Regression Forests for Efficient Anatomy Detection and Localization in CT Studies

Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu

Microsoft Research Ltd, CB3 0FB, Cambridge, UK

Abstract. This paper proposes multi-class random regression forests as an algorithm for the efficient, automatic detection and localization of anatomical structures within three-dimensional CT scans.

Regression forests are similar to the more popular classification forests, but trained to predict *continuous* outputs. We introduce a new, continuous parametrization of the anatomy localization task which is effectively addressed by regression forests. This is shown to be a more natural approach than classification.

A single pass of our probabilistic algorithm enables the direct mapping from voxels to organ location and size; with training focusing on maximizing the confidence of output predictions. As a by-product, our method produces *salient anatomical landmarks*; *i.e.* automatically selected "anchor" regions which help localize organs of interest with high confidence. Quantitative validation is performed on a database of 100 highly variable CT scans. Localization errors are shown to be lower (and more stable) than those from global affine registration approaches. The regressor's parallelism and the simplicity of its context-rich visual features yield typical runtimes of only $1s$. Applications include semantic visual navigation, image tagging for retrieval, and initializing organ-specific processing.

1 Introduction

This paper introduces the use of regression forests in the medical imaging domain and proposes a new, parallel algorithm for the efficient detection and localization of anatomical structures ('organs') in computed tomography (CT) studies.

The main contribution is a new parametrization of the anatomy localization task as a multi-variate, continuous parameter estimation problem. This is addressed effectively via tree-based, non-linear regression. Unlike the popular *classification* forests (often referred to simply as "random forests"), regression forests [1] have not yet been used in medical image analysis. Our approach is fully probabilistic and, unlike previous techniques (*e.g.* [2,3]) maximizes the confidence of output predictions. The focus of this paper is both on accuracy of prediction and speed of execution, as we wish to achieve anatomy localization in seconds. Automatic anatomy localization is useful for efficient visual navigation, initializing further organ-specific processing (*e.g.* detecting liver tumors), and semantic tagging of patient scans to aid their sorting and retrieval.

B. Menze et al. (Eds.): MICCAI 2010 Workshop MCV, LNCS 6533, pp. 106–117, 2011.

Regression-based approaches. Regression algorithms [4] estimate functions which map input variables to *continuous* outputs[1]. The regression paradigm fits the anatomy localization task well. In fact, its goal is to learn the non-linear mapping from voxels *directly* to organ position and size. [5] presents a thorough overview of regression techniques and demonstrates the superiority of boosted regression [6] with respect to *e.g.* kernel regression [7]. In contrast to the boosted regression approach in [2] maximizing confidence of output prediction is integral to our approach. A comparison between boosting, forests and cascades is found in [8]. To our knowledge only two papers have used regression forests [9,10]; neither with application to medical image analysis, nor to multi-class problems. For instance, [10] addresses the problem of detecting pedestrians v background.

Classification-based approaches. In [11] organ detection is achieved via a confidence maximizing sequential scheduling of multiple, organ-specific *classifiers*. Our single, tree-based regressor allows us to deal naturally with multiple anatomical structures simultaneously. As shown in the machine learning literature [12] this encourages feature sharing and, in turn better generalization. In [13] a sequence of PBT classifiers (first for salient slices, then for landmarks) are used. In contrast, our single regressor maps directly from voxels to organ poses. Latent, salient landmark regions are extracted as a by-product of our procedure. In [14] the authors achieve localization of organ *centres* but fail to estimate the organ extent (similarly for [10]). Here we present a more direct, continuous model which estimates the position of the walls of the bounding box containing each organ; thus achieving simultaneous organ localization and extent estimation.

Registration-based approaches. Although atlas-based methods have enjoyed much popularity [3,15,16] their conceptual simplicity is in contrast to the need for robust, cross-patient registration. Robustness is improved by multi-atlas techniques [17], at the price of slower algorithms involving multiple registrations. Our algorithm incorporates atlas information within a compact tree-based model. As shown in the result section, such model is more efficient than keeping around multiple atlases and achieves anatomy localization in only a few seconds. Comparisons with affine registration methods (somewhat similar to ours in computational cost) show that our algorithm produces lower and more stable errors.

1.1 Background on Regression Trees

Regression trees [18] are an efficient way of mapping a complex input space to continuous output parameters. Highly non-linear mappings are handled by splitting the original problem into a set of smaller problems which can be addressed with simple predictors. Figure 1 shows an illustrative 1D example where the goal is to learn an analytical function to predict the real-valued output y (*e.g.* house prices) given the input x (*e.g.* air pollution). Learning is supervised as we are given a set of training pairs (x, y). Each node in the tree is designed to split the data so as to form clusters where accurate prediction can be performed with

[1] As opposed to *classification* where the predicted variables are discrete.

