

# PROLIFERATIVE DIABETIC RETINOPATHY CHARACTERIZATION BASED ON THE SPATIAL ORGANIZATION OF VASCULAR JUNCTIONS IN FUNDUS IMAGES

*Argyrios Christodoulidis, Thomas Hurtut, Farida Cheriet*

Dept. of Computer Engineering, Polytechnique Montréal, Montréal, QC H3C 3A7, Canada  
argyrios.christodoulidis@polymtl.ca

## ABSTRACT

Proliferative diabetic retinopathy is an important public health issue with deteriorating impact on the vision of its patient. In this study, a novel approach is proposed for the characterization of abnormal vessels based on a spatial point pattern method. Points of interest corresponding to vascular junctions are detected by a perceptual organization technique, and then second-order statistical measures are computed. Significant differences ( $p < 0.05$ ) between healthy retinal regions and areas with neovascularizations were obtained, which suggests that the second-order statistics could be used as a relevant feature to discriminate the abnormal from the normal vasculature. The relevance of the new measures was also evaluated with respect to an existing set of features using classification. The inclusion of a new second-order measure increases the characterization sensitivity against the already existing features from 75.76% to 84.85%.

**Index Terms**— Diabetic retinopathy, fundus imaging, neovascularization, spatial point pattern analysis, perceptual organization

## 1. INTRODUCTION

A considerable proportion of preventable blindness in the western world is attributed to complications in vision caused by diabetic retinopathy (DR) [1]. The changes that occur under DR can lead to the proliferation of new vessels, also called neovascularizations (NVs), which are triggered in response to the ischemic retina tissue [2]. NVs signify the proliferative DR (PDR), which is a severe case of DR and different from the non-proliferative stage (NPDR) stage. Even though the prevalence of PDR is only 0.4 percent of the total screening population [3], the appearance of NVs is an emergency where immediate treatment should be administered. Telemedicine as part of national screening programs is becoming more prevalent, increasing the number of available data. In this scenario, computerized assisted detection of NVs can be beneficial for timely diagnosis of the proliferative diabetic retinopathy.

Some efforts have been placed towards automatic characterization of the NVs within the optic disc only (NVDs) [4, 5]. NVDs are well contrasted against the background, owing to the fact that the optic disc is the brightest retinal structure. However, the appearance of NVs, regardless of their site, constitute the transition to the proliferative stage, and it is thus important for the algorithms to be able to detect new vessels in all the available retinal sites. Recent studies include the detection of NVs elsewhere (NVEs) [6, 7, 8, 9]. Most of the previous work is based on extracting features for classification of the retinal images in healthy against PDR cases. These include gray-scale or binary measurements on individual vessels' appearance, morphology, geometry [4], texture [5, 6] as well as fractal analysis [8]. Another work is based on counting the number of vascular junctions from the skeletonized segmentation of the multi-scale line detector (MSLD) [10]. This measure demonstrates good discrimination power compared to a set of features selected on the gray-level and the morphology of the vessels [9].

However, the straight line pattern used in MSLD is inadequate for vessel and junction segmentation at the small vessels or the neovascularizations. A. Christodoulidis et al. [11] combined the line detector with the tensor voting framework (TVF) [12] in a multi-scale way to overcome the small vessel segmentation problem. Still, the junction detection problem is not trivial. One advantage of using TVF for junction detection, is that it can infer the information based on the perceptual organization of the neighbor structures. Therefore, it can locate the junctions even if the MSLD method fails to extract any meaningful information. In this paper we choose this method to isolate the vascular junctions from already segmented vessels. To the best of our knowledge, this is the first time where TVF is used to isolate the retinal vascular junctions that are in turn used as points of interest.

Furthermore, the existing work is limited to first-order statistics to distinguish between healthy and abnormal vessels, such as counting the total number of junction, and their density in an image, or in a region. However, their relative spatial arrangement is not taken into account. Neovascularizations are highly inter-connected vessels. Hence, their junctions are closer than healthy cases. The utilization of second-order measures coming from the spatial point pattern analysis

---

Thanks to the Natural Science and Engineering Research Council of Canada (NSERC) for funding.

field [13] can give more discriminating information about the spatial organization of the vascular junctions. Thus, in addition to junction detection we exploit second-order statistics computed on the junction spatial distribution for characterizing the vessels as healthy or abnormal. Finally, we examine the relevance of the new measures as part of a classification scheme where several features are extracted from healthy and NVs regions, NVEs and NVDs, coming from PDR images.

