Scientific Article

Feasibility of automated 3-dimensional magnetic resonance imaging pancreas segmentation

Shuiping Gou PhD [a,b], Percy Lee MD [b], Peng Hu PhD [c],
Jean-Claude Rwigema MD [b], Ke Sheng PhD [b,*]

[a] *Key Lab of Intelligent Perception and Image Understanding of Ministry of Education, Xidian University, Xi'an, Shanxi, China*
[b] *Department of Radiation Oncology, University of California, Los Angeles, California*
[c] *Department of Radiological Science, University of California, Los Angeles, California*

Abstract

Purpose: With the advent of magnetic resonance imaging (MRI) guided radiation therapy, internal organ motion can be imaged simultaneously during treatment. In this study, we evaluate the feasibility of pancreas MRI segmentation using state-of-the-art segmentation methods.

Methods and materials: T2-weighted half-Fourier acquisition single-shot turbo spin-echo and T1 weighted volumetric interpolated breath-hold examination images were acquired on 3 patients and 2 healthy volunteers for a total of 12 imaging volumes. A novel dictionary learning (DL) method was used to segment the pancreas and compared to t mean-shift merging, distance regularized level set, and graph cuts, and the segmentation results were compared with manual contours using Dice's index, Hausdorff distance, and shift of the center of the organ (SHIFT).

Results: All volumetric interpolated breath-hold examination images were successfully segmented by at least 1 of the autosegmentation method with Dice's index >0.83 and SHIFT ≤2 mm using the best automated segmentation method. The automated segmentation error of half-Fourier acquisition single-shot turbo spin-echo images was significantly greater. DL is statistically superior to the other methods in Dice's overlapping index. For the Hausdorff distance and SHIFT measurement, distance regularized level set and DL performed slightly superior to the graph cuts method, and substantially superior to mean-shift merging. DL required least human supervision and was faster to compute.

Conclusions: Our study demonstrated potential feasibility of automated segmentation of the pancreas on MRI scans with minimal human supervision at the beginning of imaging acquisition. The achieved accuracy is promising for organ localization.

* Corresponding author. Ke Sheng, PhD, Department of Radiation Oncology, University of California, Los Angeles, 200 Medical Plaza Way, Suite B265, Los Angeles, CA 90095.
E-mail address: ksheng@mednet.ucla.edu (K. Sheng)

Introduction

Patients with pancreatic adenocarcinoma have a poor prognosis, with cumulative 5-year survival of less than 5%.[1,2] Many patients present with unresectable locally advanced lesions at the time of diagnosis. Although it is

unlikely to cure pancreatic cancer with radiation therapy alone, with sufficiently high doses, it is possible to achieve local control or resectability conversion[3-8] that is correlated to significantly prolonged patient survival.[9] However, radiation doses to the pancreas are limited by surrounding radiosensitive serial organs. The goal of delivering a sufficient tumor dose is further complicated by significant organ motion[10-12] that is poorly anchored to the bony anatomy position.[13] Moreover, the large internal target volume established using 4-dimensional computed tomography (CT) has proved to be highly unreliable with high probability of under- or overestimating pancreatic motion.[14,15] The current gold standard for pancreas image guided radiation therapy is registration using implanted fiducial markers. To account for the uncertainties of using the markers to represent the pancreas location, a 3-mm margin is added to the gross tumor volume for planning target volume.[16] As a noninvasive motion management method, magnetic resonance imaging (MRI)-guided radiation therapy[17-21] is a promising alternative because of its superior soft tissue imaging capability.

To effectively use MRI in radiation therapy, target and surrounding normal tissue need to be delineated. Manual contouring for planning purposes is acceptable but impractical for intrafractional motion monitoring. Automated pancreas segmentation is particularly challenging and rarely reported. In a CT-based multiorgan segmentation study, the accuracy of pancreas segmentation was shown to be significantly lower than other abdominal organs.[22] Lower pancreas segmentation accuracy using a simple region growth, gradient, and shape control method, compared with nearby organs such as the liver and stomach based on 2-dimensional (2D) dynamic MRI images was reported.[23] Moreover, the 2D segmentation study did not capture the complex morphology of the pancreas based on a single image slice; therefore, it is not clear whether the pancreas can be automatically segmented on MRI and if the accuracy is comparable to that of the implanted fiducial markers, which requires an additional 3-mm margin. In the present study, we aim to test the feasibility using four state-of-the-art automated pancreatic segmentation methods and their potential for tumor tracking.

Methods and materials

MRI acquisition

Three pancreatic cancer patients (P1, P2, and P3) and 2 healthy volunteers were included in this study under an institutional review board protocol.

The patient MRI data were retrospectively selected and healthy volunteer data were prospectively acquired using 2 MRI techniques described as follows. The first MRI was performed using a T1-weighted 3-dimensional (3D) Fast Low Angle SHot volumetric interpolated breath-hold examination (VIBE) technique with fat suppression on a 1.5 T MRI scanner (Avanto, Siemens Medical Solutions, Erlangen, Germany) with a 6-channel body receive coil array. Data were typically acquired with the scan parameters with small variation in resolution and field of view among subjects: repetition time /echo time: 3.6/1.32 ms; axial field of view: 350×284 mm^2; flip angle = 10°; matrix dimension: 320×260; slice thickness: 2.5 mm and in-plane pixel size 1.093×1.093 mm^2.

The second MRI sequence was a T2-weighted half-Fourier acquisition single-shot turbo spin-echo (HASTE) sequence on the same 1.5 T MRI scanner. Acquisition parameters were: repetition time /echo time: 800/54 ms; axial field of view: 400×280 mm^2; flip angle = 150°; matrix dimension: 256×180; slice thickness: 6 mm, spacing between slices: 7.5 mm and in-plane pixel size 1.6×1.6 mm^2.

Note that VIBE was a 3D sequence and HASTE was a 2D multislice sequence. Both VIBE and HASTE images were acquired with breath hold at the end-of-exhalation position. These 2 sequences were of interest for MRI guided radiation therapy because they were relatively fast ($\sim$12 seconds/volume without using acceleration techniques such as compressed sensing and low-rank decomposition to exploit patient spatiotemporal coherence[24]) and showed good abdominal organ contrast. With accelerated imaging acquisition, they were potential candidates for real-time monitoring of the internal organ motion. The MRI of the 2 healthy volunteers were noncontrast only and all patients also received gadolinium-DTPA contrast resulting in pre- and post-VIBE images except 1 patient whose precontrast VIBE images was corrupted and unusable. As a result, we obtained a total of 12 image sets for the segmentation study.

In addition to VIBE and HASTE, balanced steady-state free precession is commonly used for high-speed volumetric or 2D images. However, steady-state free precession was not included in the patient abdominal MRI protocol and thus excluded from this report.

Segmentation methods

In this study, we used 4 state-of-the-art segmentation methods including mean shift merging (MSM), distance regularized level set (DRLS), graph cuts (GC), and dictionary learning (DL). These methods are described in the following section.

MSM

The edge- and region-based techniques was first proposed by Bajcsy and Pavlidis to segment natural image.[25,26] This method was then adopted for MRI scan segmentation.[27] In this model, the image was initially

partitioned into regions and then an iterative region merging process was applied to refine the segmentation result. In this study, the corresponding parameters were set: spatial bandwidth, 4 to 12; range bandwidth, 4 to 11; and minimum region area, 20 to 100.