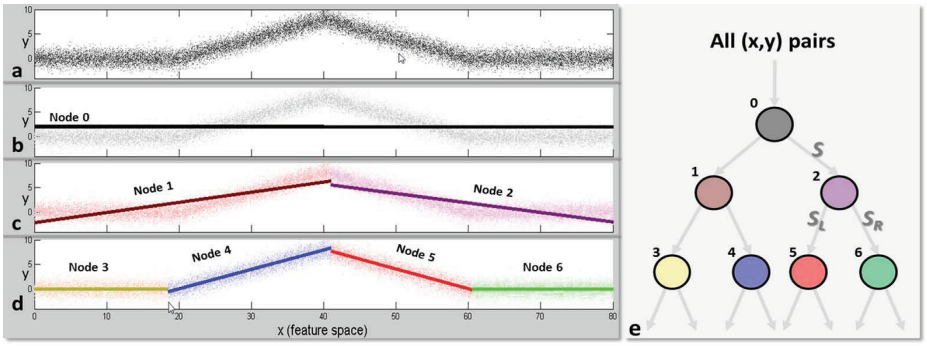

Fig. 1. Regression tree: an explanatory 1D example. **(a)** Input data points. **(b)** A single linear function fits the data badly. **(c,d)** Using more tree levels yields more accurate fit of the regressed model. Complex non-linear mappings are modelled via a hierarchical combination of many, simple linear regressors. **(e)** The regression tree.

simpler models (*e.g.* linear in this example). More formally, each node performs the test $\xi > f(x) > \tau$, with ξ, τ scalars. Based on the result each data point is sent to the left or right child.

During training, each node test (*e.g.* its parameters ξ, τ) is optimized so as to obtain the best split; *i.e.* the split that produces the maximum reduction in geometric error. The error reduction r is defined here as: $r = e(\mathcal{S}) - \sum_{\mathtt{i} \in \mathtt{L},\mathtt{R}} \omega_i e(\mathcal{S}_i)$ where $\mathcal{S}$ indicates the set of points reaching a node, and `L` and `R` denote the left and right children (for binary trees). $\omega_i = |\mathcal{S}_i|/|\mathcal{S}|$ is the ratio of the number of points reaching the i^{th} child. For a set $\mathcal{S}$ of points the error of geometric fit is: $e(\mathcal{S}) = \sum_{j \in \mathcal{S}} [y_j - y(x_j; \boldsymbol{\eta}_S)]^2$, with $\boldsymbol{\eta}_S$ the two line parameters computed from all points in $\mathcal{S}$ (*e.g.* via least squares or RANSAC). Each leaf stores the continuous parameters $\boldsymbol{\eta}_S$ characterizing each linear regressor. More tree levels yield smaller clusters and smaller fit errors, but at the risk of overfitting.

2 Multivariate Regression Forests for Organ Localization

This section presents our mathematical parametrization and the details of our multi-organ regression forest with application to anatomy localization.

Mathematical notation. Vectors are represented in boldface (*e.g.* $\mathbf{v}$), matrices as teletype capitals (*e.g.* $\mathtt{A}$) and sets in calligraphic style (*e.g.* $\mathcal{S}$). The position of a voxel in a CT volume is denoted $\mathbf{v} = (v_x, v_y, v_z)$.

The labelled database. The anatomical structures we wish to recognize are $\mathcal{C} = \{$ `heart, liver, spleen, left lung, right lung, l. kidney, r. kidney, gall bladder, l. pelvis, r. pelvis`$\}$. We are given a database of 100 scans which have been manually annotated with 3D bounding boxes tightly drawn around the structures of interest (see fig. 3a). The bounding box for the organ

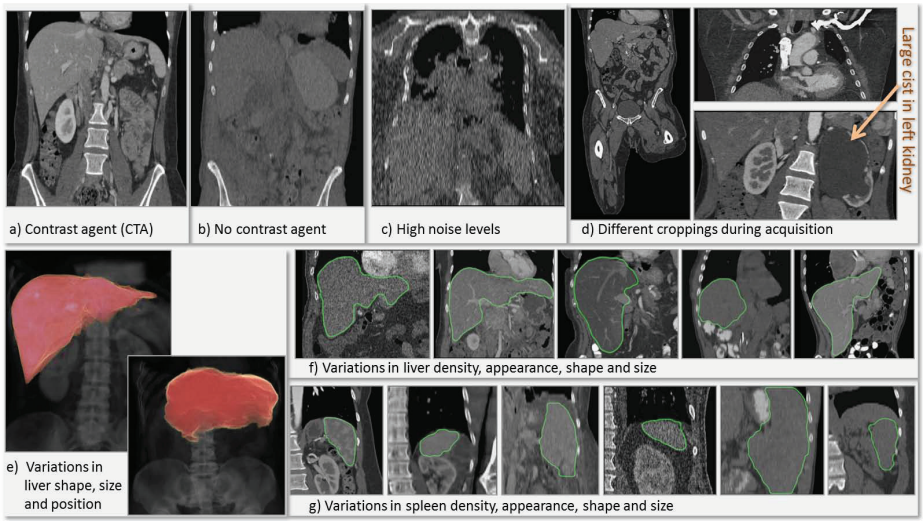

Fig. 2. Variability in our labelled database. (a,b,c) Variability in appearance due to presence of contrast agent, or noise. **(d)** Difference in image geometry due to acquisition parameters and possible anomalies. **(e)** Volumetric renderings of liver and spine to illustrate large changes in their relative position and in the liver shape. **(f,g)** Mid-coronal views of liver and spleen across different scans in our database to illustrate their variability. All views are metrically and photometrically calibrated.

