

Research Article

3D Kidney Segmentation from Abdominal Images Using Spatial-Appearance Models

Fahmi Khalifa,^{1,2} Ahmed Soliman,¹ Adel Elmaghraby,³
Georgy Gimel'farb,⁴ and Ayman El-Baz¹

¹Bioengineering Department, University of Louisville, Louisville, KY, USA

²Electronics and Communication Engineering Department, Mansoura University, Mansoura, Egypt

³Computer Engineering and Computer Science Department, University of Louisville, Louisville, KY, USA

⁴Department of Computer Science, University of Auckland, Auckland, New Zealand

Correspondence should be addressed to Ayman El-Baz; aselba01@louisville.edu

Received 19 August 2016; Revised 29 November 2016; Accepted 22 December 2016; Published 9 February 2017

Academic Editor: Po-Hsiang Tsui

Copyright © 2017 Fahmi Khalifa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kidney segmentation is an essential step in developing any noninvasive computer-assisted diagnostic system for renal function assessment. This paper introduces an automated framework for 3D kidney segmentation from dynamic computed tomography (CT) images that integrates discriminative features from the current and prior CT appearances into a random forest classification approach. To account for CT images' inhomogeneities, we employ discriminate features that are extracted from a higher-order spatial model and an adaptive shape model in addition to the first-order CT appearance. To model the interactions between CT data voxels, we employed a higher-order spatial model, which adds the triple and quad clique families to the traditional pairwise clique family. The kidney shape prior model is built using a set of training CT data and is updated during segmentation using not only region labels but also voxels' appearances in neighboring spatial voxel locations. Our framework performance has been evaluated on in vivo dynamic CT data collected from 20 subjects and comprises multiple 3D scans acquired before and after contrast medium administration. Quantitative evaluation between manually and automatically segmented kidney contours using Dice similarity, percentage volume differences, and 95th-percentile bidirectional Hausdorff distances confirms the high accuracy of our approach.

1. Introduction

Kidney segmentation from dynamic contrast-enhanced computed tomography (CT) is of immense importance for any computer-assisted diagnosis of renal function assessment, pathological tissue localization, radiotherapy planning, and so forth [1]. Nevertheless, accurate segmentation of kidney tissues from dynamic CT images is challenging due to many reasons, including data acquisition artifacts, large inhomogeneity of the kidney (e.g., cortex and medulla), large anatomical differences between subjects, similar intensities of adjacent organs, and varying signal intensities over the time course of data collection due to agent transit [2, 3]; see Figure 1.

Many automated and semiautomated approaches have been developed to address these challenges. Earlier comput-

erized renal image analysis (e.g., [4]) was usually carried out either manually or semiautomatically. Typically, a user-defined region-of-interest (ROI) is delineated in one image and for the rest of the images, image edges were detected and the model curve was matched to these edges. However, ROI placements are based on the users' knowledge of anatomy and thus are subject to inter- and intraobserver variability. Additionally, these methods are very slow, even though semiautomated techniques reduce the processing time. Traditional segmentation techniques utilizing image thresholding or region growing [5–9] have been also explored for CT kidney segmentation. For example, Pohle and Toennies [7] developed an automatic region-growing algorithm based on estimating the homogeneity criterion from the characteristics of the input images. A semiautomated method was also proposed by Leonardi et al. [9]. First, a region-growing approach

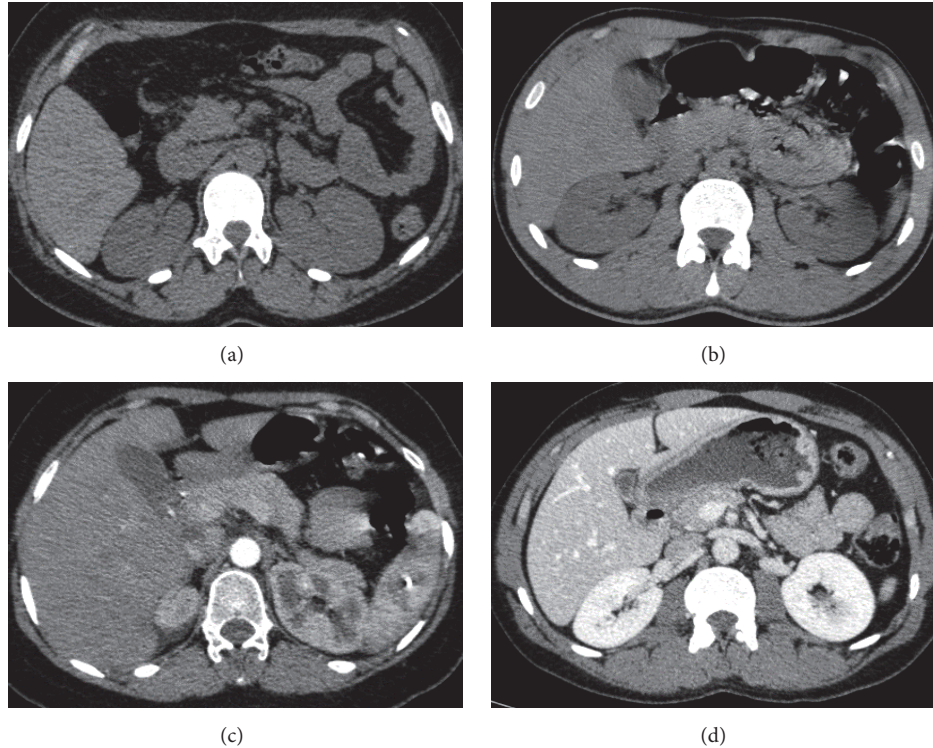


FIGURE 1: Axial cross-sectional images showing different CT data challenges: (a) low contrast, (b) intensity inhomogeneities, (c) fuzzy boundary, and (d) contrast and anatomy differences.

is performed to obtain an initial kidney segmentation from the grayscale image stack. Then, a refinement step utilizing histogram analysis is employed to the initially segmented kidney regions to reduce incorrectly segmented areas. However, these traditional methods are not accurate due to the large overlap of gray level intensity between the kidney and other surrounding tissues in addition to sensitive to initialization.

To more accurately segment abdominal CTs, recent segmentation methods consider either visual appearances, prior shapes, texture features, or hybrid techniques. In particular, Tsagaan et al. [10] presented a deformable model-based approach utilizing a nonuniform rational B-spline surface representation. Their framework incorporated statistical shape information (e.g., mean and variation) into the objective function for the model fitting process as an additional energy term.

A constrained optimization deformable contour by Wang et al. [11] exploited the degree of contour interior homogeneity as an extra constraint within the level set energy minimization framework. Lu et al. [12] developed a coarse-to-fine approach for kidney segmentation on abdominal CT images using the Chan-Vese (CV) level set method [13]. Mathematical morphology operations are performed to extract the kidney structures interactively with prior anatomy knowledge. Huang et al. [14] proposed a multiphase level set approach with multidynamic shape models to segment the kidneys on abdominal CT images. Campadelli et al. [15] proposed an automatic, gray-level based segmentation framework based on a multiplanar fast marching method.

