Automatic Detection and Segmentation of Kidneys in 3D CT Images Using Random Forests

Rémi Cuingnet[1], Raphael Prevost[1,2], David Lesage[1], Laurent D. Cohen[2], Benoît Mory[1], and Roberto Ardon[1]

[1] Philips Research Medisys, France
[2] CEREMADE, UMR 7534 CNRS, Paris Dauphine University, France

Abstract. Kidney segmentation in 3D CT images allows extracting useful information for nephrologists. For practical use in clinical routine, such an algorithm should be fast, automatic and robust to contrast-agent enhancement and fields of view. By combining and refining state-of-the-art techniques (random forests and template deformation), we demonstrate the possibility of building an algorithm that meets these requirements. Kidneys are localized with random forests following a coarse-to-fine strategy. Their initial positions detected with global contextual information are refined with a cascade of local regression forests. A classification forest is then used to obtain a probabilistic segmentation of both kidneys. The final segmentation is performed with an implicit template deformation algorithm driven by these kidney probability maps. Our method has been validated on a highly heterogeneous database of 233 CT scans from 89 patients. 80 % of the kidneys were accurately detected and segmented (Dice coefficient > 0.90) in a few seconds per volume.

1 Introduction

Segmentation of medical images is a key step to gathering anatomical information for diagnosis or interventional planning. Renal volume and perfusion, which can be extracted from CT images, are typical examples for nephrologists. However, it is often long and tedious for clinicians to segment 3D images. Automatic and fast segmentation algorithms are thus needed for practical use. It is yet still challenging to design an algorithm robust enough to noise, acquisition artifacts or leakages in neighboring organs.

Several papers in the literature tackle the problem of kidney segmentation in CT images. In [1] and [2], the authors used the Active Shape Model framework to learn the kidney mean shape and principal modes of variation, in order to constrain the segmentation. Recently Khalifa et al. [3] proposed a level-set approach, based on a new force combining shape and intensity priors as well as spatial interactions, which showed promising results. However, they were assessed on small datasets (41, 17 and 20 volumes in [1], [2] and [3] respectively). Moreover, all these algorithms are either based on a manual initialization, or tested on images already cropped around the kidney. A fully automatic method has

already been introduced by Tsagaan et al. [4], but their detection of the region of interest presents limitations. First, it relies on hard geometrical constraints, which requires knowledge on the field of view. Then, a rough search is done by template matching, which is not robust to pathologies or kidney orientation.

In this paper, we propose a fast and completely automatic method to detect and segment both kidneys in any kind of CT image: acquired at different contrast phases (or without contrast) with various fields of view, from both healthy subjects and patients with kidney tumors. Kidneys' positions are first detected with regression forests following a coarse to fine strategy (Section 2). Then a two-step segmentation is performed on cropped images around the kidneys (Section 3) using (i) a random forest to estimate a probability map of each kidney and (ii) a template deformation algorithm [5] to extract the kidney surface. Experiments and results are detailed in Section 4.

2   Kidney Detection with Regression Forests

This section presents a fast and reliable estimation of the kidneys' locations. Various approaches for anatomy detection and localization have been proposed in the literature (Section 2.1). We propose a regression-based method in two steps. The whole image is first used to provide an estimate of the region of interest (Section 2.2) which is then refined using local information only (Section 2.3).

2.1   Background on Organ Detection

Registration-based approaches using labeled atlases (e.g. [6]) have often been used for this problem. However such approaches are subject to registration errors due to inter individual variability. The robustness of the registration step can be improved by using multi-atlas or multi-template techniques [7] but at the cost of an increase in computation time.

More recently, supervised learning methods have been used for this detection problem to better take into account interindividual variability. Most ***classification-based*** detection algorithms consist in constructing a classifier whose role is to predict from local features to which organ a voxel belongs (e.g. [8]). However, by considering only local features, such approaches do not benefit from anatomical contextual information. To overcome this shortcoming, Criminisi et al. [9] used a generalization of Haar features that models contextual information. Instead of classifying each voxel, some authors consider the detection problem as finding a vector of parameters describing the organ locations. Such parameters can describe contour line or surface of an organ [10] or more simply bounding boxes around the different organs of interest [11]. The role of the classifier is then to predict whether a set of parameters is correct or not. Zheng et al. [11] used a greedy approach to avoid searching the whole parameter space, which is intractable.

Zhou et al. [12] showed that finding a set of continuous parameters from an image is by definition a multiple-output regression problem. More precisely, they

proposed a boosting ridge regression to detect and localize the left ventricle in cardiac ultrasound 2D images. ***Regression-based*** techniques do not require an exhaustive search of parameters. Other regressors such as regression forests and random ferns have also been proposed [13, 14].

