Medical Image Analysis 18 (2014) 752–771

Contents lists available at ScienceDirect

Medical Image Analysis

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

Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images

Jayaram K. Udupa [a,*], Dewey Odhner [a], Liming Zhao [a], Yubing Tong [a], Monica M.S. Matsumoto [a], Krzysztof C. Ciesielski [a,b], Alexandre X. Falcao [d], Pavithra Vaideeswaran [a], Victoria Ciesielski [a], Babak Saboury [a], Syedmehrdad Mohammadianrasanani [a], Sanghun Sin [e], Raanan Arens [e], Drew A. Torigian [c]

[a] Medical Image Processing Group, Department of Radiology, University of Pennsylvania, 423 Guardian Drive, Blockley Hall, 4th Floor, Philadelphia, PA 19104, United States
[b] Department of Mathematics, West Virginia University, Morgantown, WV 26506-6310, United States
[c] Department of Radiology, Hospital of the University of Pennsylvania, Philadelphia, PA 19104-4283, United States
[d] LIV, Institute of Computing, University of Campinas, Av. Albert Einstein 1251, 13084-851 Campinas, SP, Brazil
[e] Division of Respiratory and Sleep Medicine, Children's Hospital at Montefiore, 3415 Bainbridge Avenue, Bronx, NY 10467, United States

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form 11 April 2014
Accepted 11 April 2014
Available online 24 April 2014

Keywords:
Anatomy modeling
Fuzzy models
Object recognition
Image segmentation
Fuzzy connectedness

ABSTRACT

To make Quantitative Radiology (QR) a reality in radiological practice, computerized body-wide Automatic Anatomy Recognition (AAR) becomes essential. With the goal of building a general AAR system that is not tied to any specific organ system, body region, or image modality, this paper presents an AAR methodology for localizing and delineating all major organs in different body regions based on fuzzy modeling ideas and a tight integration of fuzzy models with an Iterative Relative Fuzzy Connectedness (IRFC) delineation algorithm. The methodology consists of five main steps: (a) gathering image data for both building models and testing the AAR algorithms from patient image sets existing in our health system; (b) formulating precise definitions of each body region and organ and delineating them following these definitions; (c) building hierarchical fuzzy anatomy models of organs for each body region; (d) recognizing and locating organs in given images by employing the hierarchical models; and (e) delineating the organs following the hierarchy. In Step (c), we explicitly encode object size and positional relationships into the hierarchy and subsequently exploit this information in object recognition in Step (d) and delineation in Step (e). Modality-independent and dependent aspects are carefully separated in model encoding. At the model building stage, a learning process is carried out for rehearsing an optimal threshold-based object recognition method. The recognition process in Step (d) starts from large, well-defined objects and proceeds down the hierarchy in a global to local manner. A fuzzy model-based version of the IRFC algorithm is created by naturally integrating the fuzzy model constraints into the delineation algorithm.

The AAR system is tested on three body regions – thorax (on CT), abdomen (on CT and MRI), and neck (on MRI and CT) – involving a total of over 35 organs and 130 data sets (the total used for model building and testing). The training and testing data sets are divided into equal size in all cases except for the neck. Overall the AAR method achieves a mean accuracy of about 2 voxels in localizing non-sparse blob-like objects and most sparse tubular objects. The delineation accuracy in terms of mean false positive and negative volume fractions is 2% and 8%, respectively, for non-sparse objects, and 5% and 15%, respectively, for sparse objects. The two object groups achieve mean boundary distance relative to ground truth of 0.9 and 1.5 voxels, respectively. Some sparse objects – venous system (in the thorax on CT), inferior vena cava (in the abdomen on CT), and mandible and naso-pharynx (in neck on MRI, but not on CT) – pose challenges at all levels, leading to poor recognition and/or delineation results. The AAR method fares quite favorably when compared with methods from the recent literature for liver, kidneys, and spleen on CT images. We conclude that separation of modality-independent from dependent aspects, organization of objects in a hierarchy, encoding of object relationship information explicitly into the hierarchy, optimal threshold-based recognition learning, and fuzzy model-based IRFC are effective concepts which allowed us to demonstrate the feasibility of a general AAR system that works in different body regions on a variety of organs and on different modalities.

* Corresponding author. Tel.: +1 215 662 6783.
E-mail address: jay@mail.med.upenn.edu (J.K. Udupa).

http://dx.doi.org/10.1016/j.media.2014.04.003
1361-8415/

1. Introduction

1.1. Background

Since the birth of radiology in 1895, the emphasis in clinical radiology has been on *human visualization* of internal structures. Although various tomographic image modalities evolved subsequently for deriving anatomic, functional, and molecular information about internal structures, the emphasis on human visualization continued and the practice of *clinical radiology* has remained mostly *descriptive* and *subjective*. Quantification is amply employed in radiology in clinical research. However, in clinical radiological practice, this is not common. In the qualitative mode, quantifiable and/or subtle image information is underutilized, interpretations remain subjective, and subtle changes at early disease stages or due to therapeutic intervention may be underestimated or missed (Torigian and Alavi, 2007). It is generally believed now that if Quantitative Radiology (QR) can be brought to routine clinical practice, numerous advances can be made including: improved sensitivity, specificity, accuracy, and precision of early disease diagnosis; more objective and *standardized* response assessment of disease to treatment; improved understanding of what is "normal"; increased ease of disease measurement and reporting; and discovery of new disease biomarkers.

To make QR a reality, we believe that computerized *Automatic Anatomy Recognition* (AAR) during radiological image interpretation becomes essential. To facilitate AAR, and hence eventually QR, and focusing only on the anatomic aspects of shape, geography, and architecture of organs, while keeping the larger goal in mind, we present in this paper a novel fuzzy strategy for building body-wide anatomic models, and for utilizing these models for automatically recognizing and delineating body-wide anatomy in given patient images.

1.2. Related work

Image segmentation – the process of recognizing and delineating objects in images – has a rich literature spanning over five decades. From the perspective of the direction in which this field is headed, it is useful to classify the methods developed to date into three groups: (a) Purely image-based, or pI approaches (Beucher, 1992; Boykov et al., 2001; Kass et al., 1987; Malladi et al., 1995; Mumford and Shah, 1989; Udupa and Samarasekera, 1996), wherein segmentation decisions are made based entirely on information derived from the given image; (b) object model-based, or OM approaches (Ashburner and Friston, 2009; Cootes et al., 2001; Heimann and Meinzer, 2009; Pizer et al., 2003; Shattuck et al., 2008; Staib and Duncan, 1992), wherein known object shape and image appearance information over a population are first codified in a model and then utilized on a given image to bring constraints into the segmentation process; and (c) hybrid approaches (Chen and Bagci, 2011; Hansegard et al., 2007; Horsfield et al., 2007; Liu and Udupa, 2009; Rousson and Paragios, 2008; Shen et al., 2011; van der Lijn et al., 2012; Zhou and Bai, 2007), wherein the delineation strengths of the pI methods are combined synergistically with the global object recognition capabilities of the OM strategies. pI algorithms predate other approaches, and they still continue to seek new frontiers. OM approaches go by various names such as statistical models and probabilistic atlases, and continue to be pursued aggressively. Particularly, atlas-based techniques have gained popularity in brain MR image segmentation and analysis (Cabezas et al., 2011). Hybrid approaches hold much promise for AAR and QR and are currently very actively investigated. Since our focus in this paper is the body torso, and since the nature of the images and of the objects and challenges encountered are different for these regions (from, for example, for the brain), our review below will focus mainly on methods developed for the torso.

Since the simultaneous consideration of multiple objects offers better constraints, in recent years, multi-object strategies have been studied under all three groups of approaches to improve segmentation. Under pI approaches, the strategy sets up a competition among objects for delineating their regions/boundaries (e.g.; Bogovic et al., 2013; Saha and Udupa, 2001). In OM approaches, the strategy allows including inter-relationships among objects in the model to influence their localization and delineation (e.g.; Cerrolaza et al., 2012; Duta and Sonka, 1998). In hybrid approaches, multi-object strategies try to strengthen segmentability by incorporating relevant information in model building, object recognition/localization, and subsequently also in delineation via the pI counterpart of the synergistic approach (Chen et al., 2012; Chu et al., 2013; Linguraru et al., 2012; Lu et al., 2012; Meyer et al., 2011; Okada et al., 2008; Shen et al., 2011; Tsechpenakis and Chatzis, 2011). Motivated by applications (such as semantic navigation) where the focus is just locating objects in image volumes and not delineating them, a separate group of methods has been emerging (Criminisi et al., 2013; Zhou and Rajapakse, 2005; Zhou et al., 2013). They use features characterizing the presence of whole organs or specific anatomic aspects of organs (such as the femoral neck and head) combined with machine learning techniques to locate objects in image volumes by finding the size, location, and orientation of rectangular bounding boxes that just enclose the anatomic entities.

The state-of-the-art in image segmentation seems to leave several gaps that hinder the development of a body-wide AAR system. First, while multi-object strategies have clearly shown superior performance for all approaches, in all published works they have been confined to only a few (three to five) selected objects and have not taken into account an entire body region or all of its major organs, the only exception being (Baiker et al., 2010), whose focus was whole body segmentation of mice on micro CT images. Second, and as a result, there is no demonstrated single method that operates on different body regions, on all major organs in each body region, and at different modalities. Third, all reported modeling strategies have a statistical framework, either as statistical models of shape and intensity pattern of appearance of objects in the image or as atlases, and none taking a fuzzy approach, except (Zhou and Rajapakse, 2005) and our previous work (Miranda et al., 2008, 2009), both in the brain only. Fuzzy set concepts have been used extensively otherwise in image processing and 3D visualization. Fuzzy modeling approaches allow bringing anatomic information in an all-digital form into graph theoretic frameworks designed for object recognition and delineation, obviating the need for (continuous) assumptions made otherwise in statistical approaches about shapes, random variables, their independence, functional form of density distributions, etc. They also allow capturing information about uncertainties at the patient level (e.g., blur, partial volume effects) and population level, and codification of this information within the model. Fourth, objects have complex inter-relationships in terms of their geographic layout. Learning this information over a population and encoding this explicitly in an object hierarchy can facilitate object localization considerably. Although several multi-object methods have accounted for this relationship indirectly, its direct incorporation into modeling, object recognition, and delineation in an anatomic hierarchical order has not been attempted. The AAR approach presented in this paper is designed to help overcome these gaps.

1.3. Outline of paper and approach

We start off by describing a novel hierarchical fuzzy modeling framework for codifying prior population information about object

assemblies in Section 2. In Section 3, we delineate methods for automatically recognizing objects in given patient images that employ these hierarchical models. We present fuzzy-connectedness-based object delineation techniques in Section 4 that employ the modified fuzzy models found at recognition as constraints in delineation. We demonstrate and evaluate the applicability of the AAR methodology in Section 5 on three different body regions – thorax, abdomen, and neck – on different modalities. A comparison to methods from recent literature, the lessons learned, our conclusions, and the challenges we encountered are examined in Section 6. The AAR approach has five unique characteristics: (1) direct hierarchical codification of the prior object geographic and geometric relationship information; (2) a "what-you-see-is-what-you-get" entirely digital fuzzy modeling strategy; (3) hierarchical object recognition strategies that go from a broader gestalt to narrower specifics in locating objects; (4) demonstrated generality of applicability of the same approach to different organ systems, body regions, and modalities; and (5) adaptability of the system to different applications.

The AAR approach is graphically summarized in Fig. 1. The body is divided into body regions $B_1,\ldots,B_K$. Models are built for each specific body region $\mathcal{B} \in \{B_1,\ldots,B_K\}$ and each population group G (whatever way G is defined). Throughout this paper, $\mathcal{B}$ and G are treated as variables, and each body region is considered separately and independent of other body regions. In Section 6, we will discuss briefly the issue of linking body regions for considering the whole body for the AAR schema. The three main blocks in Fig. 1 correspond to model building, object recognition, and object delineation. A fuzzy model $FM(O_\ell)$ is built separately for each of the L objects O_ℓ in $\mathcal{B}$, and these models are integrated into a hierarchy chosen for $\mathcal{B}$. The output of the first step is a fuzzy anatomic model $FAM(\mathcal{B}, G)$ of the body region $\mathcal{B}$ for group G. This model is utilized in recognizing objects in a given patient image I of $\mathcal{B}$ belonging to G in the second step. The hierarchical order is followed in this process. The output of this step is the set of transformed fuzzy models FM^T (O_ℓ) corresponding to the state when the objects are recognized in I. These modified models and the image I form the input to the third step of object delineation which also follows the hierarchical order. The final output is in the form of delineated objects $O_1^D,\ldots,O_L^D$, where each O_ℓ^D is a binary image.

Very preliminary versions of some of the contents of this paper appeared in SPIE Medical Imaging conference proceedings in 2011, 2012, and 2013. Those papers did not contain the full details presented here on model building. More importantly, based on earlier experience many improvements are reported in this paper, none of which appeared earlier. Further, the recognition and delineation methods presented here have many novel elements. As a result, the entire AAR approach has changed substantially. Additional differences include comprehensive evaluation and the demonstration of the AAR scheme on multiple body regions.

2. Building fuzzy model of body region $\mathcal{B}$

Notation: We will use the following notation throughout this paper. G: the population group under consideration. $\mathcal{B}$: the body region of focus. $O_1,\ldots,O_L$: L objects or organs of $\mathcal{B}$ (such as esophagus and pericardium for $\mathcal{B}$ = Thorax). $\mathcal{I} = \{I_1,\ldots,I_N\}$: the set of images of $\mathcal{B}$ for G from N subjects which are used for model building and for training the parameters of the AAR algorithms. $I_{n,\ell}$: the binary image representing the true delineation of object O_ℓ in the image $I_n \in \mathcal{I}$. $\mathcal{I}^b = \{I_{n,\ell}: 1 \leqslant n \leqslant N \,\&\, 1 \leqslant \ell \leqslant L\}$ is the set of all binary images used for model building. $FM(O_\ell)$: Fuzzy model of object O_ℓ derived from the set of all binary images $\mathcal{I}_\ell^b = \{I_{n,\ell}: 1 \leqslant n \leqslant N\}$ of O_ℓ. $FAM(\mathcal{B}, G)$: Fuzzy anatomy model of the whole object assembly in $\mathcal{B}$ with its hierarchy. $FM^T(O_\ell)$: Transformed (adjusted) $FM(O_\ell)$ corresponding to the state when O_ℓ is recognized in a given patient image I. O_ℓ^D: Delineation of O_ℓ in I represented as a binary image. Any image I will be represented by a pair $I = (C, f)$, where C denotes a 3D rectangular array of voxels, and f is a mapping f: $C \rightarrow \mathbb{I}$ where $\mathbb{I}$

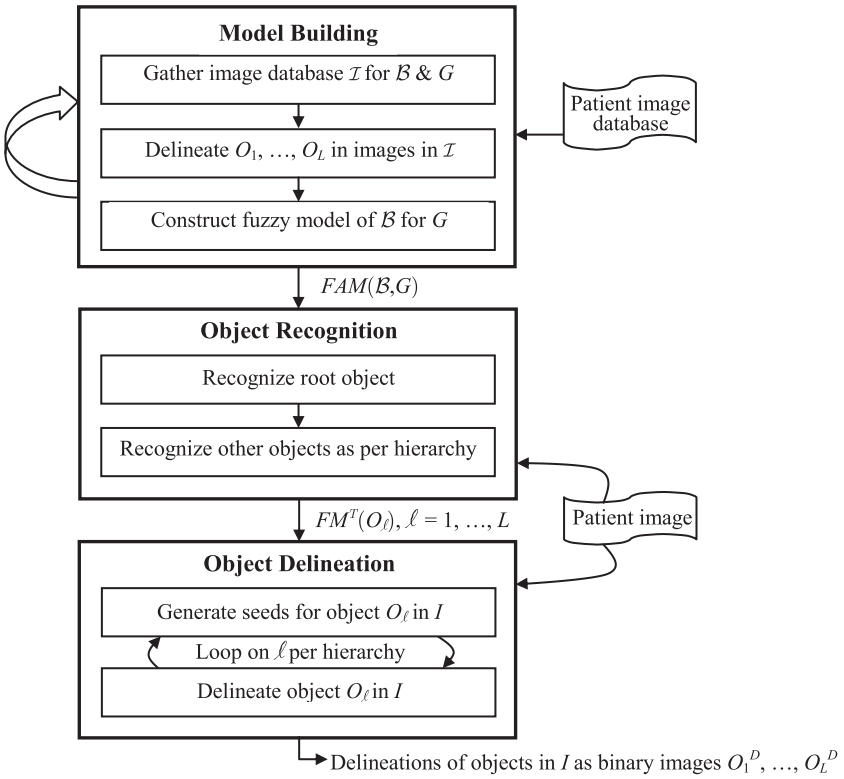

Fig. 1. A schematic representation of the AAR schema. The three main steps of model building, object recognition, and object delineation are explained in Sections 2–4.

is a set of integers[1] denoting the image intensities. For any binary image $J = (C, f_b)$, we will use $PAS(J)$ to denote the principal axes system derived from the set X of voxels of J with value 1. $PAS(J)$ is described by the geometric center of X and the eigenvectors derived from X via principal component analysis.