## 2. METHODS

### 2.1. Junction Center Detection

**Vessel Extraction:** We apply the MSLD [10] on the green channel of the retinal images whose illumination is corrected and background noise is filtered, as in [11]. We apply the line detector, adjusting the line length parameter  $W$  according to the image resolution. The overall distribution of line response values varies according to the considered image and database, so we determine the simple threshold value adaptively by fitting a Gaussian function on the distribution of the MSLD response values of each image. We use the same threshold value  $T = \lfloor \mu_{\text{Gaussian}} \rfloor + \lfloor \sigma_{\text{Gaussian}} \rfloor$  for all the available images, where  $\mu_{\text{Gaussian}}$  is the mean value and  $\sigma_{\text{Gaussian}}$  is the standard deviation of the fitted function, respectively.

**Postprocessing:** Simple thresholding the line response leads to the retention of disconnected background structures that are either small in size or irregular in shape. In order to remove this type of noise we perform a conservative morphological analysis. We suppress all the disconnected components that have  $400 < \text{Area}_{CC} < 1000$  and that are not elongated enough ( $\text{Elongation}_{CC} < 0.99$ ), where elongation is defined according to [14]. Smaller components are completely removed, while larger components are kept regardless of their shape. The values of these parameters are chosen empirically aiming for background noise suppression, and at the same time retention of NV regions that are identified as PDR.

**Tensor Voting:** We detect the junctions by decomposing the second order tensor  $\mathbf{T}$  into its eigenvalues ( $\lambda_1, \lambda_2$ ), and measuring the ballness measure ( $\lambda_2$ ) there:

$$\mathbf{T} = \lambda_1 \hat{\mathbf{e}}_1 \hat{\mathbf{e}}_1^T + \lambda_2 \hat{\mathbf{e}}_2 \hat{\mathbf{e}}_2^T = (\lambda_1 - \lambda_2) \hat{\mathbf{e}}_1 \hat{\mathbf{e}}_1^T + \lambda_2 (\hat{\mathbf{e}}_1 \hat{\mathbf{e}}_1^T + \hat{\mathbf{e}}_2 \hat{\mathbf{e}}_2^T),$$

where  $\hat{\mathbf{e}}_{1,2}$  are orthogonal eigenvectors that hold the direction of the inferred lines at each pixel location. Furthermore, scale parameter  $\sigma_{\text{TVF}}$  defines the extend of the voting field and subsequently controls the distance in which the neighbor information is taken into account. The ballness measure of the tensor quantifies the local orientation uncertainty of the segmented vessels and it exhibits local maximum there. Since we are interested in extracting information from cluttered vessels low scale value should be used ( $\sigma_{\text{TVF}} = 20$ ).

**Junction Centers Isolation:** We first identify the regional maxima by applying the height-domes [15] algorithm. We then isolate the maximum values inside these domes. We

morphologically reconstruct the original image  $I$  from  $I - h_1$ , and then we subtract the result from the input image. This way, whole regions, unlike single pixels that demonstrate local maxima, can be identified. Additionally, we select the most prominent domes, namely the regions that have height over  $h_2$ . The values for the parameters are empirically set to  $h_1 = 2600$  and  $h_2 = 1400$  and depend on the value of the scale  $\sigma_{\text{TVF}}$ . Finally, we merge closely identified junctions that belong to the same region by finding the barycenter in an  $11 \times 11$  window. Figure 1 shows an example of the previous algorithmic steps in an image with NVDs.

