A robust and accurate approach to automatic Blood Vessel detection and segmentation from Angiography X-ray images using multi-stage Random Forests

Vipin Gupta[a] and Amit Kale[b] and Hari Sundar[c]

[a]Philips Research,Bangalore,India;
[b]Siemens Information Systems Ltd.,Bangalore,India
[c]Siemens Corporate Research, Princeton, US

ABSTRACT

In this paper we propose a novel approach based on multi-stage random forests to address problems faced by traditional vessel segmentation algorithms on account of image artifacts such as stitches organ shadows etc.. Our approach consists of collecting a very large number of training data consisting of positive and negative examples of valid seed points. The method makes use of a 14×14 window around a putative seed point. For this window three types of feature vectors are computed viz. vesselness, eigenvalue and a novel effective margin feature. A random forest RF is trained for each of the feature vectors. At run time the three RFs are applied in succession to a putative seed point generated by a naiive vessel detection algorithm based on vesselness. Our approach will prune this set of putative seed points to correctly identify true seed points thereby avoiding false positives. We demonstrate the effectiveness of our algorithm on a large dataset of angio images.

Keywords: Vessel Segmentation, Random Forests, Robust Detection

1. INTRODUCTION

Blood vessels are tubular objects, which are visible in the X-ray (fluoro images) with the help of injected contrast agent. Blood vessel enhancement and segmentation is crucial for many clinical diagnostic and planning tasks. Blood vessels are used as features, for successful registration of the intra-operative fluoro images with pre-operative 3D images (CT/MR/Angio) .[1] This registration is also helpful in neurointerventions and aortic stenting procedure.

This work focuses on segmentation of cardiac blood vessels or coronary arteries. A large variety of different approaches has been proposed in literature for image enhancement and automatic segmentation of coronary arteries (blood vessels) from Angiography X-ray images[2–4] . In general, these approaches involve blood vessel enhancement using various techniques such as directional filter bank ,[5] matched filters,[6] and Hessians followed by using some heuristics to analyze these local features. These approaches can efficiently handle the multi-scale nature of blood vessel. However, when applied on data generated by widely varying settings of X-ray dosage settings and amount of contrast agent applied problems may arise in these naive approaches. These problems are further compounded by the presence of organs projections, surgical stitches and leaking contrast agents. While simple heuristics can be employed to eliminate isolated false positives, usually a cluster of false positive seeds which can happen in case of presence of surgical stitches cannot usually be eliminated (See Figure 1).

Typically vessel segmentation algorithms consist of some variant of three main stages: seed selection, DVI(Directional Vector Integration) based fiber growing, centerline extraction and vessel pruning. We assume that a simple seed selection algorithm possibly with a high false positive rate has been applied to generate likely seed points. In this paper we propose to address the problems faced by conventional methods by employing a novel approach based on multi-stage random forests to prune the seeds produced by simple vesselness based measures. Our approach consists of collecting a very large number of training samples coming from angio images which involves the user clicks on positive examples:interior of the blood vessel and negative examples : points away from the

Further author information: (Send correspondence to Amit Kale.: E-mail: kale.amit@siemens.com, Telephone: 91 80 3340 2100

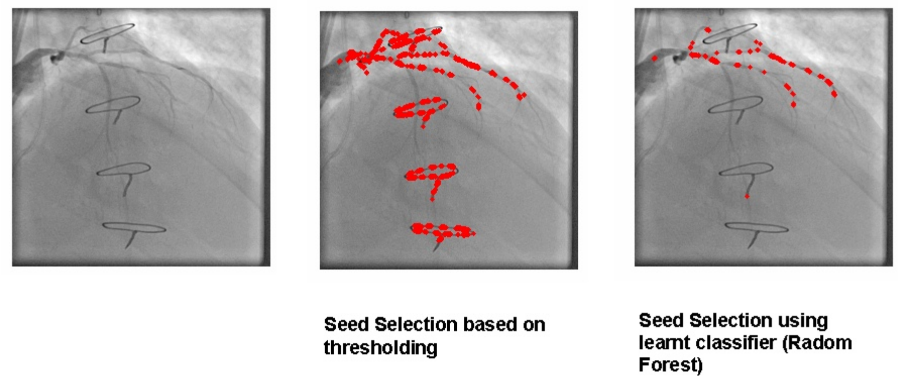

Figure 1. Seed detection for vessel detection in challenging angio images (with surgical stitches). Original Image (left) Simple vesselness based approach result (center) Proposed approach result(right)

