

# Smart Probabilistic Image Contouring Engine (SPICE)

K Bzdusek, Philips Healthcare, Madison, WI, USA

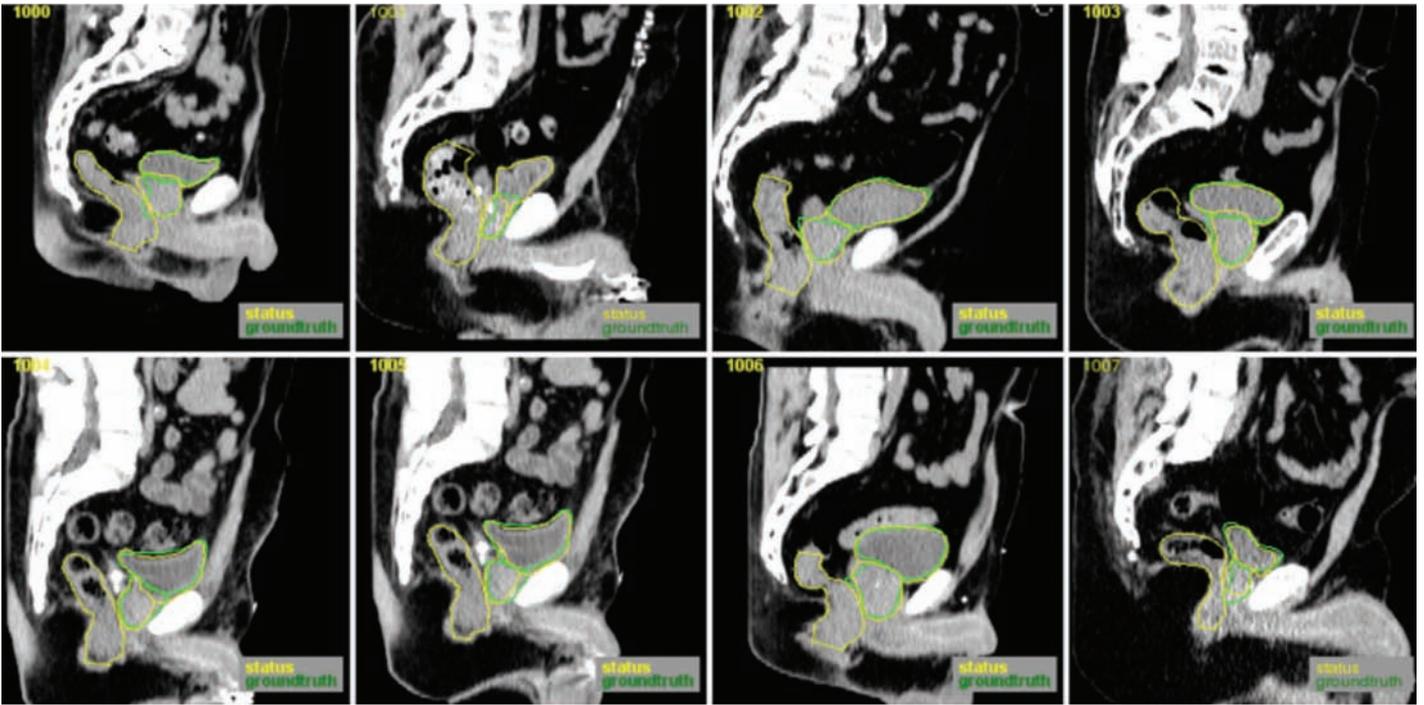
D. Bystrov, V. Pekar, J. Peters, N. Schadewaldt, H. Schulz, T. Vik, Philips Research, Hamburg, Germany

Modern radiotherapy allows for highly conformal dose distributions, especially using techniques such as intensity modulated radiation therapy (IMRT) and volumetric arc therapy (VMAT), which lead to a decrease in radiation-related toxicity. These techniques are driven by precise segmentations of organs at risk and target structures. These segmentations can be performed using manual tools but it is tedious and suffers from both interobserver and intraobserver variability. There are a number of different semi-automatic and automatic approaches that are generally either atlas-based, model-based, or hybrid. Many of these approaches are technically fully automated but the results are not always ideal, especially for tissue types that have poor contrast relative to neighboring structures. This article describes a summary of the Auto-Segmentation algorithms implemented in Pinnacle<sup>3</sup> Auto-Segmentation with SPICE. SPICE (Smart Probabilistic Image Contouring Engine) is a fully automated hybrid approach which combines several deformable registration algorithms with Model-Based Segmentation and probabilistic refinement to accurately segment normal and target tissues from head and neck, thorax, prostate, and abdominal CT images. Some results of the algorithm have been published.[1,8,9]

## Introduction

Intensity modulated radiotherapy (IMRT) and volumetric arc therapy (VMAT) have become standard treatment methods for many patients around the world. These methods offer the ability to apply high therapeutic dose to tumors while significantly reducing dose toxicities to other organs traditionally plagued by side effects.[1] An essential step in enabling these methods is to segment or delineate the organs at risk and target structures. This procedure, if done manually, is very tedious, time-consuming, and subjective, and requires a high level of concentration. A recent study documents

that an average of 2 hours 42 minutes is required to fully contour a single head-and-neck patient for IMRT or VMAT.[2] In addition to the time taken, manual contours are subjected to interobserver and intraobserver variability which may even exceed planning and setup errors.[3] Automatic segmentation aims to remove most of the manual work by providing initial structures that require minimal modifications for patient images in a consistent manner. There are many challenges to automatic segmentation including, but not limited to, poor soft tissue contrast, noise, image artifacts,



**Figure 1** Examples of typical structure variations in the male pelvis region.

contrast agents, variable field of view, organ shape, and personal preference (see **Figure 1** for examples of shape variations).

Many approaches to automated segmentation have been published.[4,5,6,7] Several methods in recent publication [1,8,9] have been combined into a single package and commercialized under the product name Pinnacle<sup>3</sup> Auto-Segmentation with SPICE. These methods consist of multiple steps generally consisting of an initial

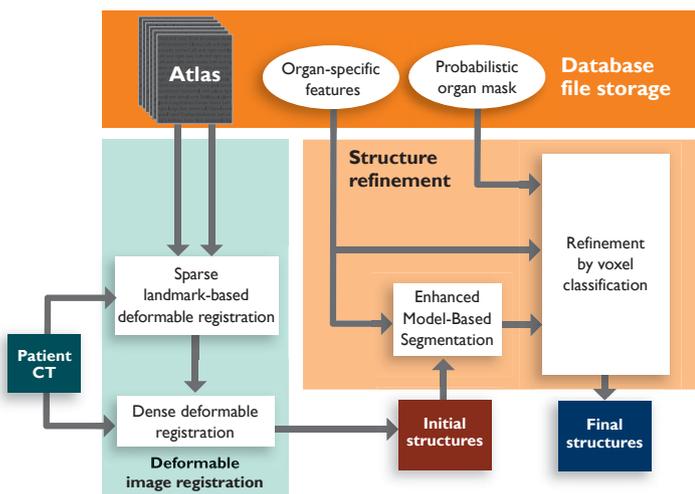
registration, dense deformable registration, and refinement. To obtain the best results, two different types of algorithm pipelines have been used – one for thorax and head-and-neck anatomies and a separate one for the abdominal and prostate anatomies. The major differentiators of this method are the high level of organ specific refinement, the lack of need to train the software as a user, and the simple integrated workflow with the planning software.

