Computer Vision and Image Understanding 113 (2009) 1053–1063

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Search strategies for shape regularized active contour

Tianli Yu [a], Jiebo Luo [b,*], Narendra Ahuja [c]

[a] Like.com, 3 Waters Park Drive San Mateo, CA 94403, USA
[b] Kodak R&D Labs, Foundation Science Center, Eastman Kodak Company, Rochester, NY 14650, USA
[c] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 9 October 2006
Accepted 14 April 2008
Available online 3 June 2008

Keywords:
Active contour
Object segmentation
Dynamic programming
Local optimization
Nonlinear shape models

ABSTRACT

Nonlinear shape models have been shown to improve the robustness and flexibility of contour-based object segmentation when there are appearance ambiguities between the object and the background. In this paper, we focus on a new search strategy for the shape regularized active contour (ShRAC) model, which adopts existing nonlinear shape models to segment objects that are similar to a set of training shapes. The search for optimal contour is performed by a coarse-to-fine algorithm that iterates between combinatorial search and gradient-based local optimization. First, multi-solution dynamic programming (MSDP) is used to generate initial candidates by minimizing only the image energy. In the second step, a combination of image energy and shape energy is minimized starting from these initial candidates using a local optimization method and the best one is selected. To generate diverse initial candidates while reducing invalid shapes, we apply two pruning methods to the search space of MSDP. Our search strategy combines the advantages of global combinatorial search and local optimization, and has shown excellent robustness to local minima caused by distracting suboptimal solutions. Experimental results on segmentation of different anatomical structures using ShRAC, as well as preliminary results on human silhouette segmentation are provided.

1. Introduction

The segmentation of anatomical structures is often a critical component of medical imaging systems, such as a computer-aided diagnosis (CAD) system or a patient information system (PIS). In chest radiography, researchers have developed numerous methods for segmenting the lung fields, rib cage, heart, clavicles, blood vessels, as well as abnormal structures such as lung nodules. However, given the projective nature of chest radiography, superimposed anatomical structures make images complicated and challenging for both radiologists and computerized systems. On the other hand, the limited domain of the problem provides an opportunity to incorporate prior knowledge of the shape, and in some cases the appearance, of the anatomical structures of interest. This prior knowledge, often in the form of object shape and/or appearance model, is used during the search for similar structures in the image. Earlier researches focus on rule-based reasoning and hand-crafted shape models [1], while more recent approaches use statistical learning tool to train the object model from a set of examples.

Landmark based active shape model (ASM) [2] is a successful method for object localization and segmentation. It uses linear shape and appearance models extracted by principle components analysis (PCA). Linear models have the problem that they admit invalid shapes when the shape distribution is in fact nonlinear. Multiple models are often needed to span a nonlinear shape distribution while excluding invalid shapes. The training of ASM is also quite tedious. A set of landmarks need to be identified and correctly matched for each example.

It is desirable to incorporate nonlinear shape priors into segmenting different anatomical structures (e.g., lung fields, clavicles, ribs) in medical images, which can remove the need to train multiple linear models. For the specific problem of lung field segmentation, it is desired that the segmentation algorithm can automatically adapt and converge to the correct shape, without any model selection in advance, e.g., to decide whether to use the left or right lung field model (Fig. 1). In addition, the method should also be robust to noise and clutter that are commonly seen in medical images.

Nonlinear shape priors have been incorporated into nonlinear ASM [3] and variational image segmentation with nonlinear shape statistics [4]. These algorithms have been shown to be more flexible and powerful in model building and therefore more successful in segmenting objects with nonlinear shape variations. However, they also pose many difficulties to the design of a search strategy that balances the use of shape prior constraints and actual image structures to find a globally optimal solution.

* Corresponding author.
E-mail addresses: tianli@like.com (T. Yu), jiebo.luo@kodak.com (J. Luo), ahuja@vision.ai.uiuc.edu (N. Ahuja).

1077-3142/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.cviu.2008.04.002

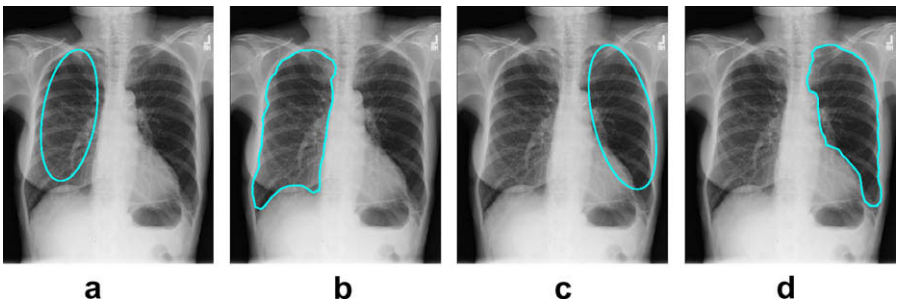

Fig. 1. Starting with the same elliptical shape, a segmentation algorithm should converge to the correct shape without model selection (e.g., to adapt to the left or the right lung field). Here we show the results of the proposed ShRAC search strategy: (a) Initial contour on the right lung. (b) Final contour on the right lung. (c) Initial contour on the left lung (note that it is the same shape as in (a)). (d) Final contour on the left lung.

In the work by Cremers et al. [4], a Mumford-Shah functional type image energy and a shape energy derived from kernel space density estimation (KSDE) are combined to construct an energy function for segmentation. The KSDE-based shape energy enforces object-specific nonlinear shape constraint, in contrast to general smoothness constraint in conventional snakes [5,6]. However, this shape energy term makes it difficult to minimize the energy function globally. Instead, gradient-descent is used in their work, which often converges to a local minimum.

Algorithms developed under the framework of ASM [2] address this problem by not minimizing a single energy function. Instead, the search iterates between two steps. In the first step, ASM searches for the best landmarks. This is equivalent to minimizing an image energy term. In the other step, the contour is regularized to the space of training shapes, and can be viewed as minimizing a shape energy term. Similar approaches are also used in medical image segmentation [7]. Although the alternating process provides certain capabilities for the algorithm to jump out of a local minimum and find a better solution, the lack of a unified energy function may introduce oscillations when trying to optimized two separate targets. Therefore, a more robust method is needed to avoid potential oscillations.

In this paper, we focus on a novel search strategy for the shape regularized active contour (ShRAC) [8], which incorporates nonlinear shape priors through kernel space density estimation, similar to [4]. Compare to [8], the new search strategy combines the advantages of combinatorial search, which provides global optimality, and gradient-based local optimization. It could avoid many problems of separately optimizing two type of energy functions as in [8], such as oscillation and bias toward the image energy. In particular, we propose to first use multi-solution dynamic programming (MSDP) to minimize the image energy as the initial candidate selection step, and then use a local optimization method that eventually minimizes a combination of the image energy and shape energy determined by the prior shape models. In addition, a new search space pruning process is used to ensure the diversities of the selected candidates and reduce the number of invalid shapes. Our combined search strategy provides improved robustness to image noise and various distracting structures that present in medical images.

Compare to [4,2], as well as other shape model constrained segmentation developed under the level set framework [9–11], our method can be thought as a multi-path search strategy, where multiple different contours are efficiently generated and evolved in the search space to find the optimal solution. Most of the previous methods mentioned above only evolve a single initial contour until convergence.

Although the main problem solved in this paper is in the context of medical image segmentation. We provide some preliminary results on human silhouette segmentation at the end of this paper to explore the feasibility of our approach to other domains.

This paper is organized as follows: Section 2 describes the formulation of ShRAC. Section 3 presents the iterative search strategy. Section 4 uses experiments in three segmentation tasks to demonstrate the robustness and versatility of ShRAC, and Section 5 gives conclusions and outlines related future work.

