Hierarchical Parsing and Semantic Navigation of Full Body CT Data

Sascha Seifert[a], Adrian Barbu[b], S. Kevin Zhou[c], David Liu[c], Johannes Feulner[a], Martin Huber[a], Michael Suehling[d], Alexander Cavallaro[e] and Dorin Comaniciu[c]

[a]Integrated Data Systems, Siemens Corporate Technology, Erlangen, Germany, [b]Department of Statistics, Florida State University, Tallahassee, FL, USA, [c]Integrated Data Systems, Siemens Corporate Research, Princeton, NJ, USA, [d] Computed Tomography, Siemens Healthcare, Erlangen, Germany, [e]University Hospital, Erlangen, Germany

ABSTRACT

Whole body CT scanning is a common diagnosis technique for discovering early signs of metastasis or for differential diagnosis. Automatic parsing and segmentation of multiple organs and semantic navigation inside the body can help the clinician in efficiently obtaining accurate diagnosis. However, dealing with the large amount of data of a full body scan is challenging and techniques are needed for the fast detection and segmentation of organs, e.g., heart, liver, kidneys, bladder, prostate, and spleen, and body landmarks, e.g., bronchial bifurcation, coccyx tip, sternum, lung tips. Solving the problem becomes even more challenging if partial body scans are used, where not all organs are present. We propose a new approach to this problem, in which a network of 1D and 3D landmarks is trained to quickly parse the 3D CT data and estimate which organs and landmarks are present as well as their most probable locations and boundaries. Using this approach, the segmentation of seven organs and detection of 19 body landmarks can be obtained in about 20 seconds with state-of-the-art accuracy and has been validated on 80 CT full or partial body scans.

Keywords: organ segmentation, anatomical landmark detection, full body parsing and navigation

1. INTRODUCTION

Whole body scanning is a common practice for diagnosis of systemic diseases such as cancer and for differential diagnosis. But the large amount of data makes it difficult for the clinician to work efficiently. Thus, automatic tools are strongly needed to support the clinician in data reading and navigation. Recently, robust solutions to the heart[1] and liver[2] segmentation have been proposed. But they are organ specific and costly to transfer to new organs. In this paper we present a new framework for hierarchically parsing whole body CT volumes and efficiently segmenting multiple organs, taking contextual information into account. Our framework enables the user to segment six organs and to detect 19 landmarks very fast and robustly in about 20 seconds. New anatomy is easy to incorporate since the framework can be trained and handles the segmentation of organs and the detection of landmarks in a unified way. The detected landmarks and segmented organs facilitate the semantic navigation inside the body.

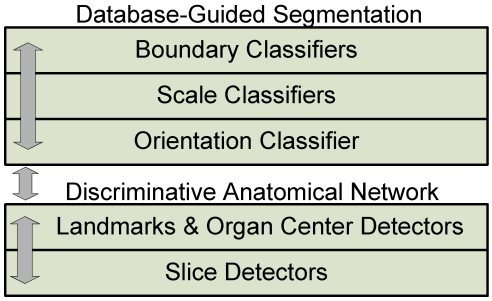


Figure 1. System Diagram.

Most approaches to segmentation of multiple structures rely on probabilistic atlases[3] that are registered non-rigidly to the current scan. Such approaches have two disadvantages. First, non-rigid registration cannot handle topological changes in the organ boundaries. Such changes occur when an organ has moved from its usual place or is greatly enlarged due to disease. Second, registration techniques are usually computationally intensive.

In this paper, we present a different approach that is based on the Marginal Space Learning[1] idea recently introduced in medical imaging. In Marginal Space Learning, the organs are detected using a sequence of learned classifiers, starting with classifiers with a few parameters (e.g., organ position without orientation and scale) and ending with a classifier that models all the desired organ parameters (e.g., position, orientation and scale). Each

learned classifier is a Probabilistic Boosting Tree[4] (PBT) with 2D[5] or 3D[6] Haar and steerable features[1] and trained by AdaBoost.[7] Other learning methods such as Probabilistic Boosting Network (PBN)[8] can be used too

The overall system architecture consists of two layers (see Fig. 1): the *Discriminative Anatomical Network* (DAN) and the database-guided segmentation module. The purpose of the DAN is to give an estimate about the scale of the patient and the portion of the volume visible. Furthermore, it detects a set of landmarks for navigating the volume. The database-guided segmentation module uses the output of the DAN for the detection of the position, the orientation, and scale of the organs visible in the given volume portion. By the use of boundary classifiers the organs are subsequently delineated.