After the images were roughly segmented into small subareas based on the presentation of native organ structures using mean shift, a human operator then defined the foreground and background on 1 2D slice. The maximal-similarity measure was then used to merge regions based on previously published rules.[28]

Level set

The level set method was first developed by Osher and Sethian to describe wave propagation.[29,30] The method was then applied for medical imaging processing[31] and has evolved to be one of the most important tools for imaging segmentation.[32] The level set evolution was derived as the gradient flow that minimizes energy functional with a distance regularization term and an external energy that drove the motion of the zero level set toward desired locations. In this study, initial contours were manually drawn on the first imaging slice and the propagated to the subsequent slices. The detail of the distance regularized level set (DRLS) method can be found in the reference.[32] Briefly, segmentation was performed based on a distance regularized level set energy function:

$$\varepsilon(\varnothing) = \mu \mathscr{R}_p(\varnothing) + \lambda \mathscr{L}_g(\varnothing) + \alpha \mathscr{A}_g(\varnothing), \qquad (1)$$

Where $\mathscr{R}_p(\varnothing)$ was the level set regularization term, $\mathscr{L}_g(\varnothing)$ computed the line integral of the function along the zero level contour, and $\mathscr{A}_g(\varnothing)$ was introduced to speed up the motion of the zero level contour in the level set evolution process. We adopted the same parameter values for μ (0.2), λ (5) from the original publication,[32] but the α value resulted in poor segmentation performance. In this study, we initially set α as [1.5, 4] and the value λ of [3, 5] for pancreas and adjusted for individual subjects based on segmentation results. Different from the other 3 methods using manual marking of the foreground and background, 1 or 2 initial pancreas 2D contours were needed for DRLS as the initialization condition.

GC

In the GC method, the MRI pixels were represented by nodes that were connected by edges that describe the dissimilarity between them. To segment the foreground from background, an optimal cut can be achieved by minimizing the cut cost function that is built into the edge weights.[33] This method has been applied to liver segmentation and showed superior performance than an active contour method.[34] In this study, we closely follow the method outlined by the previous publication.[33]

Briefly, to segment the MRI using GC, A graph $G = (V, E)$ was defined as a set of nodes (vertices V) and a set of undirected edges (E) that connected these nodes corresponding to pixels $x_i \in P$ of the image. These nodes were also connected to 2 special nodes named foreground terminal (a source S) and background terminal (a sink T). Therefore, $V = P \cup \{S, T\}$. The set of edges E consisted of 2 types of undirected edges: n-links (neighborhood links) and t-links (terminal links). Each pixel x_i had 2 t-links $\{x_i, S\}$ and $\{x_i, T\}$ connecting it to each terminal. An n-link connecting a pair of neighbors x_i and x_j was denoted by $\{x_i, x_j\}$. Therefore, $E = N \cup \{x_i, S\}, \{x_i, T\}$.

The whole image was then labeled by a binary vector $\mathscr{L} = \{l_1, l_2, ..., l_i\}$ whose components l_p specifies assignments to pixels x_i in P. Each l_i can be either "foreground (pancreas)" (1) or "background" (0). Therefore, vector $\mathscr{L}$ defined segmentation. The soft constraints imposed on boundary corresponding to the n-links and region properties corresponding to the t-links of $\mathscr{L}$ were described by the cost function (weight)

$$E(\mathscr{L}) = \lambda R(\mathscr{L}) + B(\mathscr{L}) \qquad (2)$$

where the coefficient $\lambda \geq 0$ specifies a relative importance of the region term $R(\mathscr{L})$ versus the boundary term $B(\mathscr{L})$.

$$R(\mathscr{L}) = \sum_{i \in P} R_i(l_i) \qquad (3)$$

was the regional term and $R_i(l_i)$ was the cost for assigning pixel x_i to "foreground" and "background" based on the negative log-likelihoods and the imaging histograms.

$$B(\mathscr{L}) = \sum_{\{x_i, x_j\} \in N} B_{i,j} \cdot \delta_{l_i \neq l_j} \text{ was the boundary term} \qquad (4)$$

where

$$\delta_{i \neq j} = \begin{cases} 1 & if \ l_i \neq l_j \\ 0 & if \ l_i = l_j \end{cases}, \ B_{i,j} \propto \exp\left(-\frac{(x_i - x_j)^2}{2\sigma^2}\right) \cdot \frac{1}{dist(x_i, x_j)}, \ \sigma$$

was the estimated image noise and $dist(x_i, x_j)$ the Euclidean distance between x_i and x_j. This function penalized discontinuities between pixels of similar imaging values. The image was segmented into the pancreas and background by removing edges that minimized the total cutting cost.

To define S and T terminals, the background and foreground were manually labeled in *each* slice that the pancreas presented.[33] The minimum cost cut on G was computed exactly via gradient descendent for 2 terminal graph cuts. In this study, we set the neighbor constant be as 10, terminal constant as 10^{12}. As shown previously,[35] and in our segmentation results, GC alone was insufficient to achieve high segmentation specificity and was typically used in combination with a refinement tool, such as the manifold clustering method described later.

DL

The substantial inter-subject variation of pancreas morphology[36] renders a population based atlas or dictionary ineffective for segmentation. In this study, the dictionary was *individually* established for each 3D image. We adopted a DL segmentation method originally developed for prostate segmentation[37] and modified it using manifold clustering that will be described later. To obtain enhanced imaging feature information, we calculated gray-gradient on patches of size 5 × 5 pixel,[2] and obtained 15-dimensional features. These in combination with the gray values of the 25 pixels in the patch constructed a 40-D feature vector for each patch. In addition, the texture-rich VIBE images provided additional 6-dimensional texture that can be captured using a gray-level co-occurrence matrix.

The images were first segmented to subareas using mean shift. We trained the target dictionary of pancreas D_1 and background (non-pancreas organs) dictionaries D_2-D_3 using the K-means singular value decomposition (K-SVD) algorithm on mean shift rough partition. This process was performed in 2 steps.

1. Sparse encoding: the 2D initial input image X was represented by its feature vector v, and initialize dictionary denoted by $D^0 = \{D_1^0, D_2^0, D_3^0\} = \{d_1^0, d_2^0, ..., d_M^0\}$. The optimization problem was solved using the matching pursuit method:

$$\min \left|\left|v - D^0\alpha\right|\right|^2 \quad \text{subject to} \quad \left|\left|\alpha\right|\right|_0 \leq L, \tag{5}$$

where α was the sparse representation coefficient matrix corresponding to each sampling that $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_M\}$.

2. Dictionary update: Each atom d_k of dictionary D and its sparse representation coefficient α_k was iteratively updated by solving the following optimization problem and obtaining corresponding error matrix E_k in the jth iteration,

$$\arg \min_{d_k, \alpha_k} \left|\left|E_k^j - d_k^j\alpha_k^j\right|\right|_2, \text{ Where } E_k = v - \sum_{i \neq k} d_i^{j-1}\alpha_i^{j-1} \tag{6}$$

The error matrix E_k was decomposed using single value decomposition (SVD), such that $SVD(E_k) = U\Delta V^T$.

d_k was updated with the first column of the matrix U, α_k with the first row of matrix V multiplied by $\Delta(1, 1)$. This step was iteratively performed until the convergence condition was met[38] or the maximal number of iteration of 20 was reached.

The subareas from mean-shift presegmentation corresponding to dictionary D_1, D_2, and D_3 were then labeled K_1, K_2, and K_3, respectively.

Once the dictionary was obtained from the first training slice, for a subsequent image slice N, we extracted the 46-dimensional feature vector v. The vector v was reconstructed by dictionaries D_1 to D_3, respectively, to obtain reconstructed vector v_1 to v_3. The reconstruction error e_1 to e_3 associated with reconstructed vector v_1 to v_3 were calculated as follows:

$$e_i = \left|\left|v - v_i\right|\right|^2, \ i = 1, 2, 3. \text{ and } v_i = D_i\alpha_i \tag{7}$$

The coefficient vector α_i was determined using orthogonal matching pursuit based on dictionary D_i. If the reconstruction error was greater than a given threshold R ($21 < R < 36$ in this study), slice N was used to train a new dictionary using methods described by equations (5) and (6).

