Performance of an Automated Segmentation Algorithm for 3D MR Renography

Henry Rusinek,1* Yuri Boykov,2 Manmeen Kaur,1 Samson Wong,1 Louisa Bokacheva,1 Jan B. Sajous,1 Ambrose J. Huang,3 Samantha Heller,1 and Vivian S. Lee1

The accuracy and precision of an automated graph-cuts (GC) segmentation technique for dynamic contrast-enhanced (DCE) 3D MR renography (MRR) was analyzed using 18 simulated and 22 clinical datasets. For clinical data, the error was 7.2 ± 6.1 cm³ for the cortex and 6.5 ± 4.6 cm³ for the medulla. The precision of segmentation was 7.1 ± 4.2 cm³ for the cortex and 7.2 ± 2.4 cm³ for the medulla. Compartmental modeling of kidney function in 22 kidneys yielded a renal plasma flow (RPF) error of 7.5% ± 4.5% and single-kidney GFR error of 13.5% ± 8.8%. The precision was 9.7% ± 6.4% for RPF and 14.8% ± 11.9% for GFR. It took 21 min to segment one kidney using GC, compared to 2.5 hr for manual segmentation. The accuracy and precision in RPF and GFR appear acceptable for clinical use. With expedited image processing, DCE 3D MRR has the potential to expand our knowledge of renal function in individual kidneys and to help diagnose renal insufficiency in a safe and noninvasive manner. Magn Reson Med 57:1159–1167, 2007. © 2007 Wiley-Liss, Inc.

Key words: image analysis; MR renography; magnetic resonance imaging; renal function; error analysis

One technique to determine renal function consists of the intravenous injection of radioactive tracer followed by assessment of its plasma clearance 2–4 hr later. This technique is time-consuming, requires multiple blood samples, and measures only the global glomerular filtration rate (GFR)—a disadvantage when asymmetric or unilateral renal disease is present. The gold standard technique for assessing single-kidney GFR is inulin clearance, but this procedure is too invasive and complex for routine clinical assessment. Manual segmentation of MRR images into cortical and medullary volumes is time-consuming, requires multiple blood samples, and measures only the global glomerular filtration rate (GFR)—a disadvantage when asymmetric or unilateral renal disease is present. The gold standard technique for assessing single-kidney GFR is inulin clearance, but this method is too invasive and complex for routine clinical application. Accurate segmentation of contrast-enhanced MR images remains a difficult task. Dynamic 3D MR images of the abdomen suffer from partial-volume and respiratory-motion artifacts and have a relatively low signal-to-noise ratio (SNR). Other sources of error include signal nonuniformity and wraparound artifacts. The presence of cysts and renal masses, and reduced corticomedullary differentiation in many subjects compound the difficulties.

Manual segmentation of MRR images into cortical and medullary regions is a laborious and time-consuming task. In order to improve the clinical utility of MRR, we developed a semiautomated segmentation technique based on the graph-cuts (GC) (11,12). The purpose of the present study was to estimate the accuracy, precision, and efficiency of our method. Using both simulated MRR data and human data segmented manually by a consensus of expert observers, we evaluated the performance of the GC segmentation technique in terms of its ability to measure 1) cortical and medullary volumes, 2) corresponding CTAs, and 3) functional parameters (RPF and GFR).

MATERIALS AND METHODS

We performed two sets of experiments to evaluate the GC segmentation technique: one using simulated MRR data derived from measured human data, and one using actual measured human-subject data.

Subjects

To generate simulated MRR data, we used images from six healthy volunteers (two women and four men, age = 39.2 ± 10.1 years) and eight patients (three women and
five men, age = 67.9 ± 8.4 years) with renal insufficiency manifested by serum creatinine ≥ 2 mg/dL (average = 2.9 ± 1.2 mg/dL). The method used to generate the simulated datasets is described below.

For the human-subject dataset we examined 11 consecutive MRR studies (22 kidneys, six women and five men, age = 29–90 years, mean = 60.5 ± 18.5 years). There was no overlap between these 11 subjects and the 14 subjects used to generate simulated data. Six subjects were patients who had been referred for suspected renovascular disease, three were evaluated as candidates for renal transplant donation, and two were healthy volunteers. The patients had underlying diagnoses of either chronic renal failure or hypertension. Serum creatinine levels ranged from 0.6 to 2.9 mg/dL.

