# Probabilistic Boosting Trees for Automatic Bone Removal from CT Angiography Images

Arne Militzer<sup>a,b</sup> and Fernando Vega-Higuera<sup>b</sup>

<sup>a</sup>Chair of Pattern Recognition, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
<sup>b</sup>Siemens Healthcare, Siemensstr. 1, 91301 Forchheim, Germany

## ABSTRACT

In CT angiography images, osseous structures occluding vessels pose difficulties for physicians during diagnosis. Simple thresholding techniques for removing bones fail due to overlapping CT values of vessels filled with contrast agent and osseous tissue, while manual delineation is slow and tedious. Thus, we propose to automatically segment bones using a trainable classifier to label image patches as bone or background. The image features provided to the classifier are based on grey value statistics and gradients. In contrast to most existing methods, osseous tissue segmentation in our algorithm works without any prior knowledge of the body region depicted in the image. This is achieved by using a probabilistic boosting tree, which is capable of automatically decomposing the input space. The whole system works by partitioning the image using a watershed transform, classifying image regions as bone or background and refining the result by means of a graph-based procedure. Additionally, an intuitive way of manually refining the segmentation result is incorporated. The system was evaluated on 15 CTA datasets acquired from various body regions, showing an average correct recognition of bone regions of 80% at a false positive rate of 0.025% of the background voxels.

Keywords: CT angiography, CTA, bone removal, segmentation, hierarchical classifier, boosting

# 1. INTRODUCTION

In clinical routine, CT angiography (CTA) is a standard procedure for the examination of vascular diseases. However, when visualized with Maximum Intensity Projection (MIP) or Volume Rendering (VR), target structures are often occluded by bones. In principle, there are two ways to overcome this deficit: Segment the target structures directly or segment and remove bones from the images to allow a clear view to the vessels. The approach followed here is the latter, as it appears to be the more general one and allows a visualization of vascular structures in their anatomical context.

For the task of removing osseous tissue from the image, simple thresholding techniques are not sufficient as grey values of vessels filled with contrast agent overlap with those of soft bone structures or bone marrow. A common approach to bone removal involves mimicking Digital Subtraction Angiography (DSA) by acquiring a native CT scan in addition to the CTA scan, in order to either subtract it from the CTA image after registration<sup>1,2</sup> or generate a mask from it using some thresholding or region growing techniques and apply this mask to the CTA image.<sup>3–5</sup> These algorithms often work satisfactorily for head examinations, where the relative position of bones and vessels usually does not vary during or in between the scans. However, the acquisition of the additional native CT scan is time consuming and means extra radiation for the patient.

As an alternative, approaches for bone removal based on standard CTA data have been proposed. These methods mostly include heuristic combinations of thresholding and region growing procedures, often along with morphological dilation,<sup>6,7</sup> sometimes with complex filter cascades.<sup>8</sup> Straka et al.<sup>9</sup> register the CTA image to an atlas built from hand segmented data to assign bone probabilities to image regions found with a watershed transform (WT). These algorithms have to be tuned for some training data or even specific body regions<sup>7,8</sup> and

While conducting this research, Arne Militzer was with the Institute of Neural Information Processing, Ulm University, Ulm, Germany.

Send author correspondence to arne.militzer@informatik.uni-erlangen.de.

Copyright 2009 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

http://dx.doi.org/10.1117/12.811886

thus do not generalize very well or, as in the case of the atlas based segmentation, have difficulties capturing details in the structures they try to segment.

The work most closely related to the presented approach is the hybrid watershed transform (HWT) by Hahn et al.<sup>10</sup> Their system comprises three steps: First, a variant of the watershed transform is applied to the image that during immersion collects information on each merging event between catchment basins, resulting in a merge hierarchy. The result of the second step is a sparse labeling of basins as bone or background provided by a support vector machine (SVM). To generate the final segmentation these labels are then propagated through the volume by traversing the merge hierarchy. This merging scheme also allows for manual interaction by setting bone or background markers in the image.

The general setup of the algorithm proposed here follows the intuitive three step structure by Hahn et al.<sup>10</sup> In fact, it utilizes their hierarchical 3-D watershed algorithm for partitioning the image in the first step. The focus, however, is different: While their segmentation mainly rests upon the WT and the merging scheme, here the classifier is the primary instrument to segment the image.

The rest of the paper is organized as follows: Section 2.1 describes first the set of features used for classification in the presented system and then the hierarchical classifier proposed. Section 2.2 contains a description of the postprocessing procedure currently used. In section 3 experimental results are reported, followed by our conclusions in section 4.