A stochastic level set-based framework by Khalifa et al. [16, 17] integrated probabilistic kidney shapes and image signals priors into Markov random field (MRF) for abdominal 3D CT kidney segmentation. Despite their popularity, deformable model-based methods fail in the case of excessive image noise, poor image resolution, or diffused boundaries if they do not take advantage of a priori models.

Freiman et al. [18] proposed a model-based framework utilizing maximum a posteriori-MRF (MAP-MRF) estimation of the input CT image. The MAP-MRF estimation is obtained by using a graph min-cut technique. Lin et al. [19] proposed a framework that combined region- and model-based methods. Initial kidney location is estimated using geometrical location, statistical information, and a priori anatomical knowledge. Secondly, an elliptic candidate kidney region extraction approach is proposed. Finally, an adaptive region-growing approach is employed for kidney segmentation. Spiegel et al. [20] proposed an active shape model (ASM) based framework that was combined with a curvature-based nonrigid registration approach to solve the point correspondence problem of the training data. A hybrid framework by Chen et al. [21] combined active appearance model (AAM), live wire, and graph-cuts methods for 3D abdominal organ segmentation. In general, parametric shape-based techniques depend on the existence of adequate texture features in abdominal images and may perform poorly due to noise and the lack of well-defined features. Cuingnet et al. [22] exploited random regression and classification forests for CT kidney images segmentation. Initially, global contextual

information is used to detect the kidney. This is followed by a cascade of local regression forests for refinement. Then, probabilistic segmentation maps are built using classification forest. Finally, an implicit template deformation algorithm driven by these maps is employed to obtain the final segmentation. A model-based framework by Badakhshanoory and Saeedi [23] combined low-level segmentation schemes with a statistical-based modeling approach. First, an organ space is built using a statistical model and principle component analysis. Then, each image slice of an input CT volume is segmented multiple times using a graph-based segmentation by varying segmentation parameters. Finally, a distance-based criterion from the organ space is used to choose the closest candidate as the best segmentation result. In general, knowledge-based approaches are computationally intensive, and their accuracy depends on the training data size.

Bagci et al. [24] developed a multiobject segmentation framework that integrates a statistical shape model and hierarchical object recognition into a global graph-cuts segmentation model. Wolz et al. [25] developed a hierarchical two-step atlas registration framework for multiobject segmentation. First, subject-specific priors are generated from an atlas database based on multiatlas registration and patch-based segmentation. Final segmentation is obtained using graph-cuts, incorporating high-level spatial knowledge and a learned intensity model. Another study by Okada et al. [26] performed multiobject segmentation using probabilistic atlases that combines interorgan spatial and intensity a priori models. Despite the potential to improve the segmentation accuracy due to the spatial kidney constraints from other organs, multiobject segmentation schemes require more comprehensive prior information. A semiautomated Grow-Cut algorithm by Dai et al. [27] employed a monotonically decreasing function and image gray features to propagate initial user-defined labels over all the slices to derive an optimal cut for a given CT data in space. Zhao et al. [28] proposed a sliced-based framework for 3D kidney segmentation. First, an initial segmentation is obtained using the CV approach [13]. Then, a set of contextual features (e.g., slices overlap, the distance) and multiple morphological operations are used to estimate the continuity between slices. The final segmentation is obtained by discarding the leakage and the weak edges between adjacent slices using a local iterative thresholding method. Chu et al. [29] presented an automated MAP-based multiorgan segmentation method that incorporated image-space division and multiscale weighting scheme. Their framework is based on a spatially divided probabilistic atlases and the segmentation is performed using a graph cut method. Yang et al. [30] developed on multiatlas framework using a two-step approach to obtain coarse-to-fine kidney segmentation. A coarse segmentation is obtained by registering an input down-sampled CT volume with a set of low-resolution atlas images. Then, cropped kidney images are coaligned with high-resolution atlas images using B-Splines registration. The final segmentation result is obtained by majority voting of all deformed labels of all atlas images. Liu et al. [31, 32] developed a framework for kidney segmentation on noncontrast CT images using efficient belief propagation. A preprocessing step is applied to extract anatomical landmarks to localize

kidney search regions. Then, an efficient belief propagation is used to extract the kidney by minimizing an energy function that incorporates intensity and prior shape information. However, the method was evaluated on five noncontrast CT data sets only and additional segmentation of other organs (e.g., liver, spleen) is required to determine subimages that envelope the kidneys.

In summary, during the last few years there have been numerous studies for abdominal CTs kidney segmentation. In addition to the above-mentioned limitations, current methods have the following shortcomings. Most of them are based on visual appearance and did not take into account the spatial interaction relationships. Most of the shape-based methods utilize fixed models and therefore have limited accuracy for CT data outside their training scope. Most of the existing methods work very well with contrast CTs only. Most of the energy-based methods (e.g., graph-cut) use regional and boundary information that may not exist in some (e.g., precontrast) images and may not achieve globally optimal results.

To account for these limitations, we developed a 3D kidney segmentation framework that integrates, in addition to the current CT appearance features, higher-order appearance models and adaptive shape model features into a random forests (RF) classification model [33]. The integrated features increase the ability of our framework to account for the large CT images' inhomogeneities and therefore accurately segment both contrast and noncontrast CTs. Particularly, the spatial features are based on a higher-order Markov-Gibbs random field (MGRF) model that adds to the traditional pairwise cliques [34] the families of the triple and quad cliques. The spatial-appearance kidney shape prior is an adaptive model that is updated during segmentation and accounts not only for region labels, but also intensities in neighboring spatial locations. Moreover, compared to other tissue classification methods the RF is employed due to its (i) powerful ability to learn the characteristics of complex data classes [35], (ii) less sensitivity to data outliers, (iii) ability to overcome overfitting of the training set, and (iv) ability to handle high dimensional spaces as well as large number of training examples.

A detailed description of our developed methodology for kidney segmentation from dynamic CT images including the details of the discriminative features is given in Section 2. It is worth mentioning that, in addition to our methodology presentation in [33], this paper provides (i) a more comprehensive review of the related literature work on the abdominal CT images segmentation (Section 3); (ii) detailed description of the metrics that are used for segmentation evaluation of our and compared techniques (Section 3); and (iii) expansion of the experimental results by adding an essential metric that is used to evaluate the robustness of segmentation techniques, namely, the receiver operating characteristics (ROC) (Section 4).