In the following, we consider regression forests to simultaneously detect both kidneys. Regression forests [15, 16] are particularly well adapted to this problem in clinical routine since, thanks to their tree structures, they allow very fast testing with nonlinear regression functions. Since there is no explicit regularization, random forests require a large number of training samples to avoid overfitting the training data. Here, this is not a limitation since the training samples are the voxels of the training CT scans.

2.2   Coarse Localization of the Kidneys

We consider the detection step as the problem of finding bounding boxes around both kidneys. First, we find a coarse positioning based on contextual information adapting the approach proposed by Criminisi et al. [13]. Then, the position of each box is refined based on local information.

Each bounding box is parameterized by a vector in $\mathbb{R}^6$ (two points in 3D). A random forest is trained on CT scans with known kidney bounding boxes to predict for each voxel the relative position and size of the kidneys. Since CT intensities (expressed in Hounsfield units) have direct physical meaning, as explained in [13], the features used are the mean intensities over displaced, asymmetric cuboidal regions. To allow a much faster training, we used residual sum of squares (RSS) instead of the information gain for the node optimization in the training stage. Note that optimizing the RSS comes to minimizing the trace of the covariance matrix at each node instead of its determinant. We did not notice any differences in term of prediction accuracies.

This step gives a first estimate of the kidneys' positions and sizes. By construction, the relative estimated position of the left and right kidneys are strongly correlated. Such a correlation ensures coherent results but may not reflect the whole possible interindividual variability. This might be critical when the number of subjects in the training set is low. To overcome this shortcoming, we propose a refinement step of the bounding boxes that relaxes the correlation between the two kidneys' position.

2.3   Refinement of the Region of Interest

This step consists in refining the left and right kidneys' positions based on local information only. The constraints between the kidneys' relative positions are relaxed by treating both kidneys independently. For each kidney, a regression forest is trained to predict, from every voxel located in its neighborhood, the relative position of the kidney's center.We used the same training set as in the previous step. The features used for this step are, for each voxel, its intensity and its gradient magnitude, as well its neighbors'.

For testing, only the voxels in the neighborhood of the center of the bounding box predicted by the first step are considered. As depicted in Figure 1.b, each voxel v then votes for a location $\hat{\mathbf{c}}_v$ of the kidney's center. For robustness sake, the final location estimate is $\hat{\mathbf{c}} = \arg\min_{\mathbf{c} \in \mathbb{R}^3} \sum_{v=1}^{K} \|\mathbf{c} - \hat{\mathbf{c}}_v\|_1$ where $(\hat{\mathbf{c}}_v)_{1,K}$ are the K votes with the highest probability. The final bounding box is then translated accordingly.

To ensure stability, this refinement step is constrained to very small displacements and is iterated until convergence. This can be considered as a cascaded pose regression similar to [17]. Illustration of the kidney detection is given in Figure 1 and quantitative results are reported in Section 4.
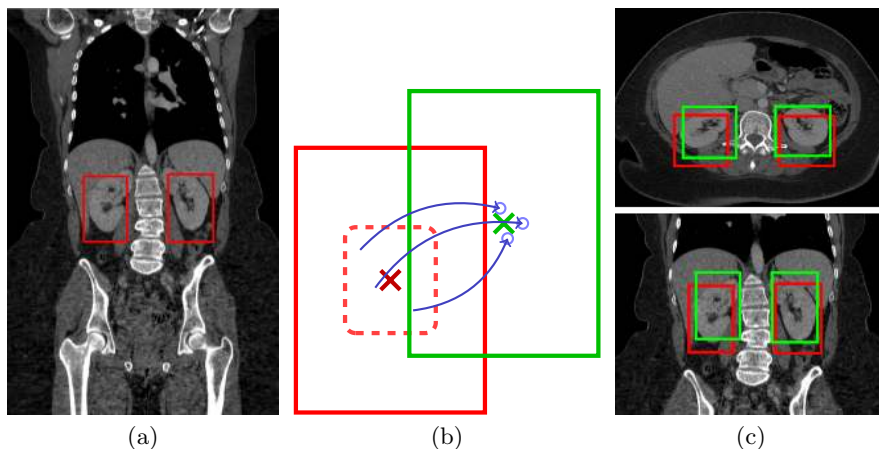

(a)   (b)   (c)

Fig. 1. Illustration of the kidney detection on a CT volume. (a) Initial bounding boxes detected using global contextual information. (b) Refinement step: voxels near the center of the initial bounding box (red) vote for its new center, using only local information. (c) Comparison between the initial (red) and refined (green) bounding box.

