Efficient Kidney Segmentation in Micro-CT Based on Multi-Atlas Registration and Random Forests

FENGJUN ZHAO[ID], PEI GAO, HAOWEN HU, XUELEI HE, YUQING HOU, AND XIAOWEI HE

School of Information Sciences and Technology, Northwest University, Xi'an 710069, China

Corresponding authors: Yuqing Hou (houyuqin@nwu.edu.cn) and Xiaowei He (hexw@nwu.edu.cn)

ABSTRACT Micro-computed tomography (micro-CT) provides an *in vivo* high-resolution preclinical imaging for murine kidneys. However, due to the relatively low dosage of X-rays, accurate and efficient segmentation of murine kidneys in micro-CT imaging remains challenging. In this paper, we proposed an efficient kidney segmentation method in micro-CT images based on multi-atlas registration (MAR) and random forests (RFs). First, we constructed a probability map of kidneys by the MAR and obtained an initial shape estimation of kidneys. We acquired the transformations of MAR based on high-contrast organs and then mapped kidneys to the new micro-CT images. Second, we extracted multiple features (including intensity, texture, and context features) from the voxels with lower probabilities and fed these features to a RF classifier. The role of RF is to fine-tune kidney boundaries after MAR. Finally, combining the initial shape with high probabilities and the fine-tuning of RF, we obtained the final segmentation of kidneys. The experiments were conducted on datasets acquired by micro-CT imaging of mice with and without the administration of contrast agent (Dataset1 and Dataset2). The results demonstrated the proposed MAR-RF outperformed the level sets, statistical-atlas registration, active shape model, and other supervised learning methods, with the Dice coefficients of 0.9766 and 0.9255, and the mean surface distances of 1.25 and 0.98 mm on Dataset1 and Dataset2, respectively. The training and prediction time of our MAR-RF were only 37.04% and 17.68% of the compared method, respectively. The proposed method has great potential for applications in other segmentation tasks of computer-aided diagnosis.

INDEX TERMS Kidney segmentation, computed tomography, multi-atlas registration, random forests.

I. INTRODUCTION

According to a survey conducted by the National Institute of Health in 2011, about 4.4 million adults in the United States suffer from renal diseases and the mortality is up to 50,476 [1]. Thus, precise diagnosis and treatment of these diseases are of great importance for improving the health conditions of the patients. In the past decade, state-of-the-art medical imaging techniques (such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US)) make it possible to investigate the anatomical information of kidneys *in vivo*. Specifically, micro-computed tomography (micro-CT) provides *in vivo* high-resolution preclinical imaging of murine kidneys. In this regard, accurate segmentation of kidneys is the critical step for the anatomical analysis. However, due to the relatively low dosage of X-rays,

the desirable delineation of renal shapes and boundaries in micro-CT imaging remains challenging.

Forouzan *et al.* [2] proposed a segmentation method for soft tissue organs based on multiple thresholds that were determined by analyzing the statistics of intensities. With spines as landmarks, Lin *et al.* [3] estimated the locations of kidneys with elliptical regions and then grew the regions to segment kidneys. However, due to the heterogeneous distribution of intensities, it is difficult to adaptively determine the thresholds and growing criteria for the above methods. Khalifa *et al.* proposed a level-set approach for kidney segmentation guided by a stochastic speed function. This method integrated both the shape prior knowledge and the spatial interaction model between the kidney voxels and its background [4]. Moreover, Wu and Sun [5] constructed

a probabilistic deformable model for kidney segmentation, which was initialized by Laws' micro-texture energies and maximum*a posteriori* (MAP) estimation. Generally, the model-based methods outperformed the thresholding and region growing techniques. Another popular method in medical image segmentation is the active shape model (ASM) [6], which constructs a statistically mean shape by iterative closest point (ICP) registration [7]. The specific shape variance for each object is determined by the standard deviation calculated by covariance matrix. Furthermore, Spiegel *et al.* and Wimmer *et al.* modified the ASM by introducing a non-rigid registration and a non-parametric framework, respectively [8], [9]. These modified methods improved the accuracy of ASM in terms of point correspondence and the adaption for kidney boundaries. Nevertheless, they are sensitive to the initialization and may not be plausible if the initial shapes are too far from the target shapes.

Many studies employed the atlas registration method to estimate organ locations from *in vivo* small animal imaging. Generally, the rigid/non-rigid registrations of high-contrast organs (such as skeleton, lungs and skin) were firstly conducted. Then, whole-body organs were mapped from an atlas to the newly scanned CT images, based on the correspondences established from the registration of high-contrast organs [10], [11]. Nevertheless, even if non-rigid transformations were adopted, a single atlas was not sufficient to compensate the inter-subject anatomical variations. Therefore, Wang *et al.* [12] proposed a statistical atlas-based approach to estimate major organs in mouse micro-CT images. They firstly registered the high-contrast organs in micro-CT images by fitting a statistical shape model, then estimated the low-contrast organs using a conditional Gaussian model. Subsequently, they modified the model of mouse atlas and make it adaptable for any body weight by changing body length and fat amount [13]. In addition, Alchatzidis *et al.* [14] established the consistency between multi-atlas registration and segmentation of target organs from the new image by introducing a pairwise Markov random field. These atlas-based methods can quickly estimate the locations of multiple organs, but cannot segment each organ accurately. Worse yet, the performance would greatly decrease when the quantity or quality of training images was insufficient.

Recently, more and more supervised learning methods were applied to the kidney segmentation, which reduce the segmentation to a problem of pixel classification by marking each pixel as kidneys or background. Cuingnet *et al.* [15] segmented kidneys by combining and refining random forests (RF) and template deformation. In their method, kidneys were localized with RF following a coarse-to-fine strategy. RF collaborating with other methods were also employed to further segment a kidney into four components (renal cortex, renal column, renal medulla, and renal pelvis) [1], [16]. Moreover, Gloger *et al.* [17] and Verma *et al.* [18] realized the volumetry and identification of renal diseases with support vector machines (SVM). The role of SVM was to automate

the procedure of conventional diagnosis of renal diseases with suitable feature extraction. Both RF and SVM require the handcrafted features that distinguish kidneys from background in CT images. It is worth mentioning that applications of deep neural networks (DNNs) in segmentation of polycystic or pathological kidneys emerged [19]–[21]. Compared with RF and SVM, DNNs directly took CT images as inputs instead of the handcrafted features, thus further reduced the user intervention during kidney segmentation. These supervised learning methods achieved relatively high segmentation accuracy. However, the computation costs for these methods are enormous, due to the time-consuming feature extraction in RF and SVM or the training of DNNs.

In this paper, we proposed an efficient kidney segmentation method in micro-CT images based on multi-atlas registration (MAR) and random forests (RF). Firstly, we constructed a probability map of kidneys with MAR and obtained an initial shape estimation of kidneys. Similar to previous methods, we acquired the transformations of MAR based on high-contrast organs and then mapped kidneys to the new micro-CT images. Then, we extracted multiple features (including intensity, texture, and context features) from the voxels with lower probabilities and fed these features to an RF classifier. The role of RF is to fine-tune the kidney boundaries after MAR. Combining the initial shape with high probabilities and the fine-tuning of RF, we obtained the final segmentation of kidneys. Our method balanced the computational cost and segmentation accuracy with the effective combination of MAR and RF. The main contributions of this paper were given as follows.

1) The effective combination took advantages of both the MAR and RF for segmenting murine kidneys in micro-CT images. The segmentation was speeded up without sacrificing the segmentation accuracy.

2) The image features we used in RF consisted of intensity, texture (Gray-level co-occurrence matrix, GLCM), and context features, which were able to distinguish kidneys from background. Although the micro-CT datasets used in this paper contained very large individual differences, the designed image features showed good discrimination capability.