Written informed consent was obtained from all subjects, and the protocol was approved by our institutional review board.

MRR Acquisition Protocol

Dynamic MRI was performed on a 1.5T system (Avanto; Siemens, Erlangen, Germany) with a maximum slew rate of 200 T/m/s, maximum gradient strength of 45 mT/m, and a torso phased-array coil. 3D T1-weighted spoiled gradient-echo imaging was performed in the oblique coronal orientation to include the abdominal aorta and both kidneys. The following parameters were used: TR = 2.8 ms, TE = 1.1 ms, flip angle = 12°, matrix = 161 × 256 × 20, FOV = 425 × 425 × 100 mm, bandwidth = 650 Hz/voxel, volume acquisition time = 3 s. The original 5-mm coronal partitions were interpolated to 40 2.5-mm slices.

Five unenhanced acquisitions were performed during a single breath-hold. A 4-ml bolus of Gd-DTPA (Magnevist; Berlex Laboratories, Wayne, NJ, USA) was then injected, followed by 20 ml of saline, both at 2 ml/s. Over 20 min, 36 3D datasets were acquired using a variable sampling schedule: 10 sets acquired every 3 s, followed by four sets every 15 s, followed by seven sets every 30 s, and ending with 15 sets every minute. We attempted to acquire the first 10 sets in one breath-hold. Before each subsequent acquisition the patients were instructed to suspend respiration at end-expiration. Oxygen via nasal cannula was routinely offered to the patients before the exam to facilitate breath-holding. For image processing, all 41 3D datasets (five acquired before and 36 acquired after injection) were evaluated.

Simulated MRR Datasets

To understand the bias involved in the use of the GC segmentation method, we simulated 18 renal 4D datasets. To generate the simulated datasets, we first generated 14 representative sets of cortical $C_\text{X}(t)$, medullary $Med_\text{X}(t)$, arterial $A(t)$, and collecting system $P(t)$ signal intensity curves by manually sampling these regions in six normal subjects ($i = 1 \ldots 6$) and eight patients ($i = 7 \ldots 14$), as described above. The enhancement curves were temporally aligned to the same arterial peak time and averaged separately for the normal controls and the patients, yielding six ideal curves for the cortex ($C_\text{X}$), medulla ($Med_\text{X}$), and renal pelvis ($P$). Three curves ($C_\text{X}(t)$, $Med_\text{X}(t)$, and $P(t)$) represented normal function, and three ($C_{\text{X}}(t)$, $Med_{\text{X}}(t)$, and $P_{\text{ab}}(t)$) represented renal dysfunction.

Next, two representative MRR datasets (one from a normal subject and one from a patient with renal insufficiency) were selected and the left and right kidneys were cropped to $96 \times 128 \times 32$ volumes. The resulting four 4D datasets were segmented into the background, cortex, medulla, and collecting system space using procedure described below. In the kidneys from the normal subject we replaced the signal intensity at each time point $t$ in all cortical voxels with $C_{\text{X}}(t)$, in all medullary voxels with $Med_{\text{X}}(t)$, and in all collecting system voxels with $P(t)$. All other voxels remained as originally acquired (Fig. 1). In the diseased kidneys from the patient we substituted $C_{\text{X}}(t)$ for cortical voxels, $Med_{\text{X}}(t)$ for medullary voxels, and $P_{\text{ab}}(t)$ for voxels representing the collecting system.

Each of these four ideal datasets was then transformed into nine more realistic sets by selective blurring and the addition of random noise (Fig. 2). Blurring the kidney portion of the ideal image was achieved by convolution with a discrete Gaussian kernel. The width of this isotropic 3D kernel was characterized by a parameter $\sigma$. In addition, pseudo-random, spatially uncorrelated noise was added with variable magnitude $n$, expressed as a
percentage of the precontrast cortical signal. Simulated 4D datasets were generated corresponding to $\sigma = 1.5$ mm, $n = 20\%$ (low blur, high noise); middle column: $\sigma = 1.5$ mm, $n = 10\%$; right column: $\sigma = 3$ mm, $n = 10\%$ (high blur, low noise). A mid-coronal slice is shown before contrast injection (top row) and at the time of maximum corticomedullary contrast (bottom row).