blood vessel, noisy regions, organ projections, surgical stitches. The method makes use of a 14×14 window a putative seed point. For this window three types of feature vectors are computed viz. vesselness, eigenvalue and a novel effective margin feature. A RF is trained for each of the feature vectors. This can be motivated as follows. Vesselness feature is a good acceptor but a bad rejector i.e. give lot of false positives. On the other hand Eigen values(EV) are good rejectors but proven to be a bad acceptor i.e it gives many missed detections. V and EV are robust features but they sometimes fail in cases of organ shadow and imaging artifacts. To overcome this, we have chosen a feature based on EM (Effective Margin). EM is computed as follows. For a given region a histogram of the intensity values is computed. In case of organ shadows, the histogram is bimodal with two nearly equal peaks, for non vessel regions it is flat where as for vessels it is bimodal with two unequal peaks. This makes it a good rejector for organ shadows. At run time, each of the classifiers is applied to a putative seed in the above order before deeming it to be a true seed point. In Section 2 we describe the proposed algorithm details and sample results. In section 4, we conclude our work and discuss future directions.

2. METHOD

As noted earlier, vessel segmentation algorithms consist of three main stages: seed selection, DVI(Directional Vector Integration) based fiber growing, centerline extraction and vessel pruning. We focus on pruning the seeds produced by a naive seed detection algorithm which we describe next.

2.1 Vessel Detection (Seed Generation)

Seed generation is the first step of the algorithm. Here, seeds are the points which are most likely to lie on the vessels.[2] The selection of pixels which qualify to be seeds depends on identifying an image property which corresponds to the tubular structure of the blood vessel. The second-order structure of local intensity variations [2] has been shown to be one such image property. It is obtained by convolution of the image by partial second order derivative of Gaussian kernel. The second order property of Hessian Eigensystem has following characteristics: - small curvature along tube direction - large curvature along perpendicular direction This translates to small value of first eigenvalue $\lambda 1$ and first eigenvector $\nu 1$ in the direction of the tube and larger value of second eigenvalue $\nu 2$ with the eigenvector perpendicular to the tubular structure. Hessian based eigen systems are usually applied as multi-scale approach in order to enhance tubular structures in a certain range of scales $\sigma 1, \sigma 2, \cdots, \sigma N$, which is defined below:

$$V = max_{\sigma \in \sigma_1, \ldots, \sigma_N}(\sigma^{\gamma}.V_{\sigma}) \tag{1}$$

Where, V_σ denotes the singlescale vesselness as a function of λ_1 and λ_2 at scale σ.

Figure 2. Examples of positive and negative examples used for training (a). Steps used in training for classifier (b)

2.1.1 Random Forest based Seed Classification

To train the RF classifier, we collect large number of positive and negative samples from x-ray angiography datasets. For each frame we hand select positive and negative examples for vessels. Examples of these positive and negative examples are shown in Fig.2.

We first compute the vesselness map, eigenvalue map and eigenvector map noted in Section 2.1for the full image. One way to build a robust seed classifier would be to concatenate these entities for positive and negative examples of seed points. A classifier can be then trained using this data. One problem with such an approach however is that it only considers pointwise information without regard to the context. Furthermore it ignores specific attributes of vesselness and eigenvalues.