3) We proposed an efficient feature extraction scheme in both the training and testing stage. In the training stage, we extracted the image features from the bounding boxes circumscribed the kidneys in training datasets. In the testing stage, only the image features in the regions with low probabilities (in probability map) were extracted.

4) The proposed method was applied to our micro-CT datasets with/without administration of contrast agent. The results showed that our method can accurately and efficiently. segment murine kidneys. The proposed method has great potential for applications in other segmentation tasks.

The rest of the paper was structured as follows. We described our method in Sections II, followed by the experiments and results in Section III. We discussed and concluded the paper in Section IV.
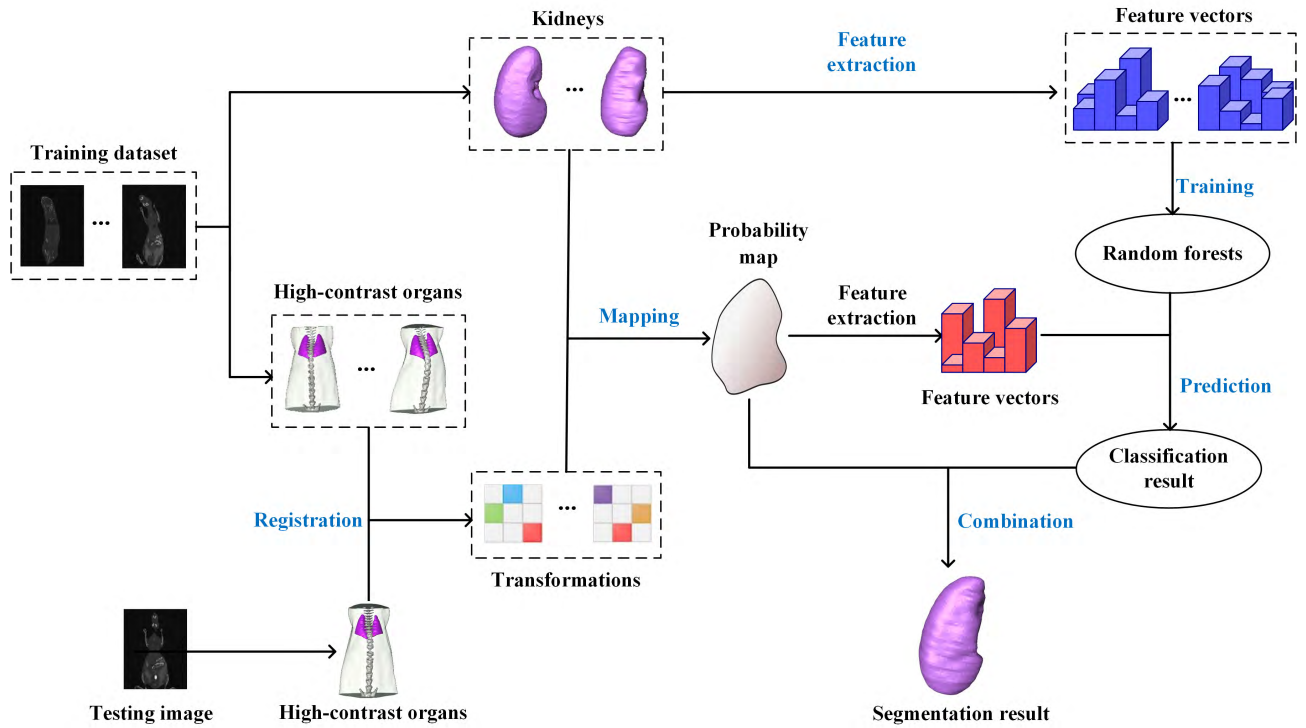
FIGURE 1. Framework of the proposed kidney segmentation method, which mainly consists of two stages: the acquisition of kidney probability map based on multi-atlas registration and kidney segmentation with random forests.

II. METHOD

The framework of the proposed method is shown in Fig. 1, which mainly consists of two stages, i.e. the acquisition of kidney probability map based on MAR and the kidney segmentation using RF. In the first stage, the high-contrast organs of the training atlases were registered one by one to the testing image, and thus the transformations are obtained. Subsequently, these transformations mapped the corresponding kidneys in the training dataset to the testing image. The probability map of the kidney was constructed based on the mapping results. In the second stage, multiple features (including intensity, texture, and context features) of training images were extracted to train an RF. The RF was applied to further classify the initial shape with low probabilities as kidneys or background. Finally, the probability map and the classification result were integrated to generate the final segmentation of the kidney.

A. ACQUISITION OF PROBABILITY MAP
1) ATLAS CONSTRUCTION

All the organs in the training dataset were interactively segmented under the supervised of experts. Multiple atlases in mouse trunk region were constructed, including high-contrast organs (skins, skeletons, and lungs) and kidneys. The high-contrast organs are easily identifiable in micro-CT images, which were labeled using a semi-automatic software incorporating thresholding, region growing, and deformable simplex mesh methods. As we all know, the shape representation

is fundamental for designing models and constructing atlas in medical image segmentation. In our method, the shape representation of the high-contrast organs was composed of a set of points distributed across the organ surface, which were extracted from labeled data using the marching cubes algorithm [22]. This is probably the simplest and the most generic method used to represent shapes [23]. The high-contrast organs were taken as the references in the atlas registration to overcome large variations in postures among individuals. To balance the registration time and the maintenance cost of organ shape, the numbers of representation points were 3400 for skin, 1250 for skeleton, and 1350 for lung. In order to obtain an accurate kidney probability map, the representation of kidneys in the training dataset was composed of the volumetric kidney voxels rather than the surface points.

Let X_i denote the ith training image, H_i denote the high-contrast organs in X_i

$$H_i = \left[X_i^{skin}; X_i^{skeloton}; X_i^{lung} \right] \tag{1}$$

where X_i^{skin}, $X_i^{skeloton}$, and X_i^{lung} denote the skin, skeleton, and lung in X_i, respectively.

$$X_i^{skin} = \left[x_{i,1}^{skin}, y_{i,1}^{skin}, z_{i,1}^{skin}; \cdots ; x_{i,N^{skin}}^{skin}, y_{i,N^{skin}}^{skin}, z_{i,N^{skin}}^{skin} \right] \tag{2}$$

with $\left(x_{i,j}^{skin}, y_{i,j}^{skin}, z_{i,j}^{skin} \right)$ being the 3D coordinates of the jth point of X_i^{skin}, and N^{skin} being the total number of points that represent the skin. Similar denotations also apply to

the skeleton, lung, and kidney. Therefore, H_i contains the 3D coordinates of $\left(N^{skin} + N^{skeleton} + N^{lung}\right) = 6000$ points.

2) ATLAS REGISTRATION AND PROBABILITY MAP CONSTRUCTION

Due to the variances in individuals, the size, shape, and location of the same organ generally vary among different images. To establish the point correspondence between each pair of atlas and the testing image, we firstly registered all the high-contrast organs in the training dataset to the ones in the testing image. Secondly, we calculated the transformation matrices from all the atlases to the testing image. Finally, we applied the transformation matrices to map the corresponding kidneys to the testing image to build a probability map of the kidney.