2. Methods

A block diagram of our kidney segmentation framework is shown in Figure 2. Our technique is based on random forests

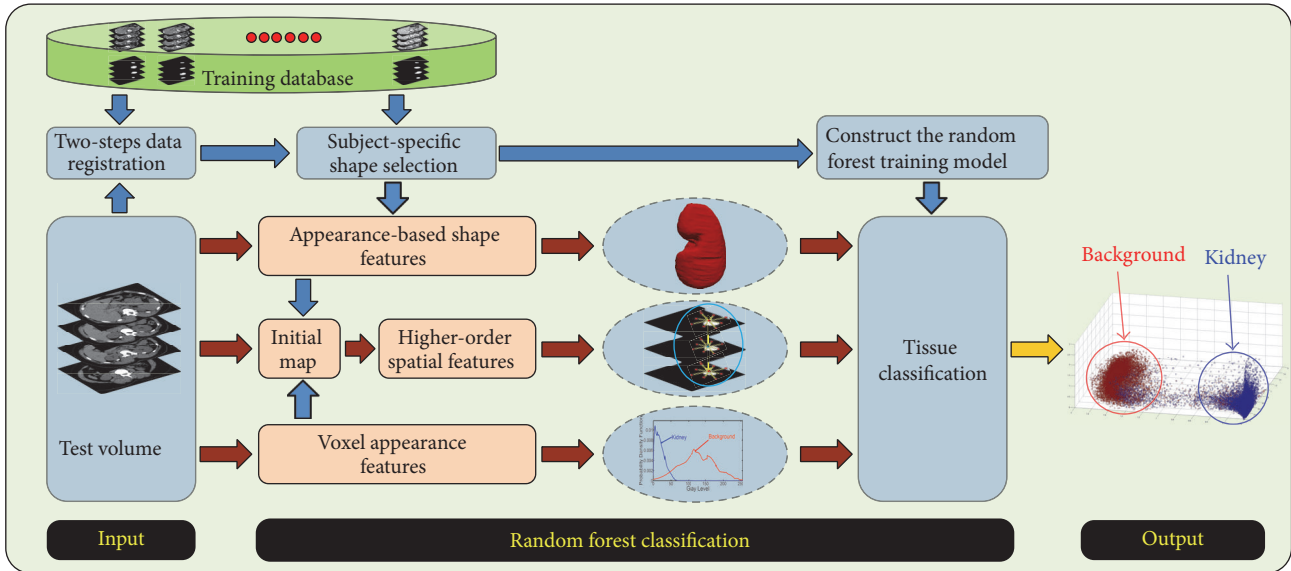


FIGURE 2: Block diagram of our kidney segmentation framework from abdominal CT images using random forest (RF).

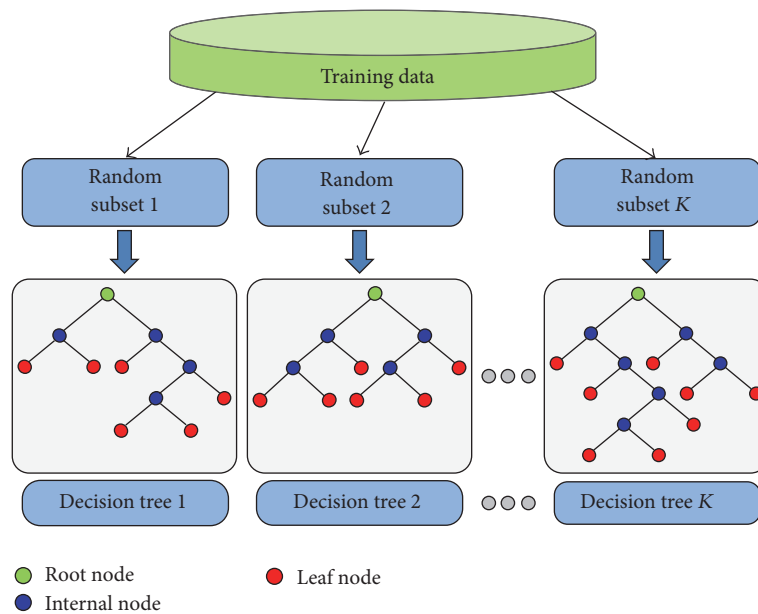


FIGURE 3: A schematic illustration of the random decision trees for random forests (RF) classification.

(RF) classification and incorporates spatial-appearance features for better separation of the CT data classes. RF is an efficient multiclass machine learning technique, which is increasingly being utilized in data clustering as well as image classification. As an ensemble learning classifier, RF typically consists of many decision trees (DTs) and combines two main concepts [36]. The first is the random selection of features and the second is “bagging” [37], which implies the training of each DT with a randomly chosen and overlapping subset of the training samples. In general, as numbers of the DTs increase the results get better. Nevertheless, there is a threshold beyond which the performance benefit from

adding more DTs will be lower than the computational cost for learning these additional DTs [38].

During the RF training phase, each DT recursively processes its randomly selected training samples’ features along a path starting from the tree’s root node using binary classification tests, as shown in Figure 3. The latter tests compare the features’ values at each internal tree node to a certain threshold that is selected using a certain criterion. A leaf node of the DT is reached if all samples belong to a single class; the number of data samples is smaller than a predefined value, or the maximum tree depth is reached [35]. Once occurred, the most frequent class label of the training

data at the node is stored for the testing phase. For testing, a given data sample is handled by applying respective tests in line with the path it traverses from the tree root node to the leaf. When a leaf node is reached, the DT casts a vote corresponding to the class assigned to this node in the training stage. Finally, a majority voting is used to class-label test samples. The final class probabilities are estimated by the fraction of votes for that class by all DTs.

In order to build an accurate RF model that provides better separation of data classes, discriminative and robust features are needed. Therefore, in this paper multiple features from the CT data are extracted, for both training and testing phases. These features include (i) first-order appearance (Hounsfield units (HUs) values) features; (ii) higher-order spatial interaction features; and (iii) appearance-based shape model features. Those features are extracted at each voxel's location $\mathbf{p} = (x, y, z)$ in the 3D arithmetic lattice $\mathbf{R} = \{(x, y, z) \mid 0 \leq x \leq X - 1, 0 \leq y \leq Y - 1, 0 \leq z \leq Z - 1\}$ supporting the grayscale CT images, $\mathbf{g} = \{g_{\mathbf{p}} : \mathbf{p} \in \mathbf{R}, g_{\mathbf{p}} \in \mathbf{Q}\}$, and their region, or segmentation maps, $\mathbf{m} = \{m_{\mathbf{p}} : \mathbf{p} \in \mathbf{R}, m_{\mathbf{p}} \in \mathbf{L}\}$. Here, $\mathbf{Q} = \{0, 1, \dots, Q - 1\}$ and $\mathbf{L} = \{\text{"KT"}, \text{"OT"}\}$ is a finite set of integer gray levels and region labels (kidney object tissues "KT" and other background tissues "OT"), respectively. Since spatial and shape features are based on probabilistic models, the first-order appearance-based features were also normalized to reduce the domination of a specific feature during RF classification. Details of the employed features are given in the following sections.

2.1. First-Order Appearance Features. The first type of features that are used in our framework is the CT voxel-appearance features. Those features were extracted at each voxel \mathbf{p} regionally from the CT data after normalization. Due to image noise presence and reconstruction artifacts, we used, at each voxel \mathbf{p} , regional intensity features in addition to the local CT Hounsfield units (HU). Namely, we used the mean HU values of a symmetric 3D cube (i.e., voxels' 26-neighbors) centered around \mathbf{p} and the mean of the HUs of a 3×3 in-plane symmetric window (i.e., voxels' 8-neighbors) centered around \mathbf{p} .

