Medical Image Analysis xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Medical Image Analysis

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

Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior

Johannes Feulner [a,b,*], S. Kevin Zhou [c], Matthias Hammon [d], Joachim Hornegger [a], Dorin Comaniciu [c]

[a] Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany
[b] Siemens Corporate Technology, Erlangen, Germany
[c] Siemens Corporate Research, Princeton, NJ, USA
[d] Radiology Institute, University Hospital Erlangen, Germany

ARTICLE INFO

Article history:
Received 23 September 2011
Received in revised form 1 November 2012
Accepted 7 November 2012
Available online xxxx

Keywords:
Lymph nodes
Chest CT
Detection
Segmentation
Spatial prior

ABSTRACT

Lymph nodes have high clinical relevance and routinely need to be considered in clinical practice. Automatic detection is, however, challenging due to clutter and low contrast. In this paper, a method is presented that fully automatically detects and segments lymph nodes in 3-D computed tomography images of the chest. Lymph nodes can easily be confused with other structures, it is therefore vital to incorporate as much anatomical prior knowledge as possible in order to achieve a good detection performance. Here, a learned prior of the spatial distribution is used to model this knowledge. Different prior types with increasing complexity are proposed and compared to each other. This is combined with a powerful discriminative model that detects lymph nodes from their appearance. It first generates a number of candidates of possible lymph node center positions. Then, a segmentation method is initialized with a detected candidate. The graph cuts method is adapted to the problem of lymph nodes segmentation. We propose a setting that requires only a single positive seed and at the same time solves the small cut problem of graph cuts. Furthermore, we propose a feature set that is extracted from the segmentation. A classifier is trained on this feature set and used to reject false alarms. Cross-validation on 54 CT datasets showed that for a fixed number of four false alarms per volume image, the detection rate is well more than doubled when using the spatial prior. In total, our proposed method detects mediastinal lymph nodes with a true positive rate of 52.0% at the cost of only 3.1 false alarms per volume image and a true positive rate of 60.9% with 6.1 false alarms per volume image, which compares favorably to prior work on mediastinal lymph node detection.

1. Introduction

Lymph nodes play an important role in clinical practice, especially in the mediastinal area (the region between the lungs). They routinely need to be considered during oncological examinations related to all kinds of cancer (Duwe et al., 2005; de Langen et al., 2006), for instance lung cancer (McLoud et al., 1992), where metastases settle in lymph nodes, but also lymphoma, which is a cancer of the lymphatic system itself. Furthermore, they are also relevant in case of inflammation in general (Drake et al., 2009).

Cancer causes affected lymph nodes to be enlarged. In order to assess the progress of the disease and to check whether treatment is effective, physicians are interested in statistics like the number of enlarged nodes or the total volume of the nodes, but also in the spatial distribution, and changes over time. Patients are commonly examined using CT (Suwatanapongched and Gierada, 2006).

Manually counting and measuring lymph nodes in the images is not only cumbersome but also error prone because annotations from different human observers and even from the same human observer vary significantly. In practice, lymph nodes are not annotated individually because it would take too much time, though the clinical value would be potentially high. Automatic detection is however challenging because lymph nodes have an attenuation coefficient similar to muscles and vessels and therefore low contrast to surrounding structures. Moreover, their shape and size varies a lot. The first author needed days of training to consistently find lymph nodes in CT volume images. Examples of mediastinal lymph nodes are shown in Fig. 1.

The topic has received increasing attention in the last five years. In Kitasaka et al. (2007), two blob detectors which are called 3-D Min-DD filter and extended 3-D Min-DD filter are used in a cascade to detect lymph nodes in abdominal CT data. A Hessian based vessel detector, the CT Hounsfield units and morphological operations

* Corresponding author at: Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany.
E-mail address: johannes.feulner@informatik.uni-erlangen.de (J. Feulner).

1361-8415/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.media.2012.11.001

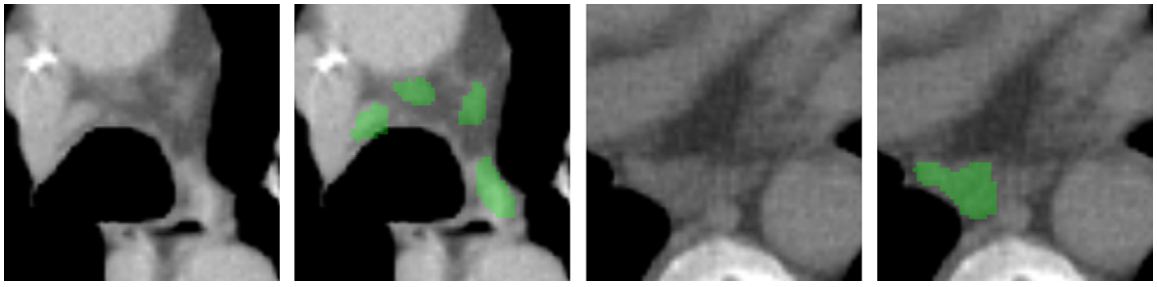

Fig. 1. Two axial cross sections of CT volumes with expert-reviewed lymph node annotations (green).

are used to reduce the number of false positives. In Feuerstein et al. (2009), a similar approach is used to detect lymph nodes in chest CT. Here, the first 3-D Min-DD filter is replaced with a Hessian based blob detector. The more expensive extended 3-D Min-DD filter is used at the second level of the cascade. For segmentation, a model based approach using mass spring models was proposed in Dornheim et al. (2006). It was also used for detection by placing models on a regular grid over the volume (Dornheim and Dornheim, 2008). A lymph node was assumed at positions where the model fitting converged with a good score. In Feulner et al. (2010, 2011) and Barbu et al. (2010), data driven approaches for lymph node detection were proposed. In all three cases, a discriminative model is trained to detect lymph nodes in CT from their appearance. In Feulner et al. (2010, 2011), the focus was on the mediastinal region, and the discriminative model was combined with prior anatomical knowledge that is modeled as a spatial prior probability. In Barbu et al. (2010), the focus was on the axillary region. Lymph nodes are also segmented by fitting a radial shape model with a Markov random field prior to the image at locations of detected lymph node centers. The result of the segmentation is used to improve the detection: A good segmentation result indicates a good detection. A similar technique is used in Feulner et al. (2011): In a final step, each detection is verified by initializing a segmentation algorithm with the detection. But instead of fitting a radial shape model, graph cuts are adapted to the problem of lymph nodes segmentation. Features are extracted from the segmentation and used to train a classifier to learn whether a segmentation is a true or a false alarm.

This paper is an extension of our prior work Feulner et al. (2011). Fig. 2 gives an overview of our system for lymph node detection and segmentation, which is based on a cascade of binary classifiers. The first two stages of the cascade use 3-D Haar like features to generate a set of candidate lymph node centers. In stage 3, the detected position candidates from stage 2 are verified by a third classifier that now uses gradient-aligned features as proposed in Barbu et al. (2010). The result is again a set of position candidates. Finally, a segmentation algorithm is initialized with the lymph node position candidates. The graph cuts method is adapted to the problem of lymph nodes segmentation by incorporating knowledge about the shape and the appearance of the node into the segmentation framework. A feature set is extracted from the segmentation, and a fourth classifier is trained to learn whether a segmentation is a true lymph node or not.

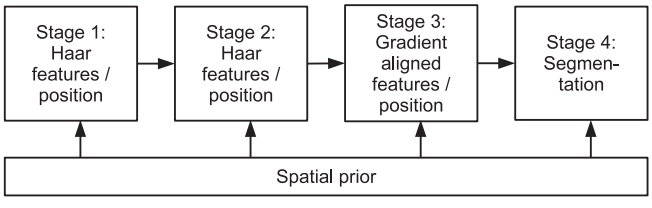

Fig. 2. Overview of the detection and segmentation system that was originally proposed in Feulner et al. (2011). Lymph nodes are detected by a cascade of binary classifiers. The first two stages detect the center of a lymph node using 3-D Haar-like features. Then, the detected positions are verified using gradient-aligned features in stage 3. In stage 4, each position candidate from stage 3 is used to initialize a segmentation algorithm, features are extracted from the segmentation, and based on these features, a classifier decides whether the detection is a true lymph node or a false alarm. At all stages, prior anatomical knowledge is included in the form of a spatial prior probability.

This work extends Feulner et al. (2011) in a number of ways. A comparison to a method was added that does not only use a single segmentation at the final verification stage 4, but instead combines hints from multiple alternative segmentations. Further additional experiments include a comparison of different types of spatial priors, and a direct comparison with prior work on mediastinal lymph node detection.

Spatial priors, or probabilistic atlases, of tissue classes are already very popular in brain imaging because of a low intra subject and also a relatively low inter subject variability of the brain in 3-D scans. Due to the rigid skull, registration techniques work generally better in the brain than in other body regions. Probabilistic brain atlases used in prior work model, however, mostly the structure of the normal brain and not common locations of tumors, or lesions in general. Examples are Moon et al. (2002), Van Leemput et al. (1999a,b), Prastawa et al. (2003), Gooya et al. (2011), and Bauer et al. (2010b). Atlases of pathologies are rarely used. In Moon et al. (2002), a normal atlas is augmented with a subject specific posterior probability map for tumor that is generated from pre and post contrast MRI. Similarly, in Prastawa et al. (2003), the tumor is first enhanced using contrast agent, and a normal atlas is then modified with the contrasted tissue. The variability in the chest is far higher than in the brain, and it is far more difficult to apply segmentation-by-registration techniques. This work uses a spatial prior of lymphatic tissue only, no prior of normal tissue is used. Instead, organs in the chest area are segmented, and the segmentation results are combined with the pathology prior. It is therefore not necessary to make the prior very accurate, nor is it necessary to very accurately register the prior to an image.

The remainder of this paper is structured as follows: Section 2 explains the different kinds of spatial priors that are used in this work to model anatomical knowledge. Section 3 describes how the region of interest containing the mediastinum is determined. Section 4 explains the first two stages of the detection cascade. In Section 5, a method is presented that segments lymph node candidates and rejects or accepts them based on the segmentation result. Section 6 presents experiments and results, and Section 7 concludes the paper.

2. Spatial prior of lymphatic tissue

Mediastinal lymph nodes are very hard to detect only from their appearance. They have a similar attenuation coefficient like muscles and vessels, and both muscles and vessels cover a much larger

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

volume of the body. Thus, an automatic detector has to cope with lots of clutter. Furthermore, the size of a lymph node can vary a lot. While healthy lymph nodes typically have a size in the range of a few millimeters up to one or two centimeters (Warwick and Williams, 1858), lymph nodes that are pathologically enlarged, for instance due to cancer or an infection, can have a size of five centimeters and more. Often, multiple enlarged lymph nodes lie directly adjacent to each other and form clusters. Detecting lymph nodes in a cluster is especially challenging because the boundary between different nodes is often not clearly visible, or not visible at all. Then, the shape of the cluster can be almost arbitrary.

Because of these difficulties, it is vital to incorporate as much prior knowledge as possible into the detection. In particular, we know that

- Lymph nodes do not appear anywhere. In the mediastinum, they always lie in fat tissue, so space inside any organ can be excluded.
- Lymph nodes are not uniformly distributed in fat tissue. Instead, it is much more likely to observe lymph nodes below the aortic arch and close to the trachea.

It turns out that exploiting this prior knowledge can help to greatly reduce the number of false detections and thus improve the overall detection performance.

In this work, this knowledge is modeled using a spatial prior probability $p(m = 1|\boldsymbol{t})$ of observing a lymph node at a given location $\boldsymbol{t} = (t_x, t_y, t_z)$, where m denotes the binary class variable. Three increasingly complex priors are proposed and compared against each other.

2.1. Automatic landmark detection and organ segmentation

While variant 1 is a trivial prior, the variants 2 and 3 depend on anatomical structures that first need to be detected in a CT volume image. We automatically find a set of 20 salient anatomical landmarks that lie mostly but not exclusively in the chest area and can be detected robustly. The detection method used here is described in Seifert et al. (2009) and Liu et al. (2010). Examples of landmarks are the bifurcation of the trachea, the bottom tip of the shoulder blade left and right, the topmost point of the aortic arch and the topmost point of the lung left and right.

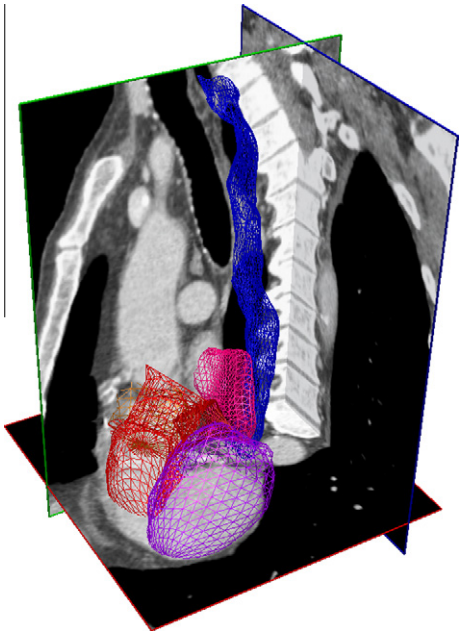

Fig. 3. Heart and esophagus model fitted to a CT volume.