In detail, we denote H_o as the high-contrast organs in the testing image. The transformation T_i was obtained by registering H_i to H_o using the iterative closest point (ICP) algorithm [24]. Then T_i was used to map X_i^{kidney} to the testing image and produce X'^{kidney}_i. Similar to X_i^{kidney}, X'^{kidney}_i contained 3D coordinates of the points representing the kidney after mapping. We created a binary matrix M_i, which is of the same size as X_i. In M_i, the value of those positions covered by the points of X'^{kidney}_i is 1 and those of the other positions is 0. $\Psi = \left\{X'^{kidney}_1, \cdots, X'^{kidney}_{N^a}\right\}$ represents all the mapped kidneys, where N^a is the number of training images. $\Omega = \{M_1, \cdots, M_{N^a}\}$ was obtained when the registration traversed all the images in the training dataset. The probability map P of the kidney in the testing image was computed as

$$P = \frac{1}{N^a} \sum_{i=1}^{N^a} M_i \qquad (3)$$

Since each element of M_i is either 0 or 1, the value of each element of P is within [0, 1]. For $\forall p \subseteq P$, if $p = 1$ then the corresponding voxel was covered by all the elements of Ψ. On the contrary, $p = 0$ means that no element of Ψ covered the corresponding voxel. Hence, if the value of p was closer to 1, the corresponding voxel in P had a higher probability of being the target, and vice versa. The right inset of Fig. 5 shows the probability map. Darker color indicates the greater probability of the voxel being the kidney.

B. KIDNEY SEGMENTATION

The probability map describes the probability that each voxel in the testing image belongs to the kidney. We can set a probability threshold and obtain the segmentation result. However, experiments showed that the segmentation results were not accurate enough, as shown in Fig. 2. The 0.1-threshold causes the issue of over-segmentation, while the 0.9-threshold results in the under-segmentation with non-smooth contours. Obviously, such segmentation was far from meeting the clinical needs. Therefore, it was necessary to further fine-tune the segmentation based on the probability map. As shown in Fig. 3, we used an RF classifier to achieve this goal.

Firstly, we selected an upper threshold T_H and a lower threshold T_L. In the probability map, voxels with probabilities
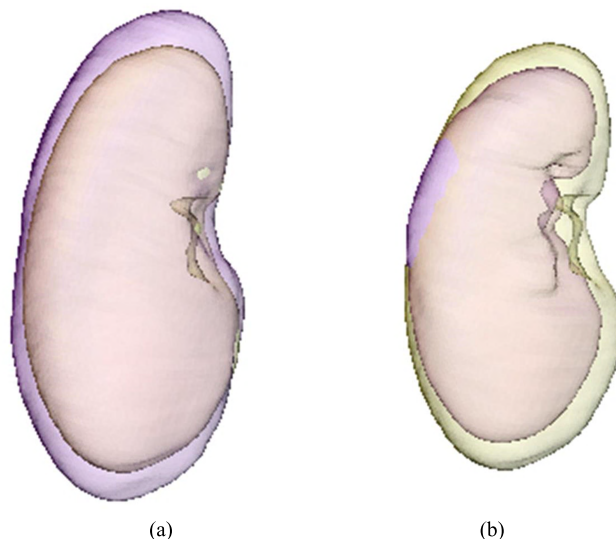

FIGURE 2. Threshold segmentation results based on the probability map, where purple represents algorithmic segmentation and yellow represents expert segmentation.
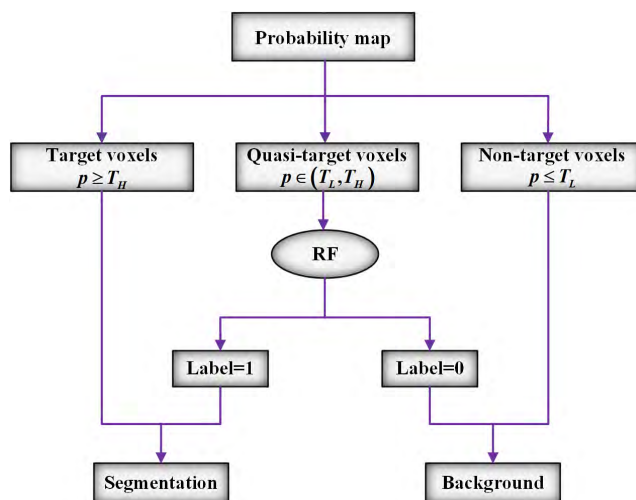

FIGURE 3. The combination MAR and RF for the kidney segmentation.

greater than T_H and smaller than T_L were directly marked as kidney voxels and background voxels, respectively. The voxels with probabilities between T_H and T_L were defined as the quasi-target voxels. Then, the trained RF classifier was applied to classify these quasi-target voxels as kidneys (label = 1) or background (label = 0). Finally, combining the voxels with high probabilities and the fine-tuning of RF, we obtained the final segmentation of the kidney. It is worth noting that the accuracy of the RF largely depended on the exquisite designed image features.

1) IMAGE FEATURES

Both 2D and 3D features were previously extracted to segment kidneys [1], [16]. The 2D features include hog features, Gabor features, Robert and Hessian features. The 3D features include mean and variance of voxels, which describe

different relationships between the voxels of interest and their surrounding ones. Moreover, Cuingnet et al. [15] segmented kidneys by an RF with intensity features and first/second order derivatives of the voxels. In our study, only the 2D features were extracted, which consisted of intensity features, texture features (gray-level co-occurrence matrices, GLCM), and context features. The intensity features and GLCM were calculated in $n_1 \times n_1$ patches centered at voxels of interest. The context features contained the average intensity, Gabor texture, and curvature. We combined these three types of features to segment kidneys in micro-CT images.

a: INTENSITY FEATURES

Studies have shown that the human visual system is sensitive to the mean and variance of image gray information [25]. However, in medical image analysis, some features that cannot be distinguished by human eyes may provide more distinguishing information, so we not only computed the mean and variance, but also the skewness and kurtosis [26]. Thus, each voxel had four intensity features, i.e. mean, variance, skewness, and kurtosis.

b: TEXTURE FEATURES

As one class of popular texture features, gray-level co-occurrence matrix (GLCM) has been widely used to evaluate image properties related to second-order statistics [27], [28]. Given the linear spatial relationship between two pixels, GLCM of an $n \times m$ image patch I_s can be generated by counting the co-occurrence gray levels as follows.

$$P(i, j) = \sum_{x=1}^{n} \sum_{y=1}^{m} \begin{cases} 1, & \text{if } I_s(x, y) = i \text{ and} \\ & I_s(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where (x, y) indicates the image coordinate. $P(i, j)$ describes the occurring frequency of two pixels constrained by the displacement vector $d = (\Delta x, \Delta y)$, one with grey-level i and the other with grey-level j. Obviously, the size of the GLCM depends on the range of values of the gray levels i and j, so it is very important to quantize the image to a certain gray level N^g before calculating the GLCM.

We calculated four most commonly used texture features from the GLCM as follows.

$$Contrast = \sum_{i,j} (i - j)^2 P(i, j) \quad (5)$$

$$Correlation = \sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) P(i, j)}{\sigma_i \sigma_j} \quad (6)$$

$$Energy = \sum_{i,j} P(i, j)^2 \quad (7)$$

$$Homogeneity = \sum_{i,j} \frac{P(i, j)}{1 + |i - j|} \quad (8)$$

where μ_i and σ_i, μ_j and σ_j are the means and standard deviations of $P_i = \sum_k P(i, k)$ and $P_j = \sum_k P(k, j)$, respectively. In order to obtain more effective texture information, different

GLCMs were calculated by setting different d. Δx and Δy are usually smaller in the practical applications. Empirically, we took four different values of d as $(0, 1)$, $(-1, 1)$, $(-1, 0)$, and $(-1, -1)$. Thus, we obtained 16 (4×4) texture features for each voxel in total.

c: CONTEXT FEATURES

Contextual information is of particular importance for medical image processing [29], which has been used to segment prostates from CT images [30] and cardiac structures from MRI [31], [32]. The context features in our paper consisted of average intensity, texture, and curvature. Fig. 4 shows the strategy for calculating context features. Centered at the current voxel, the intersections of the red lines and the three circles are the sampled points (24 in total, one indicated by '+'). At each sampling point, we extracted a $n_2 \times n_2$ patch (indicated by a square) and calculated the average intensity, texture, and curvature. The average texture was derived from the texture map of the patch. The average curvature was obtained based on the curvature map corresponding to the patch. The generation of texture map and curvature map was given as follows.
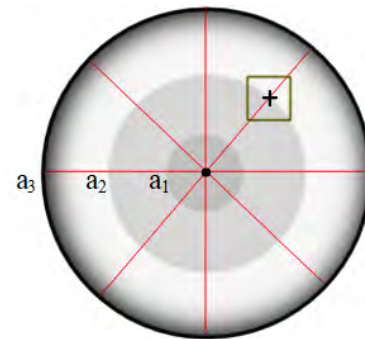
FIGURE 4. The strategy for calculating context features, where the angle between the adjacent red lines is 45°, and the radii of the three circles are a_1, a_2 and a_3, respectively.