$c \in \mathcal{C}$ is parametrized as a 6-vector $\mathbf{b}_c = (b_c^{\mathtt{L}}, b_c^{\mathtt{R}}, b_c^{\mathtt{A}}, b_c^{\mathtt{P}}, b_c^{\mathtt{H}}, b_c^{\mathtt{F}})$ where each component represents the position (in mm) of the corresponding axis-aligned wall[2]. The database comprises patients with different conditions and large differences in body size, pose, image cropping, resolution, scanner type, and possible use of contrast agents (fig. 2). Voxel sizes are $\sim 0.5 - 1.0mm$ along x and y, and $\sim 1.0 - 5.0mm$ along z. The images have not been pre-registered or normalized in any way. The goal is to localize anatomies of interest accurately and automatically, despite such large variability. Next we describe how this is achieved.

2.1 Problem Parametrization and Regression Forest Learning

Key to our algorithm is the fact that *all* voxels in a test CT volume contribute with varying confidence to estimating the position of the six walls of *all* structures' bounding boxes (see fig. 3b,c). Intuitively, some distinct voxel clusters (*e.g.* ribs or vertebrae) may predict the position of an organ (*e.g.* the heart) with high confidence. Thus, during testing those clusters will be used as reference (landmarks) for the localization of those anatomical structures. Our aim is to learn to cluster voxels together based on their appearance, their spatial context and, above all, their confidence in predicting position and size of all

[2] Superscripts follow standard radiological orientation convention: $\mathtt{L}$ = left, $\mathtt{R}$ = right, $\mathtt{A}$ = anterior, $\mathtt{P}$ = posterior, $\mathtt{H}$ = head, $\mathtt{F}$ = foot.

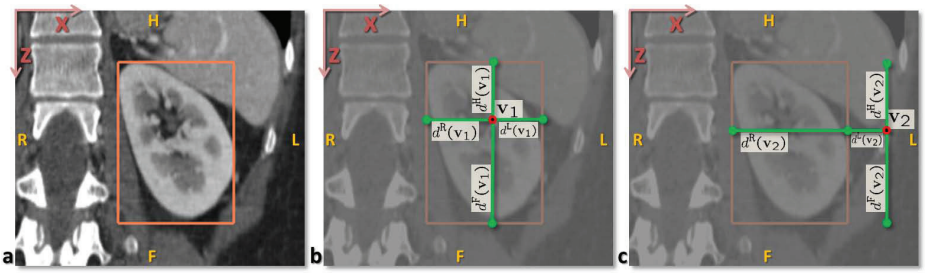

Fig. 3. Problem parametrization. (a) A coronal view of a left kidney and the associated ground-truth bounding box (in orange). **(b,c)** *Every* voxel $\mathbf{v}_i$ in the volume votes for the position of the six walls of each organ's 3D bounding box via 6 relative, offset displacements $d^{\mathrm{k}}(\mathbf{v}_i)$ in the three canonical directions x, y and z.

anatomical structures. We tackle this simultaneous feature selection and parameter regression task with a multi-class random regression forest (fig. 4); *i.e.* an ensemble of regression trees trained to predict location and size of all desired anatomical structures simultaneously.

Note that in the illustrative example in section 1.1 the goal was to estimate a two-dimensional continuous vector representing a line. In contrast, here the desired output is one six-dimensional vector $\mathbf{b}_c$ per organ, for a total of $6|\mathcal{C}|$ continuous parameters. Also note that this is very different from the task of assigning a categorical label to each voxel (*i.e.* the classification approach in [14]). Here we wish to produce confident predictions of a small number of continuous localization parameters. The *latent* voxel clusters are discovered automatically without supervised cluster labels.

