Int. J. Radiation Oncology Biol. Phys., Vol. 61, No. 3, pp. 954–960, 2005

0360-3016/05/$–see front matter

doi:10.1016/j.ijrobp.2004.11.014

PHYSICS CONTRIBUTION

COMPARISON OF HUMAN AND AUTOMATIC SEGMENTATIONS OF KIDNEYS FROM CT IMAGES

MANJORI RAO, M.S.,* JOSHUA STOUGH, B.S.,† YUEH-YUN CHI, M.S.,‡ KEITH MULLER, PH.D.,‡ GREGG TRACTON, B.S.,* STEPHEN M. PIZER, PH.D.,*† AND EDWARD L. CHANEY, PH.D.*

Departments of *Radiation Oncology, †Computer Science, and ‡Biostatistics, University of North Carolina, Chapel Hill, NC

Purpose: A controlled observer study was conducted to compare a method for automatic image segmentation with conventional user-guided segmentation of right and left kidneys from planning computerized tomographic (CT) images.

Methods and Materials: Deformable shape models called m-reps were used to automatically segment right and left kidneys from 12 target CT images, and the results were compared with careful manual segmentations performed by two human experts. M-rep models were trained based on manual segmentations from a collection of images that did not include the targets. Segmentation using m-reps began with interactive initialization to position the kidney model over the target kidney in the image data. Fully automatic segmentation proceeded through two stages at successively smaller spatial scales. At the first stage, a global similarity transformation of the kidney model was computed to position the model closer to the target kidney. The similarity transformation was followed by large-scale deformations based on principal geodesic analysis (PGA). During the second stage, the medial atoms comprising the m-rep model were deformed one by one. This procedure was iterated until no changes were observed. The transformations and deformations at both stages were driven by optimizing an objective function with two terms. One term penalized the currently deformed m-rep by an amount proportional to its deviation from the mean m-rep derived from PGA of the training segmentations. The second term computed a model-to-image match term based on the goodness of match of the trained intensity template for the currently deformed m-rep with the corresponding intensity data in the target image. Human and m-rep segmentations were compared using quantitative metrics provided in a toolset called Valmet. Metrics reported in this article include (*1*) percent volume overlap; (*2*) mean surface distance between two segmentations; and (*3*) maximum surface separation (Hausdorff distance).

Results: Averaged over all kidneys the mean surface separation was 0.12 cm, the mean Hausdorff distance was 0.99 cm, and the mean volume overlap for human segmentations was 88.8%. Between human and m-rep segmentations the mean surface separation was 0.18–0.19 cm, the mean Hausdorff distance was 1.14–1.25 cm, and the mean volume overlap was 82–83%.

Conclusions: Overall in this study, the best m-rep kidney segmentations were at least as good as careful manual slice-by-slice segmentations performed by two experienced humans, and the worst performance was no worse than typical segmentations from our clinical setting. The mean surface separations for human–m-rep segmentations were slightly larger than for human–human segmentations but still in the subvoxel range, and volume overlap and maximum surface separation were slightly better for human–human comparisons. These results were expected because of experimental factors that favored comparison of the human–human segmentations. In particular, m-rep agreement with humans appears to have been limited largely by fundamental differences between manual slice-by-slice and true three-dimensional segmentation, imaging artifacts, image voxel dimensions, and the use of an m-rep model that produced a smooth surface across the renal pelvis. © 2005 Elsevier Inc.

Image segmentation, Kidney, Treatment planning.

INTRODUCTION

Three-dimensional radiation treatment planning (3D RTP) systems require a user-created model of the patient to localize and display objects of interest, position the isocenters of the treatment beams, shape the radiation beams to conform to the outline of the target volume and avoid nearby sensitive tissues, incorporate tissue inhomogeneities into dose calculations, and compute volume-weighted metrics such as dose–volume histograms (DVHs) that are used for comparing competing treatment plans. The anatomic structures and tumor-related objects comprising the patient

Reprint requests to: Edward L. Chaney, Ph.D., Department of Radiation Oncology, University of North Carolina, Campus Box 7512, 101 Manning Dr, Chapel Hill, NC 27599-7512. Tel: (919) 966-0300; Fax: (919) 966-7681; E-mail: chaney@med.unc.edu

This research was supported by NCI P01 EB002779.

Received Jun 25, 2004, and in revised form Oct 25, 2004. Accepted for publication Nov 1, 2004.

model are defined by segmenting one or more volume images, usually computerized tomographic (CT) and magnetic resonance images. Due to the large number of departments practicing 3D RTP and the large number of patients undergoing 3D RTP every day, segmentation of medical images is a commonly performed clinical task that affects critical treatment decisions. It is likely that segmentation is performed more often as a clinical procedure in radiation oncology than for all the other medical specialties combined. Unfortunately current segmentation practice is inherently inefficient and expensive. Most methods in routine clinical practice are user-guided, slice-by-slice contouring tools that require well-trained users to achieve acceptable results for 3D RTP. Other flaws of current segmentation methods that tend toward suboptimal treatment planning include intra- and interuser variabilities (1–9), the lack of practical approaches that fully consider all three spatial dimensions, and the inability to deal with ambiguous surface localization.