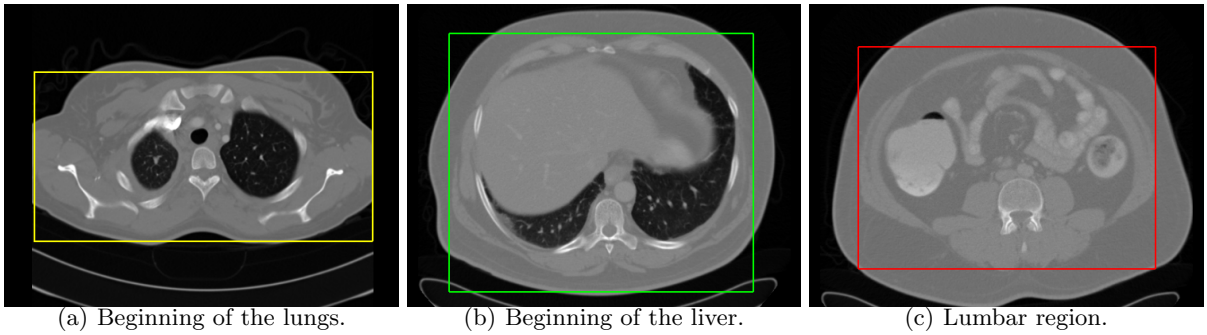
2. DISCRIMINATIVE ANATOMICAL NETWORK

The body landmarks serve three purposes. First, some landmarks have clinical importance (e.g. aortic bend) and help the clinician find valuable information about the patient. Second, some landmarks are very salient and can be robustly detected, thus help improve the system's robustness through the anatomical network. Third, the 1D landmarks / slices help speed up the detection process by serving as reference points for where to search for the other landmarks. Integrating the 1D landmarks results in a speedup of about two orders of magnitude in detecting all the landmarks and organ centers.

2.1 Slice Detectors

The CT and MR full body scans have a natural alignment of the z-axis with the body since the scans are acquired as axial cross-sections (slices). Browsing these slices, one finds that some are very salient and can be easily detected.

Based on this idea, we aligned 44 volumes at 4mm resolution along the z-axis by the z-location of the lung tips, aortic arch, liver center, right kidney center and coccyx. Then we generated 100 slices from each volume by linear interpolation between the above mentioned alignment positions. Thus for example, slice 10 corresponds in all 44 volumes to the aortic arch (see Figure 2(a)). Finally, we trained 100 slice detectors using PBT[4] and 2D Haar features,[5] with the k-th detector using slice k from all volumes as positive examples and all other slices at a distance at least 10 from slice k as negatives.



(a) Beginning of the lungs. (b) Beginning of the liver. (c) Lumbar region.

Figure 2. The three cross-sections (slices) that are used as reference for the landmark detectors.

An evaluation of the 100 detectors using 5-fold cross-validation revealed that three slices are the most salient: slice 4 - the beginning of the lungs, slice 44 - the beginning of the liver and slice 93 - in the lumbar region. Examples of these slices are showed in Figure 2. In each slice, the bounding box around the body is found by simple thresholding and pixel-counting operations. We retrained these three slices, denoted by z_0, z_1, z_2, with all the available data from 110 manually annotated volumes. The learned slice detector probabilities are denoted by $p(Z_i = z_i|I)$. The three slice detectors are connected into an anatomical network which ensures that the relative position of the slices is correct. Basically this network is a Markov Random Field (MRF) with the energy:

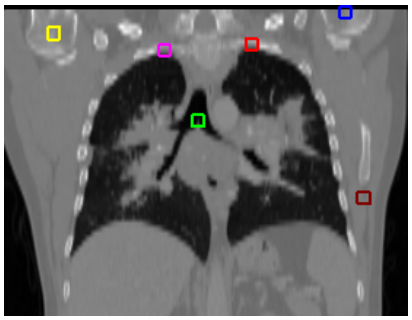
$$E(z_0, z_1, z_2) = \begin{cases} \infty & \text{if } z_1 < z_2 + D_0 \text{ or } z_0 < z_1 + D_0 \text{ or if } z_1 = -1 \text{ and } z_0, z_2 \geq 0 \\ -\sum_{i=0}^{2} \log p(Z_i = z_i|I) + (z_0 - z_1 - \mu_{01})^2/\sigma_{01}^2 + (z_1 - z_2 - \mu_{12})^2/\sigma_{12}^2 & \text{otherwise.} \end{cases}$$