### 2.2. Spatial Distribution Measures

We compute the following spatial statistics. Two versions of the **dispersion index** defined as: i) the variance over the mean number of junctions (VMR) across the healthy and the PDR group of images, and ii) the mean distance from the center of mass of junctions (VMR ( $\mu_d$ )). Additionally, we compute two versions of the  **$K^{\text{th}}$ -Nearest Neighbor index**. The first is the index described in [16], that in turn is based on the aggregation index  $R$  proposed by P.J. Clark [17], and defined as:

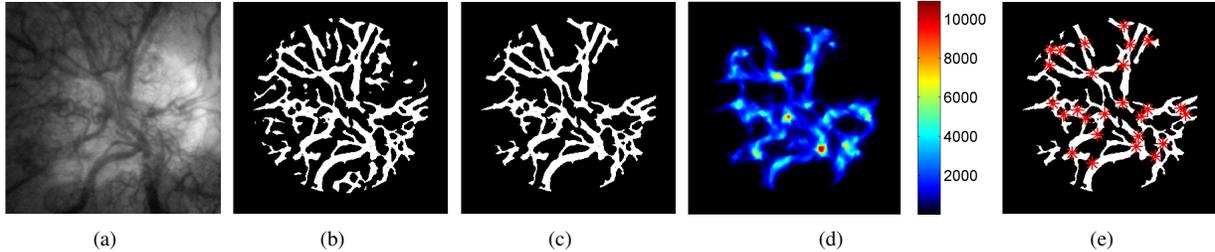
$$K^{\text{th}}\text{-NN}(K) = \frac{D(\text{observed})(K)}{D(\text{random})(K)} = \frac{\sum_{i=1}^K \mu(d_i)}{\frac{K(2K)!}{(2^K!)^2 \sqrt{\rho}}},$$

where  $\mu(d_i)$  is the mean nearest  $K^{\text{th}}$ -order neighbor distance between all the junctions,  $N$  is the number of points, and  $\rho$  is the density. The second version of the measure is just the numerator of the index, without the comparison against the random distribution which is also a second-order statistic that can quantify the spatial arrangement.

The first measure can quantify if the arrangement of points, or the area pattern, is more clustered or ordered than if it have been randomly arranged in the same region. The second measure summarizes, in pixels, the inter-junction distances. Both measures are straightforward to compute, and in contrast with the first-order measures they take into account distances. By applying these measures we can study the relative aggregation of the patterns in the two considered cases, the healthy and the PDR. Therefore, we can investigate if the assumption that the vessel junctions are denser in PDR holds. Finally, it allow us to study the spatial interactions between the junctions that give rise to the formation of the specific pattern in each case.

### 2.3. Classification Protocol

We assess the benefit of "incorporating" the newly proposed measures for classifying regions to PDR. For that reason, we extract 21 features from each region, that were first proposed in the literature [4, 9]. Furthermore, we include the following that were also proposed in [8]: local entropy, measurements



**Fig. 1.** Algorithmic steps for the junction center isolation. (a) PDR region, (b) vessel extraction after MSLD, (c) postprocessed result, (d) tensor voting ballness measure, and (e) final detected junction centers.

on the gray-level co-occurrence matrix (GLCM), and the proposed spatial distribution measures up to the 20<sup>th</sup> order for the  $K^{\text{th}}$ -NN case. Additionally, we compute the mean, median, minimum, maximum values of the new measures, as well as the slope of a linear function fitted on the proposed measures in Section 2.2. In total, 88 features thus are considered. We use a wrapper method for forward feature selection using a support vector machine (SVM) with an RBF kernel, as in [4, 9]. We dedicate 2/3 of the whole set for training and validation, and the rest for testing.

### 3. RESULTS

We perform our analysis in images collected from 5 databases, where image-level labeling is provided for the majority of the data (Diaretdb [18], HRF [19], Messidor [20], Kaggle [21], and a private one). In order to compare the healthy and the abnormal junctions we confine our analysis at a region-level.

