

3D automatic anatomy segmentation based on iterative graph-cut-ASM

Xinjian Chen^{a),b)}

Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Building 10 Room 1C515, Bethesda, Maryland 20892-1182 and Life Sciences Research Center, School of Life Sciences and Technology, Xidian University, Xi'an 710071, China

Ulas Bagci

Radiology and Imaging Sciences, Clinical Center, National Institutes of Health, Building 10 Room 1C515, Bethesda, Maryland 20892-1182

(Received 15 December 2010; revised 16 May 2011; accepted for publication 1 June 2011; published 25 July 2011)

Purpose: This paper studies the feasibility of developing an automatic anatomy segmentation (AAS) system in clinical radiology and demonstrates its operation on clinical 3D images.

Methods: The AAS system, the authors are developing consists of two main parts: object recognition and object delineation. As for recognition, a hierarchical 3D scale-based multiobject method is used for the multiobject recognition task, which incorporates intensity weighted ball-scale (b-scale) information into the active shape model (ASM). For object delineation, an iterative graph-cut-ASM (IGCASM) algorithm is proposed, which effectively combines the rich statistical shape information embodied in ASM with the globally optimal delineation capability of the GC method. The presented IGCASM algorithm is a 3D generalization of the 2D GC-ASM method that they proposed previously in Chen *et al.* [Proc. SPIE, **7259**, 72590C1–72590C-8 (2009)]. The proposed methods are tested on two datasets comprised of images obtained from 20 patients (10 male and 10 female) of clinical abdominal CT scans, and 11 foot magnetic resonance imaging (MRI) scans. The test is for four organs (liver, left and right kidneys, and spleen) segmentation, five foot bones (calcaneus, tibia, cuboid, talus, and navicular). The recognition and delineation accuracies were evaluated separately. The recognition accuracy was evaluated in terms of translation, rotation, and scale (size) error. The delineation accuracy was evaluated in terms of true and false positive volume fractions (TPVF, FPVF). The efficiency of the delineation method was also evaluated on an Intel Pentium IV PC with a 3.4 GHz CPU machine.

Results: The recognition accuracies in terms of translation, rotation, and scale error over all organs are about 8 mm, 10° and 0.03, and over all foot bones are about 3.5709 mm, 0.35° and 0.025, respectively. The accuracy of delineation over all organs for all subjects as expressed in TPVF and FPVF is 93.01% and 0.22%, and all foot bones for all subjects are 93.75% and 0.28%, respectively. While the delineations for the four organs can be accomplished quite rapidly with average of 78 s, the delineations for the five foot bones can be accomplished with average of 70 s.

Conclusions: The experimental results showed the feasibility and efficacy of the proposed automatic anatomy segmentation system: (a) the incorporation of shape priors into the GC framework is feasible in 3D as demonstrated previously for 2D images; (b) our results in 3D confirm the accuracy behavior observed in 2D. The hybrid strategy IGCASM seems to be more robust and accurate than ASM and GC individually; and (c) delineations within body regions and foot bones of clinical importance can be accomplished quite rapidly within 1.5 min. © 2011 American Association of Physicists in Medicine. [DOI: 10.1118/1.3602070]

Key words: statistical models, object recognition, image segmentation, active shape models, graph cut

I. INTRODUCTION

With the development of medical image processing methods and systems, clinical radiology places increasingly greater emphasis on quantification in routine practice. To facilitate this, computerized recognition, labeling, and delineation of anatomic organs, tissue regions, and suborgans, playing an assistive role, becomes important in clinical radiology. In spite of several decades of research and many key advances, several challenges still remain in this area. Efficient, robust and automatic anatomy segmentation (AAS) is one of these

challenges. Based on the interested region, the methods of AAS can be classified into two types: anatomy segmentation methods for the brain and anatomy segmentation methods for the body region (skull base to feet).

The body region segmentation methods could be classified into several types: model based,^{1–5} image based,^{6–13} and hybrid methods.^{14–16} Representatives in the model-based group are active contours models,^{1,2,17} active shape and appearance models (ASM, AAM).^{3–5} Active contour models^{1,2} are capable of modeling complex shapes via continuously deformable

curves. The ASM/AAM (Refs. 3–5) methods use “landmarks” to represent shape and principal component analysis (PCA) to capture the major modes of variation in shape observed in the training data sets. The image-based methods include graph cut (GC),^{6,7} level set,^{8,9} watershed,^{10,11} fuzzy connectedness,^{18,19} and live wire.^{12,13} The GC (Refs. 6 and 7) methods have been widely used in image segmentation due to their ability to compute globally optimal solutions. Graph cuts have proven to be useful multidimensional optimization tools that can enforce piecewise smoothness while preserving relevant sharp discontinuities. The level set^{8,9} method is also widely used in image segmentation. The advantage of the level set method is that one can perform numerical computations involving curves and surfaces on a fixed Cartesian grid without having to parameterize these objects. Also, the level set method makes it easy to follow shapes that change topology. The watershed method^{10,11} has interesting properties that make it useful for many different image segmentation applications: it is simple and intuitive, can be parallelized, and always produces a complete division of the image. The fuzzy connectedness methods have characteristics similar to those of graph cut methods and have additional theoretical and computational advantages. The live wire methods^{12,13} are user-steered two-dimensional segmentation methods in which the user provides recognition help and the algorithm does the delineation. In fact, all of the above image-based methods operate in the same manner. Hybrid approaches are rightfully attracting a great deal of attention now. Their premise is to combine the complementary strengths of the individual methods to arrive at a more powerful hybrid method. These methods include methods such as combination of active shape model with live wire method,¹⁴ combination of watersheds with fast region merging methods,¹⁵ and combination of shape-intensity prior models with level sets.¹⁶

Integrating shape priors into GC segmentation framework receives a great interest recently.^{20–25} For these methods, the way of generating the shape crucially affects the success of the segmentation, as delineation may leak into nonobject territories due to suboptimal recognition. Vu *et al.*²⁰ defined the shape prior as energy on a shape distance with popular level set approaches. However, the prior shape is a simple fixed shape, which may lead to the delineation results to become unpredictable along the weak edges. Freedman and Zhang²¹ incorporated a shape template into the GC formulation as a distance function. Malcolm *et al.*²² imposed a shape model on the terminal edges and performed min-cut iteratively starting with an initial contour. These two methods need an effective user interaction and they might possibly fall short in handling complex shapes of arbitrary topology of 3D objects. Leventon *et al.*²⁵ integrated a deformed shape into GC segmentation, where the shape prior is deformed based on the Gaussian distribution of some predefined geometrical shape statistics. However, this is not true in reality, because pose variance and deformations are specific to the object of interest and often having non-Gaussian distributions. Kohli *et al.*²³ presented an algorithm for performing simultaneous segmentation and 3D pose estimation of a human body from multiple views. They used a simple articulated stickman model, which together with a conditional random field is

used as the shape prior. Lempitsky *et al.*²⁴ used nonparametric kernel densities to model a shape prior and integrated into the GC. However, the computational burdens of the proposed methods are high, and the high variations pertaining to medical images are not accurately handled. The contributions of this study are as follows: Unlike all those methods, we propose a fully automatic method based on a hierarchical 3D scale-based multiobject recognition (HSMOR) framework.²⁶ Moreover, in the proposed methodology not only we perform recognition and delineation directly on 3D, but also computational cost is minimal due to without doing any search and optimization. In addition, we don't need to do the shape registration due to automatic recognition/initialization.

The AAS system, we are developing consists of three main parts: model building (training), object recognition, and object delineation. For the object recognition, HSMOR method²⁶ is used for the recognition task. For the delineation, we present an iterative graph-cut-ASM (IGCASM) algorithm, which is a 3D generalization of the 2D GC-ASM method.^{27,28} The proposed methods are tested on a clinical abdominal CT data set with 20 patients and a foot MRI data set with 11 images. The preliminary results show that: (a) it is feasible to explicitly bring prior 3D statistical shape information into the GC framework and (b) the 3D IGCASM delineation method improves on ASM and GC and can provide practical operational time on clinical images.

This paper is organized as follows. In Sec. II, the complete methodology of the proposed AAS method is described. In Sec. III, we describe a detailed evaluation of this method in terms of its recognition and delineation accuracy and efficiency on the clinical datasets. In Sec. IV, we summarize our conclusions. A preliminary version of this paper has appeared in the Conference Proceedings of the SPIE 2010 Medical Imaging Symposium.²⁸

II. GRAPH-CUT-ASM

II.A. Overview of approach

The proposed method consists of two phases: training phase and segmentation phase. In the training phase, we construct the ASM model and train the GC parameters. The segmentation phase consists of two main parts: initialization (recognition) and delineation. For the recognition part, a hierarchical 3D scale-based multiobject recognition method is used. For the delineation part, the object shape information generated from the initialization step is integrated into the GC cost computation, and an iterative GCASM method is proposed for object delineation. The proposed flowchart of our AAS system is shown in Fig. 1.

