Automated Segmentation of Kidneys from MR Images in Patients with Autosomal Dominant Polycystic Kidney Disease

Youngwoo Kim,* Yinghui Ge,* Cheng Tao,* Jianbing Zhu,* Arlene B. Chapman,† Vicente E. Torres,‡ Alan S.L. Yu,§ Michal Mrug,§ William M. Bennett,§ Michael F. Flessner,** Doug P. Landsittel,†† and Kyongtae T. Bae * for the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP)

Abstract
Background and objectives Our study developed a fully automated method for segmentation and volumetric measurements of kidneys from magnetic resonance images in patients with autosomal dominant polycystic kidney disease and assessed the performance of the automated method with the reference manual segmentation method.

Design, setting, participants, & measurements Study patients were selected from the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease. At the enrollment of the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease Study in 2000, patients with autosomal dominant polycystic kidney disease were between 15 and 46 years of age with relatively preserved GFRs. Our fully automated segmentation method was on the basis of a spatial prior probability map of the location of kidneys in abdominal magnetic resonance images and regional mapping with total variation regularization and propagated shape constraints that were formulated into a level set framework. T2-weighted magnetic resonance image sets of 120 kidneys were selected from 60 patients with autosomal dominant polycystic kidney disease and divided into the training and test datasets. The performance of the automated method in reference to the manual method was assessed by means of two metrics: Dice similarity coefficient and intraclass correlation coefficient of segmented kidney volume. The training and test sets were swapped for crossvalidation and reanalyzed.

Results Successful segmentation of kidneys was performed with the automated method in all test patients. The segmented kidney volumes ranged from 177.2 to 2634 ml (mean, 885.4 ± 569.7 ml). The mean Dice similarity coefficient ± SD between the automated and manual volume method was 0.88 ± 0.08. The mean correlation coefficient between the two segmentation methods for the segmented volume measurements was 0.97 (P < 0.001 for each crossvalidation set). The results from the crossvalidation sets were highly comparable.

Conclusions We have developed a fully automated method for segmentation of kidneys from abdominal magnetic resonance images in patients with autosomal dominant polycystic kidney disease with varying kidney volumes. The performance of the automated method was in good agreement with that of manual method.

Introduction
Autosomal dominant polycystic kidney disease (ADPKD) is the most common hereditary renal disorder and the third most common single cause of ESRD (1,2). ADPKD is characterized by bilateral enlarged kidneys containing numerous cysts that expand and compress surrounding renal parenchyma (3,4). For the assessment of the severity and the progression of ADPKD, total kidney volume (TKV) serves an important biomarker and also, inversely correlates with declining GFR (5,6). Nevertheless, the relationship between the TKV and renal function in patients with ADPKD remains complicated and controversial because of the number of confounding factors affecting the progression of renal disease (7–9).

The Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP) was established to explore the associations among the renal or cyst volume, renal function, and clinical variables and evaluate their changes during the course of the disease (2). In this study, the kidneys were segmented manually from magnetic resonance (MR) images to compute TKV, which was subsequently analyzed in association with other clinical, functional, and genetic biomarkers.

Various methods for segmenting kidneys and quantifying kidney volumes were reported from computed tomography (CT) or MR images in patients with ADPKD. King et al. (10) used manual delineation and adaptive threshold to segment ADPKD kidneys.
and cysts, respectively. In addition, stereology represents a commonly used approach to segment and quantify the volumes of kidneys and cysts (2,4,11,12). These manual segmentation methods rely heavily on the observer’s perception and manual input to complete the segmentation process. Although straightforward, the manual segmentation methods are resource intensive, time consuming, and subject to analyst bias and error. Development of an automated approach could overcome these limitations, making it highly desirable. Although previous studies have reported semiautomated methods for the segmentation of ADPKD kidneys that still required the observer’s perceptual guidance and input (2,8,13,14), to date, no study has published a fully automated segmentation of kidneys in ADPKD.