2.2. Shape Prior Features. The ultimate goal is to accurately segment the kidney from the CT data such that the extracted kidney borders closely approximate the expert manual delineation. However, due to the similar visual appearance between some kidney structures (e.g., medulla) and background, the segmentation should not rely only on image signals. Therefore, shape features of the expected kidney shape are also used in our segmentation framework. In this paper, we employed an adaptive, probabilistic kidney shape model that takes into account not only voxels' location, but also their intensity information [39, 40].

For training, a shape database is constructed using a set of training data sets that is collected from different subjects; each contains multiple CT scans acquired at different phases of contrast-enhancements. The ground truth segmentation (labeled data) of the training images is obtained by manual delineation of the kidney borders by an expert. In order to

reduce the variability across subjects and maximize overlaps of the kidneys for estimating the shape prior probability, the training grayscale images are coaligned using a two-step registration methodology. First, a 3D affine transformation is used with 12 degrees of freedom (3 for the 3D translation, 3 for the 3D rotation, 3 for the 3D scaling, and 3 for the 3D shearing) to account for global motion [41]. Second, local kidney deformations are handled using a 3D B-splines based transformation proposed in [42]. Finally, the obtained transformation parameters for each scan are applied to its binary (labeled) data to be used during segmentation to estimate the shape prior probability.

For testing, an input grayscale 3D CT kidney image, \mathbf{g}_t , to be segmented is first coaligned with the training database using the two-step registration methodology described above. Then, a subject-specific shape, \mathbf{g}_i , $i = 1, 2, \dots, N$, is extracted by computing the conventional normalized cross correlations (NCC) between the coaligned input grayscale image and all grayscale images in the database, to select the top N similar kidneys ($N = 19$ in our experiments below). Finally, visual appearances of both the input 3D grayscale CT image and the selected grayscale training images guide adapting the shape prior. Namely, the voxel-wise probabilities, $P_{s;\mathbf{p}}(l)$ for the adaptive shape prior $P_s(\mathbf{m}) = \prod_{\mathbf{p} \in \mathbf{R}} P_{s;\mathbf{p}}(m_{\mathbf{p}})$, are estimated based on the found voxels $l \in \mathbf{L}$. Let $\mathbf{v}_{i;\mathbf{p}}(l) = \{\rho : \rho \in \mathbf{R}; \rho \in \mathbf{C}_{\mathbf{p}}; |g_{i;\rho} - g_{t;\rho}| \leq \tau\}$ be a subset of similar training voxels within a search cube $\mathbf{C}_{\mathbf{p}}$ in the training image g_i , where τ is a predefined fixed signal range and $g_{t;\rho}$ is the mapped input signal. Let $v_{i;\mathbf{p}} = \text{card}(\mathbf{v}_{i;\mathbf{p}})$ denote the cardinality (number of voxels) of this subset $\mathbf{v}_{i;\mathbf{p}} = \sum_{i=1}^N \mathbf{v}_{i;\mathbf{p}}$ and $\delta(z)$ be the Kronecker's delta-function: $\delta(0) = 1$ and 0 otherwise. Then $P_{s;\mathbf{p}}(l)$ is given as [39]

$$P_{s;\mathbf{p}}(l) = \frac{1}{v_{\mathbf{p}}} \sum_{i=1}^N \sum_{\rho \in \mathbf{v}_{i;\mathbf{p}}} \delta(1 - m_{i;\rho}). \quad (1)$$

More details about the adaptive shape model can be found in [39, 40]. Our experiments were conducted using three shape features, like the voxel-appearance features. Namely, we used the $P_s(\mathbf{m})$ value at \mathbf{p} , the average $P_s(\mathbf{m})$ value for the 26 neighbors of a 3D cube around \mathbf{p} , and the average $P_s(\mathbf{m})$ of the 8 in-plane neighbors for a 3×3 symmetric window centered at \mathbf{p} .

2.3. Spatial Features. To improve the segmentation accuracy and account for the large inhomogeneity of the kidney, we incorporated into our segmentation approach the spatial features that describe the relationships between the kidney voxels and their neighbors. These relationships are described using a higher-order spatial model with analytically estimated potentials. The spatial modeling enhances the segmentation by calculating the likelihood of each voxel to be kidney or background on the basis of the initial labeling, \mathbf{m} , of the adjacent voxels, formed by a voxel-wise classification using shape and intensity values. Our spatial interactions model adds the triple and quad clique families to the traditional pairwise clique family [34] using the 18-connectivity neighborhood. Thus, it is an extension of the conventional Potts model [43],

differing only in that the potentials are estimated analytically. For more mathematical details about our higher-order spatial model, please see [33, 44]. Similar to the other features, three spatial-based features were used: the local spatial probability at \mathbf{p} and the average probabilities for a 3D cube and a 3×3 window centered around \mathbf{p} . In total, the whole segmentation approach is summarized in Algorithm 1.

Algorithm 1 (3D kidney segmentation steps).

Step 1 (data coalignment and shape database selection)

- (a) Register the input grayscale CT volume to the training database using the two-step registration in Section 2.2.
- (b) Calculate the NCC between the input coaligned data and all training volumes. Then, select the NCC-19-top ranked training samples.

Step 2 (features extraction)

- (a) Estimate the voxel-appearance features of the coaligned CT volume.
- (b) Estimate the higher-order Potts-MGRF spatial probabilities $P_G(\mathbf{m})$.
- (c) Estimate the appearance-based shape prior $P_s(\mathbf{m})$ using the method described in [39, 40].

Step 3 (RF training)

- (a) Construct the RF training model for the selected 19-top-ranked training images.

Step 4 (tissue segmentation)

- (a) Obtain the final segmentation of the input CT volume using the model in Step 3.

3. Segmentation Evaluation Metrics

The performance of our segmentation is evaluated using two metrics. The first is a volumetric-based similarity that characterizes spatial overlaps and volume differences between the segmented and “ground-truth” kidney regions. This type of metrics is important for studying area measurements, for example, total kidney volumes. The second is a distance-based metric that measures how close the edge of a segmented region is to the ground truth, that is, how accurate the shape of a segmented object is with respect to ground truth. Here, we used the Dice coefficient (DC) and percentage volume difference (PVD) to describe the volumetric-based similarity, while the bidirectional 95th-percentile Hausdorff distance (BHD_{95}) is used to characterize the distance-based error metric: $\mathbf{G} \leftrightarrow \mathbf{S}$.

Let \mathbf{G} and \mathbf{S} denote sets of ground-truth and segmented kidney voxels, respectively. The similarity volumetrics evaluate an overlap between these sets and account for cardinalities (i.e., voxel numbers) $c_i = |V_i|$ of true positive (tp), false

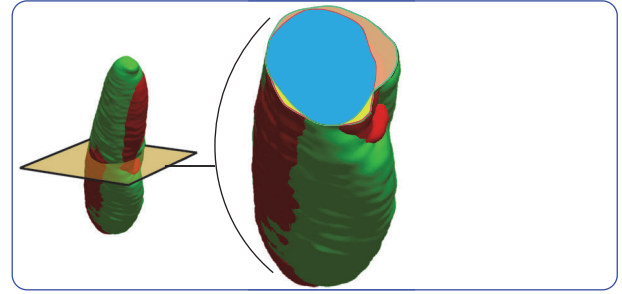


FIGURE 4: 3D illustration of DC measurement for segmentation evaluation between the ground truth \mathbf{G} and model segmentation \mathbf{S} .

positive (fp), and false negative (fn) subsets V_i ; $i \in \{\text{tp}, \text{fp}, \text{fn}\}$; see Figure 4.