2. The ShRAC model

ShRAC [8] can be considered as an extension of the conventional active contour by replacing the smoothness term with a nonlinear shape distance measure [4]. In particular, a discrete formulation is used so that efficient combinatorial search methods can be applied.

We try to find a contour c that minimizes an energy function $E(c)$, which is a weighted sum of two terms:

$$E(c) = E_{\text{shape}}(c) + wE_{\text{image}}(c) \tag{1}$$

where both $E_{\text{image}}(c)$ and $E_{\text{shape}}(c)$ depends on the shape of the contour c. $E_{\text{image}}(c)$ is the image energy term, and $E_{\text{shape}}(c)$ is the shape energy term that measures the similarity between the c and the training shapes. The image energy in ShRAC is chosen to be the intensity differences on two side of a contour. This is because in medical images such as X-ray, the appearance of an object may often be changed by other superimposed structures, but the edges on the object boundary are largely preserved. $E_{\text{image}}(c)$ can be computed as the integration of local edge strength $h_e(\cdot)$ along the contour.

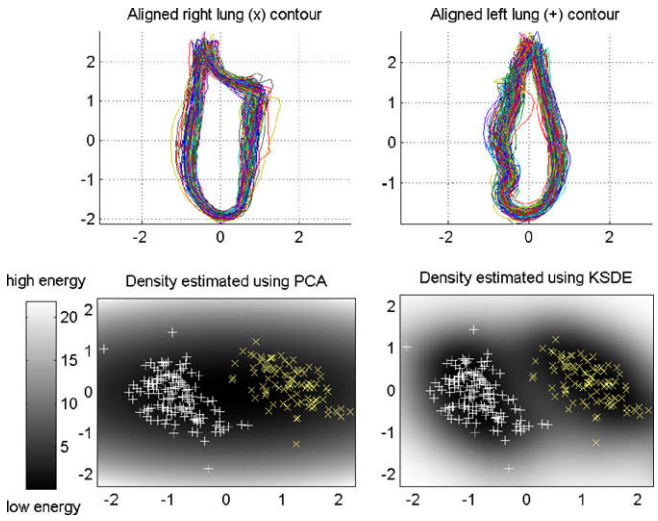

Fig. 2. Comparison of the shape energy functions learned by PCA (lower left) and KSDE (lower right) from the training shapes of both left (+) and right (x) lungs.

$$E_{\text{image}}(c) = \frac{\int h_e[v(s)]\mathrm{d}s}{L(c)} \tag{2}$$

where $v(s)$ can be viewed as a parametric representation of points on the contour. $L(c)$ is the length of the contour which is used as a normalization term. The term $h_e(\cdot)$ can be computed as the absolute or signed intensity difference between pixels on the two sides of the contour. In many medical applications, such as lung field or rib segmentation, a signed intensity difference will give better performance since the intensity differences generally have the same sign along the contour.

To use efficient combinatorial optimization methods, we represent the contour as a set of control points $c = \{v_i | i = 1, \ldots, N\}$ connected sequentially by line segments. Eq. (2) can be re-written in a discrete form as:

$$E_{\text{image}}(c) = \frac{\sum_{i=0}^{n-1} h_e[v_i\vec{v}_{i+1}] * l(v_i\vec{v}_{i+1})}{L(c)} \tag{3}$$

where $h_e[v_i\vec{v}_{i+1}]$ represents the average edge strength along $v_i\vec{v}_{i+1}$ and $l(v_i\vec{v}_{i+1})$ is the length of $v_i\vec{v}_{i+1}$. If the control points are equally spaced, $l(v_iv_{i+1})/L(c)$ becomes a constant, and can be absorbed into the weight w. This gives us the following energy function:

$$E(c) = E_{\text{shape}}(c) + w' \sum_{i=0}^{n-1} h_e[v_i\vec{v}_{i+1}] \tag{4}$$

We adopt the shape energy used in [4], which is a similarity measure based on kernel space density estimation (KSDE). KSDE computes the covariance matrix of the training data using a nonlinear kernel function instead of the usual dot product, which gives an energy function like this:

$$E_{\text{shape}}(c) = \sum_{j=1}^{r} \left(\sum_{i=1}^{m} \alpha_i^j \tilde{k}(c_i, c) \right)^2 \cdot (\lambda_j^{-1} - \lambda_\perp^{-1}) + \lambda_\perp^{-1} \cdot \tilde{k}(c, c) \tag{5}$$

In Eq. (5), α_i^j is the jth eigenvector of the centered kernel matrix, $\tilde{k}(\cdot,\cdot)$ is the centered kernel function, and c_i is the ith training example. Furthermore, $\lambda_j = (1/m)\tilde{\lambda}_j$, where $\tilde{\lambda}_j$ is the jth eigen value of the centered kernel matrix, m is the number of training examples, and $\lambda_\perp$ is a constant value chosen to replace all the smaller eigenvalues.

The KSDE model allows us to build a shape energy term that contains nonlinear variations. For example, we can put together training shapes from both left and right lungs and build a single model to represent the distribution of all the possible lung shapes.

Fig. 2 contrasts the performance of PCA and KSDE in modeling the shape energy from a set of lung shapes. Both left and right lung shapes (Fig. 2, first row) are aligned into a common coordinate space and represented as a sequence of control points. The learned Mahalanobis distances from PCA and from KSDE using (5) are plotted in the second row, along the axes spanned by the eigenvectors that correspond to the largest two PCA eigenvalues. Clearly, KDSE shows two clusters, separated by a high-energy ridge (in light gray) in the shape space. It gives a better characterization of the training shape distribution than the single cluster energy from PCA, which actually assigns the lowest energy (highest likelihood) to the invalid shapes in between the two clusters.

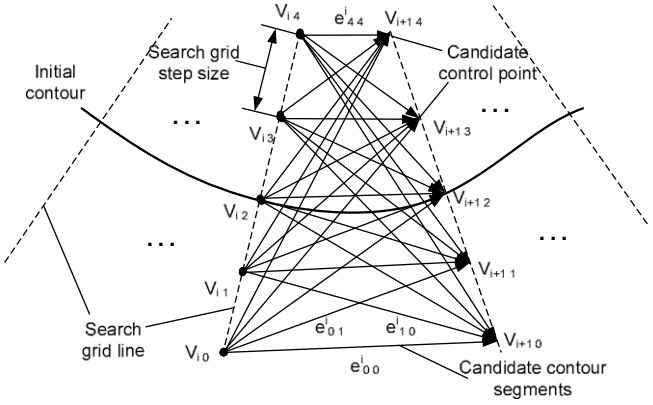

Fig. 4. A MSDP search space constructed from an initial contour using the straight-line grid. $\{V_{ij}\}$ is the candidate set for control point i and e^i_{jk} are candidate contour segments between control points i and $i+1$.