Besides the landmarks, a number of different organs are segmented. The lungs and the trachea are detected using simple thresholding followed by a morphological opening operation. The four heart chambers are segmented as described in Zheng et al. (2007). The esophagus is segmented using the approach of Feulner et al. (2009). The latter two methods both combine discriminative learning with model fitting. The esophagus is of special interest as it is often surrounded by lymph nodes, but at the same time can be confused with lymphatic tissue. Fig. 3 shows an example segmentation of the heart and the esophagus. All segmentation methods do not require user interaction.

2.2. Variant 1: constant prior

In variant 1, the probability $p(m = 1|\boldsymbol{t})$ is simply modeled to be constant

$$p_1(m = 1|\boldsymbol{t}) = \text{const.}, \tag{1}$$

which means that no prior knowledge is used. This serves as a baseline for the remaining two variants.

2.3. Variant 2: binary mask

In the second variant, the spatial prior is modeled to be proportional to a binary mask $B(\boldsymbol{t})$

$$p_2(m = 1|\boldsymbol{t}) \propto B(\boldsymbol{t}) = \begin{cases} 0 & \text{if } \boldsymbol{t} \text{ is inside an organ} \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

that labels regions that cannot contain lymph nodes with 0 and other regions with 1. The lungs, the trachea, the esophagus and the heart are excluded, i.e. labeled with zero in the mask.

2.4. Variant 3: soft prior

The third variant consists of the binary mask $B(\boldsymbol{t})$ and a probabilistic atlas

$$G(\boldsymbol{t}) \in [0, 1] \tag{3}$$

which is learned in the space of a reference patient. Non-rigid inter subject registration is used to map segmented lymph nodes from a set of test patients to the reference patient, where they are averaged. The segmentations are binary masks, and thus $G(\boldsymbol{t})$ is the spatial probability of lymphatic tissue. The learned probabilistic atlas is blurred with a Gaussian filter with a standard deviation of 12 mm which is necessary because of limited training data. This can also be considered as a Parzen estimation.

The registration is based on the set of 20 landmarks. If a landmark is not detected, e.g. because it is not visible in the image, it is omitted. A thin-plate spline (TPS) transformation (Bookstein, 1989) is created from the detected landmarks and the reference landmarks and used for the warping. During training, the transformation maps from the reference space to the current image, and for the testing phase, it maps into the other direction.

The prior is then modeled to be

$$p_3(m = 1|\boldsymbol{t}) \propto B(\boldsymbol{t})G(\boldsymbol{t}) \tag{4}$$

proportional to the product of the binary mask and the probabilistic atlas.

Fig. 4 shows examples of the different prior types along with the original volume image they were computed from. Each image in a column shows the same slice of the volume, and the slices are parallel to the coordinate planes. Fig. 4a shows the original volume. The binary prior $p_2(m = 1|\boldsymbol{t})$ shown in Fig. 4b excludes already

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

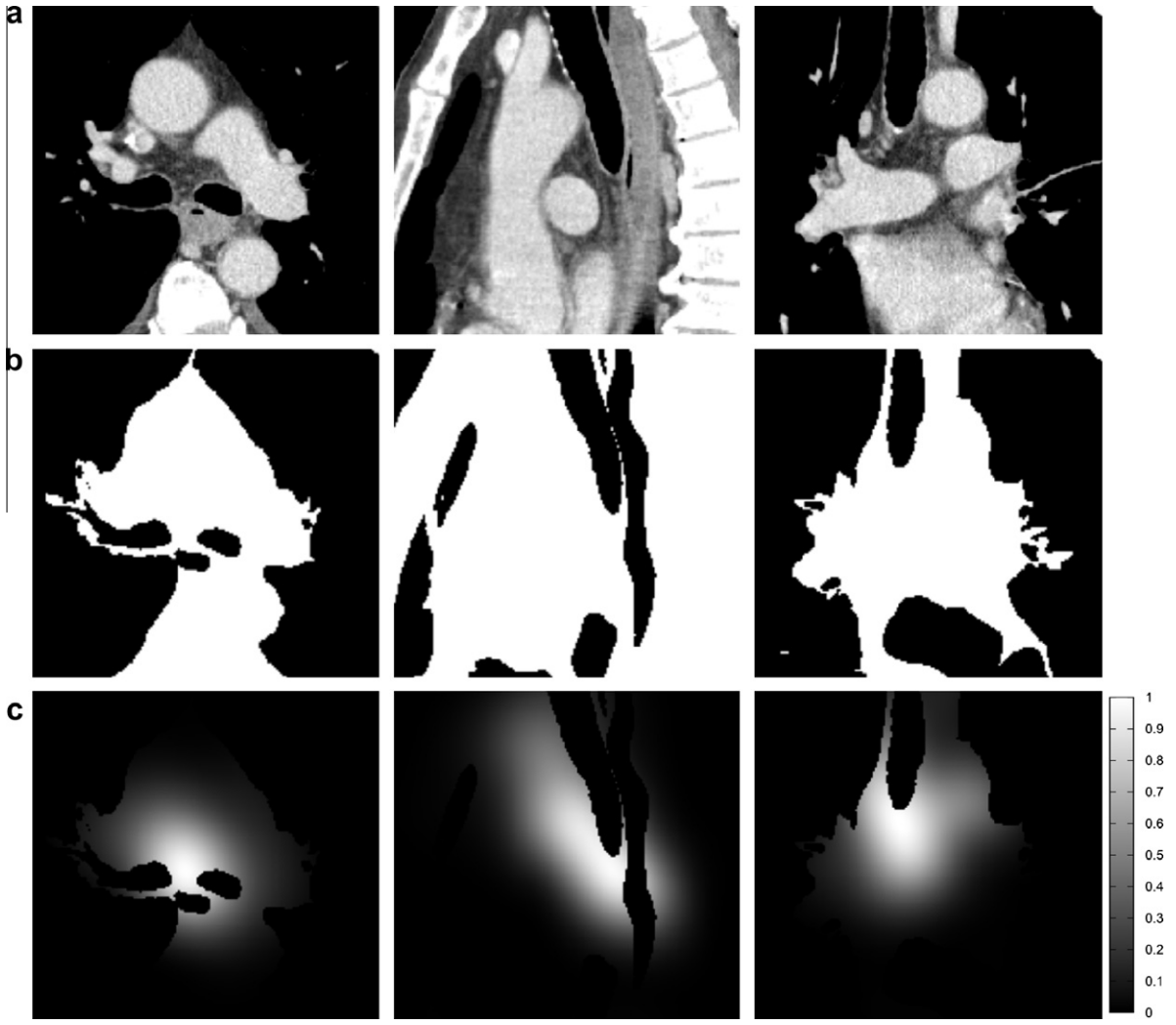

Fig. 4. Examples of different spatial priors computed for a test volume. The three columns show axis-aligned orthogonal slices of the volume. (a) The input volume. (b) The binary prior $p_2(m = 1|\boldsymbol{t})$ (see Eq. (2)) that excludes air and organs. (c) The "soft" prior $p_3(m = 1|\boldsymbol{t})$ (see Eq. (4)). As the absolute values of the prior do not matter in our model, the prior can have an arbitrary positive scaling factor.

considerable portions of the volume. The "soft" prior $p_3(m = 1|\boldsymbol{t})$ shown in Fig. 4c puts special focus on relatively small regions of the volume.

3. Region of interest detection

This work focuses on detecting mediastinal lymph nodes. Therefore, a region of interest (ROI) is automatically detected that covers the mediastinum. This ROI is anchored at the bifurcation of the trachea, which is a landmark that can be detected very robustly. It is contained in the set of landmarks that are detected as described in Section 2.1. The ROI has a fixed minimum size of $18.4 \times 18.0 \times 19.5\ \text{cm}^3$. If parts of the segmentations of the heart or the esophagus, which are used to compute the spatial prior, are outside this ROI, then it is enlarged to completely contain these segmentations. Fig. 5 shows an example of an ROI.

4. Position candidate detection

Detecting lymph nodes in one step would require to estimate many parameters simultaneously. Instead of directly searching a high dimensional parameter space, we break the detection into smaller sub problems. At first, we only detect a number of candidates of possible lymph node center positions, which is described in this section. In later steps, these candidates are verified or rejected, and used to initialize detectors for the actual lymph node segmentation.

Stage one of the detection system (Fig. 2) is a sliding window detector that uses a probabilistic boosting tree (PBT) classifier (Tu, 2005) that selects a subset of 3-D Haar-like features from a large pool as described in Tu et al. (2006). A PBT is a binary decision tree with a strong AdaBoost classifier at each node. The classifier is trained to learn the probability

$$p(m = 1|\boldsymbol{H}(\boldsymbol{t})) \tag{5}$$

of whether there is a lymph node model instance at a given position $\boldsymbol{t}$, which has similarities to the face detector proposed by Viola and Jones (2001). Here, $\boldsymbol{H}$ denotes the Haar feature vector extracted at position $\boldsymbol{t}$. These features are weighted sums of integrals of the image computed over axis aligned boxes. Haar features are used because they can be computed very efficiently by precomputing an integral image so that it is even possible to search all positions in the ROI exhaustively.

Given the output of the classifier, a set of position candidates $C_{H1} = \{\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{|C_{H1}|}\}$ is generated. If a fixed threshold θ_{H1} is used and we select all $\boldsymbol{t}$ that satisfy $p(m = 1|\boldsymbol{H}(\boldsymbol{t})) > \theta_{H1}$, we run into the problem that lots of candidates are generated at lymph nodes which are clearly visible, but we do not get any candidates at

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

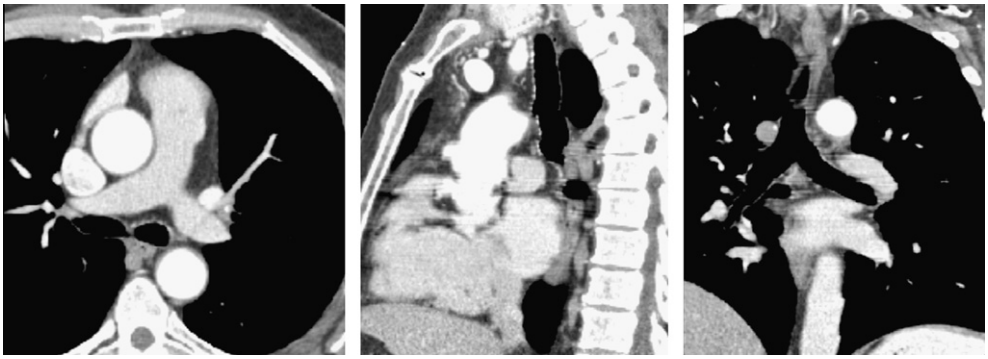

Fig. 5. Axial, sagittal and coronal slice through the region of interest of a CT scan.

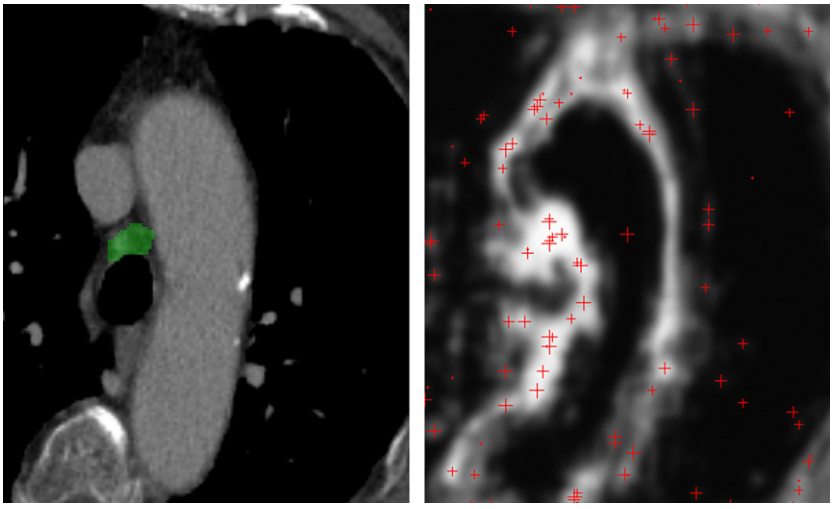

Fig. 6. Local maxima of the probability map generated by the detector are used as position candidates (marked as +). Note that this is a 2-D slice of a 3-D volume and points that look like local optima in 2-D are not necessarily local optima in 3-D.

lymph nodes which are hard to see. To overcome this, we use a technique proposed in Chen et al. (2009). First, a probability map is generated by multiplying the classifier output with the spatial prior. This map is blurred using a Gaussian filter with a standard deviation of 1.5 mm, and local maxima in the probability map are selected as candidates. An example of candidates extracted from the blurred probability map can be seen in Fig. 6 along with the CT image data it was generated from.