In this study, we developed a fully automated segmentation of kidneys from MR images in patients with ADPKD by exploiting prior knowledge of spatial location of kidneys modeled as the spatial prior probability map (SPPM) and the propagated shape constraint (PSC) that were incorporated into the level set framework (15).

Materials and Methods

Patients and MR Imaging

The study protocol for the CRISP (clinical trials registration NCT01039987; registration date: December 23, 2009) was approved by the institutional review board at each participating clinical center. Informed consent was obtained from all patients who participated in the CRISP Study. The patients in the CRISP Study were recruited from patients with ADPKD between 15 and 46 years of age with relatively preserved GFRs (i.e., a 24-hour creatinine clearance >70 ml/min per 1.73 m² or a Cockcroft-Gault estimate of creatinine clearance >70 ml/min) to participate. Men were eligible if their serum creatinine concentrations were <1.6 mg/dl, and women were eligible if their serum creatinine concentrations were <1.4 mg/dl.

The standardized magnetic resonance imaging (MRI) protocol in the CRISP Study was carried out using 1.5-T MRI scanners. The parameters of coronal T2–weighted single-shot fast spin echo MR images with fat saturation were 0.59- to 1.41-mm/pixel resolution in plane, 3-mm slice thickness, 90° flip angle, and 500–1491 ms/82–101 ms repetition time/echo time. The details of the CRISP MRI protocol are described in previous publications (2,4). After the MRI images were coded to maintain the anonymity of the participant, the images were then transferred via secure internet transmission from the four clinical centers, and the images were retrieved to a desktop personal computer workstation (3).

For our study, 60 patients with ADPKD were randomly selected from the CRISP Study MR image database. The study patients consisted of 29 men and 31 women, with their ages ranging from 27 to 59 years old (mean ±45.6 ±9.3 years old). Note that the ages of study patients increased since the enrollment (the CRISP I: mean age =33.8±8.9 years old; range =15–46 years old), because we used the MR image data from more recent follow-ups (the CRISP III: follow-up of 12 years). The TKVs of the patients ranged from 184.9 to 2563.6 ml (mean =947±579.4 ml and median =811.4 ml). Table 1 summarizes patients’ demographic and kidney volume distributions.

To generate the reference standard for the segmentation from the training and validation datasets, an experienced radiologist reviewed and separated the MR images of the abdomen into the right- and left-sided abdomen. From each set of the lateralized abdominal MR images, the boundary of each kidney was manually delineated slice by slice using the commercially available software (Analyze 12.0; Mayo Clinic, Rochester, MN).

We divided 60 study patients into two groups: the training set of 30 patients and the test set of 30 patients. The 60 study patients were first ranked according to the distribution of volumetric size measured from the manual segmentation. The rank ordering determined 30 patients who were odd ranked (D1) and 30 patients who were even ranked (D2). Initially, the D1 was used as the training set, whereas the D2 served as the test set for the automated method and evaluation of the segmentation performance. Then, these two datasets were swapped as the training and test sets and reanalyzed for a crossvalidation and data analysis.

Automated Segmentation Method

In our implementation of the automated segmentation method, we exploited the prior knowledge of the spatial locations of the kidney in MR images. The spatial locations of kidneys were modeled as the SPPM. We determined that the SPPM approach was more appropriate than a fixed shape model approach for ADPKD kidneys, because the shape and size of ADPKD kidneys were highly variable compared with normal kidneys. The key notion of using the SPPM is that kidneys are located within the abdomen MR images and spatially distributed as estimated by the probability distribution of the contours of kidneys learned from the training set. In this model framework, the shapes of kidneys may be variable and do not have to be constrained. Nevertheless, the similarity in kidney contours in neighboring MR images was enforced by the PSC and integrated into the level set framework.

The segmentation method was composed of three steps (Figure 1): SPPM construction, regional mapping, and boundary refinement. As a first step, the SPPM was constructed from the training set of the manually segmented kidney contours. When a test patient’s abdominal MR images were entered, they were first preprocessed for noise reduction and mapped according to the SPPM to obtain candidate kidney regions. The left and right candidate kidney regions were mapped separately and processed with the spatial prior-based level set that was followed by the morphologic closing operation to refine and the segmented boundary of kidney. The iterative process is denoted in Figure 1 between the SPPM–based level set and the segmented kidney corresponding to the PSC.