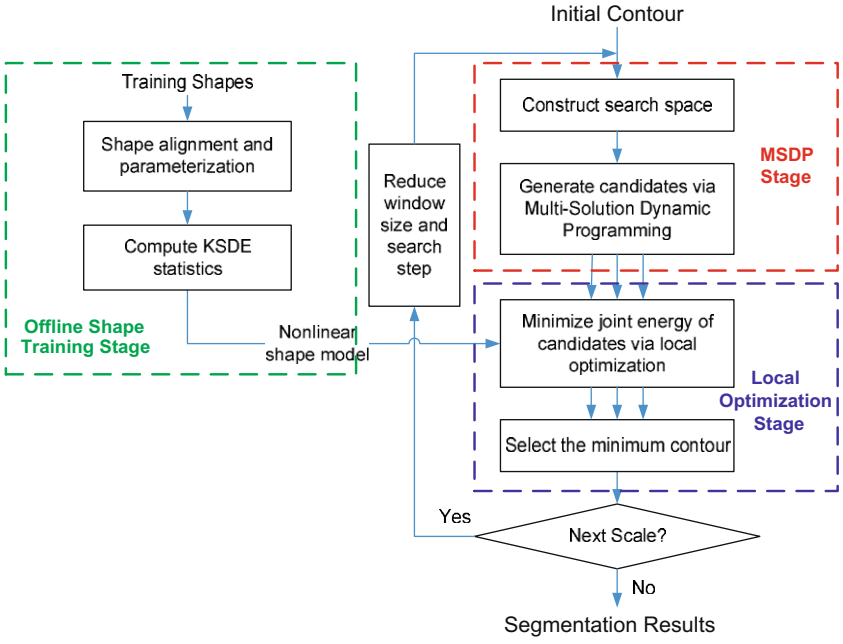

Fig. 3. The flow diagram of the proposed search strategy for ShRAC.

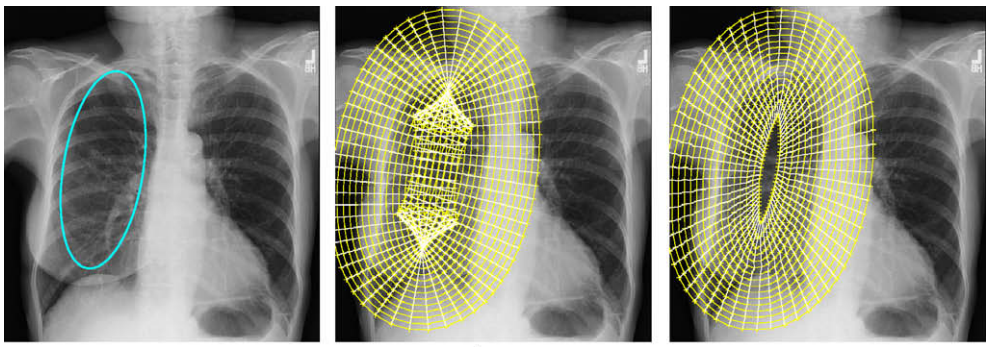

Fig. 5. Different search spaces generated from an initial contour. (a) Initial elliptical contour. (b) Search space generated using the straight-line method (directions are normal to the contour). (c) Search space generated using the distance transform method. For clarity purpose, only segments connecting two corresponding candidate points are shown.

3. Search strategy for ShRAC

3.1. Combinatorial search vs. local optimization

It is easy to see that the image energy term in the energy function (4) has a good property: it can be expressed as a sum of functions of two consecutive control points. Functions of this form can be optimized efficiently using combinatorial search methods such as Dynamic Programming. Dynamic programming guarantees to find a global minimum within polynomial time [12]. This ensures the robustness of the algorithm to initial conditions and local minima.

The shape energy term, on the other hand, has a more complex form. The addition of this term makes the whole function no longer decomposable as before, and therefore cannot be optimized by dynamic programming. Except for stochastic minimization (often impractical), we can only rely on gradient-based local methods to minimize the energy. With the presence of imaging noise and distracting structures in typical medical images, the local gradient-based optimization methods can be easily trapped in local minima.

Previous research has shown that combinatorial search strategies can be used to generate good initializations for various snake algorithms [6]. We also adopt a hybrid search strategy for ShRAC that combines the advantages of combinatorial search and local optimization. The algorithm iterates between the following two stages. In the first stage, we use multi-solution dynamic programming (MSDP) to generate initial candidates by minimizing E_{image} alone. In the second stage, we add the shape energy and use a local optimization method to minimize the entire energy in (4), starting with these candidates. After local optimization, a best contour is selected and used as the input of the next iteration. We also use a coarse-to-fine scheme: initially we use a large window to compute the image energy and a large step size in MSDP, which can be viewed as a coarse search stage, and then reduce both of them to perform finer searches. The whole search algorithm is illustrated in Fig. 3.

3.2. Multi-solution dynamic programming

In our first step, we generate a set of initial candidates for subsequent local optimization based on minimizing the image energy E_{image} alone. This is done by first constructing a search space around the initial contour and then using Dynamic Programming to select optimal contours.

Fig. 4 illustrates the construction of a search space from an initial contour. The initial contour is first aligned to the space of training contours and is represented as a set of control points (please see Section 4.1 for the details of the contour alignment of the lung field). For each initial control point, a set of candidate points can be generated based on the variations of the initial point. This search space can be represented by a graph, where directed segments are created to connect neighboring control point candidates, and they become the candidates for the contour segments. In this directed graph, each segment is assigned a weight $h_e(V_{ia}V_{(i+1)b})$ which is the local image edge strength. All possible closed contours in this directed graph constitute the search space for Dynamic Programming.

One important family of candidate sets is the set of points on a straight-line passing through the initial control points. The orientations of the lines can be along the normal to the initial contour, radial directions from the center of mass of the initial contour, or other object-specific directions (see Section 4.2 for an example). The advantages of the straight-line family is that the search grid is relative easy to generate. The disadvantage is that some lines can cross each other and create invalid contours in the search space.[1] Another method to generate candidate points that can avoid this problem is based on the signed distance transform of the initial contour. First the signed distance function $D(x,y)$ of the initial contour is computed. Then, each level set contour $\{(x,y)|D(x,y) = k\Delta d,\ k = \pm1, \pm2, \ldots\}$ is aligned and discretized using the same method as in shape training. The corresponding control points of these level set contours form a candidate set of the MSDP search space.

Fig. 5 shows two search grids constructed for lung field segmentation. Fig. 5(b) is generated from the straight-line method with direction normal to the initial contour. Fig. 5(c) shows the candidate points constructed from signed distance transform of the same initial contour. In general, search space generated from signed distance transform is more suited for objects with a blob-like structures and has many scale variations across different images.

Dynamic Programming [12] is used to efficiently find a minimum-weight contour in such search spaces. In order to generate several different initial candidates for the subsequent local optimization, we modify the dynamic programming algorithm to find the top p closed contours that minimize the image energy E_{image}. The modified algorithm, named multi-solution dynamic programming (MSDP), is summarized in Fig. 6. Note that standard Dynamic Programming is a special case of MSDP when $p = 1$.

[1] Since an optimal contour should have both low image energy and shape energy, these invalid contours usually will not be selected as the final result, and therefore it is a waste of computation to search in these spaces.

Dynamic Programming	Multi-Solution Dynamic Programming
1. For each candidate points of the first control point V_1, find a contour with minimum weights that pass through this candidate point.	1. For each candidate points of the first control point V_1, find the first p contours with minimum weights that pass through this candidate points.
1.1 for j = 2 to n, for each candidate point, compute the local cost by adding the local cost of previous candidates and the weight of the connecting edges. Select the minimum cost as the local cost of the current candidate and record the edge selected.	1.1 for j = 2 to n, for each candidate point, compute the local cost by adding the each local cost of previous candidates and the weight of the connecting edge. Select first p minimum cost among these sums as the local costs of the current candidate and record the edges selected.
1.2 compute the final cost by adding the last set of edges weights from V_{n-1} select the minimum one	1.2 compute the final const by adding the last set of edge weights from V_{n-1}, select the first p mininum one
2. Find a minimum weight contour among the n contours from step 1	2. Find k minimum weight contours from the n*p contours from step 1
3. Trace back the local edge record and output the minimum contour	3. Trace back the local edge record and output the p minimum contours

Fig. 6. The multi-solution dynamic programming algorithm.