#### 2. METHODS

After applying an intensity threshold to remove air and soft tissue, the image is first partitioned using the hierarchical 3-D watershed algorithm by Hahn et al., effectively performing a compression of the data. Subsequent stages of the algorithm work with the basins found during the WT instead of single voxels, which greatly reduces their computational complexity.

Compared to a simple rectangular subdivision, the WT has the advantage of providing meaningful regions, where all voxels within one region ideally belong to the same tissue class, so that classifying a basin and transferring its label to all the contained voxels does not introduce a big error.

#### 2.1 Classification

In this step, each basin that was found during the WT is assigned a bone probability value by a trained classifier. Based on this value all voxels in the basin are later labeled as bone or background.

The classification problem at hand has a high intra-class variability, making it difficult for a single classifier like an SVM to globally find an appropriate model for each target class and thus an optimal decision boundary in the input space. Here, the use of a hierarchical classifier, such as the probabilistic boosting tree<sup>11</sup> (PBT), is proposed. Due to its hierarchical nature, it can capture the full variability of the classes enabling a single classifier to segment osseous tissue in images of any body region, while other algorithms have to be fine tuned for different kinds of examinations.

Additionally, the PBT performs feature selection during training, making it particularly suitable for highdimensional feature spaces and providing additional insight into the usefulness of available features.

#### 2.1.1 Features

The classifier's decision is based on a set of features that is calculated for each basin and comprises two parts. The first describes standard parameters of the distribution of grey values in the basin and consists of their maximum and minimum, mean, range and standard deviation. This is similar to what Hahn et al. proposed.<sup>10,12</sup> In some cases, however, this information is not sufficient to separate bone from background regions. A strongly contrasted vessel may have a grey value distribution very similar to that of a small bone or bone marrow. Morphologically however, they can still differ significantly. Therefore the approach presented here incorporates morphological information into the decision process by adding gradient features to the basins's feature vector. With I(x, y, z) denoting the image's intensity at voxel position (x, y, z) and  $\nabla I(x, y, z)$  the corresponding intensity gradient, for a basin B these consist of:

1.  $\sum_{(x,y,z)\in B} \frac{1}{\|B\|} | \bigtriangledown I(x,y,z)|$  the sum of the gradient magnitudes of all voxels in the basin, indicating inhomogeneities or strong edges

mogeneities or strong edges.

- 2.  $|\sum_{\substack{(x,y,z)\in B}} \nabla I(x,y,z)|$  the magnitude of the summation vector of all gradients in the basin, measuring the consistency in the gradients' orientation and
- 3. three discretizations of the direction of this summation vector, namely one approximating its direction in the *x-y*-plane, one approximating the angle with the *z*-axis and one giving a rough estimate of its direction in space. This last one is calculated by dividing the surface of a unit sphere into 18 sections. The index of the section the summation vector points to is then included into the feature vector.

We also performed experiments adding a small subset of the well established in object recognition Haar features,<sup>13,14</sup> centered around the basin to be classified. However, as these did not improve classification, they will not be considered any further in this discussion.

To incorporate context information, all mentioned features are computed not only for the basin B that is to be classified but also for two bigger regions including neighbouring basins, and then combined into B's feature vector.

#### 2.1.2 The Probabilistic Boosting Tree

The PBT is a hierarchical two-class classifier proposed by Tu.<sup>11</sup> Its general structure is depicted in fig.1. The PBT operates in a divide and conquer manner, resembling a soft decision tree. The learning procedure can be seen as a hierarchical extension to the boosting principle. It recursively builds a tree, training an AdaBoost<sup>15</sup> classifier for each node. The resulting AdaBoost classifier

$$H(\mathbf{x}) = \sum_{i=0}^{N} \alpha_i h_i(\mathbf{x}),\tag{1}$$

after training  $N \in \mathbb{N}$  weak learners  $h_i$  with weights  $\alpha_i \in \mathbb{R}$ , was shown to approach logistic regression<sup>16</sup> for posterior probabilities  $p(y|\mathbf{x}), y \in \{-1, 1\}$  by

$$H(\mathbf{x}) \approx \frac{1}{2} ln \frac{p(y=+1|\mathbf{x})}{p(y=-1|\mathbf{x})},\tag{2}$$

allowing the computation of approximate posterior probabilities for a sample  $\mathbf{x}$  as

$$q(\pm 1|\mathbf{x}) = \frac{exp(\pm 2H(\mathbf{x}))}{1 + exp(\pm 2H(\mathbf{x}))}.$$
(3)

According to these probabilities the training samples are reweighted and split at the newly trained node into a positive and a negative subset, putting ambiguous samples into both sets. A sample is considered ambiguous if its posterior probability q falls into the range  $\left[\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\right]$  for a user defined  $\epsilon$ . This parameter thus controls the degree of overfitting that the PBT develops during training. The positive and negative subsets are then used to train the right and left subtrees of the node, until a predefined maximum depth is reached.