Our description in the rest of Section 2 will follow the schematic of Fig. 1. Table A1 in Appendix A lists brief anatomic definitions of all objects from all three body regions considered in this paper.

2.1. Gathering image database for $\mathcal{B}$ and G

This retrospective study was conducted following approval from the Institutional Review Board at the Hospital of the University of Pennsylvania along with a Health Insurance Portability and Accountability Act (HIPAA). The basic premise of our AAR approach is that the fuzzy anatomic model of $\mathcal{B}$ for G should reflect *near normal* anatomy. Consequently, the cleanest way of gathering image data for model building will be to prospectively acquire image data in a well-defined manner from subjects in group G who are certified to be near normal. Such an approach would be expensive and may involve radiation exposure (in case of CT imaging). For developing the concepts and testing the feasibility of AAR, therefore, we have taken a vastly less expensive and simpler approach of utilizing existing human subject image data sets. For the thoracic and abdominal body regions, a board certified radiologist (co-author DAT) selected all image data sets (CT) from our health system patient image database in such a manner that the images appeared radiologically normal for the body region considered, with exception of minimal incidental focal abnormalities such as cysts and small pulmonary nodules. Images with severe motion/streak artifacts or other limitations were excluded from consideration. For these two body regions, the population groups considered have an age range of approximately 50–60 years. This age range was selected to maximize our chances of finding sufficient number of near normal images. For the neck body region, we have utilized image data (MRI) previously acquired from normal subjects for the study of pediatric upper airway disorders. G in this instance is female subjects in the age range of 7–18. Our modeling schema is such that the population variables can be defined at any desired "resolution" in the future and the model can then be updated when more data are added.

Some organs in $\mathcal{B}$ are better defined in a slice plane different from the slice plane used for imaging others. For example, for $\mathcal{B}$ = neck, the best plane for slice imaging is sagittal for tongue and soft palate, while for the upper airways and other surrounding organs, axial slices are preferred. Our AAR methodology automatically handles organs defined in images with different orientations of digitization by representing image and object data in a fixed and common scanner coordinate system of reference.

2.2. Delineating objects of $\mathcal{B}$ in the images in the database

There are two aspects to this task – forming an operational definition of both $\mathcal{B}$ and the organs in $\mathcal{B}$ in terms of their precise anatomic extent, and then delineating the objects following the definition. These considerations are important for building consistent and reliable models, and, in the future, if similar efforts and results for body-wide models are to be combined, exchanged, and standardized.

2.2.1. Definition of body regions and objects

Each body region is defined consistently in terms of a starting and ending anatomic location. For axial slice data, these locations are determined in terms of transverse slice positions. For example, for $\mathcal{B}$ = Thorax, the body region is considered to extend axially from 5 mm below the base of the lungs to 15 mm above the apex of the lungs. Arms are not included in this study. For other orientations of slice planes in slice imaging, the same definitions are applied but translated into other planes. Similarly, each object included in $\mathcal{B}$ is defined precisely irrespective of whether it is open-ended because it straddles body regions (for example, esophagus) or closed and contained within $\mathcal{B}$ but is contiguous with other objects (for example, liver with hepatic portal vein, common hepatic artery, and bile duct). For each body region, we have created a document that delineates its precise definition and the specification of the components and boundaries of its objects. This document is used as a reference by all involved in generating data sets for model building. These definitions are summarized in the table included in Appendix A.

Each body region is carved out manually, following its definition, from the data sets gathered for it. In our notation, $\mathcal{I}$ denotes the resulting set of such standard images that precisely cover $\mathcal{B}$ as per definition. We assume the scanner coordinate system, SCS, as a common reference system with respect to which all coordinates will be expressed.

2.2.2. Delineation of objects

The objects of $\mathcal{B}$ are delineated in the images of $\mathcal{I}$, adhering to their definition, by a combination of methods including live wire, iterative live wire (Souza and Udupa, 2006), thresholding, and manual painting, tracing and correction. To minimize human labor and to maximize precision and accuracy, algorithms in terms of a proper combination of these methods and the order in which objects are delineated are devised first, all of which operate under human supervision and interaction. For illustration, in the abdomen, to delineate subcutaneous adipose tissues (SAT) as an object, the skin outer boundary ASkn (as an object) is first segmented by using the iterative live wire method. Iterative live wire is a version of live wire in which once the object is segmented in one slice, the user commands next slice, the live wire then operates automatically in the next slice, and the process is continued until automatic tracing fails when the user resorts to interactive live wire again, and so on. Subsequently, the interface between the subcutaneous and visceral adipose compartments is delineated by using also the iterative live wire method. Once these two object boundaries are delineated, the subcutaneous and visceral components are delineated automatically by using thresholding and morphological operations. On MR images, the same approach works if background non-uniformity correction and intensity standardization (Nyul and Udupa, 1999) are applied first to the images in $\mathcal{I}$. If direct delineation by manual tracing or even by using live wire is employed, the process would become complicated (because of the complex shape of the adipose and visceral compartments) and much more labor intensive.

Because of the enormity of this task, a number of trainees, some with medical and biomedical but some with engineering background, were involved in accomplishing this task. All tracings were examined for accuracy by several checks – 3D surface renditions of objects from each subject in various object combinations as well as a slice-by-slice verification of the delineations overlaid on the gray images for all images. The set of binary images generated in this step for all objects is denoted by $\mathcal{I}^b = \{I_{n,\ell}: 1 \leqslant n \leqslant N \;\&\; 1 \leqslant \ell \leqslant L\}$. The set of binary images generated just for object O_ℓ is denoted by $\mathcal{I}^b_\ell = \{I_{n,\ell}: 1 \leqslant n \leqslant N\}$.

2.3. Constructing fuzzy object models

The *Fuzzy Anatomy Model* $FAM(\mathcal{B}, G)$ of any body region $\mathcal{B}$ for group G is defined to be a quintuple:

[1] Except when we deal with fuzzy sets, which are also expressed as images for computational purposes, in which case $\mathbb{I}$ is a set of real numbers.

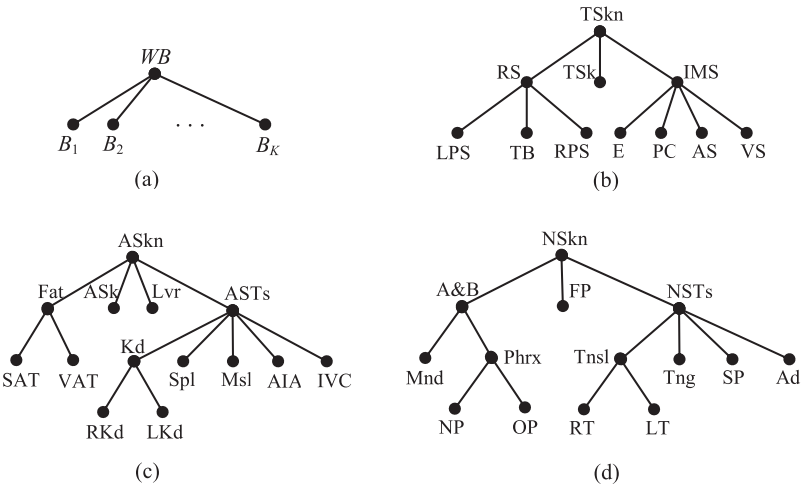

Fig. 2. (a) ***Hierarchy for whole body WB***. (b) ***Hierarchy for Thorax***. TSkn: Outer boundary of thoracic skin as an object; RS: Respiratory System; TSk: Thoracic Skeleton; IMS: Internal Mediastinum; RPS, LPS: Right & Left Pleural Spaces; TB: Trachea & Bronchi; E: Esophagus; PC: Pericardium; AS, VS: Arterial & Venous Systems. (c) ***Hierarchy for Abdomen***. ASkn: Outer boundary of abdominal skin; ASk: Abdominal Skeleton; Lvr: Liver; ASTs: Abdominal Soft Tissues; SAT & VAT: Subcutaneous and Visceral Adipose Tissues; Kd: Kidneys; Spl: Spleen; Msl: Muscle; AIA: Aorta and Iliac arteries; IVC: Inferior Vena Cava; RKd & LKd: Right and Left Kidneys. (d) ***Hierarchy for Neck***. NSkn: Outer boundary of skin in neck; A&B: Air & Bone; FP: Fat Pad; NSTs: Soft Tissues in neck; Mnd: Mandible; Phrx: Pharynx; Tnsl: Tonsils; Tng: Tongue; SP: Soft Palate; Ad: Adenoid; NP & OP: Nasopharynx and Oropharynx; RT & LT: Right and Left Tonsils.

$$FAM(\mathcal{B}, G) = (H, \mathcal{M}, \rho, \lambda, \eta). \quad (1)$$

Briefly, the meaning of the five elements of $FAM(\mathcal{B}, G)$ is as follows. H is a hierarchy, represented as a tree, of the objects in $\mathcal{B}$; see Fig. 2. $\mathcal{M}$ is a collection of fuzzy models, one model per object in $\mathcal{B}$. ρ describes the parent-to-offspring relationship in H over G. λ is a set of scale factor ranges indicating the size variation of each object O_ℓ over G. η represents a set of measurements pertaining to the objects in $\mathcal{B}$. A detailed description of these elements and the manner in which $FAM(\mathcal{B}, G)$ is derived from $\mathcal{I}$ and $\mathcal{I}^b$ are presented below.

2.3.1. Hierarchy H

This element describes the way the objects of $\mathcal{B}$ are considered ordered anatomically as a tree structure. This order currently specifies the inclusion of an offspring object O_k anatomically in the parent object O_ℓ.[2] While each $\mathcal{B}$ has its own hierarchy, $\mathcal{B}$ itself forms the offspring of a root denoting the whole body, *WB*, as shown in Fig. 2. The hierarchies devised for the three body regions studied are shown in Fig. 2. An object that is exactly a union of its offspring will be referred to as a *composite object*. Examples: RS, Fat, Kd, etc. Note that none of the skin objects is a composite object since the full body region inside the skin is not fully accounted for by the union of the offspring objects. The notion of composite objects is useful in combining objects of similar characteristics at a higher level of the hierarchy, which may make object recognition (and delineation) more effective. Thin tubular objects will be called *sparse objects*: TB, E, AS, VS, AIA, IVC, Phrx, NP, and OP. Compact, blob-like objects will be referred to as *non-sparse*: TSkn, RS, IMS, LPS, RPS, PC, ASkn, Fat, SAT, VAT, Lvr, Spl, Kd, RKd, LKd, NSkn, FP, NSTs, Tnsl, Tng, SP, Ad, RT, and LT. Some objects are a hybrid between these two types, consisting of both features. Examples: TSk, Ask, ASTs, A&B, and Mnd.

2.3.2. Fuzzy model set $\mathcal{M}$

The second element $\mathcal{M}$ in the description of $FAM(\mathcal{B}, G)$ represents a set of fuzzy models, $\mathcal{M} = \{FM(O_\ell): 1 \leqslant \ell \leqslant L\}$, where $FM(O_\ell)$ is expressed as a fuzzy subset of a reference set $\Omega_\ell \subset Z^3$ defined in the SCS; that is, $FM(O_\ell) = (\Omega_\ell, \mu_\ell)$. The membership function $\mu_\ell(v)$ defines the degree of membership of voxel $v \in \Omega_\ell$ in the model of object O_ℓ. Ideally, for any ℓ, $1 \leqslant \ell \leqslant L$, we would like the different samples of O_ℓ in different subjects to differ by a transformation $A_{n,\ell}$ involving translation, rotation, and isotropic scaling. Our idea behind the concept of the fuzzy model of an object is to codify the spatial variations in form from this ideal that may exist among the N samples of the object as a spatial fuzzy set, while also retaining the spatial relationship among objects in the hierarchical order.

Given the training set of binary images $\mathcal{I}^b_\ell$ of object O_ℓ, we determine $A_{n,\ell}$, μ_ℓ, and $FM(O_\ell)$ for O_ℓ as follows. We permit only such alignment operations, mimicking $A_{n,\ell}$, among the members of $\mathcal{I}^b_\ell$, that are executed precisely without involving search and that avoid the uncertainties of local optima associated with optimization-based full-fledged registration schemas. In this spirit, we handle the translation, rotation, and scaling components of $A_{n,\ell}$ in the following manner.

For translation and rotation, for each manifestation $I_{n,\ell}$ of O_ℓ in $\mathcal{I}^b_\ell$, we determine, within *SCS*, the principal axes system $PAS(I_{n,\ell})$ of O_ℓ. Subsequently, all samples are aligned to the mean center and principal axes[3]. The scale factor estimation is based on a linear size estimate (in *mms*) of each sample of O_ℓ and resizing all samples to the mean size. The size of O_ℓ in $I_{n,\ell}$ is determined from $\sqrt{(e_1 + e_2 + e_3)}$, where e_1, e_2, and e_3 are the eigenvalues corresponding to the principal components of O_ℓ in $I_{n,\ell}$.[4]

After aligning the members of $\mathcal{I}^b_\ell$ via $A_{n,\ell}$, a distance transform is applied to each transformed member for performing shape-based interpolation (Raya and Udupa, 1990; Maurer et al., 2003), the distances are averaged over all members, and converted through a sigmoid function to obtain the membership values μ_ℓ and subsequently $FM(O_\ell)$.

2.3.3. Parent-to-offspring relationship ρ

This element describes the parent-to-offspring spatial relationship in H for all objects in $\mathcal{B}$. Since each object O_k has a unique parent, this relationship is represented by $\rho = \{\rho_k: 1 \leqslant k \leqslant L\}$[5]. For each

[2] However, as discussed in Section 6, other arrangements are possible for H.

[3] In our empirical investigations of the AAR system, we have studied the construction and use of fuzzy models both with and without orientation alignment. See Section 5.

[4] Among several size measures we tested, such as $\sqrt[3]{\text{volume}}$, largest eigenvalue, and the length of the diagonal of the enclosing box, this measure turned out to be the most robust.

