Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2018 September 04.

Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2018 February ; 10578: . doi:10.1117/12.2294042.

Hierarchical model-based object localization for auto-contouring in head and neck radiation therapy planning

Yubing Tong[1], **Jayaram K. Udupa**[1], **Xingyu Wu**[1], **Dewey Odhner**[1], **Gargi Pednekar**[2], **Charles B. Simone II**[3], **David McLaughlin**[2], **Chavanon Apinorasethkul**[4], **Geraldine Shammo**[4], **Paul James**[4], **Joseph Camaratta**[2], and **Drew A. Torigian**[1]

[1]Medical Image Processing Group, 602 Goddard building, 3710 Hamilton Walk, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104, United States

[2]Quantitative Radiology Solutions, 3624 Market Street, Suite 5E, Philadelphia, PA 19104, United States

[3]University of Maryland School of Medicine, Department of Radiation Oncology, Maryland Proton Treatment Center, 850 W. Baltimore, MD 21201, United States

[4]Radiation Oncology Department at University of Pennsylvania, Philadelphia, PA 19104, United States

Abstract

Segmentation of organs at risk (OARs) is a key step during the radiation therapy (RT) treatment planning process. Automatic anatomy recognition (AAR) is a recently developed body-wide multiple object segmentation approach, where segmentation is designed as two dichotomous steps: object recognition (or localization) and object delineation. Recognition is the high-level process of determining the whereabouts of an object, and delineation is the meticulous low-level process of precisely indicating the space occupied by an object. This study focuses on recognition.

The purpose of this paper is to introduce new features of the AAR-recognition approach (abbreviated as AAR-R from now on) of combining texture and intensity information into the recognition procedure, using the optimal spanning tree to achieve the optimal hierarchy for recognition to minimize recognition errors, and to illustrate recognition performance by using large-scale testing computed tomography (CT) data sets. The data sets pertain to 216 non-serial (planning) and 82 serial (re-planning) studies of head and neck (H&N) cancer patients undergoing radiation therapy, involving a total of ~2600 object samples. Texture property “maximum probability of occurrence” derived from the co-occurrence matrix was determined to be the best property and is utilized in conjunction with intensity properties in AAR-R. An optimal spanning tree is found in the complete graph whose nodes are individual objects, and then the tree is used as the hierarchy in recognition. Texture information combined with intensity can significantly reduce location error for gland-related objects (parotid and submandibular glands). We also report recognition results by considering image quality, which is a novel concept. AAR-R with new features achieves a location error of less than 4 mm (~1.5 voxels in our studies) for good quality images for both serial and non-serial studies.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Keywords

Automatic anatomy recognition (AAR); optimal spanning tree; texture; image quality; head and neck cancer; radiation therapy; computed tomography (CT)

1. INTRODUCTION

Radiation therapy (RT) planning is performed to reduce radiation-induced toxicities and the risk of secondary malignancies while maximizing therapeutic effects by precisely controlling how ionizing radiation will be administered to designated portions of the body [1]. Segmentation of target structures (e.g., tumor) and organs at risk (OARs) is a key step that is required during the treatment planning process. Because manual segmentation of these structures is challenging, time-consuming, and affects patient throughput in busy radiation therapy departments, developing accurate automated methods is crucial to aid pre-treatment RT planning [2].

Image segmentation mainly consists of two related steps: object recognition (or localization), the high-level process of determining the whereabouts of an object of interest in an image, and object delineation, the meticulous low-level process of precisely indicating the space occupied by an object of interest in an image. Automatic anatomy recognition (AAR) [3] is a recently developed body-wide multiple object segmentation approach, which has previously been applied to localize different objects (and lymph zones) in different body regions such as thorax [4], abdomen [5], and pelvis [6], not only on computed tomography (CT) images but also on magnetic resonance (MR) images [7] and PET/CT images. Most of previous work with AAR involved scans from near normal subjects [8].

