material

For the experimentation made in this study, the publicly available DRIVE database [20] was used. This database has been widely used by researchers to test their vessel segmentation methodologies since, apart from being public, they provide manual segmentations for performance evaluation.

The DRIVE database comprises 40 eye-fundus color images taken with a Canon CR5 non-mydriatic 3CCD camera with a 45° field of view (FOV). Each image was captured at 768x584 pixels, 8 bits per color plane and, in spite of being offered in LZW compressed TIFF format, they were originally saved in JPEG format. The database is divided into two sets: a test set and a training set, each of them containing 20 images. The test set provides the corresponding FOV masks for the images, which are circular (approximated diameter of 540 pixels) and two manual segmentations generated by two different specialists for each image. The training set also includes the FOV masks for the images and a set of manual segmentations made by the first observer.

The set used for all the experimentation presented in this work is the test set of this database. This set, as it has been described, offers two sets of gold standard images made by two different observers. This material was necessary for making some experiments described in this paper.

IV. Experimentation

In this section, the behaviour of the CAL and Acc against some examples representative of the four previously described weaknesses of the current metric is compared.

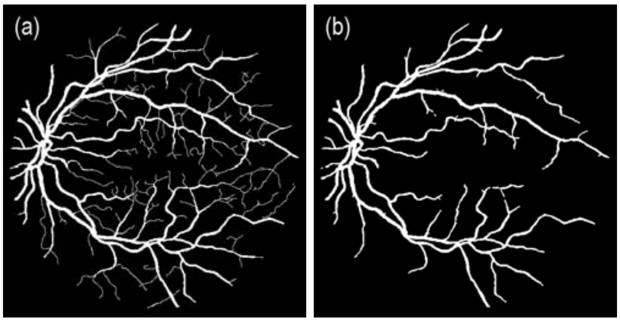

Fig. 1. (a) Gold standard. (b) Segmentation with none thin vessel detected.

A. Sensibility to Thin Vessels Segmentation

Let's define as gross vessels those that directly emerge in the optic disc and the direct ramifications of these ones, and thin vessels as the rest. Then, to assess influence of poor thin vessels detection in the Acc and CAL, let's consider the case illustrated in Figure 1. In this figure, image (b) is a vessel segmentation in which none thin vessel has been detected and all gross vessels have been perfectly detected, and image (a) is the gold standard. Then, values obtained for image (b) compared with image (a) using both functions are:

- $Acc = 0.9749$
- $CAL = 1.0 * 0.8190 * 0.7323 = 0.5997$

Note that, although none thin vessel pixel was detected, the Acc indicates that image (b) is a segmentation close to be perfect. The CAL function penalizes much more this defect.

B. Sensibility to Poor Segmentations

Let's consider three different segmentations of the same image. The first one contains the 50% of the vascularity (Figure 2, image (b)), in the second segmentation only the 25% of vessel pixels were detected (Figure 2, image (c)) and the last segmentation is an "absurd" case that represents 0% of vessel detection (Figure 2, image (d)). If image (a) is the gold standard and these segmentations are evaluated with the Acc and CAL, these results are obtained:

- 50% of vascularity detected:
  - $Acc = 0.9478$
  - $CAL = 1.0 * 0.5005 * 0.4900 = 0.2452$
- 25% of vascularity detected:
  - $Acc = 0.9220$
  - $CAL = 1.0 * 0.2547 * 0.2490 = 0.0634$
- 0% of vascularity detected:
  - $Acc = 0.8962$
  - $CAL = 0.9999 * 0 * 0 = 0$

Although segmentations (b) and (c) visually compared with the gold standard (a) are too deficient, they obtain a global measure of quality using the Acc that may be defined as excellent, as both obtained a score above 0.9. On the other hand image (d), despite of being a null segmentation, obtains with this metric a score of almost 0.9. The behaviour of the CAL similarity function is closer to what can be expected, grading low the images (b) and (c) and with 0 the image (d).