Now, another PBT classifier is used to examine the position candidates in the set C_{H1} and to reject false positives, resulting in a set C_{H2} of candidates. The second classifier is trained using exactly the same pool of 3-D Haar-like features as the first one. The difference is that the negative training examples of the second classifier are generated by scanning images using the first classifier and collecting false alarms. Thus, more focus is put on difficult cases. As their training sets differ, the first and the second classifier will in general select different features from the pool.

As described in the next section, lymph node center point candidates from the set C_{H2} are now used as seed points of a segmentation method, and the resulting segmentations are used to verify the detection result.

5. Joint detection and segmentation

Standard features such as Haar features used in Section 4 work well for a broad range of applications. But better performance can be achieved using features that are designed for a particular problem. The idea is to not extract features at predefined locations, for instance on a regular grid. Instead, in this paper we present a set of problem-specific features that is extracted from a candidate segmentation. In general, the quality of the segmentation result will be good if the segmentation was initialized with the center of a true lymph node, and otherwise it will be poor. Therefore, the segmentation result can give a valuable hint of whether the underlying detection is a true or a false positive.

We use a variant of the graph cuts method for seeded image segmentation (Boykov and Kolmogorov, 2003) to segment a lymph node given its detection. Advantages of graph cuts are that the segmentation is computationally inexpensive because it can be solved non-iteratively, and at the same time, it is flexible enough to allow adapting it to this particular problem, leading to a good performance.

In total, these segmentation based features are however computationally relatively expensive because the segmentation has to be carried out for every single candidate. Therefore, these features are used in the final step of the detection cascade. To avoid computing too many segmentations during test, an additional step is inserted into the detection cascade that reduces the number of position candidates C_{H2} generated by the second cascade level. It uses so-called gradient-aligned features. These are point features that are extracted at local maxima of the gradient magnitude. They are computationally less expensive and explained in Section 5.1 in more detail.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

5.1. Verifying detections using gradient aligned features

In order to reduce the number of lymph node position candidates in the set C_{H2} (see Section 4) that is generated by the second Haar feature based detector, the candidates are verified by a detector that uses gradient-aligned features as proposed in Barbu et al. (2010).

Starting from the presumed center $\boldsymbol{t}$ of a lymph node, 14 rays are sent out in radial direction such that three rays are parallel to the axes of the coordinate system, three are antiparallel, and the remaining eight hit the corners of an axes aligned cube placed at $\boldsymbol{t}$. Each ray is regularly sampled with a sample spacing of 1 mm. The image gradient magnitude is computed for each sample, and local maxima of the gradient magnitude above a threshold are detected along each ray. This is done for 10 different threshold values, and not only on the original image, but also on two coarser versions with a voxel spacing that is two times and four times the original spacing, respectively. This resolution hierarchy is used to make the features more robust to noise.

At each of the top three local maxima of each ray, so-called steerable features (Zheng et al., 2007) are computed, which are 24 simple point features. For a maximum located at $(m_x, m_y, m_z) \in \mathbb{R}^3$, the image intensity $I(m_x, m_y, m_z)$, the partial derivatives $\frac{\partial I}{\partial m_x}$, $\frac{\partial I}{\partial m_y}$, $\frac{\partial I}{\partial m_z}$ in all directions, the gradient magnitude $\|\nabla I\|_2$, and nonlinear variations including I^2, I^3, $\log I$, $\sqrt{\|\nabla I\|_2}$, $\|\nabla I\|_2^2$, $\|\nabla I\|_2^3$, $\log \|\nabla I\|_2$ are evaluated using nearest neighbor interpolation. Additionally, each feature is also computed on the two coarser levels of the resolution hierarchy.

The same point features are computed halfway from the center $\boldsymbol{t}$ to each of the first three local optima, for each ray. Next, the distance from the center to each of the first three local optima is used as a feature. Finally, asymmetry is captured by measuring the differences of the distances of corresponding local optima to the center for all pairs of different rays.

Now an AdaBoost classifier is trained to learn the probability $p(m = 1|\boldsymbol{A}(\boldsymbol{t}))$ of whether there is a true lymph node ($m = 1$) given the gradient aligned feature vector $\boldsymbol{A}$ extracted at position $\boldsymbol{t}$. It is used to decide if a position candidate of set C_{H2} is a true positive or a false positive. The set of candidates with the best classification score is kept and denoted with C_A.

5.2. Segmenting node-like structures using graph cuts

In this section, we adapt the graph cuts segmentation method to the problem of lymph nodes segmentation. Graph cuts are an all-purpose segmentation method, but the performance can be considerably improved by using problem-specific seeds and weights.

We adapt graph cuts segmentation in two ways. First, we select the edge weights of the graph according to boundary and object probabilities that are obtained from intensity and joint intensity histograms extracted from manually segmented data.

Next, we propose a sphere shape prior that is well suited to segment blob-like nodal structures. For approximately spherical objects, this prior solves at the same time a major problem of graph cuts, which is the "small cut" behavior: When only few seeds are given, the cheapest cut is often the one directly around the seeds.

At this point, we already have the center $\boldsymbol{t} \in C_A$ of a detected lymph node candidate from our previous detection steps. We consider a sub-image cropped from the original volume image such that $\boldsymbol{t}$ is centered in the sub image. The size of the sub-image remains fixed at $4 \times 4 \times 4\,\mathrm{cm}^3$. This is relatively large and ensures that almost all lymph nodes fit into this window.

This sub-image is now converted into a graph representation. Each voxel is a node of the graph, and neighboring voxels are connected based on a neighborhood criterion. In this work, this is either a 6-neighborhood or a 26-neighborhood. A larger neighborhood leads to a more complex graph and is computationally more expensive, but often leads to a more accurate segmentation (Boykov and Kolmogorov, 2003). There are furthermore two special nodes, the source and the sink, that are directly connected to multiple voxel nodes. This graph topology is also called s-t graph in the literature. Such a graph is illustrated in Fig. 7 for a 2-D image and a 4-neighborhood.

Each edge is associated with a capacity. The capacity between neighboring voxels i and j is denoted with β_{ij}, and the so-called unary capacities of the edges from the source or to the sink are denoted by λ_i. If $\lambda_i > 0$, then there is an edge with a capacity λ_i from the source to voxel i. If $\lambda_i < 0$, then there is an edge from voxel i to the sink with capacity $-\lambda_i$.

Let N be the number of voxels in the sub-image. Each voxel is associated with a random variable $x_i \in \{0, 1\}$ that is either one for "foreground" or zero for "background". When the minimum cut (or the maximum flow) is computed that separates the source from the sink, and x_i is set to 1 if voxel i is on the source side of the cut and to 0 otherwise, then the resulting vector $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_N)$ equals

$$\hat{\boldsymbol{x}}_{\text{undirected}} = \arg\max_{\boldsymbol{x}} \sum_i \lambda_i x_i + \frac{1}{2} \sum_{ij} \beta_{ij} \delta(x_i, x_j) \tag{6}$$

with

$$x_i \in \{0, 1\}; \quad \delta(a, b) = \begin{cases} 1 : a = b \\ 0 : \text{else} \end{cases} \tag{7}$$

This was shown in Greig et al. (1989). A high β_{ij} value reflects that voxels i and j are likely to have the same label. A high λ_i value means that, without knowing anything about its neighborhood, voxel i is more likely to be foreground.

Eq. (6) only holds for an undirected graph. In the more general case of a directed graph as used in this paper, the cost function computed by graph cuts slightly differs from (6) and becomes

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \sum_i \lambda_i x_i - \sum_{ij} \beta_{ij} x_i (1 - x_j). \tag{8}$$

In case of undirected edges, i.e. $\beta_{ij} = \beta_{ji}$, (6) and (8) are equivalent.

Since the center $\boldsymbol{t}$ of the sub-image is assumed to be the center of the lymph node, it is used as positive seed and its λ_{i_t} value is set to ∞. The boundary voxels of the sub-image are marked as negative seeds and their unary capacities are set to $-\infty$.

If all other unary capacities λ_i were set to zero, and all binary capacities β_{ij} to some positive constant, then the cost of a cut would be proportional to its surface, and the smallest cut that separates the source from the sink would simply separate the positive seed from its direct neighbors (see Fig. 8 left). This is also known as the small cut problem of graph cuts (Sinop and Grady, 2007). In this special setting, the problem can be solved by simply adding a factor $\frac{1}{r_{ij}^2}$ to the capacities β_{ij}, where r_{ij} denotes the distance of

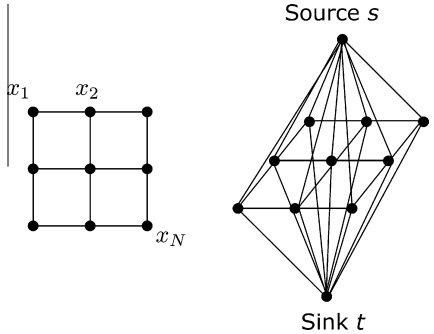

Fig. 7. Illustration of the graph of a 2-D 3 × 3 image.

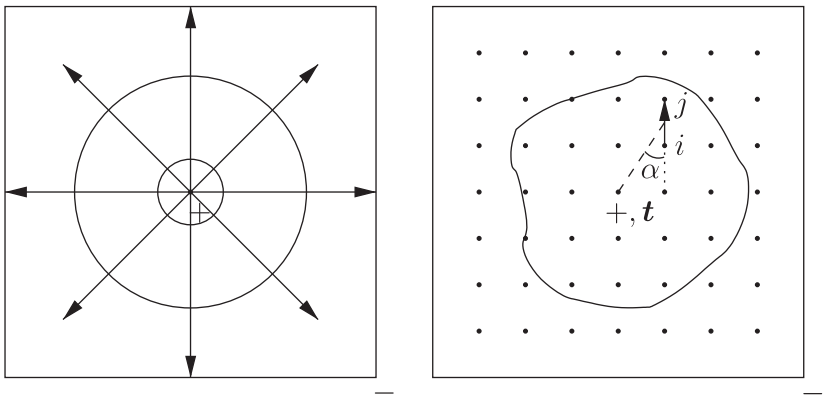

Fig. 8. Left: the total flow through surfaces that separate the positive seed (+) from the negative seeds (−), for instance concentric circles (spheres) centered at the positive seed, is constant. If the unary edge weights λ_i are nonzero only at the seeds, and if β_{ij} is constant for all edges between neighboring voxels i,j, then the cost of a cut is proportional to its length (surface). Right: illustration of $p(\text{out}_{ij})$. If the edge from voxel i to voxel j points away from the center $\boldsymbol{t}$ (i.e. $\cos\alpha \approx 1$), then it is likely to point in outward direction.

the center point of the edge from voxel i to voxel j to the positive seed at $\boldsymbol{t}$. If the original capacities β'_{ij} are constant and $\beta_{ij} = \frac{1}{r_{ij}^2}\beta'_{ij}$, then the integrated capacity $B(r)$

$$B(r) = \sum_{(i,j)\in\mathcal{B}(r)} \beta_{ij} \tag{9}$$

over a sphere centered at $\boldsymbol{t}$ is nearly constant for different radii r (it is not exactly constant because of discrete voxels). In (9), $\mathcal{B}(r)$ denotes the set of edges intersected by the sphere centered at $\boldsymbol{t}$ with radius r. Now, there is no bias any more toward a small cut. Because of the smaller surface, spherical cuts are preferred over non-spherical cuts, which is a desirable property for the purpose of segmenting a node-like structure. This method is not only simple, is also comes at no additional computational costs.

Other shape priors have been proposed for graph cuts segmentation. In Slabaugh and Unal (2005), a prior for elliptic shapes was introduced. However, the segmentation must be solved iteratively. In Funka-lea et al. (2006), a method that favors cuts that are orthogonal to the line from the current point to the center was proposed. This is effectively a prior for blob-like structures but does not solve the small cut problem. A prior for star-shaped structures and also a balloon force that corresponds to a certain boundary length was introduced in Veksler (2008). This solves the small cut problem, but the balloon force is optimized iteratively. In Das et al. (2009), the same balloon force as in Veksler (2008) is used together with a prior for compact shapes, but there is no obvious extension of the shape prior to 3-D. Wang and Siskind (2001) proposed to normalize the cost of a cut by its boundary length. This solves the bias toward a small cut. However, it also removes the bias toward a smooth surface and is therefore more prone to producing leaking segmentations. Furthermore, the minimum cut/maximum flow based global optimization technique of graph cuts is not applicable any more with this cost function and an iterative optimization scheme is required. A prior for tubular shapes was presented in Bauer et al. (2010a). Transformation into a prior for spherical shapes is straightforward, but in contrast to our proposed shape prior, the approximate radius needs to be known in advance.