II.B. Model building and parameters training

For volumetric data, there are several solutions to establish landmarks' correspondence. One popular method is to project landmark points on a spherical coordinate system, but this method is generally limited to convex objects.²⁹ In this paper, a semiautomatic landmark tagging method,

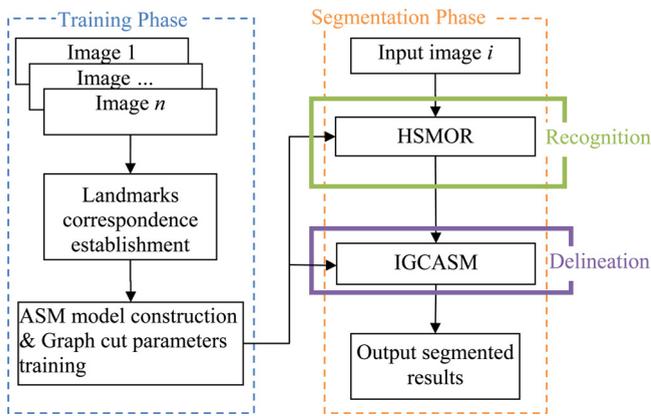


FIG. 1. The flowchart of the proposed AAS system.

equally spaced landmark tagging,³⁰ is used to establish correspondence among landmarks in our experiments. Although this method is proposed for 2D objects, and equally spacing a fixed number of points for 3D objects is much more difficult, we use this equally spaced landmark tagging technique in a pseudo-3D manner, where the 3D object is annotated slice-by-slice. In order to provide anatomical correspondence among 2D slices of 3D objects, a careful selection procedure was devised for use by an expert in the training step.²⁶ The same physical location of slices in one object does not necessarily correspond to the same physical location in another object of the same class. Therefore, experts select slices corresponding to each other in terms of physical locations within the body. This is a much simpler 1D correspondence problem, which is easier and simpler to tackle than the 2D point correspondence problem, however, requires expert's anatomy knowledge, which facilitates identifying portions of the organs or bones equivalent to each other in position.

To obtain a true shape representation of the family of an object, location, scale, and rotation effects within the family need to be filtered out. This is done by aligning shapes within the family of object (in the training set) to each other by changing the pose parameters (scale, rotation, and translation). For multiple objects, the object assemblies are aligned. PCA is then applied to the aligned N training shape vectors \mathbf{x}_i , $i=1, \dots, N$, where \mathbf{x}_i includes the coordinates of the shape boundaries. The model M is then constructed following the standard ASM procedure.³

The parameters of GC are also trained during the training stage; more details on this are given at Sec. II D 1.

II.C. Anatomical structure recognition

Recognition of anatomical structures is the first step in model-based segmentation approaches. Recognition in the model-based approach is to locate the model in the image in terms of pose (translation, scale, and orientation) using shape, texture, or both shape and texture information. This procedure involves matching model to image and calculating pose transformations regarding to the pose difference. Once recognition is handled, the resulted pose is used as an input

for delineation process. For recognition task, we use our proposed recognition method called HSMOR.²⁶ Critically, the HSMOR combines coarse to fine strategies to build an efficient model-based segmentation algorithm. To do so, we incorporate a large number of structures into the recognition algorithm to yield quick, robust, and accurate recognition. Besides, we use scale information to build reliable relationship between shape and texture patterns that facilitates accurate recognition of single and multiple objects without using optimization methods.

Different from the conventional ASM method, which incorporates samples of image information in the neighborhood of the shape boundary points, we incorporate expected appearance information within the entire interior of the annotated shape into the ASM after the construction of model assembly. In order to accomplish this, first, we encode the appearance of the gray level images to extract hierarchical geometric patterns. This encoding process is called ball-scale (*b-scale* for short) encoding.²⁶ The main idea in b-scale encoding is to determine the size of local structures at every voxel as the radius of the largest ball centered at the voxel within, which intensities are homogeneous under a prespecified region-homogeneity criterion. Although the size of a local structure is estimated using appearance information of the gray scale images, i.e., region-homogeneity criterion, b-scale images contain only rough geometric information. Incorporating appearance information into this rough knowledge characterizes scale information of local structures. Thus, it allows us to distinguish objects of even same size by their appearance information. We proposed in Ref. 26 hat extracting b-scale information from images together with corresponding appearance information of the local structures can be possible by weighting the radius of the ball centered at a given voxel with the intensity value of that voxel. As a result, object scale information is enriched with local intensity values.

The algorithm for intensity weighted b-scale estimation is presented below.

Algorithm IWOSE (Intensity Weighted Object Scale Estimation):

Input: $c \in C$ in a scene $C = (C, f)$, W_ψ , and a fixed threshold t_s .

Output: Intensity weighted b-scale $r'(c)$ at c .

Begin

Set $k = 1$

While $FO_{k,\mu}(c) \geq t_s$ **do**

Set k to $k + 1$

EndWhile

Set $r(c)$ to k

Output $r'(c) = f(c)r(c)$;

End

where $C = (C, f)$ represents a scene, C is a rectangular array of voxels, and f is a function that assigns to every voxel an image intensity value. The ball radius k is iteratively increased starting from one, and the algorithm checks for $FO_{k,\mu}(c)$, the fraction of the object containing c that is contained in the ball and μ indicates the size of the voxel in all

directions. When this fraction falls below the threshold t_s , it is considered that the ball contains an object region different from that to which c belongs. Following the recommendation in Ref. 31, $t_s = 0.85$ is chosen. Finally, the ball size is multiplied by voxel intensity. (W_ψ is a homogeneity function and we use zero-mean unnormalized Gaussian function for it, further details on how to choose homogeneity function and $FO_{k,\mu}(c)$ can be found in Refs. 31 and 32.)

Since histogram of the b-scale image contains only the information about the radius of the balls, it is fairly easy to eliminate small ball regions and obtain a few largest balls by applying simple thresholding to the b-scale scene. This simple observation leads to selection of larger or smaller objects in the scene by using only a threshold interval. For example, the first row in Fig. 2 shows different slices of the b-scale scene of an abdominal CT image and the remaining rows except the last show thresholded b-scale scenes obtained using different threshold intervals based on balls' size. We observe that the patterns pertaining to the largest balls retained after thresholding have strong correlations with the truly delineated objects shown in the last row of the Fig. 2. The truly delineated objects and patterns obtained after thresholding share some global similarities, for instance, of scale, location, and orientation. Patterns show salient characteristics because they depend on the object scale estimation,

and they are mostly spatially localized. Therefore, a concise but reliable relationship can be built using scale, position, and orientation information as parameters. Note that thresholding is applied not on the original images, but in a space generated by the object scale estimation algorithm, namely intensity weighted b-scale images. Although the range of object scale estimation is restricted to the size of objects retained in the scene, thresholding allows us to select the specific object size to be retained in the scene. In other words, a considerable amount of information is still captured in both the spatial distribution of intensities and object scale information of the image even if thresholding is not applied to the original images. As easily noticed from the forth and fifth rows of Fig. 2, the thresholded scenes have stronger correlations with the corresponding truly delineated objects shown in the last row of the same figure. Similarly, flexibility in selecting threshold intervals is a desirable property for creating a similarity group between shape and intensity structure system, because it allows us to avoid some of the redundant structures and provide a more concise basis for shape patterns.

In recognition, as the aim is to recognize *roughly* the whereabouts of an object of interest in the scene, and also since the trade-off between locality and conciseness of shape variability will be modulated in the delineation step, it will be sufficient to use concise bases produced by PCA without considering localized variability of the shapes. For the former case, on the other hand, it is certain that analyzing variations for each subject separately instead of analyzing variations over averaged ensembles leads to exact solutions where specific information present in the particular image is not lost. In order to find the translation, scale, and orientation that best align the shape structure system of the model with the intensity structure system of a given image, we learn the similarity of shape and intensity structure systems in the training images via PCA to keep track of translation and orientation differences. We use the bounding box approach to find scale similarity. In the bounding box approach, *the real physical size* of the segmented objects and the structures derived from thresholded intensity weighted b-scale images are used. For orientation analysis, parameters of variations are computed via PCA. The principal axes systems (PAS) of the shape and intensity structure systems denoted by PA_o and PA_b , respectively, have an origin and three axes representing the inertia axes of the structure. For the PAS of the same subject, the relationship function F that maps PA_b into PA_o can be decomposed into the form $F = (s, t, R)$, where $t: (t_x, t_y, t_z)$ is the translation component, s is a scale component, and $R: (R_x, R_y, R_z)$ represents three rotations. We observe that F can be split into three component functions (f_1, f_2, f_3) , corresponding to scale, translation, and rotation, respectively.