For inference, slices z_0 and z_1 are restricted to have two values: -1 (when the slice does not exist) and the location that maximizes the probability $p(Z_i = z_i|I)$. Then the last slice location z_2 is found to minimize the energy $E(z_0, z_1, z_2)$. Using this inference algorithm, the DAN is an Active Random Field.[9] Training the MRF parameters is an optimization procedure based on simulated annealing to minimize the misclassification error. In 10000 steps, the parameters were obtained with a misclassification error of 0.91%.

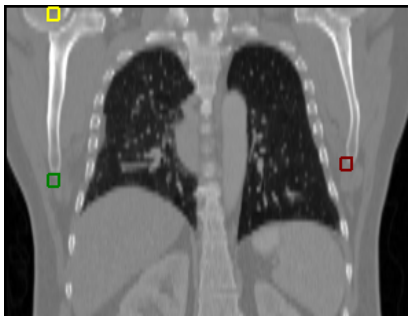
2.2 Landmark Detectors

The landmarks and organ centers are 3D points and treated the same way in our system. Below we simply call them landmarks. The landmark detectors are based on PBT and 3D Haar features.[6] Each landmark is connected to one of the three detected slices. If the slice is reported as present, the detector is started with a search range relative to the slice location. The search ranges are typically 300 times smaller than the full body, hence a speedup of two orders of magnitude is obtained.

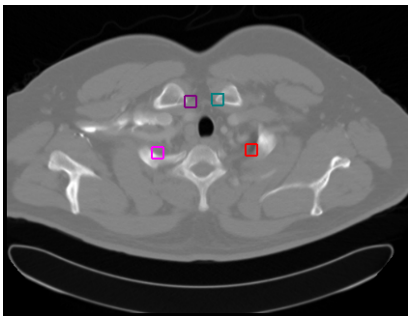
The following 19 landmarks have been trained in our system (see Fig. 3 and Fig. 4): left and right lung tips, left and right humerus heads, bronchial bifurcation, left and right shoulder blade tips, inner left and right clavicle tips, sternum tip bottom, aortic arch, left and right endpoints of rib 11, bottom front and back of the L5 vertebra, coccyx, pubic symphysis top and the left and right front corners of the hip bone.



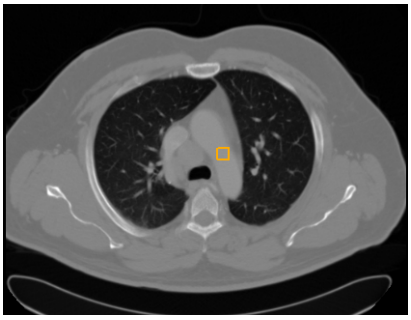
(a) Humerus heads (yellow, blue), lung tips (pink, red), bronchial bifurcation (green), left shoulder blade tip (dark red).



(b) Humerus head (yellow), shoulder blade tips (dark red, dark green).



(c) Clavicle tips (purple, turquoise), lung tips (pink, red).

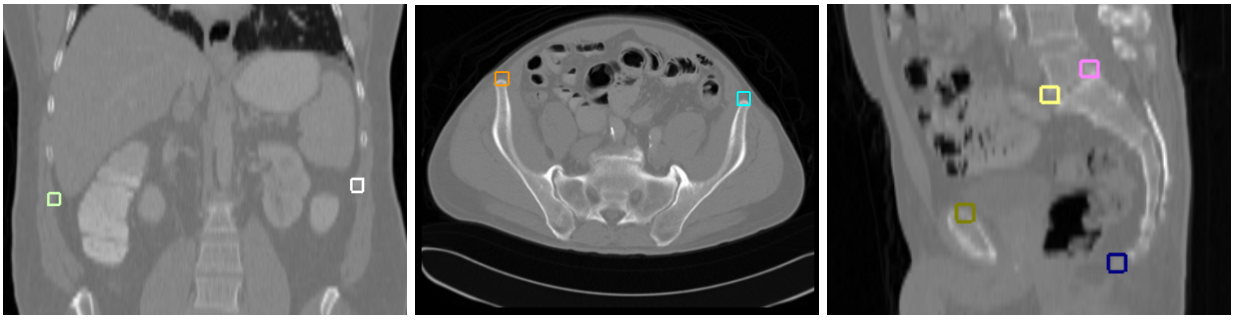


(d) Aortic arch.