Prototype versions of the package have been tested in the previously mentioned publications and the results are summarized in this article. Similar to many recent publications,[10,11] the Dice similarity score and the mean Hausdorff distance were used for quantitative evaluation.

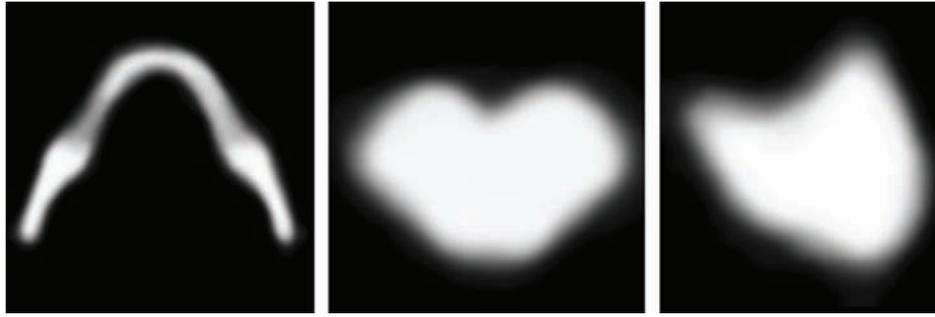
### Methods and materials

Two unique atlas segmentation pipelines were combined into SPICE. The first pipeline is customized to delineate thorax and head-and-neck patient images (see **Figure 2**). For an in-depth look at the pipeline, please refer to the publication by Qazi et al.[1]

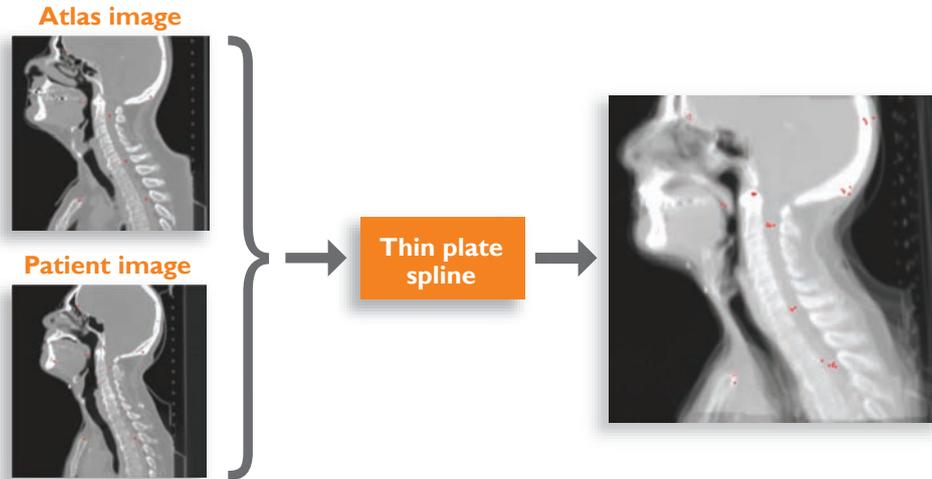
In general it consists of three main steps: a sparse landmark-based deformable image registration, a dense deformable image registration, and an individual structure refinement step (see **Figure 3**).



**Figure 2** Thorax and head-and-neck algorithm pipeline.



**Figure 3** Axial slices of the probabilistic masks for the mandible (left), brain stem (middle), and left parotid (right). Brighter areas correspond to higher probabilities.

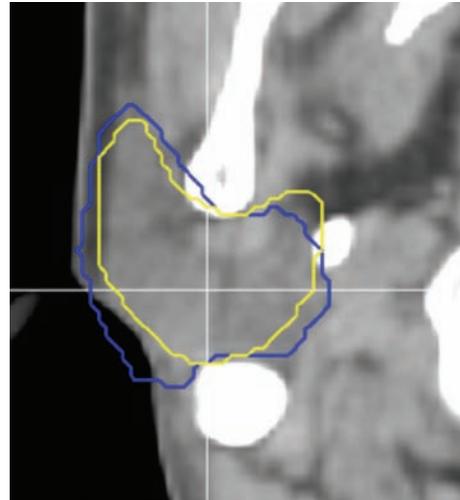


**Figure 4** A low-dimensional thin plate spline transform between the atlas and the patient images is derived from matching 14 landmark points distributed in the head and neck region.

The first step is a low-dimensional transform determined by finding 14 landmark points distributed within the head and neck region (see **Figure 4**).

These points represent the patient “pose” and compensate for global differences between the reference and the patient dataset, such as size, flexion of the neck, etc. In the second step, the initial transformation is used to initialize a dense deformable registration method (Enhanced Demons) which further initializes organ-specific deformable models.

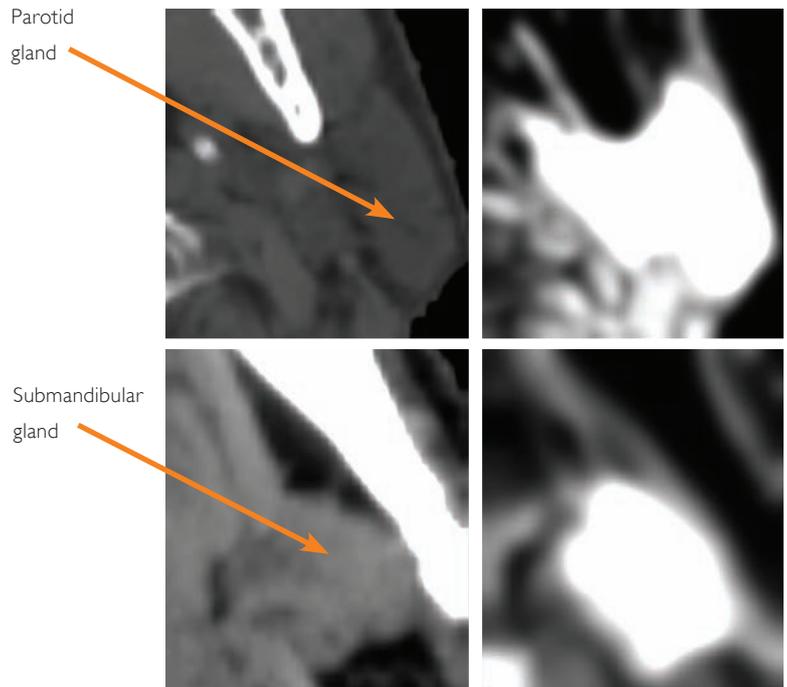
The final step is model adaptation and probabilistic refinement, the benefits of which are highlighted for a parotid gland in **Figure 5**. First, the Model-Based Segmentation (MBS) [12] is used to adapt the organ-specific deformable organ models to local image features.