The subsets contain true kidney voxels labeled as kidney, nonkidney (background) voxels labeled as kidney, and true kidney voxels labeled as background, respectively:

$$\begin{aligned} V_{\text{tp}} &= \{v : v \in \mathbf{G}, v \in \mathbf{S}\}; & c_{\text{tp}} &= |V_{\text{tp}}| \\ V_{\text{fp}} &= \{v : v \notin \mathbf{G}, v \in \mathbf{S}\}; & c_{\text{fp}} &= |V_{\text{fp}}| \\ V_{\text{fn}} &= \{v : v \in \mathbf{G}, v \notin \mathbf{S}\}; & c_{\text{fn}} &= |V_{\text{fn}}|. \end{aligned} \quad (2)$$

Obviously, $\mathbf{G} = V_{\text{tp}} \cup V_{\text{fn}}$; $\mathbf{S} = V_{\text{tp}} \cup V_{\text{fp}}$; $V_{\text{tp}} = \mathbf{G} \cap \mathbf{S}$; and $V_{\text{tp}} \cup V_{\text{fp}} \cup V_{\text{fn}} = \mathbf{G} \cup \mathbf{S}$ where \cup and \cap denote the set union and intersection, respectively. Therefore, it holds that $|\mathbf{G}| = c_{\text{tp}} + c_{\text{fn}}$; $|\mathbf{S}| = c_{\text{tp}} + c_{\text{fp}}$, and $|\mathbf{G} \cup \mathbf{S}| = c_{\text{tp}} + c_{\text{fp}} + c_{\text{fn}}$. The DC [45] and the PVD are defined as

$$\begin{aligned} \text{DC} &= 100 \frac{2c_{\text{tp}}}{2c_{\text{tp}} + c_{\text{fp}} + c_{\text{fn}}} \equiv 100 \frac{2|\mathbf{G} \cap \mathbf{S}|}{|\mathbf{G}| + |\mathbf{S}|} \\ \text{PVD} &= 100 \frac{(c_{\text{tp}} + c_{\text{fn}}) - (c_{\text{tp}} + c_{\text{fp}})}{c_{\text{tp}} + c_{\text{fn}}} \equiv 100 \frac{|\mathbf{G}| - |\mathbf{S}|}{|\mathbf{G}|}. \end{aligned} \quad (3)$$

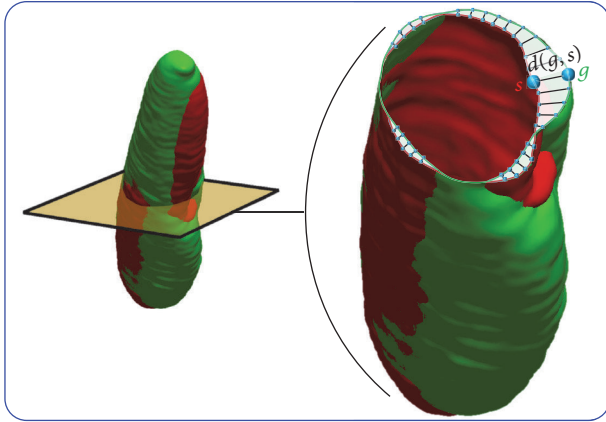
In addition to the DC and PVD, the 95th-percentile bidirectional Hausdorff distance (BHD_{95}) is used to measure dissimilarities between the \mathbf{G} and \mathbf{S} boundaries; see Figure 5. The HD from \mathbf{G} to \mathbf{S} is the maximum distance from the points g from \mathbf{G} to their closest points s in \mathbf{S} [46]:

$$\text{HD}_{\mathbf{G} \rightarrow \mathbf{S}} = \max_{g \in \mathbf{G}} \{ \min_{s \in \mathbf{S}} d(g, s) \}, \quad (4)$$

where $d(g, s)$ is the Cartesian distance between two 3D points. The HD is asymmetric, as generally $\text{HD}_{\mathbf{G} \rightarrow \mathbf{S}} \neq \text{HD}_{\mathbf{S} \rightarrow \mathbf{G}}$. The symmetric BHD between these two sets is defined as

$$\text{HD}_{\mathbf{G} \leftrightarrow \mathbf{S}} = \max \{ \text{HD}_{\mathbf{G} \rightarrow \mathbf{S}}, \text{HD}_{\mathbf{S} \rightarrow \mathbf{G}} \}. \quad (5)$$

To decrease the sensitivity to outliers, the 95th-percentile BHD is used in this paper to measure the segmentation accuracy.



■ Ground truth (G)
■ Segmentation (S)

FIGURE 5: Schematic illustration for the calculation of the Hausdorff distance between the ground truth (green) and segmented (red) objects.

4. Experimental Results

Performance assessment of our framework is carried using dynamic CT kidney data, which were collected from 20 subjects. Each subject dataset consists of three 3D CT scans obtained at the pre- and postcontrast medium administration, namely, noncontrast, postcontrast, and late contrast 3D scan. The CT data were obtained using a GE light speed plus scanner (General Electric, Milwaukee, USA). The CT data acquisition parameters were 120 KV, 250 mA, in-plane resolution: $0.64 \times 0.64 \text{ mm}^2$, slice thickness: 0.9 mm, field-of-view (FOV): 360 mm, the 3D image sizes range from $512 \times 512 \times 232$ to $512 \times 512 \times 366$. In order to minimize the effect of interobserver variability, two experts delineated the kidney borders independently on the CT images and the ground truth labels were considered as the common segmented region of their delineations.

Quantitative evaluation is performed using a leave-one-subject-out approach and the number of decision trees was set to 400. First, all the 3D CT scans (60 scans in total) from all of the 20 subjects are coregistered using our registration methodology described in Section 2.2. To segment a test subject, all of its pre- and postcontrast scans are removed from the training database. Then, the 19 NCC-top-ranked scans are selected from the remaining training scans to build the test scan adaptive shape prior, described by (1) and the method in [39, 40]. Lastly, all regional features described in Sections 2.1 and 2.3 are extracted for (i) the NCC-selected scans to build the training model of the RF; and (ii) the 3D coregistered test scan to be classified using the built RF model. The above steps are repeated for all of the 60 CT volumes of the 20 subjects.