#### 3.1. Anatomically Corresponding Regions

We apply the spatial point pattern analysis in anatomically corresponding circular regions. Generally, two regions from different images are corresponding if they are in approximately similar locations in the available field of view (FOV). A trained member isolated 114 regions with NVs from the 5 databases, each region originating from a single image. Then we transform the NV center into polar coordinates  $(R, \phi)$ . With  $R$  being the distance, in optic disc (OD) radius, between the OD center and the NV center, while  $\phi$  is the angle. In the healthy images we select the regions that are at  $R$  distance and  $\phi$  angle with respect to the OD, manually adjusting if the center is found outside of the FOV. Thus, each NV region is compared to a set of healthy regions. The area of the selected regions was set to one fifth of the total FOV area of each image. The healthy group of images contained 99 images. Overall, 5688 regions from the healthy group are compared with 114 abnormal regions. We perform the spatial analysis using the measures described in Section 2.3. Figure 2 shows an isolated region with NVDs along with the corresponding regions from the healthy cases. For each region we provide the segmented vessels with superimposed the junctions.

#### 3.2. Result on First-order Measure

The boxplot in Figure 3(a) shows the results on the number of junctions for the two cases. Generally, there is an overlap between the healthy and the diabetic groups. Even though we have a denser pattern in the PDR cases, just counting the number of junctions is not enough to highlight any differences between the healthy and the PDR group.

#### 3.3. Dispersion Indices

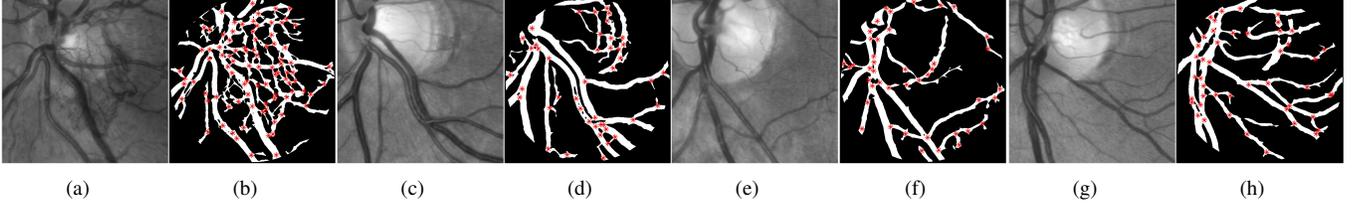
The value for the dispersion is different for the two groups for both measures (Table 1). Both the healthy and the PDR regions are over-dispersed with respect to the mean number of junctions. According to this measure, there are many instances where the count of junctions is higher or lower than the mean value of the corresponding group. Furthermore, the junctions are approximately 3 pixels closer to the center of mass of junctions, an artificial reference point, when a PDR case is encountered. These measure are not sensitive to the spatial arrangement of junctions in each region so they are invariant to changes in the underlying pattern, namely they cannot distinguish between a dense or a scattered pattern.

**Table 1.** Dispersion Indices Across the Different Cases

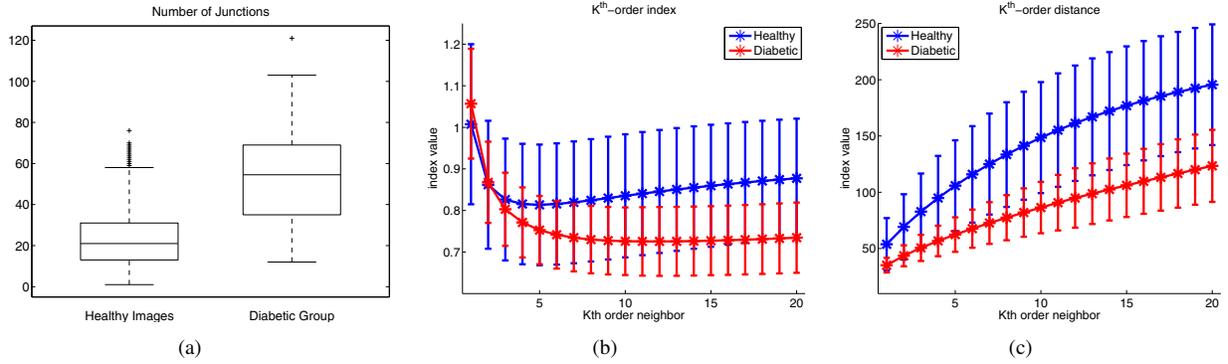
| Category       | VMR  | VMR ( $\mu_d$ ) |
|----------------|------|-----------------|
| <i>Healthy</i> | 8.95 | 11.13           |
| <i>PDR</i>     | 7.14 | 8.26            |