**Figure 5** Example of a parotid gland segmentation result from Demons deformable image registration alone (blue) which shows obvious overlap with the mandible and patient boundary. When Demons is followed by Model-Based Segmentation probabilistic refinement (yellow), these errors are no longer present.

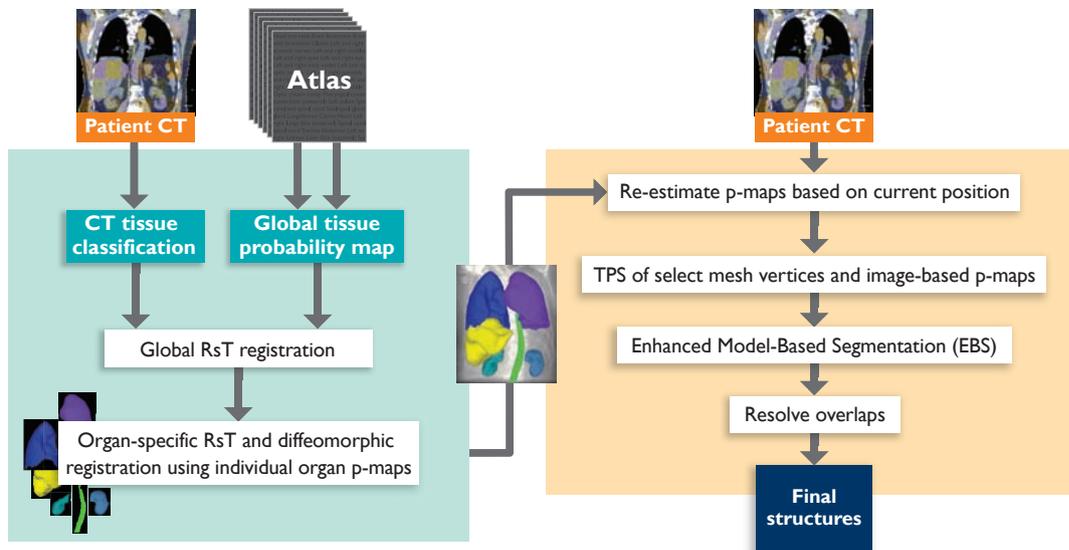
Then, depending on the structure, the models are refined or deformed within some uncertainty bounds defined by a probabilistic mask. The refinement uses the patient anatomy and voxel classification to compensate for the residual local differences (**Figure 6**).

The second pipeline is customized to delineate prostate and abdominal images. It consists of three main steps: global positioning, organ-specific positioning, and structure refinement (See **Figure 7** and **Figure 8**).



**Figure 6** Results of feature-based organ-specific image enhancement using voxel classification.

### Liver Auto-Segmentation pipeline



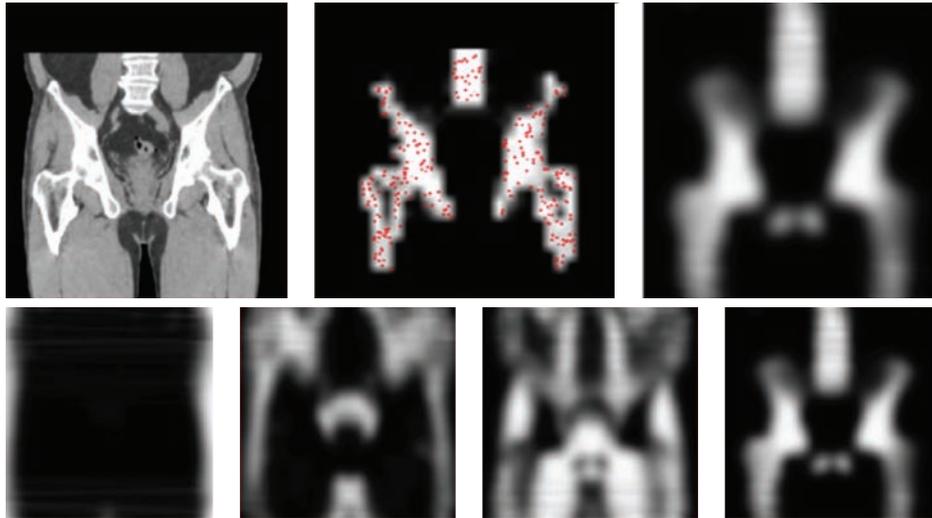
**Figure 7** Pipeline for liver Auto-Segmentation (male pelvis is similar).



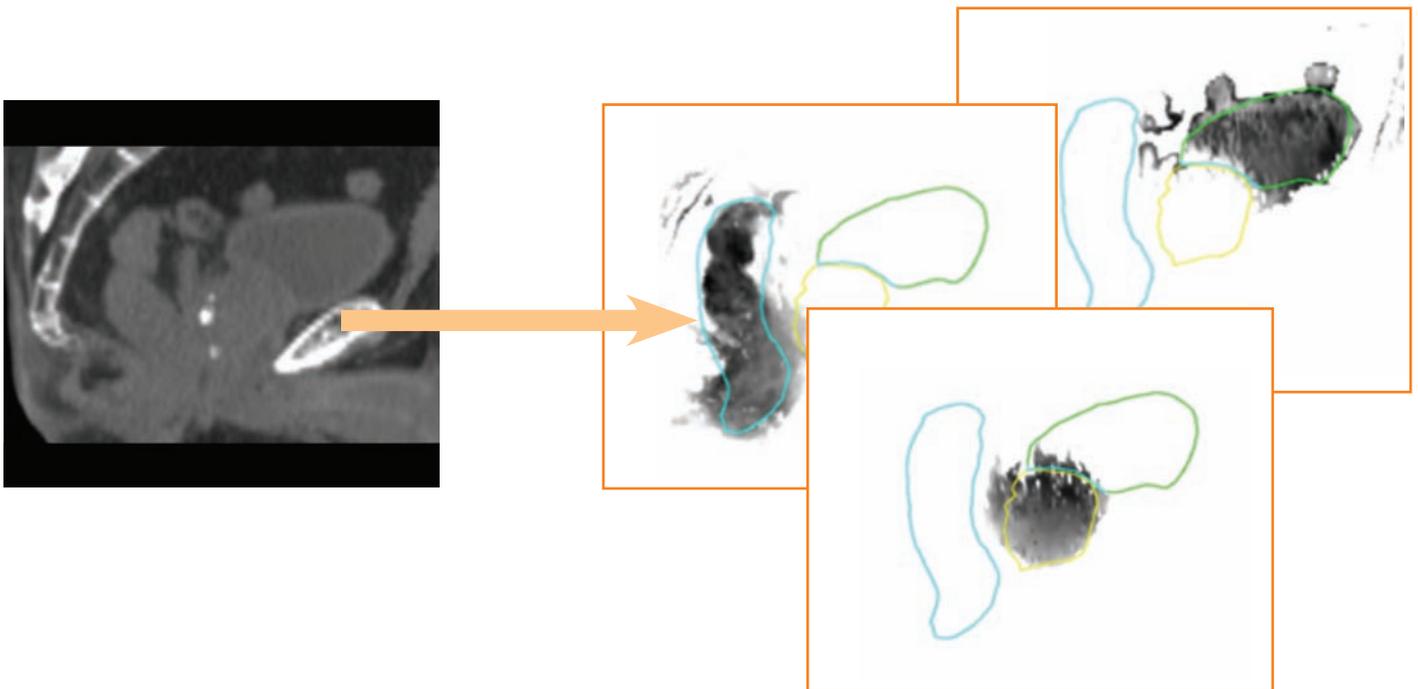
**Figure 8** Segmentation sequence from left to right:  
1. RsT, 2. Coarse deformable registration, 3. MBS.