It is important to note that the mean shift partitioned subareas generated from the training 2D image were only used to train the dictionary. For subsequence slices, each pixel was then assigned to either foreground or the background based on the reconstruction error of the sliding 5 × 5 patch centered to that pixel.

Manifold clustering constraint

Both graph cuts and dictionary learning can be used as standalone methods for segmentation but often resulted in false positivity because of similar imaging properties of nearby organs. In such cases, addition shape constraints were needed to separate organs adjacent to the pancreas.[23] The pancreas is a curvy and elongated organ with significant variation in shapes among individuals. To apply this shape constraint without losing generality, a manifold clustering constraint was imposed to remove abutting organs causing false positivity from the preliminary segmentation results.

Different from the K-mean clustering method that does not use the structure information of data set, manifold clustering reveals the low-dimensional geometry structure of the dataset that is critical to segment the elongated pancreas organ shape.[39-41]

In the study, each segmented region was represented as a manifold. The manifold clustering method was sequentially performed in 3 steps described as manifold generation, manifold distance metric, and manifold clustering. Briefly, in the first step, each manifold generated by the GC or DL segmentation and the manifold map preserved most of the necessary structures for image segmentation such as boundaries of objects. In the second step, manifold distances were computed as the shortest path along the elongated structure. For the MRI, we were given a collection of n data points $\{x_i\}_{i=1}^{n}$ lying in m different manifolds and m randomly initialized clustering centers. The beeline length between x_i and x_j, $L(x_i, x_j)$, was defined: where $dist(x_i, x_j)$ was Euclidean distance between x_i and x_j.

$$L(x_i, x_j) = \begin{cases} dist(x_i, x_j) & \text{if neighboring points and belong to the same binary mask} \\ \infty & \text{otherwise} \end{cases} \qquad (8)$$

The 8 pixels immediately adjacent to the pixel of interest were defined as neighboring pixels.

All possible path beeline lengths L from x_i to x_j composed the set P_{ij}. Thus the manifold distance $M(x_i, x_j)$ between x_i and x_j was determined using:

$$M(x_i, x_j) = \min\{P_{ij}\} \qquad (9)$$

The manifold metric can measure the shortest path along manifold, the path between a pair of points on the same manifold was consisted of shorter beelines between their neighbors, and that the two points on the different manifold should be linked by many longer beelines. In other words, the distance between the points on the different manifold was longer than that on the same manifold.

Therefore, the similarity w_{ij} between voxel x_i and x_j was defined as follows:

$$w_{ij} = \frac{1}{1 + M(x_i, x_j)} \qquad (10)$$

Where i ≠ j, when i = j, $w_{ij} = 1$.

In the last step for manifold clustering, each pixel was assigned to a cluster based on the shortest manifold distance from the pixel to a cluster center. The cluster center was iteratively updated until the change of two iteration clustering result was less than the given threshold R. Note that equation (10) was nonparametric and used without supervision. Neighbor-averaging and hole-filling were performed to refine the segmentation. The results are shown in Fig 1.

To avoid overfitting of the small sample, all segmentation methods were independently performed on individual patients without using them as a collective dataset. Furthermore, the segmentation parameters of GC and DL were tuned on healthy volunteer 1 and then used on all subjects while the MSM and DRLS parameters were individually tuned.

Evaluation of the segmentation

In addition to the automated segmentation, manual reference segmentation was performed by a physician. The segmentation results were evaluated both visually and quantitatively.

To quantitatively analyze the segmentation performance, Dice's similarity index (DI), maximal surface distance, which is better known as the Hausdorff distance[42] and the relative shift between the centers-of-the-weight of the manual and automated contours (SHIFT) were calculated between the automated and manual pancreas segmentation. DI was calculated using

$$DI = 2 \cdot \frac{|V_1 \cap V_2|}{|V_1| + |V_2|} \qquad (11)$$

Where V_1 and V_2 were the binary masks from automated and manual segmentation, respectively.

To calculate Hausdorff distance, the surface points on the automated contours were exhaustively searched to determine the minimal distance from this point to the reference manual contour surface in 3D. The organ center-of-the-weight was calculated to determine the shift between manual and automated segmentation. Statistical analysis was performed using t test on the logit transformation of DI and original values of the Hausdorff distances and shift.

Results

Visual inspection of the segmentation performance

Figure 2 show the segmentation results of a representative imaging slice for a healthy volunteer and a patient. MSM and GC tend to under- or oversegment the pancreas when there is a gradual transition of the imaging intensity to nearby organs such as the stomach. DRLS and DL were more robust than MSM and GC.

The segmentation results grouped by imaging techniques for all subjects are shown in Fig 3. All 4 methods were able to segment the pancreas in the postcontrast VIBE images but MSM clearly resulted in larger errors for noncontrast patient images. DRLS, GC, and DL are visually comparable except for the H2 and P3 HASTE images where DRLS substantially undersegmented the pancreas and GC over-segmented the pancreas.

Quantitative evaluation

There was a significant difference in the pancreas manual reference volume defined in HASTE and VIBE images. The average pancreas volume in the HASTE, VIBE pre-contrast and VIBE post contrast images were 100.9 cm³, 64.74 cm³, and 76.7 cm³, respectively. The difference is largely caused by that the VIBE sequence suppressed the fat signals and showed the non-fat portion of this organ.

Dice's index (DI)