II.C.1. Estimation of the scale function – f_1

The bounding box enclosing the objects of interest for each subject in the training set is used to estimate the real physical size of the objects in question.²⁶ The length of the

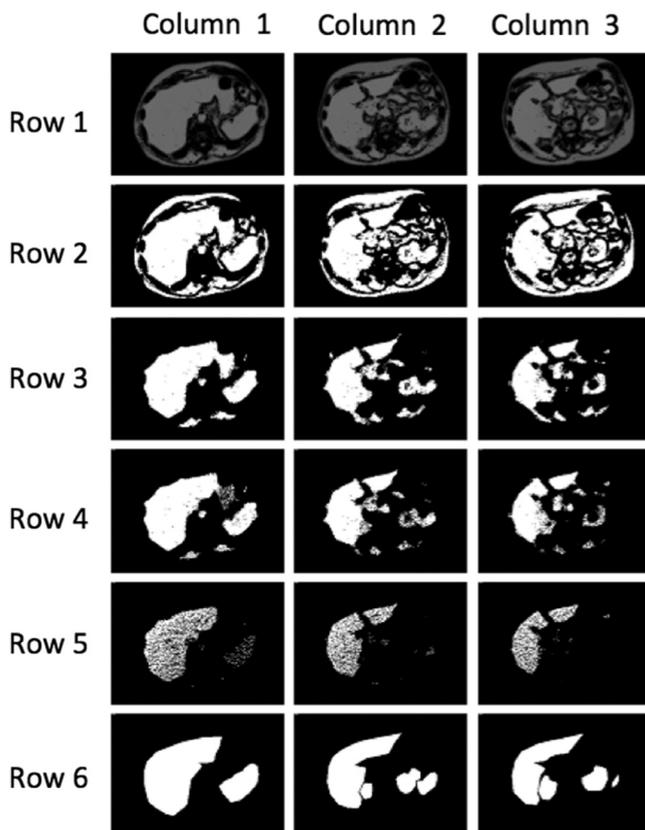


FIG. 2. Different slices of intensity weighted b-scale scenes extracted from a CT image (female subject, abdominal region) are shown in the first row. Second–fifth rows show corresponding thresholded intensity weighted b-scale scenes for increasing thresholds. The last row denotes some truly segmented objects of the abdominal region.

diagonal is used for estimating the scale parameter. The mean scale parameter s' and standard deviation of scale parameter $\text{std}(s)$ are used to obtain an interval for the estimation. We assume that the training set captures the variability of size differences such that the scale interval $[s' \pm \text{std}(s)]$ is used to estimate the scale of any given image.

Alternatively, volume or surface information of segmented objects in the training set can also be used as scale parameters. However, computational cost in the extraction of volume or surface information will be higher. In other words, if computational complexity of extracting scale parameters is denoted by $O(\text{number of voxels})$, it is obvious that the *number of voxels* necessary for computing the volume and surface are higher than the *number of voxels* necessary for computing the bounding box.

II.C.2. Estimation of the translation function – f_2

This is solely based on forming a linear relationship between the centroids of the objects of interest obtained from the binary images \mathbf{I}_i^b in the training set, and the thresholded intensity weighted b-scale images obtained from the training images. These centroids are denoted by \mathbf{c}_o^i and \mathbf{c}_b^i , respectively. By averaging the translational vector over N subjects in the training set from \mathbf{c}_b to \mathbf{c}_o , we get the mean translation vector as: $\bar{t} = (1/N) \sum_{i=1}^N (\mathbf{c}_o^i - \mathbf{c}_b^i)$. For any given test image, we simply estimate the centroid of objects in it using mean and standard deviation of t .

There are a few alternatives to estimating the centroids of shapes. The most straightforward way is as the average of the shape points' coordinates in the binary image. The centroid can also be obtained by using its gray valued image if available. Shape points' coordinates in this case are weighted by the intensity values of the voxels, leading to appearance-based centroid. The goal in the latter approach is to increase the correlation of two structures by considering not only shape features but also texture features. Without using textural information, two structures may show similar centroids, although they are far apart. Note that accuracy difference between intensity based and shape based centroids can differ a lot based on the two facts (1) the size of the object and (2) textural variability of the objects. Accuracy change due to with and without using textural information in centroid estimation for objects with small size can vary from 1 to 5 mm (for kidneys with large texture variability), and up to 10 cm difference (for livers with large texture variability).

As seen from the Fig. 3, geometric centers (shown in diamond) show no difference about the poses of the shapes as the centers are in the same coordinates for both shapes. On the other hand, appearance-based centroids (shown in square) give more accurate information about the poses of the shapes. Therefore, we use appearance-based centroids to build the f_2 component of F in our experimental setup.

II.C.3. Estimation of the orientation function – f_3

Let the normalized principal axes systems of the shape and intensity structure systems be PA_{oi} and PA_{bi} , respec-

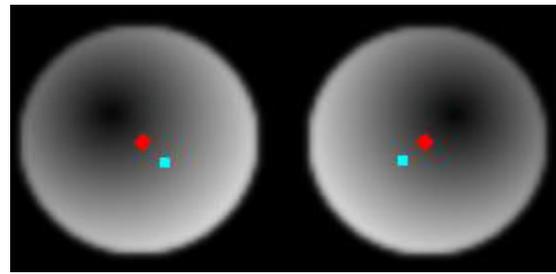


Fig. 3. Example textured shapes. Geometric and appearance based centroids are shown in diamond and square respectively. Region based centroids are obtained by weighting shape points' coordinates with the corresponding intensity values.

tively, obtained from the i th training image. Since these principal component vectors constitute an orthonormal basis and assuming that the translation between the two systems is eliminated from using $(\mathbf{c}_o^i - \mathbf{c}_b^i)$ estimated previously in translation estimation step, two systems are related by $PA_o = (R).(PA_b)$, where R is an orthonormal rotation matrix carrying information about the relative positions of shape and intensity structure systems in terms of their Euler angles. A set of N segmented training images and their corresponding intensity weighted b-scale images are used to find their PA systems so that we can relate them by computing the orthogonal rotation matrices R_i that relate PA_{oi} to PA_{bi} for $i = 1, \dots, N$. At the end, we have N rotation matrices describing how PA_b is related to PA_o for each subject in the training set. To obtain the basic population statistics over these N subjects, we need to compute the mean and standard deviation of the N rotation matrices R_i , $i = 1, \dots, N$. However, computing the statistics of circular (spherical) or directional data is not trivial.³³ Since three-dimensional orientation data are elements of the group of rotations that generally are given as a sequence of unit quaternion or as a sequence of Euler angles, etc., the group of rotations is not a Euclidean space, but rather a differentiable manifold.^{33,34} Therefore, the notion of mean or average as basic statistical definitions for this particular problem is not obvious in Euclidean space. In our case, in analogy with the mean in Euclidean space, mean rotation is defined to be the minimization of the sum of squared geodesic distances from the given rotations in spherical space. Note that the mean rotation R^* is assumed to be a point on the sphere such that the sum of squared geodesic distances between R^* and R_1, \dots, R_N is the minimum

$$\mathcal{M}(R_1, \dots, R_N) = \arg \min_{R^*} \sum_{n=1}^N d(R_n, R^*)_G^2 \quad (1)$$

where $d(\cdot)_G$ represents geodesic distance form in Riemannian manifold and $\mathcal{M}(\cdot)$ represents mean operation in the spherical space.

Figure 4 shows smooth path across rotation on the sphere where metric tensor is determined by arc-length. PA systems differ from each other only by orientation as shown in the figure. The similar colors show the corresponding Euler angles in spherical coordinate systems.

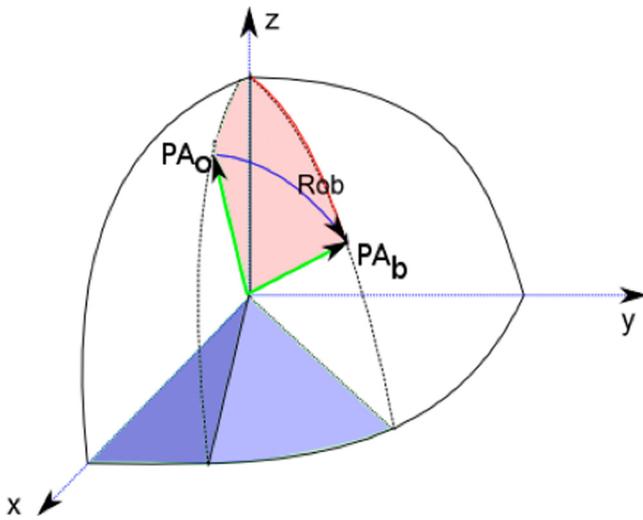


FIG. 4. The shape and intensity structure systems, PA_o and PA_b , are shown in the spherical coordinate system with their Euler angles drawn in similar colors. Any orientation difference between the PA systems requires the computation of another orthonormal rotation denoted R_{ob} , which rotates the shape structure system into alignment with the intensity structure system on the sphere.