The purpose of this paper is to introduce new features of the AAR-recognition approach. Compared with previous work on AAR in whole-body object recognition, new extensions are introduced including the following: (1) Texture properties derived from a co-occurrence matrix is employed in the AAR framework which significantly improves AAR recognition performance; (2) an optimal hierarchy for recognition is achieved by using the optimal spanning tree to minimize recognition errors; (3) illustration on a much larger scale of routine clinical RT data sets is provided, including data sets with presence of pathology and with different image qualities, involving 216 non-serial and 82 serial studies with a total of 2600 object samples (OAR samples) from head and neck (H&N) cancer patients.

2. MATERIALS AND METHODS

Image data

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) waiver. Two data sets from the Department of Radiation Oncology, University of Pennsylvania, are utilized in this study.

The first data set is from non-serial (planning CT) studies from 216 H&N cancer patients. These cases were selected from among existing patient cases. The contour data

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(segmentations of OARs) for the cases were previously created by radiation oncology dosimetrists in the process of routine clinical radiation therapy planning of these patients. The data sets constitute 54 cases gathered from each of four groups: 40–59-year-old males and females (denoted G_{M1} and G_{F1}, respectively), and 60–79-year-old males and females (denoted G_{M2} and G_{F2}, respectively). We corrected the previously drawn contours in only those cases with gross deviations from our standardized definitions of H&N OARs, and left the task of creating standardized, high-quality, model-worthy contours for future consideration.

The second data set is from serial studies from 30 patients who underwent proton beam RT (PBRT) fractionated treatment serially (re-planning CT). For each patient, we selected image data sets at 2 to 3 serial time points, accounting for a total of 82 studies. Since manual object (and tumor) contouring is impractical to perform for every serial CT scan associated with treatment fractions, ground truth segmentations drawn for clinical purposes were not available for the corresponding CT data sets. Therefore, two dosimetrists from the Penn Radiation Oncology department created manually-drawn contours on these serial studies for research purposes for the following 5 objects: Esophagus, Spinal Canal, Mandible, Orohypopharynx constrictor muscle, and Larynx. The dosimetrists were asked to record the start time and end time for each contouring session for each object.

Generally, the serial and non-serial data sets had a scene size of 512 × 512 × 110–140, and a voxel size of 0.93 × 0.93 × 1.5 mm^3 to 1.6 × 1.6 × 3 mm^3. Among the 216 image data sets for non-serial studies, 36 were selected as model-worthy, for which streak artifacts did not occur on more than 3 slices, and the other samples (> 80% of all subjects) were utilized for testing.

Objects in head and neck region

Although some object contouring guidelines available in the literature are followed by dosimetrists and oncologists, the flexibility allowed and the looseness of the definitions make the resulting contours unsuitable for building precise computational population object models. We observed this variability within the slice plane as well as in the cranio-caudal direction for long objects such as esophagus and spinal canal that cross body regions. We therefore standardized the object definitions, based on existing guidelines which are employed in RT.

The anatomic objects considered were: Skin outer Boundary (SB), Left and Right Parotid Glands and their union called Parotid Glands (LPG, RPG, PG, respectively), Left and Right Submandibular Glands and their union called Submandibular Glands (LSG, RSG, SG, respectively), Esophagus (ES), Larynx (LX), Spinal Canal (SC), Mandible (MD), and Orohypopharynx constrictor muscle (OHP). We further sub-divided object SB into an inferior portion below the neck (SBi) and a superior portion in the neck (SBs) by developing an automatic algorithm. The reason for this division was that SBs has far less subject-to-subject variation than SBi due to the different extent to which the upper extremities are included/excluded in different subjects. Figure 1 shows object samples from 5 model-worthy data sets for each of 9 objects SC, ES, LPG, RPG, LSG, RSG, LX, MD, and OHP, which are compositely surface rendered.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Fuzzy model building

The *Fuzzy Anatomy Model* of the H&N body region B for a group G, $FAM(B, G) = (H, M, \rho, \lambda, \eta)$, was built from the binary and gray scale images following mostly the methodology in [3] using the model-worthy data sets. Note that H denotes a hierarchical arrangement of the objects; M is a set of fuzzy models with one model for each object; ρ represents the parent to offspring relationship in G in the hierarchy; λ is a set of scale ranges (one for each object); η includes a host of parameters representing object properties such as the range of variation of size, image intensity, etc., of each object.