The DIs of automated segmentation results are compared to the manual segmentation in Table 1. DL was superior to the other three methods in 9 out of 12 image sets. DRLS results better overlap with the manual
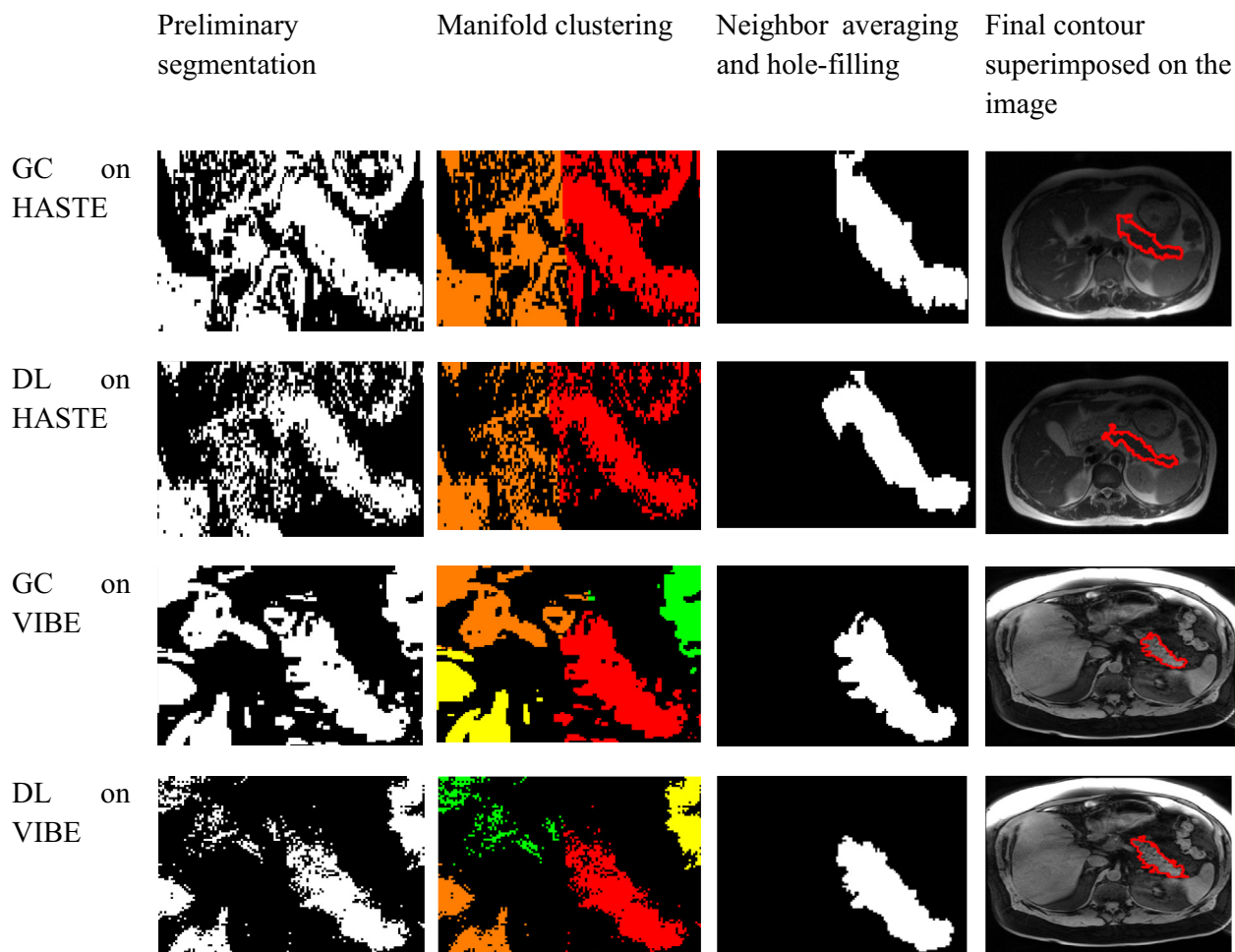
Figure 1 Preliminary segmentation results using GC and DL on VIBE and HASTE images and subsequent refinement using manifold clustering, neighbor averaging, and hole filling. The manifold clustering results; different clusters are shown in different colors. DL, dictionary learning; GC, graph cuts; HASTE, half-Fourier acquisition single-shot turbo spin-echo; VIBE, volumetric interpolated breath-hold examination.

contours for the two healthy volunteer VIBE images. GC resulted in superior segmentation results for a patient HASTE image. On average for all images, DL resulted in more accurate pancreas segmentation (DI = 0.83) than MSM (DI = 0.72), DRLS (DI = 0.80) and GC (DI = 0.78). The segmentation accuracy based on VIBE images were substantially higher than that based on the HASTE images, two of which resulted in lower than 0.7 DI in all methods. Since on average, DL is superior to the other methods, the *P* values were calculated using the DL as the reference. The results show that DL is statistically more accurate than the other 3 methods for DI.

Maximum surface distances and shift of automated contours relative to the manual contours

Table 2 shows the Hausdorff distance and SHIFT calculation results. The best performing Hausdorff distance is between 7.9 mm and 32 mm, indicating large mismatch on at least a small surface of all automated segmentation methods. MSM results were significantly worse than the other three methods. DL resulted in smallest SHIFT error (1.7 mm), followed by DRLS, GC and MSM. Automated segmentation methods failed to segment two challenging HASTE images of H2 and P3 with both low DI and surface distance agreement due to low contrast, signals from fat, the lack of imaging texture and thick slice thickness (7.5 mm).

Figure 4 shows the 3D rendering of the segmentation results of a typical HASTE and VIBE images. DRLS, GC, and DL were able to closely match the manual segmentation result but MSM showed a more substantial deviation, which was consistent with the quantitative analyses. Quantitative measurement such as the Hausdorff distance was sensitive to over- or undersegmentation of small areas such as the one indicated by the arrow in Figure 4B despite overall good estimation of the organ. The
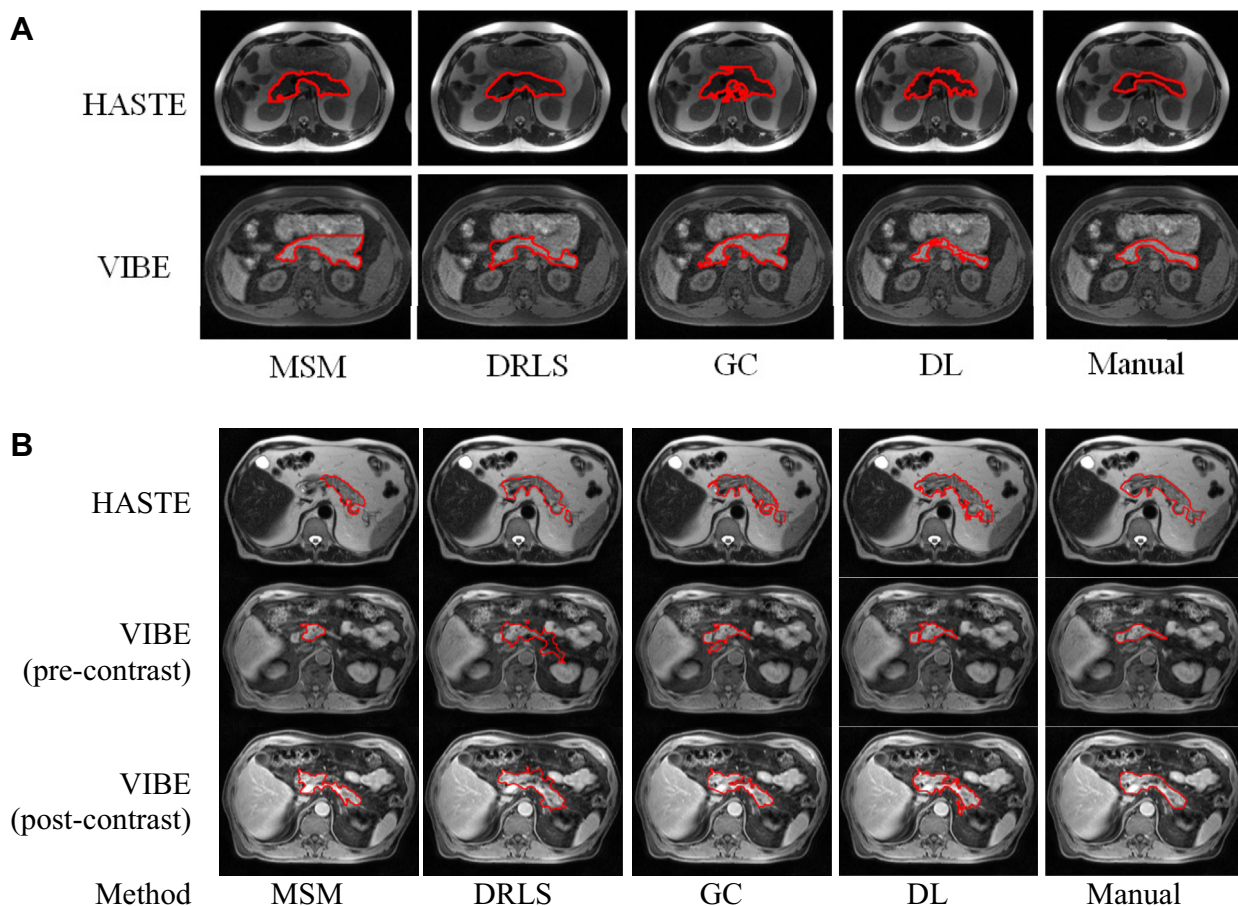
Figure 2   Automated segmentation results for a health volunteer (A) and a patient (B). DL, dictionary learning; DRLS, distance regularized level set; GC, graph cuts; HASTE, half-Fourier acquisition single-shot turbo spin-echo; MSM, mean shift merging; VIBE, volumetric interpolated breath-hold examination.

statistical analysis shows that MSM is significantly inferior but the differences between the other three methods are insignificant.