Forest training. The training process constructs each regression tree and decides at each node how to best split the incoming voxels. We are given a subset of all labelled CT volumes (the training set), and the associated ground-truth organ bounding boxes (fig. 3a). The size of the forest T is fixed and all trees are trained in parallel. Each voxel is pushed through each of the trees starting at the root. Each split node applies the following binary test $\xi_j > f(\mathbf{v};\boldsymbol{\theta}_j) > \tau_j$ and based on the result sends the voxel to the left or right child node. $f(.)$ denotes the feature response computed for the voxel $\mathbf{v}$. The parameters $\boldsymbol{\theta}_j$ represent the visual feature which applies to the j^{th} node. Our visual features are similar to those in [10,14,19], *i.e.* mean intensities over displaced, asymmetric cuboidal regions. These features are efficient and capture spatial context. The feature response is $f(\mathbf{v};\boldsymbol{\theta}_j) = |F_1|^{-1} \sum_{\mathbf{q}\in F_1} I(\mathbf{q}) - |F_2|^{-1} \sum_{\mathbf{q}\in F_2} I(\mathbf{q})$; with F_i indicating 3D box regions and I the intensity. F_2 can be the empty set for unary features. Randomness is injected by making available at each node only a random sample of all features. This technique has been shown to increase the generalization of tree-based predictors [1]. Next we discuss how to select the node test.

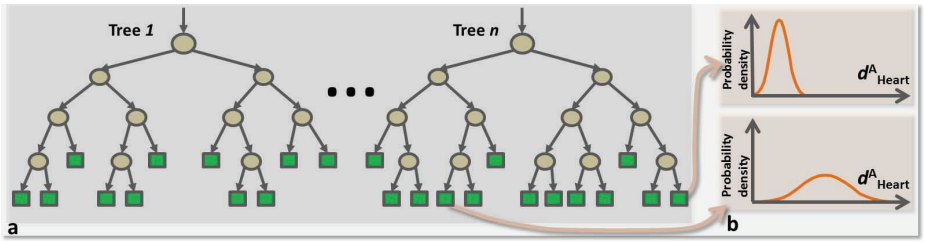

Fig. 4. A regression forest is an ensemble of different regression trees. Each leaf contains a distribution for the continuous output variable/s. Leaves have associated different degrees of confidence (illustrated by the "peakiness" of distributions).

Node optimization. Each voxel $\mathbf{v}$ in each training volume is associated with an offset $\mathbf{d}_c(\mathbf{v})$ with respect to the bounding box $\mathbf{b}_c$ for each class $c \in \mathcal{C}$ (see fig. 3b,c). Such offset is denoted: $\mathbf{d}_c(\mathbf{v}) = (d_c^{\mathtt{L}}, d_c^{\mathtt{R}}, d_c^{\mathtt{A}}, d_c^{\mathtt{P}}, d_c^{\mathtt{H}}, d_c^{\mathtt{F}}) \in \mathbb{R}^6$, with $\mathbf{b}_c(\mathbf{v}) = \hat{\mathbf{v}} - \mathbf{d}_c(\mathbf{v})$ and $\hat{\mathbf{v}} = (v_x, v_x, v_y, v_y, v_z, v_z)$. As with classification, node optimization is driven by maximizing an information gain measure, defined as: $IG = H(\mathcal{S}) - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i H(\mathcal{S}_i)$ where H denotes entropy, $\mathcal{S}$ is the set of training points reaching the node and $\mathtt{L}, \mathtt{R}$ denote the left and right children. In classification the entropy is defined over distributions of discrete class labels. In regression instead we measure the purity of the probability density of the real-valued predictions. For a single class c we model the distribution of the vector $\mathbf{d}_c$ at each node as a multivariate Gaussian; *i.e.* $p(\mathbf{d}_c) = \mathcal{N}(\mathbf{d}_c; \overline{\mathbf{d}_c}, \Lambda_c)$, with the matrix Λ_c encoding the covariance of $\mathbf{d}_c$ for all points in $\mathcal{S}$. The differential entropy of a multivariate Gaussian can be shown to be $H(\mathcal{S}) = \frac{n}{2}(1 + \log(2\pi)) + \frac{1}{2}\log|\Lambda_c(\mathcal{S})|$ with n the number of dimensions ($n = 6$ in our case). Algebraic manipulation yields the following regression information gain: $IG = \log|\Lambda_c(\mathcal{S})| - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i \log|\Lambda_c(\mathcal{S}_i)|$. In order to handle simultaneously all $|\mathcal{C}| = 10$ anatomical structures the information gain is adapted to: $IG = \sum_{c\in\mathcal{C}} \left(\log|\Lambda_c(\mathcal{S})| - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i \log|\Lambda_c(\mathcal{S}_i)|\right)$ which is readily rewritten as

$$IG = \log|\Gamma(\mathcal{S})| - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i \log|\Gamma(\mathcal{S}_i)|, \quad \text{with } \Gamma = \text{diag}\left(\Lambda_1, \cdots, \Lambda_c, \cdots, \Lambda_{|\mathcal{C}|}\right). \quad (1)$$

Maximizing (1) encourages minimizing the determinant of the $6|\mathcal{C}| \times 6|\mathcal{C}|$ covariance matrix Γ, thus decreasing the uncertainty in the probabilistic vote cast by each cluster of voxels on each organ pose. Node growing stops when IG is below a fixed threshold, too few points reach the node or a maximum tree depth D is reached (here $D = 7$). After training, the j^{th} split node remains associated with the feature $\boldsymbol{\theta}_j$ and thresholds ξ_j, τ_j. At each leaf node we store the learned mean $\overline{\mathbf{d}}$ (with $\mathbf{d} = (\mathbf{d}_1, \cdots, \mathbf{d}_c, \cdots, \mathbf{d}_{|\mathcal{C}|})$) and covariance Γ, (fig. 4b).