The Gabor filter was employed to extract average texture features. Given the current image patch I, the texture map Tex_I can be calculated as

$$Tex_I = I \otimes G_{\gamma, \omega}(x, y) \quad (9)$$

where $\otimes$ denotes the 2D convolution, and $G_{\gamma, \omega}(x, y)$ is the Gabor filter function, which can be expressed as

$$G_{\gamma, \omega}(x, y) = \alpha \times g[\alpha(x \cos \theta + y \sin \theta),$$
$$\alpha(-x \sin \theta + y \cos \theta)] \quad (10)$$

where α is the scale factor, θ is the direction factor, and $g(x, y)$ is defined as

$$g(x, y) = \frac{\gamma}{2\pi \sigma^2} \times \exp\left[-\frac{x^2 + \gamma^2 y^2}{2\sigma^2} + 2\pi j(\lambda_x x + \lambda_y y)\right] \quad (11)$$

with γ being the spatial aspect ratio, which determines the ellipticity of the support of the Gabor function. σ is the

standard deviation of the Gaussian kernel. λ_x and λ_y are shifting parameters in the frequency domain. Empirically, we set $\alpha = 1$, $\theta = \pi/2$, $\gamma = 0.5$, $\sigma = 2$, and $\lambda_x = \lambda_y = 0.1$.

The curvature is another commonly used feature to distinguish different tissues [33]. In our paper, the curvature map Cur_I of I was calculated as

$$Cur_I = \frac{Cur_I^{up}}{Cur_I^{down}} \tag{12}$$

with

$$Cur_I^{up} = \frac{\partial^2 I}{\partial y^2}\left[1 + \left(\frac{\partial I}{\partial x}\right)^2\right]$$
$$-2\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\frac{\partial^2 I}{\partial x \partial y} + \frac{\partial^2 I}{\partial x^2}\left[1 + \left(\frac{\partial I}{\partial y}\right)^2\right] \tag{13}$$

and

$$Cur_I^{down} = 2 \times \left[1 + \left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2\right]^{\frac{3}{2}} \tag{14}$$

As the above description shows, each voxel corresponds to 72 (24×3) context features. Combining 4 intensity features and 16 texture features, we finally obtained a 92 ($4+16+72$) dimensional feature vector.

2) SUPERVISED LEARNING USING RANDOM FORESTS

As we all know, it is time-consuming to extract multiple features (i.e. the intensity, texture, and curvature features) over the whole volumetric training data. Moreover, large amount of features also increase the burden of RF training. Therefore, to make full use of training atlases, we proposed an efficient feature extraction strategy in both the training and testing stage of kidney segmentation with an RF classifier.

In the training stage, we extracted the image features from the bounding boxes of the kidneys in the training dataset. As shown in the left inset of Fig. 5, the smallest hexahedron circumscribed the kidney obtained from the corresponding atlas was regarded as the region of interest (ROI). The multiple features were only extracted within the ROI. In addition, a sampling strategy among training images was adopted. We assumed $\Gamma = \{F_1, \cdots, F_{N^a}\}$, with F_i denoting the set of the feature vectors in X_i. We uniformly selected feature vectors from each element of Γ to build a set F with less feature vectors used for RF training.

$$F = \left\{Sam\left(F_1, N^{F_1}/N^a\right), \cdots, Sam\left(F_{N^a}, N^{F_{N^a}}/N^a\right)\right\} \tag{15}$$

where $Sam(A, a)$ represents the operation of randomly selecting a elements from A, and N^{F_i} denotes the number of feature vectors in F_i. The sampling number is defined as N^{F_i}/N^a (where N^a is the number of training images) aiming at preserving the importance of the feature vectors from each training image, i.e., the feature vectors in F_i accounted for the same proportion in Γ and F.
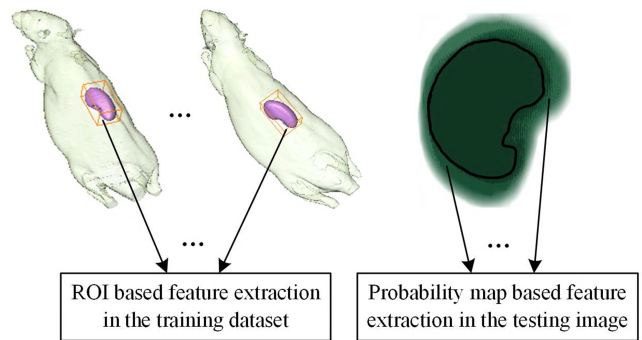
FIGURE 5. The feature extraction scheme. The left and the right insets show the feature extraction scheme in the training stage and testing stage, respectively.

In the testing stage, only the multiple features of the quasi-target voxels (whose probability within the interval (T_L, T_H)) were extracted (the right inset of Fig. 5). It was because the voxels with higher probability (larger than T_H) and the voxels with lower probability (smaller than T_L) were directly taken as kidneys and background, respectively. As a result, the number of voxels requiring feature extraction was reduced dramatically. This strategy undoubtedly speeded up the supervised learning and thus improved the efficiency of kidney segmentation.
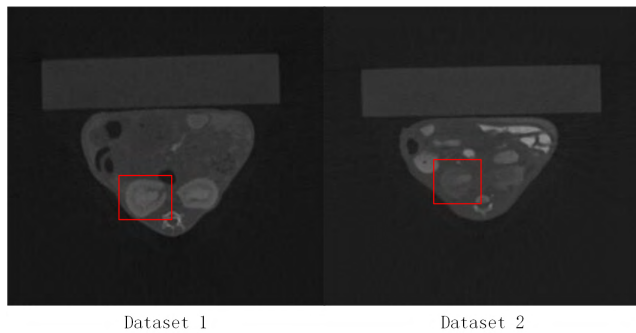
Considering both the accuracy and efficiency, random forest (RF) was selected as the supervised learning classifier. RF is an ensemble learning method for classification, which is composed of a plenty of decision trees and outputs the classification results by the voting of individual trees. A decision tree is a collection of nodes and leaves with each node being a weak classifier and each leaf storing *a posterior* probability distribution. Training an RF was finding the best parameters for the weak classifiers by optimizing each node with a random subspace of the feature space on a random subset of the whole training set. Then, the trained random forest was used to classify the quasi-target voxels as kidneys or background. It is noted that the number (N^t) and depth (D^t) of trees are the key indicators affecting the classification performance of an RF. In fact, they affect not only the accuracy of the segmentation, but also its efficiency. Finally, combining the voxels with high probabilities after MAR and the classification results of RF, we obtained the final segmentation of the kidney.