Other performance comparisons

For the DL and GC, the same segmentation parameters were used for all subjects. The results were overall consistent with occasional large errors as reported in the results. MSM and LS required parameter tuning for individual subjects as the segmentation results were more sensitive to the parameter selection. Because DL, GC, and DRLS were close in performance, it is worth to consider their different initialization conditions and processing times. DL required simplest manual marking on *1* 2D slice and the shortest supervision time on the order of a few seconds. GC requires manual marking of the organ of interest on *every* 2D slice that can take up to 1 minute. DRLS requires accurate contours of the organ of interest on 1 to 2 2D slices, a process that can take more than 1 minute. All automated segmentation methods were

implemented using Matlab (V2013a) on a core i7 computer with 16 GB RAM. With DL, feature extraction was time consuming (121 second), but it only needed to be performed once for the entire image series. Furthermore, the time can be significantly reduced using a graphic processing unit to 0.38 seconds. For online applications, the DL and GC segmentation times were 0.45 seconds and 0.89 seconds, respectively, in comparison to DRLS, which took 7.6 seconds to perform.

Discussion

To the best of our best knowledge, automated pancreas segmentation based on 3D MRI has not been previously reported. In this study, we applied 4 state-of-the-art segmentation methods including a MSM, DRLS, GC and DL method. The last 2 methods were further subjected to a shape constraint using manifold clustering. Among these methods, DRLS, GC, and DRLS showed superior accuracy to the published pancreas segmentation study based on CT images with contrast,[22] indicating the potential
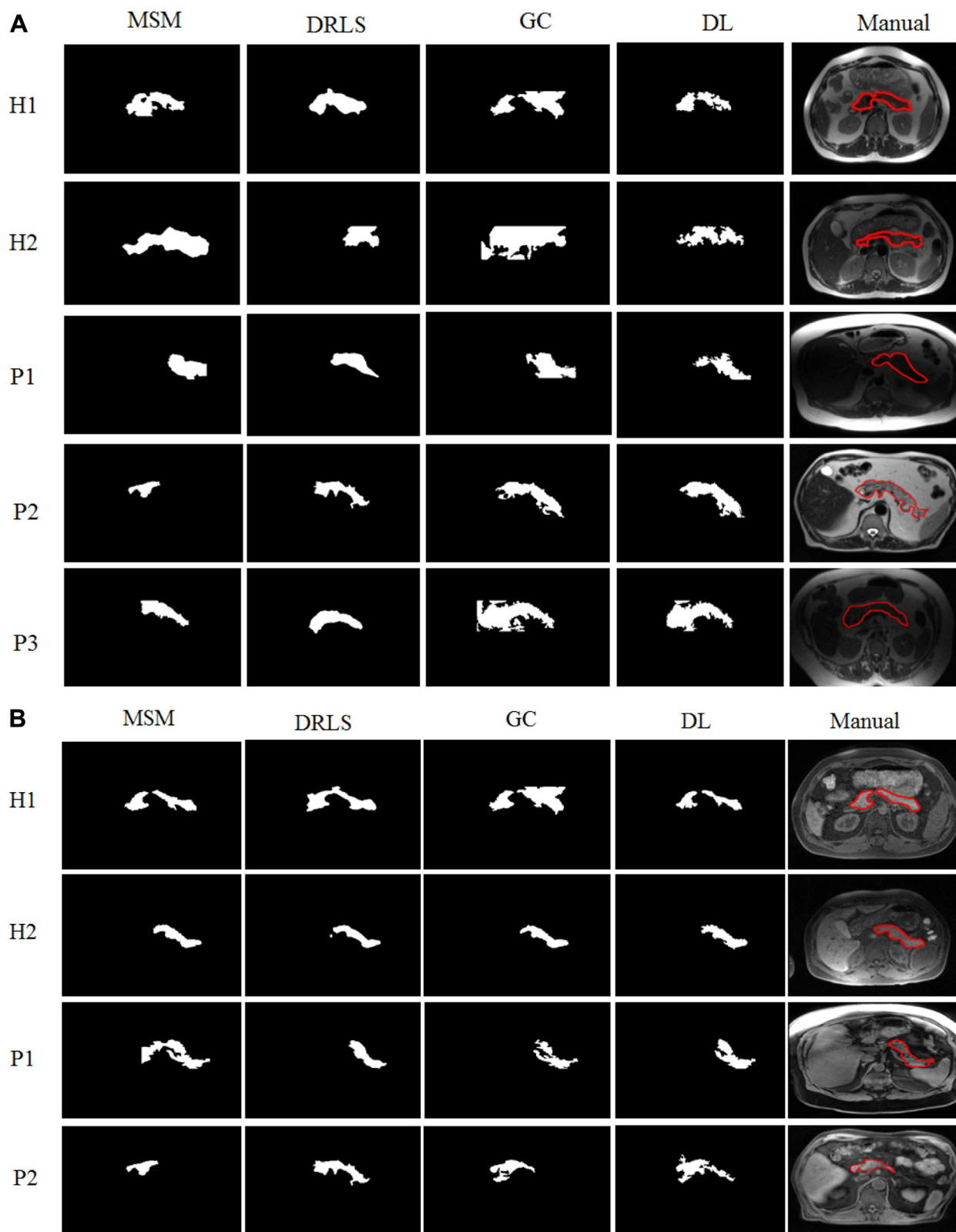
Figure 3 Segmentation comparison for different imaging techniques. Automated segmentation results are shown in binary masks and the manual segmentation results are shown as superimposed contours. (A) Segmentation results for half-Fourier acquisition single-shot turbo spin-echo images. (B) Segmentation results for precontrast VIBE images. (C) segmentation results for post-contrast VIBE images. DL, dictionary learning; DRLS, distance regularized level set; H1, H2, 2 healthy volunteers; MSM, mean shift merging; P1, P2, P3, 3 pancreatic cancer patients; VIBE, volumetric interpolated breath-hold examination.
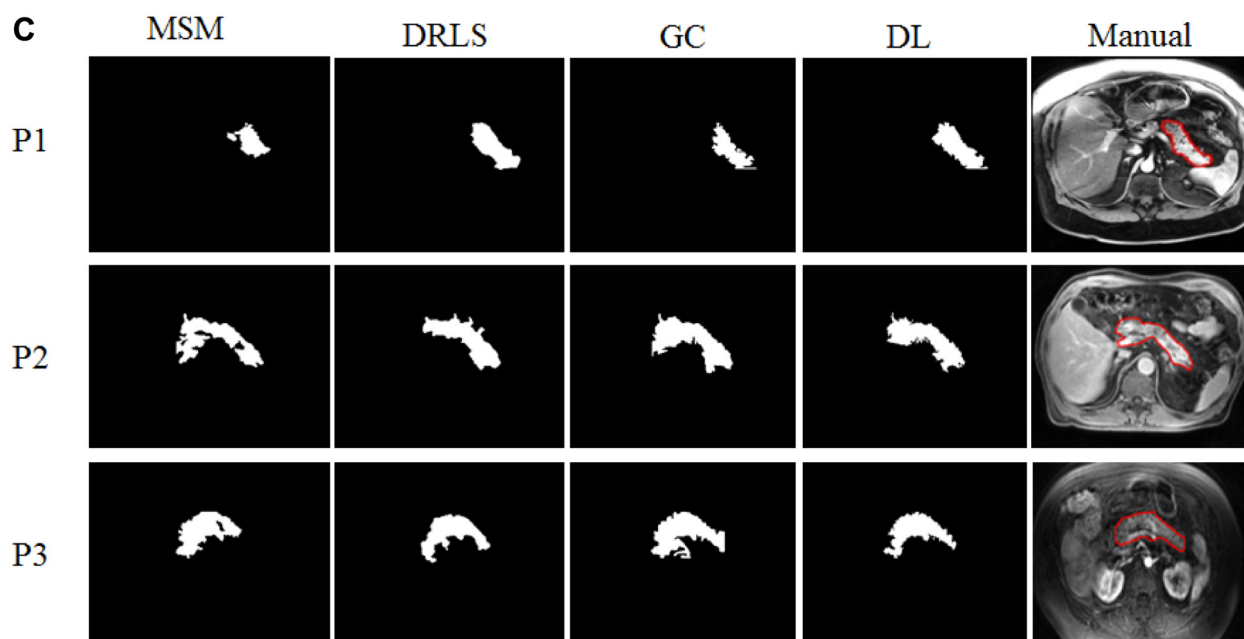
Figure 3   (*continued*).