II.D. Organ delineation

The input of our delineation part is the recognized result from the organ recognition part. We propose an IGCASM method for the organ's delineation. The IGCASM algorithm effectively combines the rich statistical shape information embodied in ASM (Ref. 3) with the globally optimal 3D delineation capability of the GC method.

II.D.1. Shape integrated GC

For GC segmentation, we represent the image as a six-connectivity graph $G(V, E)$. Boykov's α -expansion method⁶ is used as the optimization method. In the traditional GC method, the energy function that is minimized usually consists of two parts: data penalty and boundary penalty terms. In this paper, we propose a new graph cut energy function, which additionally consists of a 3D shape term.

$$E(f) = \sum_{p \in P} (\alpha \cdot D_p(f_p) + \beta \cdot S_p(f_p)) + \sum_{p \in P, q \in N_p} \gamma \cdot B_{p,q}(f_p, f_q), \quad (2)$$

where P is the set of pixels p , N_p is the set of pixels in the neighborhood of p , α, β, γ are the weights for the data term, shape term S_p , and boundary term, respectively, satisfying $\alpha + \beta + \gamma = 1$. These components are defined as follows:

$$D_p(f_p) = \begin{cases} -\ln P(I_p|O), & \text{if } f_p = \text{object label} \\ -\ln P(I_p|B), & \text{if } f_p = \text{background label} \end{cases}, \quad (3)$$

$$B_{p,q}(f_p, f_q) = \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{d(p, q)} \cdot \delta(f_p, f_q), \quad (4)$$

and

TABLE I. Different numbers of landmarks for different objects are listed.

CT-data	Number of landmarks	MRI-data	Number of landmarks
Skin	75	Skin	70
Liver	35	Calcaneus	35
Left kidney	20	Talus	35
Right kidney	20	Tibia	35
Spleen	25	Cuboid	35
—	—	Navicular	35

$$\delta(f_p, f_q) = \begin{cases} 1, & \text{if } f_p \neq f_q \\ 0, & \text{otherwise,} \end{cases}$$

where I_p is the intensity of pixel p , object label is the label of the object (foreground). $P(I_p|O)$ and $P(I_p|B)$ are the probability of intensity of pixel p belonging to object and background, respectively, which are estimated from object and background intensity histograms during the training phase, more details are given below $d(p, q)$ is the Euclidian distance between pixels p and q , and σ is the standard deviation of the intensity differences of neighboring voxels along the boundary.

$$S_p(f_p) = 1 - \exp\left(-\frac{d(p, x_o)}{r_o}\right), \quad (5)$$

where $d(p, x_o)$ is the distance from pixel p to the set of pixels, which constitute the interior of the current shape x_o of object O . (Note that if p is in the interior of x_o , then $d(p, x_o) = 0$.) r_o is the radius of a circle that just encloses x_o . The linear time method of Ref. 35 was used in this paper for computing this distance.

During the training stage, the histograms of intensity for each object are estimated from the training images. Based on this, $P(I_p|O)$ and $P(I_p|B)$ can be computed. As for parameters α, β , and γ in Eq. (1), since $\alpha + \beta + \gamma = 1$, we estimate only α and β by optimization of accuracy as a function of α and β and set $\gamma = 1 - \alpha - \beta$. We use the gradient decent method³⁶ for the optimization. $Accu(\alpha, \beta)$ represents the algorithm's accuracy (here, we use the true positive volume fraction.³⁷ α and β are initialized to 0.35 each, and $Accu(\alpha, \beta)$ is optimized over the training data set to determine the best α and β .

II.D.2. Minimizing E with graph cuts

Let G be a weighted graph (V, E) , where V is a set of nodes and E is a set of weighted arcs. Given a set $T \subseteq V$ of k terminal nodes, a cut is a subset of edges $C \subseteq A$ such that no path exists between any two nodes of T in the residue graph $G(V, E \setminus C)$. In our implementation, we segment the object by the α -expansion method.⁵

The graph is designed as follows. We take $V = P \cup L$, i.e., V contains all the pixel nodes and multiple terminals corresponding to the labels in L that represent objects of interest plus the background. $A = A_N \cup A_T$, where A_N is the n -links, which connect pixels p and q ($p \in P, q \in N_p$) and

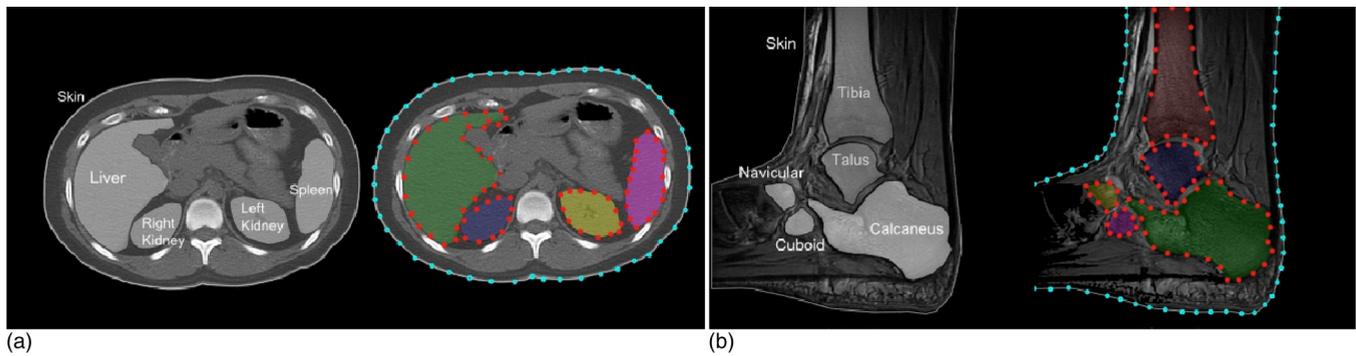


FIG. 5. (a) A CT slice of the abdominal region with selected objects (skin, liver, spleen, and left and right kidneys) is shown on the left. Annotated landmarks for the selected objects are shown on the right. (b) An MRI slice of the foot with selected objects (skin, navicular, calcaneus, tibia, talus, and cuboid) is shown on the left. Annotated landmarks for the selected objects are shown on the right.

with a weight of w_{pq} . A_T is the set of t -links, which connect pixel p and terminals, $\ell \in L$ and with a weight of $w_{p\ell}$. The desired graph with cut cost $|C|$ equaling $E(f)$ is constructed using the following weight assignments:

$$w_{pq} = \gamma \cdot B_{p,q} \quad (6)$$

$$w_{p\ell} = K - (\alpha \cdot D_p(\ell) + \beta \cdot S_p(\ell)), \quad (7)$$

where K is constant that is large enough to make the weights $w_{p\ell}$ positive.

II.D.3. IGCASM

For IGCASM, we assume that the recognized shapes are sufficiently close to the actual boundaries. It then determines what the new position of the landmarks of the objects represented in x_{in} (initialized shape results) should be such that the minimum graph cut cost is the smallest possible.

Algorithm IGCASM:

Input: Initialized shapes x_{in} .

Output: Resulting shapes x_o and the associated object boundaries.

Begin

While number of iterations $< k$ **do**

 Perform GC segmentation based on ASM shapes x_{in} (one object at a time);

 Compute the new position of the landmarks by moving each landmark in x_{in} to the point closest on the GC boundary of the corresponding object; call the resulting shapes x_{new} ;

If no landmarks moved, then set x_{new} as x_o and **break**;

Else subject x_{new} to the constraints of model M , and set the result as x_{in} .

EndWhile

If number of iterations = k , set x_{in} as x_o .

 Perform one final GC segmentation based on x_o , and get the associated object boundaries.

End

From our experimental results, we found our algorithm is usually ended before three iterations due to good recognition results. Based on that, we set $k=6$. It means “stop” happened at most time. Landmarks usually moved far (5–6 pixels) at the beginning and less (1–2 pixels) at the end.

One may surmise, if it is possible to analyze theoretically if the altered GC formulation obeys the submodularity criteria. Although the achieved results by the proposed method strongly indicates that the proposed GC functional is submodular, it is very difficult to take the analysis all the way through since the parameter space depends not just on the intensity characteristics of the images but their spatial distributions and the shape characteristics. The submodularity analysis of the proposed GC functional is outside the scope of this paper and will be regarded as a future extension of this work.