The term optimized in (8) can also be considered as being proportional to the logarithm of the *a posteriori* probability $p(\boldsymbol{x}|I_1 \cdots I_N)$

$$\log p(\boldsymbol{x}|I_1 \cdots I_N) \propto \sum_i \lambda_i x_i - \sum_{ij} \beta_{ij} x_i (1 - x_j) \tag{10}$$

of a pairwise Markov random field, as described in Greig et al. (1989) for the case of a directed graph. This motivates selecting the capacity λ_i according to the logarithm of the probability that the isolated voxel i is foreground, and selecting β_{ij} based on the logarithm of the probability of observing the object boundary between the voxels i and j. The range of the attenuation coefficients of lymph nodes is restricted. This allows to estimate these probabilities according to intensity histograms.

Here, we set the unary capacity λ_i to

$$\lambda_i = \log \frac{p^u(x_i = 1|I_i)}{1 - p^u(x_i = 1|I_i)} \tag{11}$$

the logarithm of the odds that voxel i is foreground given its intensity value I_i. The probability $p(x_i = 1|I_i)$ is estimated non-parametrically using a histogram. u is a normalizing constant that is used to balance the influence of the unary and binary capacities. It was set to 0.13 in the experiments. Then, λ_i approaches $-\infty$ as $p(x_i = 1|I_i)$ approaches zero. The effect is that it is infinitely expensive to label voxel i as foreground if it is for sure a background voxel. Similarly, λ_i approaches ∞ as $p(x_i = 1|I_i)$ approaches one, meaning that labeling voxel i as foreground if it belongs for sure to the foreground leads to an infinite gain of the cost function (8).

As pointed out above, the binary capacity β_{ij} should be set according to the distance of the edge from voxel i to voxel j to the central seed at $\boldsymbol{t}$, but also according to the orientation of the edge, the intensities at i and j, and common intensity jumps at the border of lymph nodes in order to use all available information. Here, it is set to

$$\beta_{ij} = -\frac{1}{r_{ij}^2} \cdot \frac{1}{d_{ij}} \log[p(\text{out}_{ij}) p(x_i = 1, x_j = 0|I_i, I_j)], \tag{12}$$

where

$$p(\text{out}_{ij}) = \frac{\cos\alpha_{ij} + 1}{2} \tag{13}$$

is the estimated probability that the edge from voxel i to voxel j is pointing in outward direction. In (13), α_{ij} is the angle between the edge from i to j and the line from the positive seed to the center of the edge. Thus, $\cos\alpha_{ij} = 1$ if the edge is pointing away from the central seed, and $\cos\alpha_{ij} = -1$ if it is pointing toward the center. See Fig. 8 (right) for an illustration. The term $p(x_i = 1, x_j = 0|I_i, I_j)$ denotes the probability of observing the object boundary between the adjacent voxels i and j given the intensity I_i of the voxel that is assumed to be inside and I_j of the voxel that it assumed to be outside the segmentation. The term d_{ij} in (12) denotes the euclidean distance of the voxels i and j. It is relevant when a neighborhood system is used that does not only include the six direct neighbors.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

Note that β_{ij} is directed, as this allows to incorporate additional knowledge about the object boundary. If the edge from voxel i to j is pointing in inward direction, then $p(\text{out}_{ij}) = 0$ and therefore $\beta_{ij} = \infty$. It models that the interior of the lymph node is expected to be closer to the center than the exterior of the node. If both $p(\text{out}_{ij}) = 1$ and $p(x_i = 1, x_j = 0|I_i, I_j) = 1$, then $\beta_{ij} = 0$, meaning that cutting this edge comes at no costs.

In (12), $p(x_i = 1, x_j = 0|I_i, I_j)$ can be expressed as

$$p(x_i = 1, x_j = 0|I_i, I_j) = \frac{p(x_i = 1, x_j = 0, I_i, I_j)}{p(I_i, I_j)}. \quad (14)$$

Both $p(x_i = 1, x_j = 0, I_i, I_j)$ and $p(I_i, I_j)$ are estimated non-parametrically using joint intensity histograms. $p(I_i, I_j)$ is set to the number of neighboring voxels with intensities I_i and I_j divided by the number of neighboring voxels. $p(x_i = 1, x_j = 0, I_i, I_j)$ is computed by first counting the number of neighboring voxels with the properties that voxel i is inside a lymph node and has an intensity of I_i, and voxel j is outside any lymph node and has an intensity of I_j, and then dividing this number by the number of neighboring voxels. However, $p(x_i = 1, x_j = 0, I_i, I_j)$ is sparse because of a limited number of training examples of points on the boundary of lymph nodes. Therefore, $p(x_i = 1, x_j = 0|I_i, I_j)$ is smoothed with a Gaussian filter with $\sigma = 40$ HU, which is effectively a Parzen estimation. Fig. 9 shows the estimated probability $p(x_i = 1, x_j = 0|I_i, I_j)$. All histograms have 400 equally spaced bins in each dimension, where the lowest bin corresponds to -1024 HU and each bin is 4 HU wide. Voxels with an intensity outside this range are dropped and not added to the histogram. During test, $p(x_i = 1, x_j = 0|I_i, I_j)$ is set to zero if either I_i or I_j is outside the range. The implied error is tolerable, because lymph nodes and directly adjacent structures are, with very few exceptions, inside this relatively wide range.

5.3. Segmentation based features

We now have the candidate segmentation that was initialized with the detected lymph node center $\boldsymbol{t}$. As final stage in the detection cascade, an AdaBoost classifier is trained with features extracted from the segmentation to learn whether $\boldsymbol{t}$ is a true lymph node or a false detection.

The first kind of features is histogram based: Given a binary segmentation mask image, a hierarchy of normalized histograms of the intensity values inside the segmentation is computed. The histogram at the first level has 256 bins. Each bin is one Hounsfield unit wide, and the first bin corresponds to -128 HU. Lymph nodes typically fall into this range of HU values. At the next level, the number of bins is halved, and the width of each bin is doubled. In total, seven levels are used. The entry of each bin of each pyramid level is a scalar feature.

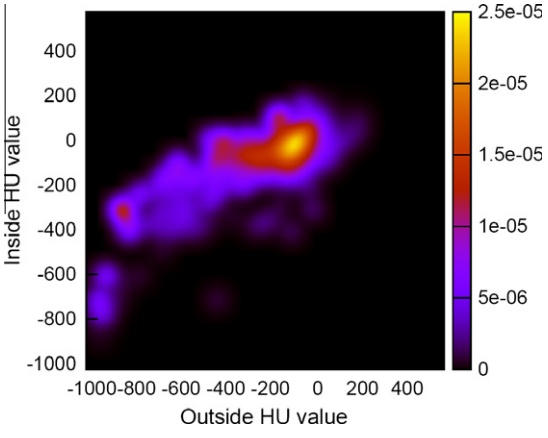

Fig. 9. Estimate of the probability $p(x_i = 1, x_j = 0|I_i, I_j)$. It is asymmetric because the interior of a lymph node has usually a higher attenuation coefficient than its surroundings.

The second kind of features are again based on a hierarchy of histograms, but the histograms are now computed from the 3 mm wide neighborhood of the segmentation. The neighborhood is determined using morphological operations. Additionally, we use the second, third and fourth central moments of the histograms both inside and outside the segmentation.

Next, 100 points are, with equal probability, randomly sampled with replacement from the surface voxels of the segmentation. As proposed in Barbu et al. (2010), the gradient is computed at each point, and the points are sorted by their gradient magnitude. The sorting is necessary to enumerate the points. At each point, the normal to the surface is computed, and the normal is sampled at seven positions with a spacing of 1 mm between the samples. At each sample, steerable features (see Section 5.1) are computed. All scalar features at all samples at all normals at all points are added to the feature pool.

Furthermore, features are used to capture the relative position of the lymph node center $\boldsymbol{t}$ within the tight axes-aligned bounding box of the segmentation. A relative position $\boldsymbol{t}'$ of $\boldsymbol{t}$ inside this box is computed that is normalized to lie in $[-0.5, 0.5]$ for each dimension. A value of 0 indicates that $\boldsymbol{t}'$ is centered, and values of -0.5 and 0.5 indicate that $\boldsymbol{t}'$ lies on the bounding box wall in this dimension. The minimum relative distance to any wall of the box, the difference of the maximum and the minimum distance to any wall, and the relative distance averaged over the three dimensions are used as features.

Finally, the volume, the surface, the sphericity, the maximum flow value and the maximum flow divided by the surface are used. In total, the feature pool contains 51,436 features. The vector containing these features extracted at position $\boldsymbol{t}$ is denoted with $\boldsymbol{D}(\boldsymbol{t})$.

As in Section 5.1, an AdaBoost classifier is used to select a small subset of the feature pool and is trained to learn the probability $p(m = 1|\boldsymbol{D}(\boldsymbol{t}))$ of whether $\boldsymbol{t}$ is a true lymph node and not a false positive given the segmentation based features. Note that, apart from the maximum flow features, the described feature set does not depend on the segmentation method and can therefore also be used in combination with other segmentation techniques, as described in the following section.

5.4. Alternative segmentation methods

In Section 5.2, a graph cuts based segmentation method was presented that was specially designed for the problem of lymph nodes segmentation.

It is however interesting to see how the detection performance is affected if this graph cuts based segmentation method is replaced with a simpler segmentation method. We therefore also use both graph cuts with standard weights and a watershed based segmentation method as a baseline. Features extracted from these segmentations are used in the same way as described in Section 5.3 to train a classifier to distinguish true positives and false positives.

5.4.1. Graph cuts with standard weights

It is popular in the literature (Boykov and Funka-Lea, 2006) to use graph cuts segmentation with unary weights

$$\lambda_i = 0 \quad (15)$$

set to zero for all voxels i except for the seeds. This means that no prior knowledge about the foreground or background intensities is available and is therefore not specific to a certain problem. The binary edge weight β_{ij} from voxel i to voxel j is commonly set to

$$\beta_{ij} = \exp\left(\frac{-(I_i - I_j)^2}{2\sigma_\beta^2}\right), \tag{16}$$

where σ_β is a constant that is typically set according to the noisiness of the input data. These weights are symmetric and simply mean that the object boundary is probably at image intensity jumps. The effect of a low value of σ_β is that the edge capacities β_{ij} quickly approach zero already at moderate intensity jumps. The cut is then more susceptible to noise, and also numeric problems can arise. If, on the other hand, σ_β is set to a high value, cuts at intensity jumps are more likely, and the surface of the cut becomes more important. As a result, the method is more prone to the small cut problem. Unless mentioned otherwise, we use a value of σ_β = 16 HU, which is a good compromise.

5.4.2. Hierarchical watershed segmentation

The watershed transformation (Beucher and Meyer, 1992) is another popular low level method for image segmentation.

In order to enhance relevant edges, the input image is windowed with a soft tissue window (center: 16 HU, width: 400 HU). An edge image E is generated by computing the gradient magnitude of the windowed image I_W. To reduce the susceptibility to noise, the gradients are computed by convolving the image with the first derivatives of a 3-D Gaussian g with a standard deviation of 2 mm in each direction:

$$E(\boldsymbol{t}) = \|\nabla g(\boldsymbol{t}) * I_W(\boldsymbol{t})\|_2 \quad \text{with } \nabla g(\boldsymbol{t}) \in \mathbb{R}^3. \tag{17}$$

The convolution in (17) is carried out component-wise.

In this work, we use an hierarchical segmentation method (Beucher, 1994). In the first step, it generates a very oversegmented mosaic image using the standard watershed transform of the edge image E. In order to reduce the number of watershed regions that are generated in this first step, E is thresholded: Voxels below a threshold θ_{WS} are set to θ_{WS}. Local minima of E serve as seeds in the watershed transform, and the effect of the thresholding is that seeds are merged if they are close to each other and have a low absolute value. The threshold θ_{WS} is set to $0.01e_{max}$, where e_{max} is the maximum value that occurs in the edge image E.