Three readers (see below) used the GC software to segment the 36 simulated datasets. The learning bias was minimized by starting the segmentations with the poorest-quality images and proceeding in order of increased spatial resolution and decreased noise level.

Spatial Alignment of the Left and Right Kidneys
Prior to image segmentation, a registration algorithm was applied to separately align the left and right kidneys across multiple 3D acquisitions performed during repeated breath-holds. First, a locally developed program was used to remove patient-identifying header information and convert the original DICOM images into a format consisting of

FIG. 2. Example of a simulated normal kidney at different levels of noise and blur. Left column: $\sigma = 1.5$ mm, $n = 20\%$ (low blur, high noise); middle column: $\sigma = 1.5$ mm, $n = 10\%$; right column: $\sigma = 3$ mm, $n = 10\%$ (high blur, low noise). A mid-coronal slice is shown before contrast injection (top row) and at the time of maximum corticomedullary contrast (bottom row).

FIG. 3. Initial preprocessing involves cropping of the left and right kidneys. From acquired 4D MRR data (left column) the user selects two smaller subvolumes that encompass the right kidney (middle column) and the left kidney (last column). Top row: Sample image acquired just before injection of contrast. Bottom row: The same mid-coronal slice, acquired during the arterial phase, shows improved contrast between the renal cortex and medulla.
a text header file (containing the matrix size, voxel dimensions, and timing information) and the “raw” image file in which signal intensities were stored as 16-bit integers. The user then specified two separate subvolumes of $96 \times 128 \times 32$ voxels each to encompass the right and left kidneys (Fig. 3). To compensate for respiratory motion, the subvolumes for each of the subsequent 40 time points were spatially aligned with the first time point using the mutual information coregistration method (13,14). The transformations used in this process were constrained to translations in three directions. All subsequent processing was applied to the coregistered dynamic datasets.

Manual Segmentation–Reference Standard

To generate a reference standard for the human-subject MRR data, two experienced individuals collaborated to manually segment each of the 22 kidneys into cortical and medullary regions. These observers used an interactive, locally developed Unix-based software package that provides the user with the ability to view and scroll through the 3D data and construct volumes of interest (VOIs) in arbitrary planes (15). The software provides the user with an electronic paintbrush and eraser tools, and the ability to overlay user-defined masks on arbitrarily selected volumes acquired at different times. The initial VOI, $M_{\text{kid}}$, was created for the entire kidney. Renal cysts (visible on 6/22 kidneys) were excluded from $M_{\text{kid}}$. Using an arterial phase dataset that showed good corticomedullary differentiation, a second VOI, $M_{\text{cx}}$, was traced for the cortex. A third VOI, $M_P$, was constructed to include the renal pelvis, collecting system, and intrarenal fat. $M_P$ was traced on either precontrast (fat) or late excretory (collecting system) volumes. The renal medulla was then generated using a mask-combining tool as: $M_{\text{med}} = M_{\text{kid}} - (M_{\text{cx}} \cup M_P)$ (here $\cup$ is the set union, and $-$ is the set difference).

The manual segmentation took on the average of 2.5 hr per kidney. In spite of its labor-intensive nature, considerable effort was made to provide accurate ground reference masks. The two readers jointly segmented three representative kidneys (normal, cystic, and atrophic case) in order to identify potential problems. Subsequently, one investigator segmented the remaining kidneys. The results were then examined by the second reader who noted areas of disagreements and introduced corrections. Finally, in a joint session the corrected masks were reviewed to resolve areas of disagreement.

![FIG. 4. Based on the seed regions painted by the observer (solid red = kidney seed, solid blue = background seeds), the GC algorithm computes the binary partition of the volume into the kidney mask (translucent red) and background (translucent blue). After viewing the result, the user may place additional seeds to improve the segmentation.](image)

![FIG. 5. a: By operating within the kidney mask and using primarily images from the vascular phase (which provide good corticomedullary differentiation), the operator segments the cortex. b: Seeds are placed over cortical tissue (solid red), and background (solid blue) seeds are placed over the medulla and collecting system. c: The results of the GC algorithm are shown to the observer as translucent red and blue overlays. After viewing the cortical mask (translucent red), the user may place additional red and blue seeds to improve the segmentation. d: A similar process is applied in the third step to separate the collecting system from the medulla using primarily the image from a late excretory phase.](image)
For the simulated dataset analysis, the reference standard consisted of the predefined regions of interest (ROIs) used to generate the datasets.