H_{opt} from training data sets

The tree structure for H is determined by finding an optimal hierarchy, H_{opt}, using a strategy which is more advanced than in our previous work. In this strategy, for each parent-child object pair (P, C), assuming that the location of P is known, C is localized in the image and the error in location is found. The mean of this error over all training images is found and used as the cost for the arc (P, C) in a complete graph involving all objects. Then, an optimal spanning tree is found in this graph, which is used as the hierarchy H_{opt}. The found optimal hierarchy for H&N is shown in Figure 2.

Recognition procedure with optimal threshold and texture

After fuzzy object models were built offline, the AAR object recognition/localization procedure (AAR-R) proceeds as follows. AAR-R finds the root object first in the hierarchy H_{opt} by calling R-ROOT. We always use SB as the root object since it is much easier to find accurately than other objects. Subsequently, AAR-R localizes the offspring objects by calling R-OBJECT in a breadth-first order in H_{opt}. R-OBJECT combines optimally the prior information in $FAM(B, G)$ with the information in the given image I in locating each object.

At the model building stage, for each object O_k, an optimal threshold interval is estimated by performing a recognition rehearsal (training) step wherein the best pose for the fuzzy model $FM(O_k)$ over all training images is determined by minimizing the sum of the false positive and false negative region volumes over all training images. Subsequently, in the actual recognition process on test image I, this learned optimal threshold is used for finding the best pose for $FM(O_k)$. Object texture properties derived from the image [9] are used for some objects instead of/in addition to image intensity. The above process for finding optimal threshold is then performed correspondingly in the image/texture image. Image and/or texture property that is best suited for each object is determined, and this information is stored in the model $FAM(B, G)$ in the element η and used at recognition. In this study, we found texture property "maximum probability of occurrence" derived from the co-occurrence matrix to be optimal for gland related objects where it made significant difference in accuracy of localization.

3. RESULTS

AAR-Recognition evaluation and image quality

Qualitative evaluation—Figure 3 shows the recognition results of the objects of interest by overlaying the recognized models onto the original images. Recognition on images with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

good quality (artifacts or major deviations in not more than 3 slices) is shown in the top row of Figure 3, and recognition on images with low quality is shown in the bottom row of Figure 3. Note how even in the presence of extreme artifacts objects are localized quite accurately.

Quantitative evaluation—Location error (LE) and scale error (SE) are used to quantitatively evaluate recognition results. Location error is the distance (in mm) of the geometric center of the object model at recognition to the known true geometric center of the object. Scale error is the ratio of the estimated object size to its true size. The ideal values for these factors are 0 mm and 1, respectively.

Recognition results were derived from experiments E1 and E2 by considering image quality as described via an image quality score (IQS) in this study. A method to express object quality score (OQS) and IQS is devised, which is reported in a separate paper submitted to this symposium (#10576-85).

Briefly, OQS is computed based on quality grades assigned to individual object regions in an image by considering several quality factors such as streak artifacts, image noise, image contrast enhancement (which means the presence of administered contrast material in the object), shape distortion, object intensity deviation, and mouth and neck posture deviation, etc. A logical predicate then generates OQS for each object in an image, and another logical predicate uses OQS of all objects in an image to generate IQS for an image. Two trained observers performed object quality grade assessment on all serial and non-serial studies.

(E1) involved application of model $FAM(B, G_M)$ to high quality data sets (objects with artifacts and other distortions in not more than 3 slices) in group G_M; (E2) is similar to (E1), but applied to low quality data sets. The objects involved in E1 were more-or-less free of streak artifacts and other major artifacts. The objects may have come from any data sets in G_M with any IQS value. Although the objects were nearly free of artifacts, they may be still affected by pathology. There were 361 object samples used in E1 from group G_M out of a total of 1140 samples (32%). The recognition on high quality images and low-quality images for serial studies is 3.84 mm vs. 14.02 mm, respectively, for location error (LE), and 1.03 vs. 1.04, respectively, for scale error (SE). Overall, the accuracy of recognition (object localization) on those high quality data sets is around 1.5 voxels (< 4 mm). Recognition on images with high quality achieved much lower location error and scale error than that on images with low quality. Table 1 summarizes some early results from this work. We get similar results on serial data set.