This framework may be reformulated using non-parametric distributions, with pros and cons in terms of regularization and storage. We have found our parametric assumption not to be restrictive since the multi-modality of the input space is captured by our hierarchical piece-wise Gaussian model.

Discussion. Equation (1) is an information-theoretical way of maximizing the confidence of the desired continuous output *for all* organs, without going through intermediate voxel classification (as in [14] where positive and negative examples of organ centres are needed). Furthermore, this gain formulation enables testing different context models; *e.g.* imposing a *full* covariance Γ would allow correlations between all walls in all organs, with possible over-fitting consequences. On the other hand, assuming a *diagonal* Γ (and diagonal class covariances Λ_c) leads to uncorrelated output predictions. Interesting models live in the middle ground, where Γ is sparse but correlations between selected subgroups of classes are enabled, to capture *e.g.* class hierarchies or other forms of spatial context. Space restrictions do not permit a more detailed description of these issues.

Forest testing. Given a previously unseen CT volume $\mathcal{V}$, each voxel $\mathbf{v} \in \mathcal{V}$ is pushed through each tree starting at the root and the corresponding sequence of tests applied. The voxel stops when it reaches its leaf node $l(\mathbf{v})$, with l indexing leaves across the whole forest. The stored distribution $p(\mathbf{d}_c|l) = \mathcal{N}(\mathbf{d}_c; \overline{\mathbf{d}}_c, \Lambda_c)$ for class c also defines the posterior for the absolute bounding box position: $p(\mathbf{b}_c|l) = \mathcal{N}(\mathbf{b}_c; \overline{\mathbf{b}}_c, \Lambda_c)$, since $\overline{\mathbf{b}}_c(\mathbf{v}) = \hat{\mathbf{v}} - \overline{\mathbf{d}}_c(\mathbf{v})$. The posterior probability for $\mathbf{b}_c$ is now given by

$$p(\mathbf{b}_c) = \sum_{l \in \tilde{\mathcal{L}}} p(\mathbf{b}_c|l) p(l). \tag{2}$$

$\tilde{\mathcal{L}}$ is a subset of all forest leaves. Here we select $\tilde{\mathcal{L}}$ as the set of leaves which have the smallest uncertainty (for each class c) and contain 1% of all test voxels. Finally $p(l) = 1/|\tilde{\mathcal{L}}|$ if $l \in \tilde{\mathcal{L}}, 0$ otherwise. This is different from averaging the output of all trees (as done *e.g.* in [9,10]) as it uses the most confident leaves, independent from which tree in the forest they come from.

Anatomy detection. The organ c is declared present in the scan if $p(\mathbf{b}_c = \tilde{\mathbf{b}}_c) > \beta$, with $\beta = 0.5$.

Anatomy localization. The final prediction $\tilde{\mathbf{b}}_c$ for the absolute position of the c^{th} organ is given by the expectation $\tilde{\mathbf{b}}_c = \int_{\mathbf{b}_c} \mathbf{b}_c p(\mathbf{b}_c) d\mathbf{b}_c$.

3 Results, Comparisons and Validation

This section assesses the proposed algorithm in terms of its accuracy, runtime speed and memory efficiency; and compares it to state of the art techniques.

Accuracy in anatomy localization. Qualitative results on automatic anatomy localization within previously unseen, whole-body CT scans are shown in fig. 5.

Quantitative evaluation. Localization errors are shown in table 1. The algorithm is trained on 55 volumes and tested on the remaining 45. Errors are defined as absolute difference between predicted and true wall positions. The table aggregates results over all bounding box sides. Despite the large data variability we