#### 3.4. $K^{\text{th}}$ -Order Nearest Neighbor Indices

The two measures based on  $K^{\text{th}}$ -NN ( $K$ ) take into account the inter-junction distances. The  $K^{\text{th}}$ -order index includes the  $K = 1$  case which only examines the nearest neighbor. Higher order nearest neighbors better highlight the differences between the two considered groups. The graphs in Figures 3(b) and 3(c) show the results. For the first case, values under one show that the arrangement of points, or the area pattern, is more clustered than the random distribution. In PDR, we have a generally more clustered pattern than the healthy



**Fig. 2.** Isolated, segmented areas, and junctions for analysis in HRF database. (a-b) Isolated PDR region, (c-h) anatomically corresponding healthy areas together with superimposed their isolated junctions (red stars).



**Fig. 3.** (a) Boxplot of junction counts. (b)  $K^{\text{th}}$ -order index measurements over the expected random distribution of the junctions. (c)  $K^{\text{th}}$ -order distance of the junctions for each group.

regions (20<sup>th</sup>-order Mann-Whitney U-test:  $p < 0.05$ ). Furthermore, we show that the mean inter-junction distance, when it is not compared against the random distribution, is considerable lower for the diabetic than the healthy group (20<sup>th</sup>-order Mann-Whitney U-test:  $p < 0.05$ ).

### 3.5. Classification

We apply a 10-fold feature selection process on two different feature sets: (a) including 39 features proposed in the literature, and (b) including 88 features, with 49 newly proposed. In both cases the following features are selected: mean vessel wall gradient ( $Wall(\mu_{\nabla})$ ), number of junctions obtained by skeletonizing the segmented vessels ( $Skel(J)$ ), mean gray level ( $Seg(\mu_I)$ ), mean local gray level ( $\mu_I$ ), and local entropy ( $E$ ). Additionally, in the (a) case the number of vessel orientations ( $Dir_N$ ) is included, while in the (b) case the  $K^{\text{th}}$ -NN (2) without the random comparison is added. Replacing the number of vessel orientations with the proposed second-order feature improves the classification sensitivity by approximately 9%, and slightly the accuracy (Table 2).

**Table 2.** Classification Performance per Feature Set Case

| Selected Features | Sensitivity | Specificity | Accuracy | Precision |
|-------------------|-------------|-------------|----------|-----------|
| (a) feature set   | 75.76%      | 99.94%      | 99.48%   | 96.15%    |
| (b) feature set   | 84.85%      | 99.94%      | 99.66%   | 96.55%    |

## 4. CONCLUSION AND DISCUSSION

In this paper we proposed a new method for the characterization of the retinal blood vessels. This is important in the direction of assisting the physician in the effective detection of this category of vessels. For the first time, TVF is applied for the vessel characterization problem as well as the distribution analysis of junctions with measures that originate from the spatial point pattern analysis theory. We observe significant differences between healthy and regions with PDR when we apply our analysis. Finally, including the  $K^{\text{th}}$ -NN ( $K$ ) distance measure improves the classification performance.

Accurate a priori information about the vessel location is crucial for the junction detection step. TVF infers the information from the neighborhood so it can partially compensate for missed segmentation at junctions which is not the case for junction detection from the skeleton. In parallel, the post-processing and the center isolation step ensure the suppression of irrelevant junctions. However, their parameter values are empirically set. Future work will focus towards the application of the proposed methodology for accomplishing the detection task. This could be achieved by training the features on non-proliferative diabetic retinopathy cases and then testing on whole images. Moreover, higher order tensors could be used to resolve the orientation uncertainty problem, and thus distinguish junctions from crossings. This could allow to further study the differences in the distribution of different kinds of vessel junctions under healthy or abnormal cases.