SPPM Construction. A set of kidney masks segmented manually by the radiologist was used for the SPPM construction. A volume–based spatial normalization using Lanczos interpolation was carried out to normalize the variation in the fields of views among the patients in the training set (16). The constructed SPPM is illustrated in Figure 2: the median slice with the probability color scale is in Figure 2A, and the surface rendering representation at P=0.50 is in Figure 2B. The equation for the SPPM generation is in Supplemental Appendix 1.
Regional Mapping. Each test patient’s abdominal MR images were first preprocessed by the application of total variation (TV) regularization to reduce image noise and improve image signal homogeneity (17). The TV regularization process also increases the regional connectivity of voxels representing the kidney. The parameters of TV regularization were 0.1 and 100 for regularization (λ) and iteration, respectively. The original MR and TV regularized images are shown in Figure 3.

After the TV regularization, the test image was preprocessed to compute the magnitudes of image gradients. The magnitude of image gradients, which is similar to contours often used in maps, represented the base image features for the extracted boundary of the candidate kidney regions. To obtain the full range of direction, the magnitudes of image gradients were calculated in three different directions.

The magnitudes of image gradients were multiplied to the SPPM to generate the map of candidate kidney regions (Figure 3). The three-dimensional volume of the magnitudes was projected onto the two-dimensional coronal plane (i.e., from anterior to posterior direction of the abdomen) and then, the one-dimensional axial plane (i.e., from superior to inferior direction). With the one-dimensional signal obtained by the aforementioned two projections, the separation between the left and right kidney regions was determined by using the method by Otsu (18) from the bimodal distribution of the signals. However, in this implementation, to simplify the segmentation process, the abdominal MR images were divided into the right and left abdominal MR images, and the image processing was applied separately to the right and left abdominal MR images to segment the right and left kidney, respectively.

Boundary Refinement. The segmentation of kidney was determined when the candidate boundaries satisfied the constraints of the SPPM (i.e., a priori construction of the kidney regional map) and the PSC (i.e., smoothness of boundary neighboring slices) that were incorporated into the level set framework (19). In this model, the constraints were expressed in energy terms, and the segmentation was achieved as a particular case of minimizing energies. The total energy represented the sum of the three energy terms corresponding to the test image set, SPPM, and PSC. The SPPM or PSC energy term was expressed as a function of the probability distribution and the Heaviside function, where the variable of the Heaviside function represented the level set function, defining the evolving contour of segmentation. The motion equation was derived from the aforementioned total energy functional by using the Euler–Lagrange equation. Additional descriptions of the definition of the energy terms and the derivation of the motion equation are in Supplemental Appendix 1.

Finally, the evolved contour was postprocessed with a morphologic closing operation to improve the anatomic connectivity and smoothness of kidney boundary.

Data Analyses

The performance of the automated segmentation method was evaluated and compared with the manual method by means of two metrics: the Dice similarity coefficient (DSC) (20), which measures how closely two independently segmented kidneys match geometrically when they are superimposed onto each other, and the intraclass correlation coefficient of the segmented kidney volume measurements between the two methods. A formula for calculating DSC is in Supplemental Appendix 1.

For the crossvalidation of the training and test sets, the data analyses and evaluations were performed twice: E₁ and E₂. The E₁ was conducted with the D₁ dataset for the training set and the D₂ dataset for the test set, whereas the E₂ was performed with the D₂ dataset for the training set and the D₁ dataset for the test set. In addition, Bland–Altman analysis was used to visually assess systemic differences and estimate the bias and limits of agreement (LOAs). The statistical analysis was performed in Stata (version 12; StataCorp., College Station, TX); Bland–Altman plots were created using the batplot command.

