

Hierarchical Detection of Multiple Organs Using Boosted Features

Samuel Hugueny and Mikaël Rousson

Department of Imaging and Visualization
Siemens Corporate Research, Princeton, NJ, USA

Abstract. We propose a framework for fast and automated initialization of segmentation algorithms in Computed Tomography images. Based on the idea that time-consuming voxel classification should be done only on spatially constrained areas, we build classifiers at body and slice levels which quickly define a constrained region of interest. Voxel classification is then performed by a divide-and-conquer strategy using a probabilistic-boosting tree. In addition, this framework can incorporate additional information on the volume, if available, such as the position of another organ to improve its accuracy and robustness. The framework is applied to seed extraction in kidneys and liver.

1 Introduction

Statistical model-based segmentation algorithms such as active contours [1], region competition [2] or graph cuts [3] are considered powerful and effective, and are widely used in medical image segmentation [4,5]. They share a common particularity, namely, they need to be provided with one or multiple seeds, or an initial contour. This initialization step is often performed manually, hence preventing them to be fully automated.

A possible approach to automate the process is machine learning. Being able to combine abstractly various sources of information, machine learning techniques (bagging trees [6], neural networks [7], support vector machines [8], boosting trees [9], *etc.*) have a major role in computer assisted diagnosis (CAD). For the purpose of seed extraction from Computed Tomography (CT) 3D scans, direct application of learning algorithms either at image or voxel level is bound to fail. At image level, one would need a large training set of CT scans (hundreds), which is not easily available. At voxel level, inter-organ similarities and intra-organ variability prevent learning to be fast and robust. Moreover, the amount of voxels pin a 3D scan makes a direct approach computationally expensive, both in the training stage and in the detection phase. Even with divide-and-conquer strategies on the voxel population, detectors are hard to train and their robustness is questionable.

However, the organization of the body structures is well defined and remains mostly consistent from one patient to another. Modelization of this high-level knowledge of dependencies between organs and tissues can help reduce the complexity of the seed extraction problem. Large parts of the body/scan can be discarded, hence leading to a faster and better decision boundary. Previous works

in this direction includes use of deformable models [10], or computation of probability maps based on a previous segmentation of the lungs [11].

In this paper, we present a framework for fully automated multi-organ detection in abdominal CT images, based on the Adaboost algorithm. We propose to progressively incorporate spatial knowledge in the learning process, first by classifying horizontal slices, then by discarding slice areas based on prediction of the organ of interest's position, and finally by performing a voxel classification. This leads to an important reduction of both training and detection time as well as an improved accuracy. At each step, features are defined and normalized when possible, so that they are independent of the actual intensity dynamic and patient characteristics. Detectors are therefore robust to variations due to unhealthy organs or contrast agents. We also define filters which takes into account predefined landmarks, if available, such as results from a previous run of the algorithm on a different organ.

Section 2 details the framework learning procedures, section 3 describes the features used at each step, section 4 shows how detection can be improved with the prior knowledge of landmarks, while section 5 presents some of our results.

2 Learning

2.1 Adaboost

Adaboost, invented by Freund and Schapire [12] and its variants have been successfully applied to many problems in vision and machine learning. Friedman and al. [13] have shown that Adaboost approach the posterior $p(y|x)$ by selecting and combining a set of weak classifiers into a strong classifier. At each iteration, it increases the weights of the previously misclassified samples, so that the next selected weak classifier will perform better on "hard" examples. The final hypothesis is a weighted linear combination of the T hypotheses whose weights are inversely proportional to their training errors: $H(x) = \sum_{t=1}^T \alpha_t h_t(x)$ with $h_t(x)$ being a weak classifier and α_t its weight. A classifier trained with the standard procedure is used for slice classification. In the detection phase, only voxels which belong to slices detected as positives are passed to the voxel classifier. However, using standard Adaboost at voxel level is questionable for different reasons:

- although Adaboost asymptotically converges to the target distribution, it needs to pick hundreds of weak classifiers, thus being computationally expensive. Since a 3D CT image contains millions of pixels with great diversity, picking a single sampled training set would make training a matter of weeks, and detection a matter of minutes.
- the order in which features are selected is not preserved although it may correspond to high level semantics and be useful for the understanding of patterns.