Splitting the training set effectively divides the input space at each node, such that nodes deeper in the tree solve a specialized version of the original decision problem by considering only a small part of the input space, while higher nodes make rough decisions on the full input space.

This mechanism, combined with the feature selection performed by the AdaBoost training, also allows an assessment of the usefulness of single features for the decision making process. AdaBoost iteratively trains a set of weak learners, in each iteration incorporating the one with the best classification performance into the final classifier. If decision stumps are used as weak learners, this selection process means picking the single most discriminating feature in each iteration. Thus, features that are selected by the AdaBoost procedure in high tree



Fig. 1. Structure of the PBT. Each node holds an AdaBoost classifier with a number of decision stumps as base classifiers.

nodes, carry high level information, allowing a rough decision, while features that are picked in deeper nodes are useful for detailed decisions on more special cases.

Applying the PBT to classify a new pattern  $\mathbf{x}$  works in perfect analogy to the training. It is fed into the tree at the root node. The node's AdaBoost classifier calculates its posteriors  $q(\mathbf{x})$  and depending on the result the pattern is recursively passed into the subtrees. At each node the results from the subtrees are combined and returned to the calling node above. That way the PBT combines the classification results of its internal nodes into the overall approximate posterior distribution  $\tilde{p}(y|\mathbf{x})$ . This probability value can then be used to directly trade off the tree's sensitivity vs. its specificity via a single threshold.

# 2.2 Postprocessing

The result of the classification step is a single value for each basin denoting its probability of belonging to the bone class. To increase robustness and to retain a high specificity, all basins are divided into *bone*, *background* and *unknown* basins using two user adjustable thresholds, according to their probability values. Basins labeled as *unknown* receive their final class label from the graph-based postprocessing step.

In this step, a graph is constructed from the watershed information, where each basin yields one graph node. Edges correspond to neighbourhood relations between nodes' basins and are derived from the merge events that occurred during immersion. An edge's weight is calculated as the Euclidean distance between the two connected basins in the space spanned by their grey value statistical features. Eventually, for each node labeled as *unknown*, Dijkstra's algorithm for shortest paths<sup>17</sup> is run to determine the closest node in the graph that already has a final class label. This label is then adopted by the node and subsequently by its basin.

At this point, the user can manually refine the segmentation by clicking into the image to assign a class label to a voxel and its basin, which in turn triggers a re-evaluation of the graph-based procedure.

#### **3. RESULTS AND DISCUSSION**

To evaluate the proposed system 15 CTA volumes (8 head or neck, 5 torso, 2 leg-runoffs) were used. From these images, ground truth for the experiments was generated in two steps:

- 1. Osseous tissue was segmented with a commercially available semi-automatic bone-removal tool<sup>18</sup> to generate a bone mask.
- 2. The WT was run on the original datasets and all found basins were labeled as bone or background if their overlap with the bone mask generated before was  $\geq$  or < 80%, respectively. For the following procedures only these labeled basins were considered instead of the original bone mask.

As the sole purpose of this preliminary study was to examine the PBT's general bone classification abilities, the mask generation was not done by medical experts, so that some mislabelings typically occurred in the ground

truth. However, as this was the case in both the datasets that were used for training and for testing, the learning performance on these (imperfect) data could still be evaluated, though a slightly better classification result can be expected for properly labeled training sets.



Fig. 2. ROC curves showing the influence of the gradient features on the classification performance. The solid curve was generated training and testing the classifier with the full feature set, for the dashed one the gradient features were not included in the data.



Fig. 3. ROC curves illustrating the PBT's adaptation to data from specific body regions. To generate the dashed and dotted curves in (a), the PBT was only trained on datasets showing head or neck, in (b) these were generated training the PBT on body datasets only. In both cases the dashed curve shows the testing results for head and neck data, the dotted curves are the results from tests with body datasets. The solid curves relate these results to a classifier that was trained and tested using datasets from all body regions.