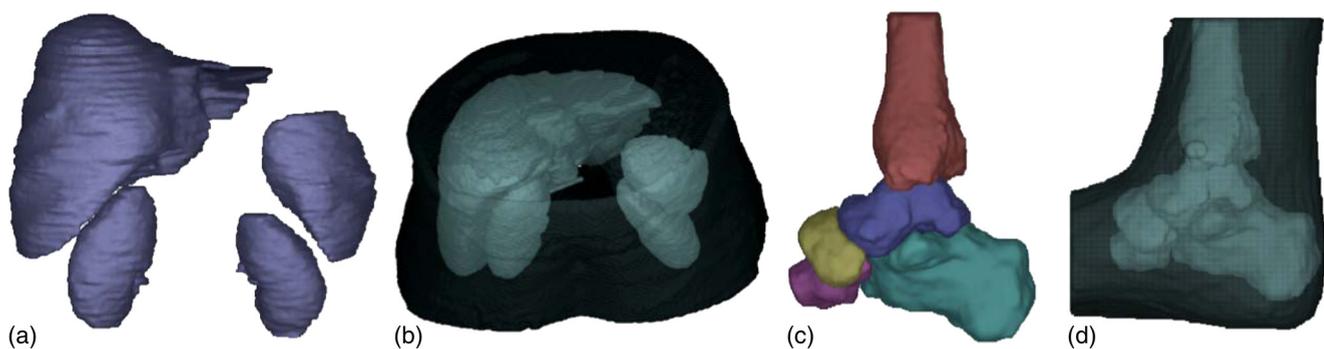


FIG. 6. Mean shape is generated using 3D-ASM for multiple objects of the abdominal region. (a) Mean shapes of liver, spleen, right and left kidneys. (c) Mean shapes of calcaneus, talus, tibia, navicular, and cuboid. (b) and (d) Mean shapes of skin boundaries for the objects presented in (a) and (c), respectively.

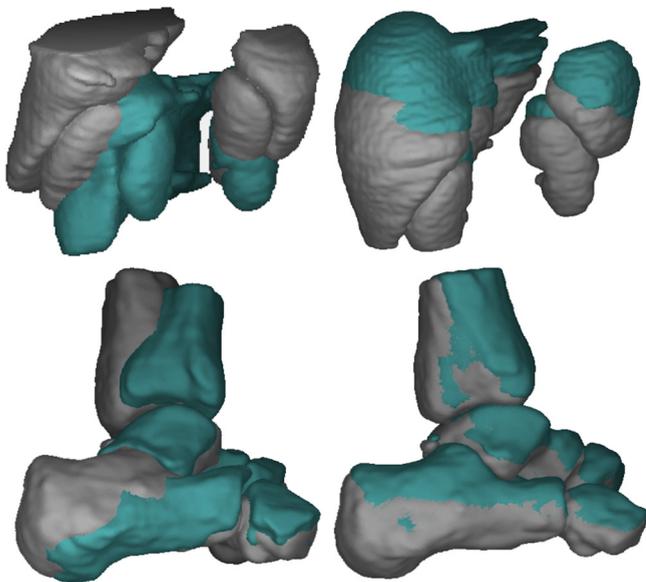


FIG. 7. (First column) the model assembly (MA) is overlaid with the organs/objects of one subject prior to recognition. (Second column) positioned MA for the subject is shown after recognition.

III. EXPERIMENTAL RESULTS

In this section, we demonstrate qualitatively, through image display, and quantitatively, through evaluation experiments, the extent of effectiveness of the IGCASM. Performance of the proposed methodology has been evaluated on two datasets: male and female abdominal organs in low-resolution CT images pertaining to 20 patients, and foot MR images pertaining to 11 patients. Our method of evaluation, based on the framework of Ref. 37 will focus on the analysis of accuracy, and efficiency of IGCASM. We will consider manual segmentation performed in these different data sets. We used a manual delineation method to constitute a surrogate of true segmentation. For all the data sets, expert radiologists labeled the data in a semiautomatic way using the live wire method¹² in pseudo-3D manner (slice-by-slice).

III.A. Image data sets and model building

For CT images, we used whole body PET-CT scans of ten female and ten male patients, who underwent prior abdominal CT imaging for clinical purposes in University of Pennsylvania. The voxel size of the CT images is $1.17 \text{ mm} \times 1.17 \text{ mm} \times 1.17 \text{ mm}$ (interpolated from 5 mm slices). Due to all clinically important reasons, we focus on selecting

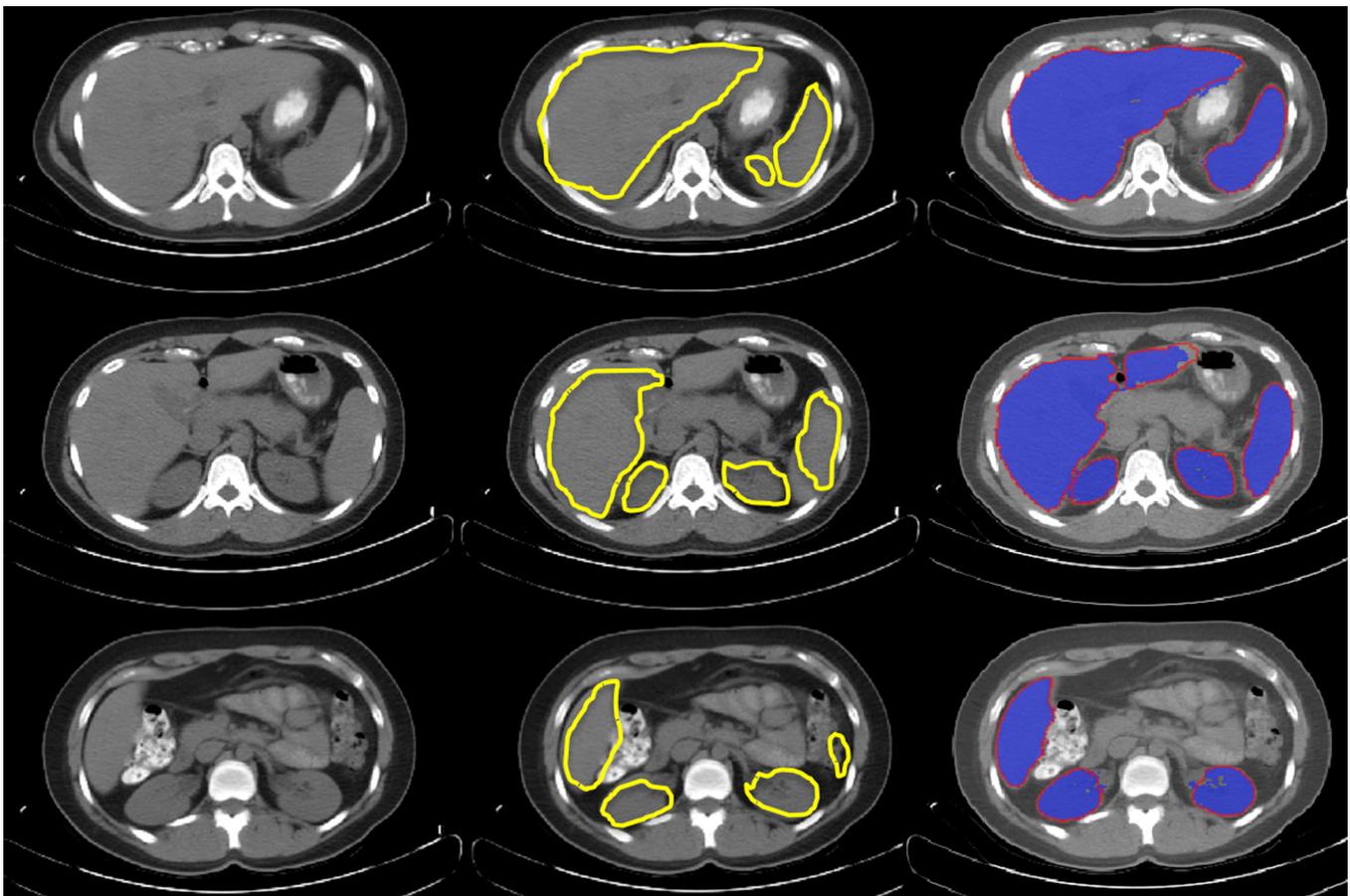


FIG. 8. The experimental results for multiorgan segmentation are shown in three different anatomical levels for CT abdominal dataset. The first column shows original images slices; the second column indicates the recognized organs; and the third column shows the delineation results yielded by the proposed IGCASM. The contours in third column shows manually delineated organ boundaries. All of the images have been cropped for the best view and original image size is (512×512) .

TABLE II. Quantitative measure of the proposed recognition method.

Quantitative measures	Mean translation error (in mm)	Mean orientation error (degree)—(x)	Mean orientation error (degree)—(y)	Mean orientation error (degree)—(z)	Mean scale error
Female abdominal data	8.0133 ± 1.9666	$0.6319^\circ \pm 3.3861$	$0.3611^\circ \pm 4.0832$	$0.2789^\circ \pm 9.4338$	0.03 ± 0.0100
Male abdominal data	10.7566 ± 1.9276	0.0938 ± 12.0647	0.3208 ± 3.0894	0.6203 ± 3.5691	0.05 ± 0.0250
Foot MRI	3.5709 ± 4.2200	0.0292 ± 0.7088	0.3576 ± 2.0739	0.0209 ± 9.4049	0.025 ± 0.0050

as many healthy organs as we can select from the abdominal regions of individuals attended the scanning sessions. Selection process had been performed with the help of experts, who were with expertise in supervising and interpreting radi-

ology examinations, and then, we have selected the following four objects from each subject: liver, left kidney, right kidney, and spleen. Body skin is also considered in our study to constraint the search space for the four selected objects.