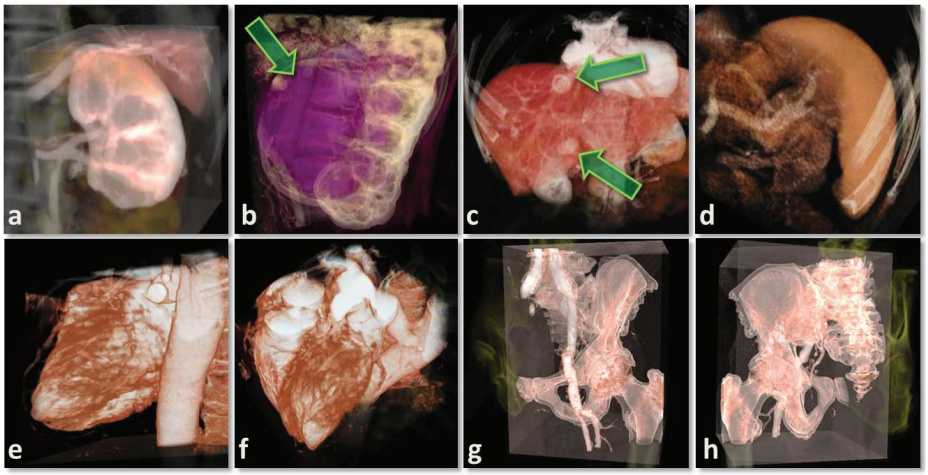

Fig. 5. Qualitative results showing the use of our automatic anatomy localizer for semantic visual navigation within 3D renderings of large CT studies. **(a)** The automatically computed bounding box for a healthy left kidney rendered in 3D. **(b)** As before but for a diseased kidney. **(c)** Automatically localized liver showing hemangiomas. **(d)** Automatically localized spleen, **(e, f)** heart and **(g, h)** left pelvis. Once each organ has been detected the 3D camera is repositioned, the appropriate cropping applied and the best colour transfer function automatically selected.

obtain a mean error of only $\sim 1.7cm$ (median $\sim 1.1cm$), sufficient to drive many applications. On average, errors along the z direction are about twice as large as those in x and y. This is due both to reduced resolution and larger variability in cropping along the z direction. Consistently good results are obtained for different choices of training set as well as different training runs.

Testing each tree on a typical 512^3 scan takes approximately $1s$ with our C++ implementation; and all trees are tested in parallel. Further speed-ups can be achieved with more low-level code optimizations.

Comparison with affine, atlas-based registration. One key aspect of our technique is its speed; important *e.g.* for clinical use. Thus, here we chose to compare our results with those obtained from a comparably fast atlas-based algorithm, one based on global registration. From the training set a reference atlas is selected as the volume which when registered with all *test* scans produced the minimum localization error. Registration was attained using the popular MedInria[3] package. We chose the global registration algorithm (from the many implemented) and associated parameters that produced best results on the *test* set. Such algorithm turned out to be block-matching with an affine transformation model. Note that optimizing the atlas selection and the registration algorithm on the test set produces results which are biased in favor of the atlas-based technique and yields a much tougher evaluation ground for our regression algorithm.

[3] `www-sop.inria.fr/asclepios/software/MedINRIA/`

Table 1. Regression forest results. Bounding box localization errors (in mm).

organ	heart	liver	spleen	left lung	right lung	left kidney	right kidney	gall bladder	left pelvis	right pelvis	*across all organs*
mean	15.4	17.1	20.7	17.0	15.6	17.3	18.5	18.5	13.2	12.8	**16.7**
std	15.5	16.5	22.8	17.2	16.3	16.5	18.0	13.2	14.0	13.9	**17.0**
median	9.3	13.2	12.9	11.3	10.6	12.8	12.3	14.8	8.8	8.4	**11.5**

Table 2. Atlas-based results. Bounding box localization errors (in mm).

organ	heart	liver	spleen	left lung	right lung	left kidney	right kidney	gall bladder	left pelvis	right pelvis	*across all organs*
mean	24.4	34.2	36.0	27.8	27.0	39.1	28.3	27.6	23.4	22.4	31.3
std	27.0	59.3	57.2	29.9	27.6	55.6	53.3	26.7	43.3	43.5	50.0
median	15.5	16.4	20.1	15.7	18.0	25.7	15.4	19.8	10.9	11.8	17.2

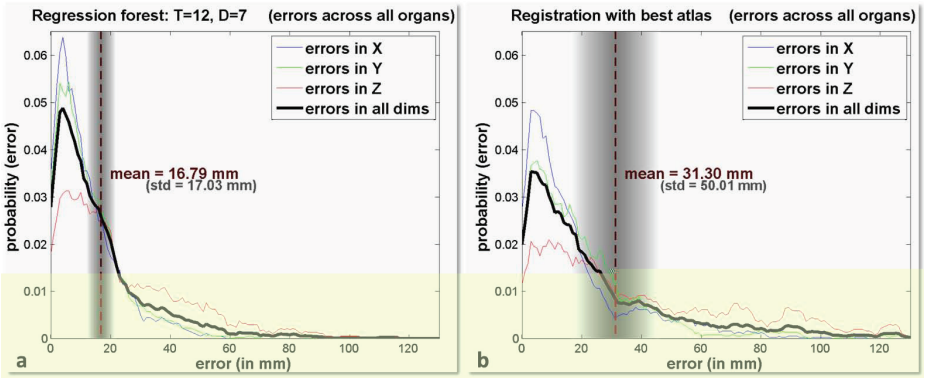

Fig. 6. Comparison with atlas-based registration. Distributions of localization errors for **(a)** our algorithm, and **(b)** the atlas-based technique. The atlas-induced errors show more mass in the tails, which is reflected by a larger standard deviation (std). The width of the vertical shaded band is proportional to the standard deviation.