To allow an assessment of the PBT's performance, receiver operating characteristic (ROC) curves were generated by switching off the graph-based postprocessing step and varying a threshold on the classifier output. Each experiment was repeated 10 times, each time using 5 randomly chosen datasets for classifier training and the remaining ones for testing.

Figure 2 shows the influence of the proposed gradient features on the classification performance, plotting the classifier's sensitivity  $(\frac{correctly \ classified \ bone \ volume}{total \ bone \ volume})$  against its false positive rate 1-specificity (1 - specificity)

 $\frac{correctly\ classified\ background\ volume}{total\ background\ volume}$ ). Classification performance is visibly improved if the described gradient features are included into the feature set. In total, a recognition rate of 80% of the bone volume can be achieved with 0.025% of the non-bone volume falsely classified as bone.

The next set of experiments aimed at examining the classifier's ability to deal with the high variability in the classification domain at hand. Previous approaches to bone segmentation were designed only for specific body regions or at least needed several parameter sets for different kinds of examinations.<sup>7,8</sup> To see whether the PBT can overcome this deficit, it was trained and evaluated with data from different body regions separately. For the ROC curves in fig.3(a) it was trained using only head and neck datasets. The evaluation shows a clear adaptation to these images, while datasets from other body regions induce more errors. Figure 3(b) shows the same effect for classifiers trained only on body datasets. The best performance in both cases however, shows the classifier that was trained on mixed data from all body regions (solid reference curve). The hierarchical subdivision of the input space performed during the tree building procedure not only compensates for the bigger problem space, the full tree is actually better than two specialized ones. The learning procedure can use the available additional data to improve the decision making process. Thus, leaving the separation of the feature space entirely to the hierarchical learning procedure proves to be superior to a manual separation.



Fig. 4. Volume rendering of segmentation results of the proposed system for a leg runoff. (a) is the initial bone mask, prior to the postprocessing. (b) is the corresponding mask after double thresholding and the graph-based label propagation. In (c) this mask was applied to the input image to remove the bones, (d) shows the final result after setting 10 markers.

Some segmentation results of the presented system including the postprocessing steps are depicted in fig.4. Interval thresholding together with label propagation did not bring a quantitative advantage in comparison to standard thresholding, since for a good threshold setting about 95% of the unlabeled basins found their label within 2 node expansions (the remaining ones usually needed up to 10 expansions). Qualitatively on the other hand, interval thresholding does make a difference when examining only a specific image region: Figure 4(a) shows a segmented bone mask without the postprocessing, applying just a threshold to the classifier output. Figure 4(b) shows the same image with interval thresholding and the graph-based label propagation in place. Though some bone sections are missed by the algorithm, the mask is much better than before along the *femoral arteries*. Figure 4(c) is the corresponding bone removal result after applying this mask to the input image. The



Fig. 5. Typical bone removal result for a head dataset, displayed as MIP. The *vena innominata* was partly removed (arrows).

remaining errors can be removed with just a few manually set markers (fig.4(d)).

The bone removal result in fig.5 illustrates one of the system's remaining challenges. In this case the *vena innominata* was partly mislabeled as bone during postprocessing. Sometimes it would be recognized as bone by the classifier, because it is very bright due to its high concentration of contrast agent. Similar mislabelings occur for the sometimes highly contrasted *vena cava* and aortic stents. These cases simply cannot be captured by the features used for classification yet, as they show a grey value distribution and shape very similar to that of tubular bones.

#### 4. CONCLUSION

The PBT has proven itself useful for the problem of segmenting osseous tissue in CTA images. The hierarchical problem solving strategy does improve the classification compared to using several classifiers that were fine tuned for different body regions independently. Also, including the proposed gradient features into the decision helped reduce errors. An analysis of the used features shows however, that while gradient features are often picked in deeper tree nodes, the most important features still are the ones describing the grey value distribution, especially the maximum grey value of the basin. As the classifier based on grey value distribution features has reached its limit, further investigation of new, advanced features is necessary to further reduce classification errors. An additional point we would like to address in the future is to improve performance by more complex postprocessing. Finally, a sound evaluation of the segmentation performance of the entire system based on clinically validated ground truth data will be performed.