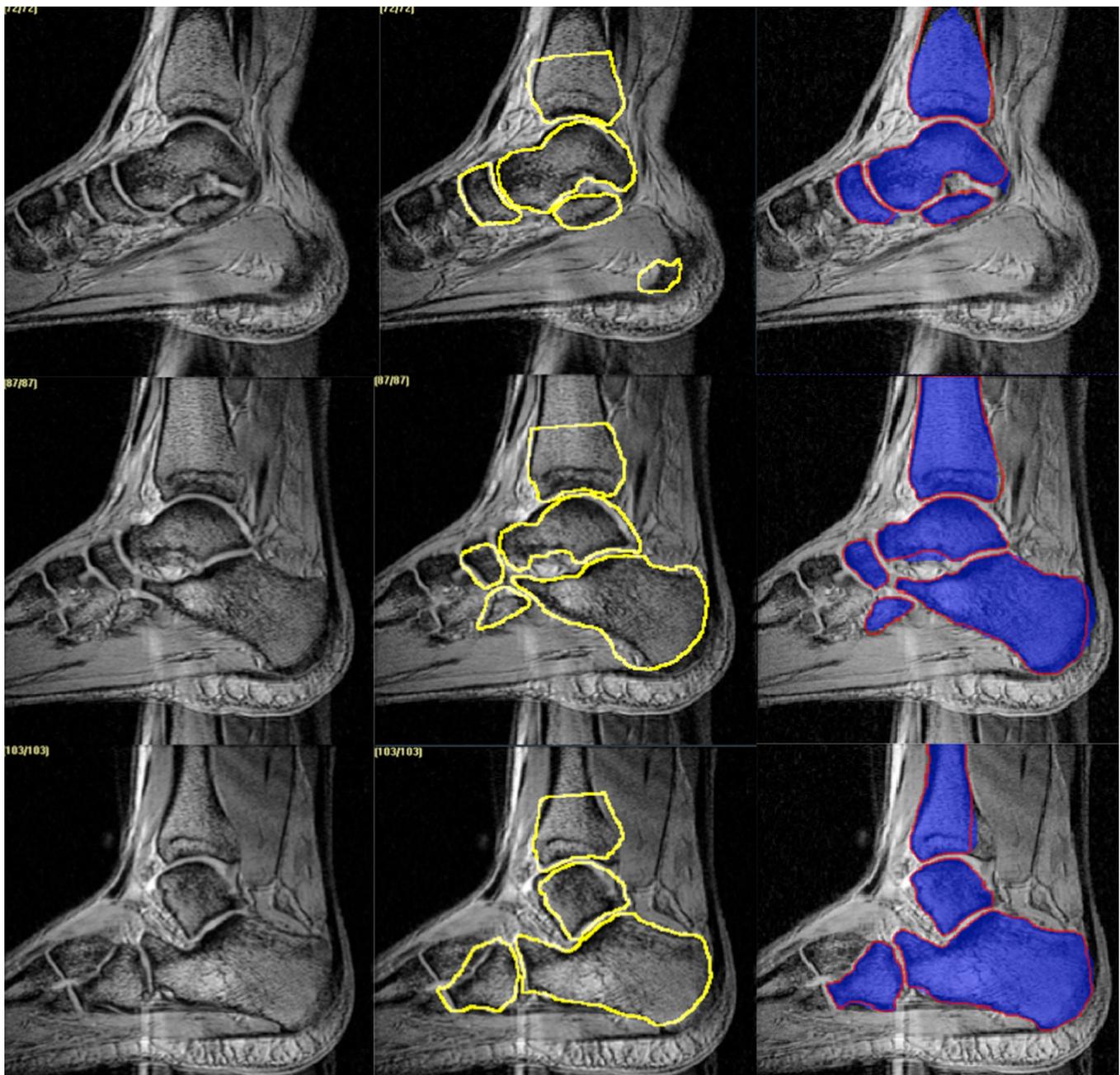


FIG. 9. The experimental results for multiorgan segmentation are shown in three different anatomical levels for foot MRI dataset. The first column shows original images slices; the second column indicates the recognized organs; and the third column shows the delineation results yielded by the proposed IGCASM. The contours in third column shows manually delineated bone boundaries. All of the images have been cropped for the best view and original image size is (512×512) .

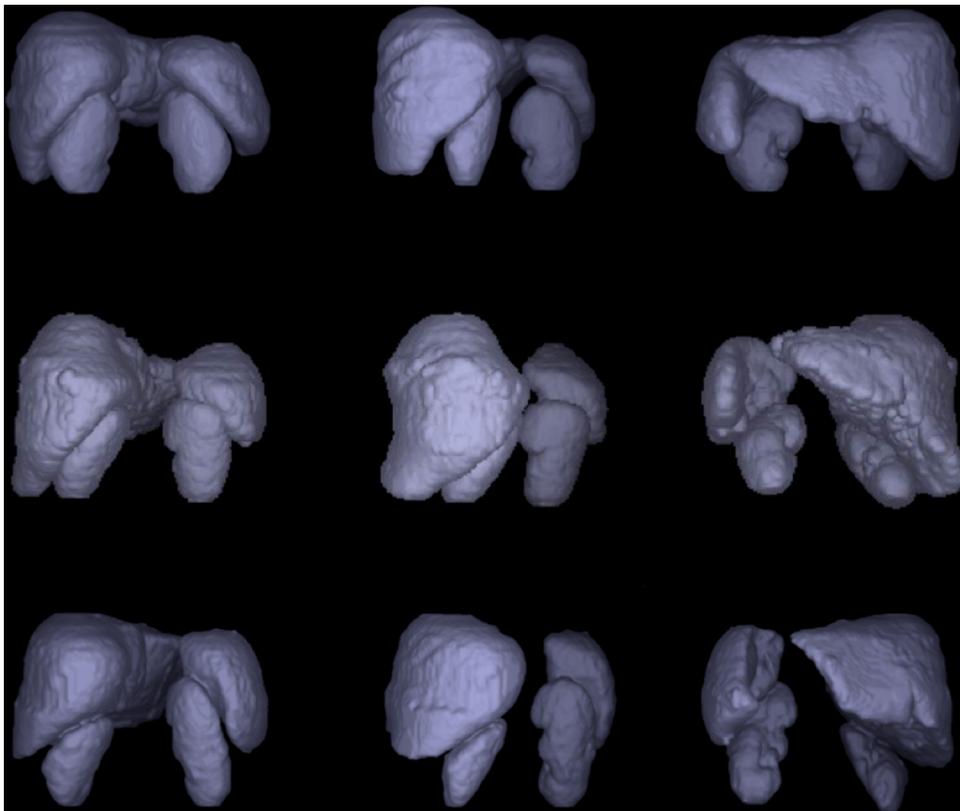


FIG. 10. Three different views of delineation results for three examples: CT abdominal organs by IGCASM.

In addition, we also used 11 sets of MR foot images to conduct segmentation experiments. Magnetic resonance imaging (MRI) becomes widespread in the diagnosis and treatment of many musculoskeletal diseases of the ankle and foot. Anatomical structures in the bones and soft tissues are demonstrated before they become evident at other imaging modalities. The data were acquired on a commercial 1.5T GE MRI machine, by using a coil specially designed for the study.³⁸ During each acquisition, the foot of the subject was locked in a nonmagnetic device. This allows the control of orientation and the motion of the foot. The imaging protocol used a 3D steady-state gradient echo sequence with a TR/TE/Flip angle = 25 ms/10 ms/25°. The voxels are of size $0.55 \times 0.55 \times 0.55 \text{ mm}^3$ (interpolated from slices 1.5 mm apart). The slice orientation was roughly sagittal. Apart from the skin body, we select all the bones clearly seen in sagittal orientation, i.e., talus, navicular, calcaneus, cuboid, and tibia.

Different numbers of landmarks are used for different objects considering their size, as listed in Table I. Figure 5 shows annotated landmarks for five different objects (skin, liver, right kidney, left kidney, and spleen) in a CT slice of the abdominal region, and six different objects (skin, talus, navicular, calcaneus, cuboid, and tibia) in a MRI slice of the foot.

Figure 6 shows multiobject 3D ASMs for the abdominal organs and foot bones. Note that the mean shapes of the objects do not have any overlap among them. This is because, in the training part, objects are not aligned separately, their spatial relations before and after alignment do not change.