The resulting errors (computed on the same test set) are reported in table 2. They show much larger error mean and standard deviation (about double) than our approach. Registration is achieved in between $90s$ and $180s$ per scan, on the same dual-core machine (*cf.* our algorithm runtime is $\sim 6s$ for $T = 12$ trees).

Figure 6 further illustrates the difference in accuracy between the two approaches. In the registration case larger tails of the error distribution suggest a less robust behavior[4]. This is reflected in larger values of the error mean and standard deviation and is consistent with our visual inspection of the registrations. In fact, in $\sim$ 30% cases the process got trapped in local minima and

[4] Because larger errors are produced more often than in our algorithm.

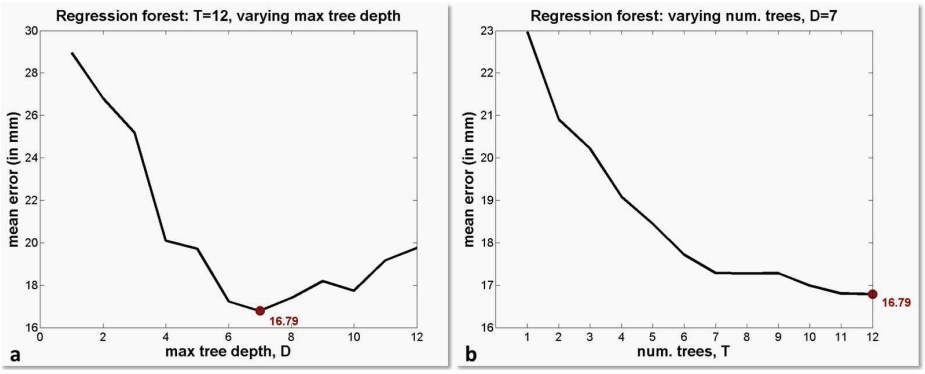

Fig. 7. Mean error as a function of forest parameters. (a) with varying maximum tree depth D and fixed forest size $T = 12$. **(b)** with fixed tree depth $D = 7$ and varying forest size T. All errors are computed on previously unseen test scans.

produced grossly inaccurate alignment. Those cases tend not to get improved when using a local registration step[5], while adding considerably to the runtime.

A regression forest with 6 trees takes $\sim$ 10MB of memory. This is in contrast with the roughly 100MB taken by each atlas. The issue of model size and runtime efficiency may be exacerbated by the use of more accurate and costly multi-atlas techniques [17]. Finally, in our algorithm increasing the training set usually decreases the test error without affecting the test runtime, while in multi-atlas techniques increasing the number of atlases linearly increases the runtime.

Comparison with voxel-wise classification. When compared to the classification approach in [14] we have found that our regression techniques produces errors less than half than those reported in [14] (on identical train and test sets) which, in turn demonstrated better accuracy than GMM and template-based approaches. In addition our regression algorithm computes the position of each wall (rather than just the organ centre), thus enabling approximate extent estimation.

Accuracy as function of forest parameters. Fig. 7 shows the effect of tree depth and forest size on accuracy. Trees deeper than 7 levels lead to over-fitting. This is not surprising as over-training with large trees has been reported in the literature. Also, as expected increasing the forest size T produces monotonic improvement without overfitting. No tree pruning has been employed here.

Automatic landmark detection. Fig 8 shows anatomical landmark regions automatically selected to aid organ localization. Given a trained tree and a chosen organ class (*e.g.* `left kidney`) we choose the two leaves with highest confidence. Then, we take the feature boxes (sect. 2.1) associated with the two closest ancestors (blue circles in fig. 8a) and overlay them (in green in fig. 8b,c) onto

[5] Which tends not to help escaping bad local minima.