[5] It also encodes *WB* to body region relationships, although this is not taken into account in our current implementation. See comments in Section 6.

O_k, ρ_k codifies the mean position as well as the orientation relationship between O_k and its parent over N samples. We adopt the convention that ρ_1 denotes the relationship of the root object of $\mathcal{B}$ relative to SCS. Let $GC_{n,\ell}$ be the geometric center of O_ℓ in $I_{n,\ell}$. Then, the mean positional relationship $P_{\ell,k}$ between O_ℓ and O_k is considered to be the mean of the vectors in the set $\{GC_{n,k} - GC_{n,\ell}: 1 \leqslant n \leqslant N\}$. To find the mean orientation $Q_{\ell,k}$, we make use of the eigenvectors $E^1_{n,\ell}, E^2_{n,\ell}$, and $E^3_{n,\ell}$ of the shape of O_ℓ in $I_{n,\ell}$ estimated over all N samples. We take an average of each $E^i_{n,\ell}$ over N samples for $i = 1, 2, 3$. However, for some n and i, $E^i_{n,\ell}$ may be more than 90 degrees from the average, in which case we replace $E^i_{n,\ell}$ by $-E^i_{n,\ell}$ while simultaneously replacing $E^j_{n,\ell}$ by $-E^j_{n,\ell}$ for some j different from i so as to keep the system right-handed. We then recalculate the average, and repeat until the eigenvector is within 90 degrees of the average. Then, starting from either the first or the third eigenvector, whichever has the eigenvalue farther from the second, we normalize and make the others orthogonal to it. $Q_{\ell,k}$ is then taken to be the transformation that aligns the eigenvector system of the parent O_ℓ with that mean orientation. This method guarantees a robust orientation estimate despite the 180-degrees switching property of eigenvectors.

In order not to corrupt ρ_k by the differences in size among subjects, before estimating ρ_k, the parent O_ℓ and all offspring objects O_k of O_ℓ are scaled with respect to the center $GC_{n,\ell}$ of O_ℓ as per a common scale factor, estimated for O_ℓ via the method described above. The reasoning behind this scaling strategy is that an object and its entire offspring should be scaled similarly to retain their positional relationship information correctly.

2.3.4. Scale range λ

The fourth element λ of $FAM(\mathcal{B}, G)$ is a set of scale factor ranges, $\lambda = \{\lambda_\ell = [\lambda^b_\ell, \lambda^h_\ell] : 1 \leq \ell \leq L\}$, indicating the size variation of each object O_ℓ over its family $\mathcal{I}^b_\ell$. This information is used in recognizing O_ℓ in a given image to limit the search space for its pose; see Section 3.

2.3.5. Measurements η

This element represents a set of measurements pertaining to the object assembly in $\mathcal{B}$. Its purpose is to provide a database of normative measurements for future use. We are not exploring this aspect in this paper. However, this element also serves to improve our knowledge about object relationships (in form, geographical layout, etc. in $\mathcal{B}$) and thence in constructing better hierarchies for improving AAR. We will discuss this briefly in Section 5.

There are several parameters related to object recognition (Section 3) and delineation (Section 4), some of which are image modality specific. (They are identified by $\mathcal{T}^m_1$ and Th_ℓ in Section 3 and $\sigma_{\psi O}$, $m_{\varphi O}$, $m_{\varphi B}$, $\sigma_{\varphi O}$, and $\sigma_{\varphi B}$ in Section 4.) The values of these parameters are also considered part of the description of η. The definition of these parameters and the process of their estimation are described at relevant places in Sections 3 and 4 for ease of reading, although their actual estimation is done at the model building stage.

The fuzzy anatomy model $FAM(\mathcal{B}, G)$ output by the model building process is used in performing AAR on any image I of $\mathcal{B}$ for group G as described in Sections 3 and 4.

3. Recognizing objects

We think of the process of what is usually referred to as "segmenting an object in an image" as consisting of two related phenomena – object recognition (or localization) and object delineation. *Recognition* is a high-level process of determining the whereabouts of the object in the image. Given this information for the object, its *delineation* is the meticulous low-level act of precisely indicating the space occupied by the object in the image. The design of the entire AAR methodology is influenced by this conceptual division. We believe that without achieving acceptably accurate recognition it is impossible to obtain good delineation accuracy. The hierarchical concept of organizing the objects for AAR evolved from an understanding of the difficulty involved in automatic object recognition. Once good recognition accuracy is achieved, several avenues for locally confined accurate delineation become available, as we discuss in Section 4. The goal of recognition in AAR is to output the pose (translation, rotation, and scaling) of $FM(O_\ell)$, or equivalently the pose-adjusted fuzzy model $FM^T(O_\ell)$, for each O_ℓ in a given test image I of $\mathcal{B}$ such that $FM^T(O_\ell)$ matches the information about O_ℓ present in I optimally.

The recognition process proceeds hierarchically as outlined in the procedure *AAR-R* presented below. In Step R_1, the root object is recognized first by calling algorithm *R-ROOT*[6]. Then, proceeding down the tree represented by H in the breadth-first order, other objects are recognized by calling algorithm *R-OBJECT*. The latter makes essential use of the parent fuzzy model and the parent-to-offspring relationship ρ encoded in $FAM(\mathcal{B}, G)$.

Procedure *AAR-R*

Input: An image I of $\mathcal{B}$, $FAM(\mathcal{B}, G)$.
Output: $FM^T(O_\ell)$, $\ell = 1, \ldots, L$.
Begin
R1. Call *R-ROOT* to recognize the root object in H;
R2. *Repeat*
R3. Find the next offspring O_k to *recognize* in H (see text);
R4. Knowing $FM^T(O_\ell)$, ρ_k, and λ_k, call *R-OBJECT* to recognize O_k;
R5. *Until* all objects are covered in H;
R6. Output $FM^T(O_\ell)$, $\ell = 1, \ldots, L$;
End

Two strategies are described here for each of algorithms *R-ROOT* and *R-OBJECT*. The first, a global approach, does not involve searching for the best pose. We call this the *One-Shot Method* since the model pose is determined directly by combining the prior information stored in $FAM(\mathcal{B}, G)$ and information quickly gathered from the given image I. The one-shot method is used as initialization for a more refined second method called *Thresholded Optimal Search*.

3.1. One-shot method

A threshold interval Th_1 corresponding to the root object O_1 is applied to I followed by a morphological opening operation to roughly segment O_1 to produce a binary image J. The purpose of the morphological operation is to exclude as much as possible any significant extraneous material, such as the scanner table and patient clothing, from J. Then the transformed model $FM^T(O_1)$ is found by applying a transformation $\mathcal{T}^m_1$ to $FM(O_1)$. $\mathcal{T}^m_1$ is devised to express the mean relationship between the roughly segmented O_1 and the true segmentation of O_1 represented in the binary images $I_{n,1} \in \mathcal{I}^b$. The estimation of $\mathcal{T}^m_1$ is done at the model building stage of AAR as mentioned in Section 2.3. To determine $\mathcal{T}^m_1$, similar thresholding and morphological operations are performed on each gray image I_n in the training set to obtain a rough segmentation of O_1, denoted $J_{n,1}$, in I_n. The relationship between this rough segmentation $J_{n,1}$ and the true segmentation $I_{n,1}$ of O_1 in $\mathcal{I}^b$ is found as a transformation $\mathcal{T}_{n,1}$ that maps $PAS(J_{n,1})$ to $PAS(I_{n,1})$. The mean,

[6] We assume that the field of view in I fully encloses the root object. For the hierarchies shown in Fig. 2, the root object is the skin outer boundary which is typically more-or-less, although not perfectly, fully included within the imaging field of view. See also Section 6 for further comments.

denoted $\mathcal{T}_1^m$, of such transformations over all training images is then found.

Once the root object O_1 is recognized, the poses for other objects in I in the hierarchy H are determined by combining (in the sense of composition) $\mathcal{T}_1^m$ with the parent to offspring relationship information stored in ρ_k for each parent-offspring pair. The transformed models $FM^T(O_\ell)$ are then found from this information.

3.2. Thresholded optimal search

This is a strategy to refine the results obtained from the one-shot method. Its premise is that the overall image intensity of the objects in $\mathcal{B}$ can be characterized by threshold intervals[7] such that, at the model's pose corresponding to the best match of the model with an underlying object in the given test image I, the mismatch between the thresholded result and the model is minimal. For MR images for this approach to make sense, it is essential to correct for background intensity non-uniformities first followed by intensity standardization (Nyul and Udupa, 1999).

Suppose that at the model building stage, the optimal threshold interval Th_ℓ for each object O_ℓ has already been determined automatically from the training image set. We will explain below how this is accomplished. Then, at the recognition stage, the threshold for O_ℓ is fixed at this learned value Th_ℓ. Starting from the initial pose found by the one-shot method, a search is made within the pose space for an optimal pose $\boldsymbol{p}^*$ of the fuzzy model over I that yields the smallest sum of the volume of false positive and false negative regions, where the model itself is taken as the reference for defining false positive and negative regions. Specifically, let $FM^{\boldsymbol{p}}(O_\ell)$ denote the fuzzy model of O_ℓ at any pose $\boldsymbol{p}$, expressed as an image, and let J denote the binary image resulting from thresholding I at Th_ℓ. Then[8],

$$\boldsymbol{p}^* \in \arg\min_{\boldsymbol{p}}(|FM^{\boldsymbol{p}}(O_\ell) - J| + |J - FM^{\boldsymbol{p}}(O_\ell)|). \quad (2)$$

Image subtraction here is done in the sense of fuzzy logic, and $|x|$ denotes the fuzzy cardinality of x, meaning that it represents the sum total of the membership values in x. The search space to find $\boldsymbol{p}^*$ is limited to a region around the initial pose. This region is determined from knowledge of ρ_k and its variation and the scale factor range λ_k. For the positional vector, we search in an ellipsoid with its axes in the coordinate axis directions and with length four times the standard deviation of the corresponding coordinate. When searching in orientation space, we search in an ellipsoid with its axes in the direction of the eigenvectors of the rotation vector distribution (covariance matrix) and with length four times the square root of the corresponding eigenvalue. (A rotation vector has magnitude equal to the angle of rotation and direction along the axis of right-handed rotation. The rotation referred to is the rotation of $Q_{\ell,k}$ required to bring it into coincidence with $E^i_{n,\ell}$.) For the scale factor, we search in an interval of size four times the standard deviation of the scale factor.

3.2.1. Determining Th_ℓ at the model building stage

To estimate Th_ℓ, we run a rehearsal of the recognition method described above as follows, essentially for attempting to learn the recognition process. Imagine we already built $\mathcal{M}$ and estimated ρ and λ. Suppose that we now run the recognition process on the training images. Since we do not know the optimal threshold but have the true segmentations, the idea behind this learning of the recognition process is to test recognition efficacy for each of a number of threshold intervals t and then select the interval Th_ℓ that yields the best match of the model with the known true segmentations for each O_ℓ. That is, if $J_n(t)$ is the binary image resulting from thresholding the training image I_n at t, then

$$Th_\ell \in \arg\min_{p,\,t} \sum_n |(J_n(t) \times FM^{\boldsymbol{p}}(O_\ell)) - I_{n,\ell}| + |I_{n,\ell} - (J_n(t) \times FM^{\boldsymbol{p}}(O_\ell))|. \quad (3)$$

Here, $\times$ denotes fuzzy intersection. In words, the optimal threshold Th_ℓ is found by searching over the pose space over all training data sets and all thresholds the best match between the true segmentation of O_ℓ with the result of thresholding I_n restricted to the model. In our implementation, 81 different values of the intervals are searched (9 for each end of the interval). The 9 positions for the lower end are the 5th, 10th, ..., 45th percentile values of the cumulative object intensity histogram determined from the training image set. Similarly, for the upper end, the positions are 55–95th percentile values.

To summarize, the thresholded optimal search method starts the search process from the initial pose found by the one-shot method. It uses the optimal threshold values Th_ℓ determined at the training stage for each object O_ℓ and finds the best pose for the fuzzy model of O_ℓ in the given image I by optimally matching the model with the thresholded version of I. The only parameters involved in the entire recognition process are the thresholds Th_ℓ, one threshold interval per object, and $\mathcal{T}_1^m$. Their values are automatically determined in the model building stage from image and binary image sets $\mathcal{I}$ and $\mathcal{I}^b$ and they become part of the model $FAM(\mathcal{B}, G)$ itself.

4. Delineating objects

Once the recognition process is completed and the adjusted models $FM^T(O_\ell)$ are output for a given image I of $\mathcal{B}$, delineation of objects is performed on I in the hierarchical order as outlined in the procedure *AAR-D* presented below. As in recognition, in Step *D*1, the root object is first delineated by calling *D-ROOT*. *AAR-D* then proceeds in the breadth-first order to delineate other objects by calling *D-OBJECT*.

Procedure *AAR-D*

Input: An image I of $\mathcal{B}$, $FAM(\mathcal{B}, G)$, $FM^T(O_\ell)$, $\ell = 1, \ldots, L$.
Output: O_ℓ^D, $\ell = 1, \ldots, L$.
Begin
*D*1. Call *D-ROOT* to delineate the root object in H;
D2. *Repeat*
*D*3. Traverse H and find the next offspring O_k to delineate in H (see text);
*D*4. Knowing delineation of O_ℓ, call *D-OBJECT* to delineate O_k in I;
*D*5. *Until* all objects are covered in H;
*D*6. Output O_ℓ^D, $\ell = 1, \ldots, L$;
End

For *D-ROOT* and *D-OBJECT*, we have chosen an algorithm from the fuzzy connectedness (FC) family in view of the natural and intimate adaptability of the FC methods to prior information coming in the form of fuzzy sets. In particular, since we focus on the problem of delineating one object at a time, for both Steps *D*1 and *D*4, we have selected the linear-time Iterative Relative FC (IRFC) algorithm of (Ciesielski et al., 2012) for separating each object O_ℓ from its background. Our novel adaptations are in incorporating fuzzy model information into the IRFC formulation and in making the latter fully automatic. These modifications are described below.

[7] All thresholds are assumed to represent intervals in this paper unless specified otherwise.

[8] Since *arg min* is a set, "$\in$" means one of the values chosen from the set is assigned to $\boldsymbol{p}^*$.

4.1. Fuzzy model-based IRFC (FMIRFC)

There are two aspects that need to be addressed to fully describe the *FMIRFC* algorithm: *affinity function* and *seed specification*. Affinity is a local concept indicating the degree of connectedness of voxels locally in terms of their spatial and intensity nearness. In the FC family, this local property is grown optimally into a global phenomenon of object connectedness through the notion of path strengths.

4.1.1. Affinity function

The FC framework (Udupa and Samarasekera, 1996; Ciesielski et al., 2012) is graph-based. An ordered graph (C, α) is associated with the given image $I = (C, f)$ where α is an adjacency relation on C such as 6-, 18-, or 26-adjacency. Each ordered pair (c, d) of adjacent voxels in α is assigned an affinity value $\kappa(c, d)$ which constitutes the weight assigned to arc (c, d) in the graph. To each path π in the graph (or equivalently in I) in the set of all possible paths $\Pi_{a,b}$ between two voxels a and b of C, a strength of connectedness $K(\pi)$ is determined, which is the minimum of the affinities along the path. The *connectivity measure* $K^*(a, b)$ between a and b is then defined to be $K^*(a, b) = \max\{K(\pi): \pi \in \Pi_{a,b}\}$. The notion of connectivity measure can be generalized to the case of "between a set A and a voxel b" by a slight modification: $K^*(A, b) = \max\{K(\pi): \pi \in \Pi_{a,b} \ \& \ a \in A\}$. By using a fast algorithm to compute $K^*(A, b)$, the machinery of FC allows a variety of approaches to define and compute "objects" in images by specifying appropriate affinity functions and seed sets. In particular, in IRFC, two seed sets A_O and A_B are indicated for an object O and its background B, respectively. Then the object indicated by A_O is separated optimally from the background indicated by A_B by an iterative competition in connectivity measure between A_O and every voxel $c \in C$ and between A_B and c. In published IRFC methods, A_O and A_B are specified usually with human interaction.