### Segmentation Using GC

In the GC approach, 3D voxels are represented as nodes of a graph (11,12). The neighboring voxels \( p = (p_x, p_y, p_z) \) and \( q = (q_x, q_y, q_z) \) are linked using directed edges \( e(p,q) \) that are assigned cost based on combination of two quantities: 1) the similarity of their signal intensity \( S \), expressed as a monotonic function of \( |S(p) - S(q)| \), and 2) the angle between \( e(p,q) \) and the vector of local image gradient. Using a mouse-controlled paintbrush, the operator places the seeds on two regions \( O' \) and \( B' \) that represent the object and background. The algorithm then considers partitions of the image in two disjoint subsets, the object \( O \) and the background \( B \), with \( O' \subset O \) and \( B' \subset B \), so that the “cost” of this boundary between \( O \) and \( B \) reaches the minimum value. The formula used to express the boundary’s cost includes 1) a measure of its surface area, and 2) the total cost of all severed edges from nodes in \( O \) to nodes in \( B \). A combinatorial optimization algorithm is used to globally minimize the cost function (16). For our cropped 96 × 128 × 32 data the algorithm takes approximately 4 s to calculate the optimum graph.

The GC algorithm is applied in three phases. First, the user separates the kidney from the background that includes fatty tissue in the peripelvic region (Fig. 4). Next, operating within the kidney mask, the user separates the cortex from the rest of the kidney (Fig. 5). Finally, the collecting system is separated from the medulla. The software was implemented to provide users with the means to improve the segmentation result in each phase by planting additional seed voxels in either the object or the background, and iterating the GC algorithm based on revised constraints. Thus, with a sufficient number of iterations, the user can refine the resulting masks (as with the manual approach) while avoiding the need to “paint” each voxel. On average, each renal segmentation phase took three iterations, each averaging five paintbrush strokes to place the seeds on diverse (coronal, axial, and sagittal) sections through the kidney.

In addition to planting representative seeds, at each segmentation phase the cost of edges is tuned to bias the result towards a boundary with properties that are known a priori. For example, in the first phase, which is based on precontrast images (Fig. 4), the edge cost is assigned to favor finding of a dark object within a brighter background. This is reversed in the second phase, when we bias the minimum-cost cut towards the boundary between the brighter cortex and darker medulla (Fig. 5a).

Our implementation of GC was written in C++ (Visual Studio version 2003; Microsoft, Redmond, WA, USA) using max-flow/min-cut software library (17). The program was run on a personal computer with a 2.4 GHz Intel Pentium 4 processor and 1 GB of RAM.

Following three training sessions, three independent readers segmented each of the 22 dynamic datasets. The readers had respectively several weeks, 2 years, and 7 years of experience in viewing renal MR images. Each user segmented each kidney, saved the cortical and medullary masks, and recorded the total time spent performing the segmentation.

### Deriving Renal Physiological Parameters From MRR

Measured signal intensities \( Cx(t) \) and \( Med(t) \) were first converted to Gd-DTPA concentrations using a two-step process (18). Briefly, the conversion is based on a sequence-specific calibration curve \( f \) that relates measured signal to longitudinal relaxation times \( T \):

\[
S = g(f(T)). \quad [1]
\]

The function \( f \) is determined empirically by scanning a Gd-DTPA-doped water phantom. We ignore \( T \) effects because of the short 1.0-ms TE used and the relatively low expected Gd-DTPA peak uptake (<0.5 mM) in the renal cortex and medulla. Equation [1] includes a scaling factor or “gain” \( g \) that is time invariant and can be computed from the preinjection signal intensity \( S' \) and the preinjection longitudinal relaxation time \( T' \):

\[
g = \frac{S'}{f(T')} \quad [2]
\]

where \( T' \) is measured using a single breath-hold, low-flip-angle, segmented inversion-recovery prepared TrueFISP technique (19). Thus an arbitrary signal \( S \) can be converted into the corresponding \( T \) by inverting Eq. [1]:

\[
T = f^{-1}\left(\frac{S}{g}\right). \quad [3]
\]

If \( f \) is monotonic, its inverse can be computed numerically. In the next step the longitudinal relaxation time \( T \) is converted to Gd-DTPA concentration \( c \) using

\[
\frac{1}{T} = \frac{1}{T'} + r_c, \quad [4]
\]

where \( r \) is the specific relaxivity of Gd-DTPA assumed to be 4.5 (mM s)\(^{-1} \) (20).