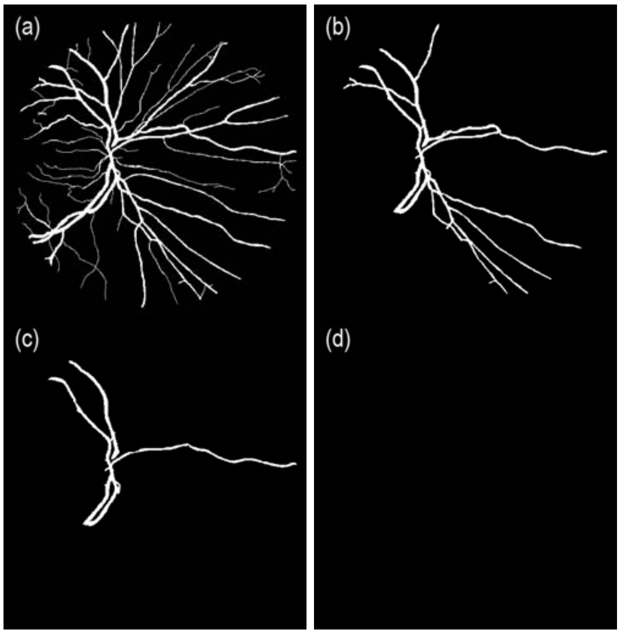

Fig. 2. (a) Gold standard. (b) Segmentation of (a) with 50% of vascularity detected. (c) Segmentation of (a) with 25% of vascularity detected. (d) Segmentation of (a) with 0% of vascularity detected.

C. Dependence with the Gold Standard

This aspect is studied by comparing two manual segmentations of the same image made by the two different observers. In Figure 3, image (a) is a gold standard made by the first observer, and image (b) is the gold standard made by the second one. Both images were modified to avoid differences in vessels presence or absence and significative differences in vessels length. By this way, only slight variations in vessels tracing or width can be studied. Image (c) is the colour composition of both images in which black colour indicates coincidence in background, pixels in yellow colour are coincidences of both segmentations in vessels, and red and green colours show disagreements in vessels; red pixels are from the first observer and green ones from the second observer. Then, considering image (b) as a segmentation to be evaluated, and image (a) as the gold standard image to be compared with, the following measures of quality are obtained:

- $Acc = 0.9689$
- $CAL = 1.0 * 0.9312 * 0.9565 = 0.8907$

On the other hand, image (d) is a degradated version of image (a), generated by cutting and erasing many vessels. The quality values obtained for this case are:

- $Acc = 0.9888$
- $CAL = 0.9998 * 0.8725 * 0.8272 = 0.7216$

As it can be checked, paradoxically, although any human observer would say that image (d) is clearly better segmentation than (b) when visually comparing with the gold standard image (a), the Acc indicates the opposite. That is result of the sensitivity of this metric to the previously discussed variations. As results indicate, this sensitivity is reduced with the CAL function.

TABLE I
BEHAVIOUR COMPARISON OF THE SUBJECTIVE POLLING, Acc AND CAL ON THE SEGMENTATION CASES DESCRIBED IN THIS PAPER. DATA IN TERMS OF AVERAGE VALUES

	Fig.1 (b)-(a)	Fig.2 (b)-(a)	Fig.2 (c)-(a)	Fig.2 (d)-(a)	Fig.3 (b)-(a)	Fig.1 (d)-(a)
Poll	0.4375	0.1750	0.10	0.0	0.925	0.6125
Acc	0.9749	0.9478	0.9220	0.8962	0.9689	0.9888
CAL	0.5997	0.2452	0.0634	0.0	0.8907	0.7216