- the re-weighting scheme of Adaboost may cause samples previously correctly classified to be misclassified again. In the case of a rare event detection problem, we want to be able to discard a large amount of samples in the early stages, so that later stages deal with samples that are similar to the targeted data.

These issues are naturally addressed by a divide-and-conquer (e.g. tree or cascade) approach.

2.2 Probabilistic Boosting Tree (PBT)

Voxel classification is performed using a *Two-class Probabilistic Boosting Tree* (PBT) as presented in [14]. The PBT procedure automatically constructs a tree in which each node is a standard Adaboost classifier (a combination of weak classifiers) that can deal with complex distributions while being resistant to overfitting. Training samples are then divided into two new sets according to their response to the learnt classifier and used to train a left sub-tree and a right sub-tree. Confusing data are passed further down, leading to the expansion of the tree. PBT naturally embeds clustering and has applications in classification, detection, recognition and segmentation [15]. Table 1 details the learning process.

The detection phase is consistent with the learning phase. At each node, a sample is passed down to the right subtree if its response to the node's strong classifier is positive (meaning that the voxel belongs to the organ of interest), to the left subtree if not, until it reaches a leaf. The probability returned is the ratio of positives samples that reached this leaf in the training process.

3 Defining Features for Seed Detection

Since seed detection is a voxel classification problem, training has ultimately to be made at voxel level. A voxel-based training faces the following problems: huge amount of voxels, intensity similarities between organs and tissues, variations between acquisitions, lack of organ tissue homogeneity within and among different image slices both in shape and texture, abnormalities of unhealthy tissues, absence of the organ of interest from the considered scan.

Most of these problems can be addressed by an *ad hoc* reduction of the number of potential seeds. Since very general anatomical considerations (on one of the body, beneath the diaphragm, ...) help predicting an organ position, we propose to constrain a spatial domain of interest using boosted classifiers at body and slice level. To be independent of scale factors, a normalized coordinate system of the patient's body is defined for the two axial dimensions. Because CT scans do not always represent the same part of a patient's body in the vertical dimension, we cannot have the same normalization in z. The axial dimensions on one

Table 1. PBT procedure used to learn the voxel classifier. The only difference with the procedure presented in [14] is the sampling stage (2) introduced to speed up trainings.

<p>Procedure for training of a tree with maximum depth of L:</p> <ol style="list-style-type: none"> 1. Given: a set of images $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$. X_k is a region of interest of a particular image. Y_k is the corresponding ground truth. Training set is $S = \{(x_1, y_1), \dots, (x_p, y_p)\}$, where $x_i \in X_k$ and $y_i = 0, 1$ for negative and positive respectively. 2. Sample a training set $S_{sampled}$ of labeled examples $\{(x'_1, y'_1), \dots, (x'_n, y'_n)\}$ with respect to the number of positives and negatives in S. w_i is the weight of the training sample. 3. Initialize weights $w_i = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively. 4. Normalize weights in $S_{sampled}$. 5. In $S_{sampled}$, train a strong classifier using the standard Adaboost procedure. Exit early if the training error at classifier $\epsilon_t > \theta$ (e.g. $\theta = 0.45$). 6. If the current tree depth is L, then exits. 7. Initialize two empty sets S_{left} and S_{right}. 8. For each sample in S compute the probability $q(1 (x_k, i_k))$ and $q(-1 (x_k, i_k))$ using the learned strong classifier learned on $S_{sampled}$. 9. if $q(1 x_i) > \frac{1}{2}$ then $(x_i, y_i, 1) \rightarrow S_{right}$ else $(x_i, y_i, 1) \rightarrow S_{left}$ 10. Repeat the procedure recursively for S_{left} and S_{right}.