Results

The mean DSCs ±SD of the two crossvalidation outcomes, E₁ and E₂, were 0.88 ± 0.08 and 0.88 ± 0.09, respectively, and 0.88 ± 0.08 on average. Figure 4 illustrates two
examples of evolving contours with the motion equation defined in equation A7 in Supplemental Appendix 1. The initial contours, which are denoted by multiple circles diffusely laid over the image, progressively merge or phase out to converge onto the boundary of kidney. Some other structures with high T2 signal in the image (e.g., liver cysts and cerebrospinal fluid) mimicking renal cysts were successfully excluded from the final segmented renal boundary.

The segmentation outcomes from kidneys of varying size and cyst burden are shown with superimposed manual and automated boundaries in Figure 5. Although the segmentation results by the manual and automated methods show some differences in finer boundary depiction, particularly at regions with highly irregular protrusion of cysts, the overall boundaries determined with the two methods are in good agreement.

The scatterplots for the volume measurements of the kidneys segmented by the manual versus the automated method are shown in Figure 6, A and B; one plot is from validation E1, whereas the other plot is from E2, respectively. The intraclass correlation coefficients were 0.97 ($P<0.001$; 95% confidence interval [95% CI], 0.95 to 0.98) from E1 and 0.96 ($P<0.001$; 95% CI, 0.94 to 0.98) from E2. The combined correlation coefficient was 0.97 ($P<0.001$; 95% CI, 0.95 to 0.98). The slope of the regression ($P<0.001$) indicates that the automated method minimally but statistically insignificantly overestimates the volume compared with the manual method. Although the high level of average agreement and overall variability were explained (e.g., with high intraclass correlations), Bland–Altman analyses revealed large differences for several individual measurements (Figure 6, C–F). More specifically, the LOAs were $-236.90$ and $295.75$ ml for E1 and $-268.74$ and $315.53$ ml for E2 in terms of the difference between the estimated TKVs. Regarding the percentage difference, in which TKV difference was normalized by the estimated volume of the manual segmentation, in estimated TKVs, the LOAs were $-28.77\%$ and $37.17\%$ for E1 and $-37.22\%$ and $43.20\%$ for E2. A relatively even spread of differences was observed between the LOAs over most of the range, with some outliers in the differences. These outliers corresponded to several right kidneys with borders that were obscured by extensive liver cysts located in the posterior liver and difficult to differentiate from the liver border. Some of the liver cysts were included in the segmented kidneys, resulting in the overestimation of the kidney volume.

**Discussion**

Quantitative imaging plays a crucial role for monitoring ADPKD progression, clinical management, and targeted therapeutic trials (21). In particular, kidney volume measured from MR images is an important imaging biomarker for the assessment of ADPKD disease progression (2,4). However, the measurement of kidney volume from patients with ADPKD remains challenging, because the morphology and size of kidneys vary greatly between patients at different stages of ADPKD progression.

In a typical clinical setting, when a patient undergoes ultrasound imaging of the kidney, a crude estimate of size is usually provided as the length of kidney measured linearly from the most superior to the most inferior tip of the kidney. CT or MR images offer a more accurate measurement of kidney volume from cross-sectional image analysis for clinical applications, such as kidney transplant planning. However, this cross-sectional imaging approach is more difficult and requires more time-consuming processes in image analysis than using ultrasound imaging. A simple but less precise approach of approximating kidney volume from CT or MR images uses the ellipsoid formula that is on the basis of the multiplication of three orthogonally measured longest dimensions of kidney (Supplemental Appendix 2) (22). In a typical clinical setting, the kidney volume from CT or MR images is measured manually by a trained expert who delineates the kidney border over the CT or MR images. Although the segmentation and volume measurement of the kidney by this method are straightforward, it is laborious and subject to considerable interobserver variability and error.
To overcome some of the limitations of the manual method, groups have developed various automated and semiautomated techniques for the segmentation of kidneys from CT or MR images (23–25). These methods were primarily designed to segment kidneys of normal morphology and size and are not applicable to ADPKD kidneys that present with a wide range of variations in shape and size. However, reports have described the use of some semiautomated methods for the segmentation of ADPKD kidneys (2,8,13,14). The operation of these semiautomated methods relied on the expert user who reviewed MR images, identified the kidney region, and placed seed positions over the region for region growing. The kidney was then segmented as a region with similar signal intensity. The user had to carefully monitor the entire segmentation process, edit, and finalize the segmentation outcome. In comparison, our new segmentation method is a fully automated approach without requiring user interaction during the segmentation process. To our knowledge, this is the first study reporting a fully automated segmentation of ADPKD kidneys from cross-sectional radiologic imaging studies.