## REFERENCES

- Kwon, S. M., Kim, Y. S., Kim, T.-S., and Ra, J. B., "Digital subtraction CT angiography based on efficient 3D registration and refinement," *Computerized Medical Imaging and Graphics* 28(7), 391–400 (2004).
- [2] Jayakrishnan, V. K., White, P. M., Aitken, D., Crane, P., McMahon, A. D., and Teasdale, E. M., "Subtraction helical CT angiography of intra- and extracranial vessels: Technical considerations and preliminary experience," *American Journal of Neuroradioly* 24(3), 451–455 (2003).

- [3] Venema, H. W., Hulsmans, F. J. H., and den Heeten, G. J., "CT angiography of the circle of willis and intracranial internal carotid arteries: Maximum intensity projection with matched mask bone elimination:feasibility study," *Radiology* 218(3), 893–898 (2001).
- [4] van Straten, M., Venema, H. W., Streekstra, G. J., Majoie, C. B. L. M., den Heeten, G. J., and Grimbergen, C. A., "Removal of bone in CT angiography of the cervical arteries by piecewise matched mask bone elimination," *Medical Physics* **31**(10), 2924–2933 (2004).
- [5] Tomandl, B., Hammen, T., Klotz, E., Ditt, H., Stemper, B., and Lell, M., "Bone-subtraction CT angiography for the evaluation of intracranial aneurysms," *American Journal of Neuroradiology* 27(1), 55–59 (2006).
- [6] Alyassin, A. M. and Avinash, G. B., "Semiautomatic bone removal technique from CT angiography data," in [Medical Imaging 2001: Image Processing], Sonka, M. and Hanson, K. M., eds., Proc. SPIE 4322, 1273–1283 (2001).
- [7] Fiebich, M., Straus, C. M., Sehgal, V., Renger, B. C., Doi, K., and Hoffmann, K. R., "Automatic bone segmentation technique for CT angiographic studies," *Journal of Computer Assisted Tomography* 23(1), 155–161 (1999).
- [8] Lee, Y., Tsai, D.-Y., Inomata, S., Yokoyama, R., Hara, T., Fujita, H., Kanematsu, M., Iwama, T., and Hoshi, H., "An automated bone removal technique for extraction of cerebral vessels from head CT angiography," in [Biomedical Engineering 2005], Proc. IASTED Internat. Conf.: Biomedical Engineering, 22–25 (2005).
- [9] Straka, M., Cruz, A. L., Köchl, A., Šrámek, M., Gröller, E., and Fleischmann, D., "3D watershed transform combined with a probabilistic atlas for medical image segmentation," *Journal of Medical Informatics & Technologies* 6 (2003).
- [10] Hahn, H. K., Wenzel, M. T., Drexl, J., Zentis, S., and Peitgen, H.-O., "HWT hybrid watershed transform: optimal combination of hierarchical interactive and automated image segmentation," in [Medical Imaging 2007: Image Processing], Pluim, J. P. W. and Reinhardt, J. M., eds., Proc. SPIE 6512 (2007).
- [11] Tu, Z., "Probabilistic Boosting-Tree: Learning discriminative models for classification, recognition, and clustering," in [International Conference on Computer Vision 2005], Proc. ICCV 2, 1589–1596 (2005).
- [12] Hahn, H. K., Wenzel, M. T., Konrad-Verse, O., and Peitgen, H.-O., "A minimally-interactive watershed algorithm designed for efficient CTA bone removal," in [*Computer Vision Approaches to Medical Image Analysis*], Beichel, R. and Sonka, M., eds., *LNCS* 4241, 178–189, Springer Verlag (2006).
- [13] Tu, Z., Zhou, X., Bogoni, L., Barbu, A., and Comaniciu, D., "Probabilistic 3D polyp detection in CT images: The role of sample alignment," in [*IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006*], Proc. CVPR 2, 1544–1551 (2006).
- [14] Viola, P. and Jones, M., "Rapid object detection using a boosted cascade of simple features," in [IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2001], Proc. CVPR 1 (2001).
- [15] Freund, Y. and Schapire, R. E., "A decision-theoretic generalization of on-line learning and an application to boosting," in [*Proc. of the Second European Conference on Computational Learning Theory*], Vitányi, P. M. B., ed., *Proc. EuroCOLT* (1995).
- [16] Friedman, J., Hastie, T., and Tibshirani, R., "Additive logistic regression: a statistical view of boosting," Annals of Statistics 28(2), 337–407 (2000). Special Invited Paper.
- [17] Dijkstra, E. W., "A note on two problems in connexion with graphs," Numerische Mathematik 1(1), 269–271 (1959).
- [18] Siemens AG, Healthcare Sector (2007).