3   Kidney Segmentation

Even when the image is cropped to a region Ω around the kidney, its segmentation remains a challenging task: (i) kidneys are composed of different tissues (cortex, medulla, sinus) resulting in different image intensities, (ii) surrounding organs may touch the kidney without a clear boundary, (iii) the contrast phase of the CT image is unknown. For all these reasons, it is not possible to solely rely on the image intensity, and we rather use it simultaneously with other features.

3.1   Probability Estimation via Random Forests

In addition to regression, random forests can also be used to perform classification [15, 16]. We trained a random forest classifier to predict, for each voxel $\mathbf{x}$

of the previously detected bounding box, the probability $P(\mathbf{x})$ of belonging to a kidney. This random forest combines different image features: intensity and first/second order derivatives of the voxel and its neighbors. Decision stumps were used as weak classifiers and the impurity criterion was the Gini index [15, 16]. Such probability maps are shown in Figures 2.a and 2.c. Independently from the contrast-phase, the whole kidney tissues are enhanced, whereas the confusing adjacent structures are removed.

3.2 Initialization of the Segmentation

For the sake of robustness to interindividual variability and to pathologies, we only assumed that kidneys have a bean shape that can be roughly approximated by an ellipsoid. The segmentation algorithm is thus initialized with the ellipsoid $\mathcal{E} = \{\mathbf{x} \in \mathbb{R}^3 \mid (\mathbf{x}-\mathbf{c}_\mathcal{E})^T \mathcal{M}_\mathcal{E}^{-1}(\mathbf{x}-\mathbf{c}_\mathcal{E}) = 1\}$, where $\mathbf{c}_\mathcal{E} = \int_\Omega P(\mathbf{x})\,\mathbf{x}\,d\mathbf{x}$ denotes the weighted barycenter and $\mathcal{M}_\mathcal{E}$ is proportional to the weighted covariance matrix $\int_\Omega P(\mathbf{x})\,(\mathbf{x} - \mathbf{c}_\mathcal{E})(\mathbf{x} - \mathbf{c}_\mathcal{E})^T\,d\mathbf{x}$.

3.3 Implicit Template Deformation

We followed the framework introduced in [5] to deform the ellipsoid $\mathcal{E}$. A model-based approach is here particularly suited because (i) kidneys usually have very smooth shapes, (ii) we want the algorithm to reasonably extrapolate the boundary when the probability map is uncertain. Hereafter we recall the main principles of the adapted model-based deformation algorithm.

Given a working image $I : \Omega \to \mathbb{R}$ and the initial ellipsoid $\mathcal{E}$ defined by an implicit function ϕ, we find a transformation $\psi : \Omega \to \Omega$ such that the image gradient flux across the surface of the deformed ellipsoid $\mathcal{E}(\psi) = (\phi \circ \psi)^{-1}(0)$ is maximum. Denoting $\vec{\mathbf{n}}$ the normal vector, the segmentation energy is then

$$E_s(\psi) = \int_{\mathcal{E}(\psi)} - \left\langle \vec{\nabla} I(\mathbf{x}) \, , \, \vec{\mathbf{n}}(\mathbf{x}) \right\rangle \, d\mathbf{x} + \lambda \mathcal{R}(\psi) \, , \tag{1}$$

$\mathcal{R}(\psi)$ is a regularization term which prevents large deviations from the original ellipsoid. The transformation is decomposed as $\psi = \mathcal{L} \circ \mathcal{G}$ where

- $\mathcal{G}$ is a global linear transformation, which may correct or adjust the center, orientation and scales of the initial ellipsoid;
- $\mathcal{L}$ is a non-rigid local deformation, expressed using a displacement field $\mathbf{u}$ such that $\mathcal{L}(\mathbf{x}) = \mathbf{x} + (\mathbf{u} * K_\sigma)(\mathbf{x})$. K_σ is a Gaussian kernel that provides built-in smoothness, at a given scale σ.

