Medical Image Analysis 32 (2016) 216–232

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A fully automated tortuosity quantification system with application to corneal nerve fibres in confocal microscopy images

Roberto Annunziata[a,*], Ahmad Kheirkhah[c], Shruti Aggarwal[c], Pedram Hamrah[b,c], Emanuele Trucco[a]

[a] Computer Vision and Image Processing Group, VAMPIRE Project, School of Science and Engineering (Computing), University of Dundee, Dundee, UK
[b] Boston Image Reading Center and Cornea Service, New England Eye Center, Tufts Medical Center, Boston, USA
[c] Ocular Surface Imaging Center, Cornea Service, Massachusetts Eye & Ear Infirmary, Department of Ophthalmology, Harvard Medical School, Boston, USA

ARTICLE INFO

Article history:
Received 19 November 2015
Revised 20 April 2016
Accepted 20 April 2016
Available online 22 April 2016

Keywords:
Tortuosity
Cornea
Multiscale
Segmentation
Curvature
Automated

ABSTRACT

Recent clinical research has highlighted important links between a number of diseases and the tortuosity of curvilinear anatomical structures like corneal nerve fibres, suggesting that tortuosity changes might detect early stages of specific conditions. Currently, clinical studies are mainly based on subjective, visual assessment, with limited repeatability and inter-observer agreement. To address these problems, we propose a fully automated framework for image-level tortuosity estimation, consisting of a hybrid segmentation method and a highly adaptable, *definition-free* tortuosity estimation algorithm. The former combines an *appearance* model, based on a *Scale and Curvature-Invariant Ridge Detector* (SCIRD), with a *context* model, including multi-range learned context filters. The latter is based on a novel tortuosity estimation paradigm in which discriminative, *multi-scale* features can be automatically learned for specific anatomical objects and diseases. Experimental results on 140 *in vivo* confocal microscopy images of corneal nerve fibres from healthy and unhealthy subjects demonstrate the excellent performance of our method compared to state-of-the-art approaches and ground truth annotations from 3 expert observers.

1. Introduction

The tortuosity of curvilinear structures in the human body has received much attention since the seminal papers by Edington (1901) and Cairney (1924).

Numerous studies have reported correlations between several pathologies and the tortuosity of a wide range of anatomical structures such as retinal vessels (Heneghan et al., 2002; Cheung et al., 2011; Sasongko et al., 2011, 2015; Longmuir et al., 2010; Muraoka et al., 2014; Maude et al., 2014), intracerebral vessels (Bullitt et al., 2005) and conjunctival blood vessels Owen et al. (2008), but also the coronary (Eleid et al., 2014), iliac (Coulston et al., 2014), carotid (Bogunović et al., 2012) and aortic (Franken et al., 2015) arteries, the optic nerve (Ji et al., 2013) and corneal nerve fibres (Kallinikos et al., 2004; Hamrah et al., 2010; Kurbanyan et al., 2012; Hamrah et al., 2013; Edwards et al., 2014).

The pathologies investigated affect a large portion of the population worldwide and include diabetes, diabetic retinopathy and diabetic neuropathy (Sasongko et al., 2011, 2015; Edwards et al., 2014), retinopathy of prematurity (Heneghan et al., 2002; Wilson et al., 2008), malignant gliomas (Bullitt et al., 2004), facioscapulohumeral muscular dystrophy (Longmuir et al., 2010), spontaneous coronary artery dissection (Eleid et al., 2014), central retinal vein occlusion (Muraoka et al., 2014), children with neurofibromatosis type 1 (Ji et al., 2013). In particular, tortuosity has been investigated in corneal diseases such as unilateral herpes zoster (Hamrah et al., 2013), herpes simplex keratitis (Hamrah et al., 2010), acute acanthamoeba and fungal keratitis (Kurbanyan et al., 2012) and diabetic neuropathy (Kallinikos et al., 2004; Edwards et al., 2014).

In several studies (e.g., Hamrah et al., 2013; Eleid et al., 2014; Muraoka et al., 2014; Lagali et al., 2015), tortuosity was assessed by experienced specialists typically grading structures or whole images on a 3–5 level scale or as normal/abnormal based on qualitative, albeit structured, protocols (Oliveira-Soto and Efron, 2001). Regardless of the specific anatomical object of interest, such assessment is subjective and can lead to substantial inter-observer variability and possibly non-negligible intra-observer variability, thus limiting the sensitivity of the assessment scheme. Moreover, requiring direct inspection by specialists limits the amount of images that can be analyzed and make large screening programs unfeasible or at least very expensive.

* Corresponding author.
E-mail address: r.annunziata@dundee.ac.uk (R. Annunziata).

http://dx.doi.org/10.1016/j.media.2016.04.006
1361-8415/

Several definitions of tortuosity have been proposed to try and quantify tortuosity automatically (Hart et al., 1999; Dougherty and Varro, 2000; Bullitt et al., 2003; Kallinikos et al., 2004; Wilson et al., 2008; Grisan et al., 2008; Trucco et al., 2010), but no single definition is widely accepted, possibly because tortuosity has different characteristics for different anatomical structures. We argue therefore that a highly adaptable tortuosity estimation algorithm, learning key features for specific image types and structures (*definition-free*) would be a promising and effective new research target.

However, efficiency would still be limited by manual segmentation, especially with today's image resolutions, e.g. 3500×2300 pixels for standard fundus camera images (Budai et al., 2009). Fast and accurate curvilinear structure segmentation is therefore needed, but different characteristics of tortuous curvilinear structures across image modalities makes segmentation challenging. In fact, tortuousity violates one of the basic assumptions of most tubular structure detectors, namely *locally straight tubular shape* (Frangi et al., 1998; Soares et al., 2006; Law and Chung, 2008; Hannink et al., 2014). Further issues include the presence of non-target structures (clutter), low resolution, noise and non-uniform illumination.