4. CONCLUSIONS

This paper introduces new features of AAR-recognition and illustrates object localization from AAR in head and neck cancer patients on CT images including: (1) combining texture and intensity information in AAR-recognition; (2) achieving optimal hierarchy based on the optimal spanning tree for recognition, which in turn is based on recognition error as a cost function; and (3) illustrating AAR-R for localizing objects on a large number (nearly 300) of serial and non-serial head and neck CT data sets acquired for radiation therapy along with a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

large number (2600) of object samples. This is by far the largest study on real-world clinical CT data sets, and demonstrates object localization accuracy as a function of object/image quality. AAR-R can effectively localize objects within about 1.5 voxels of true location in the presence of common pathology and when streak artifacts (such as from dental fillings) are in not more than 3 slices. Surprisingly, our recognition process is quite robust to the presence of low to moderate amount of streak artifact. Texture properties combined with intensity properties significantly improve recognition performance. Further exploration of the relationships between image quality and image segmentation will be performed in the future. This may improve our understanding and facilitate the evaluation of different object localization algorithms on large scale data sets.

Acknowledgments

This research was funded by an NSF grant IIP1549509.

References

1. Bock M, Eriksson K, Forsgren A, Hardemark B. Toward robust adaptive radiation therapy strategies. Med Phys. 2017; 44:2054–2065. [PubMed: 28317129]
2. Raudaschl PF, Zaffino P, Sharp GC, Spadea MF, Chen A, Dawant BM, et al. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. Med Phys. 2017; 44(5):2020–2036. [PubMed: 28273355]
3. Udupa JK, Odhner D, Zhao L, Tong Y, Matsumoto MM, Ciesielski KC, et al. Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images. Med Image Anal. 2014; 18(5):752–771. [PubMed: 24835182]
4. Wang HQ, Udupa JK, Odhner D, Tong YB, et al. Automatic anatomy recognition in PET/CT Images-A novel approach. Medical Physics. 2016; 43(1):613–629. [PubMed: 26745953]
5. Tong YB, Udupa JK, Odhner D, et al. Abdominal Adiposity Quantification at MRI via Fuzzy Model–Based Anatomy Recognition. Proceeding of SPIE, Medical Imaging. 2013; 8672:8672R1–7.
6. Liu Y, Udupa JK, Tong YB, et al. Definition and anatomic anatomy recognition of lymph node zones in the abdomen and pelvis on CT images. SPIE2016, Medical Imaging. 2016; 9785:9785151–97852156.
7. Tong YB, Udupa JK, Odhner D, et al. Recognition of Upper Airway and Surrounding Structures at MRI in Pediatric PCOS and OSAS. Proceeding of SPIE, Medical Imaging. 2013; 8670:86702S1–7.
8. Huang LD, Udupa JK, Odhner D, Tong YB, et al. Automatic anatomy recognition on CT images with pathology. SPIE2016, Medical Imaging. 2016; 9785:97851S1–6.
9. Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision. Thomson-Engineering; 2007.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