(e) Aortic arch (orange), sternum tip bottom (brown).

Figure 3. Detected landmarks in the thoracic region.

To obtain a fast and robust system, the landmarks are connected in a graph (network). Information regarding the location of each landmark is propagated across the edges of the graph, which not only speeds up detection, but also increases detection accuracy. This is because the edges encode geometric relationships such as “to the right of”, “close to”, etc., and thus constrain the search problem into a smaller domain.



(a) Endpoints of rib 11.

(b) Hip bone tips front.

(c) L5 vertebra endplate border (yellow, pink), pubica symphysis (olive), coccyx (blue).

Figure 4. Detected landmarks in the abdominal region.

Denote the landmark detectors as $\phi_i(x_i|V)$, where random variables x_i, $i = 1, ..., N$, denote the unknown position, scale, and orientation parameters of the N landmarks, depending on the input volume V. The landmark detectors are incorporated into the DAN through a MRF as follows:

$$P(x_1, ..., x_N) = \frac{1}{Z} \prod_i \phi_i(x_i|V) \prod_{(i,j)\in E} \psi_{ij}(x_i, x_j)$$

where E denotes the set of edges in the network between landmarks, and $\psi_{ij}(x_i, x_j)$ expresses the pairwise geometric relationship between landmarks. Based on the Belief Propagation algorithm, the marginal distribution of x_i is $P(x_i) \propto \phi_i(x_i|y_i) \prod_{j\in E(i)} m_{ji}(x_i)$, where $E(i)$ contains the neighbors of landmark i in the network, and $m_{ji}(.)$ is computed iteratively according to the schedule:

$$m_{ij}(x_j) = \sum_{x_i} \Big(\phi_i(x_i|y_i) \psi_{ij}(x_i, x_j) \prod_{k\in E(i)\setminus j} m_{ki}(x_i) \Big)$$

3. DATABASE-GUIDED SEGMENTATION

The organ segmenters contain a trained object detector together with a shape refinement module. Currently, seven organs are implemented: heart, liver, both kidneys, spleen, bladder, and prostate (Example see Fig. 5). For improved performance the object detection works on a lower resolution optimized for the specific organ, e.g., 6mm for liver, 4mm for kidneys, 3mm for bladder. The boundary refinement is performed on the original high resolution volume to get a more accurate delineation. For organs with a large shape variability such as the liver or the spleen, the refinement is hierarchically done, i.e., the result of the first delineation on a coarse resolution is taken as an input for the refinement on the high resolution volume. Therefore a volume pyramid is built, shared by the DAN and all the organ segmenters, to manage the different resolution volumes used during the parsing.

3.1 Object Detection

The integrated object detector finds the position $x = (x_1, x_2, x_3)$, orientation $\theta = (\theta_1, \theta_2, \theta_3)$ and scale $s = (s_1, s_2, s_3)$ of the organ. The problem can be formulated as

$$(\hat{x}, \hat{\theta}, \hat{s}) = arg\ max\ p(x, \theta, s|V). \tag{1}$$

Instead of a search in a 9D space, we use the inference scheme, *Marginal Space Learning*, which reduces the size of the searching space by marginal space inference and sequentially propagates to the whole space:

$$p(x, \theta, s|V) = p(x|V)p(\theta|x, V)p(s|x, \theta, V). \tag{2}$$

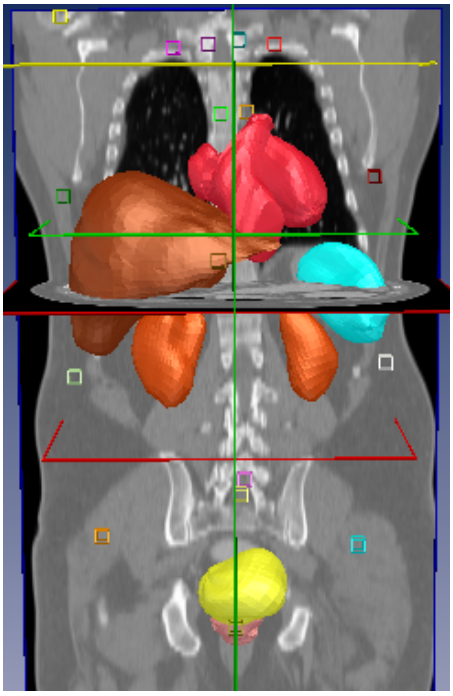


Figure 5. Whole body scan with organs and 19 body landmarks automatically obtained by our method. Segmented organs are heart (red), liver (brown), kidneys (orange), spleen (turquoise), bladder (yellow), and prostate (pink).

For an improved performance the detectors take contextual information coming from the anatomical network into account. Therefore we extend (2) with the prior probability

$$p(x|\{\phi_k\}) = \prod_k p(x|\phi_k)$$

given a set of landmark detectors $\{\phi_k\}$ specific for each organ:

$$p(x, \theta, s|V) = p(x|\{\phi_k\})p(x|V)p(\theta|x, V)p(s|x, \theta, V). \tag{3}$$

Currently, we assume a unified probability for the prior and use the following landmark-organ relationships:

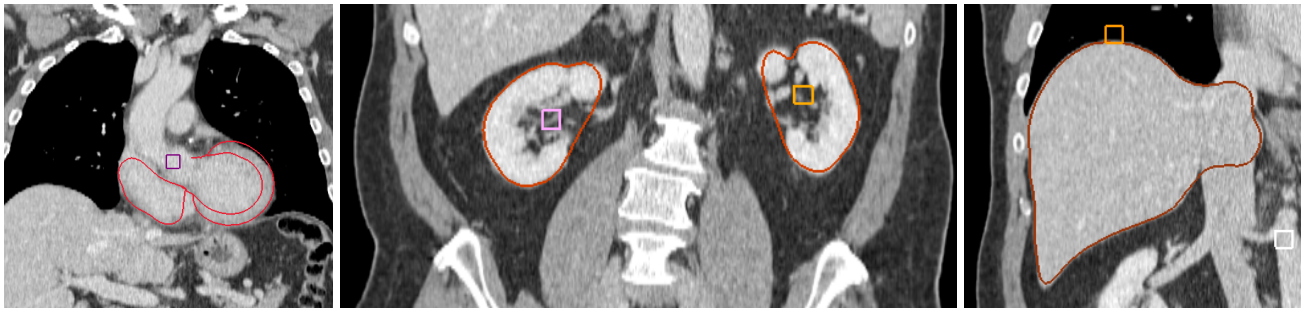
1. *Heart:* The aortic root landmark (see Fig. 6(a)) is the junction between the left ventricle and the aortic arch which is the landmark closest to the geometrical center of the heart. With 328 annotations and a median error of 3.84 mm the landmark is quite stable and reliable to serve as prior for the heart detection.

2. *Liver:* The liver dome is the maximal point directing into the right lung. The high contrast at the border of the lung facilitates its detection (see Fig. 6(c)). The median error of this landmark is 2.7mm.

3. *Kidneys:* The renal hilum of the kidney is the recessed central fissure which transmits the vessels, nerves, and ureter. It is a highly discriminative anatomical structure and with a median error of 3.5mm smaller than the kidney itself and therefore appropriate for use as prior for the kidney detection (see Fig. 6(b)).