Cross-sectional segmentation results in the axial, sagittal, and coronal views using our technique are demonstrated in Figure 6 for CT data from four subjects at different contrast-enhancement phases. The 3D kidney surface is constructed

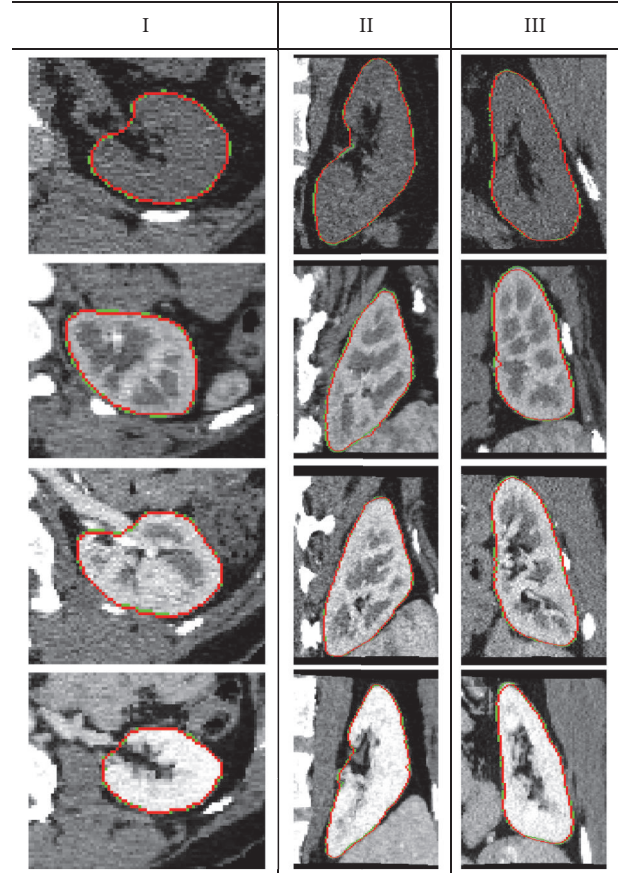


FIGURE 6: Cross-sectional axial (I), sagittal (II), and coronal (III) segmentation results of our approach for multiple subjects at different contrast-enhancement phases, showing reliable determination of kidney borders (red) compared with the ground truth (green) contours.

TABLE 1: Segmentation accuracy of our method compared with Zhang et al. [47] approach based on the DC, PVD, and BHD_{95} metrics. Note that DC, PVD, BHD_{95} , and SD stand for Dice coefficient, percentage volume difference, bidirectional 95th-percentile Hausdorff distance, and standard deviation, respectively.

Metric	Segmentation method		<i>p</i> value
	Our Mean \pm SD	Zhang et al. [47] Mean \pm SD	
DC (%)	97.27 ± 0.83	91.60 ± 2.29	$\leq 10^{-4}$
BHD_{95} (mm)	0.93 ± 0.49	5.36 ± 1.12	$\leq 10^{-4}$
PVD (%)	2.92 ± 2.21	5.00 ± 3.28	$\leq 10^{-4}$

by accounting for the object labels in the output of the RF classifier. Followed by a postprocessing step using a 3D median filter to smooth the noisy output labels of the classifier. The segmentation accuracy of our framework is assessed using the evaluation metrics described in Section 3. The overall accuracy for all subjects in terms of mean and standard deviation is summarized in Table 1.

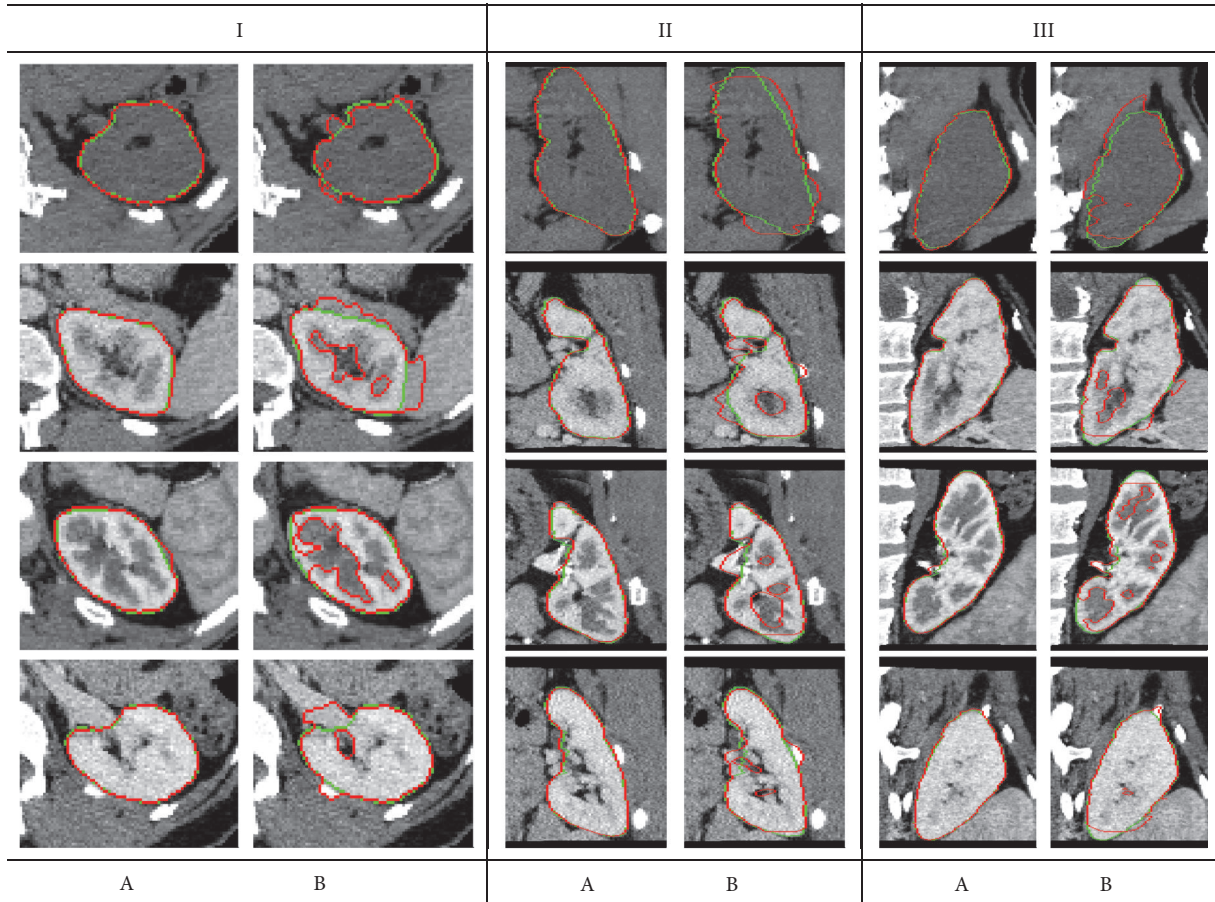


FIGURE 7: Cross-sectional axial (I), sagittal (II), and coronal (III) segmentation results from multiple subjects at different contrast-enhancement phases of our approach (A) and the approach proposed by Zhang et al. [47] (B). The red and green contours refer to model segmentation and the ground truth, respectively.