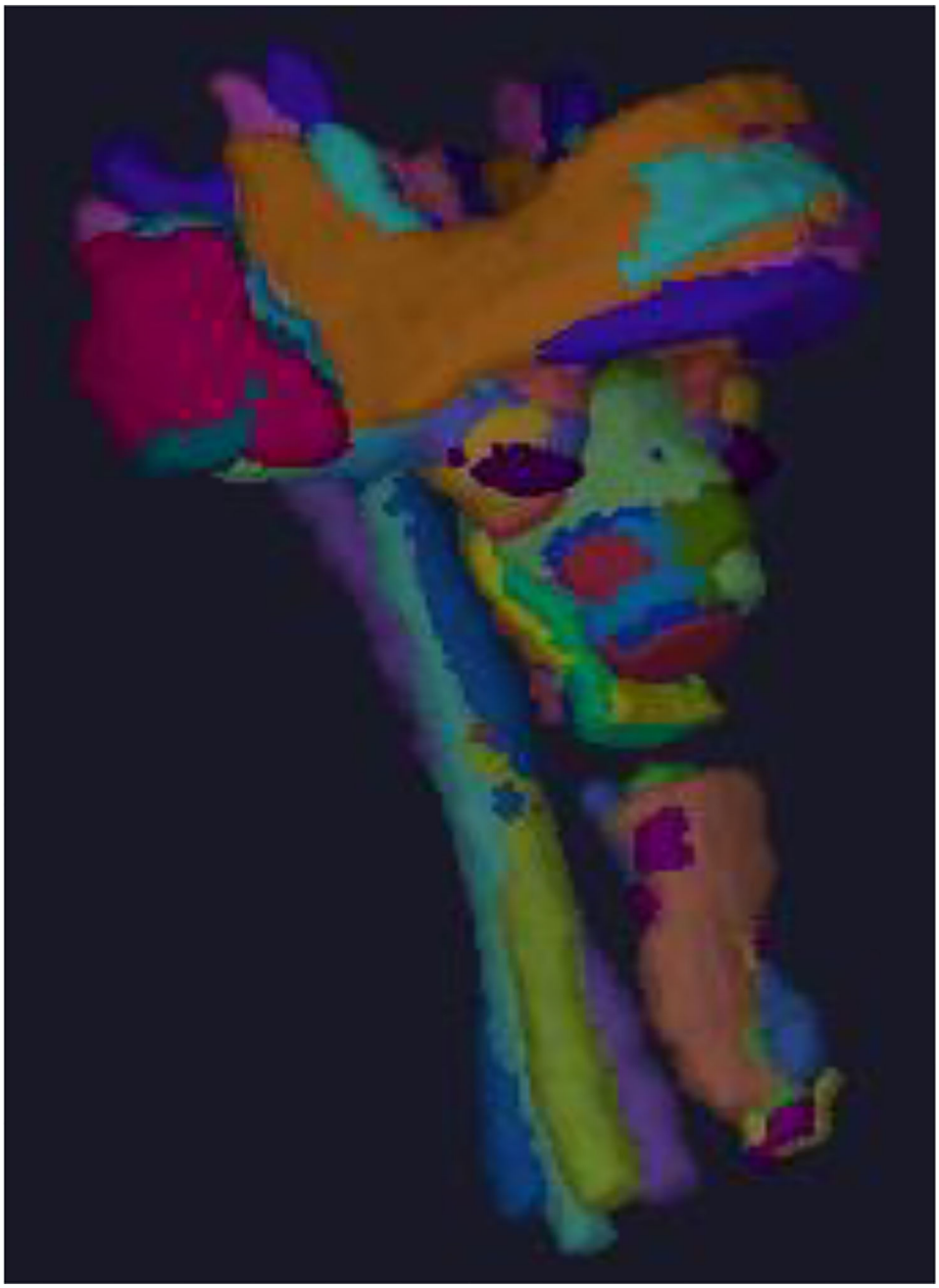

Figure 1.
5 object male samples from 5 model-worthy data sets for each of 9 objects (SC, ES, LPG, RPG, LSG, RSG, LX, MD, OHP): Different colors indicate 45 different object samples in each panel.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