We propose to address these issues as follows: The feature vector is computed by vectorizing a window taken across neighborhood of the sample. Experimentally, a neighborhood size of 14x14 was found to be optimal. The feature vector is reduced in dimensionality by choosing feature values for every alternate location in the neighborhood grid. Note that choosing a neighboring region without taking account the orientation of the vessel can cause large intra-class variability in the positive examples in the training data. The next issue in the seed pruning stage is the choice of classifier.

We use a popular pattern classification algorithm viz. Random Forests (introduced by Leo Breiman .[7] This entails growing multiple binary randomized trees using a large amount of training set. Random Forests have been proven to be one of the best classifiers available till date and they have the following advantages:

- Generalization error of a random forest converges.
- Randomness injected at each node by subsampling of features, gives reduced variance and robustness to outliers
- Robust to mislabeling of training data (empirically proved) as compared to best of the classification algorithms (e.g. Adaboost)
- Easily parallelizable in contrast to boosting based approaches. Each randomized tree can be grown in parallel and final label can be computed just by voting.
- Can handle and incorporate intra-class multiple variations in the classifier design.

Specifically, in the context of seed pruning, a RF based classifier can automatically perform clustering of the multimodality in the positive examples as regards their orientation. One way to learn the RF classifier is to simply concatenate the above feature values into a high dimensional feature vector. However the following observations were made about the vesselness and eigen value features:

- Vesselness(V) feature is a good acceptor but proven to be a bad rejector i.e. give lot of false positives.
- On the other hand Eigen values(EV) are good rejectors but proven to be a bad acceptor i.e it gives many missed detections.

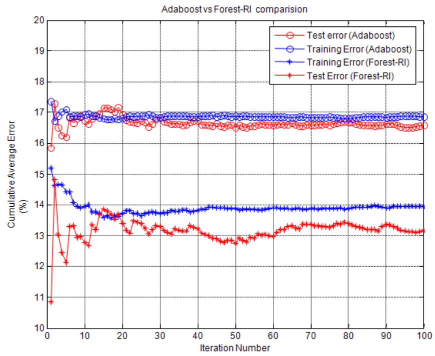

Figure 3. Training and Testing errors obtained for random forest and Adaboost based classifiers. Note that the RF paradigm using its implicit clustering is better capable of handling intraclass variability.

This makes it more sensible to train separate classifiers based on vesselness and eigen values separately and apply them in succession on the test seed. We train standalone RF classifiers on the vesselness and eigenvalue features. At run time we first evaluate a putative seed using the vesselness RF followed by the EigenValue RF. V and EV are robust features but they sometimes fail in cases of organ shadow and imaging artifacts. To overcome this, we have chosen a feature based on EM (Effective Margin). EM is computed as follows. For a given region a histogram of the intensity values is computed. In case of organ shadows, the histogram is bimodal with two nearly equal peaks, for non vessel regions it is flat where as for vessels it is bimodal with two unequal peaks. This makes it a good rejector for organ shadows. Hence a third RF based on the EM feature is applied to the result of the first two stages.

3. RESULTS

As noted earlier, the RF classifier is applied to the putative seeds computed using a vesselness measure. The training data consists of 7000 examples from 28 angio datasets. Each forest is trained by building N randomized trees. Experimentally the best accuracy is obtained for $N = 100$. Each tree is trained by a set of 30% training samples chosen randomly from complete set. Around 20 features are selected randomly at each node for finding the classification criterion. Based on the feature usage i.e. each feature is picked how many times, we computed the variable importance map.[7] This variable importance map illustrates the features of high importance and those which can be removed from feature set without loss of accuracy. Around 80% of examples are chosen randomly from complete training set S which form set S_{tr}, and rest 20% is kept for testing which forms the test set S_t. A forest of 100 trees is built by learning from set S_{tr}. The training error is computed by testing the forest on S_{tr} samples and test set error is computed using out-of-bag testing on S_t samples. This procedure is repeated 100 times to get the average training and testing error. The advantage of doing stagewise application of RFs is shown in Figure **??**. As we can see significant reduction of false positives is carried out by the cascade of random forests. We also compared our approach to a competing strong classifier based on Adaboost. The error plots are shown in Figure 3. A sample result is shown in Figure 5. The result shows the final segmentation by growing the seeds generated by pruning the original set of seeds generated using the second order Hessians. Note that regions of organ shadows and other noisy regions are eliminated using the multi-stage RF approach.