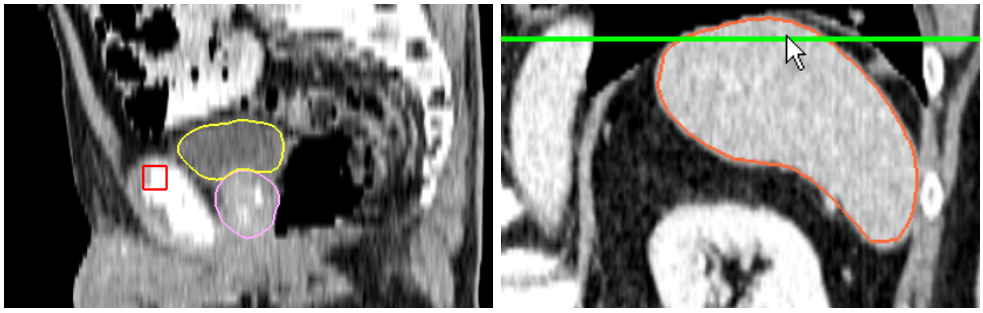
4. *Spleen:* We did not find a discriminant relevant landmark in the vicinity of the spleen, but the liver slice gives a good estimate of the upper border of the spleen (see Fig. 7(b)).

5. *Bladder* and *prostate*: The *Pubica Symphisis* is a bony, highly discriminant structure, close to the bladder and prostate, which is used by physicians as a landmark to locate both structures in CT. Hence, we use it as a prior for both organs (see Fig. 7(a)). Currently. the median error 7.1mm is quite high but sufficient as a starting point for the object detection.



(a) aortic root ↔ heart (b) kidney hilum ↔ kidney (c) liver dome ↔ liver

Figure 6. Landmark-organ relationships (HU window [-240, 320]).



(a) pubica symphisis ↔ bladder and prostate (b) liver slice ↔ spleen

Figure 7. Landmark-organ relationships (HU window [-240, 320]).

3.2 Shape Refinement

The shape refinement module consists of a trained boundary model and a PCA shape model. The multi-chamber heart segmenter and the liver segmenter have been described in previous work[1,2] and are integrated into our system. The shape model is hierarchical and is built by recursively applying a downsampling process.[2]

The refinement module is sequentially invoked for every organ and mainly differs in number of iterations and mesh pyramid levels. The shape of the organ is represented by the closed triangle mesh $\mathcal{M}_{Organ}(P,T)$, where $P = \{p_i \in \mathbb{R}^3\}_{i=1}^{n}$ is the set of n mesh points, and $T = \{t_i \in \mathbb{Z}_+^3\}_{i=1}^{m}$ is the set of m triangle indices. The meshes are built using a canonical mapping from the organ surface to a unit sphere. To describe the shape of the organs we use the parametrical formulation $S(\theta,\phi) : [0,\pi)x[-\pi/2,\pi/2] \rightarrow \mathbb{R}^3$, which allows us to sample S in the space of the spherical coordinates (θ,ϕ), which results in a dense mesh. Practically, we cut S into slices and uniformly sample along each slice to generate P. These reference meshes provide point correspondence for the training to build the mean shape $\mathcal{M}'_{Organ}$ and shape space. We get a mesh error reduction of approximately 20% on the training set, if we use a surface-to-point correspondence approach[10] on the generation of the shape model.

The hierarchical mesh refinement for each individual organ works as follows:

1. The mesh $\mathcal{M}_{Organ}$ is initialized with the mean shape $\mathcal{M}'_{Organ}$, aligned in the volume using x, θ, s from the object recognition stage. The number of mesh points is optimized with regard to the dimensions and the shape variability of the organ.

2. The boundary classifier is then used to move every point p_i of $\mathcal{M}_{Organ}$ to the optimal position where estimated boundary probability is maximized. The advantage of the learning based boundary detection is its robustness for highly heterogeneous backgrounds.

3. To avoid an unnatural shape, the adjusted mesh $\mathcal{M}_{Organ}$ is projected onto a shape subspace.[11]

4. The mesh is upsampled by the use of *Thin Plate Splines*[12] (TPS) to get a finer mesh with a doubled number of points for the next refinement iteration.

5. Step 2-4 is repeated as many times as defined for the considered organ.

6. Finally, in the last iteration, step 4 is omitted and instead, the resulting dense mesh $\mathcal{M}_{Organ}$ is smoothed by a windowed sinc function interpolation kernel.[13] The effect is to 'relax' the mesh, making the cells shaped more regular and the vertices more evenly distributed.

4. RESULTS AND CONCLUSION

Evaluation of the system is done at three levels, namely slice, landmark and organs.

At the slice level, in 586 out of 588 volumes slices were correctly reported, therefore resulting in an accuracy of 99.7%.