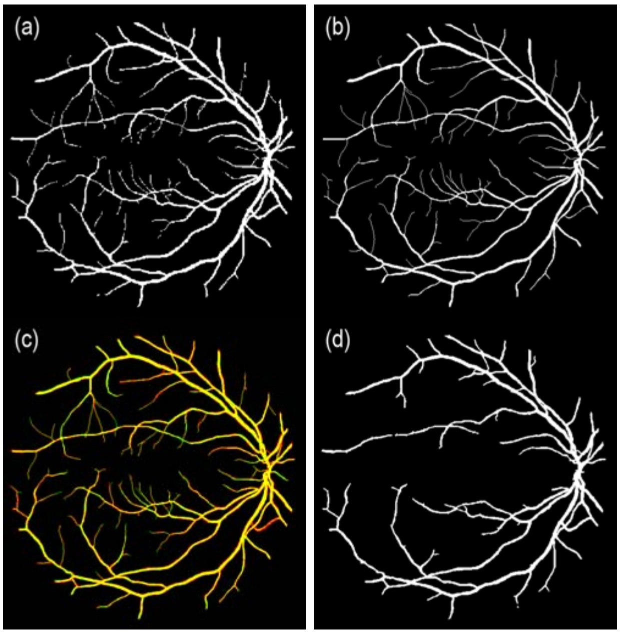

Fig. 3. (a) Gold standard from the 1st observer. (b) Gold standard from the 2nd observer. (c) Color composition of (a) and (b). (d) Degradated version of (a).

D. Degree of Matching with Human Perception

For evaluating this issue, 20 observers were asked for marking quality of all segmentation cases proposed in Figures 1, 2 and 3. Results obtained with this opinion poll are summarized in Table 1. In it, each column represents a segmentation case. Each case is codified as *Fig.n (x)-(y)*, where *n* is the number of the figure and *(x)* and *(y)* are the gold standard and the segmentation used in that case respectively. As it can be checked, the new proposal significatively enhances correlation between subjective and objective quality evaluation.

V. DISCUSSION AND CONCLUSION

In this paper, a new similarity function for assessing quality of retinal vessel segmentations has been presented. This measure is derived from the authors' observation that three main properties describe the vascularity: connectivity, area and total length. Weaknesses of the Acc as a metric for this purpose have been identified and discussed. Experiments made show that the new proposal presents better behaviour in those situations.

This work was motivated by the authors' interest in the field of retinal vessel segmentation. Developing segmentation algorithms, the authors realized that any evident visual enhancing had very low incidence in quality measurements and that these measurements did not match with what they expected according with their subjective perception. It was also found that quality values were always very high and that difference between results visually good and poor was only a few hundredths. Furthermore, sometimes the authors obtained measurements that indicated worse results for some versions of the algorithms when they expected an enhancing.

For all mentioned, the new similarity function presented here constitutes a good complement to the metrics based on contingency tables for binary classification for the evaluation of global quality of retinal vessel segmentations.

The CAL function has been designed for measuring segmentation quality of a certain structure and therefore is not applicable in general cases. However, generalizing, the applied concept of measuring descriptive properties may be useful for designing other specialized similarity functions for enhancing segmentation quality assessment of other complex shapes.

ACKNOWLEDGMENT

The authors would like to thank J. J. Staal and colleagues for making their database publicly available.

REFERENCES

[1] T. Walter and J. C. Klein, "Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques," in *Medical Data Analysis*, J. Crespo, V. Maojo, and F. Martin, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 282-287. ser. Lecture Notes in Computer Science.

[2] F. Zana and J. C. Klein, "Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation," *IEEE Trans. Image Processing*, vol. 10, pp. 1010-1019, 2001.

[3] C. Heneghan, J. Flynn, M. OKeefe, and M. Cahill, "Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis," *Med. Image Anal.*, vol. 6, pp. 407-429, 2002.

[4] A. M. Mendonça and A. Campilho, "Segmentation of Retinal Blood Vessels by Combining the Detection of Centerlines and Morphological Reconstruction," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1200-1213, Sept. 2006.