The computational complexity of the original Dynamic Programming algorithm is $O(nm^2)$, where n is the number of control points and m is the number of candidates for each control points. The MSDP's complexity is $O(pnm^2)$, which means it grows linearly with the number of candidate contours needed.

3.3. Search space pruning for MSDP

To accommodate possible object scale variations in different images, the search space of MSDP is usually constructed such that the candidate control points cover a large area of the image. However, the search space with every possible connecting segment between candidate control points may contain many contours that has very large shape energy (low probability in the training data). To limit the number of these invalid contours, we first prune the search space based on the segment orientations of the initial contour. All the connecting segments in the search graph are compared to the corresponding segment in the initial contour. Those segments with orientation differences larger than a threshold are removed from the search graph (see Fig. 7). This threshold controls how different a contour in the search space can be compared with the initial contour. We select this threshold empirically and keep it fixed across all the input images. This procedure allows the computation to be focused on those candidates that have similar shapes to the initial contour. It is especially useful in feeding the shape information back into the MSDP search space.

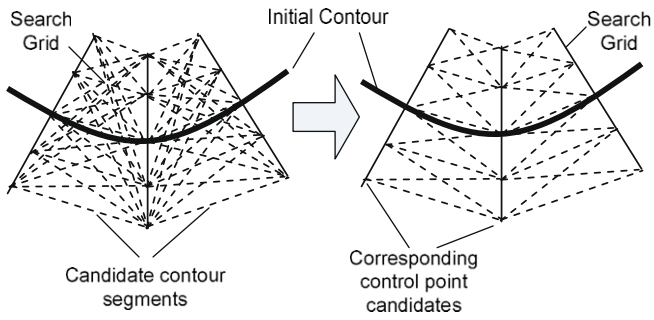

Fig. 7. Search space pruning to reduce invalid shapes: segments that has large orientation difference with the initial contour are removed from the search space.

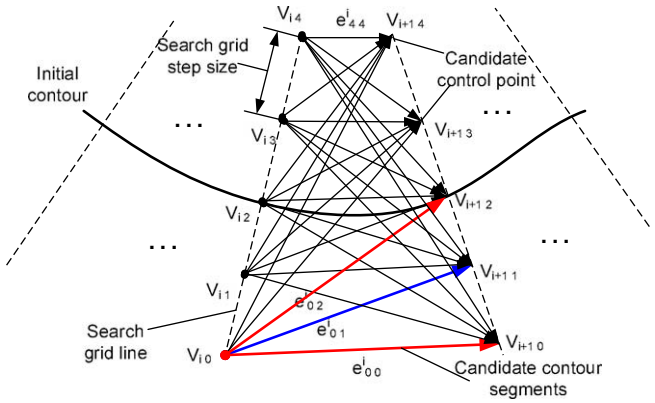

Fig. 8. Search space pruning to reduce clustered solutions: blue segment e^i_{01} is only preserved if its image energy is less than both of the red segments e^i_{00} and e^i_{02}. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A second pruning method is designed to improve the diversity of solutions obtained by MSDP. Because the image energy is usually continuous with respect to the control point positions, the top ranking contours from MSDP often cluster together and might not include the optimal contour for the entire energy. Medical images are intrinsically noisy and ambiguous, and often many competing interpretations exist. It is critical to preserve multiple, and more importantly, distinct solutions when only image energy

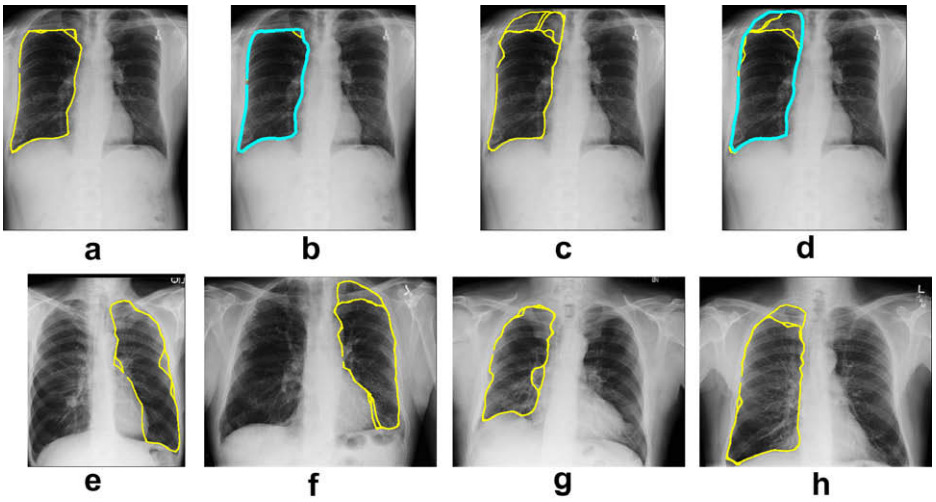

Fig. 9. Effect of clustered solution pruning. (a) Initial candidates selected by MSDP without pruning. (b) Local optimization result of the candidates from (a). (c) Initial candidates selected by MSDP with pruning. (d) Local optimization result of candidates from (c). The bold blue contours are the finally selected solutions. (e–h) Other examples of candidate contours generated by MSDP in the pruned search space. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

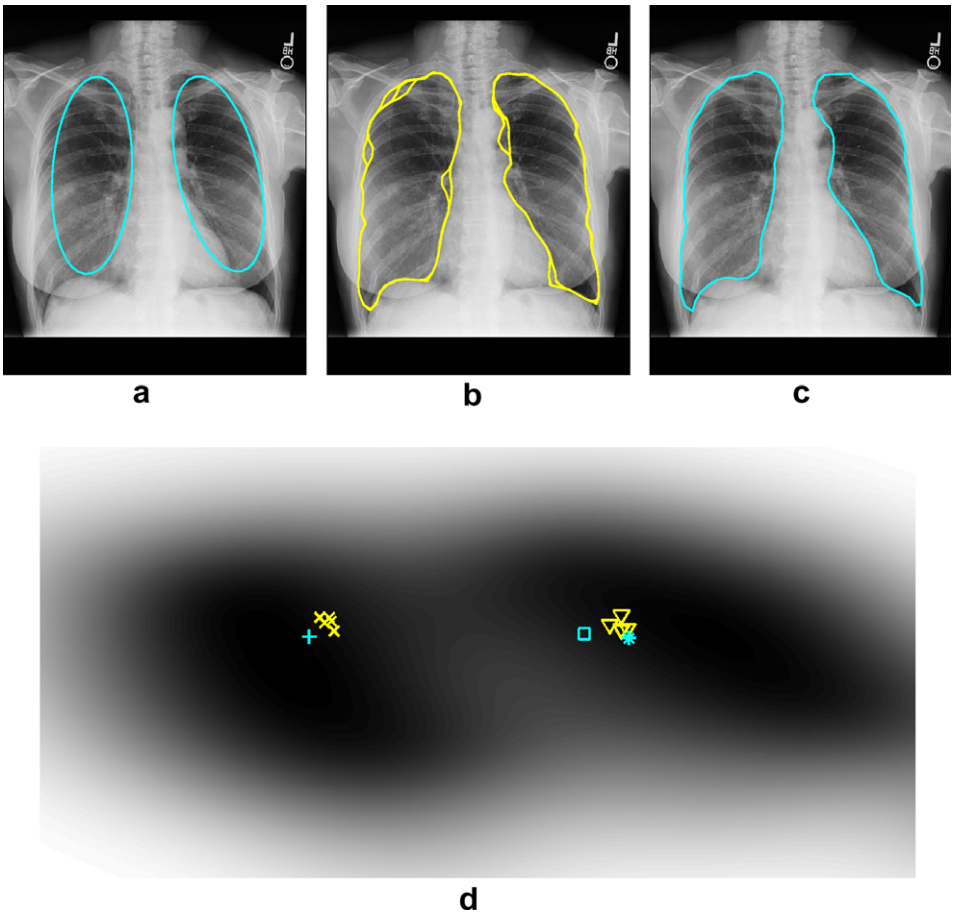

Fig. 10. Segmentation progress of our search strategy for ShRAC. (a) Initial contours (same shape for both lungs). (b) Contour candidates generated by MSDP at the coarse scale. (c) Final segmentation result at the fine scale of our algorithm. (d) The corresponding contours in the shape space (only the projection on the first two PCA eigenvectors of the training data is shown). □: Initial contour, ×: MSDP results for the left lung, ▽: MSDP results for the right lung, +: final contour for the left lung, ∗: final contour for the right lung.