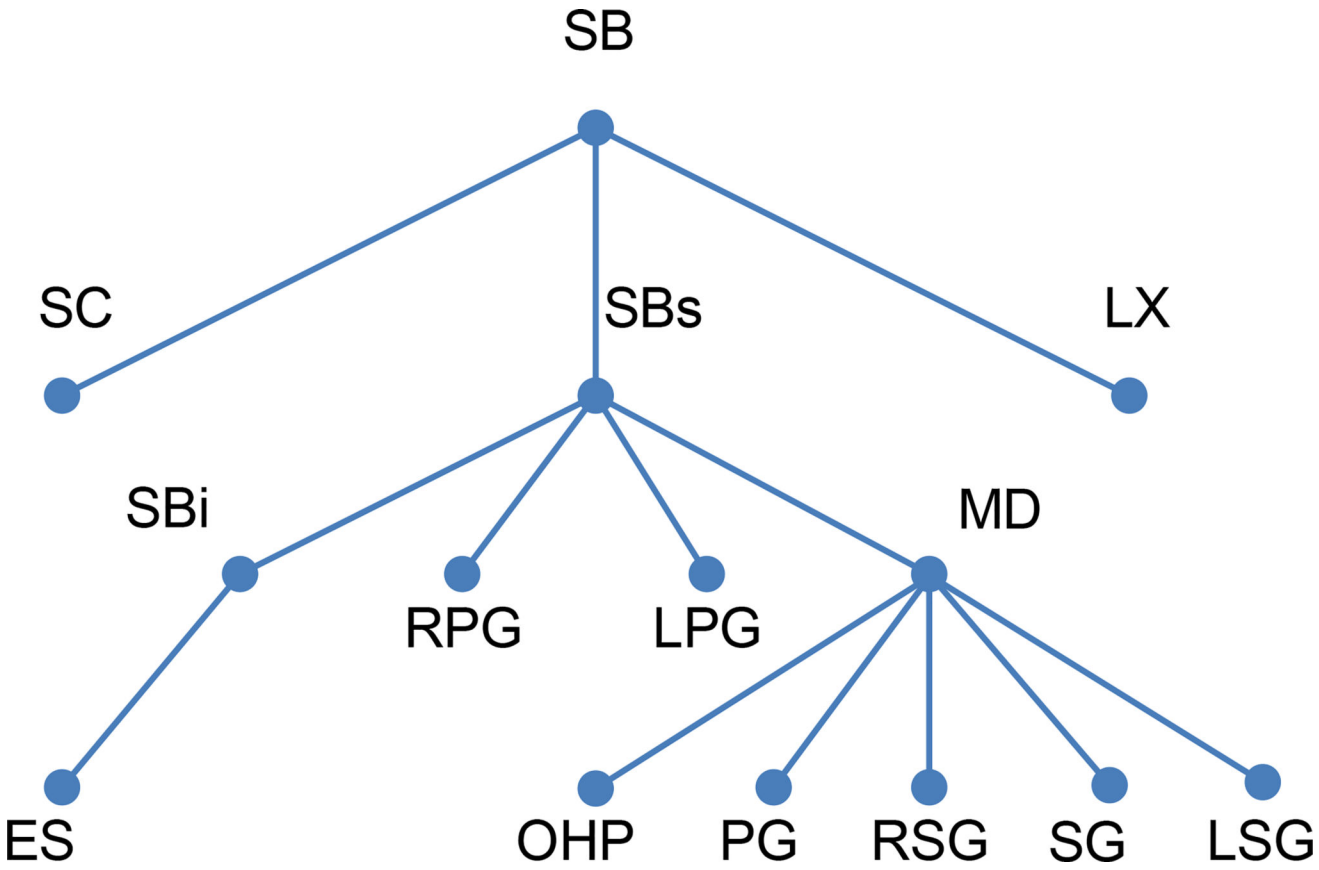

Figure 2.

Optimal hierarchy found for the H&N objects in this study. SB: Skin outer Boundary, SBs: SB superior part, Sbi: SB inferior part, SC: Spinal Canal, LX: Larynx; LPG, RPG, PG: Left & Right Parotid Glands and their union, LSG, RSG, SG: Left & Right Submandibular Glands and their union, MD: Mandible, OHP: Orohypopharynx constrictor muscle, ES: Esophagus.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