In *FMIRFC*, affinities $\kappa_O(c, d)$ and $\kappa_B(c, d)$ for O and B are designed separately. Subsequently they are combined into a single affinity κ by taking a fuzzy union of κ_O and κ_B. Each of κ_O and κ_B has three components. The description below is for κ_O. The same applies to κ_B.

$$\kappa_O(c,d) = \omega_1\psi_O(c,d) + \omega_2\varphi_O(c,d) + \omega_3\gamma_O(c,d). \tag{4}$$

Here, $\psi_O(c, d)$ represents a *homogeneity component* of affinity, meaning, the more similar image intensities $f(c)$ and $f(d)$ are at c and d, the greater is this component of affinity between c and d. As commonly done in the FC literature, we set

$$\psi_O(c,d) = exp[-(f(c) - f(d))^2/2\sigma_{\psi_O}^2], \tag{5}$$

where $\sigma_{\psi O}$ is a homogeneity parameter that indicates the standard deviation of intensities within object O. $\varphi_O(c, d)$, the *object feature component*, on the other hand, describes the "degree of nearness" of the intensities at c and d to the intensity $m_{\varphi O}$ expected for the object O under consideration. Denoting the standard deviation of object intensity by $\sigma_{\varphi O}$, this nearness is expressed by

$$\varphi_O(c,d) = exp[-(\max\{(f(c) - m_{\varphi_O})^2, (f(d) - m_{\varphi_O})^2\}/2\sigma_{\varphi_O}^2]. \tag{6}$$

The third component γ_O incorporates *fuzzy model information* into affinity by directly taking the larger of the two fuzzy model membership values $\mu_O(c)$ and $\mu_O(d)$ at c and d for the object,

$$\gamma_O(c,d) = \max\{\mu_O(c), \mu_O(d)\}. \tag{7}$$

Finally, a combined single affinity κ on I is constructed by

$$\kappa(c,d) = \max\{\kappa_O(c,d), \kappa_B(c,d)\}. \tag{8}$$

The weights in (4) are chosen equal and such that they add up to 1. The homogeneity parameter is set equal for object and background ($\sigma_{\psi O} = \sigma_{\psi B}$) and estimated from uniform regions in the training images (after leaving out high gradient regions), as commonly done in the FC literature (Saha and Udupa, 2001). The remaining parameters ($\sigma_{\varphi O}$, $\sigma_{\varphi B}$, $m_{\varphi O}$, $m_{\varphi B}$) are estimated automatically from the training data sets from the knowledge of O and B regions for each object.

4.1.2. Seed specification

Seed sets A_O and A_B are found by a joint criterion of a threshold for image intensity and for model membership for each of O and B. The threshold interval Th_O for O is the same as the one used for recognition, namely Th_ℓ. The threshold interval Th_B for background is a union of similar threshold intervals for the background objects. (In principle, all objects other than O can be considered to be background objects of O; however, in practice, only the anatomic neighbors of O matter.) The only new parameters are $Th_O^{\mathcal{M}}$ and $Th_B^{\mathcal{M}}$ used as model thresholds for indicating A_O and A_B, respectively. These parameters are used as follows:

$$\begin{aligned} A_O &= \{v \in C : f(v) \in Th_O \ \& \ \mu_O(v) \in Th_O^{\mathcal{M}}\}, \\ A_B &= \{v \in C : f(v) \in Th_B \ \& \ \mu_B(v) \in Th_B^{\mathcal{M}}\}. \end{aligned} \tag{9}$$

Algorithm *FMIRFC*

Input: Image I of $\mathcal{B}$, $FAM(\mathcal{B}, G)$, $FM^T(O_\ell)$ at recognition. Below, we assume $O = O_\ell$.
Output: O_ℓ^D.
Begin
*FC*1. Determine background B of O;
*FC*2. Retrieve affinities κ_O and κ_B from $FAM(\mathcal{B}, G)$;
*FC*3. Compute combined affinity κ;
*FC*4. Retrieve thresholds Th_O, Th_B, $Th_O^{\mathcal{M}}$, and $Th_B^{\mathcal{M}}$ from $FAM(\mathcal{B}, G)$ and determine seed sets A_O and A_B in I via (9);
*FC*5. Call the IRFC delineation algorithm with κ, A_O, A_B, and I as arguments;
*FC*6. Output image O_ℓ^D returned by the IRFC algorithm;
End

In our implementation, $Th_O^{\mathcal{M}}$ is fixed at [0,0.9] and [0,0.5] for non-sparse and sparse objects, respectively, and $Th_B^{\mathcal{M}}$ is set to [0,0].

Finally, we summarize the *FMIRFC* algorithm as shown in the box display.

5. Illustrations, experiments, results, and discussion

We will describe the image data sets in Section 5.1, present model-construction related results in Section 5.2, and illustrate and evaluate recognition and delineation results in Sections 5.3 and 5.4.

5.1. Image data

The data sets used for the three body regions are summarized in Table 2.

Data sets *DS*1 and *DS*2 are from CT and are selected from our hospital patient image database, and were verified to be of acceptable quality and radiologically normal, with exception of minimal incidental focal abnormalities, in the body regions for which they are chosen. Note the typical clinical resolution for pixel size (~1 mm) and slice spacing (5 mm) in these data sets and hence the challenge for object recognition and delineation. Our goal in focusing on these data was to challenge the AAR system to perform on typical clinical data sets. *DS*3 is from an on-going research project investigating the association of Polycystic Ovary Syndrome

Table 2
Summary of data sets used in the experiments.

Data identifier	Body region $\mathcal{B}$	Group G (age)	Number of subjects N	Image modality	Imaging protocol details	Image information
DS1	Thorax	50–60 male	50 normal	CT	Contrast-enhanced, axial, breath-hold	$512 \times 512 \times 51$–69, $0.9 \times 0.9 \times 5$ mm^3
DS2	Abdomen	50–60 male	50 normal	CT	Contrast-enhanced, axial, breath-hold	$512 \times 512 \times 38$–55, $0.9 \times 0.9 \times 5$ mm^3
DS3	Neck	8–17 male & female	15 normal	MRI	T2-weighted, axial & T1- & T2-weighted sagittal. T2: TR/TE = 8274.3/82.6 ms, T1: TR/TE = 517.7/7.6 ms	$400 \times 400 \times 35$–50, $0.5 \times 0.5 \times 3.3$ mm^3
DS4	Abdomen	8–17 male & female	14 6 normal, 8 obese patients	MRI	T2-weighted, axial. TR/TE = 1556.9/84 ms	$400 \times 400 \times 45$–50, $0.7 \times 0.7 \times 6$ mm^3

with Obstructive Sleep Apnea in obese pediatric female subjects (Arens et al., 2011). It consists of both axial and sagittal acquisitions and a mix of T1- and T2-weighted images. *DS*1-*DS*3 represent the three body regions for which the hierarchy of organs was depicted in Fig. 2. *DS*4 (Wagshul et al., 2013), however, is used for testing the ability of the AAR method to rapidly prototype an application by using existing models for the same body region. In this case, models built from *DS*2 from CT are deployed on *DS*4 from MRI.

In all data sets, any extra slices falling outside the body region $\mathcal{B}$ as per definition are removed manually first. Note the variation in the size of the body region in Table 2 (expressed roughly as slice spacing $\times$ number of slices). In the case of MRI, the resulting images are processed, first to suppress background non-uniformities and subsequently to standardize the image intensities (Nyul and Udupa, 1999). Standardization is a post-acquisition image processing technique which significantly minimizes the inter-subject and intra- and inter-scanner image intensity variations for the same tissue and achieves tissue-specific numeric meaning for MR images. It has been shown to significantly improve the accuracy of delineation algorithms (Zhuge and Udupa, 2009). It is done separately for each MRI protocol and body region. For *DS*1 and *DS*2, one half of the image data sets were used for model building, which included the estimation of the parameters of the recognition and delineation algorithms ($\mathcal{T}_1^m$, Th_ℓ, $\sigma_{\psi O}$, $m_{\varphi O}$, $m_{\varphi B}$, $\sigma_{\varphi O}$, and $\sigma_{\varphi B}$), and the remaining data sets were used for testing the methods. For *DS*3, the train-test sets were set up as 11 and 4, and this was repeated 30 times for different choices of 11 and 4 data sets. For *DS*4, all data sets were used for testing, and model building was based on one half of the data sets in *DS*2. This provided an interesting scenario for the challenge for the AAR method, in that, models built from normal CT data sets for one patient group were used for performing AAR on MRI data sets from normal subjects and patients from another group.

5.2. Model building

In Fig. 3, the organs defined in the image of one of the subjects employed in model building are displayed for each body region in different combinations of the organs. We have examined all data sets under *DS*1–*DS*3 in this manner which has helped us in properly understanding the organ relationships. This is crucial for devising effective hierarchies, recognition strategies, and delineation algorithms.

Fig. 4 displays fuzzy models $FM(O_\ell)$ of objects in various combinations for the three body regions. Since the volumes are fuzzy, they are volume rendered by using an appropriate opacity function. Note that although the models appear blurred, they portray the overall shape of the objects they represent and the object relationships. From consideration of the difficulties in model building, recognition, and delineation, we divided objects in the body into sparse, non-sparse, and hybrid groups. Sparse objects pose special challenges for recognition and delineation, stemming mostly from difficulties in model building. We will come back to these issues in Sections 5.3 and 5.4. Variations in the form, shape, and orientation of sparse objects cause them to overlap far less, or often not at all, compared to non-sparse objects, when forming the model by gathering fuzzy overlap information. In other words, the models tend to diffuse or become too fuzzy. For example, in AS (thorax), the descending aortic portion extends from superior to inferior. However, this part is often either bent from the vertical or is crooked, and the pattern of the brachiocephalic and subclavian arteries arising from the aortic arch is different. If the variation is just in orientation only, then aligning by orientation may produce sharper models. But the issue is not one of producing less fuzzy models but of building models that have the right/correct amount of fuzziness so that the recognition process will be least misguided by the model[9]. We will say more on this in Section 6. To study the effect of orientation alignment, we display in Fig. 5 models created without and with orientation adjustment, for several sparse as well as non-sparse objects from all three body regions. The volume renditions were created with exactly the same settings for each object for its two versions of models. Orientation adjustment does not produce any dramatic difference in the models created, although close scrutiny reveals that the model definition improves slightly; examine especially LPS, AIA, AS, and Lvr.

Relating to the fifth element η of $FAM(\mathcal{B}, G)$, we show in Tables 3–5 correlations among objects in their size for the three body regions[10]. Object size is determined as explained in Section 2.3. As may be expected, bilateral organs, such as LPS and RPS, LKd and RKd, and LT and RT, are strongly correlated in size. That is, their sizes go together, whatever way they may be related to the subject's body size. There are also other interesting strong, poor (or no), and even weak negative, correlations, as highlighted in the tables; for example, TSk with RS and RPS; VS with TB, PC, and E; ASkn with ASTs, SAT and Msl; ASTs with SAT and Msl; Msl with SAT; NSkn with A&B; Ad with NSkn, FP, NP, and SP. Although we have not explored the utility of such information in this paper, we envisage that this and other information will be useful in devising hierarchies more intelligently than guided by just anatomy, and hence in building better $FAM(\mathcal{B}, G)$.

5.3. Object recognition

Results for recognition are summarized in Figs. 6–8 and Tables 6–9 for the different body regions. Figs. 6–8 and Tables 6–8 illustrate recognition results for the three body regions for the best

[9] This dilemma of the disconnection between model building and recognition is common to all model/atlas-based methods and is the real challenge in automatic recognition of sparse and hybrid objects.

[10] For this analysis, we have used all image data sets since the information provided by this analysis does not influence at present the testing of AAR algorithms for recognition and delineation.

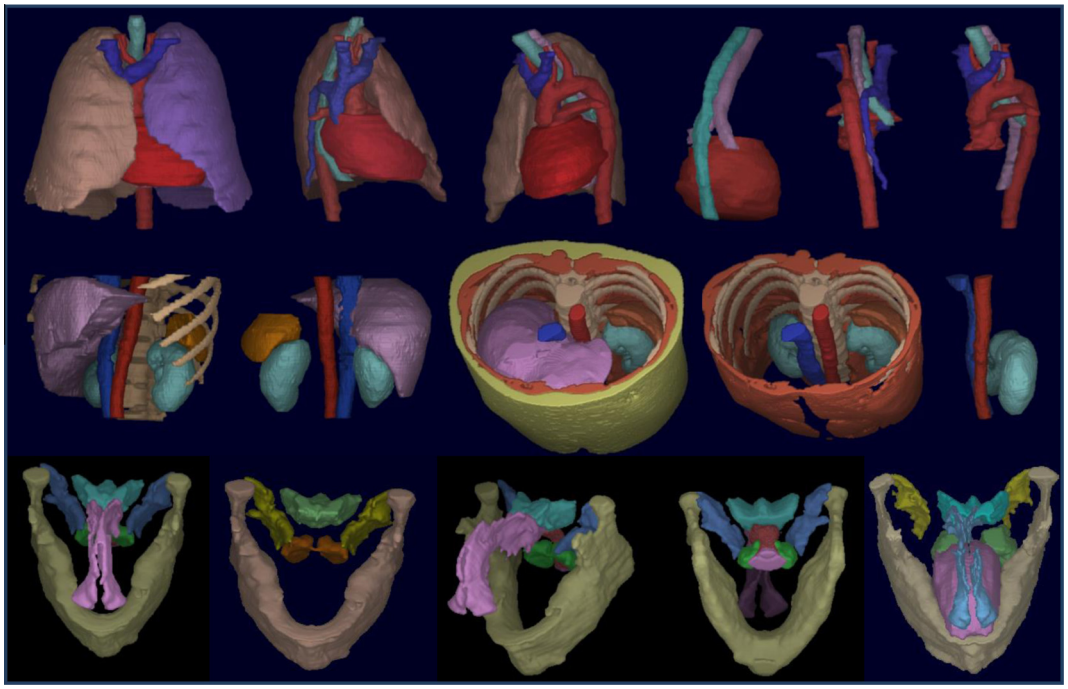

Fig. 3. Organs from one training set for each body region are displayed via surface rendering. For each row, objects in one picture are listed as {...}. Top row: Thorax. 3rd picture: {RPS, TB, E, AS, VS, PC}. Middle row: Abdomen. 3rd picture: {Ask, Lvr, LKd, IVC, AIA, Spl, SAT, Msl}. Bottom row: Neck. 5th picture: {Mnd, Tng, NP, OP, Ad, FP, Tnsl}.

set up involving orientation adjustment selectively for different objects. The alignment strategy was as follows for the different objects in these results.

