Neurocomputing 126 (2014) 71–77

Contents lists available at ScienceDirect

Neurocomputing

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

Random forest active learning for AAA thrombus segmentation in computed tomography angiography images

Josu Maiora, Borja Ayerdi, Manuel Graña *

Computational Intelligence Group, UPV/EHU, Spain

ARTICLE INFO

Article history:
Received 27 March 2012
Received in revised form
21 November 2012
Accepted 10 January 2013
Available online 2 August 2013

Keywords:
Active learning
Random forests
CTA image segmentation
Abdominal aortic aneurysm segmentation

ABSTRACT

Image segmentation of 3D Computed Tomography Angiography (CTA) is affected by a variety of noise conditions that may render ineffective image segmentation procedures that have been developed and validated on a collection of training CTA data when applied on new CTA data. The approach followed in this paper to tackle this problem is to provide an Active Learning based interactive image segmentation system which will allow quick volume segmentation, with minimal intervention of a human operator. Image segmentation is achieved by a Random forest (RF) classifier applied on a set of image features extracted from each voxel and its neighborhood. An initial set of labeled voxels is required to start the process, training an initial RF. The most uncertain unlabeled voxels are shown to the human operator to select some of them for inclusion in the training set, retraining the RF classifier. The approach is applied to the segmentation of the thrombus of Abdominal Aortic Aneurysm (AAA) in CTA data (of patients), showing that the CTA volume can be accurately segmented after few iterations requiring a small labeled data sample.

1. Introduction

An abdominal Aortic Aneurysm (AAA) is a local dilation of the Aorta that occurs between the renal and iliac arteries. The weakening of the aortic wall leads to its deformation and the generation of a thrombus. 3D Computerized Tomography Angiography (CTA) allows minimally invasive visualization of the Aorta's lumen, thrombus and calcifications. The segmentation of the AAA thrombus is a challenging task due to the low contrast of signal intensity values between the AAA thrombus and its surrounding tissue, as can be appreciated in Fig. 1. Furthermore, the AAA thrombus shows great shape variability, both intra and inter-subjects, so that little prior information is available to guide the segmentation. General reviews of blood vessel segmentation methods are given in [1,2].

AAA thrombus segmentation methods reported in the literature need a lot of human interaction or *a priori* information one way or the other. A method based on Active Shape Models is described in [3] which needs initial manual labeled landmark points in one slice. The initial contour is propagated to neighboring slices on the basis of grayscale similarities. A classification approach that needs an initial manual segmentation of the Aorta lumen is proposed in [4]. The classifier features are grayscale profiles extracted from the normal to the lumen surface following a careful manual sampling procedure. In [5] an initial rough specification of the aneurysm surface is refined by means of level set segmentation driven by an *a priori* model and the likelihood estimation provided by Support Vector Machine classifiers trained on voxel location, intensity and texture features. In [6] a deformable NURBS model is driven by a probability map built from a Gaussian Mixture Model trained on selected samples. This approach needs an initial manual lumen segmentation and intensity renormalization to avoid convergence mishaps of the NURBS model adaptation. In [7] the AAA thrombus after endovascular repair is detected following a radial model approach needing the specification of the lumen centerline and some manually tuned correction performed on the polar coordinate representation of the image. A graph-cut approach constrained by a geometrical model is proposed in [8], needing a previous lumen segmentation and centerline computation. The approach iterates labeling and geometric model re-estimation, which are costly processes.

The approach followed in this paper for AAA thrombus segmentation is to build a voxel[1] classifier into AAA thrombus or background classes [8,4,5]. Classification approaches need careful selection and labeling of training data samples from the available data. In response to this issue, Active Learning [9] tries to achieve the most accurate classification using the smallest possible training set, minimizing the user interaction needed to label the training samples. Active Learning starts with a minimal training sample, adding new labeled samples in an iterative process.

* Corresponding author. Tel.: +34 943018044.
E-mail address: manuel.grana@ehu.es (M. Graña).

[1] In this paper we work with 3D images, so that each image element is a voxel instead of a pixel.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.01.051

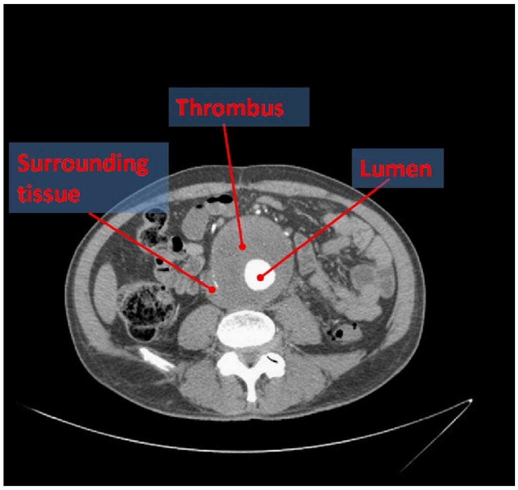

Fig. 1. CTA axial slice depicting thrombus and lumen of AAA. Blood in lumen is highlighted due to contrast, but thrombus intensity levels are similar to other surrounding tissues.