III.B. Recognition

Figure 7 demonstrates the effectiveness of the recognition method by displaying the abdominal organs and the MA for one particular example. Figure 7 (left) displays 3D surface renditions of the organs and MA before the recognition method is applied. Figure 7 (right) similarly shows 3D renditions of the organs and the MA after the recognition method is applied. Note that the differences between the principal axes and the centroids of the objects in the left are reduced considerably in the right.

The recognition results for one particular patient in three different anatomical levels were shown in second column of Fig. 8. Note that slices of organs or objects in the beginning or end of stack may not have significant overlaps with the model; however, this situation never exceeds a few slices. As previously mentioned, aim in this step is to roughly localize the model to the data so that delineation can capture local details. We can observe that the recognition results are pretty good by visual checking. In addition to the qualitative results, we report the recognition accuracies in terms of quantitative measures. Table II summarizes the recognition results in three categories: mean translation error, mean orientation errors in three directions, and mean scale error showing size differences among objects from different subjects. We can find that all the objects are recognized with the mean translation error less than 8 mm, and their orientation errors are less than 10° , even much smaller in x and y directions. A possible reason for this is that the spatial resolution in the z direction is lower than in other directions. For the scale component, since initially we align all

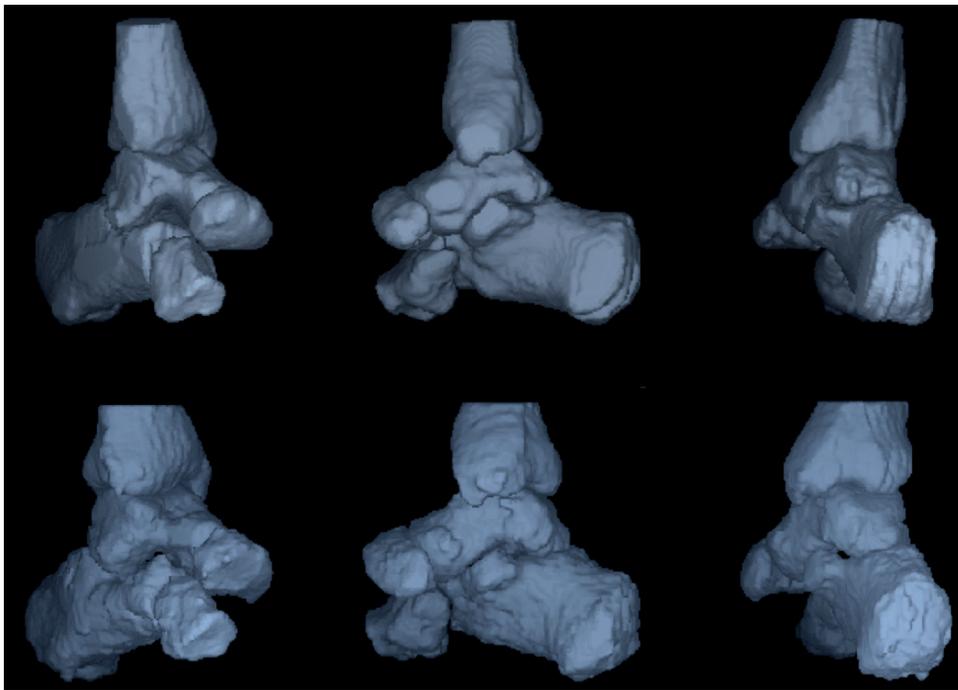


FIG. 11. Three different views of delineation results for two examples: MRI foot bones by IGCASM.

the objects in seven-dimensional affine space as described in landmarking process, the size differences within the subjects are uniformly handled. Very tight interval is obtained for the scale range, which is reported as (0.97–1.07) on average and the mean scale error is about 0.03.

III.C. Delineation

Figures 8 and 9 show the delineation results for one particular patient in three different anatomical levels for CT and MRI dataset, respectively. Furthermore, Figs. 10 and 11 show 3D views of delineation results by IGCASM from three different views on three examples for CT and MRI dataset, respectively. We can observe that the delineation results are pretty good by visual checking.

In addition to the qualitative results, we also report delineation accuracies in terms of quantitative measures. Here, we focus on the analysis of accuracy and efficiency of IGCASM and compare it with ASM. Accuracy relates to how well the segmentation results agree with the true delineation of the objects. Efficiency indicates the practical viability of the method, which is determined by the amount of time required for performing computations.

The results of delineation accuracy are expressed in true positive volume fractions (TPVF) and false positive volume fractions (FPVF).³⁷ TPVF indicates the fraction of the total amount of tissue in the true delineation by the method. FPVF denotes the amount of tissue falsely identified by the method. The delineation results are expressed in Table III. We can find that comparing with ASM method, the delineation accuracy

TABLE III. Mean and standard deviation of delineation results as TPVF, FPVF for 3D ASM, and IGCASM on CT abdominal and foot MRI data.

Data set	TPVF(%)		FPVF(%)	
	3D ASM	IGCASM	3D ASM	IGCASM
Liver	82.52 ± 1.93	92.16 ± 1.03	0.96 ± 0.11	0.25 ± 0.06
Left kidney	86.15 ± 1.51	93.39 ± 0.96	0.85 ± 0.12	0.19 ± 0.05
Right kidney	85.36 ± 1.52	93.55 ± 0.92	0.89 ± 0.13	0.20 ± 0.03
Spleen	84.32 ± 1.88	93.47 ± 1.28	0.91 ± 0.16	0.23 ± 0.07
Average on four organs	84.59 ± 1.71	93.01 ± 1.05	0.90 ± 0.13	0.22 ± 0.05
Calcaneus	83.76 ± 1.86	94.63 ± 0.91	1.03 ± 0.27	0.33 ± 0.12
Talus	84.32 ± 1.36	94.89 ± 0.97	0.92 ± 0.35	0.27 ± 0.09
Tibia	82.85 ± 1.91	92.36 ± 1.27	1.21 ± 0.29	0.28 ± 0.06
Cuboid	82.22 ± 1.95	93.68 ± 1.11	1.12 ± 0.32	0.25 ± 0.08
Navicular	83.65 ± 2.12	93.17 ± 1.29	1.16 ± 0.36	0.26 ± 0.07
Average on five bones	83.36 ± 1.84	93.75 ± 1.11	1.07 ± 0.32	0.28 ± 0.08

by proposed IGCASM gets much improved. The accuracy for kidneys and spleen segmentation can be achieved about 93.5% in TPVF, and 0.21% in FPVF; and the delineation accuracy for liver segmentation is about 92.1% in TPVF, and 0.25 in FPVF. The accuracy for calcaneus and talus segmentation can be achieved about 94.6% in TPVF, and 0.30% in FPVF; and the delineation accuracy for cuboid, tibia, and navicular segmentation is about 93.0% in TPVF, and 0.25 in FPVF. It is important to notice that the segmentations for these organs and bones are done simultaneously. Here, not only the accuracy of IGCASM is better than those from ASM but also it provides more comprehensive recognition strategies to initiate the delineation algorithm without doing any search. We did not show the accuracy for GC as the traditional GC method is not a fully automatic method, rather an interactive method, and input seeds are required manually. If we compare our proposed method to traditional GC but now without inputting seeds to make it fully automatic, the performance is predictably very poor. The whole delineation process for these four organs and five foot bones takes only about 78 and 70 s, respectively, on average on an Intel Pentium IV PC with a 3.4 GHz CPU machine. Notice that this time does not include the time required for recognition, which is around 50 s on average. In addition, the running time of the delineation is comparable to the traditional ASM searching method, which is about 70 s for abdominal organs, and 65 s for foot bones.

IV. CONCLUDING REMARKS

We propose an automatic anatomy segmentation method for the body region. The AAS system, we are developing consists of two main parts: object recognition and object delineation. In this paper, a hierarchical 3D scale-based multiobject method is used for the multiobject recognition task. For object delineation, an iterative graph-cut-ASM algorithm is proposed, which effectively combines the rich statistical shape information embodied in ASM with the globally optimal delineation capability of the GC method. The presented IGCASM algorithm is a 3D generalization of the 2D GC-ASM method that we proposed previously in Ref. 27. The proposed methods are tested on a CT dataset comprised of images obtained from 20 patients of clinical abdominal scans and a foot MRI dataset with 11 patients. Our goal was to segment the liver, spleen, left kidney, and right kidney simultaneously for abdominal CT dataset, and calcaneus, tibia, cuboid, talus, and navicular for foot MRI dataset. The recognition accuracies in terms of translation, rotation, and scale error over all organs are about 8 mm, 10° and 0.03, and over all foot bones are about 3.5709 mm, 0.35° and 0.025, respectively. The accuracy of delineation over all organs for all subjects as expressed in TPVF and FPVF is 93.01% and 0.22%, and all foot bones for all subjects are 93.75% and 0.28%, respectively. In summary, the experimental results show: (a) the incorporation of shape priors into the GC framework is feasible in 3D as demonstrated previously for 2D images; (b) our preliminary results in 3D confirm the accu-

racy behavior observed in 2D. The hybrid strategy IGCASM seems to be more robust and accurate than ASM and GC individually; (c) delineations within body regions of clinical importance can be accomplished quite rapidly within 1.5 min.