### Estimates of RPF and GFR

The resulting cortical and medullary CTA curves were analyzed using a multicompartmental model to derive five independent parameters: the RPF, GFR, vascular volumes of the cortex and medulla, and rate of water absorption (10).

### Data Analysis

The volumes \( E_{\text{over}} \) of false-positive (oversegmented) and \( E_{\text{under}} \) of false-negative (undersegmented) voxels were used to assess segmentation accuracy. The total segmentation error \( E \) was defined as \( E = E_{\text{over}} + E_{\text{under}} \). The precision of segmentation was assessed by measuring the absolute disparity \( D \) of tissue volumes \( V_1, V_2, V_3 \) measured by three independent observers:
renal cortex was correlated with medullary undersegmentation.

FIG. 6. For simulated MRR images, the oversegmentation of the edulla, while noise,

\[ D = \frac{(|V_1 - V_2| + |V_1 - V_3| + |V_2 - V_3|)}{3}. \]  

[5]

We analyzed the accuracy and precision of the signal intensity and concentration curves that corresponded to the segmented volumes. Accuracy was defined as the absolute value of the signal intensity (concentration) difference between data resulting from GC and manual segmentation, averaged over all time points. Precision was measured as the absolute disparity of signal intensity (concentration) across pairs of observers, using an expression similar to Eq. [5]. The accuracy and precision values were then normalized to the average of corresponding curves taken over all time points. Multiple regression analysis was used to identify factors (observer, kidney function, body side, noise level, and blur level) contributing to \( E \).

We also measured the precision and accuracy of model parameters RPF and GFR that resulted from MRR processed with the GC segmentation technique compared to those determined from manually segmented data. For precision, an equation similar to Eq. [5] was used, with estimates of GFR and RPF replacing the volumes.

RESULTS

For simulated MRR data, including all levels of blur and noise, \( E_{\text{aver}} \) was 8.2 ± 7.3 cm\(^3\) (mean ± standard deviation (SD)) for the renal cortex and 2.7 ± 2.4 cm\(^3\) for the medulla, while \( E_{\text{under}} \) was 5.3 ± 3.5 cm\(^3\) for the cortex and 4.8 ± 4.1 cm\(^3\) for the medulla. Expressed as a percentage of the true volume, the oversegmentation error was 9.5% ± 8.6% for the renal cortex and 11.6% ± 10.4% for the medulla, while the undersegmentation error was 6.3% ± 4.3% for the cortex and 21.3% ± 18.3% for the medulla. The scatterplot shown in Fig. 6 demonstrates that cortical oversegmentation is related to medullary undersegmentation. Among potential factors (observer, kidney function, body side, noise level, blur level) contributing to segmentation error \( E \), the only significant predictor of \( E \) in both the cortex and medulla was the blur level (cortex: \( F = 66.0, P < 0.001 \), medulla: \( F = 64.3, P < 0.001 \); see Table 1).

The average precision \( D \) in measuring the volume of simulated renal cortex using the GC method was 6.3 cm\(^3\). Precision in segmenting the medulla averaged 3.9 cm\(^3\) (Table 2). The spatial blur was the only significant factor that influenced segmentation of both the cortex (\( F = 7.0, P < 0.001 \)) and the medulla (\( F = 3.6, P = 0.015 \)).

Manual segmentation (the reference method) of 22 kidneys in human subjects resulted in a wide range of volumes (98.1–240.3 cm\(^3\)), averaging 183.9 ± 30.5 cm\(^3\). There was no significant difference between the volumes of the left and right kidneys. The volumes of the renal cortex were 121.8 ± 22.9 cm\(^3\) (range = 62.9–160.8 cm\(^3\)), and the volumes of the medulla were 53.8 ± 9.8 cm\(^3\) (range = 31.3–69.9 cm\(^3\)).