III. EXPERIMENTS AND RESULTS
A. DATASET AND EXPERIMENTAL ENVIRVONMENT
All animal procedures were in accordance with the Northwest University approved animal protocol. The mice were randomly divided into two groups with 6 and 35 mice for each group, respectively. The mice in the first group were imaged after the administration of the X-ray contrast agent, while the mice in the second group were directly imaged without contrast agent. The micro-CT system was comprised of an X-ray tube (HAMAMATSU L9181-02, Hamamatsu, Japan) with a minimum focal spot size of 5 μm, and a CCD

X-ray detector (Dexela 1512, PerkinElmer, MA) with a 1944×1536 pixel array and pixel size of 74.8 μm. The 3D micro-CT images were reconstructed with 3DMed software 4.0 (www.3dmed.net). After imaging of the two groups of mice, we obtained two datasets denoted as Dataset1 and Dataset2, respectively. The examples of the two datasets were shown in Fig. 6. It is obvious the kidney segmentation on the micro-CT images without contrast agent (Dataset2) is more challenging than that on images with contrast agent (Dataset1). The segmentation method was implemented in MATLAB 2016a. Computational experiments were conducted on a Workstation with Intel Core i5, 3.3GHz CPU, 8GB RAM under Windows 10. Multithreading technology was applied in the MAR and the feature extraction.

TABLE 1. The basic parameters of the proposed method (MAR-RF).

Parameter	Used at location	Value
$N^{skin}, N^{skeleton}, N^{lung}$	*Atlas Construction*	3400, 1250, 1350
T_H, T_L	*Kidney Segmentation*	0.9, 0.1
n_1	*Image Features*	15
N^g, d	*Texture Features*	256, [(0, 1), (-1, 1), (-1, 0), (-1, -1)]
α, θ	Eq. (10)	1, $\pi / 2$
$\gamma, \sigma, \lambda_x, \lambda_y$	Eq. (11)	0.5, 2, 0.1, 0.1
a_1, a_2, a_3, n_2	*Context Features*	3, 7, 10, 3

FIGURE 6. Examples of Dataset1 and Dataset2, respectively. The kidney (delineated with the red box) in Dataset2 has a more blurred boundary and a more uneven grayscale distribution than that in Dataset1.

B. BASIC PARAMETER SETTINGS AND EVALUATION

Since our method involves many parameters, Table 1 lists the basic parameters and their values which were tuned by trial-and-error. The remaining parameters are set separately or calculated by the corresponding formula. All relevant experiments in this article use this parameter setting. All the experimental results were obtained with the leave-one-out test. The segmentation accuracy was measured by comparing the segmentation with the manually segmented ground truth. Dice coefficient (Dice) and mean surface distance (D_{surf}) were used for the quantitative assessment. Their definitions are given as follows,

$$\text{Dice} = 2 \frac{|R_a \cap R_e|}{|R_a| + |R_e|}$$

$$D_{surf} = \underset{i}{\text{Mean}} \left(\underset{j}{\min} \left(V_a^i - V_e^j \right) \right)$$

where R_a and R_e represent the segmented kidney and ground truth, respectively. V_a and V_e represent the surface points of the segmented kidney and ground truth, respectively, with i and j as the point indices of the two surfaces. Dice coefficient better manifested the segmentation performance by measuring the overlapping between the segmented kidney and ground truth. Mean surface distance aimed to measure the

surface distance between the segmented kidney and ground truth.

C. EXPERIMENTAL RESULTS

We compared our proposed method with the distance regularized level set evolution (DRLSE) [34], active shape model (ASM) with non-rigid registration [8], statistical atlas registration (SAR) [12], as well as MAR with other classifiers, such as K-nearest neighbor (MAR-KNN) and linear support vector machine (MAR-LSVM). The average kidney shape was used to initialize the DRLSE, which saved the time spending on the slice-by-slice initialization. MAR-KNN and MAR-LSVM had the same segmentation process as MAR-RF. We tested MAR with these classifiers to evaluate the influence of different classifiers. The number of atlases for our proposed method was 5 ($N^a = 5$), and the number and depth of trees in RF were 20 ($N^t = 20$) and 10 ($D^t = 10$), respectively. For fair comparison, we adjusted the parameters of all the above methods and the optimal parameters were used in the comparative experiment.

The segmentation results of different methods are shown in Fig. 7. Due to the administration of X-ray contrast agent, all the segmentation methods but SAR performed well on Dataset1. It has little visual differences among the proposed method (MAR-RF), DRSLE, ASM, and the combinations of MAR with two other supervised learning methods. The SAR failed to segment kidney accurately on Dataset1 because of the great difference between the statistical shape and the target shape. For Dataset2 (without administration of X-ray contrast agent), the segmented kidneys by DRLSE, SAR, and ASM have obvious displacements compared to the ground truth. However, the results of the combinations of MAR with three supervised learning methods (including our proposed MAR-RF) still approximated the ground truth kidneys well.

Furthermore, we quantitatively validated the proposed method, and the results are shown in Fig. 8. The Dice coefficients of MAR-KNN, MAR-LSVM, and MAR-RF on Dataset1 are close and they are higher than that of DRSLE, SAR, and ASM (Fig. 8 (a)). However, the Dice coefficient of MAR-RF obviously exceeds that of other methods on
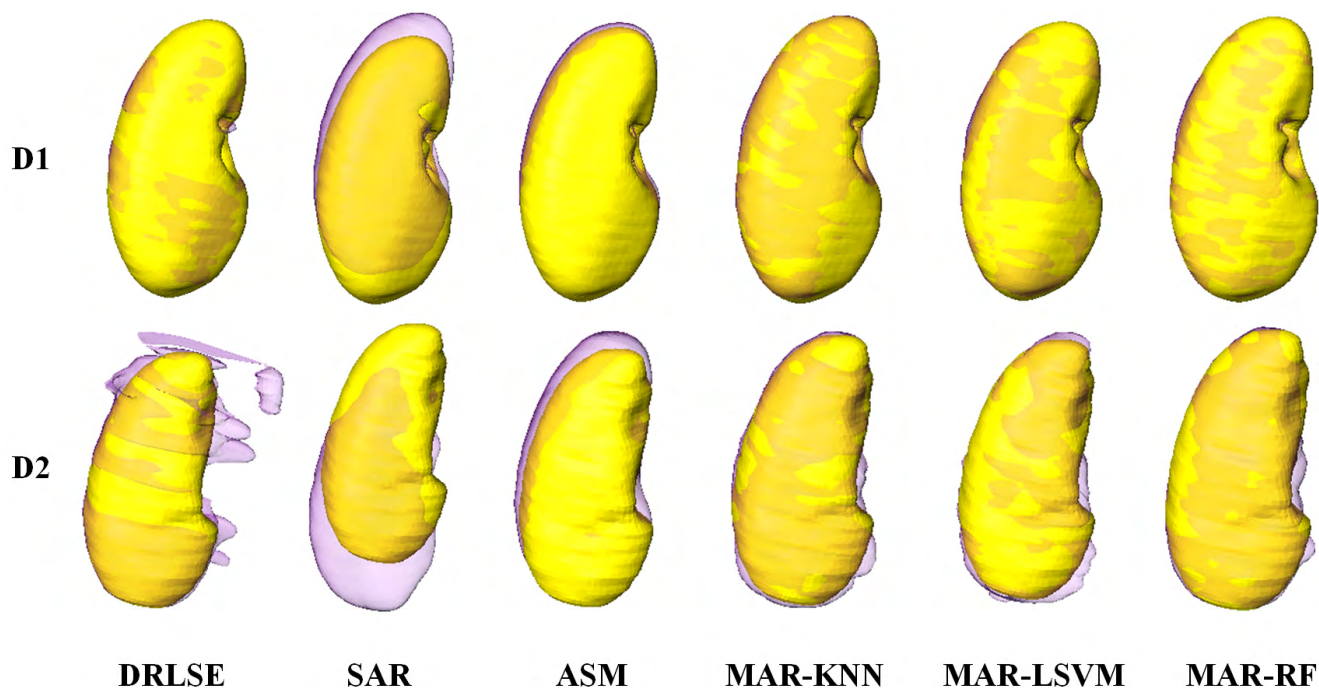
FIGURE 7. The kidney segmentation results with different methods on two datasets (Dataset1: D1, and Dataset2: D2). The insets from the left to right columns are the segmentation results with DRLSE, SAR, ASM, MAR-KNN, MAR-LSVM, and MAR-RF, respectively. Purple indicates the segmentation by each method, yellow indicates the ground truth, and orange indicates their overlap.