advantages of using MRI for pancreas image guided radiation therapy. Three metrics were used to evaluate the segmentation results. SHIFT is useful to determine the organ centroid. Hausdorff distance is sensitive to maximal disagreement between the 2 sets of contours. DI describes the volume overlap. In the study, we achieved high DI but large Hausdoff distance is observed. By combining the 3 measurements, one can draw a conclusion that the manual and automated contours generally agree with each other but there are very small segments of contours, such as the one shown in Fig 4, that disagree. A relative shift in the contours of <2 mm and volume overlap index greater than 0.83 was achievable using the best segmentation method. The accuracy is useful to location for pancreas motion tracking aiming at millimeter resolution but may be insufficient for adaptive radiation therapy that adjusts the planning target volume shapes in real time considering the large Hausdorff distance.

MSM performed substantially worse than the other 3 methods, indicating that although this method is valuable for preliminary segmentation, it alone is insufficient to segment complex abdominal organs such as the pancreas. The differences among these 3 methods are relatively subtle. DL appears to be more accurate at defining the irregular organ boundaries and the DRLS is more robust at avoiding large surface distance errors as shown by its smaller Hausdorff distance measurement. Both GC and DL relied on morphological refinement using manifold cluster. DL and DRLS were shown to be slightly more robust and accurate than GC.

Our study also revealed that the T1-weighted VIBE images were better suited for automated segmentation than the T2-weighted HASTE images. Automated segmentation methods not only tended to oversegment a larger pancreas volume based on the HASTE image compared with the manual contours, they systematically failed to segment 2 HASTE images, indicating that fine tuning of each method may be insufficient to overcome this problem. There are several reasons contributing to the difference. First, VIBE images' richer texture information helped methods relying in textures. Second, VIBE image resolution is higher than that of HASTE particularly in the slice thickness direction, resulting in less abrupt changes in the pancreas 2D morphology between slices. Last, suppression of fat signals in VIBE contributed to a better organ boundary definition.

In this study, the performance of automated segmentation was not significantly affected by whether the subject was a volunteer or a patient who has disease-altered morphology. The accuracy of automated segmentation based on precontrast VIBE was only slightly inferior to that of the postcontrast VIBE images. This is important because the contrast may not be used on a daily basis for image guided radiation therapy.

There are aspects other than the geometrical matching between the automated and manual segmentation to consider. All automated segmentation methods in this study relied on some level of human supervision that is needed only at the beginning of imaging acquisition to allow automated segmentation of all subsequent volumes. The levels of supervision vary among segmentation methods. MSM/DL requires the least intervention of labeling on 1 slice of the image, GC requires labeling of all slices, and DRLS requires complete contour on one slice. For applications such as gated radiation therapy, time is essential. For a patient who breathes at 20 beats/minute and a

Table 1   Average DI comparison of 3 methods on all slices

	Imaging technique	MSM	DRLS	GC	DL
H1	HASTE	0.6918	0.7566	0.6902	**0.7727**
	VIBE pre	0.6542	**0.7879**	0.6557	0.7323
H2	HASTE	0.5345	0.5785	0.5543	**0.5941**
	VIBE pre	0.7920	**0.8300**	0.7960	0.8094
P1	HASTE	0.5596	0.7796	0.6232	**0.8766**
	VIBE pre	0.7345	0.8499	0.8466	**0.8599**
	VIBE post	0.6541	0.8726	0.8909	**0.8960**
P2	HASTE	0.7740	0.8762	**0.9176**	0.8957
	VIBE pre	0.8595	0.8851	0.8883	**0.9040**
	VIBE post	0.8234	0.8419	0.8911	**0.9167**
P3	HASTE	0.6781	0.6385	0.6928	**0.6990**
	VIBE post	0.8228	0.7521	0.7957	**0.8354**
Average of all 5 subjects	HASTE	0.6476	0.7259	0.6956	**0.7676**
	VIBE pre	0.7601	**0.8382**	0.7967	0.8264
	VIBE post	0.7668	0.8222	0.8592	**0.8827**
	Total	0.7248	0.7954	0.7838	**0.8256**
P value compared with DL based on logit (DI)		.002	.03	.049	—

The best performance for each raw is shown in bold face.
DI, Dice's index; DL, dictionary learning; DRLS, distance regularized level set; GC, graph cuts; HASTE, half-Fourier acquisition single-shot turbo spin-echo; H1, H2, 2 healthy volunteers; MSM, mean shift merging; P1, P2, P3, 3 pancreatic cancer patients; VIBE, volumetric interpolated breath-hold examination.

minimum of 5 samples per breathing cycle, the image acquisition and automated segmentation needs to be performed in 0.6 second. Because human supervision is only needed at the beginning of image acquisition, the more important factor is the time needed for 3D MRI acquisition and automated segmentation. Currently, 3D MRI acquisition takes ~20 seconds for an abdominal volume. Many investigators are working on accelerating the acquisition. For example, the feasibility of an order of magnitude acceleration using low rank decomposition was demonstrated.[24] For the more promising DL segmentation, the time to build the dictionary can be reduced to 0.38 seconds using a graphic processing unit and it needs to be done only once. Afterwards, the time to segment a volume is 0.45 seconds on a single processor computer. The time may be further reduced using a multiprocessor computer and optimization of the code for tumor motion tracking. An additional advantage of DL is that the segmentation

Table 2   Maximum surface distances and shift of the automated segmentation compared with the manual segmentation

	Imaging technique	Hausdorff distance (mm)				SHIFT (mm)			
		MSM	DRLS	GC	DL	MSM	DRLS	GC	DL
H1	HASTE	26.3	13.2	19.5	**12.8**	3.5	1.7	2.4	**1.5**
	VIBE	37.8	**16.9**	36.5	30.7	2.3	**1.1**	3.5	2.3
H2	HASTE	38.4	**26.4**	35.8	35.9	2.1	**1.6**	2.6	2.7
	VIBE	41.5	**14.2**	25.5	41.4	2.6	1.4	2.4	**1.2**
P1	HASTE	41.9	14.7	33.2	**8.5**	2.3	1.4	4.5	**1.3**
	VIBE pre	61.8	12.1	**11.7**	11.7	2.9	0.6	1.8	**0.5**
	VIBE post	53.2	11.0	**7.9**	15.7	4.2	0.9	1.6	**0.8**
P2	HASTE	46.0	32.3	21.2	**16.5**	2.5	1.2	1.7	**1.1**
	VIBE pre	32.9	**20.3**	32.3	29.1	1.4	1.8	1.9	**1.2**
	VIBE post	36.4	27.4	24.9	**24.7**	3.1	2.4	1.6	**1.3**
P3	HASTE	42.1	**32.4**	45.7	45.2	2.6	3.4	4.3	**2.3**
	VIBE post	24.8	27.4	38.4	**18.7**	1.8	2.5	2.3	**1.7**
Average		40.3	**20.7**	27.7	24.2	2.6	1.7	2.3	**1.5**
P value compared with DL		.008	.31	.28	—	.01	.075	.0024	—