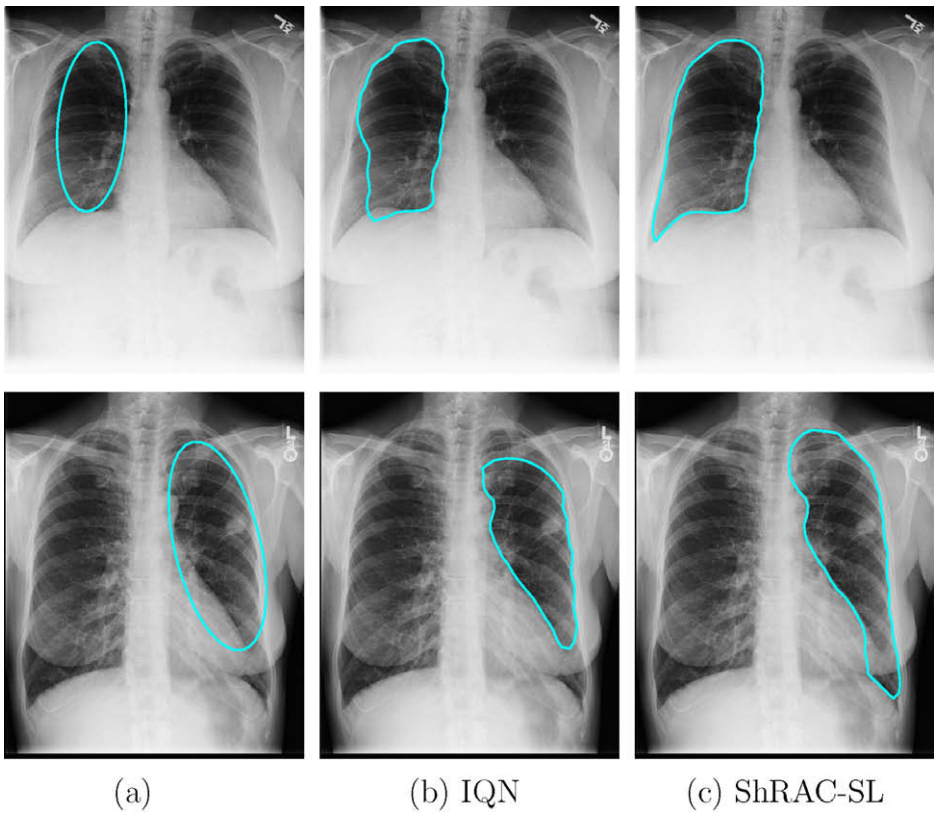

Fig. 11. Comparison of IQN and ShRAC-SL for lung field segmentation. (a) Initial contours. (b) Segmentation result by IQN. (c) Segmentation result of ShRAC-SL.

is minimized. If we apply local optimization starting from similar candidates, the final solution often converges to the same minimum (which may not be the global minimum), thus defeating the purpose of using MSDP.

We propose to further prune the search space of MSDP so that we can obtain a diverse set of initial candidates with a small p. We refer to this process as Clustered Solution Pruning, which is similar in spirit to the procedure employed in [13] where non-distinct proposals are pruned by minimizing the approximate Kullback–Leibler (KL)-divergence. In practice, we use a simple yet effective scheme. For each directed segment e^i_{jk} in the search graph (Fig. 8), we compare it with its two neighboring segments e^i_{jk+1} and e^i_{jk-1}; the segment is pruned from the search graph if it is not the minimum among the three segments.

It is easy to see that after pruning, each candidate point still has at least one segment connecting to a candidate point in the next control point set, therefore the search space still contains at least one valid solution. This pruning reduces the search space to only candidate segments with locally minimum image energy. In essence, this makes the search favor local minima of the image energy. By starting from multiple locally minimum solutions to the image energy, the subsequent local optimization of the whole energy will have a better chance to reach a global minimum solution. Fig. 9(a)–(d) shows an example where the proposed pruning method helps MSDP to find more diversified initial candidates that eventually lead to the desired solution. Fig. 9(e)–(h) shows four other examples of multiple candidate contours returned by MSDP in the pruned search space.

3.4. Overall computational complexity

As mentioned in Section 3.2, the time complexity of MSDP is $O(pnm^2)$, where p is the number of candidates selected by MSDP (5 in our experiments), n is the number of control points (60 in our experiments), m is the number of candidate control points (usually 20). This is a rather small-scale dynamic program problem (roughly equivalent to a dynamic programming based stereo matching algorithm on a single scan line of 300 pixels with 20 disparities), which can be easily handled by modern CPUs. A more reasonable estimate of the time complexity is to use the number of total energy function evaluations since most of the CPU time is actually spent on these evaluations. The MSDP search space needs nm^2 local edge strength evaluations. Since Each image energy evaluation includes n evaluation of local edge strength, the entire search space is equivalent to m^2 times evaluation of the image energy defined in Eq. (3). The total computational cost of our algorithm is therefore about 400 image energy evaluations, plus $p = 5$ local optimization from the candidate contours selected by MSDP (each include about 100 evaluation of the entire energy function as defined in Eq. (4)). For the lung field segmentation, it takes about 30 s to find an optimal contour on a Pentium 4 2.4 GHz machine implemented in matlab code.

4. Experimental results

4.1. Lung field segmentation

First, we use lung field segmentation on X-ray images to demonstrate the power of our search strategy in segmenting object consists of widely different sets of shapes as a single class. We trained a single shape model, using manually segmented image masks of both left and right lungs (200 shapes). A set of 60 control points is used to represent the contour of a lung field (either left or right). Each shape is aligned to a coordinate system that has the origin at the shape center and y-axis coincides with the shape's major axis. It is also scaled to have unit area. Such normalization

ensures that the shape training is done in a common coordinate system. The two intersections of a contour with the x-axis are used as anchor points and the other 58 control points are evenly distributed between these anchor points (29 on each side).

Our segmentation algorithm follows the flow diagram in Fig. 3. We chose to perform the coarse-to-fine search at two scales. The initial search area covered by MSDP is set to be large enough to account for possible poor initializations (Fig. 5). We tested the search space generated by the straight-line (SL) method and the distance transform (DT) method. In the rest of the paper, we will refer these two methods as ShRAC-SL and ShRAC-DT, respectively. In MSDP, we empirically chose the top five contours as the initial candidates for local optimization. The local optimization method we used is the BFGS Quasi-Newton method for unconstrained nonlinear minimization in Matlab [14].

To show the internal progress of our algorithm, we plotted the initial contours (Fig. 10(a)), MSDP results at the coarse scale (Fig. 10(b)) and the final segmented contours (Fig. 10(c)) of both left and right lung fields on one test image. The transition of these contours in the shape space is shown in Fig. 10 (similar to Fig. 2, the shape space is projected onto the first two PCA eigenvectors of the training data). Fig. 10 shows that while MSDP automatically generates diversified contours that are close to the correct shape clusters (left or right) in the shape space, the overall algorithm will eventually drive the contour to further minimize the joint energy.