This decomposition allows $\mathcal{R}$ to be pose-invariant and constrains only the non-rigid deformation : $\mathcal{R}(\psi) = \mathcal{R}(\mathcal{L}) = \int_\Omega \|\mathcal{L} - Id\|^2 = \int_\Omega \|\mathbf{u} * K_\sigma\|^2$. Finally, using Stokes formula, E_s can be rewritten as

$$E_s(\psi) = - \int_\Omega H(\phi \circ \mathcal{L} \circ \mathcal{G}) \, \Delta I + \lambda \int_\Omega \|\mathbf{u} * K_\sigma\|^2 \, , \tag{2}$$

where H is the Heaviside function and Δ is the Laplacian operator. This energy is minimized, with respect to the parameters of $\mathcal{G}$ and each component of the vector field $\mathbf{u}$, through a gradient descent. Note that since the energy in (2) is not convex, the resulting segmentation depends on the initialization. Hence, we first apply this algorithm to the probability map ($I = P$) in order to reach an appropriate local minimum (e.g. no leaks in surrounding tissues). The segmentation is finally refined on the original CT volume, with a higher shape constraint parameter λ and a finer scale σ (Figures 2.b and 2.d).

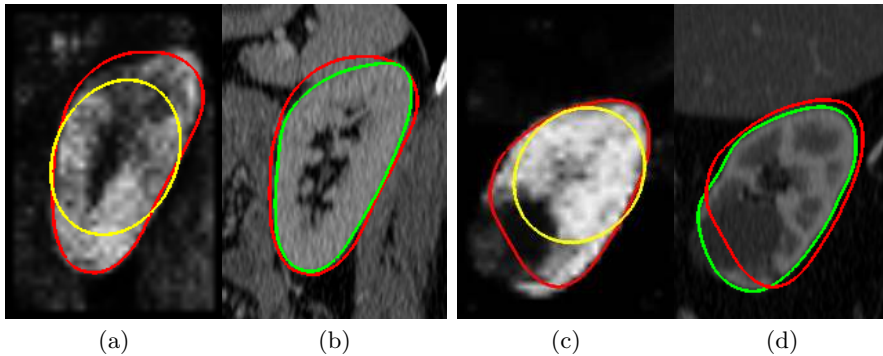| (a) | (b) | (c) | (d) |

Fig. 2. Illustration of the two-step kidney segmentation on two cases: (a-b) non-contrasted volume of a healthy patient, (c-d) contrast-enhanced image of a kidney with a tumor. The kidney probability maps (a) and (c) are learned with a random forest, and used to coarsely segment the kidney (red) by deforming an initial ellipsoid (yellow). The segmentation is then refined (green) using the original volumes (b) and (d).

4   Experiments and Results

The validation of our method was performed on a representative clinical dataset of 233 CT volumes from 89 subjects including diseased patients. The scans were contrast-enhanced or not and with various fields of view and spatial resolutions. They have between 33 and 973 (mean: 260) 512×512 slices with slice (resp. interslice) resolutions ranging from 0.5 to 1 mm (resp. 0.5 to 3.0 mm). 16% of the kidneys were slighlty truncated, but were nevertheless included in the evaluation to keep it clinically representative. The database was split into a training set of 54 volumes from 26 randomly selected patients, and a testing set composed of the other 179 volumes from 63 patients.

The proposed algorithm used 3 regression forests and 2 classification forests. Each forest was composed of 7 trees with a maximum tree depth $d = 15$ and a minimal node size $n = 100$. We did not notice a high sensitivity of the results to these parameters value. The whole training procedure lasts ~ 5 hours. Times are indicated for a C++ implementation (3.0 GHz dual-core, 4 Go RAM).

Kidney Detection Detection errors were defined as the absolute difference between predicted and true wall positions averaged over all the bounding box sides. The distance between the predicted bounding box center and the ground truth was also used to assess the detection accuracy. These results are given in Table 1 and compared to those reported in [13]. The refinement step (Section 2.3), for a low extra time cost, greatly increases the accuracy of the bounding box detection (e.g. the median center error is divided by 3).

Table 1. Detection results reported as: Mean ± Standard-deviation (Median)

Detection	Walls error (mm)		Center error (mm)		Time (s)
	Left	Right	Left	Right	Left+Right
[13]	17 ± 17 (13)	19 ± 18 (12)	–	–	–
Coarse	12 ± 7 (10)	13 ± 6 (11)	23 ± 14 (20)	26 ± 13 (23)	2.1 ± 0.5 (2.0)
Refined	7 ± 10 (5)	7 ± 6 (6)	11 ± 18 (6)	10 ± 12 (7)	2.8 ± 1.7 (2.4)

Automatic Segmentation The results of the automatic segmentation including the detection step were compared to the ground truth using the Dice index. Figure 3 shows the histograms of the scores for both kidneys. 80 % of the kidneys were correctly detected and segmented (Dice > 0.90). The algorithm failed in only 6% of the cases (Dice < 0.65). The total execution time is around 10 s.