The best performers are shown in bold face.
DL, dictionary learning; DRLS, distance regularized level set; GC, graph cuts; HASTE, half-Fourier acquisition single-shot turbo spin-echo; H1, H2, 2 healthy volunteers; MSM, mean shift merging; P1, P2, P3, 3 pancreatic cancer patients; SHIFT, shift of the center of the organ; VIBE, volumetric interpolated breath-hold examination.
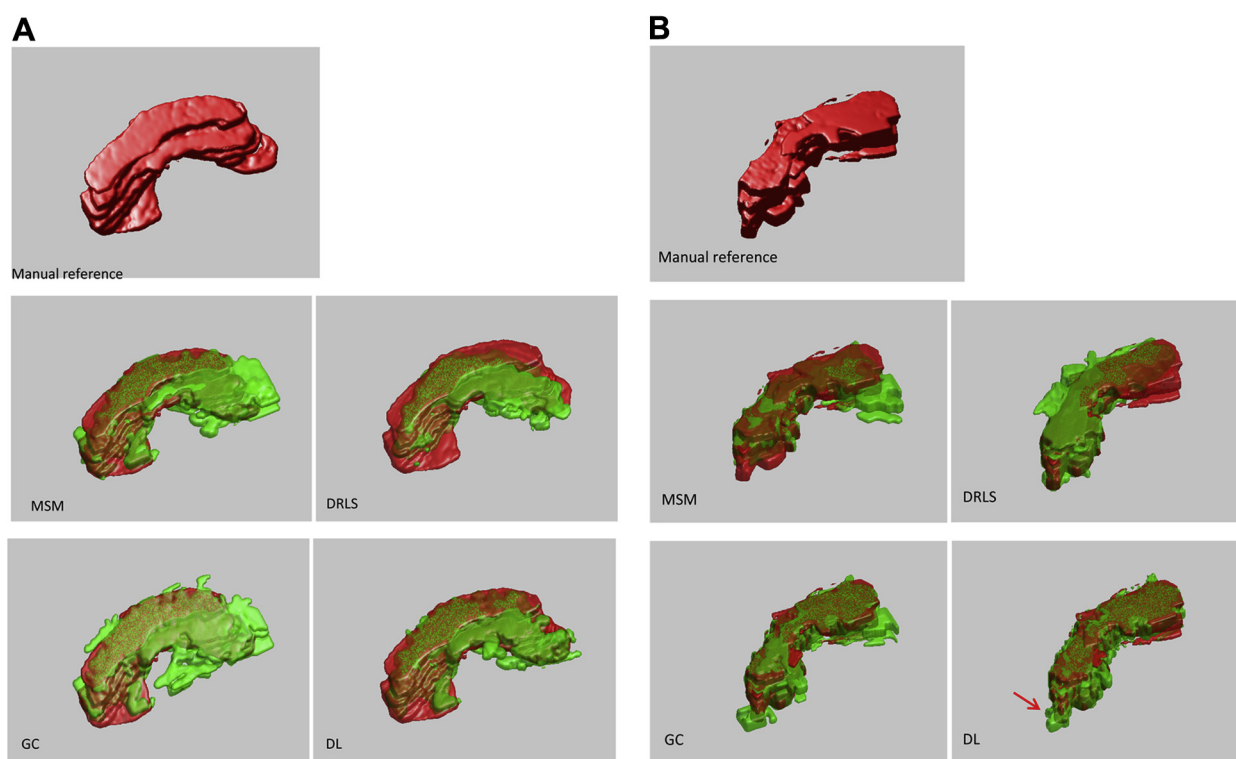
Figure 4   (A) A 3D rendering of the pancreas contour based on a half-Fourier acquisition single-shot turbo spin-echo image. (B) A 3D rendering of the pancreas contour based on a volumetric interpolated breath-hold examination image. DL preserves more details in the organ boundaries but occasionally includes abutting tissues as shown denoted by the arrow. DL, dictionary learning; DRLS, distance regularized level set; GC, graph cuts; MSM, mean shift merging.

parameters appear to be robust for the same MRI sequence without requiring adjustment for different image volumes, which is desirable for online segmentation applications.

A different approach for automated segmentation is using deformable image registration to propagate an initial set of manual segmentations. However, the accuracy of abdominal MRI deformable registration has not been well-studied. It will be a topic of interest to compare the 2 approaches, particularly when real time dynamic volumetric MRI becomes available.

There are several limitations in the study. First, the sample size is small. More subjects are needed to show the robustness of segmentation. Second, to avoid overfitting, the segmentation parameters were not tuned based on the collective dataset. By collectively tuning the parameters on a larger dataset, the accuracy may be improved. Third, geometrical metrics were used to evaluate the segmentation accuracy. DI, Hausdorff distance, and SHIFT each evaluates an aspect of segmentation but ultimately, the accuracy should be evaluated by the radiation dosimetry.

Conclusions

Our study demonstrated potential feasibility of automated segmentation of the pancreas on MRI scans with minimal human supervision at the beginning of imaging acquisition. The achieved accuracy is promising for organ localization. We showed that a volume overlap index greater than 0.83 are achievable on T1-weighted VIBE MRI scans comparing the automated and manual segmentation results and error less than 2 mm to identify the center of the contour on all images using dictionary learning. Considering the computational speed advantage and the low human supervision requirement, DL is the preferred segmentation method for potential real-time MRI segmentation. Currently, the tested methods are less robust to segment the T2-weighted HASTE MRI. A larger patient cohort is needed to test the robustness of the automated segmentation methods.

References

1. Bjerregaard JK, Mortensen MB, Jensen HA, Nielsen M, Pfeiffer P. Prognostic factors for survival and resection in patients with initial nonresectable locally advanced pancreatic cancer treated with chemoradiotherapy. *Int J Radiat Oncol Biol Phys*. 2012;83: 909-915.
2. Krechler T, Horejs J, Ulrych J, et al. [Current status of pancreatic cancer diagnosis]. *Cas Lek Cesk*. 2011;150:587-593.
3. Koong AC, Christofferson E, Le QT, et al. Phase II study to assess the efficacy of conventionally fractionated radiotherapy followed by a stereotactic radiosurgery boost in patients with locally advanced pancreatic cancer. *Int J Radiat Oncol Biol Phys*. 2005;63:320-323.

4. Koong AC, Le QT, Ho A, et al. Phase I study of stereotactic radiosurgery in patients with locally advanced pancreatic cancer. *Int J Radiat Oncol Biol Phys*. 2004;58:1017-1021.

5. Schellenberg D, Kim J, Christman-Skieller C, et al. Single-fraction stereotactic body radiation therapy and sequential gemcitabine for the treatment of locally advanced pancreatic cancer. *Int J Radiat Oncol Biol Phys*. 2011;81:181-188.