The recognition accuracy is expressed in terms of position and size. The position error is defined as the distance between the geometric centers of the known true objects in $\mathcal{I}^b$ and the center

$$\begin{aligned}&\text{Non-sparse \& hybrid objects}: \text{RS}, \text{LPS}, \text{RPS}, \text{IMS}, \text{TSk}, \text{Ask}, \text{Kd}, \text{Spl}, \text{Msl}, \text{LKd}, \text{RKd}, \text{A\&B}, \text{FP},\\ &\qquad \text{NSTs}, \text{Mnd}, \text{Tnsl}, \text{Tng}, \text{SP}, \text{Ad}, \text{RT}, \text{LT} - \text{no orientation alignment}.\\ &\text{Sparse objects}: \text{TB}, \text{E}, \text{AS}, \text{VS}, \text{AIA}, \text{IVC}, \text{Phrx}, \text{NP}, \text{OP} - \text{orientation alignment by all axes}.\end{aligned} \tag{10}$$

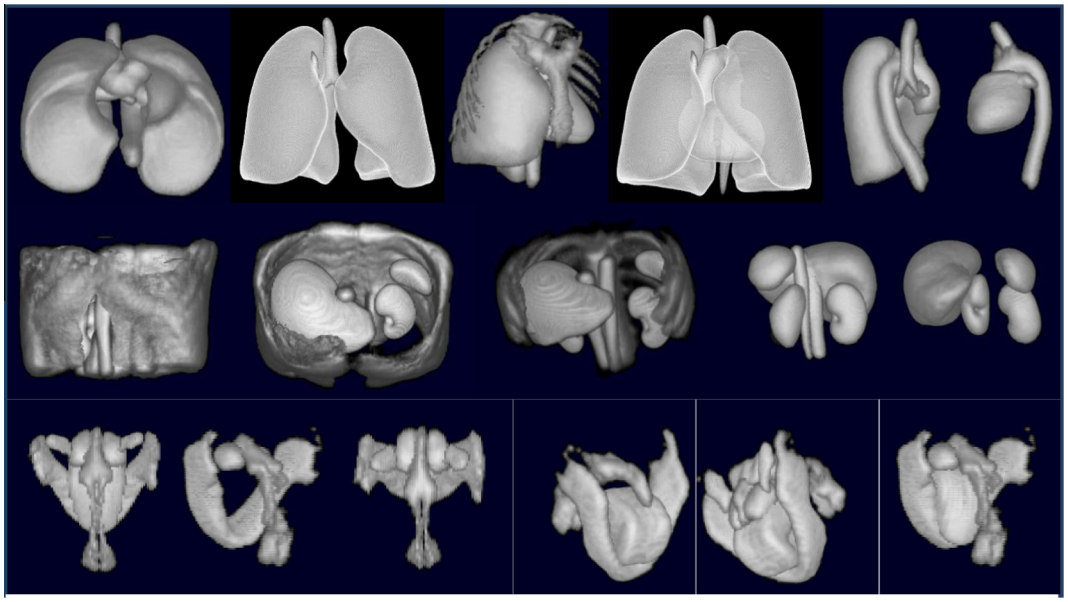

Fig. 4. Volume renditions of fuzzy models of objects in different combinations for the three body regions. For each row, objects in one picture are listed as {...}. Top row: Thorax. 5th picture: {LPS, AS, TB}. Middle row: Abdomen. 3rd picture: {ASk, Lvr, LKd, RKd, AIA, IVC, Spl}. Bottom row: Neck: 5th picture: {Mnd, Tng, NP, OP, Ad, FP}.

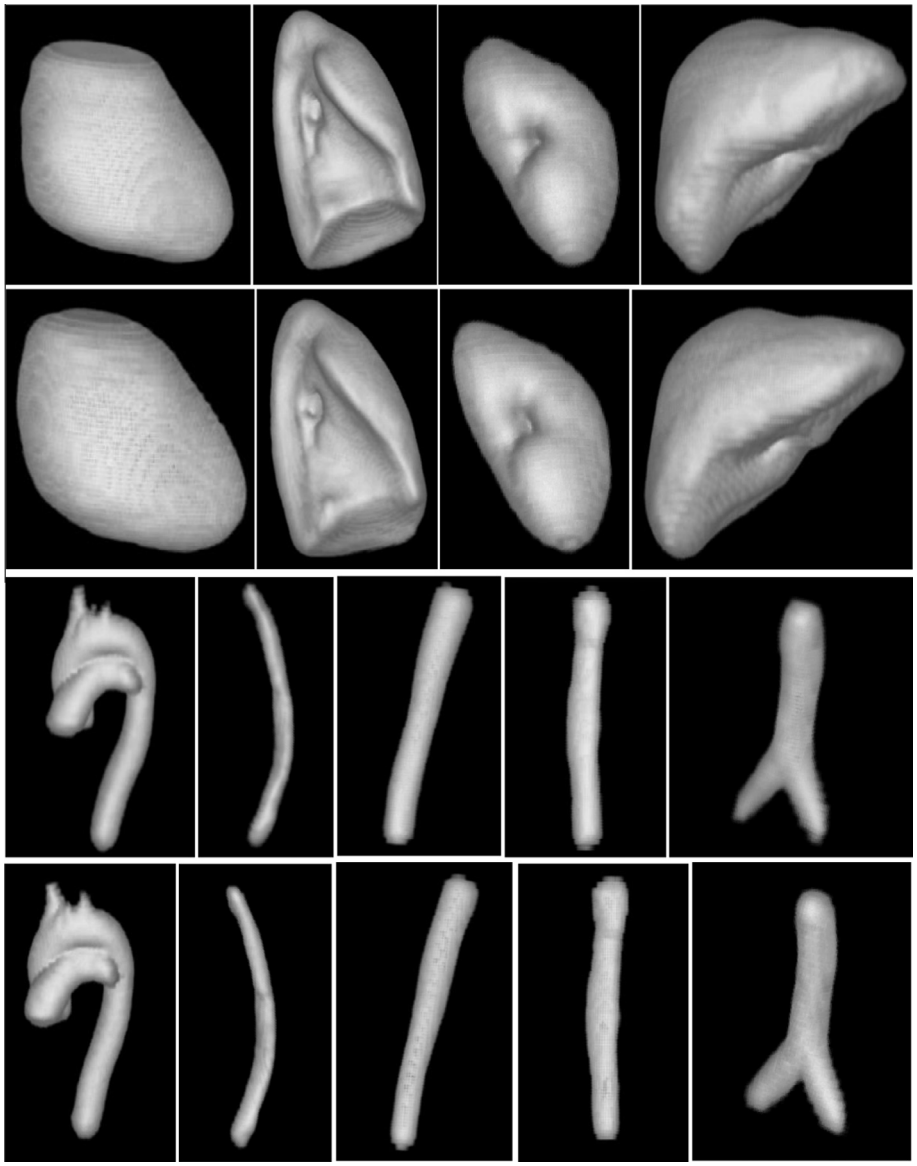

Fig. 5. Volume renditions of fuzzy models created without (Rows 1 and 3) and with (Rows 2 and 4) orientation alignment for several non-sparse (Rows 1 and 2) and sparse (Rows 3 and 4) objects. Row 1: PC, RPS, LKd, Lvr. Row 3: AS, E, AIA, IVC, TB.

of the adjusted fuzzy model $FM^T(O_\ell)$. The size error is expressed as a ratio of the estimated size of the object at recognition and true size. Values 0 and 1 for the two measures, respectively, indicate perfect recognition. Note in Figs. 6–8 that the model bleeds into adjacent tissue regions with some membership value since it is fuzzy. This should not be construed as wrong segmentation. The main issue is if the model placement via recognition is accurate enough to yield good delineation. Similarly and due to the slice visualization mode, sparse object components may appear to be missed or to present with low membership values.

Although we have not conducted extensive experiments to test all possible arrangements for orientation alignment for non-sparse and sparse objects, generally we found that orientation adjustment for non-sparse objects does not improve recognition results. In some cases like PC, it may actually lead to deterioration of results. In our experience, the set up in (10) turned out to be an excellent compromise from the viewpoint of accuracy of results and efficiency. For comparison, we demonstrate in Table 9 recognition results for the thorax with no orientation adjustment for any object in both model building and recognition.

Size error is always close to 1 for all body regions and objects. Generally, recognition results for non-sparse objects are excellent with a positional error of mostly 1–2 voxels. Note that for *DS*1 and *DS*2, voxels are quite large.[11] We observed that, the positional accuracy within the slice plane is better than across slices. In other words, errors listed in the tables are mostly in the third dimension in which voxel size is large. Orientation adjustment improves recognition somewhat for some sparse objects, but has negligible effect for non-sparse objects, at least in the thorax.

The recognition results for the MRI data set *DS*4 are demonstrated in Fig. 9 and Table 10. Again, since the model is fuzzy, it will encroach into adjacent tissue regions with some membership value. Since our goal here was just to measure subcutaneous adiposity, the hierarchy was simplified as shown in Fig. 9. Again the position error is 1–2 voxels. These results are particularly noteworthy since they are generated by using the models built from image

[11] Since recognition results do not improve much with finer discretization of the model but only increase computation for recognition, we construct models with isotropic voxels of side equal to one half of the largest dimension of the voxels in the original data. Thus for *DS*1 and *DS*2, the model voxels are of size $2.5 \times 2.5 \times 2.5$ mm^3.

Table 3
Size correlation among objects of the Thorax.

	TSkn	RS	TSk	IMS	RPS	TB	LPS	PC	E	AS	VS
TSkn	1										
RS	0.76	1									
TSk	0.76	0.93	1								
IMS	0.48	0.76	0.71	1							
RPS	0.6	0.92	0.88	0.75	1						
TB	0.06	0.41	0.5	0.56	0.59	1					
LPS	0.64	0.93	0.87	0.74	0.96	0.57	1				
PC	0.47	0.51	0.45	0.65	0.28	0.11	0.3	1			
E	0.42	0.65	0.56	0.58	0.72	0.58	0.78	0.18	1		
AS	0.44	0.53	0.49	0.71	0.54	0.24	0.51	0.35	0.35	1	
VS	0.3	0.31	0.35	0.34	0.34	0.09	0.34	−.01	0.05	0.42	1

Table 4
Size correlation among objects of the Abdomen.

	ASkn	ASk	ASTs	Lvr	SAT	Msl	Spl	RKd	LKd	AIA	IVC
ASkn	1										
ASk	0.68	1									
ASTs	0.9	0.8	1								
Lvr	0.61	0.48	0.58	1							
SAT	1	0.69	0.92	0.61	1						
Msl	0.91	0.79	0.99	0.63	0.94	1					
Spl	0.62	0.43	0.61	0.51	0.65	0.62	1				
RKd	0.53	0.64	0.57	0.61	0.51	0.6	0.34	1			
LKd	0.53	0.56	0.52	0.51	0.49	0.54	0.34	0.87	1		
AIA	0.6	0.85	0.7	0.27	0.58	0.68	0.49	0.51	0.5	1	
IVC	0.32	0.58	0.47	0.29	0.32	0.46	0.3	0.38	0.36	0.67	1

Table 5
Size correlation among objects of the Neck.

	NSkn	A&B	FP	Mnd	NP	OP	Tng	SP	Ad	LT	RT
NSkn	1										
A&B	0.89	1									
FP	0.76	0.81	1								
Mnd	0.75	0.96	0.83	1							
NP	0.39	0.12	−.06	−.12	1						
OP	0.63	0.59	0.44	0.54	0.14	1					
Tng	0.83	0.75	0.76	0.66	0.19	0.65	1				
SP	0.5	0.27	0.23	0.14	0.46	0.26	0.37	1			
Ad	−.2	0.61	−.19	0.1	−.29	−.06	−.07	−.19	1		
LT	0.61	0.56	0.58	0.48	0.28	0.5	0.64	0.25	−.1	1	
RT	0.61	0.56	0.58	0.48	0.28	0.5	0.64	0.25	−.1	1	1

data sets acquired from a different modality, namely CT, and for a different group with an age difference of about 40 years and with a different gender. This underscores the importance of understanding the dichotomy between recognition and delineation. Recognition is a high-level and rough process which gives anatomic context. The models do not have to be, and we argue should not be, detailed attempting to capture fine details. Obtaining the anatomic context is a necessary step for achieving accurate delineation. It is important to note here that for the cross modality operation to work in this manner, the MR image intensities must be standardized (Nyul and Udupa, 1999).

5.4. Object delineation

Sample delineation results are displayed in Figs. 10–13 for *DS*1–*DS*4. Delineation accuracy statistics for these data sets, expressed as false positive and false negative volume fractions (FPVF, FNVF) as well as mean Hausdorff distance (HD) between the true and delineated boundary surfaces, are listed in Tables 11–14. The HD measure is defined as the mean over all test subjects of the median of the distances of the points on the delineated object boundary surface from the true object boundary surface.

Delineation results for VS (Thorax) are not presented since the recognition accuracy for VS is not adequate for reliable delineation. We note that the delineation of 21 non-sparse objects achieves a mean FPVF and FNVF of 0.02 and 0.08, respectively, and a mean HD of 0.9 voxels, which are generally considered to be excellent. Six sparse objects also achieve good delineation outcome, with the above mean measures reading 0.05, 0.15, and 1.5, respectively. However, sparse objects VS, E, IVC, Mnd, and NP pose challenges for effective delineation. Often, even when their recognition is effective, it is difficult to guarantee placement of seed sets A_O and A_B appropriately within and outside these objects because of their sparse nature. In *DS*3 (MR images of neck), it is very difficult to properly delineate Mnd, NP, and OP because of their poor definition in the image. To test the effectiveness of the models created from these data (*DS*3) in segmenting the same objects on CT data of a group of three different pediatric subjects, we devised a simple

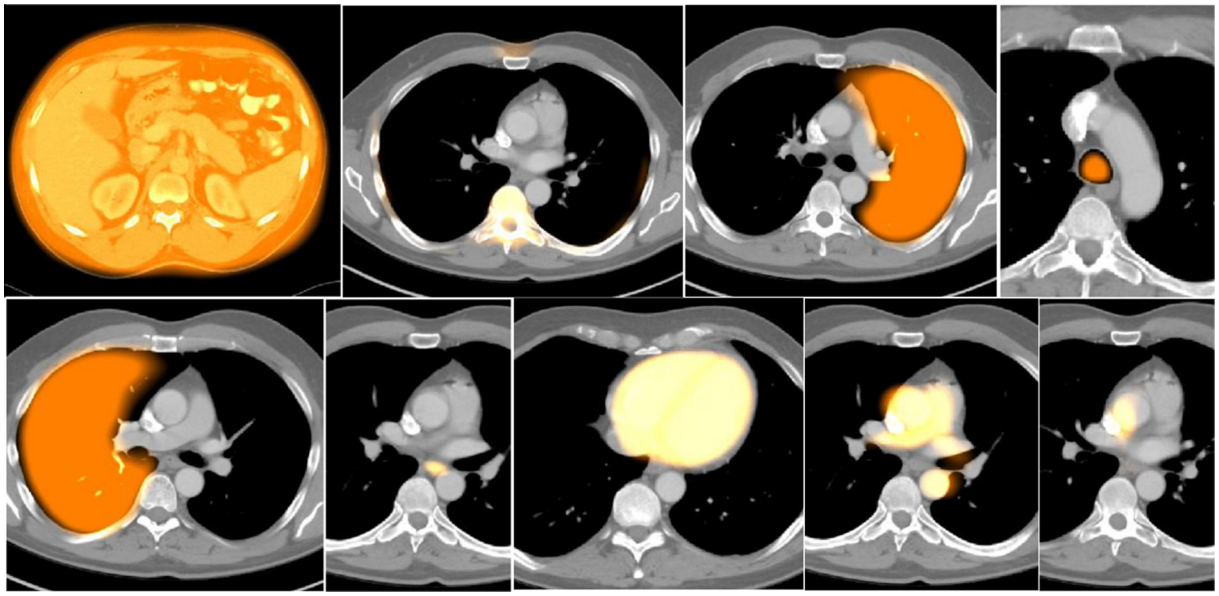

Fig. 6. Sample recognition results for Thorax for the alignment strategy shown in (10). Cross sections of the model are shown overlaid on test image slices. Left to right: TSkn, TSk, LPS, TB, RPS, E, PC, AS, VS.