In this paper, as for the recognition method, the multiobject strategy³⁹ is applied. As demonstrated in 2D images,³⁹ with increasing the number of objects in the model, both recognition and delineation accuracy can get significantly improved. Here, the skin object is included in the model to help the organ initialization.

Some more ideas underlying IGCASM can be further discussed. In this paper, we proposed a hybrid method IGCASM, which combines the model-based method (ASM) and image-based method (GC), and aim to combine the complementary strengths of ASM and GC. For the recognition, we incorporate intensity weighted b-scale²⁶ information into the ASM model. This is accomplished by b-scale encoding.²⁶ With the intensity weighted b-scale information, it can distinguish the objects of even same size. Actually this is a bit similar to the concept of the active AAM. By combining GC with AAM,⁴⁰ the whole system performance of recognition and delineation accuracy may get improved due to more texture information contained in AAM model than ASM model. These will be investigated in the near future. In this paper, we have not addressed the issue of handling abnormalities due to diseases or treatment. We believe that modeling should be (and perhaps can be) done only of normality, and through its knowledge, abnormality should be detected and delineated in given patient images. This is a topic of our current research.

ACKNOWLEDGMENTS

We would like to thank Prof. Jayaram K. Udupa for his nice suggestions and advices, and thank Dr. Drew A. Torigian for providing the data.

^aAuthor to whom correspondence should be addressed. Electronic mail: chenx6@mail.nih.gov; Telephone: 301-594-8139; Fax: (301) 480-9827.

^b<http://xinjianchen.wordpress.com>

¹M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.* **1**, 321–331 (1987).

²T. McInerney and D. Terzopoulos, "Deformable models," in *Handbook Medical Imaging: Processing and Analysis*, edited by I. Bankman (Academic, San Diego, 2000).

³T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models – their training and application," *Comput. Vis. Image Underst.* **61**, 38–59 (1995).

⁴T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**, 681–685 (2001).

⁵T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image Vis. Comput.* **12**, 355–366 (1994).

⁶Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(12), 1222–1239 (2001).

⁷Y. Boykov, and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004).

⁸D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: Integrating Color, Texture, Motion and Shape," *Int. J. Comput. Vis.* **72**(2), 195–215 (2007).

- ⁹R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(2), 158–175 (1995).
- ¹⁰J. B. T. M. Roerdink and A. Meijster, "The watershed transform: Definitions, algorithms and parallelization strategies," *Fund. Inform.* **41**, 187–228 (2001).
- ¹¹V. Grau, A. U. J. Mewes, M. Alcañiz, R. Kikinis, and S. K. Warfield, "Improved watershed transform for medical image segmentation using prior information," *IEEE Trans. Med. Imaging* **23**(4), 447–458 (2004).
- ¹²A. X. Falcao, J. K. Udupa, S. Samarasekera, and S. Sharma, "User-steered image segmentation paradigms: Live wire and live lane," *Graphical Models Image Process.* **60**, 233–260 (1998).
- ¹³E. N. Mortensen, B. S. Morse, W. A. Barrett, and J. K. Udupa, "Adaptive boundary detection using live-wire two-dimensional dynamic programming," in *IEEE Proceedings of Computers in Cardiology* (IEEE Computer Society, Durham, NC, 1992), pp. 635–638.
- ¹⁴J. M. Liu and J. K. Udupa, "Oriented active shape models," *IEEE Trans. Med. Imaging* **28**(4), 571–584 (2009).
- ¹⁵K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Process.* **7**(12), 1684–1699 (1998).
- ¹⁶J. Yang and J. S. Duncan, "3D image segmentation of deformable objects with joint shape-intensity prior models using level sets," *Med. Image Anal.* **8**(3), 285–294 (2004).
- ¹⁷V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.* **22**(1), 61–79 (1997).
- ¹⁸J. K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *Graphical Models Image Process.* **58**(3), 246–261 (1996).
- ¹⁹J. K. Udupa, P. K. Saha, and R. A. Lotufo, "Relative fuzzy connectedness and object definition: Theory, algorithms, and applications in image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 1485–1500 (2002).
- ²⁰N. Vu and B. S. Manjunath, "Shape prior segmentation of multiple objects with graph cuts," in *CVPR 2008* (IEEE Computer Society, Anchorage, AK, 2008), pp. 1–8.
- ²¹D. Freedman and T. Zhang, "Interactive graph cut based segmentation with shape priors," in *CVPR 2005* (IEEE Computer Society, San Diego, CA, 2005), pp. 755–762.
- ²²J. Malcolm, Y. Rathi, and A. Tannenbaum, "Graph cut segmentation with nonlinear shape priors," in *IEEE International Conference on Image Processing* (ICIP, San Antonio, TX, 2007), pp. 365–368, ISSN: 1522-4880.
- ²³P. Kohli, J. Rihan, M. Bray, and P. H. S. Torr, "Simultaneous segmentation and pose estimation of humans using dynamic graph cuts," *Int. J. Comput. Vis.* **79**(3), 285–298 (2008).
- ²⁴V. Lempitsky, A. Blake, and C. Rother, "Image segmentation by branch-and-mincut," *ECCV 2008, Part IV, LNCS 5305*, pp. 15–29.
- ²⁵M. Leventon, E. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours. In: *CVPR 2000* (2000).
- ²⁶Bağcı, U., Udupa, J. K. and Chen, X. J., "Ball-Scale Based Multi-Object Recognition in Hierarchical Framework", *Proc. SPIE*, **7623**, 762345 (2010).
- ²⁷X. J. Chen, J. K. Udupa, D. A. Torigian, and A. Alavi, "GC-ASM: Synergistic integration of active shape modeling and graph-cut methods," *Proc. SPIE*, **7259**, 72590C1–72590C-8 (2009).
- ²⁸X. Chen, J. K. Udupa, U. Bagci, A. Alavi, and D. Torigian, "3D automatic anatomy recognition based on iterative graph-cut-ASM," *SPIE Medical Imaging Paper No. 7625-63*, 2010.
- ²⁹C. Brechbuhler, G. Gerig, and O. Kubler, "Parameterization of closed surfaces for 3D shape description," *Comput. Vis. Image Underst.* **62**, 154–170 (1995).
- ³⁰C. G. Small, *The Statistical Theory of Shape* (Springer, New York, 1996).
- ³¹P. K. Saha, J. K. Udupa, and D. Odhner, "Scale-based fuzzy connected image segmentation: Theory, algorithms, and validation," *Comput. Vis. Image Underst.* **77**(2), 145–174 (2000).
- ³²P. Saha and J. K. Udupa, "Scale-based diffusive image filtering preserving boundary sharpness and fine structures," *IEEE Trans. Med. Imaging* **20**(11), 1140–1155 (2001).
- ³³M. Moakher, "Means and averaging in the group of rotations," *SIAM J. Matrix Anal. Appl.* **24**(1), 1–16 (2002).
- ³⁴U. Bagci, J. K. Udupa, and X. Chen, "Orientation estimation of anatomical structures in medical images for object recognition," *Proc. SPIE Med. Imaging* **7962**, 79622L (2011).
- ³⁵K. C. Ciesielski, X. Chen, J. K. Udupa, and G. J. Grevera, "Linear time algorithms for exact distance transform," *J. Math. Imaging Vis.* **39**, 193–209 (2010).
- ³⁶J. A. Snyman, *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms* (Springer, New York, 2005).
- ³⁷J. K. Udupa, V. R. Leblanc, Y. Zhuge, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn, "A framework for evaluating image segmentation algorithms," *Comput. Med. Imaging Graph.* **30**(2), 75–87 (2006).
- ³⁸J. K. Udupa, B. E. Hirsch, H. J. Hillstrom, G. R. Baurer, and J. B. Kneeland, "Analysis of in vivo 3-D internal kinematics of the joints of the foot," *IEEE Trans. Biomed. Eng.* **45**, 1387–1396 (1998).
- ³⁹X. Chen, J. K. Udupa, A. Alavi, and D. A. Torigian, "Automatic anatomy recognition via multiobject oriented active shape models," *Med. Phys.* **37**(12), 6390–6401 (2010).
- ⁴⁰X. Chen, J. Yao, Y. Zhuge, and U. Bagci, "3D automatic image segmentation based on graph cut-oriented active appearance models," in *Proceedings of 2010 IEEE 17th International Conference on Image Processing* (IEEE Signal Processing Society, Hong Kong, China, Sept. 26–19, 2010), pp. 3653–3656.