Dataset2. From Fig. 8 (b), it can be clearly seen that the mean surface distance of MAR-RF is the smallest on both datasets. The Dice coefficient of MAR-RF achieved of 0.9766 and 0.9255 on Dataset1 and Dataset2, meanwhile the mean surface distances reduced to 1.25mm and 0.98mm, respectively. The comparison demonstrates that the proposed method (MAR-RF) is superior to other methods for the murine kidney segmentation.

In general, the supervised learning method consumed more time due to the voxel-wise feature extraction compared with other methods. Nevertheless, we have speeded up the proposed method with our efficient feature extraction strategy in both the training and testing stage. Besides the combinations of MAR with three supervised learning methods, the supervised method (RF) without using the MAR in [26] was also implemented in our comparative experiment. The computation times of different methods are given in Table 2. It can be seen that both the training time and the prediction time of our proposed method (MAR-RF) are smaller than the RF without MAR, with the training time and prediction time of MAR-RF being only 37.04% and 17.68% of the latter. Since MAR-RF, MAR-KNN, and MAR-LSVM all took advantage of multi-atlas information and employed the same strategy during feature extraction, their computation time are relatively close to each other.

D. EVALUATION OF PARAMETERS ON MAR-RF
Among all the parameters, the number of atlases in MAR, the image features, and the number and depth of decision trees in RF are critical for the overall performance of the proposed

TABLE 2. The results of computation times for different kidney segmentation methods.

	MAR-RF	MAR-KNN	MAR-LSVM	RF
Training Time (min)	32.49 ± 0.25	32.96 ± 0.15	36.85 ± 0.33	88.50 ± 0.82
Prediction Time (min)	2.75 ± 1.12	3.55 ± 1.25	3.01 ± 1.40	15.55 ± 0.25

MAR-RF. Therefore, the influences of these parameters were evaluation based on a series of quantitative experiments.

1) THE COMPARISON OF DIFFERENT FEATURES
As mentioned earlier, image features are one of the most important factors in determining the classification accuracy of supervised learning (e.g. RF). To demonstrate the effectiveness of the proposed image features including intensity features, texture features, and context features, we conducted the comparative experiment with different image features. The results of kidney segmentation with different combinations of image features are shown in Fig. 9.

When the three types of features were individually used for the kidney segmentation, the context features ranked the first in terms of surface error. Among the combinations of two types of features, the combination of intensity and context features had the better segmentation accuracy. Nevertheless, our combination of all the three types of features
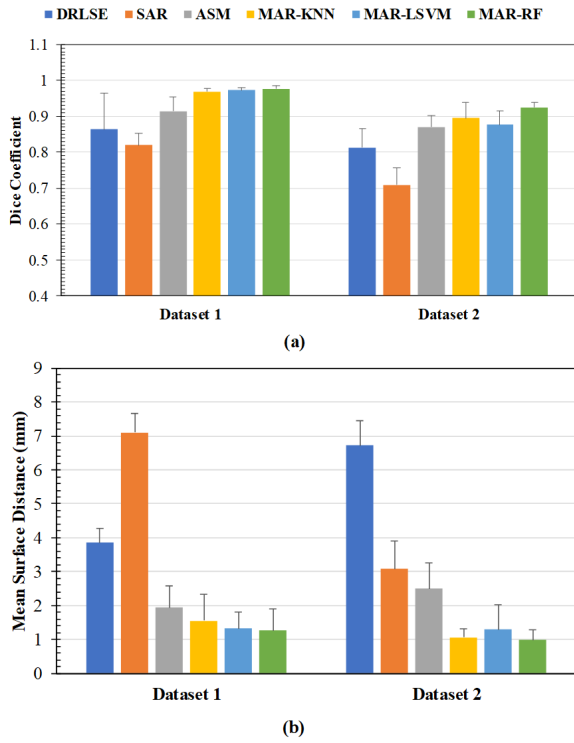
FIGURE 8. The quantitative comparison of accuracies of different methods with Dice coefficient and Mean surface distance.

(including intensity features, texture features, and context features) still ranks the first from visual inspection in Fig. 9.
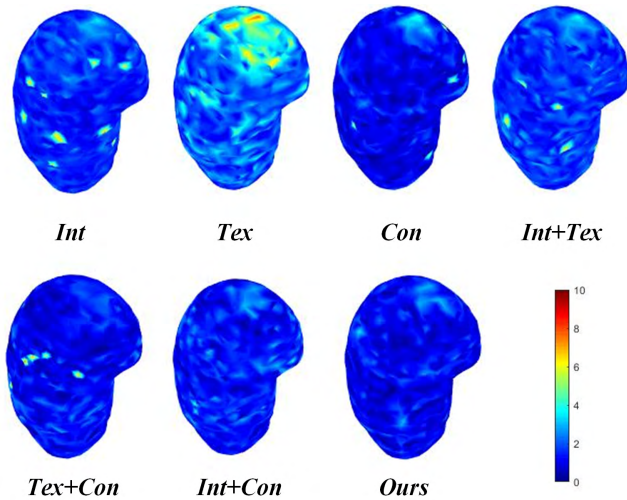

FIGURE 9. The surface errors caused by using different combinations of image features, including Intensity features (*Int*), texture features (*Tex*), and context features (*Con*). The closer the color is to blue, the closer the surface of segmented kidney is from that of the ground truth.

We also quantitatively evaluated the performance of our proposed image features with the other combinations of features. The values of Dice coefficient and mean surface distance for different features are shown in Fig. 10. Compared with the intensity and texture features, the context features
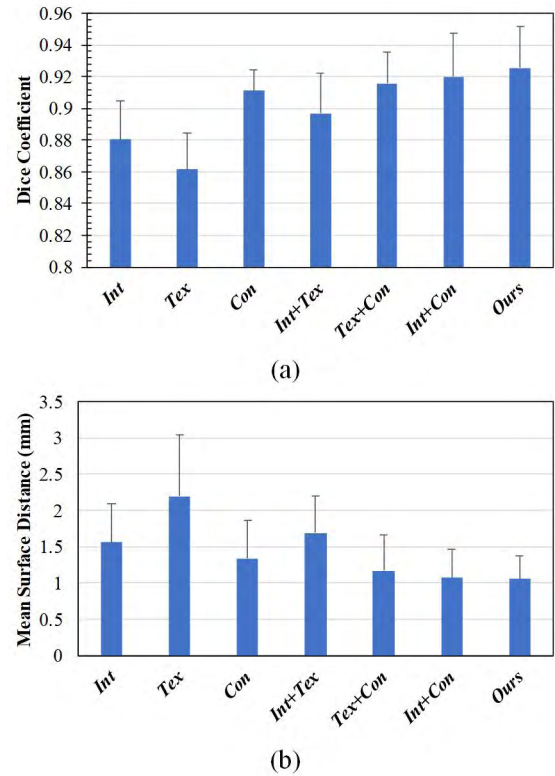

**FIGURE 10.** The effect of the combination of image features on the performance of MAR-RF.

have the highest Dice coefficient (Fig. 10 (a)) and lowest surface distance (Fig. 10 (b)). This is probably because the context features contain the information of local intensity, texture, and curvature, which fully exploit the spatial relationship of voxels. Nevertheless, it does not mean that other image features are unimportant. In fact, the combination of texture features with intensity or texture features has higher Dice coefficient and lower surface distance. Specifically, when combining all the three types of features, the Dice coefficient and surface distance achieved its highest and lowest values, respectively. Therefore, the effectiveness of our proposed features was quantitatively validated.