Using the manual segmentation as reference, the error in GC segmentation of human subjects was 7.2 ± 6.1 cm\(^3\) for the renal cortex and 6.5 ± 4.6 cm\(^3\) for the medulla. As with the synthetic data, there was a systematic undersegmentation of renal medulla, averaging 2.9 cm\(^3\), accompanied by a 0.4 cm\(^3\) oversegmentation of the cortical region. The interobserver disparity \( D \) was 7.1 ± 4.2 cm\(^3\) for the cortical volume and 7.2 ± 2.4 cm\(^3\) for the medulla. For six kidneys with cysts, there was a trend (but no statistically significant difference) of improved segmentation accuracy and precision as compared to kidneys without cysts.

By overlaying segmented masks for the renal cortex and renal medulla on the acquired 4D datasets, CTA curves were determined for the renal cortex and renal medulla. Figure 7 shows the average signal intensity after the time axes were aligned to match the peaks of the cortical curves. For the cortex, on average, there was good agreement between enhancement curves derived using manual and GC segmentation. There was a small but noticeable underestimation of the medullary signal and preinjection cortical signal measured with the GC method (Fig. 7).

The mean values and SD, in cm\(^3\), across simulated 4D datasets, are given.
The temporal distribution of the error, before averaging and taking the absolute value of the difference, is shown in Fig. 8. There is a strong correlation across time points, with the sign change in only two out of 22 kidneys (Fig. 8). A similar tendency of the differences being consistently positive or negative was observed for the medullary region and the concentration curves. The mean absolute error in signal intensity (across all data) was 3.9% ± 2.1% for the cortex and 4.3% ± 1.1% for the medulla. The interobserver disparity in the cortical signal was 3.2% ± 1.2%, while for the medullary signal it was 4.4% ± 1.9%

Figure 9 shows the average CTA in the cortex and medulla after the time axes were aligned to match the peaks of the cortical enhancement curves. Because of our nonuniform acquisition scheme, this process also required data resampling. The error in Gd-DTPA concentration due to segmentation was 12.8% ± 6.3% for the cortex and 9.3% ± 2.7% for the medulla. The interobserver disparity in Gd-DTPA concentration was 10.2% ± 6.1% for the cortex and 10.3% ± 4.5% for the medulla.

Out ultimate goal was to assess the impact of different observers’ segmentation on estimates of renal function. These estimates are based both on measured volumes of the renal cortex and medulla, and on the analysis of cortical and medullary enhancement curves. Compartmental modeling of kidney function in 22 kidneys yielded RPFs in the range of 154 – 833 ml/min and single-kidney GFRs in the range of 33.3– 83.1 ml/min. The error was 7.5% ± 4.5% for RPF and 13.5% ± 8.8% for GFR. Interobserver disparity was 9.7% ± 6.4% for RPF and 14.8% ± 11.9% for GFR.

It took the observers 20.8 ± 3.7 min to segment a kidney using the GC method. Additionally, approximately 5 min were needed for cropping and coregistration. The manual segmentation required to provide reference values averaged 2.5 hr per kidney.

**DISCUSSION**

Despite the widespread use of functional renal imaging, only a few attempts have been made to assess the performance of renal segmentation. Most previous studies reported on the accuracy of measuring the volume of the entire kidney by either ultrasound or computed tomography (CT) (21–25).

Dynamic MRI following an intravenous injection of commonly available Gd-DTPA contrast provides a wealth of new information about renal function. Partial assessment of kidney function can be done through the analysis of the slopes of transit-time curves (2). Upslope modeling can be done by sampling a small representative cortical and medullary region (26). A more complete quantification of the filtration process, however, requires full segmentation of each kidney compartment because 1) the signal from a small tissue sample may not be representative of the rest of the organ, and 2) clinically, the most significant parameter is not the regional filtration rate (i.e., the rate per
unit volume) but its integral over the entire organ. Also, sampling of small regions results in suboptimal SNR levels.