At the landmark level, in each of 80 evaluated volumes all landmarks were detected in all volumes, without false positive detections. The error histogram is shown in Figure 8, with an average of 7mm and 95% of errors less than 18.5mm. Landmark and slice detection together take about 400ms on a regular PC.

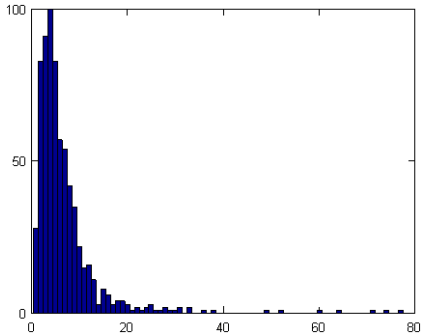


Figure 8. Histogram of landmark errors in mm.

At the organ level, the precision of the organ segmentation was evaluated with cross-validation using a mesh-to-mesh error metric (see Table 1). The heart and liver detectors and segmenters are thoroughly tested and already in use in practical applications. The training of the kidney, spleen, bladder and prostate segmenters is ongoing. Consistent with our expectations, the mesh error decreases as the number of annotations increases. In terms of detection speed, the approach using the anatomical network, yielded a speedup of about 50%, compared with one using detectors individually. Future work involves the inclusion of other organs, landmarks, bones and vessels.

Table 1. Evaluation results of the precise organ segmentation.

organs	annotations	mean mesh error [mm]	std. mesh error [mm]	detect. time [s/vol]	segm. time [s/vol]
heart	323	1.41[1]	0.45[1]	1.07	3.55
liver	226	1.26	0.53	0.67	16.0
spleen	203	2.14	1.23	1.06	0.20
kidney right	173	1.03	0.50	0.67	0.20
kidney left	163	1.15	0.35	0.67	0.20
bladder	53	1.35	0.43	1.03	0.57
prostate	40	2.55	0.77	2.00	0.59

REFERENCES

[1] Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., and Comaniciu, D., "Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning and Steerable Features," *ICCV* (2007).

[2] Ling, H., Zhou, K., Zheng, Y., Georgescu, B., Suehling, M., and Comaniciu, D., "Hierarchical, Learning-based Automatic Liver Segmentation," *CVPR* (2008).

[3] Thompson, P., MacDonald, D., Mega, M., Holmes, C., Evans, A., and Toga, A., "Detection and Mapping of Abnormal Brain Structure with a Probabilistic Atlas of Cortical Surfaces.," *Journal of Computer Assisted Tomography* **21**(4), 567 (1997).

[4] Tu, Z., "Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering," *ICCV* **3**(5) (2005).

[5] Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., and Poggio, T., "Pedestrian detection using wavelet templates," *Proc. Computer Vision and Pattern Recognition* **97**, 193–199 (1997).

[6] Tu, Z., Zhou, X., Barbu, A., Bogoni, L., and Comaniciu, D., "Probabilistic 3D Polyp Detection in CT Images: The Role of Sample Alignment," *CVPR* , 1544–1551 (2006).

[7] Schapire, R. and Singer, Y., "Improved Boosting Algorithms Using Confidence-rated Predictions," *Machine Learning* **37**(3), 297–336 (1999).

[8] Zhang, J., Zhou, S., McMillan, L., and D. Comaniciu, D., "Joint real-time object detection and pose estimation using probabilistic boosting network," in [*Proc. CVPR*], (June 2007).

[9] Barbu, A., "Training an active random field for real-time image denoising," Invited Talk, Max Planck Institute, Saarbrucken, Germany (July 16th 2008).

[10] Vos, F. M., de Bruin, P. W., Aubel, J. G. M., Streekstra, G. J., Maas, M., van Vliet, L. J., and Vossepoel, A. M., "A statistical shape model without using landmarks," in [*ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3*], 714–717, IEEE Computer Society, Washington, DC, USA (2004).

[11] Cootes, T. F., Hill, A., and Taylor, C. J., "The use of active shape models for locating structures in medical images," *Image and Vision Computing* **12**, 355–366 (1994).

[12] Bookstein, F. L., "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 567–585 (1989).

[13] Taubin, G., Zhang, T., and Golub, G., "Optimal surface smoothing as filter design," in [*Computer Vision – ECCV '96*], 283–292 (1996).