## References

- [1] Sobha Sivaprasad, Bhaskar Gupta, Roxanne Crosby-Nwaobi, and Jennifer Evans, "Prevalence of diabetic retinopathy in various ethnic groups: a worldwide perspective," *Surv. Ophthalmol.*, vol. 57, no. 4, pp. 347–370, 2012.
- [2] Joan W Miller, Anthony P Adamis, and Lloyd Paul Aiello, "Vascular endothelial growth factor in ocular neovascularization and proliferative diabetic retinopathy," *Diabetes Metab. Rev.*, vol. 13, no. 1, pp. 37–50, 1997.
- [3] Sam Philip, Alan D Fleming, Keith A Goatman, Sofia Fonseca, Paul McNamee, et al., "The efficacy of automated disease/no disease grading for diabetic retinopathy in a systematic screening programme," *Br. J. Ophthalmol.*, vol. 91, no. 11, pp. 1512–1517, 2007.
- [4] Keith A Goatman, Alan D Fleming, Sam Philip, Graeme J Williams, John A Olson, and Peter F Sharp, "Detection of new vessels on the optic disc using retinal photographs," *IEEE Trans. Med. Imag.*, vol. 30, no. 4, pp. 972–979, 2011.
- [5] Carla Agurto, Honggang Yu, Victor Murray, Marios S Pattichis, Sheila Nemeth, et al., "A multiscale decomposition approach to detect abnormal vasculature in the optic disc," *Comput. Med. Imag. Graph.*, vol. 43, pp. 137–149, 2015.
- [6] Carla Agurto, Victor Murray, Eduardo Barriga, Sergio Murillo, Marios Pattichis, et al., "Multiscale anfm methods for diabetic retinopathy lesion detection," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 502–512, 2010.
- [7] Siti Syafinah Ahmad Hassan, David BL Bong, and Mallika Premsenthil, "Detection of neovascularization in diabetic retinopathy," *J. Digit. Imaging*, vol. 25, no. 3, pp. 437–444, 2012.
- [8] Jack Lee, Benny Chung Ying Zee, and Qing Li, "Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy," *PLoS one*, vol. 8, no. 12, pp. e75699, 2013.
- [9] RA Welikala, MM Fraz, J Dehmeshki, A Hoppe, V Tah, S Mann, et al., "Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy," *Comput. Med. Imag. Graph.*, vol. 43, pp. 64–77, 2015.
- [10] U.T. Nguyen, A Bhuiyan, L. A. Park, and K. Ramamohanarao, "An effective retinal blood vessel segmentation method using multi-scale line detection," *Pattern Recognition*, vol. 46, no. 3, pp. 703 – 715, 2013.
- [11] Argyrios Christodoulidis, Thomas Hurtut, Housseem Ben Tahar, and Farida Cheriet, "A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images," *Comput. Med. Imag. Graph.*, vol. 52, pp. 28–43, 2016.
- [12] Gerard Medioni and Sing Bing Kang, *Emerging topics in computer vision*, Prentice Hall PTR, 2004.
- [13] Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan, *Statistical analysis and modelling of spatial point patterns*, vol. 70, John Wiley & Sons, 2008.
- [14] Robert M. Haralick and Linda G. Shapiro, *Computer and Robot Vision*, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1992.
- [15] Luc Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, 1993.
- [16] HR Thompson, "Distribution of distance to nth neighbour in a population of randomly distributed individuals," *Ecology*, vol. 37, no. 2, pp. 391–394, 1956.
- [17] Philip J Clark and Francis C Evans, "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology*, vol. 35, no. 4, pp. 445–453, 1954.
- [18] Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen, Lasse Lensu, Iiris Sorri, et al., "The diaretdb1 diabetic retinopathy database and evaluation protocol," in *BMVC*, 2007, pp. 1–10.
- [19] Jan Odstřcilik, Radim Kolar, Attila Budai, Joachim Hornegger, Jiri Jan, et al., "Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database," *IET Image Processing*, vol. 7, no. 4, pp. 373–383, 2013.
- [20] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Laÿ, Béatrice Cochener, et al., "Feedback on a publicly distributed image database: the messidor database," *Image Analysis and Stereology*, vol. 33, no. 3, pp. 231–234, 2014.
- [21] "https://www.kaggle.com/c/diabetic-retinopathy-detection/data," Accessed: 2016-11-03.