In the next step, segmentations with different "flooding levels" l_{WS} are generated by merging neighboring watershed regions. The higher l_{WS} is, the more regions are merged, and the less oversegmented is the resulting segmentation. l_{WS} is a relative value in the range of [0,1]. A value of zero means that no regions are merged, and one means that all regions are merged into a single one. It corresponds to an absolute flooding level L_{WS} with

$$L_{WS} = l_{WS} e_{max}. \tag{18}$$

Region A floods region B of the edge image (i.e. A and B are merged) at level L_{WS} if the relative depth of region A is lower than L_{WS} and region B is across the lowest part of the boundary of region A. The relative depth of a region in the edge image is here the lowest value

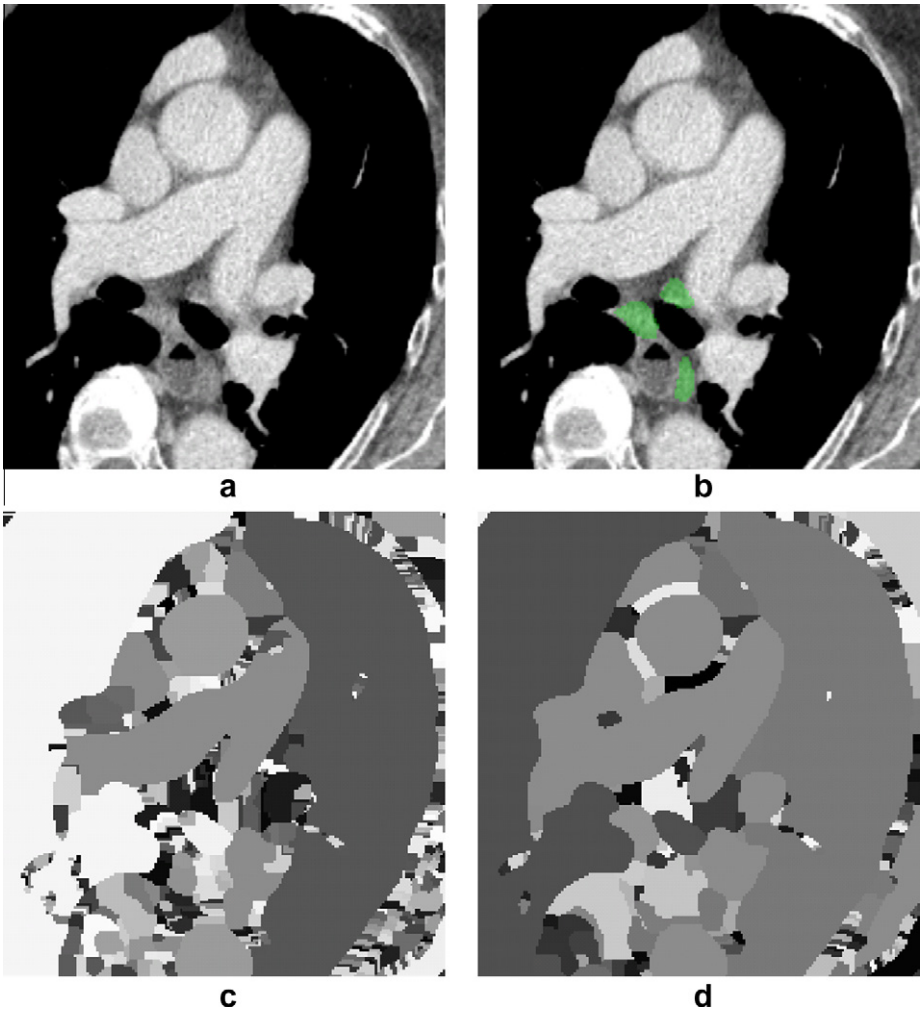

Fig. 10. Example of a slice of a 3-D hierarchical watershed segmentation. (a) An axial slice of a CT volume. (b) Manual ground truth segmentations of lymph nodes (green). (c–d) Slices of two 3-D watershed segmentations with flooding levels l_{WS} of 0.1 (c) and 0.2 (d). The image intensity of each region is randomly chosen.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

on the region boundary minus the lowest value in the region. In this work, we use two levels with $l_{WS} = 0.1$ and $l_{WS} = 0.2$. Example segmentations can be seen in Fig. 10.

The segmentation based features $\boldsymbol{D}(\boldsymbol{t})$ are now extracted from the region that contains the center $\boldsymbol{t}$ of the detected lymph node that is to be verified. Note that, in contrast to the graph cuts method, neither the segmentation nor the segmentation based features as described in Section 5.3 depend on the location of $\boldsymbol{t}$ within the region, apart from the three features that capture the relative position of $\boldsymbol{t}$ inside the bounding box of the segmentation. Therefore, the segmentation of the whole image region of interest is precomputed. Furthermore, the computed features of a region are cached. Thus, they do not need to be recomputed if another lymph node is verified whose center falls into the same region.

5.5. Combining clues from alternative segmentations

Instead of only using different segmentation methods alternatively, we also explored if the performance of the system can be improved by combining different segmentations. Even though our proposed segmentation method described in Section 5.2 is already tuned to the problem of lymph nodes segmentation, it is possible that segmentations generated with simpler methods still contain valuable information that helps to distinguish true lymph nodes from false alarms.

Here, different segmentation methods are initialized with the center of a lymph node detection. The features described in Section 5.3 are extracted from all segmentations and added to a common feature pool. For instance, if the proposed graph cuts segmentation method is combined with two hierarchy levels of the hierarchical watershed segmentation, three alternative segmentations are generated, and the number of features will be three times larger compared to when only a single segmentation is used. It can also be viewed as treating the type of the segmentation method as a feature.

If N_S different segmentation methods are combined, the new feature vector

$$\boldsymbol{D}_{\text{combined}}(\boldsymbol{t}) = (\boldsymbol{D}_1(\boldsymbol{t}), \ldots, \boldsymbol{D}_{N_S}(\boldsymbol{t})) \tag{19}$$

is simply the concatenation of the vectors extracted form the single segmentations. The classifier is then trained on the joint feature pool to learn

$$p(m = 1|\boldsymbol{D}_{\text{combined}}(\boldsymbol{t})) \tag{20}$$

the probability of whether there is a true lymph node given the joint features of the segmentations initialized with the presumed lymph node center $\boldsymbol{t}$.

5.6. Integrating the prior

Section 2 explained how a spatial prior $p(m = 1|\boldsymbol{t})$ of lymphatic tissue can be modeled, and Section 4 and the preceding subsections of Section 5 explained how lymph nodes are detected from their appearance using discriminative learning techniques. In this subsection, the spatial prior and the discriminative model are combined.

As the scores of the different position detectors are very dependent, and the scores of the first detectors are implicitly contained in the scores of the later ones because weak candidates are rejected at early and intermediate levels, we only use the score of the last detector. This score is then combined with the spatial prior. Thus, we are interested in the probability

$$p(m = 1|\boldsymbol{D}, \boldsymbol{t}) \tag{21}$$

of observing a lymph node given the segmentation based features $\boldsymbol{D}$ and the position $\boldsymbol{t}$. With Bayes' rule, (21) can be reformulated as

$$p(m = 1|\boldsymbol{D}, \boldsymbol{t}) = \frac{p(\boldsymbol{D}, \boldsymbol{t}|m = 1)p(m = 1)}{p(\boldsymbol{D}, \boldsymbol{t})}. \tag{22}$$

For simplification, we assume that the feature vector $\boldsymbol{D}$ is statistically independent from the position $\boldsymbol{t}$, and that both are also independent under the condition $m = 1$. This is an approximation as $\boldsymbol{D}$ obviously depends on $\boldsymbol{t}$, but $\boldsymbol{t}$ determines $\boldsymbol{D}$ only for a certain image. The assumption is justified by the fact that the spatial prior clearly improves the performance as we will see, which means that $\boldsymbol{D}$ does not contain much information about $\boldsymbol{t}$. Now (22) may be transformed into

$$p(m = 1|\boldsymbol{D}, \boldsymbol{t}) = \frac{p(\boldsymbol{D}|m = 1)p(\boldsymbol{t}|m = 1)p(m = 1)}{p(\boldsymbol{D})p(\boldsymbol{t})} \tag{23}$$

$$= \frac{p(m = 1|\boldsymbol{D})p(m = 1|\boldsymbol{t})}{p(m = 1)}, \tag{24}$$

which is proportional to the product of the spatial prior $p(m = 1|\boldsymbol{t})$ and the segmentation feature based detection score $p(m = 1|\boldsymbol{D})$. This product serves as final detection score.

In earlier stages of the detection cascade, the classifier score is also multiplied with the spatial prior, which can be justified analogously in each case. These scores, however, affect the final score only indirectly.

6. Results

The proposed methods have been evaluated on 54 CT datasets showing the chest area. All scans were taken from patients suffering from lymphoma. The slice spacing was 1 mm, and the intra-slice resolution was typically in the range between 0.7 mm and 0.9 mm. The images were reconstructed using a soft-tissue kernel.

All datasets were resampled to an isotropic $1 \times 1 \times 1$ mm^3 resolution. The mediastinal lymph nodes were manually segmented, and the segmentations were reviewed by a radiologist. In total, 1086 lymph nodes were annotated.

The detection performance was evaluated using threefold cross-validation. For each fold, the spatial prior, the classifiers and the graph cuts weights for the segmentation were trained on the training data and evaluated on the test data. The classifiers were only trained on lymph nodes that have a minimum size of 10 mm in at least two dimensions. Smaller lymph nodes are usually not pathologic (de Langen et al., 2006) and were therefore neglected. The set of manual segmentations contained six huge cases with a size exceeding 5 cm. These were mostly not single nodes but a cluster of lymph nodes that were densely packed so that the boundaries became invisible. Such cases were removed from the training set in order not to distract the detector with few extreme examples. Among the segmented lymph nodes, 289 were used for training. In order to achieve a better generalization and to avoid overfitting, the training data was mirrored by all three coordinate planes, resulting in $2^3 = 8$ times more training examples. For testing, only the original data was used.

In the testing phase, a lymph node is considered as detected if the center $\boldsymbol{t}$ of a detection is inside the tight axis-aligned bounding box of the lymph node. This criterion for a true positive detection is referred to as "in box". A lymph node is considered as a false negative (FN) if its size is at least 10 mm and it is not detected.

Occasionally, two or more detections are close together. In order to reduce the number of such double detections, the detected centers are spatially clustered and merged. Two detections are merged if their distance is below a distance threshold $\theta_d = 6$ mm. The confidence value of the merged detection is set to the sum of the original ones.

The way positive and negative training samples are generated considerably affects the detection performance. The problem with positive training examples is that the manual lymph node

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

Table 1
Features and parameter settings of the classifiers of the detection pipeline: the number of levels in the PBT classifier, the number of weak classifier per AdaBoost node, and the number of detection candidates generated at each stage. In stage three and four, only a single AdaBoost classifier is used.

Classifier	Features	Tree levels	Weak class.	Candidates
Stage 1	Haar	2	20	Not fixed
Stage 2	Haar	2	20	2000
Stage 3	Gradient aligned	1	270	200
Stage 4	Segmentation based	1	270	100

segmentations are often not convex. The main reason is that it is often not decidable where one lymph node ends and another one begins because there is no visible boundary. The straightforward approach would be to take the point of gravity or the center of the bounding box as positive example. However, this point is often close to the lymph node's boundary or even outside the actual node. As a solution, a depth map is computed for each ground truth lymph node. The map contains the shortest distance to the surface for each voxel. Local maxima of the depth map that have a minimum distance of 2 mm from the surface are selected as positive training samples.

The negative training samples of the first stage are generated by randomly sampling the training images, but no candidates are generated inside ground truth lymph nodes and in regions where the spatial prior has a value of zero because these regions are not considered during test. This avoids confusing the detector with data it will never see in the testing phase. In later stages, the negative training examples always come from the false positive detections of the previous stage. Thus, the classifiers get specialized on the difficult examples.