To evaluate the performance of the algorithm, we implemented another algorithm as a proxy of the method in [4]. It is a coarse-to-fine algorithm, which at each scale uses the same local optimizer as used in ShRAC. The algorithm starts with a large windows size to compute image energy and its gradient, and gradually reduces them at each iteration. We refer to this algorithm as IQN (Iterative Quasi-Newton) method.

Both of these algorithms are initialized by a program that computes a rough centerline of each lung field (by searching for valleys in the image). The centerline is used to scale and orient an elliptical shape to roughly cover the lung field. This initialization also helps us handle X-ray images of different resolution. Note the initialization algorithm is not supposed to be robust and it sometimes fails to cover the entire lung field correctly.

Fig. 11 shows two X-ray images segmented by ShRAC-SL and IQN. The results show that IQN is more susceptible to local minima and therefore is more dependent on initialization. On the other hand, ShRAC shows good robustness with respect to initialization. Fig. 12 shows more segmentation results of the ShRAC-DT algorithm using the distance transform generated search space. We include both successful (a–d) and failure cases (e and f) to provide an

Table 1
Accuracy comparison (mean and standard deviation of the overlap ratio Ω) of the three segmentation algorithms on lung field segmentation for two data sets (CR—136 images, JSRT—247 images)

	IQN	ShRAC-SL	ShRAC-DT	ASM
CR	0.841 ± 0.081	0.891 ± 0.037	0.896 ± 0.039	0.893 ± 0.060
JSRT	0.881 ± 0.070	0.907 ± 0.033	0.914 ± 0.034	0.920 ± 0.022

IQN: Iterative Quasi-Newton method. ShRAC-SL: ShRAC with the search space generated by the straight-line method. ShRAC-DT: ShRAC with the search space generated by the distance transform method. ASM: Active shape model.

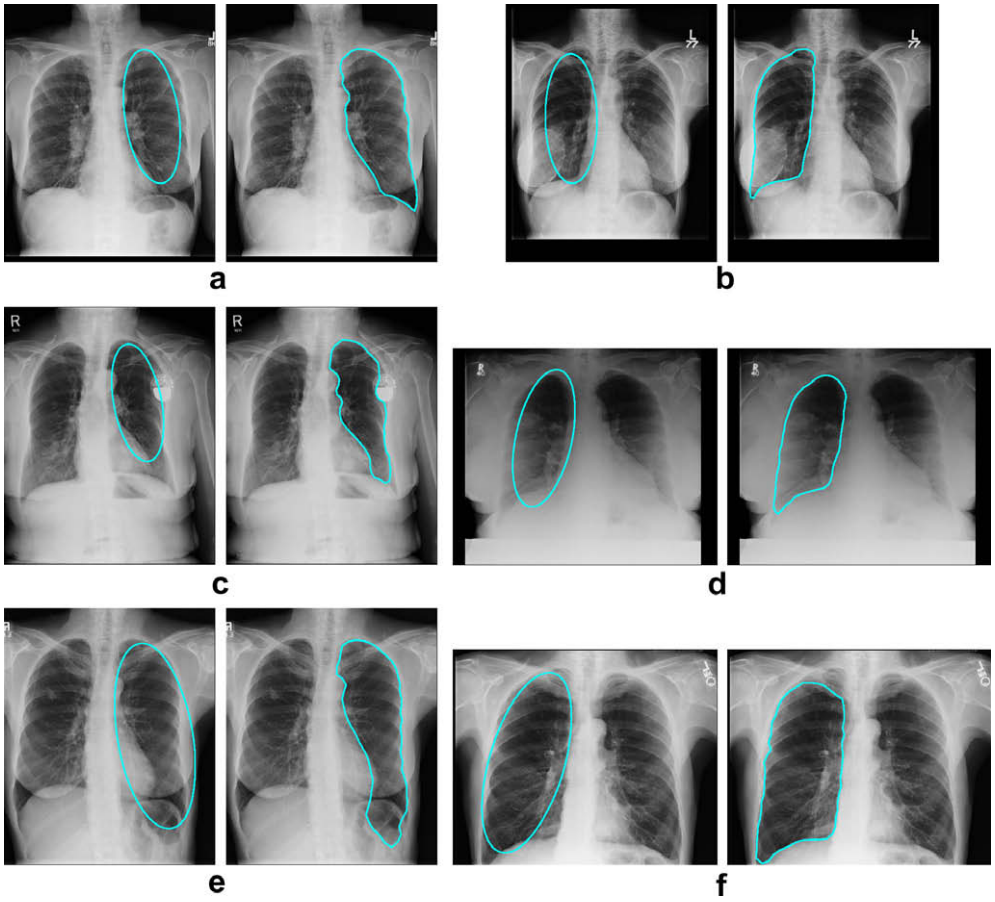

Fig. 12. Robustness of the ShRAC-DT search strategy to various structures and imaging conditions. The last row shows example failure modes of the algorithm. Note that even in the failure cases, the segmented contours were adversely influenced by the strong yet misleading edges, but still correspond to valid shapes.

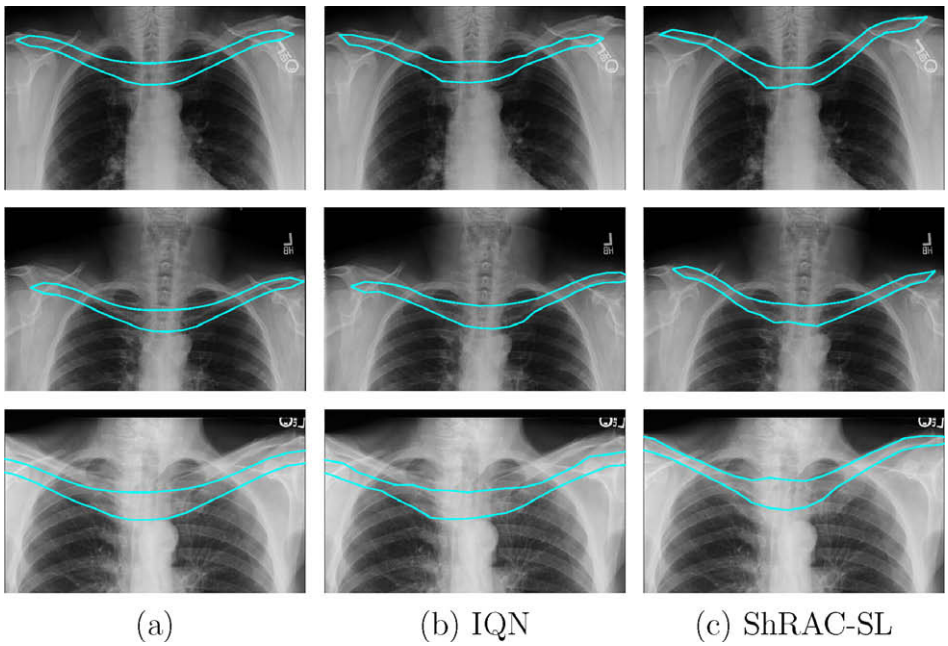

Fig. 13. Comparison of IQN and ShRAC-SL for clavicle segmentation. (a) Initial contours. (b) Segmentation results of IQN. (c) Segmentation results of ShRAC-SL. The average vertical displacement over 100 images (size 512×625, only upper half are shown in the figure) using IQN and ShRAC-SL is 13.6 pixels and 2.3 pixels, respectively.