hand and the vertical dimension on the other hand are treated differently. Our approach consists of three steps:

1. Slice classification: a classifier that gives a probability that a particular slice intersects with the organ of interest is trained. A permissive decision threshold is set to obtain an almost perfect recall.
2. Slice area selection: given an easily computed segmentation of the body, a scaled coordinate system is defined for each slice, so that regions are defined univocally in every slice. By looking up the organ coordinates in the training images, we identify regions where the organ of interest has very little chance to be observed. This process discards roughly 70 % of a slice pixel for vertically oriented organs such as kidneys.
3. Voxel classification: a voxel classifier is trained, on a large neighborhood of the organ.

Center $O_{p,k}$ and dimensions $w_{p,k}, h_{p,k}$ are determined for each slice $S_{p,k}$ in volume V_p using a slice-by-slice segmentation of the body (obtained by a simple contour-tracing algorithm). These are used to define on each slice a coordinate system, with the center of the slice as its origin and the X and Y axes as its basis axes. Coordinates are normalized by the slice dimensions, so that every voxel in the slice has coordinates between -1 and 1 , regardless of the patient's size or corpulence (Fig. 1).

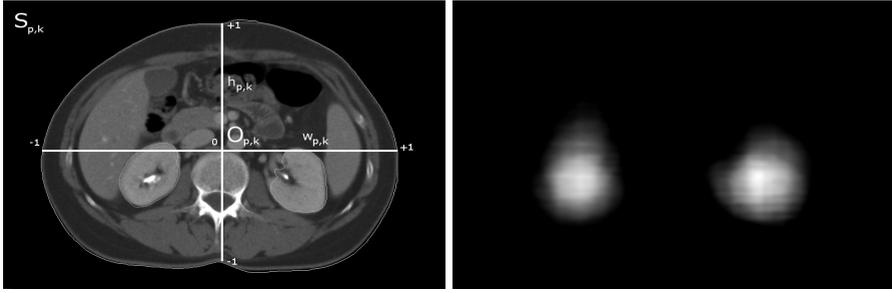


Fig. 1. Left: CT slice with coordinate system and reference segmentation used for kidneys. A similar coordinate system is defined for every slice. Right: observation map for both kidneys obtained using 20 volumes. Bottom left corner is $(-1,-1)$, upper right is $(1,1)$ in the slice coordinate system.

3.1 Slice Classification

Vertical normalization would require to have the patient’s size or two easily detected anatomical landmarks in the CT scan, which cannot be guaranteed. We propose to train an Adaboost classifier that determines the probability for a slice to intersect with the organ of interest. Features (Table 2) are designed to identify spatial organization, similarities in shape, appearance, symmetry, relative intensities compared to those of the entire body, comparisons with neighbours, entropy, *etc.* They are computed both for the entire slice and for sub-windows. Since they are slice-level features, they may be more computationally expensive than voxel features. Features are fed to a two-class Adaboost procedure, together with a training set of slices. Slices are labeled positive if they intersect with the organ of interest, negative if they do not. After regularization based on the fact that organs are connex, the obtained detectors discard a vast majority of slices, while having a nearly perfect recall.

3.2 Slice Area Selection

Using this coordinate system we report the 2D normalized coordinates of every voxel that belongs to the organ of interest in each training image on an “occurrence map”, from which the consistency of the organ’s position in the slice is observable. Based on this map we define a domain where voxels are likely to be positive, according to the number of positive observations at their locations. A permissive bounding box is defined on each slice using the observation map. For kidneys, this step discards at once $\frac{4}{5}$ of voxels. Of course, the smaller the horizontal sections of the organ are, the more voxels are discarded.