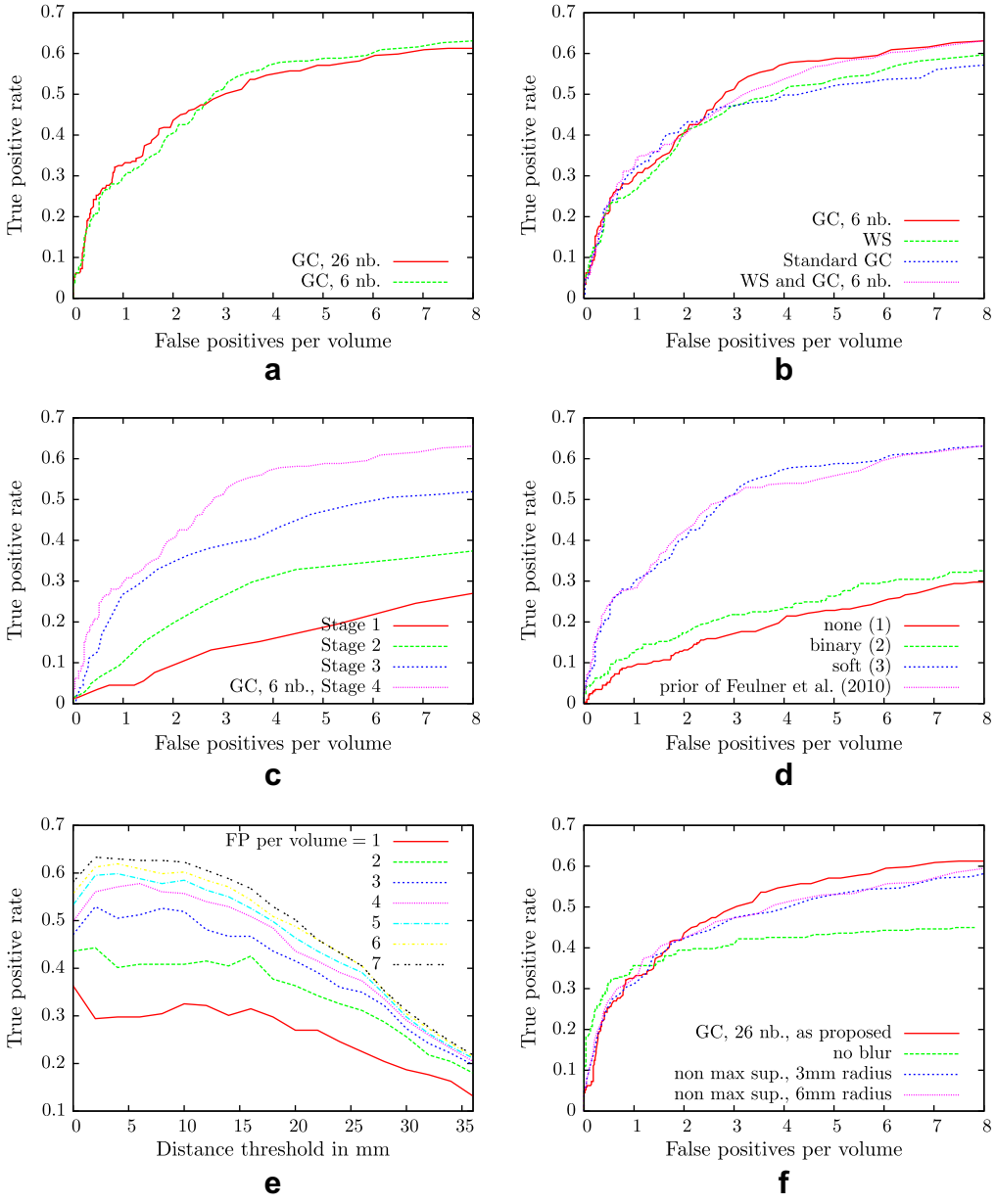

Fig. 11. Detection performance of different methods and different parameter settings. (a) Influence of the neighborhood size. (b) Influence of the segmentation method that is either the proposed one (red), watershed (green), graph cuts (GC) with standard weights (blue), or both watershed and the proposed graph cuts method (pink). (c) Performance at different stages of the cascade. (d) Effect of using different prior types. (e) Influence of the threshold used for clustering. (f) Effect of blurring the probability map, and of using non-maximal-suppression instead of clustering. See text for further details.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

Table 1 lists the parameter settings and the feature types of the classifiers in the detection pipeline that were used in the experiments, unless otherwise mentioned.

Fig. 11a–e shows the detection performance of different methods and different parameter settings as free-response receiver operating characteristic (FROC) curves.

6.1. Influence of the graph neighborhood size

We examined how the size of the graph neighborhood in the segmentation step affects the detection performance (Fig. 11a). The two curves show the detection performance with a 6-neighborhood and a 26-neighborhood. The performance is very similar, indicating that the neighborhood size does not significantly affect the detection rate. However, we noticed that a larger neighborhood leads to smoother segmentations.

6.2. Influence of the number and the type of alternative segmentations

Fig. 11b shows how the segmentation method used in stage four of the cascade affects the detection performance. The performance of our proposed graph cuts segmentation method is shown in red.[1] The blue curve shows the performance of a graph cuts segmentation with standard weights as described in Section 5.4.1. We tested whether our proposed method performs significantly better by computing the integrated difference of the two FROC curves in the range of [0,8] false positives per volume independently for each of the three cross-validation folds, and then applied a one-tailed one-sample t-test with two degrees of freedom. The difference is significant on a 90% confidence level ($P = 0.917$). Apart from evaluating alternative segmentation methods, we also did experiments with combined segmentations (see Section 5.5). First, we use two watershed segmentations with different flooding level thresholds $l_{WS} = 0.1$ and $l_{WS} = 0.2$ (see Section 5.4.2). Given a detected lymph node center candidate $\boldsymbol{t}$, the segmentation based features are extracted two times, once for each segmentation. The resulting feature vectors are called $\boldsymbol{D}_{WS}^{(0.1)}(\boldsymbol{t})$ and $\boldsymbol{D}_{WS}^{(0.2)}(\boldsymbol{t})$. In both cases, the features are extracted from the region that contains $\boldsymbol{t}$. The green curve shows the performance of a detector that is trained on the combined feature pool

$$\boldsymbol{D}_{\mathrm{WS}}(\boldsymbol{t}) = \left(\boldsymbol{D}_{\mathrm{WS}}^{(0.1)}(\boldsymbol{t}), \boldsymbol{D}_{\mathrm{WS}}^{(0.2)}(\boldsymbol{t})\right). \tag{25}$$

When used as a baseline, our proposed method (red curve) does not perform significantly better ($P = 0.778$) in detecting lymph nodes, but it usually generates meaningful segmentations, while the watershed method generates two alternative segmentations, and both are typically of poor quality. Next, we combined the features $\boldsymbol{D}_{\mathrm{WS}}(\boldsymbol{t})$ of the two watershed segmentations with features $\boldsymbol{D}_{\mathrm{GC}}(\boldsymbol{t})$ extracted from our proposed graph cuts segmentation into a feature vector

$$\boldsymbol{D}_{\mathrm{WS,GC}}(\boldsymbol{t}) = (\boldsymbol{D}_{\mathrm{WS}}(\boldsymbol{t}), \boldsymbol{D}_{\mathrm{GC}}(\boldsymbol{t})). \tag{26}$$

The resulting detection performance is shown as pink curve. Surprisingly, the detection performance does not further increase if both the graph cut and the hierarchical watershed segmentation are taken into account, indicating that the watershed segmentation does not contain a significant amount of additional information if the graph cuts segmentation is known.

6.3. Most useful segmentation based features

The AdaBoost classifier in stage 4 selects 270 out of the pool containing 51,436 scalar segmentation based features and ranks them according to their classification error. During threefold cross-validation with features extracted from the proposed graph cuts segmentation with a six neighborhood, the feature ranked first in all three folds is the ratio of the segmentation surface and the cost of the minimum cut. The third central moment of the intensity histogram of the voxels outside but in the direct vicinity of the segmentation is among the top 11 features in all three cases. The fourth central moment is in two of the three cases among the selected features, while moments of the intensity histogram inside the segmentation are interestingly not selected.

6.4. Performance at different pipeline stages

Fig. 11c shows the detection performance at different levels of the cascade (see Fig. 2). When four false alarms per volume image are allowed, the detection rates at stages one to four are 0.15, 0.31, 0.43 and 0.57. The performance improves considerably from stage to stage. In particular, the final segmentation based verification step clearly improves the detection performance.

6.5. Influence of the spatial prior

In order to examine how the type of the spatial prior affects the overall detection performance, the system was evaluated with the three different variants explained in Section 2. The results are shown in Fig. 11d. If no spatial prior is used (red curve), there is a large amount of false alarms among the detections, leading to a poor performance. Using a binary spatial prior (see Eq. (2)) that excludes organs and regions filled with air from search slightly improves the detection performance (green curve) on our data, although the improvement is not significant ($P = 0.831$). Using the "soft" prior (see Eq. (4)) that is a product of the binary prior and a probabilistic atlas greatly reduces the amount of false alarms (blue curve). When compared to using no spatial prior, the detection rate rises from 0.21 to 0.57 at four false alarms per volume, which corresponds to an increase of 171%. The improvement is significant on a 99% confidence level ($P = 0.996$). This demonstrates the importance of including prior anatomical knowledge in order to solve this challenging detection problem. Additionally, the detection performance is shown when the more complicated prior of Feulner et al. (2010) is used that is a combination of multiple probabilistic atlases. As the performance does not further increase, we propose using the simpler "soft" prior in order not to introduce unnecessary complexity.

6.6. Effect of clustering

In Fig. 11e, the detection performance of our proposed method is shown for different values of the parameter θ_d that is the distance threshold up to which close detection candidates are merged. Each curve in Fig. 11e corresponds to a different number of allowed false alarms and therefore to a certain point on the FROC curve. It can be seen that when three or more false alarms are allowed, a moderate value of θ_d in the range of 2–10 mm leads to a better detection performance. The performance degrades for higher values of θ_d because then more and more detections are merged that belong to different lymph nodes.

6.7. Comparing clustering with non-maximal suppression

Fig. 11f shows the effect of replacing the clustering step of our method with the conceptually similar technique of non-maximal suppression as used in Barbu et al. (2010) with a radius of 3 mm and 6 mm (blue and pink curves). Non-maximal suppression rejects detections in the neighborhood of a detection that has locally the highest score. Clustering performs slightly better on our data,

[1] For interpretation of color in Figs. 1, 10–12, and 14, the reader is referred to the web version of this article.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

possibly because the scores of the detections within a cluster are added, while in case of non-maximal suppression, only the score of the locally best detection is taken into account. Thus, if non-maximal suppression is used, a detection is not ranked higher if it is surrounded by many weaker detections. The difference is not significant, though.

6.8. Effect of blurring the probability map

As described in Section 4, the probability map generated in the first stage of the detection cascade is blurred, before local optima are selected as detection candidates. The green curve of Fig. 11f shows the performance of the system when the blurring step is omitted. Although the FROC curve steeply raises at the beginning, it quickly saturates and does not reach a high detection rate. We observed that, without blurring, a huge number of candidates is generated by the first stage because the probability map is noisy and contains many local optima. Thus, the first stage has almost no effect. Additionally, the resulting top-ranked detection candidates produced by the later stages are often close together and correspond to a few clearly visible lymph nodes, such that no candidates remain on lymph nodes that are less obvious.

6.9. Performance of published methods on our data

In Fig. 12a, the presented method is compared to the method we previously published in Feulner et al. (2010). There, the lymph nodes are not segmented, and stages 3 and 4 are replaced with a bounding box detector. Our current method performs better, indicating the benefit of combining detection and segmentation. The improvement is significant on a 90% confidence level ($P = 0.947$). Furthermore, we reimplemented the method of Feuerstein et al. (2009) and evaluated it on our data, in the same way as we evaluated our method. The result is shown in Fig. 12b (red curve). It generates a very high number of false alarms. Next, we combined the method of Feuerstein et al. (2009) with our "soft" prior. The detection score generated by Feuerstein et al. (2009) is the response of the "extended 3-D Min-DD" filter proposed by Kitasaka et al. (2007). This filter outputs a central difference of HU values, that can be either positive of negative. The prior can therefore not be directly multiplied with the detection score. Instead, we transformed the filter output with a sigmoid function, and multiplied the transformed filter output with the prior:

$$q(\boldsymbol{t}) = p_3(m = 1|\boldsymbol{t}) \frac{1}{1 + \exp\left(-\frac{f(\boldsymbol{t})-\gamma}{\eta}\right)}. \quad (27)$$

Here, $f(\boldsymbol{t})$ is the response of the extended 3-D Min-DD filter at position $\boldsymbol{t}$, $p_3(m = 1|\boldsymbol{t})$ is the soft spatial prior, $q(\boldsymbol{t})$ is the final detection score, and $\gamma = 500$ HU and $\eta = 100$ HU are parameters of the sigmoid function that were tuned empirically. The resulting FROC curve can be seen in Fig. 12b (blue). The prior improves the detection rate for a fixed number of false alarms remarkably. The performance is, however, still below the performance of our proposed system, and also takes far longer to process a volume (about 20 min).