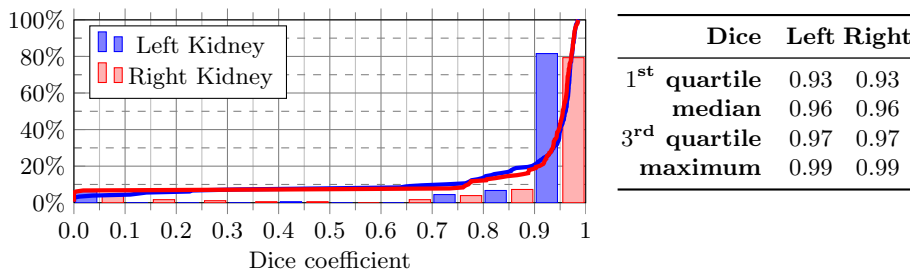| Dice	Left	Right
1$^{\text{st}}$ **quartile**	0.93	0.93
median	0.96	0.96
3$^{\text{rd}}$ **quartile**	0.97	0.97
maximum	0.99	0.99

Fig. 3. Distribution of the Dice coefficient between the ground truth and the automatically segmented kidneys. Red and blue lines show the cumulative distribution.

5   Conclusion

This paper presented a fully automatic method to detect and segment both kidneys in any CT volume using random regression and classification forests. Regression forests were used to estimate the kidneys' positions. A classification forest was then used to obtain a probability map of each kidney. The segmentation was carried out with an implicit template deformation algorithm. The

full automation and the execution time are compatible with clinical routine. Results show that our method provides an accurate segmentation in 80% of the cases despite the highly heterogeneous database. Remaining cases were mostly due to pathological kidneys not represented in the training set. Such cases could be quickly corrected by the clinician, since the chosen model-based deformation algorithm [5] allows user interactions. We also emphasize the generality of our framework, that could be as future work extended to other organs.

References

1. Spiegel, M., et al.: Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration. Comput Med Imaging Graph **33**(1) (2009) 29–39
2. Li, X., et al.: Renal Cortex Segmentation Using Optimal Surface Search with Novel Graph Construction. In: MICCAI. Volume 6893 of LNCS. Springer (2011) 387–94
3. Khalifa, F., et al.: 3D Kidney Segmentation from CT Images Using a Level Set Approach Guided by a Novel Stochastic Speed function. In: MICCAI. Volume 6893 of LNCS. Springer (2011) 587–94
4. Tsagaan, B., et al.: An Automated Segmentation Method of Kidney Using Statistical Information. In: MICCAI. Volume 2488 of LNCS. Springer (2002) 556–63
5. Mory, B., et al.: Real-Time 3D Image Segmentation by User-Constrained Template Deformation. In: MICCAI. LNCS, Springer (2012) *to appear*
6. Fenchel, M., et al.: Automatic Labeling of Anatomical Structures in MR FastView Images Using a Statistical Atlas. In: MICCAI. Volume 5241 of LNCS. (2008) 576–84
7. Isgum, I., et al.: Multi-atlas-based segmentation with local decision fusion: Application to cardiac and aortic segmentation in CT scans. IEEE Trans Med Imaging **28**(7) (2009) 1000–10
8. Montillo, A., et al.: Entangled Decision Forests and their Application for Semantic Segmentation of CT Images. In: Inf Process Med Imaging, Springer (2011) 184–96
9. Criminisi, A., et al.: Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In: MICCAI Workshop PMMIA. (2009)
10. Georgescu, B., et al.: Database-guided segmentation of anatomical structures with complex appearance. In: CVPR. Volume 2., IEEE (2005) 429–36
11. Zheng, Y., et al.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans Med Imaging **27**(11) (2008) 1668–81
12. Zhou, S., et al.: Image based regression using boosting method. In: ICCV. Volume 1., IEEE (2005) 541–48
13. Criminisi, A., et al.: Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In: Proc. of the 2010 international MICCAI conference on Medical computer vision. MCV'10, Springer (2011) 106–117
14. Pauly, O., et al.: Fast Multiple Organ Detection and Localization in Whole-Body MR Dixon Sequences. In: MICCAI. Volume 6893 of LNCS., Springer (2011) 239–47
15. Breiman, L.: Random forests. Machine learning **45**(1) (2001) 5–32
16. Criminisi, A., et al.: Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning. Technical report, Microsoft Research (2011)
17. Dollar, P., et al.: Cascaded pose regression. In: CVPR, IEEE (2010) 1078–85