3.3 Features Used for Voxel Classification

We use the three dimensional version of the Haar filters combined with the use of an integral image [16]. The computational cost of computing Haar filters is

Table 2. Examples of features used for slice classification. They are designed as cues of symmetry (h_3), shape (h_1, h_5), appearance (h_2, h_6). Features are computed for the entire slice and for rectangular subwindows.

–	$h_1(S_{p,k}) = \frac{\text{card}\{x/x \text{ in } S_{p,k}\}}{4w_{p,k}h_{p,k}}$
–	$h_2(S_{p,k}) = \text{percentage of air voxels in } S_{p,k}$
–	$h_3(S_{p,k}) = \frac{\text{card}\{x \in S_{p,k} / \vec{x} \cdot \vec{i}_{p,k} > 0\}}{\text{card}\{x \in S_{p,k} / \vec{x} \cdot \vec{i}_{p,k} < 0\}}$
–	$h_4(S_{p,k}) = \vec{i}_{p,k} \cdot \vec{i}_{p,k-1}$
–	$h_5(S_{p,k}) = \frac{w_{p,k+1}}{w_{p,k-1}}$
–	$h_6(S_{p,k}) = \text{Entropy}(S_{p,k})$
$S_{p,k}$ is a slice, $C_{p,k} = (O_{p,k}, \vec{i}_{p,k}, \vec{j}_{p,k})$ the slice coordinate system.	

constant since we only need to sum up the values of corners of the Haar filter in the integral volume. Training is made on voxels contained in the bounding boxes defined by vertical pruning, using a variant of Probabilistic Boosting Tree.

4 Hierarchical Detection

Furthermore, a better accuracy can be achieved by using a segmentation (ground-truth) of a different organ in the training process for the organ of interest, or any anatomical landmark. This information (position, shape) can be incorporated at each of the three previously presented stages:

- Slice area selection: by placing the center of the coordinate system at the center of the ground-truth, we can define another “occurrence map” by reporting the coordinates of every voxel that belongs to the organ of interest in each scan of the training set.
- Slice characterization: features using the size and the position of the reference structure are introduced in the learning process.
- Voxel classification: purely geometrical features using the position of the center of the reference structure are introduced in the learning process

Dependencies of relative positions, sizes, shapes and histograms between organs are, if present, automatically selected by the learning process, assuming features are relevant enough. Being specific to the patient and the acquisition, this information can serve as a reference for the rest of the image. An incremental detection process which relies on the segmentation of previously segmented organs to perform the seed detection of the next one can be defined.

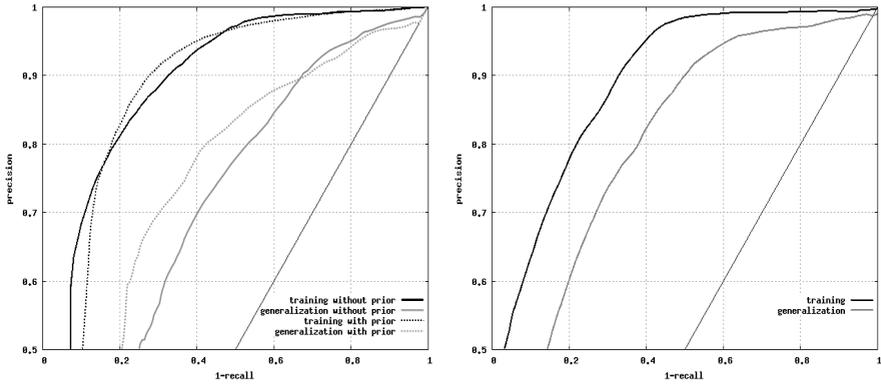


Fig. 2. Training and validation curves for the left kidney, with and without prior knowledge of the right kidney position (left) and liver (right). Straight line is the first bisector.