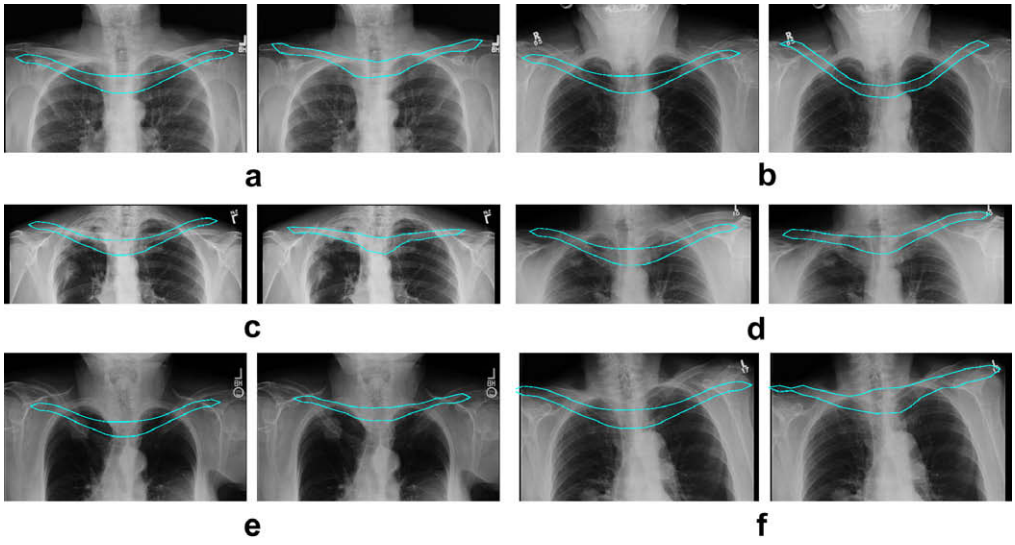

Fig. 14. More clavicle segmentation examples using ShRAC-SL. (e and f) Some failure modes of the algorithm. In (e), the clavicle shape is too different from the normal shapes in the training set. In (f), a poor initialization together with some distracting structures on the left make the algorithm select the wrong contour.

Table 2
Comparison of average vertical displacement of control points to the ground truth on clavicle segmentation (100 images with size 512×625)

	IQN	ShRAC-SL
Avg. disp.	13.6 (pixels)	2.3 (pixels)

visual evaluation of the robustness of the ShRAC model and search strategy under various structural and imaging variations. In particular, the result in (e) was adversely influenced by the additional misleading edges from the large air bubble below and additional overlapping structure (breast) above the desired boundary, while (f) was primarily due to the poor contrast above the clavicles. Still, the incorrect contours correspond to valid shapes. Note that ShRAC was able to overcome poor overall contrast in (d) and the same overlapping structures when there is no complication from prominent air bubbles.

To compare the performances quantitatively, we ran these two algorithms on two data sets (CR—136 images and JSRT—247 images) for which manually segmented ground truth is available. We use the following overlap ratio Ω to measure the performance:

$$\Omega = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP} + \mathrm{FN}) \tag{6}$$

where TP stands for true positive (the area correctly classified as object), FP for false positive (the area incorrectly classified as object), and FN for false negative (the area incorrectly classified as background).

The overlap ratios obtained by these two algorithms on the two data sets are listed in Table 1 (excluding training images). For these two data sets, we also list the performance data of an ASM imple-

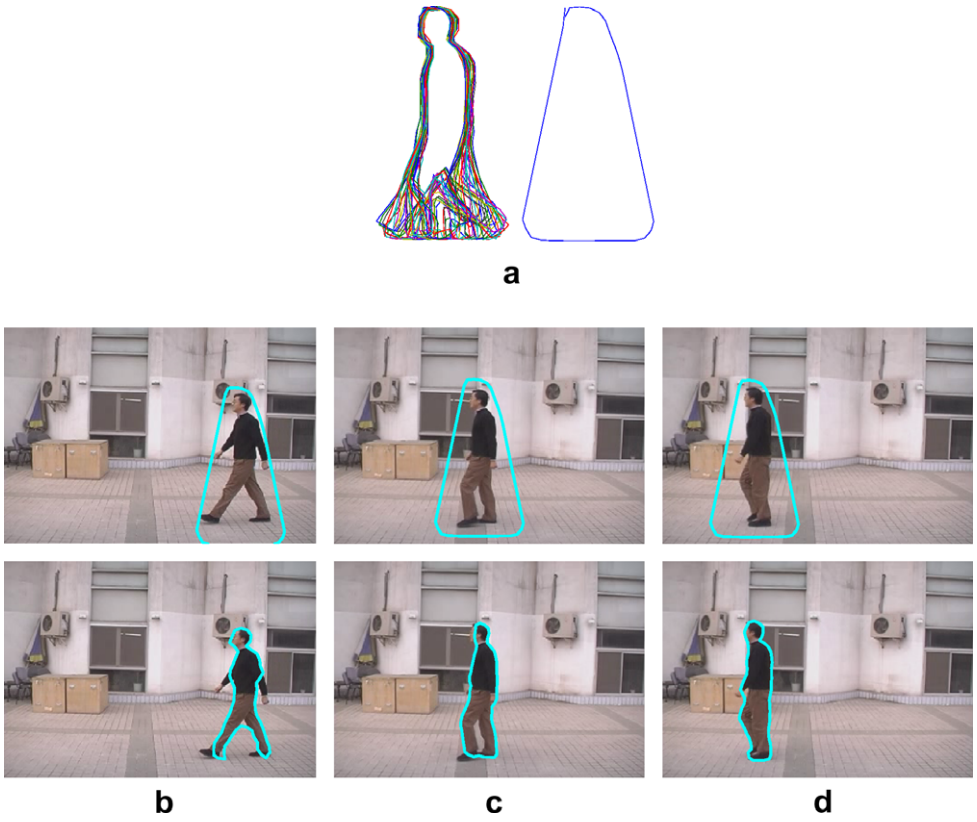

Fig. 15. Human silhouette segmentation experiment. (a) Aligned training contours of the human silhouette (without arm) and their bounding polygon (right). (b–d) Segmentation results of ShRAC-SL for different pose in the walking sequence, 1st row: initial contour, 2nd row: final contour.

mentation from another study. Recall that the ASM algorithm requires linear models, so two separate models are built for left lung and right lung, and a model selection stage is used to select the correct model before the algorithm begins. The ASM algorithm also requires more elaborate labeling because many landmarks need to be identified and meticulously located to properly train the object appearance model.

In Table 1, ShRAC clearly outperforms IQN. Our search strategy in ShRAC leads to 2.6–5.0% increase of overall segmentation performance for straight-line method and 3.0–5.8% for distance transform method. It is noteworthy that the improvement is due to capturing of shape details near the lung boundary, and therefore translates into substantial visual improvement in the segmentation. For a blob-like object such as the lung field, getting the bulk of the lung field right is almost trivial compared to locating the tips and corners accurately; for example, the dramatic improvement shown in Fig. 9 amounts to only a 10% change in the overlap ratio. Our search strategy for ShRAC achieves comparable performance as the two-model (separate for the right and left lungs) ASM algorithm that requires accurate landmarks in training. Along with the other advantages mentioned earlier, this shows the promise of ShRAC as a more flexible and powerful segmentation tool.

4.2. Clavicle segmentation

To further verify the effectiveness of our search strategy, we also applied it to segmenting clavicles. Segmenting clavicles is quite challenging because there are many similar structures in the vicinity, including the first three ribs and shoulder joints. Since the two clavicles are always symmetric, we merge them together and represent them as a single contour (Fig. 13).