The first step performs a rotation, translation, and isotropic scaling (RsT) registration of a tissue probability atlas to a tissue classified target image (**Figure 9**).

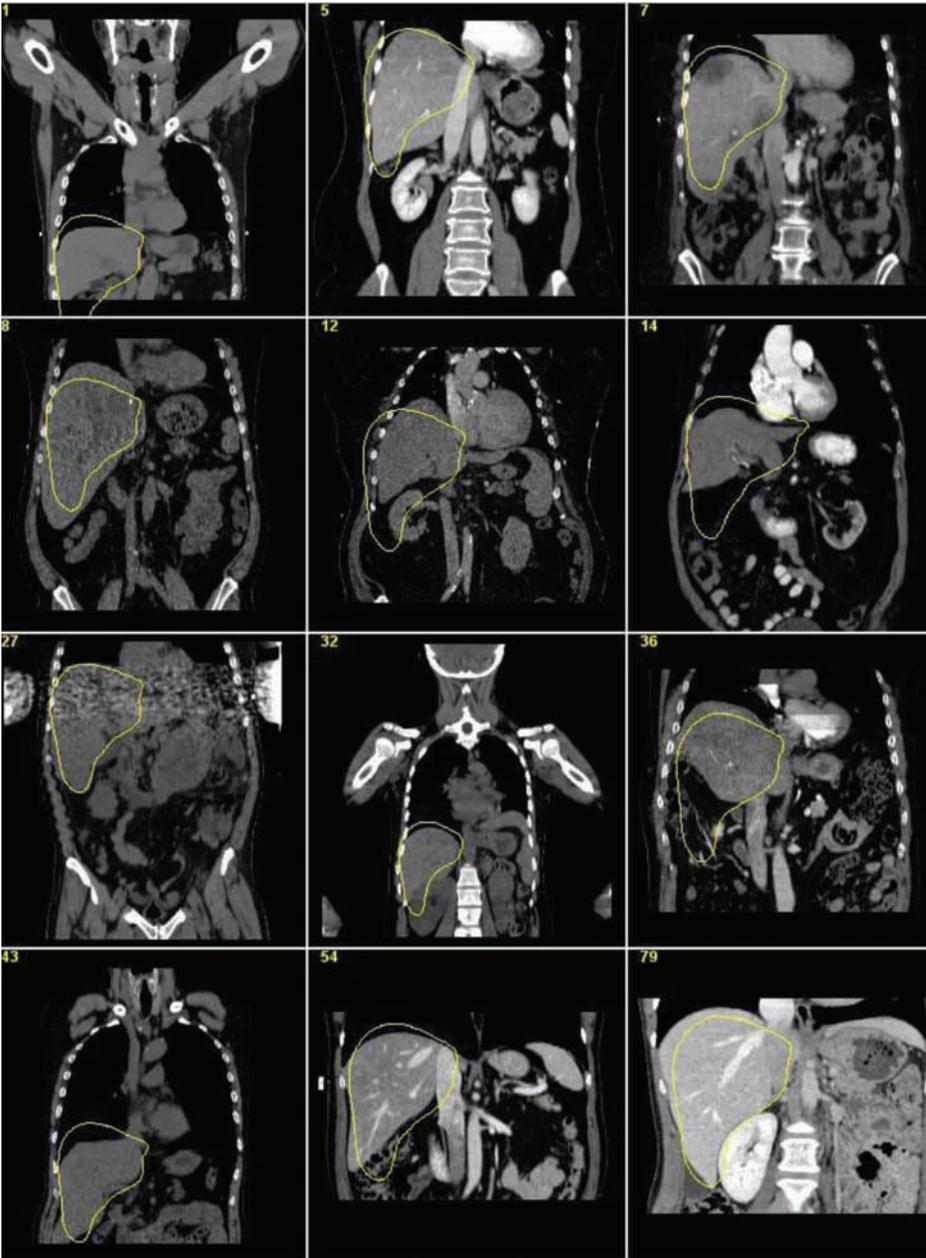
In the second step, all the organs are positioned simultaneously using the tissue classified image and organ-specific probability maps or p-maps (**Figure 10**).



**Figure 9** Upper row (left to right): (a) CT planning image  $u(x)$ , (b) derived bone probability map  $P(t = \text{bone} | u(x))$ , (c) trained bone probability atlas  $q(x | t = \text{bone})$ . Lower row (left to right): trained atlases for tissues: air, fat, water, and bone.[9]



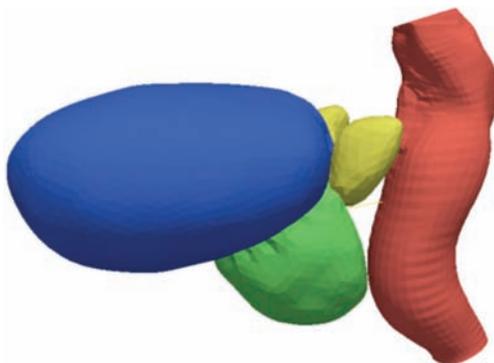
**Figure 10** Organ probability maps for rectum, bladder, and prostate based on RsT registration.[9]



Initially, an RsT transform is used followed by a diffeomorphic deformation with a small number of degrees of freedom (e.g., 8 degrees of freedom for liver). Surface mesh models of the organ are deformed to the probability map using a thin plate spline. At first, a small number of control points are used and they are iteratively increased to shrink-fit to the probability map while avoiding other organs (see **Figure 11**). The third step uses MBS to attract the surface mesh triangles to trained image features.