4. CONCLUSION

In this paper we proposed a multi-stage random forest framework for improving blood vessel segmentation in angiography images. We show that our approach is able to handle multiple kinds of noises like surgical stitches and organ shadows systematically. The main advantage of employing a machine learning based framework is that thresholds can be automatically learned and optimized. This is in contrast to methods like [2] which choose heuristic values for thresholds based on anecdotal information. The other advantage is improved and consistent performance over variety of datasets containing different kinds of noise sources. This ability is attributed to

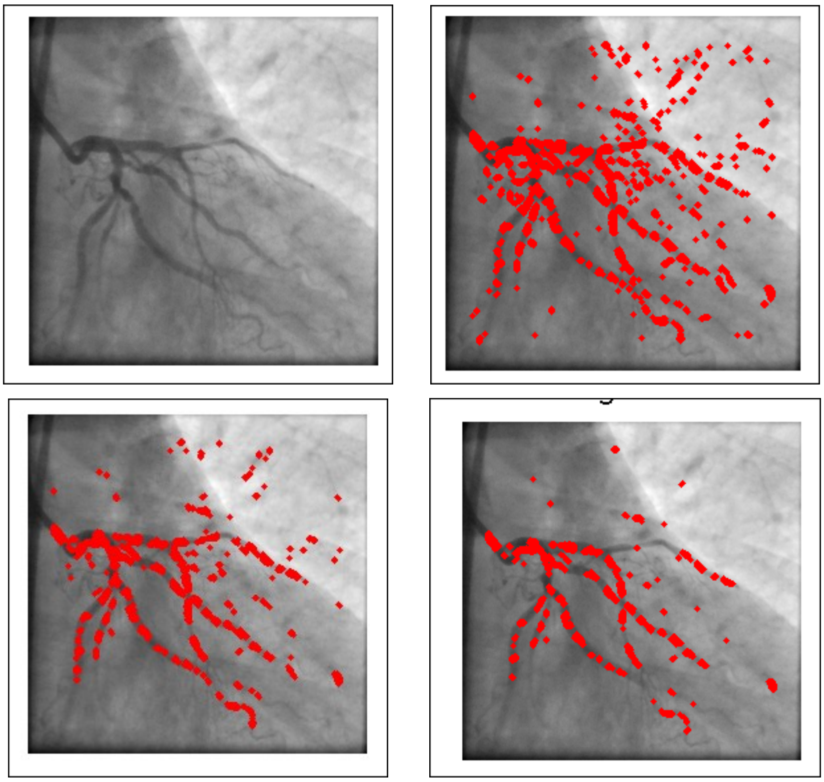

Figure 4. (a) Angio Image (b)Seeds Detected using Vesselness (c) Seeds obtained by applying the first stage RF based on vesselness (d) Seeds obtained after applying the Eigenvalue RF. Note that using context information, the first stage itself can eliminate some false positives. Further application of the Eigenvalue RF prunes most of the false positives.

extensive training given to build the classifiers, and classifiers being of ensemble type. The other advantage we found is the context learning ability of these classifiers, as features are fed for the small patch and not on single point itself as in state-of-the-art algorithm.

REFERENCES

[1] Sundar, H., Khamene, A., Xu, C., F.Sauer, and C.Davatzikos, "A novel 2d-3d registration algorithm for aligning fluoro images with 3d pre-op ct/mr imagesa," *SPIE Medical Imaging* (2006).

[2] M.Schneider and H.Sundar, "Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration," *Proceedings of ISBI* (2010).

[3] Koller, T. M., "Multiscale detection of curvilinear structures in 2-d and 3-d image data," *Computer Vision*, 864–869 (1995).