6. Yovino S, Maidment BW 3rd, Herman JM, et al. Analysis of local control in patients receiving IMRT for resected pancreatic cancers. *Int J Radiat Oncol Biol Phys*. 2012;83:916-920.

7. Barney BM, Olivier KR, Macdonald OK, Fong de Los Santos LE, Miller RC, Haddock MG. Clinical outcomes and dosimetric considerations using stereotactic body radiotherapy for abdominopelvic tumors. *Am J Clin Oncol*. 2012;35:537-542.

8. Onishi H, Ozaki M, Kuriyama K, et al. Stereotactic body radiotherapy for metachronous multisite oligo-recurrence: A long-surviving case with sequential oligo-recurrence in four different organs treated using locally radical radiotherapy and a review of the literature. *Pulm Med*. 2012;2012:713073.

9. Zakharova OP, Karmazanovsky GG, Egorov VI. Pancreatic adenocarcinoma: Outstanding problems. *World J Gastrointest Surg*. 2012;4:104-113.

10. Liu F, Erickson B, Peng C, Li XA. Characterization and management of interfractional anatomic changes for pancreatic cancer radiotherapy. *Int J Radiat Oncol Biol Phys*. 2012;83:e423-e429.

11. Gwynne S, Wills L, Joseph G, John G, Staffurth J, Hurt C, Mukherjee S. Respiratory movement of upper abdominal organs and its effect on radiotherapy planning in pancreatic cancer. *Clin Oncol (R Coll Radiol)*. 2009;21:713-719.

12. van der Geld YG, Senan S, van Sornsen de Koste JR, Verbakel WF, Slotman BJ, Lagerwaard FJ. A four-dimensional CT-based evaluation of techniques for gastric irradiation. *Int J Radiat Oncol Biol Phys*. 2007;69:903-909.

13. Yang W, Fraass BA, Reznik R, et al. Adequacy of inhale/exhale breathhold CT based ITV margins and image-guided registration for free-breathing pancreas and liver SBRT. *Radiat Oncol*. 2014; 9:11.

14. Lens E, van der Horst A, Kroon PS, et al. Differences in respiratory-induced pancreatic tumor motion between 4D treatment planning CT and daily cone beam CT, measured using intratumoral fiducials. *Acta Oncol*. 2014;53:1257-1264.

15. Minn AY, Schellenberg D, Maxim P, et al. Pancreatic tumor motion on a single planning 4D-CT does not correlate with intra-fraction tumor motion during treatment. *Am J Clin Oncol*. 2009;32: 364-368.

16. Murphy JD, Christman-Skieller C, Kim J, Dieterich S, Chang DT, Koong AC. A dosimetric model of duodenal toxicity after stereotactic body radiotherapy for pancreatic cancer. *Int J Radiat Oncol Biol Phys*. 2010;78:1420-1426.

17. Crijns SP, Raaymakers BW, Lagendijk JJ. Proof of concept of MRI-guided tracked radiation delivery: Tracking one-dimensional motion. *Phys Med Biol*. 2012;57:7863-7872.

18. Santos DM, St Aubin J, Fallone BG, Steciw S. Magnetic shielding investigation for a 6 MV in-line linac within the parallel configuration of a linac-MR system. *Med Phys*. 2012;39:788-797.

19. Cervino LI, Du J, Jiang SB. MRI-guided tumor tracking in lung cancer radiotherapy. *Phys Med Biol*. 2011;56:3773-3785.

20. Kok JG, Raaymakers BW, Lagendijk JJ, Overweg J, de Graaff CH, Brown KJ. Installation of the 1.5 T MRI accelerator next to clinical accelerators: Impact of the fringe field. *Phys Med Biol*. 2009;54: N409-N415.

21. Raaijmakers AJ, Raaymakers BW, Lagendijk JJ. Magnetic-field-induced dose effects in MR-guided radiotherapy systems: Dependence on the magnetic field strength. *Phys Med Biol*. 2008;53: 909-923.

22. Chu C, Oda M, Kitasaka T, et al. Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images. *Med Image Comput Comput Assist Interv*. 2013;16:165-172.

23. Gou S, Wu J, Liu F, et al. Feasibility of automated pancreas segmentation based on dynamic MRI. *Br J Radiol*. 2014;87:20140248.

24. Sarma M, Hu P, Rapacchi S, et al. Accelerating dynamic magnetic resonance imaging (MRI) for lung tumor tracking based on low-rank decomposition in the spatial-temporal domain: A feasibility study based on simulation and preliminary prospective undersampled MRI. *Int J Radiat Oncol Biol Phys*. 2014;88:723-731.

25. Bajcsy M, Mintz M, Liebman E. A common framework for edge detection and region growing. *Univ Penn GRASP Lab Technol*. 1986;Report 6.

26. Pavlidis T, Liow YT. Integrating region growing and edge-detection. *IEEE Trans Patt Anal Machine Intell*. 1990;12:225-233.

27. Xuan JH, Adali T, Wang Y. Segmentation of magnetic resonance brain image: Integrating region growing and edge detection. *Int Conf Image Proc Proceed*. 1995;I-III:C544-C547.

28. Ning JF, Zhang L, Zhang D, Wu CK. Interactive image segmentation by maximal similarity based region merging. *Patt Recognit*. 2010;43:445-456.

29. Osher S, Fedkiw RP. Level set methods: An overview and some recent results. *J Comp Phys*. 2001;169:463-502.

30. Sethian JA. A fast marching level set method for monotonically advancing fronts. *Proc Natl Acad Sci U S A*. 1996;93:1591-1595.

31. Malladi R, Sethian JA, Vemuri BC. Shape modeling with front propagation—A level set approach. *IEEE Trans Patt Anal Machine Intell*. 1995;17:158-175.

32. Li CM, Xu CY, Gui CF, Fox MD. Distance regularized level set evolution and its application to image segmentation. *IEEE Trans Image Proc*. 2010;19:3243-3254.

33. Boykov Y, Funka-Lea G. Graph cuts and efficient N-D image segmentation. *Int J Computer Vision*. 2006;70:109-131.

34. Massoptier L, Casciaro S. Fully automatic liver segmentation through graph-cut technique. 2007 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2007; Vols. 1-16:5243-5246.

35. Alba X, Figueras IVRM, Lekadir K, Tobon-Gomez C, Hoogendoorn C, Frangi AF. Automatic cardiac LV segmentation in MRI using modified graph cuts with smoothness and interslice constraints. *Magn Reson Med*. 2014;72:1775-1784.

36. Longnecker D. Anatomy and Histology of the Pancreas. Pancreapedia: Exocrine Pancreas Knowledge Base, 2014. http://dx.doi.org/10.3998/panc.2014.3.

37. Gao Y, Liao S, Shen D. Prostate segmentation by sparse representation based classification. *Med Phys*. 2012;39:6372-6387.

38. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Proc*. 2006;54:4311-4322.

39. Erdt M, Kirschner M, Drechsler K, Wesarg S, Hammon M, Cavallaro A. Automatic pancreas segmentation in contrast enhanced ct data using learned spatial anatomy and texture descriptors. 2011 8th *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2011:2076-2082.

40. Shimizu A, Kimoto T, Kobatake H, Nawano S, Shinozaki K. Automated pancreas segmentation from three-dimensional contrast-enhanced computed tomography. *Int J Comp Assist Radiol Surg*. 2010;5:85-98.

41. Jiang HY, Tan HQ, Fujita H. A hybrid method for pancreas extraction from CT image based on level set methods. *Comp Math Methods Med*. 2013:479516.

42. Sim DG, Kwon OK, Park RH. Object matching algorithms using robust Hausdorff distance measures. *IEEE Trans Image Process*. 1999;8:425-429.