In some cases structures are optimized as a group (see **Figure 12**) and if there are any organ overlaps they are resolved as a post-processing step.

**Figure 11** Example of liver results: a mean surface model of the liver positioned with similarity transform in different images of varying FOV, content, and contrast from different scanners.[8]



**Figure 12** Four-organ models grouped for MBS segmentation. Combined model solves problems with local ambiguities due to poor or variable soft tissue contrast.[9]

For each treatment area, the algorithm segments a group of structures (see **Table 1**).

Because the pipelines require a large amount of organ-specific training and configuration, this process is currently unsuitable for users to add new structures or variations based on their own patients.

To accommodate common differences in the way structures are drawn, some structures have variations provided. A list of the structure variations is provided in **Table 2**.

The head and neck structures were primarily based on ground truth from Princess Margaret Hospital, Toronto, Canada. Several other variations (e.g., parotid and submandibular glands) were added later. The manual delineations of lymph node levels (I–IV) and four important organs at risk in the head and neck region, mandible, brainstem, submandibular, and parotid glands were done by an expert, following the guidelines in the publication by Grégoire et al.[13] The other variations were trained using ground truth data from the University Medical College Groningen, The Netherlands, and Odense University Hospital, Denmark. Thorax ground truths were based on structures summarized from various patients from London Regional Cancer Center, Canada; Washington University, St. Louis, USA; Odense University Hospital; University Medical Center Groningen; and University Hospital Würzburg, Germany. Learning for pelvis and liver used a much larger sample of data from many institutions and was not based on any single set of ground truth patients.

## SPICE atlases and structures

<p><b>Head and neck</b></p> <ul style="list-style-type: none"> <li>• Brain</li> <li>• Brainstem</li> <li>• Brain with brainstem</li> <li>• Glottis</li> <li>• Left and right acoustic nerves</li> <li>• Left and right cochleae</li> <li>• Left and right eyes</li> <li>• Left and right eye lens</li> <li>• Left and right neck nodes</li> <li>• Left and right optic nerves</li> <li>• Left and right parotids</li> <li>• Left and right parotids with ducts</li> <li>• Left and right submandibular glands</li> <li>• Mandible</li> <li>• Optic chiasm</li> <li>• Oral cavity</li> <li>• Pharyngeal constrictor</li> <li>• Skin (external)</li> <li>• Soft palate</li> <li>• Spinal canal and spinal cord</li> <li>• Sublingual glands</li> <li>• Thyroid gland</li> </ul>	<p><b>Lung/thorax</b></p> <ul style="list-style-type: none"> <li>• Carina</li> <li>• Heart</li> <li>• Left and right lungs</li> <li>• Skin (external)</li> <li>• Spinal canal and spinal cord</li> <li>• Trachea</li> </ul> <p><b>Abdomen</b></p> <ul style="list-style-type: none"> <li>• Left and right kidneys</li> <li>• Liver</li> <li>• Skin (external)</li> <li>• Spinal canal and spinal cord</li> </ul> <p><b>Male pelvis</b></p> <ul style="list-style-type: none"> <li>• Bladder</li> <li>• Prostate</li> <li>• Rectum</li> <li>• Left and right femoral heads</li> <li>• Seminal vesicles</li> <li>• Skin (external)</li> </ul>
---	---

**Table 1** Atlases and structures that can be tailored to suit individual operators.

Head and neck	
Brain	One brain variation includes the brainstem and one variation doesn't
Parotids	Three different sets of ground truth data were used to generate the probabilistic masks and atlas meshes
Submandibular glands	Three different sets of ground truth data were used to generate the probabilistic masks and atlas meshes
Spinal cord and canal	Spinal cord is an erosion of the spinal canal
Thorax	
Heart and heart short	Designed to match a range of ground truth data
Spinal cord and canal	Spinal cord is an erosion of the spinal canal
Prostate	
Prostate vs. prostate small	Prostate small is a 2 mm lateral erosion of the prostate
Rectum vs. rectum short	The rectum was trained based on the common extents of one set of ground truth data and the rectum short was truncated to match another set of ground truth data
Liver	
No variations	

**Table 2** Summary of structure variations by anatomy.

## Validation

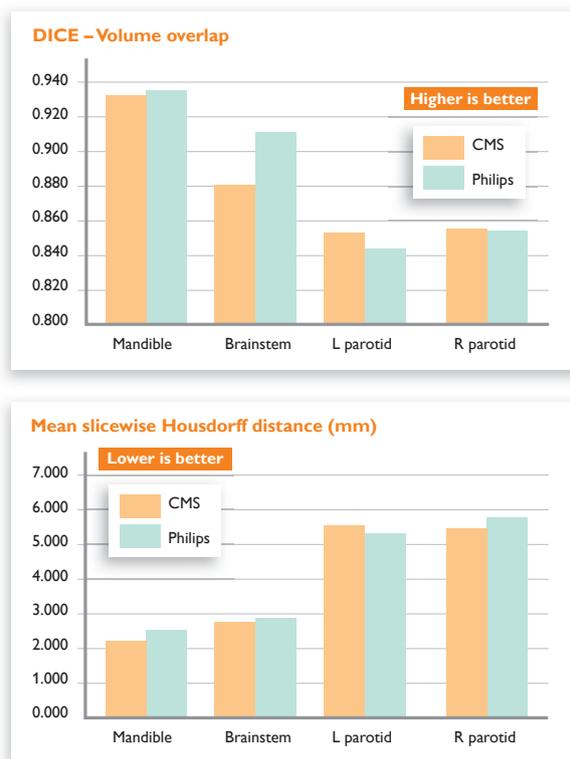
Several validation studies were published based on alpha versions of the SPICE algorithms. These publications use two common metrics used to evaluate the quality of segmentations: volume overlap fraction or the Dice similarity coefficient (DSC), [14] and a geometrical metric, the median symmetric Hausdorff distance (HD), which is evaluated slicewise. The DSC measure is defined by the following equation:

$$DSC = \frac{2 |V_{expert} \cap V_{automatic}|}{|V_{expert}| + |V_{automatic}|}$$

where  $V_{expert}$  is the expert delineation, and  $V_{automatic}$  is the result of Auto-Segmentation. The DSC measure varies between 0-1, where 0 implies no overlap and 1 represents two identical regions with perfect overlap. Statistical volumetric measures such as DSC can give

a good estimate of expert agreement; however, it is insensitive to the exact position of errors in the segmentation. Hausdorff distance, on the other hand, estimates the degree of mismatch by measuring the distance between the expert and auto-segmented contours.