In order to demonstrate the high accuracy of our kidney segmentation framework, we compare it with the image segmentation method that was proposed by Zhang et al. [47], which has a freely available software package and thus avoids reimplementing an existing method. Figure 7 demonstrates sample segmentation results comparing our method versus the approach proposed in [47] on multiple subjects. The results in Figure 7 show reliable determination of the kidney borders of our technique compared to Zhang et al. [47] method. Additionally, a summary of the overall segmentation accuracy of our and Zhang et al. [47] methods, with respect to the ground truth delineation, for all data sets, is given in Table 1. According to the higher DC and lower HD_{95} and PVD values in Table 1, our technique performs notably better compared with [47]. This has been documented using the statistical significance of the statistical paired t -test as shown in Table 1 (p value is < 0.05).

In addition to the segmentation evaluation metrics described in Section 2.2, the robustness of our segmentation framework is assessed using the receiver operating characteristics (ROC) [48] as an alternate metric to evaluate the performance of segmentation systems. Generally, the ROC analysis assesses the sensitivity of a segmentation

method relative to the choice of its operating point (e.g., a classification threshold). This is achieved by plotting the relationship between the true positive and false positive rates for different operating points. Figure 8 shows the ROC curves of our method and Zhang et al. [47] approach. The figure clearly demonstrates that our technique attained higher performance compared with [47], as evidenced by the area under the ROC curve (AUC) of 0.96 compared with 0.92 for Zhang et al. approach [47].

5. Conclusions

In conclusion, a random forests-based framework is proposed for 3D kidney segmentation from dynamic contrast enhanced abdominal CT images. In order to account for large kidney inhomogeneity and nonlinear intensity variation of the dynamic CT data, our framework integrated two spatial-appearance features, namely, the higher-order spatial interactions features and appearance-based adaptive shape prior features, in addition to the Hounsfield appearance features. Qualitative and quantitative evaluation results confirmed reliable kidney tissue segmentation using our approach at different contrast-enhancement phases of agent transit. This

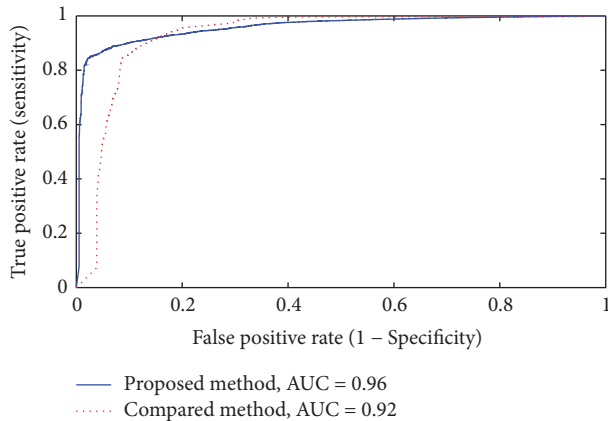


FIGURE 8: The ROC curves for our segmentation method and the method proposed in [47]. The “AUC” stand for the area under the curve.

has been evaluated on CT data sets collected from 20 subjects using both volumetric and distance-based evaluation metrics. In the future work we will investigate the addition of other features (e.g., scale space, local binary patterns). Also, we plan to test our framework on larger data sets to assess its accuracy, robustness, and limitation. Ultimately, we plan to include this segmentation approach into a kidney-dedicated CAD system for early detection of acute renal transplant rejection and treatment planning.

Competing Interests

The authors declare that they have no competing interests.

References

- [1] X. Chen, R. M. Summers, and J. Yao, “Automatic 3D kidney segmentation based on shape constrained GC-OAAM,” in *Proceedings of the Medical Imaging 2011: Image Processing*, Lake Buena Vista, Fla, USA, February 2011.
- [2] D. L. Pham, C. Xu, and J. L. Prince, “Current methods in medical image segmentation,” *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.
- [3] F. Khalifa, A. El-Baz, G. Gimel’farb, R. Ouseph, and M. A. El-Ghar, “Shape-appearance guided level-set deformable model for image segmentation,” in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR ’10)*, pp. 4581–4584, Istanbul, Turkey, August 2010.
- [4] E. Widjaja, J. W. Oxtoby, T. L. Hale, P. W. Jones, P. N. Harden, and I. W. McCall, “Ultrasound measured renal length versus low dose CT volume in predicting single kidney glomerular filtration rate,” *The British Journal of Radiology*, vol. 77, no. 921, pp. 759–764, 2004.
- [5] S.-W. Yoo, J.-S. Cho, S.-M. Noh, K.-S. Shin, and J.-W. Park, “Organ segmentation by comparing of gray value portion on abdominal CT image,” in *Proceedings of the 5th International Conference on Signal Processing*, pp. 1201–1208, Beijing, China, August, 2000.
- [6] S.-J. Kim, S.-W. Yoo, S.-H. Kim, J.-C. Kim, and J.-W. Park, “Segmentation of kidney without using contrast medium on abdominal CT image,” in *Proceedings of the 5th International Conference on Signal Processing*, vol. 2, pp. 1147–1152, Beijing, China, August 2000.
- [7] R. Pohle and K. Toennies, “A new approach for model-based adaptive region growing in medical image analysis,” in *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, pp. 238–246, Warsaw, Poland, September 2001.
- [8] R. Pohle and K. D. Toennies, “Self-learning model-based segmentation of medical images,” *Image Processing & Communications*, vol. 7, no. 3-4, pp. 97–113, 2001.
- [9] V. Leonardi, V. Vidal, J.-L. Mari, and M. Daniel, “3D reconstruction from CT-scan volume dataset application to kidney modeling,” in *Proceedings of the 27th Spring Conference on Computer Graphics (SCCG ’11)*, pp. 111–120, ACM, Viničné, Slovakia, April 2011.
- [10] B. Tsagaan, A. Shimizu, H. Kobatake, and K. Miyakawa, “An automated segmentation method of kidney using statistical information,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part I*, vol. 2488 of *Lecture Notes in Computer Science*, pp. 556–563, Springer, Berlin, Germany, 2002.
- [11] X. Wang, L. He, and W. Wee, “Deformable contour method: a constrained optimization approach,” *International Journal of Computer Vision*, vol. 59, no. 1, pp. 87–108, 2004.
- [12] J. Lu, J. Chen, J. Zhang, and W. Yang, “Segmentation of kidney using CV model and anatomy priors,” in *Proceedings of the Medical Imaging, Parallel Processing of Images, and Optimization Techniques (MIPPR ’07)*, Proceedings of SPIE, pp. 678–911, Wuhan, China, November 2007.
- [13] T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.
- [14] Y.-P. Huang, P.-C. Chung, C.-L. Huang, and C.-R. Huang, “Multiphase level set with multi dynamic shape models on kidney segmentation of CT image,” in *Proceedings of the IEEE Biomedical Circuits and Systems Conference (BioCAS ’09)*, pp. 141–144, Beijing, China, November 2009.
- [15] P. Campadelli, E. Casiraghi, and S. Pratisoli, “A segmentation framework for abdominal organs from CT scans,” *Artificial Intelligence in Medicine*, vol. 50, no. 1, pp. 3–11, 2010.
- [16] F. Khalifa, A. Elnakib, G. M. Beache et al., “3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function,” in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2011*, pp. 587–594, Springer, 2011.
- [17] F. Khalifa, G. Gimel’farb, M. Abo El-Ghar et al., “A new deformable model-based segmentation approach for accurate extraction of the kidney from abdominal CT images,” in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP ’11)*, pp. 3393–3396, September 2011.
- [18] M. Freiman, A. Kronman, S. J. Esses, L. Joskowicz, and J. Sosna, “Nonparametric iterative model constraint graph mincut for automatic kidney segmentation,” in *Proceedings of the 13th International Conference on Medical Image Computing and Computer-assisted Intervention: Part III (MICCAI ’10)*, pp. 73–80, Beijing, China, September 2010.
- [19] D.-T. Lin, C.-C. Lei, and S.-W. Hung, “Computer-aided kidney segmentation on abdominal CT images,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 59–65, 2006.