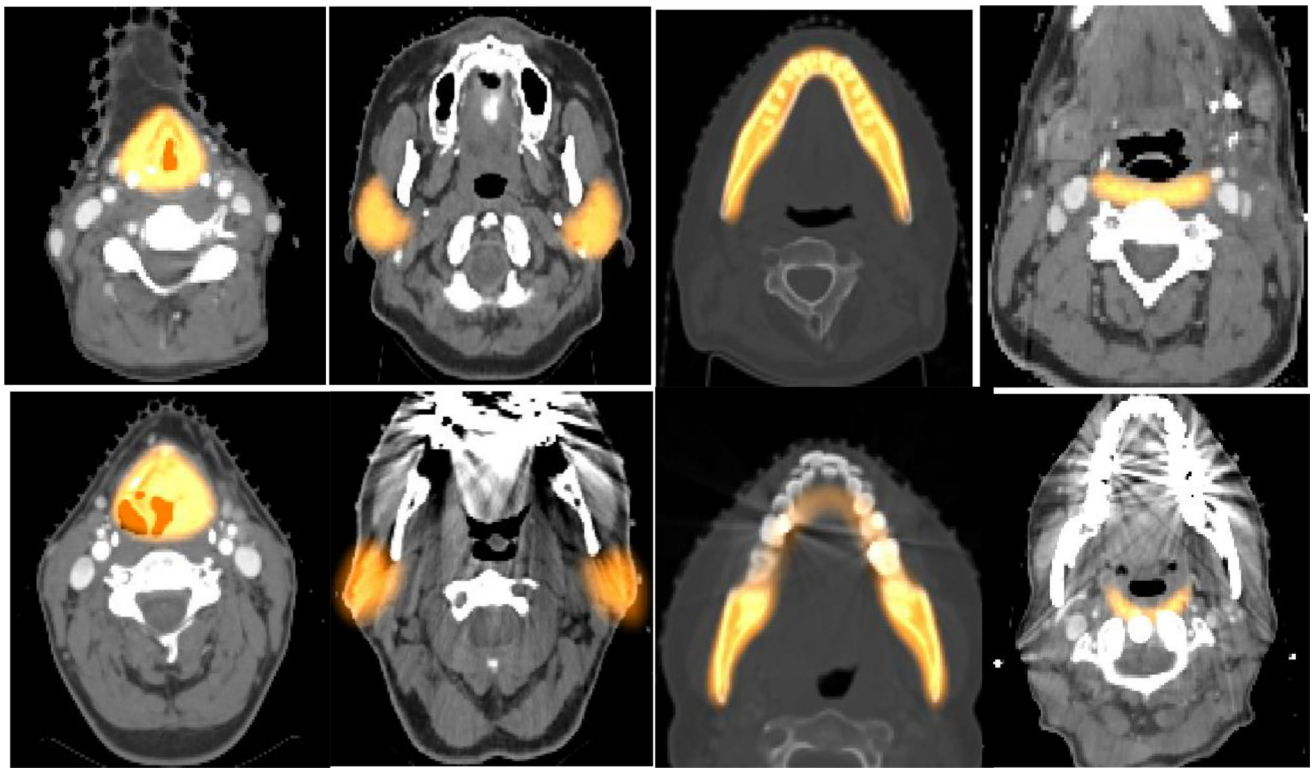

Figure 3.
Top row shows recognition results on good quality CT image data for key objects: LX, PG, MD, and OHP. Bottom row shows recognition results on low quality CT image data for the same key objects.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Location error in mm (LE) and scale error (SE) for recognition for experiments E1 and E2. Mean and SD values over tested samples for some objects are listed. Column labeled All shows mean and SD values over all objects.

		SB	SBs	LX	RPG	RSG	All
E1	LE	4.86	2.84	3.23	4.24	3.11	3.84
		0.82	1.28	2.04	1.62	1.64	1.41
	SE	1.0	0.99	1.29	1.11	0.93	1.03
		0.03	0.02	0.12	0.08	0.14	0.08
E2	LE	9.25	10.75	16.93	14.44	17.64	14.02
		3.07	4.29	8.05	5.22	7.93	6.12
	SE	1.02	0.98	1.13	1.20	0.90	1.04
		0.04	0.03	0.17	0.13	0.17	0.11

Proc SPIE Int Soc Opt Eng. Author manuscript; available in PMC 2018 September 04.