The head and neck publication by Qazi et al. [1] used the same patients as two segmentation challenges [10,11] that predated it. The results from CMS, the winner of both challenges, was summarized and compared to the results in Qazi paper in **Figure 13**. The publications compared the Dice similarity score (DSC) and the mean Hausdorff distance between auto-segmented and expert drawn contours for the mandible, brainstem, and parotid glands. The comparison shows that the differences between the two algorithms for all organs are within a few percent for DSC scores and within 1 mm for Hausdorff distances.



**Figure 13** Average DICE score (left) and average mean slicewise Housdorff distance (right) for ten patients derived from literature. [1,10,11]

The male pelvis atlas was tested using the DSC metric and presented at ESTRO9. The mean DSC scores were all above 0.75 and visual inspection suggested all structures were useful except for five seminal vesicles (**Table 3**).

Organ	No. cases	Dice score
Bladder	76	0.91
Prostate	75	0.78
Rectum	16	0.89
Seminal vesicles	21	0.80

**Table 3** Summary of prostate segmentation evaluation. [9]

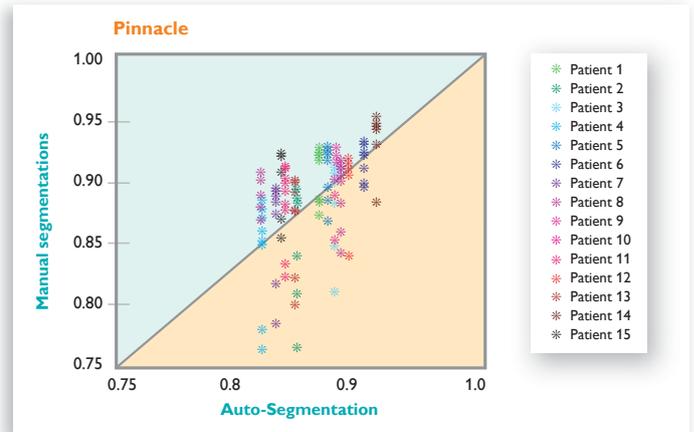


**Figure 14** Upper row: expert ground truth (each organ individually). Lower row: Automatic segmentation with combined model. Dice score: 0.95 bladder; 0.81 prostate, 0.94 rectum, 0.86 seminal vesicles.

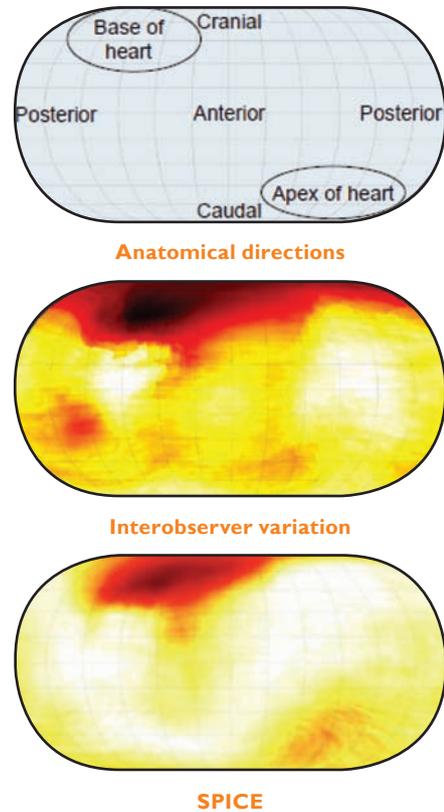
**Figure 14** shows an example of physician-drawn contours compared to the auto-segmented contours.

A multi-center analysis of heart segmentations for use in breast cancer patients was performed.[15] The heart was outlined for 15 patients by 10 observers from five centers in Denmark and the UK. A comparison of the DSC scores for automatic segmentation and manual segmentations with an average of 0.88 was obtained – as shown in **Figure 15**. The scatter around the  $y=x$  line shows that, in terms of DSC, the SPICE segmentations are indistinguishable from the group of manual segmentations.

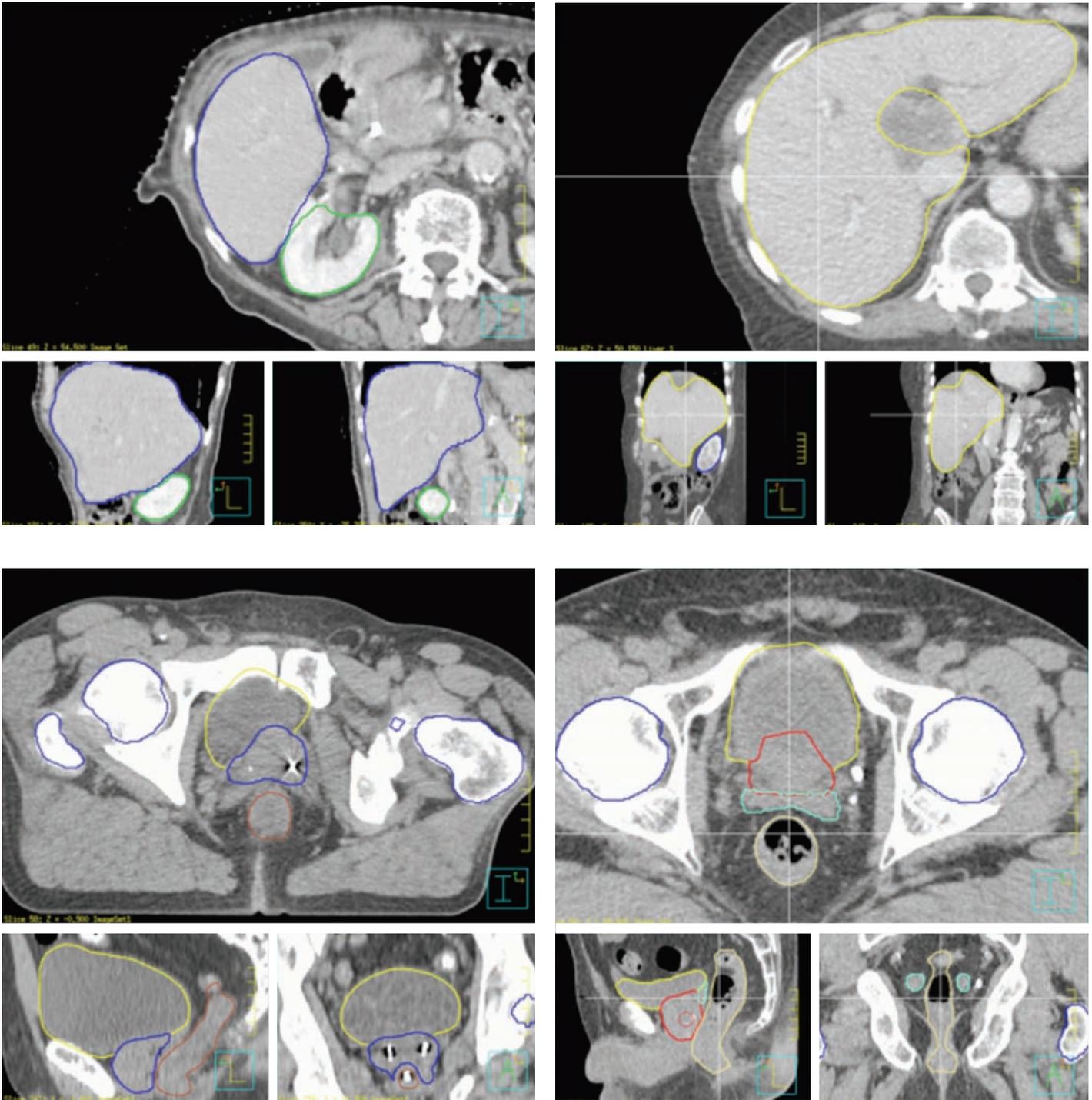
Spatial distance maps are shown in **Figure 16**. The main deviation between automatic and manual heart segmentations is in the cranial region at the base of the heart. This deviation falls within the rather large interobserver variability present in this region. The fraction of distance deviation lying within the interobserver variation was calculated with a mean value of 0.83 (range: 0.67–0.90).