Table 2 lists image databases that were used for evaluation in prior and this work, and Table 3 shows a comparison of the detection performances reported in prior work along with the performance of our method. The comparability is, however, limited because of different data, different criterions for a detection, different body regions and different minimum lymph node sizes used for evaluation. In Kitasaka et al. (2007) and Feuerstein et al. (2009), a lymph node is considered as detected if there is overlap between the segmentation and the detection. Here, this criterion is called "overlap". This error measure is however a suboptimal choice because a single huge detection covering the whole volume would result in a true positive rate (TPR) of 100% with zero FP, although the detection is obviously meaningless. Therefore, we measured the performance with the "in box" criterion mentioned earlier in this section. Both Kitasaka et al. (2007) and Feuerstein et al. (2009) report a very high number of false alarms also on their data. In Dornheim and Dornheim (2008), very good results are reported, but the method was evaluated on a single dataset. In Barbu et al. (2010), good results are reported for the axillary region. Lymph nodes in the axillary regions are however easier to detect because they are mostly isolated in fat tissue and less surrounded by clutter as in the mediastinal region.

In order to compare the automatic detection results with the performance of a human, we did an experiment on the intra-human observer variability. Ten of the CT volumes were annotated a second time by the same person a few months later. The first

Table 2
Databases used for evaluation in prior work and this paper along with minimum size of a false negative lymph node.

Database	Body region	Num. vol.	Size (mm)
Kitasaka et al. (2007)	Abdomen	5	>5.0
Feuerstein et al. (2009)	Mediastinum	5	>1.5
Dornheim and Dornheim (2008)	Neck	1	>8.0
Barbu et al. (2010)	Axillary	101	>10.0
This work	Mediastinum	54	>10.0
Intra-obs. var.	Mediastinum	10	>10.0

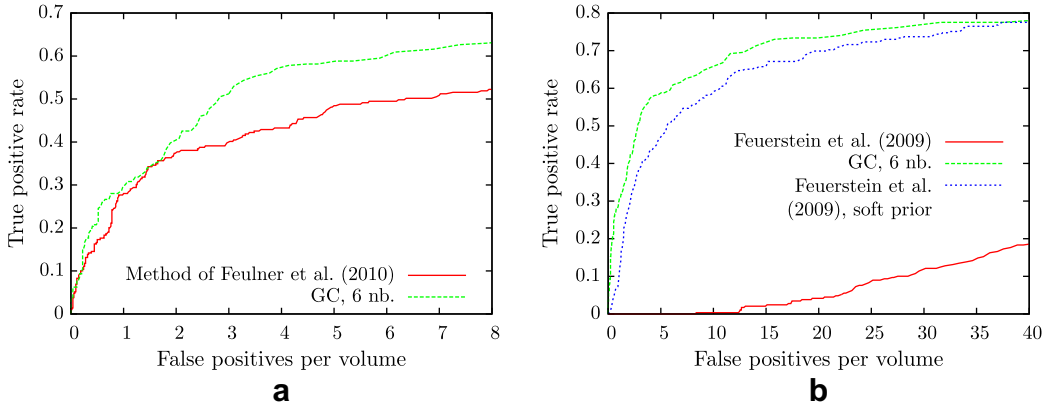

Fig. 12. Comparison with prior work. (a) Comparison of the proposed graph cuts based segmentation method and a 6-neighborhood with the method of Feulner et al. (2010). (b) The method of Feuerstein et al. (2009), either as it is or extended with our "soft" prior, compared to our graph cuts based method with a 26-neighborhood. Note that the range of the horizontal axis differs from the other plots.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

Table 3
Detection results compared to state of the art methods. The second column lists the criterion for a true positive detection. See text for details.

Method	TP crit.	TP	FP	FN	TPR (%)	FP per vol.
Kitasaka et al. (2007)	Overlap	126	290	95	57.0	58
Feuerstein et al. (2009)	Overlap	87	567	19	82.1	113
Dornheim and Dornheim (2008)	Unknown	29	9	0	100	9
Barbu et al. (2010)	In box	298	101	64	82.3	1.0
This method	In box	153	167	136	52.9	3.1
This method	In box	176	332	113	60.9	6.1
Intra-obs. var.	In box	23	8	19	54.8	0.8

segmentations served as ground truth, and the second ones were considered as detections. TPR and FP were measured in the same way as for the automatic detection. The TPR was 54.8% with 0.8 false positives per volume on average. While 0.8 FP is very low, a TPR of 54.8% shows that finding lymph nodes in CT is quite challenging also for humans. Fig. 13 shows the first and the second segmentations for one of the 10 datasets.

The proposed "soft" spatial prior relies on a set of landmarks and on segmentations of organs in the chest area (see Section 2.1). Detection and segmentation errors can therefore affect the quality of the prior, and thus the lymph node detection performance of the system. There was no overlap between the 54 datasets used for evaluation, and the training data of the segmentation and landmark detection methods. Furthermore, the image data was acquired from patients covering a normal spectrum of variation, apart from the fact that all are lymphoma patients. The quality of the detections and segmentations is therefore realistic, normal errors were present in the data. While the landmark detection step is very robust, there are cases among the 54 datasets where a segmentation method partially failed. A less accurate segmentation, however, not necessarily degrades the lymph node detection accuracy. To estimate if there is a correlation between the segmentation quality and the lymph node detection performance, we sorted all volumes by the ratio

$$R = \frac{\mathrm{TP} + 1}{\mathrm{P}}, \tag{28}$$

where TP is the number of lymph nodes detected by the graph cuts based method with a 26-neighborhood at a fixed detection threshold, and P is the number of lymph nodes in a dataset. A high value of R indicates a good detection performance on a particular dataset. TP is incremented by one to rank datasets with TP = 0 higher if P is low. Then, we checked the segmentation quality of the first 10 and last 10 sorted datasets. The segmentation quality in six of the 10 volumes with highest R, and in five of the 10 volumes with lowest R, can be considered as very good. In the other volumes, at least one segmentation has some inaccuracies. The similar performance of the segmentation methods on the two sets indicates that the segmentation quality is not a bottleneck of our method. The probabilistic atlas $G(\boldsymbol{t})$ (see Eq. (3)) is in fact not very detailed. It rather roughly highlights the regions where lymph nodes can be expected, as can be seen in Fig. 4c. Therefore, normal and also larger registration errors up to about a centimeter are tolerable. More available training data would both enable to train a more complex discriminative model, and also to generate a more accurate probabilistic atlas. In this case, the accuracy of registering the atlas $G(\boldsymbol{t})$ to an image can become more important. It could be further improved using intensity-based registration techniques. It is also conceivable to cluster the patients, for instance based on their size, and to use a different atlas for each cluster, which could also improve the registration accuracy.

The computational requirements of the proposed methods are shown in Table 4. They were measured on a standard dual core PC with 2.67 GHz. Detecting the landmarks takes 7.0 s, and segmenting the heart and the esophagus takes 20.7 s. The proposed "soft" prior (type 3) takes 2.9 s to compute, which is about 23 times faster than computing the more complicated prior of Feulner et al. (2010). Detecting and segmenting the lymph nodes takes 26.0 s if the proposed graph cuts segmentation with a 6-neighborhood system is used. With a 26-neighborhood, this step takes considerably longer (75.2 s). In total, detecting and segmenting the lymph nodes from a CT volume image takes less than a minute (56.6 s) or 1 min 46 s depending on the neighborhood system.

Fig. 14 shows example detections on unseen data. The second column shows detection results along with the corresponding segmentations that were generated using the proposed graph cuts method and a 26-neighborhood. The small boxes indicate the center of a detected lymph node. Some segmentations do not have a visible detection because it lies in another slice. For comparison, the third column shows bounding boxes detected by the method of Feulner et al. (2010). The manual ground truth segmentations are shown in the fourth column. In rows (a–e), the proposed method properly detected and segmented the lymph nodes (second column). Row (f) shows examples of false positive detections. False positives lie especially on vessels, which can look similar to lymph nodes. The method of Feulner et al. (2010) detects the clearly visible lymph nodes and some of the less clearly visible ones. There

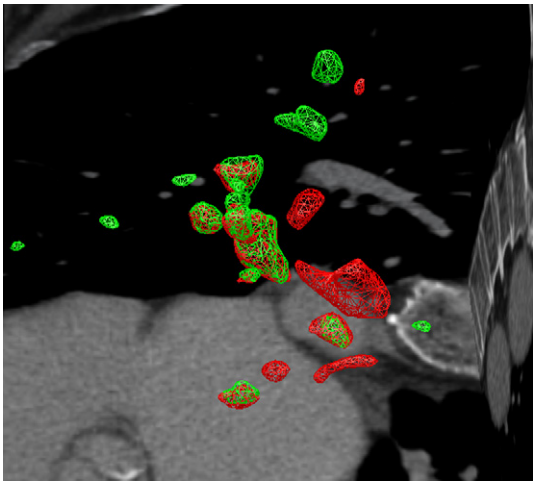

Fig. 13. Two manual segmentations from the same person.

Table 4
Computational requirements of the single steps of the presented method in seconds.

Landmark detection	7.0
Organ segmentation	20.7
Computing the prior ("soft", Eq. (4))	2.9
Detection and segmentation (6-neighborhood/26-neighborhood)	26.0/ 75.2
Total (6-neighborhood/26-neighborhood)	56.6/105.8

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

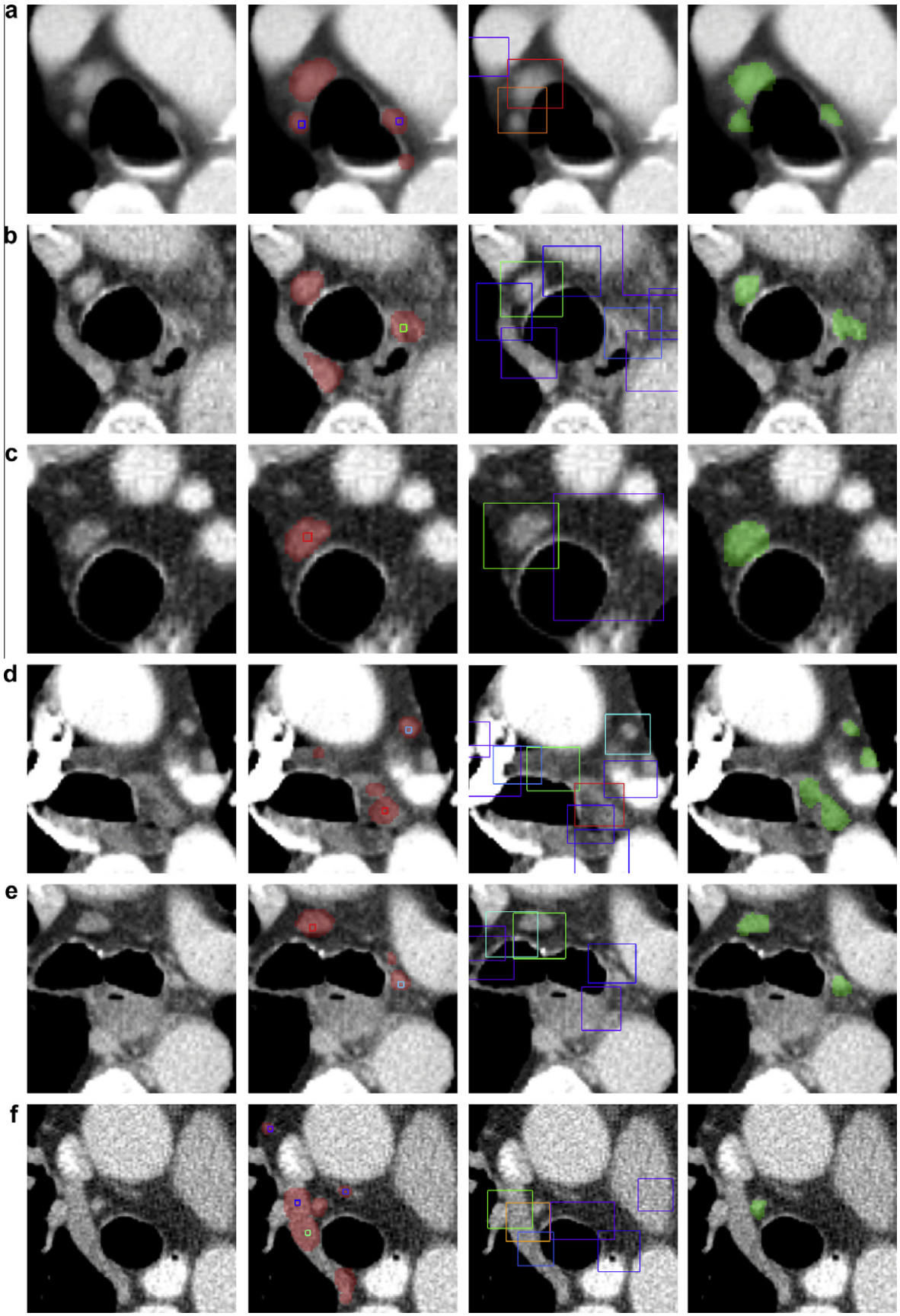

Fig. 14. Detection and segmentation examples on unseen data shown in 2-D. First column: plain CT slices. Second column: detections (small colored boxes) and resulting segmentations (red) of the proposed graph cuts based method with a 26-neighborhood. Third column: bounding boxes detected by the method described in Feulner et al. (2010) for comparison. In both columns, the detection score is color coded in HSV color space. Violet means lowest, red means highest score. Fourth column: manual ground truth segmentations (green).