[5] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Med. Imag.*, vol. 8, pp. 263-269, 1989.

[6] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, pp. 203-210, 2000.

[7] 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 Trans. Biomed. Eng.*, vol. 49, pp. 168-172, 2002.

[8] M. Al-Rawi and H. Karajeh, "Genetic algorithm matched filter optimization for automated detection of blood vessels from digital retinal images," *Comput. Methods Programs Biomed.* vol. 87, pp. 248253, 2007.

[9] M. Al-Rawi, M. Qutaishat, and M. Arrar, "An improved matched filter for blood vessel detection of digital retinal images," *Comput. Biol. Med.* vol. 37, pp. 262267, 2007.

[10] M. G. Cinsdikici and D. Aydın, "Detection of blood vessels in ophthalmoscope images using MF/ant (matched filter/ant colony) algorithm," *Comput. Methods Programs Biomed.*, vol. 96, no. 2, pp. 85-95, Nov. 2009.

[11] G. G. Gardner, D. Keating, T. H. Williamson, and A. T. Elliott, "Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool," *Br. J. Ophthalmol.*, vol. 80, pp. 940-944, 1996.
[12] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea and retinal blood vessels from digital colour fundus images," *Br. J. Ophtalmol.*, vol. 83, pp. 902-910, 1999.
[13] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *SPIE Med. Imag.*, J.M. Fitzpatrick and M. Sonka, Eds., 2004, vol. 5370, pp. 648-656.
[14] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, pp. 501-509, 2004.
[15] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar Jr., H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, pp. 1214-1222, 2006.
[16] E. Ricci and R. Perfetti, "Retinal Blood Vessel Segmentation Using Line Operators and Support Vector Classification," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1357-1365, Oct. 2007
[17] X. Jiang and D. Mojon, "Adaptive Local Thresholding by Verification-Based Multithreshold Probing with Application to Vessel Detection in Retinal Images," *IEEE Trans. Pattern Analy. Mach. Intell.*, vol. 25, no. 1, pp. 131-137, Jan. 2003.
[18] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, June 2006.
[19] J. Serra, *Image Analysis and Mathematical Morphology*. London: Academic Press, 1982, vol. 1.
[20] Research section, digital retinal image for vessel extraction (DRIVE) database University Medical Center Utrecht, Image Sciences Institute, Utrecht, The Netherlands [Online]. Available: http://www.isi.uu.nl/Research/Databases/DRIVE

Arturo Aquino Engineer in Computer Sciences from the High Technical School in Computer Sciences of the University of Sevilla (Spain), in 2007, received the M.S. Degree in Computer Engineering and Networks from the University of Granada (Spain). Since 2007 he has been pursuing the Ph.D. Degree student at the University of Huelva (Spain).

Manuel Emilio Gegúndez received the B.S. Degree in Mathematics from the University of Sevilla (Spain), in 1992, and the Ph.D. Degree in Mathematics from the University of Sevilla in 1996. He is currently an Associate Professor at the Department of Mathematics of the University of Huelva (Spain). His current research interests include Computer Vision, Identification and Knowledge Extraction.

José Manuel Bravo was born in Huelva (Spain), in 1972. He received the Ph.D. degree in Computer Science from the University of Seville, in 2004. He is currently with the Department of Electronic, Computer Science and Automatic Engineering of the University of Huelva. His current research interests include Predictive Control, Guaranteed State Estimation and Identification, Computer Vision and Digital Image Processing.

Diego Marín received the B.S. Degree in Physics from the University of Sevilla (Spain), in 1995, and the Ph. D. Degree in Physics from the University Complutense of Madrid (Spain), in 2003. Since 2003, he is a lecturer at the Department of Electronic, Computer Science and Automatic Engineering of the University of Huelva (Spain). His current research interests include Computer Vision, Digital Image Processing and System Modelling and Identification.