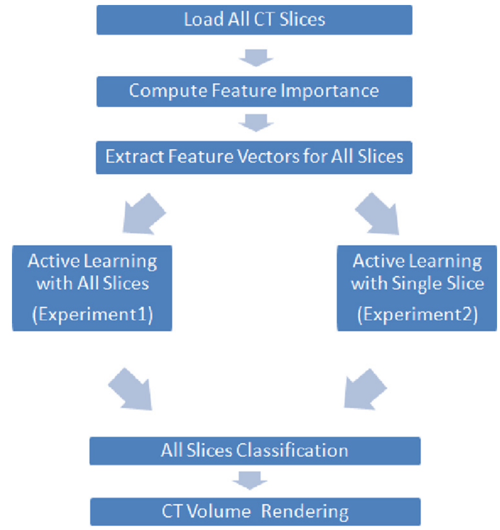

Fig. 2. Pipeline of the experimental setup for the Active Learning segmentation process.

Aiming to provide the greatest increase in classifier accuracy [10], the additional samples are selected according to some classification uncertainty measure.[2] Besides its benefits in economy of computation and data labeling, Active Learning is also useful when the underlying data statistics are non-stationary, so that the classifier built at one time instant may not be optimal later on. Active Learning has been successfully applied to the classification of remote sensing images [11–13], and image retrieval based on semisupervised Support Vector Machines [14]. Active Learning inspiration for the selection of a minimal collection of training images is proposed in [15] for the development of combined generative and discriminative models in the segmentation of CT scans. An active feedback approach is used in [16] to improve the classification based annotation of radiographs. From a clinical application point of view, Active Learning may allow the expert radiologist to obtain fast and accurate segmentations with minimal interaction, despite strong changes in the CTA data.

We perform voxel classification using a Random forest (RF) classifier based on intensity features computed on each voxel's neighborhood. These features are maximum, minimum, median and Gaussian weighted average of the 2D neighborhoods of the voxel of increasing radius. The RF is convenient for this task because of its superior reported performance [17], simplicity of training and robustness, which have been shown in reported applications to medical image segmentation. Specific examples are myocardium delineation in 3D ultrasound (US) imaging of adult hearts [18], brain tissue segmentation [19,20] and segmentation of soft organs in abdominal and thoracic CT volumes [21,22]. In [23] we provided some first results of the approach proposed in this paper. Moreover, the hybridization [24,30,31] of Active Learning and RF is possible because RF allows for the straightforward definition of a classification uncertainty measure, namely, the variance of the individual tree classifiers' outputs.

The specific contributions of the approach proposed in this paper relative to the state of the art of AAA thrombus segmentation algorithms are as follows: (1) the need for human intervention in the selection of samples and labeling is reduced to a minimum by Active Learning, (2) use of RF allows quick learning and adaptation to incremental training datasets, (3) there is no requirement of a priori information or geometric models, (4) feature extraction does not require sophisticated data processing, (5) the adaptation of the classifier to new data does not require skilful data processing, only picking the most uncertain voxels over a data visualization.

The experimental setup for validation is illustrated in Fig. 2. The paper tests two validation strategies. On one hand (Experiment 1), we train separate RF classifiers by Active Learning on each CT slice known to contain part of the thrombus. On the other hand (Experiment 2), we apply one RF classifier trained by Active Learning on the volume's central slice to the remaining slices of the volume, in order to test the generalization power of the approach. In both the experiments, the Active Learning oracle providing the samples' labels in the reported experiments is the ground truth provided by manual segmentation.

The structure of the paper is as follows: Section 2 describes the learning and feature selection methods. Section 3 describes the experimental set-up. Section 4 provides the experimental results. Finally, Section 5 provides our conclusions and some further work ideas.

2. Learning and feature selection

2.1. Random forest classifiers

Random forest (RF) algorithm is a classifier [25] that encompasses bagging [26] and random decision forests [27,28] is being used in a variety of applications [17]. RF became popular due to its simplicity of training and tuning while offering a similar performance to boosting. Consider a RF as a collection of decision tree predictors, built so that they are as much decorrelated as possible, denoted as

$$\{h(\mathbf{x};\psi_t); t=1,\dots,T\},$$

where $\mathbf{x}$ is a d-dimensional random sample of random vector X, ψ_t are independent identically distributed random vectors whose nature depends on their use in the tree construction, and each tree casts a unit vote to find the most popular class of input $\mathbf{x}$. RF captures complex interaction structures in data, and are proposed [25] to be resistant to both over-fitting of data when individual trees are very deep and no pruned, and under-fitting when individual trees are too shallow.

[2] The classification uncertainty measure does not require actual knowledge of the data sample label, thus no double-dipping is incurred.