We used 30 manually segmented contours as the training examples. We only use the straight-line method for search space generation because the distance transform method does not capture the major variations of the clavicle structure. The search directions for clavicles are restricted to the vertical directions since the bones are very thin and lack of horizontal features. The same lung field centerlines are used to estimate a rough position and scale of the initial mean shape. Fig. 13 shows three results obtained by ShRAC-SL and IQN. Locating bone structures is even more difficult for a local optimization algorithm such as IQN because there are many local minima created by other bones in the neighborhood. However, the MSDP-based ShRAC is capable of jumping out of local minima and finding superior solutions. More segmentation results can be found in Fig. 14.

For quantitative evaluation, we computed the vertical displacement of the extracted control points to their projections in the vertical direction on the manually segmented ground truth contours. The average vertical displacement over 100 images using IQN and ShRAC-SL are summarized in Table 2.

4.3. People silhouette segmentation

We also performed some preliminary experiments in a different problem domain, namely segmenting the human silhouette.[2] We used a set of 32 manually outlined side view contours of walking people (see Fig. 15(a)) to train a KSDE shape model, and applied our search strategy to segment the human silhouette in the image. The contour of the human silhouette changes dramatically due to

[2] Human gait dataset courtesy of National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences.

the arm and leg articulation, which causes problems when parameterizing the contour into a common representation. To use the closed contour representation with fixed number of control points, we exclude the arms from the silhouette and only train a model with the upper body and legs. This also limits the complexity of the shape model and reduce the point correspondence problems between different poses. Fig. 15(a) shows the contours used for training. Clearly, the shape variation concentrates on the legs and is highly nonlinear.

To process each frame independently, we used the bounding polygon of the training contours as the initial contour, whose scale is fixed to approximately the scale of the person in the video. The initial contour was manually placed around the person (Fig. 15(b)–(d)). We use ShRAC-SL with search direction normal to the initial contour for segmentation. To be more efficient in the searching process, the search steps for control points in the contour is programmed to be large in the leg parts and small in the upper body, which corresponds to the variations exhibit in the training data. The last row of Fig. 15 contains the final segmentation results. Given the background clutters in the image, these preliminary results are quite promising. Depth [15], appearance [16], and motion [17] information can be used as more reliable methods for the initialization. We are working on automatic initialization of the segmentation and more rigorous experimental evaluation. It is worth noting that contour-based silhouette characterization was used in [15], while parts-based articulated models were adopted in [16,18]. The methods described in [16,18] can also handle different poses and viewpoints.

5. Conclusions and future work

We have presented a robust search strategy for the shape regularized active contour which uses a nonlinear prior shape model. Our search strategy for ShRAC is implemented as a coarse-to-fine algorithm that combines the advantages of combinatorial search and local optimization. It uses multi-solution dynamic programming to generate initial candidates that have minimal image energies, and then uses local gradient-based minimization of the entire energy to select the final optimal contour. Two types of prunings are applied to the discrete search space to reduce the number of invalid shapes as well as to prevent the clustering of MSDP results. The experiments on lung field and clavicle segmentation demonstrate robustness of our search strategy to initialization and distracting structures in medical images.

In many cases, shape-based similarity is critical to the segmentation task in applications when there are appearance ambiguities between the object of interest and the background. The search strategies proposed in this study embed both the shape similarity and image feature matching to provide improved segmentation performance. There are a few possible extensions. Incorporating powerful appearance models, preferably nonlinear, will be of major interest. Besides a closed contour model such as ShRAC, dynamic programming can also be applied to any model that has a tree structure [18,19], extending the proposed multi-stage search strategy to this general tree structure will also be very useful. Another direction is to extend ShRAC to 3D object segmentation, where Dynamic Programming can be replaced by other combinatorial search methods such as Graph Cuts [20]. We are also exploring applications of ShRAC outside the medical imaging domain.

References

[1] B.V. Ginneken, B.M.H. Romeny, M.A. Viergever, Computer-aided diagnosis in chest radiography: a survey, IEEE Trans. Med. Imaging 20 (12) (2001) 1228–1241.
[2] T.F. Cootes, C.J. Taylor, D. Cooper, J. Graham, Active shape models-their training and application, Comput. Vis. Image Understanding 61 (1) (1995) 38–59.
[3] S. Romdhani, S. Gong, A. Psarrou, A multi-view non-linear active shape model using kernel pca, in: Proceedings of 10th British Machine Vision Conference, 1999, pp. 483–492.
[4] D. Cremers, T. Kohlberge, C. Schnorr, Shape statistics in kernel space for variational image segmentation, Pattern Recognit. 36 (9) (2003) 1929–1943.
[5] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, Int. J. Comput. Vis. 1 (4) (1988) 321–331.
[6] J. Liang, T. McInerne, D. Terzopoulos, United snakes, Med. Image Anal. 10 (2006) 215–233.
[7] D. Shen, Y. Zhan, C. Davatzikos, Segmentation of prostate boundaries from ultrasound images using statistical shape model, IEEE Trans. Med. Imaging 22 (4) (2003) 539–551.
[8] T. Yu, J. Luo, A. Singhal, N. Ahuja, Shape regularized active contour based on dynamic programming for anatomical structure segmentation, in: Proceedings of SPIE Medical Imaging 2005: Image Processing, 2005, pp. 419–430.
[9] M. Leventon, E. Grimson, O. Faugeras, Statistical shape influence in geodesic active contours, in: IEEE Conference on Computer Vision and Pattern Recognition 2000—Volume 1, IEEE Computer Society, 2000, pp. 316–323.
[10] J. Xie, Y. Jiang, H. tat Tsui, Segmentation of kidney from ultrasound images based on texture and shape priors, IEEE Trans. Med. Imaging 24 (1) (2005) 45–57.
[11] M. Rousson, N. Paragios, Shape priors for level set representations, in: European Conference on Computer Vision 2002, LNCS, vol. 2351, 2002, pp. 78–92.
[12] A.A. Amini, T.E. Weymouth, R.C. Jain, Using dynamic programming for solving variational problems in vision, IEEE Trans. Pattern Anal. Mach. Intell. 12 (9) (1990) 855–867.
[13] Z. Tu, S.-C. Zhu, Image segmentation by data-driven markov chain monte carlo, IEEE Trans. Pattern Anal. Mach. Intell. 24 (5) (2002) 657–673.
[14] The MathWorks, Inc., Matlab Optimization Toolbox, 2004. Available from: <http://www.mathworks.com/access/helpdesk/help/toolbox/optim/>.
[15] L. Zhao, Dressed human modeling, detection, and parts localization, Tech. Rep. CMU-RI-TR-01-19, The Robotics Institute, Carnegie Mellon University, 2001.
[16] D. Ramanan, D.A. Forsyth, A. Zisserman, Strike a pose: tracking people by finding stylized poses, in: CVPR'05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)—Volume 1, IEEE Computer Society, Washington, DC, USA, 2005, pp. 271–278.
[17] J. Zhang, R. Collins, Y. Liu, Representation and matching of articulated shapes, in: CVPR'04: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)—Volume 2, IEEE Computer Society, 2004, pp. 342–349.
[18] J. Zhang, Y. Liu, J. Luo, R. Collins, Body localization in still images using hierarchical models and hybrid search, in: CVPR'06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Washington, DC, USA, 2006, pp. 1536–1543.
[19] P.F. Felzenszwalb, Representation and detection of deformable shapes, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2) (2005) 208–220.
[20] V. Kolmogorov, R. Zabih, What energy functions can be minimized via graph cuts?, IEEE Trans Pattern Anal. Mach. Intell. 26 (2) (2004) 147–159.