[4] Sato, Y., "3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images.," *Medical Image Analysis* **2**, 143–168 (1998).

[5] Frangi, "Multi-scale vessel enhancement filtering," *MICCAI* **1496**, 130–137 (1998).

[6] S.Chaudhuri and S.Chatterji, "Detection of blood vessels in retinal images using two-dimensional matched filters," *IEEE Trans. Medical Imaging* **8**, 263–269 (1989).

[7] L.Breiman, "Random forests," *Machine Learning*, Springer (2001).

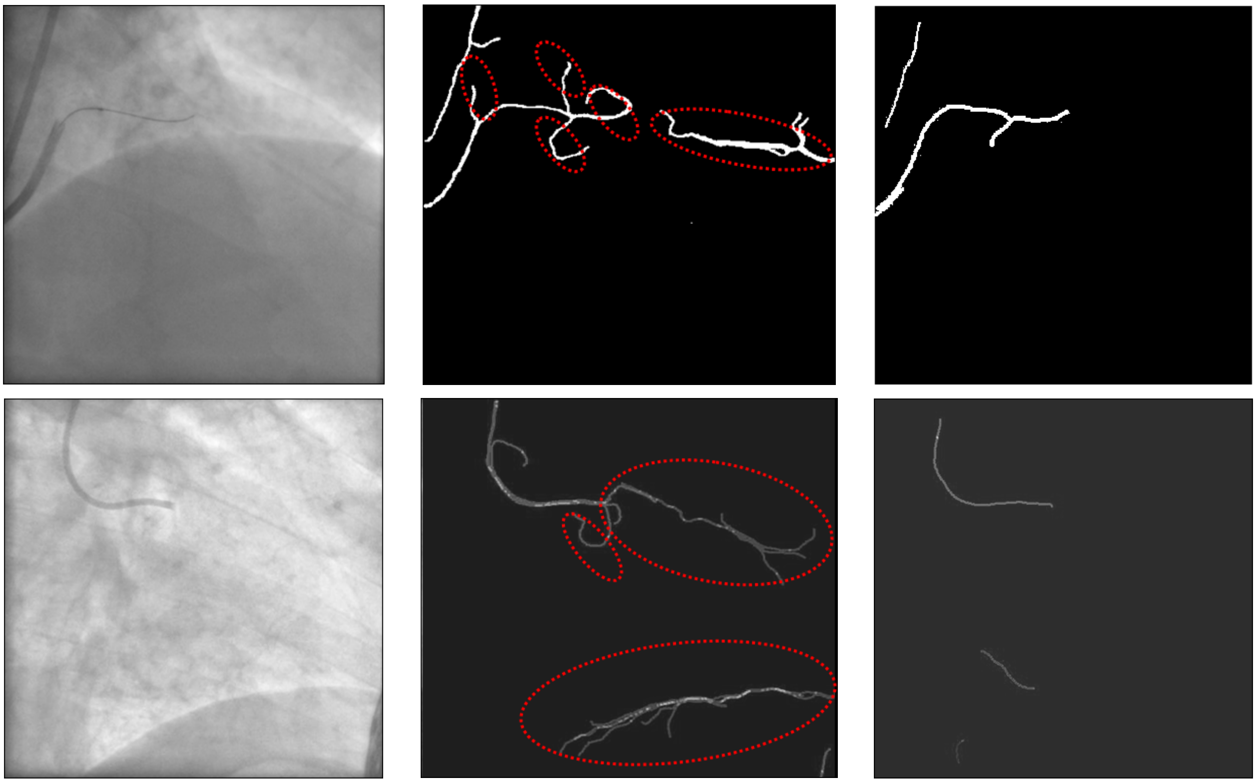

Figure 5. Vessel segmentation result using our proposed method. Left Image shows the input image. Center Image shows the result using the second order Hessian Subsystem. Right Image shows the result after growing the seeds selected using the multistage RF. Note that noisy regions and organ shadows are suppressed using our approach.