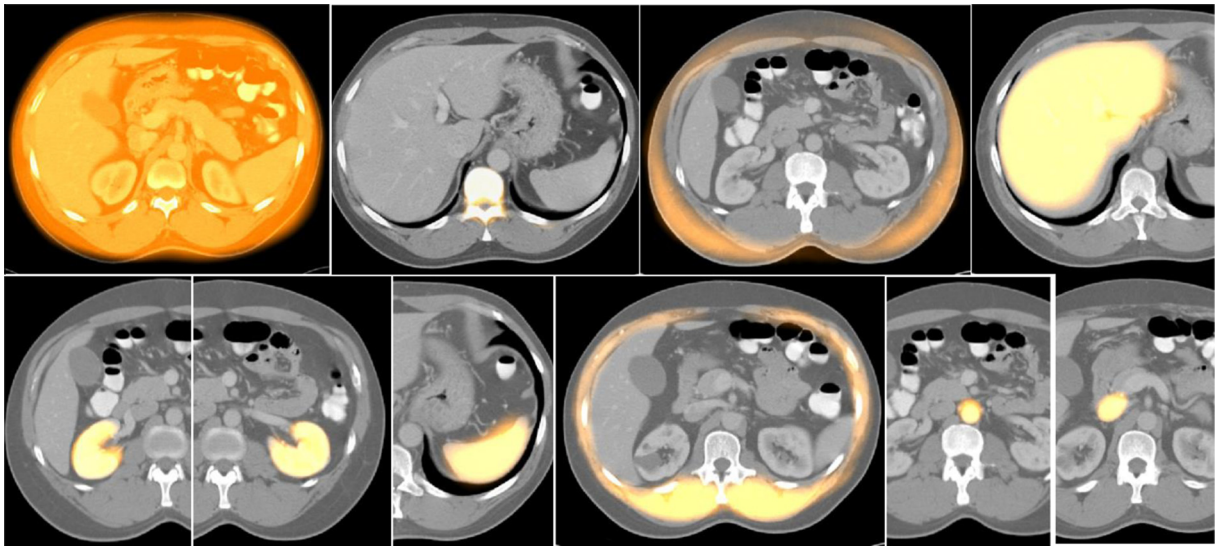

Fig. 7. Sample recognition results for Abdomen for the alignment strategy shown in (10). Cross sections of the model are shown overlaid on test image slices. Left to right: ASkn, ASk, SAT, Lvr, RKd, LKd, Spl, Msl, AIA, IVC.

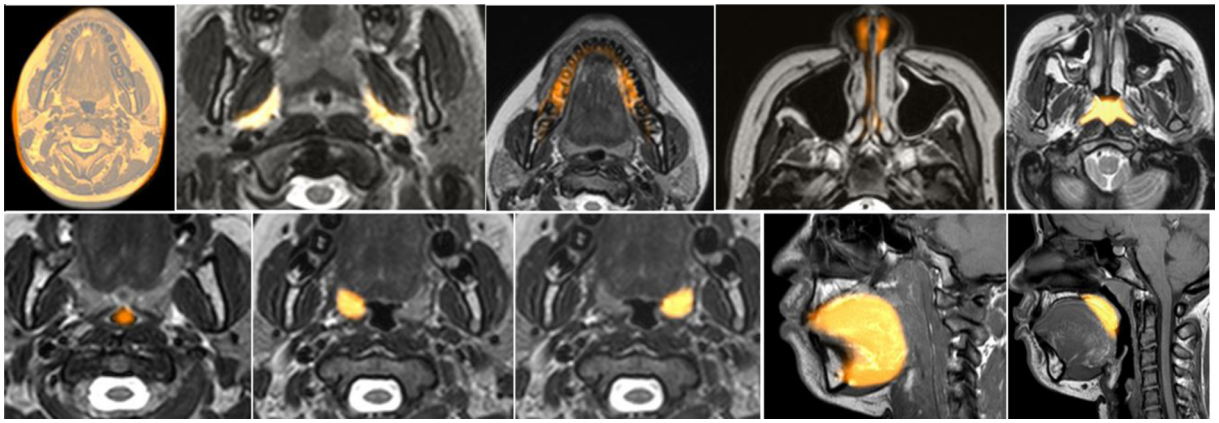

Fig. 8. Sample recognition results for Neck for the alignment strategy shown in (10). Cross sections of the model are shown overlaid on test image slices. Left to right: NSkn, FP, Mnd, NP (note that NP is a combination of nasal cavity and nasopharynx), Ad, OP, RT, LT, Tng, SP.

Table 6
Recognition results (mean, standard deviation) for Thorax for the strategy in (10). ("Mean" excludes VS.)

	TSkn	RS	TSk	IMS	LPS	TB	RPS	E	PC	AS	VS	Mean
Location error (mm)	3.9	5.5	9.0	5.6	6.3	11.6	10.4	9.8	8.6	10.7	31.8	8.1
	1.5	2.3	5.0	3.5	3.1	5.0	4.7	4.8	5.0	5.4	12.0	4.0
Size error	1.0	0.99	0.96	0.95	0.97	0.91	0.98	0.9	0.95	1.01	0.77	0.96
	0.01	0.02	0.05	0.05	0.03	0.06	0.04	0.14	0.05	0.08	0.06	0.06

Table 7
Recognition results (mean, standard deviation) for Abdomen for the strategy in (10).

	ASkn	SAT	ASk	Lvr	ASTs	Kd	Spl	Msl	AIA	IVC	RKd	LKd	Mean
Location error (mm)	5.9	20.2	11.7	7.9	7.2	10.6	11.6	7.7	8.2	8.7	11.3	7.3	9.8
	3.4	8.5	7.9	5.4	3.0	9.8	13.9	3.6	2.8	7.2	11.6	7.4	7
Size error	1.0	0.97	0.96	0.93	1.0	0.94	1.2	1.01	1.1	1.15	0.97	0.93	1.01
	0.02	0.03	0.06	0.07	0.02	0.09	0.19	0.03	0.13	0.1	0.1	0.08	0.07

Table 8
Recognition results (mean, standard deviation) for Neck for the strategy in (10).

	NSkn	A&B	FP	NSTs	Mnd	Phrx	Tnsl	Tng	SP	Ad	NP	OP	RT	LT	Mean
Location error (mm)	3	7.8	4.2	4.8	12.5	10.4	2.8	4.9	5.1	1.8	11.1	10	2.9	2.3	5.96
	1.2	3.8	2.1	2.1	3.7	4.5	1.8	2.8	1.8	0.8	6.8	8.7	2.2	2.1	1.96
Size error	1	0.9	1	0.92	0.74	0.8	1	1.02	0.93	0.9	0.65	0.74	0.92	0.9	0.93
	0.01	0.03	0.03	0.06	0.05	0.04	0.1	0.06	0.24	0.12	0.07	0.2	0.11	0.12	0.04

Table 9
Recognition results for Thorax with no orientation alignment. ("Mean" excludes VS.)

	TSkn	RS	TSk	IMS	LPS	TB	RPS	E	PC	AS	VS	Mean
Location error (mm)	3.9	5.5	9	5.6	6.3	8	10.4	14.2	8.6	8.1	33.6	8.0
	1.5	2.3	5	3.5	3.1	6.5	4.7	10.5	5	7.5	15.1	4.9
Size error	1.01	0.99	0.96	0.95	0.97	0.83	0.98	0.85	0.95	0.99	0.77	0.95
	0.01	0.02	0.05	0.05	0.03	0.08	0.04	0.12	0.05	0.08	0.06	0.05

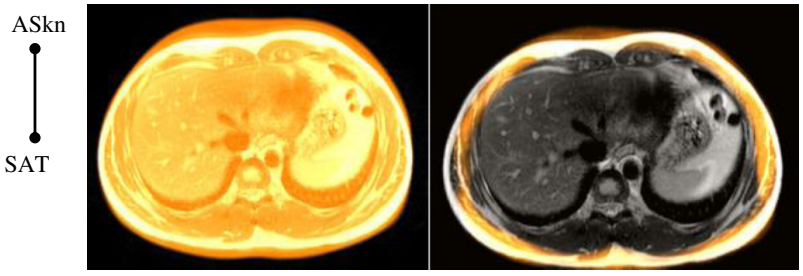

Fig. 9. The hierarchy used (left) and sample recognition results for *DS*4 (right) with model cross section overlaid on test image slices for ASkn and SAT.

Table 10
Recognition accuracy for the objects shown in Fig. 9.

	ASkn	SAT
Position error (mm)	4.6	12.97
	2.5	5.3
Size error	1.01	1
	0.05	0.03

hierarchy with NSkn as the root and with Mnd, NP, and OP as its offspring objects. The delineation results obtained for these four objects were excellent, with a mean FPVF of 0, 0.01, 0, and 0.02, and mean FNVF of 0.01, 0.01, 0.02 and 0.1, respectively.

5.5. Comparison with a non-hierarchical approach

To study the effect of the hierarchy and the knowledge encoded in it on recognition, we list in Table 15 the recognition performance of a non-hierarchical approach. The results are shown for Thorax wherein each object is recognized on its own by using the same fuzzy models $FM(O_\ell)$ as used in the hierarchical AAR system. The initial pose for search is taken to be the center of the image and search range covers roughly the whole body region with the scale factor range the same as that for the hierarchical approach. In comparison to the hierarchical approach (Tables 6 and 9), it is clear that non-hierarchical recognition performance is much worse.

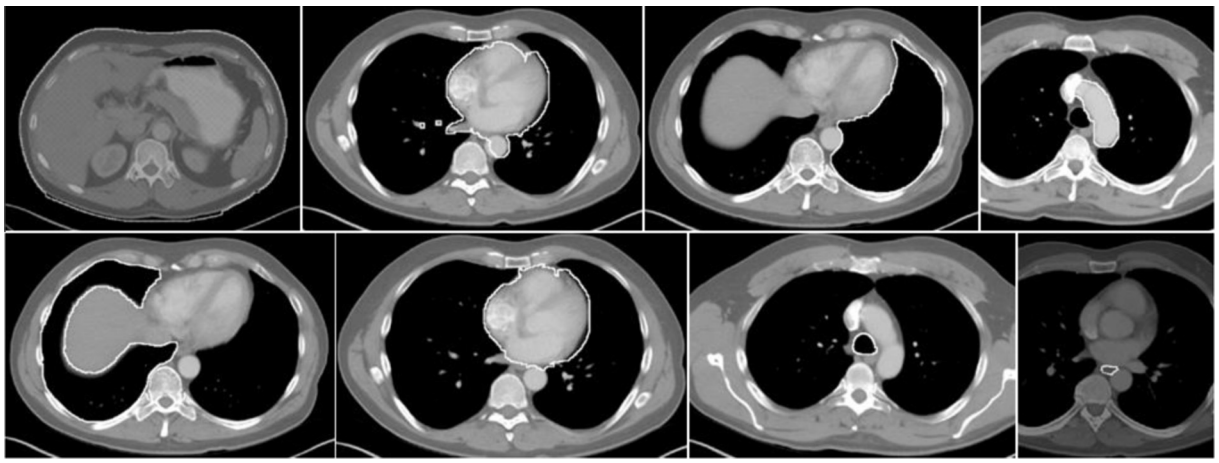

Fig. 10. Sample delineation results for Thorax. Left to right: TSkn, IMS, LPS, AS, RPS, PC, TB, E.

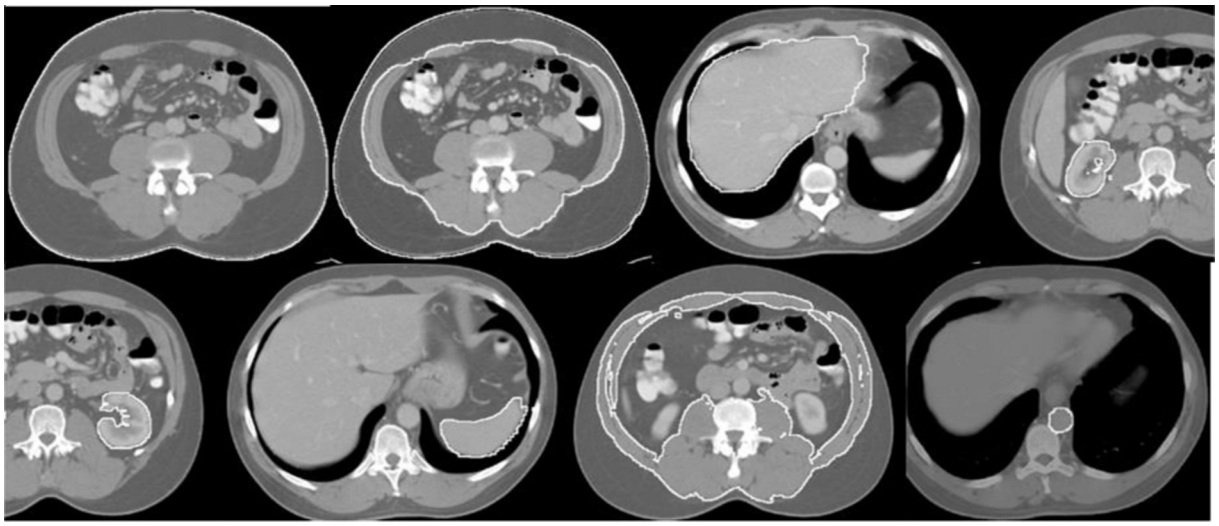

Fig. 11. Sample delineation results for Abdomen. Left to right: ASkn, SAT, Lvr, SAT, RKd, LKd, Spl, Msl, AIA.

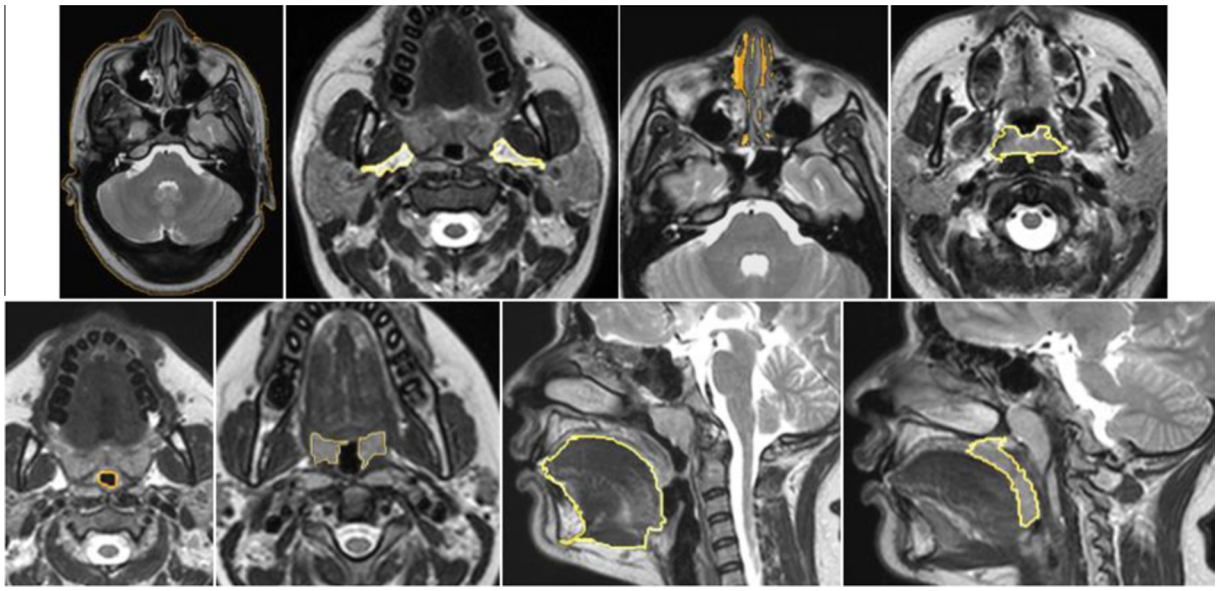

Fig. 12. Sample delineation results for Neck. Left to right: NSkn, FP, NP, OP, RT, LT, Tng, SP, Ad.