Coulam et al. (27) used manual segmentation followed by thresholding to determine the accuracy of total renal parenchymal volume and medullary fraction in eight pigs imaged with a contrast-enhanced sequence on a 1.5T system. The precision of medullary fraction measurement using a signal-intensity thresholding algorithm was found to be poor, with a coefficient of variation of 56% as compared to 22% from postmortem measurement. We know of no other attempt to estimate segmentation errors for renal substructures.

We have employed a semiautomated segmentation method based on GC that 1) requires the user to specify the samples of the tissue of interest, and 2) allows further, repeated user interactions after the output of the algorithm is viewed (11). In three steps that process different phases of renal enhancement, the algorithm provides globally optimal binary masks for the entire kidney, the cortex, and the medulla (Figs. 4 and 5). We believe that the iterative nature of our software contributed to the low volumetric error (E averaging less than 4% of kidney volume), at the cost of larger disparity \( D \) across observers.

The ability to compute globally optimal 3D cuts gives the algorithm a competitive edge over simpler heuristics (such as thresholding or region-growing) that often fail for images with low SNR. Similarly to the variational method using level-sets (28), GC computes an object boundary that is optimal with respect to the geometrically motivated criteria of "clinging" to image edges and compactness (16). It is known that GC methods can compute globally optimal segmentation even for 3D data, while level-sets generate local minima that may be sensitive to initialization (29).

Our validation study first used simulated images to examine the impact of image noise and spatial resolution on segmentation error. While noise did not significantly affect the performance of the GC algorithm, the decrease of spatial resolution from 1.5 to 3 mm resulted in a near doubling of the segmentation error and a significant reduction in precision (Tables 1 and 2). This finding suggests that to optimize an MRR acquisition, one should strive to reduce the slice thickness. The relatively large (5 mm interpolated to 2.5 mm) slice thickness is the main source of tissue blurring in our acquisition sequence. In addition to random noise and spatial blur, potential sources of error include wraparound artifacts and RF-coil-related signal nonuniformities.

When applied to human-subject MRR images, the segmentation error and precision (as assessed by interobserver disparity) for the renal cortex and renal medulla were all on the order of 4% of the kidney volume. However, when expressed relative to the volumes of corresponding structures, the error was several times larger for the medulla than for the cortex because the medulla represents less than one-third of the kidney volume. The presence of cysts did not affect segmentation error or precision.

Imperfect segmentation resulted in an absolute signal intensity error also on the order of 4% of the true signal. For a given kidney, the signed errors, or the differences between measured and true signal before taking its absolute value, are correlated across time (Fig. 8), suggesting the presence of a systematic artifact that affects the GC method but not the manual segmentation by human experts.

The error and precision in Gd-DTPA concentration in the cortex and the medulla were on the order of 10%. These errors are larger than errors in volume and signal intensity, confirming the finding of relatively large numerical variability involved in rigorous computations of Gd-DTPA concentration (30). We also noted a systematic bias (the measured concentration of Gd-DTPA is larger than the true concentration), which seems to be related to the systematic oversegmentation of the cortex. Figure 7 demonstrates that our algorithm underestimates the preinjection signal \( S' \) in the renal cortex. This causes a reduction in gain \( g \) (see Eq. [2]) and a systematic overestimation of cortical concentration (Fig. 9).

When the CTA curves are analyzed by our modeling procedure, the errors average 7.5% for RPF and 13.5% for GFR, with interobserver disparity (precision) being slightly larger. Thus, it appears that the use of a pharmacokinetic model does not further amplify the error in CTA curves, most likely due to averaging of errors from 72 samples (36 cortical and 36 medullary) to five independent model parameters.

An ideal, fully automated algorithm would have perfect precision and no interobserver variability. For our implementation of GC, the interobserver disparity was as large as or larger than the difference between an observer and the ground-truth values. This large variability may be due to the high sensitivity of the algorithm to the choice of seeds, or (more likely) is a consequence of the iterative implementation of the GC algorithm, whereby the user is allowed to place additional hard seeds after viewing segmentation masks.

In spite of the limitations, the accuracy and precision of renal segmentation appear to be acceptable for clinical needs. The system reduces processing time from several hours to 20 min. With expedited image processing, MRR has the potential to expand our knowledge of renal function in individual kidneys and to help diagnose different types of renal insufficiency in vivo in a safe and noninvasive manner.

REFERENCES