**Figure 15** On the x-axis, the average Dice coefficient between the automatic and each of the 10 manual delineations are shown for each of the 15 patients. On the y-axis, the same calculation is performed by selecting one manual delineation and comparing this to the remaining delineations (manual and automatic). This is done for each manual delineation, producing 10 averaged Dice coefficient per patient. Points above the  $y=x$  line are manual delineations which, on average, agree more with the remaining delineation than the automatic delineation, and opposite for points below the  $y=x$  line.[15]



**Figure 16** Distance between median manual heart segmentation and the automatic segmentation by SPICE is shown on the bottom. Half the interdecile range (distance between 10th to 90th percentile) in manual segmentations is shown in the middle. Each of these maps are mean maps for all 15 patients. Deviation between the automatic and median manual segmentation should be compared with the variation in manual segmentations.[15]



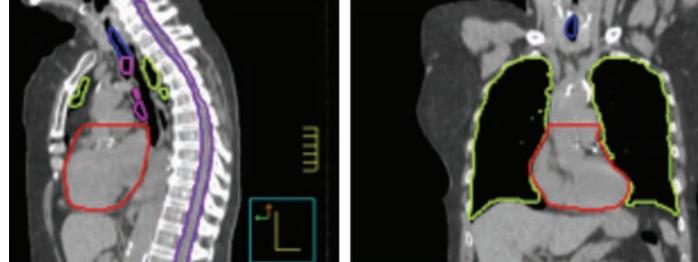
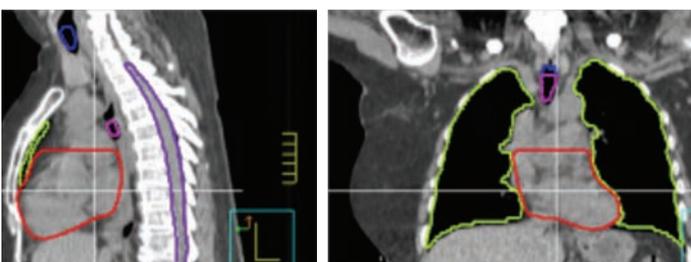
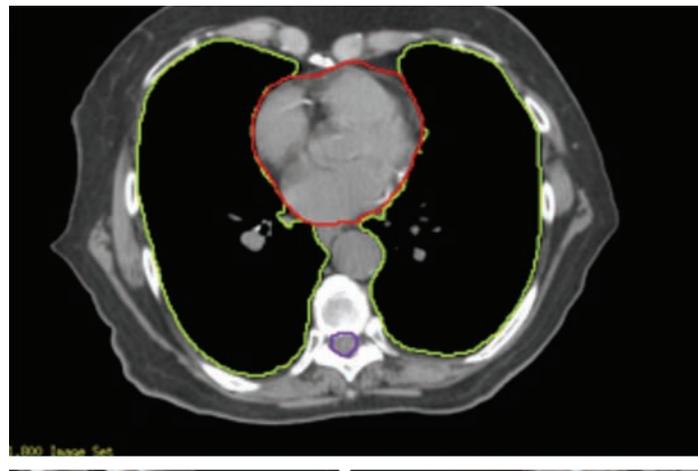
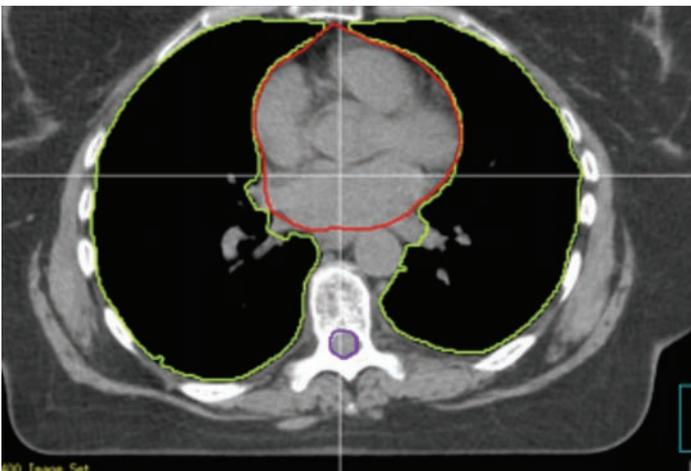
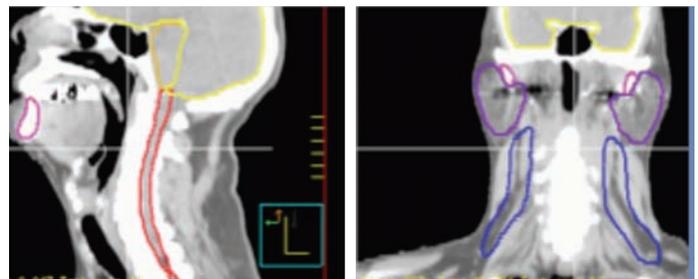
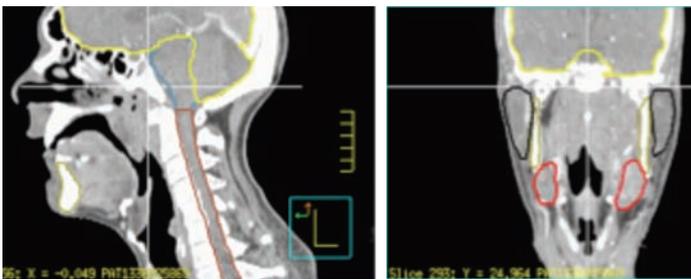
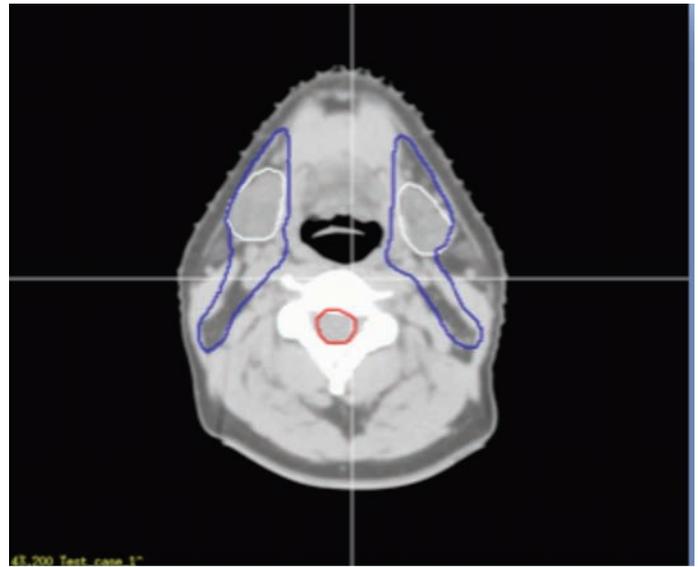
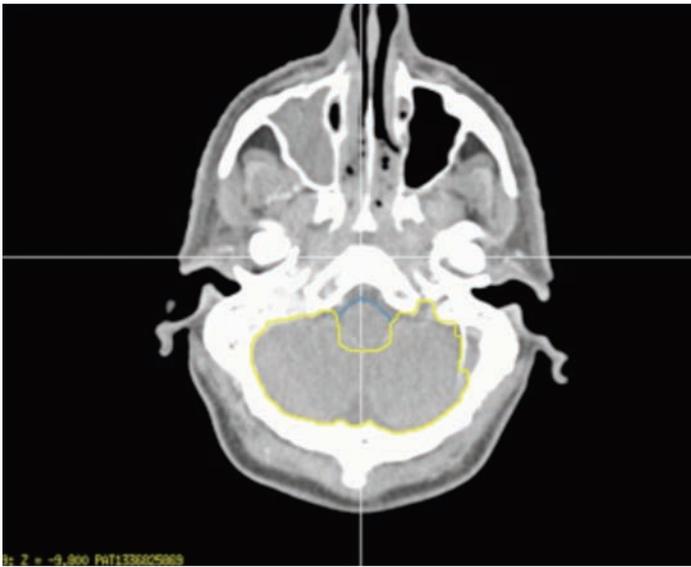
**Figures 17 – 20** Random sample of patients segmented using the SPICE algorithm for each anatomical atlas with no additional edits.