- [20] M. Spiegel, D. A. Hahn, V. Daum, J. Wasza, and J. Hornegger, "Segmentation of kidneys using a new active shape model generation technique based on non-rigid image registration," *Computerized Medical Imaging and Graphics*, vol. 33, no. 1, pp. 29–39, 2009.
- [21] X. Chen, J. K. Udupa, U. Bagci, Y. Zhuge, and J. Yao, "Medical image segmentation by combining graph cuts and oriented active appearance models," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2035–2046, 2012.
- [22] R. Cuingnet, R. Prevost, D. Lesage, L. D. Cohen, B. Mory, and R. Ardon, "Automatic detection and segmentation of kidneys in 3D CT images using random forests," *Medical image computing and computer-assisted intervention : MICCAI International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 15, no. 3, pp. 66–74, 2012.
- [23] H. Badakhshannoory and P. Saedi, "A model-based validation scheme for organ segmentation in CT scan volumes," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2681–2693, 2011.
- [24] U. Bagci, X. Chen, and J. K. Udupa, "Hierarchical scale-based multiobject recognition of 3-D anatomical structures," *IEEE Transactions on Medical Imaging*, vol. 31, no. 3, pp. 777–789, 2012.
- [25] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1723–1730, 2013.
- [26] T. Okada, M. G. Linguraru, M. Hori, R. M. Summers, N. Tomiyama, and Y. Sato, "Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors," *Medical Image Analysis*, vol. 26, no. 1, pp. 1–18, 2015.
- [27] G.-Y. Dai, Z.-C. Li, J. Gu, L. Wang, X.-M. Li, and Y.-Q. Xie, "Segmentation of kidneys from computed tomography using 3D fast growcut algorithm," *Applied Mechanics and Materials*, vol. 333–335, pp. 1145–1150, 2013.
- [28] E. Zhao, Y. Liang, and H. Fan, "Contextual information-aided kidney segmentation in CT sequences," *Optics Communications*, vol. 290, pp. 55–62, 2013.
- [29] C. Chu, M. Oda, T. Kitasaka et al., "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2013*, pp. 165–172, Springer, 2013.
- [30] G. Yang, J. Gu, Y. Chen et al., "Automatic kidney segmentation in CT images based on multi-atlas image registration," in *Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '14)*, pp. 5538–5541, August 2014.
- [31] J. Liu, M. G. Linguraru, S. Wang, and R. M. Summers, "Automatic segmentation of kidneys from non-contrast CT images using efficient belief propagation," in *SPIE Proceedings of the Medical Imaging: Computer-Aided Diagnosis*, vol. 8670, Orlando, Fla, USA, February 2013.
- [32] J. Liu, S. Wang, M. G. Linguraru, J. Yao, and R. M. Summers, "Computer-aided detection of exophytic renal lesions on non-contrast CT images," *Medical Image Analysis*, vol. 19, no. 1, pp. 15–29, 2015.
- [33] F. Khalifa, A. Soliman, A. C. Dwyer, G. Gimel'farb, and A. El-Baz, "A random forest-based framework for 3D kidney segmentation from dynamic contrast-enhanced CT images," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '16)*, pp. 3399–3403, IEEE, Phoenix, Ariz, USA, September 2016.
- [34] A. A. Farag, A. S. El-Baz, and G. Gimel'farb, "Precise segmentation of multimodal images," *IEEE Transactions on Image Processing*, vol. 15, no. 4, pp. 952–968, 2006.
- [35] D. Mahapatra, "Analyzing training information from random forests for improved image segmentation," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1504–1512, 2014.
- [36] T. K. Ho, "Random decision forests," in *Proceedings of the International Conference on Document Analysis and Recognition*, vol. 1, pp. 278–282, Montreal, Canada, August 1995.
- [37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [38] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, "How mantrees in a random forest?" in *Proceedings of the International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 154–168, 2012.
- [39] A. Soliman, F. Khalifa, A. Elnakib et al., "Accurate lungs segmentation on CT chest images by adaptive appearance-guided shape modeling," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 263–276, 2017.
- [40] F. Khalifa, A. Soliman, A. Takieldein et al., "Kidney segmentation from CT images using a 3D NMF-guided active contour model," in *Proceedings of the IEEE 13th International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '16)*, pp. 432–435, Prague, Czech Republic, April 2016.
- [41] F. Khalifa, G. M. Beache, G. Gimel'farb, J. S. Suri, and A. El-Baz, "State-of-the-art medical image registration methodologies: a survey," in *Handbook of Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, vol. 1, chapter 9, pp. 235–280, Springer, New York, NY, USA, 2011.
- [42] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, "Deformable medical image registration: setting the state of the art with discrete methods," *Annual Review of Biomedical Engineering*, vol. 13, pp. 219–244, 2011.
- [43] F. Y. Wu, "The Potts model," *Reviews of Modern Physics*, vol. 54, no. 1, pp. 235–268, 1982.
- [44] F. Khalifa, G. M. Beache, M. A. El-Ghar et al., "Dynamic contrast-enhanced MRI-based early detection of acute renal transplant rejection," *IEEE Transactions on Medical Imaging*, vol. 32, no. 10, pp. 1910–1927, 2013.
- [45] K. H. Zou, S. K. Warfield, A. Bharatha et al., "Statistical validation of image segmentation quality based on a spatial overlap index," *Academic Radiology*, vol. 11, no. 2, pp. 178–189, 2004.
- [46] G. Gerig, M. Jomier, and M. Chakos, "Valmet: a new validation tool for assessing and improving 3D object segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2001: 4th International Conference Utrecht, The Netherlands, October 14–17, 2001 Proceedings*, vol. 2208 of *Lecture Notes in Computer Science*, pp. 516–523, Springer, Berlin, Germany, 2001.
- [47] Y. Zhang, B. J. Matuszewski, L.-K. Shark, and C. J. Moore, "Medical image segmentation using new hybrid level-set method," in *Proceedings of the 5th International Conference BioMedical Visualization: Information Visualization in Medical and Biomedical Informatics (MediVis '08)*, pp. 71–76, London, UK, July 2008.
- [48] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.