2) THE NUMBER OF ATLASES
The results of different numbers of atlases in our MAR-RF are shown in Fig. 11. When less than five atlases were used ($N^a < 5$), both the Dice coefficient and time consumption (T_c) increased as the number of atlases (N^a) increased. There is no significant increase in Dice when $N^a > 5$, while the T_c increases significantly. Therefore, the best trade-off between the accuracy and time consumption is achieved when $N^a = 5$.

3) THE NUMBER AND DEPTH OF TREES
Table 3 shows the variation of segmentation accuracy and training time (T_t) of MAR-RF with different numbers of trees in RF. When $N^t = 20$, the training time is T.
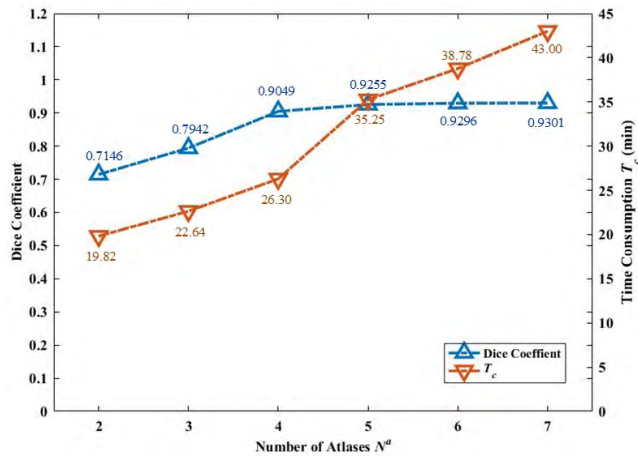
FIGURE 11. The effect of the different numbers of atlases on the performance of MAR-RF. The horizontal axis is the number of atlases and the vertical axes are the Dice coefficient (left) and time consumption (T_c) (right).

TABLE 3. The effect of number of trees in RF on segmentation accuracy and training time for MAR-RF.

N^t	5	10	15	20	50	100
Dice	0.8635	0.8906	0.9135	0.9255	0.9279	0.9298
T_t	0.31 T	0.54 T	0.77 T	T	2.41 T	4.69 T

Considering both the segmentation accuracy and time consumption, it can be concluded from Table 3 that the most appropriate value for N^t is 20. According to the same verification method, the best value of D^t is 10 (Table 4).

TABLE 4. The effect of depth of trees in RF on segmentation accuracy and training time for MAR-RF.

D^t	3	5	7	10	20	50
Dice	0.7535	0.8404	0.9002	0.9255	0.9230	0.9258
T_t	0.86 T	0.86 T	0.87 T	T	1.12 T	1.66 T

IV. DISCUSSION AND CONCLUSION

The proposed MAR-RF is a two-step, coarse-to-fine segmentation strategy. In the first step, multi-atlas registration (MAR) was used to compensate the individual differences in the morphology of the kidney. The coarse estimation of kidneys was obtained from the probability map. In the second step, the individual differences of kidneys were further fine-tuned by the random forest (RF) fed with the discriminative features. Therefore, MAR is different from SAR and ASM, which require more training images. This is why five atlases in our MAR-RF are enough to obtain a comparatively accurate segmentation results (Figs 7 and 8).

The registration method of multiple atlases was the ICP algorithm, which might be the simplest registration method for point cloud data. As a rigid registration method, ICP is suitable for tissues and organs without large morphologically alterations. Based on the results of MAR, we obtained a probability map providing a coarse estimation of kidney locations. Finer segmentation was performed by the random forest fed with multiple image features. In this sense, few mistakes were allowed when registering an atlas to the new image in our proposed segmentation method. Non-rigid registration methods might contribute to more accurate location estimation, but the time consumption would inevitably increase. Considering both the segmentation accuracy and time consumption, the ICP algorithm was used for atlas registration in this paper.

The texture features used in our MAR-RF were based on gray-level co-occurrence matrix (GLCM), with the quantization gray level (N^g) as one of the most important parameters. Smaller N^g corresponds to smoother image features with less computation time, while larger N^g covers more detailed image features at the expense of time. In our method, N^g still had a large value ($N^g = 256$, as shown in Table 1) for two reasons. Firstly, the texture feature extraction was based on small patches, which makes counting symbiotic gray levels with less time consumption. Secondly, thanks to our feature extraction scheme, the number of voxels requiring feature extraction is small. These two reasons can weaken the effect of a large N^g on the segmentation efficiency. Therefore, we set $N^g = 256$ to preserve more image details and ensured higher segmentation accuracy.

The sampling radii used to calculate the context features were set as $a_1 = 3$, $a_2 = 7$, $a_3 = 10$ (Fig.4). The number of sampling points on each circle was 8. These setting were obtained through trial-and-error based on specific segmentation target on specific datasets. In general, the sampling setting are tried under the following principles. Firstly, large radii may lead to a large amount of image information being incorporated in context features, which may fail to accurately describe the attributes of the target organ. Secondly, small radii corresponds to the context features with a large amount of redundant image information, such that the spatial relationship between voxels cannot be fully exploited. Finally, the number of sampling points on each circle is also a trade-off as the sampling radii. Taking both the distinguishability of different voxels and computation efficiency into account, we set the above sampling radii and number of sampling points.

As a conclusion, we proposed an efficient kidney segmentation method in micro-CT images based on multi-atlas registration (MAR) and random forests (RF). Firstly, we constructed a probability map of kidneys with MAR and obtained an initial shape estimation of kidneys. Secondly, we extracted image features from the voxels with lower probability and fed these features to an RF classifier. Combining the initial shape with high probabilities and fine-tuning of RF, we obtained the final segmentation of kidneys. The validation

experiments were conducted on micro-CT imaging of mice with and without administration of X-ray contrast agent (i.e. Dataset1 and Dataset2). The results demonstrated that our proposed method MAR-RF outperformed the other methods (such as DRLSE, SAR, and ASM), as well as other supervised learning methods (such as MAR-KNN and MAR-LSVM). Therefore, the accuracy and efficiency of our proposed method were validated. Our proposed method has great potential for applications in other segmentation tasks of computer-aided diagnosis.

REFERENCES

[1] C. Jin *et al.*, "3D fast automatic segmentation of kidney based on modified AAM and random forest," *IEEE Trans. Med. Imag.*, vol. 35, no. 6, pp. 1395–1407, Jun. 2016.

[2] A. H. Foruzan, R. A. Zoroofi, M. Hori, and Y. Sato, "Liver segmentation by intensity analysis and anatomical information in multi-slice CT images," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 4, no. 3, pp. 287–297, 2009.

[3] D.-T. Lin, C.-C. Lei, and S.-W. Hung, "Computer-aided kidney segmentation on abdominal CT images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 59–65, Jan. 2006.

[4] F. Khalifa *et al.*, "3D kidney segmentation from CT images using a level set approach guided by a novel stochastic speed function," in *Proc. Int. Conf. Med. Image Comput. Comput., Assist. Intervent. (MICCAI)*, vol. 6893, 2011, pp. 587–594.

[5] C.-H. Wu and Y.-N. Sun, "Segmentation of kidney from ultrasound B-mode images with texture-based classification," *Comput. Methods Programs Biomed.*, vol. 84, nos. 2–3, pp. 114–123, Dec. 2006.

[6] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 924–933, Aug. 2002.

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, 1995.

[8] 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," *Comput. Med. Imag. Graph.*, vol. 33, no. 1, pp. 29–39, Jan. 2009.

[9] A. Wimmer, G. Soza, and J. Hornegger, "A generic probabilistic active shape model for organ segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput., Assist. Intervent.*, vol. 5762, 2009, pp. 26–33.