Fully automated tortuosity estimation frameworks have been proposed for retinal blood vessels, brain vasculature and corneal nerve fibres (Heneghan et al., 2002; Joshi et al., 2010; Scarpa et al., 2011; Koprowski et al., 2012). Typically, inaccuracies in the segmentation are the main source of inaccurate tortuosity estimates (e.g., Scarpa et al., 2011). Moreover, the aforementioned methods are based on mathematical definitions of tortuosity providing *fixed* models of a subjective perception. These measures combine features such as the number of inflection points along the structure, and the length to chord ratio. We argue that such hand-crafted tortuosity definitions limit the accuracy of an automated framework of general value.

1.1. Contributions

We propose a novel, fully automated framework for tortuosity estimation and demonstrate its performance with corneal nerve fibres in *in vivo* confocal microscopy (IVCM) images. Our system relies on a fast, robust and accurate segmentation algorithm and on a novel formalization of tortuosity estimation. Validation is carried out both in terms of overall agreement with three experienced specialists, and of performance of each system module compared to several state-of-the-art methods. In addition, we quantify the accuracy in terms of tortuosity estimation when replacing manual segmentation (normally taken as ground truth for segmentation) with the proposed automated segmentation.

Our novel segmentation algorithm is specifically designed for tortuous and fragmented structures. It combines the benefits of an *appearance* model based on the *Scale and Curvature Invariant Ridge Detector* (SCIRD) and a *context* model, a multi-scale hierarchy based on learned context filters.

Finally, we propose a novel, adaptable framework for tortuosity estimation by learning flexibly important features for specific image modalities and structures. This is accomplished by tortuosity feature selection (FS) and supervised classification.

1.2. Related work

1.2.1. Curvilinear structure segmentation

Many solutions have been proposed for curvilinear structure segmentation. Challenges include low signal-to-noise ratio at small scales, confounding non-target structures, non-uniform illumination and complex configurations (see Lesage et al. (2009) for an extensive review).

The extraction of adequate features is a fundamental aspect and relies mostly on the following three approaches.

(1) Hand-crafted features (HCFs). These methods are based on hand-designed filters modeling local geometrical properties of ideal tubular structures. Eigenvalue decomposition of the Hessian matrix was employed by Frangi et al. (1998), Santamaría-Pang et al. (2007), Martinez-Perez et al. (2007), Annunziata et al. (2015a), maximum projections over each scale were used to make the approach scale invariant. These projections were then used to build the well-known tubularity measure called *vesselness* by Frangi et al. (1998). However, performance tend to degrade at crossings or bifurcations since this approach only looks for elongated structures. To overcome this issue, Hannink et al. (2014) proposed to segment crossings/bifurcations with multiscale invertible orientation scores and apply vesselness filters to maps of the latter. Optimally Oriented Flux (OOF) was recently proposed by Law and Chung (2008) to improve detection of adjacent structures with vesselness measures. OOF is based on the computation of an optimal projection direction minimizing the inward oriented flux at the boundary of localized circles (2-D) or spheres (3-D) of different radii. This projected flux can be regarded as a likelihood that a pixel is part of a tubular structure. Tubularity measures can be obtained by combining the eigenvalues of the OOF Hessian matrix. Other successful HCFs rely on Gabor wavelets (aka Morlet wavelets) as proposed, for instance, by Soares et al. (2006) who exploited their directional selectiveness to detect oriented structures and achieve fine tuning to specific frequencies. HCFs have also been used by Honnorat et al. (2011) to compute a local tubularity measure feeding a graphical model. While HCFs are typically fast, their assumptions might be violated in real images by highly fragmented and tortuous structures, which limits detection performance and consequently the accuracy of tortuosity estimation.

(2) Fully learned architectures (FLA). These are designed to overcome HCFs modeling issues by learning object representations directly from training data, with excellent performance reported on several tasks (Farabet et al., 2013; Krizhevsky et al., 2012). FLA learn automatically the specific characteristics of the training set, such as inter-pixel dependencies. However, training complex learning architectures requires high-performance hardware and/or optimised implementations, and more importantly large datasets to avoid overfitting, which are not always available for medical images. For this reason, Becker et al. (2013) have recently proposed a less complex solution employing gradient boosting to learn convolutional filters and boosting weights simultaneously, applied with success to retinal blood vessels and neurites. Sironi et al. (2014) have used the responses of convolutional filters learned by sparse coding (henceforth, SC) as input features to multiple regressors trained to predict the distance from the centerline.

(3) Hybrid methods (HM). This solution combines HCFs with filters learned by FLAs, exploiting the efficiency of fast HCFs while limiting the amount of learned filters. The first HM applied to tubular structures was proposed by Rigamonti and Lepetit, 2012) and combines OOF with learned *appearance* filters, i.e. learned on the original training images. It employs convolutional SC to learn 9 appearance filters. Quantitative results show a clear improvement over methods based only on HCFs, achieving the same level of performance obtained with a filter bank of 121 filters learned via SC, at a limited computational cost. However, applying learned filters at the same *layer* where HCFs are then applied (i.e. original image patches) may result in redundant learned filters, potentially damaging the discrimination power of the feature set.

After feature extraction, or in combination with feature learning in the case of fully learned architecture, a supervised classifier is typically employed to find optimal decision boundaries in the