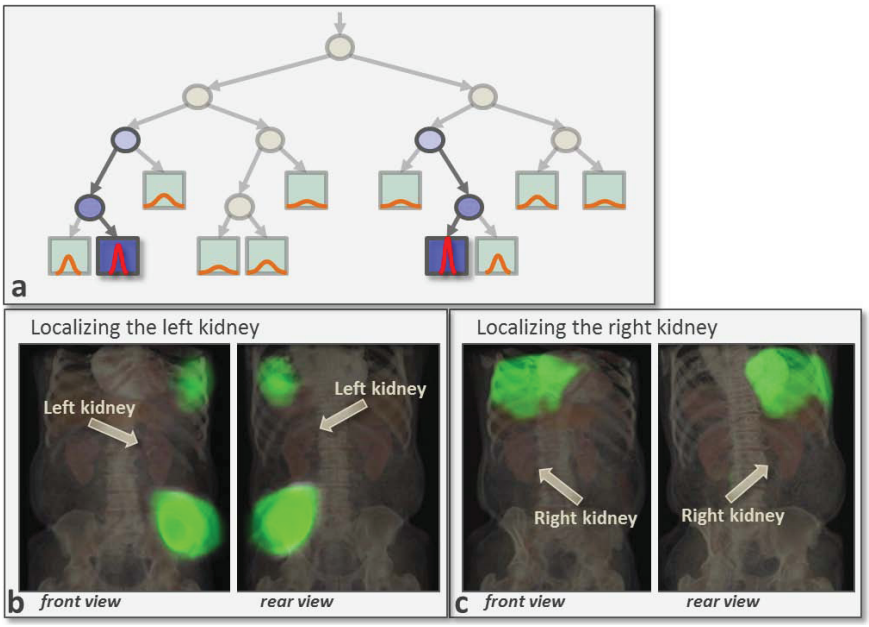

Fig. 8. Automatic discovery of salient anatomical landmark regions. (a) Given an organ class, the leaves associated to the two most confident distributions and two ancestor nodes are selected (in blue). **(b,c)** The corresponding feature boxes are overlayed (in green) on 3D renderings. The highlighted green regions correspond to anatomical structures which are automatically selected by the system to infer the position of the kidneys. See video in `http://research.microsoft.com/apps/pubs/default.aspx?id=135411`.

volumetric renderings, using the points reaching the leaves as reference. The green regions represent the anatomical locations which are used to estimate the location of the chosen organ. In this example the bottom of the left lung and the top of the left pelvis are used to predict the position of the left kidney. Similarly, the bottom of the right lung is used to localize the right kidney. Such regions correspond to meaningful, visually distinct, anatomical landmarks. They have been computed without any ground truth labels nor manual tagging.

4 Conclusion

Anatomy localization has been cast here as a non-linear regression problem where *all* voxels vote for the position of all anatomical structures. Location estimation is obtained via a multivariate regression forest algorithm which is shown to be more accurate and efficient than competing registration-based techniques.

At the core of the algorithm is a new information-theoretic metric for regression tree learning which enables maximizing the confidence of the predictions over the position of all organs of interest, simultaneously. Such strategy produces accurate predictions as well as meaningful anatomical landmark regions.

Accuracy and efficiency have been assessed on a database of 100 diverse CT studies. Future work includes exploration of different context models and extension to using other imaging modalities and non-parametric distributions.

References

1. Breiman, L.: Random forests. Technical Report TR567, UC Berkeley (1999)
2. Zhou, S.K., Zhou, J., Comaniciu, D.: A boosting regression approach to medical anatomy detection. In: IEEE CVPR, pp. 1–8 (2007)
3. Fenchel, M., Thesen, S., Schilling, A.: Automatic labeling of anatomical structures in MR fastView images using a statistical atlas. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 576–584. Springer, Heidelberg (2008)
4. Hardle, W.: Applied non-parametric regression. Cambridge University Press, Cambridge (1990)
5. Zhou, S., Georgescu, B., Zhou, X., Comaniciu, D.: Image-based regression using boosting method. In: ICCV (2005)
6. Friedman, J.: Greedy function approximation: A gradient boosting machine. The Annals of Statistics 2(28) (2001)
7. Vapnik, V.: The nature of statistical learning theory. Springer, Heidelberg (2000)
8. Yin, P., Criminisi, A., Essa, I., Winn, J.: Tree-based classifiers for bilayer video segmentation. In: CVPR (2007)
9. Montillo, A., Ling, H.: Age regression from faces using random forests. In: ICIP (2009)
10. Gall, J., Lempitsky, V.: Class-specific Hough forest for object detection. In: IEEE CVPR, Miami (2009)
11. Zhan, Y., Zhou, X.S., Peng, Z., Krishnan, A.: Active scheduling of organ detection and segmentation in whole-body medical images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 313–321. Springer, Heidelberg (2008)
12. Torralba, A., Murphy, K.P., Freeman, W.T.: Sharing visual features for multiclass and multiview object detection. IEEE Trans. PAMI (2007)
13. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Sühling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Pluim, J.P.W., Dawant, B.M. (eds.) SPIE (2009)
14. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI Workshop on Probabilistic Models for Medical Image Analysis (2009)
15. Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H.: Multi-organ segmentation in three-dimensional abdominal CT images. Int. J. CARS 1 (2006)
16. Yao, C., Wada, T., Shimizu, A., Kobatake, H., Nawano, S.: Simultaneous location detection of multi-organ by atlas-guided eigen-organ method in volumetric medical images. Int. J. CARS 1 (2006)
17. Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B.: Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in ct scans. IEEE Trans. Medical Imaging 28(7) (2009)
18. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. Chapman and Hall/CRC (1984)
19. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. In: IJCV (2009)