[10] M. Baiker *et al.*, "Atlas-based whole-body segmentation of mice from low-contrast micro-CT data," *Med. Image Anal.*, vol. 14, no. 6, pp. 723–737, Dec. 2010.

[11] D. Xiao *et al.*, "Mouse whole-body organ mapping by non-rigid registration approach," *Proc. SPIE*, vol. 7965, p. 79650E, Mar. 2011.

[12] H. Wang, D. B. Stout, and A. F. Chatziioannou, "Estimation of mouse organ locations through registration of a statistical mouse atlas with micro-CT images," *IEEE Trans. Med. Imag.*, vol. 31, no. 1, pp. 88–102, Jan. 2012.

[13] H. Wang, D. B. Stout, and A. F. Chatziioannou, "A deformable atlas of the laboratory mouse," *Mol. Imag. Biol. MIB, Pub. Acad. Mol. Imag.*, vol. 17, no. 1, pp. 18–28, Feb. 2015.

[14] S. Alchatzidis, A. Sotiras, E. I. Zacharaki, and N. Paragios, "A discrete MRF framework for integrated multi-atlas registration and segmentation," *Int. J. Comput. Vis.*, vol. 121, no. 1, pp. 169–181, Jan. 2017.

[15] 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," in *Proc. Int. Conf. Med. Image Comput. Comput., Assist. Intervent.*, vol. 7512, 2012, pp. 66–74.

[16] C. Jin, F. Shi, D. Xiang, L. Zhang, and X. Chen, "Fast segmentation of kidney components using random forests and ferns," *Med. Phys.*, vol. 44, no. 12, pp. 6353–6363, Dec. 2017.

[17] O. Gloger, K. Tönnies, B. Mensel, and H. Völzke, "Fully automatized renal parenchyma volumetry using a support vector machine based recognition system for subject-specific probability map generation in native MR volume data," *Phys. Med. Biol.*, vol. 60, no. 22, pp. 8675–8693, Nov. 2015.

[18] J. Verma, M. Nath, P. Tripathi, and K. K. Saini, "Analysis and identification of kidney stone using Kth nearest neighbour (KNN) and support vector machine (SVM) classification techniques," *Pattern Recognit. Image Anal.*, vol. 27, no. 3, pp. 574–580, Jul. 2017.

[19] T. L. Kline *et al.*, "Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys," *J. Digit. Imag.*, vol. 30, no. 4, pp. 442–448, Aug. 2017.

[20] K. Sharma *et al.*, "Automatic segmentation of kidneys using deep learning for total kidney volume quantification in autosomal dominant polycystic kidney disease," *Sci. Rep.*, vol. 7, May 2017, Art. no. 2049.

[21] Y. Zheng, D. Liu, B. Georgescu, D. Xu, and D. Comaniciu, "Deep learning based automatic segmentation of pathological kidney in CT: Local versus global image context," in *Deep Learning and Convolutional Neural Networks for Medical Image Computing*, L. Lu, Y. Zheng, G. Carneiro, and L. Yang, Eds. Cham, Switzerland: Springer, 2017, pp. 241–255.

[22] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *ACM SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.

[23] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, 2009.

[24] P. J. Besl and D. N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 239–256, Feb. 1992.

[25] M. Petrou, V. A. Kovalev, and J. R. Reichenbach, "Three-dimensional nonlinear invisible boundary detection," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3020–3032, Oct. 2006.

[26] D. Mahapatra, "Semi-supervised learning and graph cuts for consensus based medical image segmentation," *Pattern Recognit.*, vol. 63, pp. 700–709, Mar. 2017.

[27] X. Ou, W. Pan, and P. Xiao, "*In vivo* skin capacitive imaging analysis by using grey level co-occurrence matrix (GLCM)," *Int. J. Pharmaceutics*, vol. 460, nos. 1–2, pp. 28–32, Jan. 2014.

[28] S. Beura, B. Majhi, and R. Dash, "Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer," *Neurocomputing*, vol. 154, pp. 1–14, Apr. 2015.

[29] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744–1757, Oct. 2010.

[30] W. Li, S. Liao, Q. Feng, W. Chen, and D. Shen, "Learning image context for segmentation of the prostate in CT-guided radiotherapy," *Phys. Med. Biol.*, vol. 57, no. 5, pp. 1283–1308, Mar. 2012.

[31] D. Mahapatra and Y. Sun, "Integrating segmentation information for improved MRF-based elastic image registration," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 170–183, Jan. 2012.

[32] D. Mahapatra, "Cardiac MRI segmentation using mutual context information from left and right ventricle," *J. Digit. Imag.*, vol. 26, no. 5, pp. 898–908, 2013.

[33] D. Mahapatra *et al.*, "Automatic detection and segmentation of Crohn's disease tissues from abdominal MRI," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2332–2347, Dec. 2013.

[34] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, Dec. 2010.

FENGJUN ZHAO received the B.S. degree in electronic engineering and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2010 and 2015, respectively.

Since 2015, he has been a Lecturer with the School of Information Science and Technology, Northwest University, Xi'an. He has published over 10 international journal articles and owns more over 10 inventions. His research interests include medical imaging segmentation, quantitative analysis of angiogenesis, coronary plaque detection, and computer-aid diagnosis of breast cancer.

PEI GAO received the B.S. degree in electronic information engineering from Northwest University, Xi'an, China, in 2015, where he is currently pursuing the M.S. degree in signal and information processing. He owns more over two inventions. His current research interests include the medical images segmentation based on statistical atlas and machine learning methods.

HAOWEN HU received the B.S. degree in electronic information science and technology from Northwest University, Xi'an, China, in 2016, where he is currently pursuing the M.S. degree in signal and information processing. He has published one paper indexed by EI. His current research interests include organ segmentation in small animals and human heart extraction from CTA imaging based on active shape model, random forests, and convolutional neural networks.
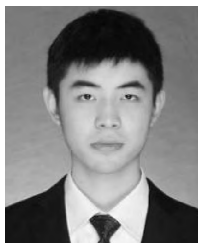
XUELEI HE received the B.S. degree in electronic and information engineering from Wuhan University, Wuhan, China, in 2015. He is currently pursuing the Ph.D. degree in computer application technology with Northwest University, Xi'an, China.

He has published one paper on the *Journal of Biomedical Optics*, and he has also applied one Chinese invention patent. His current research interests include the development of system and algorithm in bioluminescence tomography, fluorescence molecular tomography, and X-ray luminescence tomography.

YUQING HOU received the B.S. degree from Northwest University, Xi'an, China, in 1984, and the M.S. degree from the Xi'an Institute of Optics and Precision Mechanics, CAS, Xi'an, in 1990. She is currently a Professor with the School of Information Sciences and Technology, Northwest University.

From 1993 to 2002, she was a Government-Sponsored Student with Doshisha University, Japan. From 2009 to 2010, she was a Government-Sponsored Visiting Scholar with the University of Louisville, USA. She has published over 20 international journal papers, and owns over 10 Chinese invention patents. Her current research interests include digital image processing, big data in healthcare and radiomics, platform development of big data system, and big data application solution. She is a member of the DSP Application Expert Committee of the Chinese Electronic Society.

XIAOWEI HE received the M.S. degree from the School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China, in 2005, and the Ph.D. degree in pattern recognition and intelligent system from the School of Life Sciences and Technology, Xidian University, Xi'an, in 2011.

Since 2016, he has been a Professor with the School of Information Sciences and Technology, Northwest University, Xi'an. He has published over 50 international journal papers, and owns over 20 Chinese invention patents. His current research interests include medical imaging processing, 3-D molecular imaging, and artificial intelligence.

Dr. He is a member of Youth Committee of the Biomedical Photonics Commission, Chinese Optical Society. He is also a Council Member of the Shaanxi Graphics and Graphics Society and the Shaanxi Signal Processing Society.

• • •