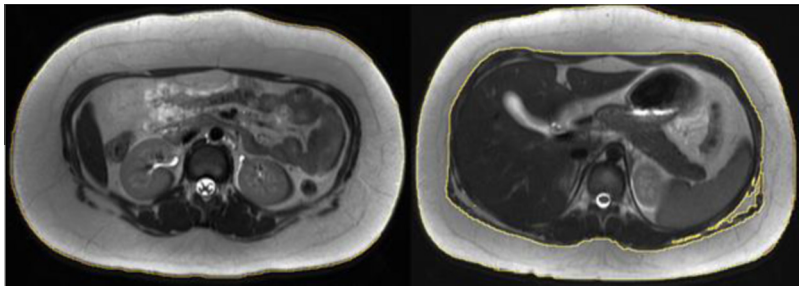

Fig. 13. Sample delineation results for *DS*4. ASkn (left) and SAT (right).

Table 11
Delineation results for Thorax (mean & standard deviation).

	TSkn	RS	TSk	IMS	LPS	RPS	E	PC	TB	AS
FPVF	0.02	0.0	0.19	0.03	0.01	0.01	0.0	0.01	0.01	0.01
	0.02	0.0	0.05	0.01	0.03	0.02	0.0	0.00	0.00	0.00
FNVF	0.05	0.06	0.13	0.07	0.04	0.04	0.49	0.09	0.16	0.17
	0.06	0.04	0.07	0.07	0.02	0.02	0.19	0.06	0.14	0.17
HD (mm)	3.6	1.24	10.6	6.2	2.9	2.1	3.1	3.5	5.2	5.3
	4.5	0.42	2.4	1.8	8.8	4.7	0.87	1.3	1.8	2.5

Table 12
Delineation results for Abdomen (mean & standard deviation).

	ASkn	ASk	Lvr	ASTs	SAT	RKd	LKd	Spl	Msl	AIA
FPVF	0.01	0.06	0.04	0.12	0.05	0.00	0.01	0.0	0.13	0.01
	0.00	0.01	0.02	0.05	0.03	0.00	0.01	0.0	0.03	0.0
FNVF	0.05	0.14	0.1	0.15	0.12	0.13	0.1	0.13	0.09	0.13
	0.08	0.09	0.05	0.09	0.02	0.04	0.02	0.03	0.08	0.03
HD (mm)	1.7	6.9	5.3	1.74	1.6	2.4	5.4	6.8	2.5	5.6
	2.7	1.5	1.6	1.0	0.8	1.1	4.8	6.0	1.1	1.8

Table 13
Delineation results for Neck (mean & standard deviation).

	NSkn	FP	Mnd	NP	OP	RT	LT	Tng	SP	Ad
FPVF	0.0	0.0	0.01	0.01	0.0	0.01	0.01	0.02	0.01	0.0
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.01	0.01	0.0
FNVF	0.0	0.1	0.49	0.32	0.2	0.06	0.06	0.02	0.08	0.07
	0.01	0.05	0.08	0.2	0.02	0.02	0.01	0.01	0.01	0.04
HD (mm)	2.8	0.83	3.3	3.8	7.6	3.3	3.2	8.4	8.03	2.2
	0.06	0.53	0.56	1.01	2.4	0.62	1.4	1.92	4.0	0.3

Table 14
Delineation results for *DS*4.

	ASkn	SAT
FPVF	0.0	0.06
FNVF	0.03	0.01
HD (mm)	1.7	3.9

5.6. Computational considerations

Program execution times are estimated on a Dell computer with the following specifications: 4-core Intel Xeon 3.6 GHz CPU with 8 GB RAM and running the Linux-jb18 3.7.10–1.16 operating system. Mean computational times for the AAR steps are listed in Table 16. Model building includes the construction of fuzzy models and the estimation of ρ, λ, and all parameters related to recognition and delineation, including the optimal threshold parameters Th_ℓ. This latter step takes about 12 s per object. As seen from Table 16, each of the three main operations takes under 1 min per object. Among these operations, only the time for model building depends on the number of training data sets, while recognition and delineation are independent of this factor. On average, model building times per object per training data set for Thorax, Abdomen, and Neck are, respectively, 1.4 s, 1.7 s, and 1 s. In statistical atlas based methods, the computational time for image registration becomes the bottleneck. Our calculation taking Elastix as a representative registration tool kit (Klein et al., 2010) indicates that the creation of a single atlas for each of the 11 objects of the Thorax at a reduced image resolution of $2.5 \times 2.5 \times 2.5$ mm^3 for the 25 training data sets of *DS*1 would take about 23.5 h compared to 6.4 min for the AAR system. The time per object for recognition and delineation

Table 15
Recognition results for Thorax: non-hierarchical approach (mean & standard deviation).

	TSkn	RS	TSk	IMS	LPS	TB	RPS	E	PC	AS	VS	Mean
Location error (mm)	10.5	12.9	21.1	27.7	91.4	53.3	72.3	42.4	45.5	23.1	82.2	43.8
	9.5	13.1	21.8	9.8	10.8	20.9	12.9	34.5	12.5	15.2	33.8	17.7
Size error	1.0	1.01	0.96	0.92	0.8	0.82	0.8	0.86	0.9	0.97	0.81	0.9
	0.02	0.09	0.08	0.07	0.09	0.06	0.07	0.14	0.06	0.11	0.08	0.08

Table 16
Mean computational time in *seconds* per object for different operations and body regions.

Operation	Thorax	Abdomen	Neck
Model building	35	42	24
Object recognition	30	46	6
Object delineation	47	56	24

can also take several minutes for these methods. Even with 100 data sets for training and 15 objects in a body region, the total time needed for the AAR model building step would be about 40 min, whereas atlas building may take days to complete especially when multi-atlas strategies are used.

5.7. Comparison with other methods

The publications reporting works that are directly related to our work in spirit are (Baiker et al., 2010; Chu et al., 2013; Criminisi et al., 2013; Lu et al., 2012; Linguraru et al., 2012; Okada et al., 2008; Zhou et al., 2012). In Table 17, we present a comparison to our AAR system based on the results reported in these works. We note that a quantitative grading/understanding of the methods is impossible since the data sets used, acquisition protocols and resolutions, considered objects, training and test data set subdivisions, cross validation strategies, and computing platforms are all different in these methods. Interestingly, a commonality among them is that they all focused on CT image data sets.

Among these methods, (Chu et al., 2013; Linguraru et al., 2012; Lu et al., 2012; Okada et al., 2008) comprise one group wherein the body region of focus was the pelvis or abdomen, with 3–5 objects considered for segmentation. They all employ an object localization step, which is achieved either through an atlas (Chu et al., 2013; Linguraru et al., 2012; Okada et al., 2008), statistical shape models (Okada et al., 2008), or machine learning techniques (Lu et al., 2012), and subsequently a delineation step that uses graph cuts (Chu et al., 2013; Linguraru et al., 2012), information theory (Lu et al., 2012), and MAP or ML estimation (Chu et al., 2013; Okada et al., 2008). In the second group (Criminisi et al., 2013; Zhou et al., 2012), the aim is only to locate the objects via machine learning techniques. The third group is constituted by (Baiker et al., 2010), the only work that considered body-wide organs, but in mice, using a kinematic model of the skeletal joints to localize objects relative to different skeletal components.

We observe that, for the same objects (liver, kidneys, and spleen), our results are comparable to, often better than, the current results from literature, especially considering the 5 mm slice spacing and the equal training-to-test data set proportion for our evaluation. We conclude that the development of a general AAR system that can be readily applied and adapted to different body regions, multitudes of organs, and modalities has not yet been demonstrated. Perhaps some of the above methods can be made to work in this general manner. However, we believe that this may require considerable further development and innovation.

6. Concluding remarks

In this paper, we presented a general body of methods for automatic anatomy recognition and delineation whose principles are not tied to any specific body region, organ system, or imaging modality. We took a fuzzy approach for building the models and attempted to harness as much specific anatomic information as possible to be embedded into the fuzzy anatomic model. We demonstrated the generality of the approach by examining the performance of the same AAR system on three different body regions using CT and MR image data sets. We also illustrated the potential of the system for rapid prototyping by demonstrating its adaptability to a new application on a different modality (*DS*4). Our system is set up to operate fully automatically. All image modality-specific parameters needed – threshold intervals for objects in $\mathcal{B}$ for recognition and affinity parameters for delineation – are estimated automatically from the training data sets. When a new application is sought at a modality different from those considered in the anatomy model $FAM(\mathcal{B}, G)$, a few sample segmentations of the objects of interest and the matching images are needed for relearning

Table 17
A comparison with the current methods from the literature that are related to our work. Unknown and irrelevant entries are indicated by "∼".

Method	Objects	Voxel size (mm^3)	Training-to-test data proportion	Location error (mm)	Region overlap (Dice, Jackard Index (JI), etc.)
Lu et al. (2012)	Prostate, bladder, rectum	∼ × ∼ × 0.8–5	141–47, 4-fold	2.4–4.2	∼
Linguraru et al. (2012)	Liver, spleen, kidneys	$(0.5–0.9)^2 \times 1–5$	27–1, 28-fold	0.8–1.2	90.9–94.8%
Okada et al. (2008)	Liver, vena cava, gallbladder	0.7 × 0.7 × 2.5	20–8	(for liver) 1.5–2.8	88%
Chu et al. (2013)	Liver, spleen, pancreas, kidneys	$(0.55–0.82)^2 \times 0.7–1$ (estimated)	90–10, 10-fold	∼	56% (pancreas-JI) to 95.2% (liver-Dice)
Criminisi et al. (2013)	26 anatomic structures in the torso	$(0.5–1)^2 \times 1–5$	318–82	9.7–19.1 (mean for each structure)	∼
Zhou et al. (2012)	12 organ regions in thorax, abdomen, pelvis	$(0.6–0.7)^3$	300–1000	6–14 for mode locations	∼
Baiker et al. (2010)	Brain, heart, kidneys, lungs, liver, skeleton	$(0.332)^3$	MOBY atlas, 26 datasets	∼	47–73%

these image intensity-related parameter values (specifically, Th_ℓ and the affinity parameters). All other modality-independent aspects of the model do not need retraining. In the case of MRI, images from each separate MRI protocol have to be standardized for image intensity so that setting up these parametric values becomes sensible. Separation of modality-independent from dependent aspects, organization of objects in a hierarchy, encoding object relationship information into the hierarchy, optimal threshold-based recognition learning, and fuzzy model-based IRFC are novel and powerful concepts with consequences in recognition and delineation, as we demonstrated in this paper.

While the above strengths of this AAR system are quite unique as revealed in our literature review, the system has some limitations at present. First, we have not studied the performance of the system on patient images that contain significant pathology. However, we note that *DS4* indeed includes image data sets of patients who are obese. Note also that these image data sets are from a very different age and gender group and on a different imaging modality from those used to build $FAM(\mathcal{B}, G)$. We believe that it is essential to make the system operate satisfactorily on normal or near-normal images before testing it on images with diverse pathologies. As such, we are currently in the process of testing the system on organs and organ systems with significant pathology in all three body regions focusing on specific disease processes.

Second, the accuracy is inadequate for some sparse objects for recognition (VS, IVC) and delineation (E, Mnd, NP). Also we have

Table A1
Anatomic definitions of organs considered in this paper.

	Acronym	Definition of object
Thoracic objects		
Thoracic skin	TSkn	The outer boundary of the thoracic skin (arms excluded). The interior region constitutes the entire thoracic body region. The inferior boundary is defined to be 5 mm below the base of the lungs and the superior boundary is defined to be 15 mm above the lung apices
Thoracic skeleton	TSk	All skeletal structures contained in the thoracic body region, including the spine, ribs, sternum, and the portions of the scapulae and clavicles that are inside the body region
Respiratory system	RS	Grouping of RPS, LPS, and TB
Right lung	RPS	The outer boundary of the right lung along the right pleura
Left lung	LPS	The outer boundary of the left lung along the left pleura
Trachea and bronchi	TB	The outer boundary of the trachea and bronchi from the superior thoracic trachea to the distal main stem bronchi
Internal mediastinum	IMS	Grouping of PC, E, AS, and VS
Pericardial region	PC	Region within the boundary of pericardial sac. The superior aspect is defined by the branching of the main pulmonary artery
Esophagus	E	The outer boundary of the esophagus from the superior aspect of thorax to the level of gastric cardia
Arterial system	AS	The outer boundary of the ascending aorta, aortic arch, descending thoracic aorta, pulmonary arteries, innominate artery, proximal left common carotid artery, and proximal left subclavian artery. The superior aspect is defined by the branching of the innominate artery
Venous system	VS	The outer boundary of the superior vena cava, right and left brachiocephalic veins, and azygos vein
Abdominal objects		
Abdominal skin	ASkn	The outer boundary of the abdominal skin. The interior region constitutes the entire abdominal body region. The superior boundary is defined by the superior aspect of the liver. The inferior boundary is defined by the bifurcation of the abdominal aorta into the common iliac arteries
Abdominal skeleton	ASk	All skeletal structures contained in the abdominal body region, including lumbar spine and portion of the inferior ribs within the body region
Soft tissue	ASTs	Grouping of Kd, Spl, Msl, AIA, IVC
Kidneys	Kd	Grouping of RKd and LKd
Right kidney	RKd	The outer boundary of the right kidney. All external blood vessels are excluded
Left kidney	LKd	The outer boundary of the left kidney. All external blood vessels are excluded
Spleen	Spl	The outer boundary of the spleen. All external blood vessels are excluded
Muscle	Msl	The outer boundaries of the abdominal musculature, including the rectus abdominis, abdominal oblique, psoas, and paraspinal muscles
Abdominal aorta	AIA	The outer boundary of the abdominal aorta. The superior and inferior slices of AIA are the same as those of the abdominal region
Inferior vena cava	IVC	The outer boundary of the inferior vena cava. The superior and inferior slices of IVC are the same as those of the abdominal region
Liver	Lvr	The outer boundary of the liver. The intrahepatic portal veins and hepatic arteries are included in this region
Fat	Fat	Grouping of SAT and VAT
Subcutaneous adipose tissue	SAT	Adipose tissue in the subcutaneous region in the abdomen
Visceral adipose tissue	VAT	Adipose tissue internal to the abdominal musculature
Neck objects		
Head and Neck skin	NSkn	The outer boundary of the head and neck skin, where the interior region constitutes the entire head and neck body region. The superior boundary is defined by a level 6.6 mm above the superior aspect of the globes. The inferior boundary is defined by a level 6.6 mm inferior to the inferior aspect of the mandible
Air and Bone	A&B	Grouping of Mnd and Phrnx
Mandible	Mnd	The outer boundary of the mandible
Pharynx	Phrx	Grouping of NP and OP
Nasopharyngeal airway	NP	The outer contour of the nasal and nasopharyngeal air cavity, extending to the inferior aspect of the soft palate
Oropharyngeal airway	OP	The outer contour of the oropharyngeal air cavities, extending from the inferior aspect of the soft palate to the superior aspect of the epiglottis
Fat pad	FP	The outer boundary of the parapharyngeal fat pad
Neck soft tissues	NSTs	Grouping of Tnsl, Tng, SP, Ad
Palatine tonsils	Tnsl	Grouping of RT and LT
Right palatine tonsil	RT	The outer boundary of the right palatine tonsil
Left palatine tonsil	LT	The outer boundary of the left palatine tonsil
Tongue	Tng	The outer boundary of the tongue
Soft palate	SP	The outer boundary of the soft palate
Adenoid tissue	Ad	The outer boundary of the adenoid tissue

not considered in this paper other important and challenging sparse objects such as the adrenal glands, pancreas, and the spinal cord. If recognition is inadequate, delineation will become unacceptable because it becomes impossible to appropriately initialize the delineation process and to exploit the model for making up for missing boundary information in the image in delineation. When we closely examined these cases, it became clear that there are fundamental challenges in the model building stage itself for sparse objects. Generally we found that sparse objects have much greater variation than their non-sparse counterparts in form, topology, and geographic layout, compared to their size. As an example, consider AS and VS (Thorax). The descending aortic portion of AS is often straight and directed vertically downward while in some subjects it may be inclined, curved, or even tortuous, with other portions, especially the aortic arch, not varying much. The branching pattern of the left and right brachiocephalic veins and the course of the azygos vein in VS also vary considerably. In view of such difficulties, we have come to the realization that sparse objects should not be modeled directly from their precise shape information in the binary image set $\mathcal{I}^b$, instead only their rough super form (such as a minimal super set that subsumes such variations) should be utilized in model building. We are exploring the use of rough sets (Maji and Pal, 2012) for this purpose.