5 Results and Discussion

Our database is composed of 50 CT volumes from various patients and institutions. 20 of them, randomly chosen, are used for training, the 30 remaining for testing. Ground-truths are obtained by a supervised level set algorithm. Detectors are trained for left kidney, with and without prior knowledge of the right kidney’s barycenter position. Figure 2 shows results curves for each detector. Since our task is detection rather than segmentation, we do not expect an excellent recall but a good precision. Precision at level of confidence α is defined by:

$$p(\alpha) = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} = \frac{\text{card}\{(x, 1)/H(x) > \alpha\}}{\text{card}\{(x, y)/H(x) > \alpha\}}$$

where $0 \leq \alpha \leq 1$ is a threshold and $0 \leq H(x) \leq 1$ the response of the H classifier for the x sample.

A retrieval rate of 30% of the organ voxels is possible with a precision higher than 90 % for both liver and kidneys, which proved sufficient to initialize a level set algorithm. In the case of a kidney, a random guess in the entire image would have a precision and a recall lower than 1 %. As expected, knowledge of another organ position improves the average accuracy. Moreover, since positively detected voxels are spatially constrained, geometrical post-processing should in the future lead to an increased precision. Detection time on a standard computer is around 400 000 voxels per second. Robustness is acceptable given the size of our training set and better results could be achieved with larger datasets.

In this paper, a framework for automated seed extraction is introduced. Learning is carried out in a hierarchical way by divide-and-conquer strategies.

Experiments are reported on liver and kidney. Results show that it is a valid and robust approach to automatically extract seeds in multiple organs. Furthermore, observed recall rates and intrinsic spatial concentration of the retrieved seeds suggest that feature improvement and larger databases could lead to application in segmentation. Future work also includes extending the method to more organs and simultaneous seed extraction in several organs.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
2. Zhu, S., Yuille, A.: Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(9), 884–900 (1996)
3. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 1222–1239 (2001)
4. Chakraborty, A., Staib, L., Duncan, J.: Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Transactions on Medical Imaging* 15(6), 859–870 (1996)
5. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In: *ICCV*, pp. 105–112 (2001)
6. Hothorn, T., Lausen, B.: Bagging tree classifiers for laser scanning images: a data- and simulation-based strategy. *Artificial Intelligence in Medicine* 27(1), 65–79 (2003)
7. Sharkey, A., Sharkey, N., Cross, S.: Adapting an ensemble approach for the diagnosis of breast cancer. In: *Proceedings of the 6th International Conference on Artificial Neural Networks*, pp. 281–286 (1998)
8. Chang, R.F., Wu, W.J., Moon, W.K., Chou, Y.H., Chen, D.R.: Support vector machines for diagnosis of breast tumors on US images. *Academic radiology*, 189–197 (2003)
9. Tu, Z.: Probabilistic 3D polyp detection in CT images: The role of sample alignment pp. 1544–1551 (2006)
10. Karssemeijer, N., van Erning, L.J.T.O., Eijkman, E.G.J.: Recognition of organs in CT-image sequences: a model guided approach. *Computers and Biomedical Research* 21(5), 434–448 (1988)
11. Zhou, X.: Constructing a probabilistic model for automated liver region segmentation using non-contrast x-ray torso ct images. In: Larsen, R., Nielsen, M., Sporning, J. (eds.) *MICCAI 2006*. LNCS, vol. 4190, pp. 856–863. Springer, Heidelberg (2006)
12. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: *European Conference on Computational Learning Theory*, pp. 23–37 (1995)
13. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. In: *Dept. of Statistics, Stanford Univ. Technical Report* (1998)

14. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition and clustering. In: 10th IEEE International Conference on Computer Vision (2005)
15. Zheng, S., Tu, Z., Yuille, A., Reiss, A., Dutton, R., Lee, A., Galaburda, A., Dinov, I., Thompson, P., Toga, A.: A learning-based algorithm for automated extraction of the cortical sulci (2006)
16. Viola, P., Jones, M.: Face recognition using boosted local features (2003)