The development of automatic three-dimensional (3D) segmentation methods is motivated by several considerations, including economic pressure to improve efficiency and contain costs and the clinical need to improve accuracy and reproducibility to steer user-directed planning decisions and inverse treatment planning algorithms consistently in the right direction. Deformable shape models are a general class that is showing great promise for automatic segmentation of normal anatomic structures. Kass *et al.* (10) first described a straightforward method based on deformable two-dimensional contours popularly known as snakes. A useful survey of snakes is found in the study by McInerney and Terzopoulos (11). Collections of articles on early deformable models can be found in the book by ter Haar Romeny (12) and in proceedings of conferences such as CVRMed '95 (13) and CVRMed-MRCAS '97 (14); the topic is also investigated in studies by Montagnat and Delingette (15), McInerney and Terzopoulos (11, 16), Jones and Metaxas (17), and Vehkomäki *et al.* (18). However, in order for classic snake-like deformable contours to be robust and reproducible in the clinical setting, the initial guesses for shape and position of the target object essentially must be equivalent hand-drawn contours. This requirement effectively precludes the possibility of replacing hand contouring with snakes. Statistically grounded deformable shape models that can be trained to capture *a priori* information about the probability distributions of target object shapes overcome many problems presented by classic snake-like methods. A special issue of the Institute of Electrical and Electronics Engineers' (IEEE) journal *Transactions on Medical Imaging* (19) on model-based analysis of medical images has a collection of articles on a number of these methods.

The more sophisticated deformable shape methods use explicit geometric models to represent object shape. Such models represent *a priori* information that can be used in a statistical framework for matching the model against a target image. For objects with predictable shapes such as normal anatomic structures, the model can be thought of as representing a shape that is typical for the target object. For example, an m-rep is a model of the mean shape that can deform, within the limits imposed by the probability distribution on target shapes, to match the shape of a corresponding object in a target image. The statistical framework for driving the deformation is reviewed briefly below and discussed in greater detail by Pizer *et al.* (20, 21), Fletcher *et al.* (22), and Lu *et al.* (23).

In this article, we discuss the results of an observer study comparing automatic and human segmentations of left and right kidneys from planning CT images. The objective was to compare m-reps against experienced humans to judge whether m-reps produce reasonable segmentations. To accomplish this, we conducted a biostatistically rigorous comparison of m-reps against two exemplars from the population of experienced humans. Kidneys were selected for this study because they are relatively unchallenging for trained humans to contour and thus an acceptable reference standard is easily defined, and because of their importance for treatment planning. They also are a challenging initial objective for automatic methods because they are located in a crowded soft-tissue environment with bony structures nearby. Segmentation was performed in this study using medial models called m-reps (20, 21). M-reps have a number of strengths that are well matched to the task of segmenting normal structures from medical images for radiotherapy treatment planning (24).

METHODS AND MATERIALS

M-reps

Detailed discussions of the structure, building, training, and deformation of m-reps can be found in articles by Pizer *et al.* (20, 21). For completeness and continuity, brief discussions relevant to the kidney m-reps used in this study are presented below.

The simplest 3D shape is a single figure without subfigures, i.e., indentations or protrusions. For this study, the combined kidney parenchyma and renal pelvis were treated as a single figure. Such an object is described using an m-rep model comprising a grid of atoms that implies a 3D surface, as shown in Fig. 1. The centers of

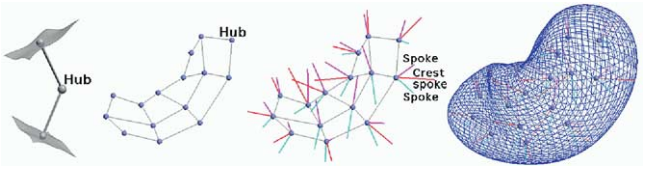

Fig. 1. Frame 1: Medial atom with two equal-length spokes that touch points on surface patches on opposite sides of the object and thus define object width at the location of the atom. Frame 2: A medial sheet of a kidney as viewed from an oblique angle. The sheet is represented as a 5 × 3 grid of medial atoms with only the atom hubs displayed. Frame 3: Medial grid with spokes displayed. Internal atoms have two spokes (magenta and cyan) and atoms on the edge of the grid have a third spoke (red) that defines the radius of curvature of the crest of the object. Frame 4: Wire-frame rendering of the surface implied by the medial sheet.