We assessed the performance of our method by the crossvalidation of the training and test sets using the manual segmentation by a radiologist expert as the reference standard. The crossvalidation showed excellent correlations across kidneys of varying size and cyst burden. A major challenge for the automated segmentation of kidneys from abdominal MR images was to separate and exclude other organs adjacent to the kidneys, such as the liver, spleen, and vertebrae, from the segmented outcome of kidney regions. In particular, when the liver is diffusely replaced by liver cysts, some liver cysts near the right kidney are located closer to the renal boundary, obliterating the border between the liver and kidney (Figures 4 and 5). The difficulty of segmenting these kidneys was confirmed in Bland–Altman analysis, in which the images outside the line of agreement were exclusively from the right kidneys with borders that were largely obscured by extensive liver cysts in the posterior surface of the liver. Some of liver cysts were included in the segmented kidneys and contributed to the overestimation of kidney volume. In addition, when the spleen juxtaposes the left kidney, the two organs may be difficult to separate, because their parenchyma tend to have similar signal intensities (Figure 5). Increased ADPKD severity results in cyst expansion, size increase, and shape irregularity, deforming the contour of the kidney and exacerbating the difficulty of separating organs.

Given the heterogeneity and uncertainty of ADPKD kidney morphology, we designed our segmentation method to not require a priori shape of the kidney. Our
segmentation method is highly flexible and applicable to kidneys with varying morphologies and sizes. The algorithm of the method was on the basis of the prior statistical information of kidney location (SPPM), regional MR signal intensity, and boundary constraint propagated between neighboring image slices (PSC). The SPPM globally guided to the initial candidate kidney region using a priori statistical distribution of kidney regions from the training set. The information from regional MR signal intensity was used to constrain the potential candidate kidney regions. Additional refinement for the candidate kidney regions was achieved from the PSC in that the kidney boundaries maintained similarity and smooth transition between adjacent slices. The spatial prior probability rather than shape knowledge was exploited to address unpredictable and diverse morphologies of ADPKD kidneys (26).

With the \textit{a priori} constraint terms formulated in the level set framework, the proposed method successfully segmented the kidney regions. The agreement in segmentation between the automated method and the radiologist expert was 88% in the shape congruence and 97% in the volume measurements. The automated program successfully detected the boundaries of small left kidneys with paucity of renal cysts at the proximity of the spleen as well as right kidneys with substantial burden of liver cysts adjacent to the kidney borders.

We anticipate that our automated method is applicable to the automated measurement of kidney volumes in a large-scale ADPKD study. When MR images are acquired at different time points in a longitudinal study, in addition to measuring kidney volume changes, we could superimpose the automatically segmented kidneys over the different time points to compare and determine the regional temporal changes in kidney growth and morphology. The detected regional differences in kidney morphology may allow us to investigate the individual patient variation in kidney growth as the disease progresses. Moreover, when we need to segment individual renal cysts from an ADPKD kidney to investigate the characteristics of renal cyst distribution and morphology, the kidney has to be segmented first. The kidney boundaries segmented with our automated method could be used as a precursor for the segmentation of renal cysts as shown in a previous study (12).

Although a fully automated segmentation program is a powerful tool to estimate kidney volumes with minimal operator intervention, it should be used with two key precautions. First, high-quality MR images are essential for accurate and precise volume measurements, regardless of the choice of segmentation and measurement methods. High-quality MR images may be even more critical for the fully automated segmentation that is performed with no supervision by an image analysis expert. As long as the quality of MR images is assured, the segmentation program is indifferent to the type of MRI scanners on which the images are acquired. Second, for practical use of the automated segmentation program, we expect the user to review the anatomy of MR images before running the program. This automated segmentation program was not designed or developed to process rare aberrant renal morphology (e.g., malrotated or horseshoe kidneys). These rare morphologies should be prescreened and measured.