### Product highlights

A number of commercial auto segmentation solutions are available, but they usually are limited by lack of integration with the planning and simulation software, the need for tedious user atlas training, user parameter and atlas selection, lack of sophisticated refinement (direct reliance on accurate deformable image registration), etc.

SPICE avoids these limitations and provides the user with:

- Accurate, fully automatic contouring for the head and neck, thorax, male pelvis, and abdomen without requiring user algorithm training
- Simple workflow with limited user intervention
- Complete integration with the data base and planning system but independent process so the user interface is free for other work



Figures 21 – 24 Random sample of patients segmented using the SPICE algorithm for each anatomical atlas with no additional edits.

## References

1. Qazi A et al. "Auto-segmentation of normal and target structures in head and neck." *Medical Physics* 38, no. 11 (October 2011): 6160-6170.
2. Harari PM, Song S, and Tomé WA. "Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer." *Int. J. Radiat. Oncol., Biol., Phys.* 77 (2010):950-958.
3. Hess E and Weiss CF. "The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy." *Strahlenther. Onkol* 179 (2003):21-30.
4. Costa MJ, Delingette H, et al. "Automatic Segmentation of Bladder and Prostate Using Coupled 3D Deformable Models." Edited by Ourselin N, Ayache S, et al. *MICCAI (Springer)* 4791 (2007):252-260.
5. Dowling J, Frapp J, et al. "Fast Automatic Multi-atlas Segmentation of the Prostate from 3D MR Images." Edited by Dowling A, Madabhushi J, et al. *Prostate Cancer Imaging (Springer)* 6963 (2011):10-21.
6. Li W, Liao S, et al. "Learning Image Context for Segmentation of Prostate in CT-Guided Radiotherapy." Edited by G., Martel G, Fichtinger AL, et al. *MICCAI (Springer)* 6893 (2011):570-578.
7. Heimann T, van Ginneken B, et al. "Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets." *IEEE Trans. Med. Imaging* 28, no. 8 (2009):1251-1265.
8. Vik T, Bystrov D, et al. "A New Method for Robust Organ Positioning in CT Images." *ISBI (IEEE)*, 2012.
9. Bystrov D, et al. "Simultaneous Fully Automatic Segmentation of Male Pelvic Risk Structures." *ESTRO Poster*, 2012.
10. Pekar V, Allaire S, Kim JJ, and Jaffray DA. "Head and neck autosegmentation Challenge." *MIDAS Journal* 703 (2009).
11. Pekar V, Allaire S, Qazi AA, Kim JJ, and Jaffray DA. "Head and neck auto-segmentation challenge: Segmentation of the parotid glands." *Medical Image Analysis for the Clinic: A Grand Challenge (CreateSpace)*, 2010:273-280.
12. Pekar V, McNutt TR, and Kaus MR. "Automated model-based organ delineation for radiotherapy planning in prostatic region." *Int. J. Radiat. Oncol., Biol., Phys.* 60 (2004):973-980.
13. Grégoire V, Eisbruch A, Hamoir M, and Levendag P. "Proposal for the delineation of the nodal CTV in the node-positive and the post-operative neck." *Radiother. Oncol* 79 (2006):15-20.
14. Dice LR. "Measures of the amount of ecologic association between species." *Ecology* 26 (1945):297-302.
15. Lorenzen EL, and Brink C. "Automatic Segmentation of Heart Evaluated With Multi-Institution Inter-Observer Variation." *ESTRO Poster*, 2012.

Please visit [www.philips.com/SPICE](http://www.philips.com/SPICE)



© 2012 Koninklijke Philips Electronics N.V.  
All rights are reserved.

Philips Healthcare reserves the right to make changes in specifications and/or to discontinue any product at any time without notice or obligation and will not be liable for any consequences resulting from the use of this publication.

Philips Healthcare is part of Royal Philips Electronics

[www.philips.com/healthcare](http://www.philips.com/healthcare)  
[healthcare@philips.com](mailto:healthcare@philips.com)

Printed in The Netherlands  
4522 962 86221 \* JUN 2012