The AAR methodology seems to have definite computational advantages over atlas-based approaches. Further, in atlas-based methods, it is perhaps much more challenging to incorporate the extensive object-level knowledge that the AAR approach exploits at various stages for recognition and delineation. These incorporations constitute highly non-linear and discontinuous phenomena which are effected in intensity, geometric, and topological spaces. The kinematic model employed in (Baiker et al., 2010) is a good analogy of how one may encode object relationships via a model that are difficult to emulate through continuous and smooth image/atlas deformations.

Some of the avenues we are currently exploring for the proposed AAR approach are delineated below.

In this paper, we did not address the problem of automatically determining the body region $\mathcal{B}$ following the definition of $\mathcal{B}$ within the given data set. As demonstrated in (Chen et al., 2012), it is possible to determine the slices delimiting a body region $\mathcal{B}$ automatically based on slice profiles. Furthermore, the information about the relationship between $\mathcal{B}$ and *WB* can also be encoded into the hierarchy as illustrated in Fig. 2(a) for each $\mathcal{B}$.

The use of composite objects often leads to better recognition accuracy. This is because the multiple objects contained in a composite object offer tighter constraints in recognition search. The aspect of how objects can be grouped to achieve optimum recognition results needs investigation. A related topic is how to device optimal hierarchies for a given body region. The hierarchies we have considered so far are anatomically motivated. Perhaps there are "optimal" hierarchies from the view point of achieving the best recognition (and hence, delineation) results. In such an investigation, matters of how objects should be grouped as well as ordered in the hierarchy can both be addressed simultaneously using graph optimization techniques.

We have set up the *AAR-R* and *AAR-D* procedures in a general way. Recognition and delineation algorithms other than those we have tested can be used independently for *R-ROOT* and *R-OBJECT* and for *D-ROOT* and *D-OBJECT* within the same hierarchical set up. Similar to composite object recognition, delineation done simultaneously for multiple objects, unlike the one-object-at-a-time approach of *AAR-D*, may improve overall accuracy.

Computationally, there are three expensive operations in the AAR system – image interpolation, distance transform, and the delineation algorithm (*FMIRFC*). To make recognition and delineation operate in practical time in a clinical setting, implementations of these operations will have to be sped up. Toward this goal, we are studying GPU implementations of these operations. GPU implementations of some fuzzy connectedness algorithms have already been published (Zhuge et al., 2011, 2013).

Finally, along the lines of the study underlying *DS*4, we are exploring the adaptation of the AAR system to several clinical applications.

Acknowledgements

The research reported in this paper is partly supported by a DHHS Grant HL 105212. Data set *DS*3 was acquired under support from an NIH Grant R01-HD-53693-01A2. A.X. Falcao was supported by a Brazilian Grant CNPq Proc. 303673/2010-9.

Appendix A

See Table A1.

References

Arens, R., Sin, S., Nandalike, K., Rieder, J., Khan, U.I., Freeman, K., Wylie-Rosett, J., Lipton, M.L., Wootton, D.M., McDonough, J.M., Shifteh, K., 2011. Upper airway structure and body fat composition in obese children with obstructive sleep apnea syndrome. Am. J. Respir. Crit. Care Med. 183, 782–787.

Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. Neuroimage 45, 333–341.

Baiker, M., Milles, J., Dijkstra, J., Henning, T.D., Weber, A.W., Que, I., Kaijzel, E.L., Lowik, C.W.G.M., Reiber, J.H.C., Lelieveldt, B.P.F., 2010. Atlas-based whole-body segmentation of mice from low-contrast Micro-CT data. Med. Image Anal. 14, 723–737.

Beucher, S., 1992. The watershed transformation applied to image segmentation. In: 10th Pfefferkorn Conference on Signal and Image Processing in, Microscopy and Microanalysis. Scanning Microscopy International, Cambridge, UK, pp. 299–314.

Bogovic, J.A., Prince, J.L., Bazin, P.L., 2013. A multiple object geometric deformable model for image segmentation. Comput. Vis. Image Underst. 117, 145–157.

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23, 1222–1239.

Cabezas, M., Oliver, A., Llado, X., Freixenet, J., Cuadra, M.B., 2011. A review of atlas-based segmentation for magnetic resonance brain images. Comput. Methods Programs Biomed. 104, 158–177.

Cerrolaza, J.J., Villanueva, A., Cabeza, R., 2012. Hierarchical statistical shape models of multiobject anatomical structures: application to brain MRI. IEEE Trans. Med. Imaging 31, 713–724.

Chen, X.J., Bagci, U., 2011. 3D automatic anatomy segmentation based on iterative graph-cut-ASM. Med. Phys. 38, 4610–4622.

Chen, X.J., Udupa, J.K., Bagci, U., Zhuge, Y., Yao, J., 2012. Medical image segmentation by combining graph cut and oriented active appearance models. IEEE Trans. Image Process. 21 (4), 2035–2046.

Chu, C., Oda, M., Kitasaka, T., Misawa, K., Fujiwara, M., Hayashi, Y., Wolz, R., Rueckert, D., Mori, K., 2013. Multi-organ segmentation from 3D abdominal CT images using patient-specific weighted-probabilistic atlas. In: SPIE Medical Imaging. SPIE, pp. 86693Y-86691–86693Y-86697.

Ciesielski, K.C., Udupa, J.K., Falcao, A.X., Miranda, P.A.V., 2012. Fuzzy connectedness image segmentation in graph cut formulation: a linear-time algorithm and a comparative analysis. J. Math. Imaging Vis. 44, 375–398.

Cootes, T.F., Edwards, G.J., Taylor, C.J., 2001. Active appearance models. IEEE Trans. Pattern Anal. Mach. Intell. 23, 681–685.

Criminisi, A., Robertson, D., Konukoglu, E., Shotton, J., Pathak, S., White, S., Siddiqui, K., 2013. Regression forests for efficient anatomy detection and localization in computed tomography scans. Med. Image Anal. 17, 1293–1303.

Duta, N., Sonka, M., 1998. Segmentation and interpretation of MR brain images: an improved active shape model. IEEE Trans. Med. Imaging 17, 1049–1062.

Hansegard, J., Urheim, S., Lunde, K., Rabben, S.I., 2007. Constrained active appearance models for segmentation of triplane echocardiograms. IEEE Trans. Med. Imaging 26, 1391–1400.

Heimann, T., Meinzer, H.P., 2009. Statistical shape models for 3D medical image segmentation: a review. Med. Image Anal. 13, 543–563.

Horsfield, M.A., Bakshi, R., Rovaris, M., Rocca, M.A., Dandamudi, V.S.R., Valsasina, P., Judica, E., Lucchini, F., Guttmann, C.R.G., Sormani, M.P., Filippi, M., 2007. Incorporating domain knowledge into the fuzzy connectedness framework: application to brain lesion volume estimation in multiple sclerosis. IEEE Trans. Med. Imaging 26, 1670–1680.

Kass, M., Witkin, A., Terzopoulos, D., 1987. Snakes-active contour models. Int. J. Comput. Vis. 1, 321–331.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W., 2010. Elastix: a toolbox for intensity based medical image registration. IEEE Trans. Med. Imaging 29, 196–205.

Linguraru, M.G., Pura, J.A., Pamulapati, V., Summers, R.M., 2012. Statistical 4D graphs for multi-organ abdominal segmentation from multiphase CT. Med. Image Anal. 16, 904–914.
Liu, J.M., Udupa, J.K., 2009. Oriented active shape models. IEEE Trans. Med. Imaging 28, 571–584.
Lu, C., Zheng, Y., Birkbeck, N., Zhang, J., Kohlberger, T., Tietjen, C., Boettger, T., Duncan, J.S., Zhou, S.K., 2012. Precise segmentation of multiple organs in CT volumes using learning-based approach and information theory. Med. Image Comput. Comput. Assist. Interv. 15, 462–469.
Maji, P., Pal, S.K., 2012. Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging. John Wiley & Sons, Inc., New York.
Malladi, R., Sethian, J.A., Vemuri, B.C., 1995. Shape modeling with front propagation – a level set approach. IEEE Trans. Pattern Anal. Mach. Intell. 17, 158–175.
Maurer, C.R., Qi, R.S., Raghavan, V., 2003. A linear time algorithm for computing exact Euclidean distance transforms of binary images in arbitrary dimensions. IEEE Trans. Pattern Anal. Mach. Intell. 25, 265–270.
Meyer, C., Peters, J., Weese, J., 2011. Fully automatic segmentation of complex organ systems: example of trachea, esophagus and heart segmentation in CT images. In: SPIE Medical Imaging. SPIE, pp. 796216-796211–796216-796211.
Miranda, P.A.V., Falcao, A.X., Udupa, J.K., 2008. Clouds: a model for synergistic image segmentation. ISBI, 209–212.
Miranda, P.A.V., Falcao, A.X., Udupa, J.K., 2009. Cloud Bank: a multiple clouds model and its use in MR brain image segmentation. ISBI, 506–509.
Mumford, D., Shah, J., 1989. Optimal approximations by piecewise smooth functions and associated variational-problems. Commun. Pure Appl. Math. 42, 577–685.
Nyul, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn. Reson. Med. 42, 1072–1081.
Okada, T., Yokota, K., Hori, M., Nakamoto, M., Nakamura, H., Sato, Y., 2008. Construction of hierarchical multi-organ statistical atlases and their application to multi-organ segmentation from CT images. Med. Image Comput. Comput. Assist. Interv. 11, 502–509.
Pizer, S.M., Fletcher, P.T., Joshi, S., Thall, A., Chen, J.Z., Fridman, Y., Fritsch, D.S., Gash, A.G., Glotzer, J.M., Jiroutek, M.R., Lu, C.L., Muller, K.E., Tracton, G., Yushkevich, P., Chaney, E.L., 2003. Deformable M-reps for 3D medical image segmentation. Int. J. Comput. Vision 55, 85–106.
Raya, S.P., Udupa, J.K., 1990. Shape-based interpolation of multidimensional objects. IEEE Trans. Med. Imaging 9, 32–42.
Rousson, M., Paragios, N., 2008. Prior knowledge, level set representations and visual grouping. Int. J. Comput. Vision 76, 231–243.
Saha, P.K., Udupa, J.K., 2001. Relative fuzzy connectedness among multiple objects: theory, algorithms, and applications in image segmentation. Comput. Vis. Image Underst. 82, 42–56.
Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39, 1064–1080.
Shen, T.A., Li, H.S., Huang, X.L., 2011. Active volume models for medical image segmentation. IEEE Trans. Med. Imaging 30, 774–791.
Souza, A., Udupa, J.K., 2006. Iterative live wire and live snake: new user-steered 3D image segmentation paradigms. In: SPIE Medical Imaging. SPIE, pp. 1159–1165.
Staib, L.H., Duncan, J.S., 1992. Boundary finding with parametrically deformable models. IEEE Trans. Pattern Anal. Mach. Intell. 14, 1061–1075.
Torigian, D.A., Alavi, A., 2007. The evolving role of structural and functional imaging in assessment of age-related changes in the body. Semin. Nucl. Med. 37, 64–68.
Tsechpenakis, G., Chatzis, S.P., 2011. Deformable probability maps: probabilistic shape and appearance-based object segmentation. Comput. Vis. Image Underst. 115, 1157–1169.
Udupa, J.K., Samarasekera, S., 1996. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. Graph. Models Image Process. 58, 246–261.
van der Lijn, F., de Bruijne, M., Klein, S., den Heijer, T., Hoogendam, Y.Y., van der Lugt, A., Breteler, M.M.B., Niessen, W.J., 2012. Automated brain structure segmentation based on atlas registration and appearance models. IEEE Trans. Med. Imaging 31, 276–286.
Wagshul, M.E., Sin, S., Lipton, M.L., Shifteh, K., Arens, R., 2013. Novel retrospective, respiratory-gating method enables 3D, high resolution, dynamic imaging of the upper airway during tidal breathing. Magn. Reson. Med. http://dx.doi.org/10.1002/mrm.24608.
Zhou, Y.X., Bai, J., 2007. Atlas-based fuzzy connectedness segmentation and intensity non-uniformity correction applied to brain MRI. IEEE Trans. Biomed. Eng. 54, 122–129.
Zhou, J., Rajapakse, J.C., 2005. Segmentation of subcortical brain structures using fuzzy templates. NeuroImage 28, 915–924.
Zhou, X., Wang, S., Chen, H., Hara, T., Yokoyama, R., Kanematsu, M., Fujita, H., 2012. Automatic localization of solid organs on 3D CT images by a collaborative majority voting decision based on ensemble learning. Comput. Med. Imaging Graph. 36, 304–313.
Zhou, X., Yamaguchi, S., Zhou, X., Chen, H., Hara, T., Yokoyama, R., Kanematsu, M., Fujita, H., 2013. Automatic organ localization on 3D CT images by using majority-voting of multiple 2D detections based on local binary patterns and Haar-like features. In: SPIE Medical Imaging. SPIE, pp. 86703A-86701–86703A-86707.
Zhuge, Y., Udupa, J.K., 2009. Intensity standardization simplifies brain MR image segmentation. Comput. Vis. Image Underst. 113, 1095–1103.
Zhuge, Y., Cao, Y., Udupa, J.K., Miller, R.W., 2011. Parallel fuzzy connected image segmentation on GPU. Med. Phys. 38 (7), 4365–4371.
Zhuge, Y., Ciesielski, K.C., Udupa, J.K., Miller, R.W., 2013. GPU-based relative fuzzy connectedness image segmentation. Med. Phys. 40 (1), 011903-1–011903-10.