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001

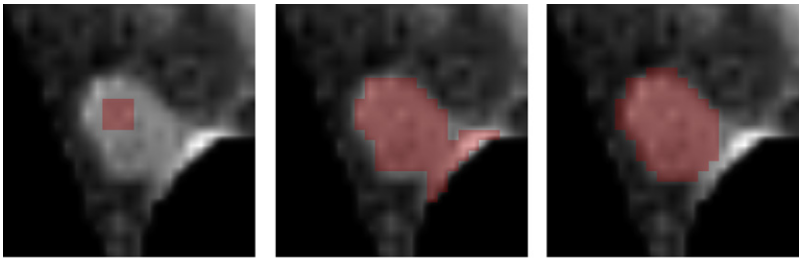

Fig. 15. Manually initialized segmentations with different edge capacities. (a) Standard graph cuts weights with $\sigma_\beta = 32$ HU. (b) $\sigma_\beta = 16$ HU. (c) Proposed graph cuts segmentation method.

are generally more false alarms compared to the proposed method, even though the false alarms mostly have a lower detection score than the true positives.

Fig. 15 shows 2-D slices of three example segmentations that were manually initialized with the same seed. In subfigure (a) and (b), the segmentation was done using graph cuts and standard weights (see Eqs. (15) and (16)). In (a), the parameter σ_β was set to 32 HU, and in (b), it was set to $\sigma_\beta = 16$ HU. A high value of σ_β allows voxel pairs across the object boundary that have a more similar attenuation coefficient. Thus, the surface of the cut becomes more important, and the segmentation is more likely to collapse. If, on the other hand, σ_β is set to a high value, the image is more likely to be cut at locations with a high gradient magnitude. However, the surface of the cut becomes less important, which can cause the segmentation to leak into neighboring structures. Subfigure (c) shows the segmentation result of our proposed segmentation method. The radial weighting prevents the segmentation from collapsing. But rugged segmentations are still penalized because of their larger surface. Thus, blob-like cuts are preferred, and the segmentation is less likely to leak into other structures. The neighborhood size is 26 in all three cases.

7. Conclusion

We have presented a method that automatically detects mediastinal lymph nodes in 3-D CT image data, which is a challenging problem due to low contrast to surrounding structures and clutter. We approach the problem from two sides: First, we heavily rely on prior anatomical knowledge, which is modeled as a spatial prior probability and learned from annotated data. Next, this is combined with a discriminative model of the lymph node appearance. A detector is trained that consists of multiple classifiers that are used in a cascade. In the first stages of the cascade, a set of lymph node center candidates is generated. Then, a detected center point serves as seed for segmenting the lymph node. A feature set is proposed that is extracted from the segmentation. It is used to train a classifier to learn whether the detected lymph node is a true or a false positive. Thus, the segmentation helps to improve the detection performance by rejecting false alarms. We propose to segment the lymph nodes using a graph cuts method adapted for blob-like structures that requires a single seed point as input. This is compared to standard graph cuts and also to a watershed segmentation.

Evaluation on 54 datasets showed that the spatial prior greatly improves the detection performance. When a fixed number of false alarms is allowed, the detection rate is well more than doubled when a prior is used. It also turned out that the simpler "soft" prior leads to a similar detection performance as the more complicated prior described in Feulner et al. (2010). We therefore propose using the simpler one that can also be computed about 12 times faster. The experiments further showed that the segmentation based verification step considerably reduces the number of false alarms. The best detection performance was achieved using the proposed graph cuts based segmentation method. Interestingly, the detection performance could not further be improved by generating additional watershed candidate segmentations for a detected lymph node center and joining the features from all segmentations. This indicates that the additional segmentations do not contain additional information about whether the detection is a true lymph node or not.

The proposed method can detect and segment lymph nodes with a TPR of 52.0% at the cost of 3.1 FP per volume image and with a TPR of 60.9% at 6.1 FP per volume within 56.6 s. This TPR is similar to the intra-observer variability of a human that has a TPR of 54.8% with, however, only 0.8 FP per volume. It also compares favorably to prior work on mediastinal lymph node detection.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.media.2012.11.001.

References

Barbu, A., Suehling, M., Xu, J., Liu, D., Zhou, S.K., Comaniciu, D., 2010. Automatic detection and segmentation of axillary lymph nodes. In: MICCAI (1), Lecture Notes in Computer Science, LNCS 6361, Beijing, China, pp. 28–36.

Bauer, C., Pock, T., Sorantin, E., Bischof, H., Beichel, R., 2010a. Segmentation of interwoven 3d tubular tree structures utilizing shape priors and graph cuts. Medical Image Analysis 14, 172–184.

Bauer, S., Seiler, C., Bardyn, T., Buechler, P., Reyes, M., 2010b. Atlas-based segmentation of brain tumor images using a markov random field-based tumor growth model and non-rigid registration. In: Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 4080–4083.

Beucher, S., 1994. Watershed, hierarchical segmentation and waterfall algorithm. In: Serra, J., Soille, P. (Eds.), Proc. Mathematical Morphology and its Applications to Image Processing, Computational Imaging and Vision, vol. 2. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 69–76.

Beucher, S., Meyer, F., 1992. Mathematical Morphology in Image Processing. In: The Morphological Approach to Segmentation: The Watershed Transformation, First ed. CRC Press, New York, NY, USA, pp. 433–482 (Chapter 12).

Bookstein, F., 1989. Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 11, 567–585.

Boykov, Y., Funka-Lea, G., 2006. Graph cuts and efficient n-d image segmentation. International Journal of Computer Vision 70, 109–131.

Boykov, Y., Kolmogorov, V., 2003. Computing geodesics and minimal surfaces via graph cuts. In: IEEE International Conference on Computer Vision (ICCV), Nice, France, pp. 26 –33.

Chen, T., Zhang, W., Good, S., Zhou, K.S., Comaniciu, D., 2009. Automatic ovarian follicle quantification from 3d ultrasound data using global/local context with database guided segmentation. In: IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, pp. 795 – 802.

Das, P., Veksler, O., Zavadsky, V., Boykov, Y., 2009. Semiautomatic segmentation with compact shape prior. Image and Vision Computing 27, 206–219.

Dornheim, J., Seim, H., Preim, B., Hertel, I., Strauß, G., 2006. Segmentation of neck lymph nodes in ct datasets with sTable 3d mass-spring models. In: MICCAI (2),

Lecture Notes in Computer Science, LNCS 4191, Copenhagen, Denmark, pp. 904–911.

Dornheim, L., Dornheim, J., 2008. Automatische detektion von lymphknoten in ct-datensätzen des halses. In: Bildverarbeitung für die Medizin. Informatik aktuell. Springer, Berlin, Germany, pp. 308–312, ISBN 978-3-540-78639-9.

Drake, R., Vogl, A.W., Mitchell, A.W.M., 2009. Gray's Anatomy for Students, second ed. Churchill Livingstone.

Duwe, B.V., Sterman, D.H., Musani, A.I., 2005. Tumors of the mediastinum. Chest 128, 2893–2909.

Feuerstein, M., Deguchi, D., Kitasaka, T., Iwano, S., Imaizumi, K., Hasegawa, Y., Suenaga, Y., Mori, K., 2009. Automatic mediastinal lymph node detection in chest ct. In: SPIE Medical Imaging, Orlando, Florida, USA, p. 72600V.

Feulner, J., Zhou, S.K., Cavallaro, A., Seifert, S., Hornegger, J., Comaniciu, D., 2009. Fast Automatic Segmentation of the Esophagus from 3D CT Data Using a Probabilistic Model. In: MICCAI (1), Lecture Notes in Computer Science, LNCS 5761, London, UK, pp. 255–262.

Feulner, J., Zhou, S.K., Hammon, M., Hornegger, J., Comaniciu, D., 2011. Segmentation based features for lymph node detection from 3-d chest ct. In: Machine Learning in Medical Imaging (MLMI), MICCAI workshops, Lecture Notes in Computer Science, LNCS 7009, Toronto, Canada, pp. 91–99.

Feulner, J., Zhou, S.K., Huber, M., Hornegger, J., Comaniciu, D., Cavallaro, A., 2010. Lymph node detection in 3-d chest ct using a spatial prior probability. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, pp. 2926–2932.

Funka-lea, G., Boykov, Y., Florin, C., p. Jolly, M., Moreau-gobard, R., Ramaraj, R., Rinck, D., 2006. Automatic heart isolation for ct coronary visualization using graph-cuts. In: IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, Arlington, VA, USA, pp. 614–617.

Gooya, A., Pohl, K., Bilello, M., Biros, G., Davatzikos, C., 2011. Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In: Fichtinger, G., Martel, A., Peters, T. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2011, Lecture Notes in Computer Science, vol. 6892. Springer, Berlin/Heidelberg, pp. 532–540.

Greig, D.M., Porteous, B.T., Seheult, A.H., 1989. Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society, Series B (Methodological) 51, 271–279.

Kitasaka, T., Tsujimura, Y., Nakamura, Y., Mori, K., Suenaga, Y., Ito, M., Nawano, S., 2007. Automated extraction of lymph nodes from 3-d abdominal ct images using 3-d minimum directional difference filter. In: MICCAI (2), Lecture Notes in Computer Science, LNCS 4792, Brisbane, Australia, pp. 336–343.

de Langen, A.J., Raijmakers, P., Riphagen, I., Paul, M.A., Hoekstra, O.S., 2006. The size of mediastinal lymph nodes and its relation with metastatic involvement: a meta-analysis. European Journal of Cardio-Thoracic Surgery 29, 26–29.

Liu, D., Zhou, K., Bernhardt, D., Comaniciu, D., 2010. Search strategies for multiple landmark detection by submodular maximization. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2831–2838.

McLoud, T.C., Bourgouin, P.M., Greenberg, R.W., Kosiuk, J.P., Templeton, P.A., Shepard, J.A., Moore, E.H., Wain, J.C., Mathisen, D.J., Grillo, H.C., 1992. Bronchogenic carcinoma: analysis of staging in the mediastinum with ct by correlative lymph node mapping and sampling. Radiology 182, 319–323.

Moon, N., Bullitt, E., van Leemput, K., Gerig, G., 2002. Model-based brain and tumor segmentation. In: 16th International Conference on Pattern Recognition, 2002, Proceedings, pp. 528 – 531.

Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., Gerig, G., 2003. Automatic brain tumor segmentation by subject specific modification of atlas priors. Academic Radiology 10, 1341–1348.

Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D., 2009. Hierarchical parsing and semantic navigation of full body ct data. In: SPIE Medical Imaging, Orlando, Florida, USA, p. 725902.

Sinop, A., Grady, L., 2007. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In: IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, pp. 1–8.

Slabaugh, G., Unal, G., 2005. Graph cuts segmentation using an elliptical shape prior. In: IEEE International Conference on Image Processing (ICIP), Genoa, Italy, pp. II–1222–5.

Suwatanapongched, T., Gierada, D.S., 2006. Ct of thoracic lymph nodes. Part I: Anatomy and drainage. British Journal of Radiology 79, 922–928.

Tu, Z., 2005. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: IEEE International Conference on Computer Vision (ICCV), Beijing, China, pp. 1589–1596.

Tu, Z., Zhou, X., Bogoni, L., Barbu, A., Comaniciu, D., 2006. Probabilistic 3d polyp detection in ct images: the role of sample alignment. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, pp. 1544–1551.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999a. Automated model-based tissue classification of mr images of the brain. IEEE Transactions on Medical Imaging 18, 897–908.

Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999b. Automated model-based tissue classification of mr images of the brain. IEEE Transactions on Medical Imaging 18, 897–908.

Veksler, O., 2008. Star shape prior for graph-cut image segmentation. In: European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science, LNCS 5304, Marseille, France, pp. 454–467.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Kauai, HI, USA, pp. 511 – 518.

Wang, S., Siskind, J., 2001. Image segmentation with minimum mean cut. In: IEEE International Conference on Computer Vision (ICCV), Vancouver, Canada, pp. 517–524.

Warwick, R., Williams, P.L., 1858. Gray's Anatomy, 35th ed. Longman, London, UK.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2007. Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features. In: IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, pp. 1–8.

Please cite this article in press as: Feulner, J., et al. Lymph node detection and segmentation in chest CT data using discriminative learning and a spatial prior. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.11.001