Figure 4. \textbf{Graphical illustration of polycystic kidney segmentation.} Evolution of contours are shown with iterations at (A) zero, (B) five, (C) 10, (D) 20, (E) 30, and (F) 50. The contours are drawn as solid lines overlaid on the noise–reduced original magnetic resonance images.
manually or with a semiautomated program. In addition, it would be necessary to display segmentation results and provide a user with image editing tools to revise the outcome of the automated segmentation for the estimation of accurate kidney volumes while incorporating an expert’s input. In particular, severe burden of liver cysts in the posterior surface of the liver may compromise the accuracy of automated segmentation of the right kidney.

This study has several limitations. First, although the measurement accuracy is an important parameter determining the performance of a segmentation method, it was not studied. In this technical development study, we mainly focused on the development of a new automated segmentation method and the evaluation of the measurement precision between the automated and manual methods. Second, the performance of a segmentation method is dependent on the training set. A good training set should contain a wide range of variations that are expected from the test population. The good result of our crossvalidation study indicated that the training and test sets were relatively well balanced in our study cohort. Third, the sample size of the study was relatively small. The main objective of the study was to develop an automated segmentation method with no user input in the segmentation process and then, evaluate the performance of this segmentation method. Despite the small sample size, we believe that the patients with ADPKD in our study showed a good representation of commonly observed ranges of disease severity in the ADPKD population and yielded a good segmentation outcome from MR images acquired in the general patient population with ADPKD.

In conclusion, we have developed a fully automated method for the segmentation of kidneys from abdominal MR images in patients with varying degrees of severity of ADPKD. The performance of the automated method was in good agreement with that of the manual segmentation method.

**Acknowledgments**

We thank the radiologists, nephrologists, radiology technologists, imaging engineers, and study coordinators in the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP). The CRISP Study is supported by National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health Cooperative Agreements DK056943, DK056956, DK056957, and DK056961; National Center for Research Resources General Clinical Research Centers Grants RR000039 (to Emory University), RR00585 (to Mayo College of Medicine), RR23940 (to Kansas...

---

Figure 5. | Polycystic kidney segmentation results for visual assessment. The manually and automatically segmented autosomal dominant polycystic kidney disease kidneys are depicted with red and green solid lines, respectively. The yellow lines separate the right and left kidneys. (A) The automated method successfully segmented the kidney boundaries closely in contact with (B and C) liver cysts and (D) the spleen.
Figure 6. | Statistical analysis of segmentation results comparing the manual and automated methods. (A and B) Scatterplots illustrating the statistical analysis in terms of volume measurements comparing the manual and automated methods and Bland–Altman plots regarding (C and D) the difference between estimated total kidney volumes and (E and F) the percentage difference in estimated total kidney volumes at the two crossvalidations of $E_1$ and $E_2$. The regressions and correlation coefficients are denoted near the trend lines in (A) and (B).

University Medical Center), and RR000032 (to the University of Alabama at Birmingham); and National Center for Research Resources Clinical and Translational Science Awards RR025008 (to Emory University), RR024150 (to Mayo College of Medicine), RR033179 (to Kansas University Medical Center), RR025777 and UL1TR000165 (to the University of Alabama at Birmingham), and
Disclosures
A.B.C. is a consultant to Otsuka Corporation. V.E.T. received research support from Otsuka Corporation. M.M. is a consultant to Otsuka Corporation and Alexion Pharmaceuticals. K.T.B. is a consultant to Kadlec LLC. No other disclosures were reported.

References

Received: August 3, 2015 Accepted: December 21, 2015

Published online ahead of print. Publication date available at www.cjasn.org.

This article contains supplemental material online at http://cjASN.asnjournals.orgLOOKUP/suppl/DOI:10.2215/CJN.08300815-/-/DCSupplemental.