

Automatic Detection and Segmentation of Axillary Lymph Nodes

Adrian Barbu¹, Michael Suehling², Xun Xu², David Liu²,
S. Kevin Zhou², and Dorin Comaniciu²

¹ Statistics Department, Florida State Univ., Tallahassee, FL 32306, USA
abarbu@stat.fsu.edu

<http://stat.fsu.edu/~abarbu>

² Siemens Corporate Research, Princeton, NJ 08540, USA

Abstract. Lymph node detection and measurement is a difficult and important part of cancer treatment. In this paper we present a robust and effective learning-based method for the automatic detection of solid lymph nodes from Computed Tomography data. The contributions of the paper are the following. First, it presents a learning based approach to lymph node detection based on Marginal Space Learning. Second, it presents an efficient MRF-based segmentation method for solid lymph nodes. Third, it presents two new sets of features, one set self-aligning to the local gradients and another set based on the segmentation result. An extensive evaluation on 101 volumes containing 362 lymph nodes shows that this method obtains a 82.3% detection rate at 1 false positive per volume, with an average running time of 5-20 seconds per volume.

1 Introduction

Lymph node (LN) analysis is a difficult task and accounts for a significant part of daily clinical work in Radiology. In particular, automatic lymph node detection and segmentation is important for cancer staging and treatment monitoring. Lymph nodes nearby primary cancer regions are routinely assessed by clinicians to monitor disease progress and effectiveness of the cancer treatment. The assessment is usually based on 3D Computed Tomography (CT) data. When the cancer treatment is successful, the lymph nodes decrease in size. Since finding the lymph nodes is time consuming and highly dependent on the observer's experience, a system for automatic lymph node detection and measurement is desired. For follow-up studies, the system could further report the size change for each major lymph node.

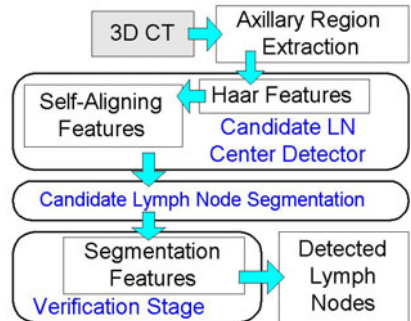


Fig. 1. Diagram of the axillary lymph node detection system

There is a limited amount of work directed to automatic lymph node detection [4,5,7]. These works target mediastinal [5], abdominal [7] and neck [4] lymph nodes while our work targets axillary lymph nodes. The axillary lymph nodes are far from airways or major vessels, so a segmentation of the vessels or airways is not necessary.

Previous work [5] relied on Hessian-based blobness measures and other filters for removing lymph node false positives. Our work comes from a learning perspective, where the appropriate features that discriminate the lymph nodes from the negatives are learned from training examples. Blobness measures could also be included in the feature set of our system.

The idea of coupling segmentation with detection has been recently proposed in the computer vision literature [8]. Our work also combines segmentation with object detection, but in a different way. First, our segmentation method produces a defined object boundary whereas [8] has a fuzzy boundary. Second, our work is oriented towards detecting 3D lymph nodes, which have a high degree of shape variability. In contrast, [8] detects 2D objects of specific shapes such as cars, cows and humans. Third, the segmentation is constructed differently in our work, using a Gaussian MRF and gradient descent as opposed to [8] where it is constructed in a greedy way from a number of patches.

Fourth, our work constructs segmentation-based features that are used to train a classifier, whereas [8] obtains a probability from the segmentation hypotheses by voting.

The diagram of the proposed solid lymph node detection and segmentation method is shown in Fig. 1. For speed and accuracy, the axillary regions are extracted automatically as described in Section 2.1. About 1500 lymph node center candidates per axillary region are generated using a two-stage detector described in Section 2.3. Each candidate is segmented as described in Section 2.4. Finally, a verification stage described in Section 2.5 gives the final result.

2 Proposed Method for Solid Lymph Node Detection and Segmentation

The proposed lymph node detection and segmentation system first detects candidate lymph node centers using a learning based approach. Each candidate is used by a segmentation module to extract a candidate lymph node boundary. A learning-based verification stage uses features obtained from the data and the extracted boundary to score the candidates and keep only the best ones.

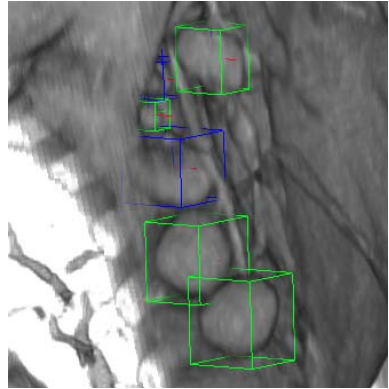


Fig. 2. The lymph nodes are marked with bounding boxes and labeled as solid (green) and non-solid (blue)

2.1 Axillary Region Extraction

To constrain the search, the two axillary regions are detected and cropped automatically; the axillary lymph node detection is performed on these two cropped subvolumes. The axillary subvolumes are obtained by first detecting the lung tips, with an approach similar to [9]. This can be done very reliably and is not the object of this paper. Relative to the two lung tip locations (x, y, z) , subvolumes of size $220 \times 220 \times 220$ voxels (at 1.5mm resolution) are cropped, with the upper-left corner at $(x + 20, y - 135, z - 131)$ for the left lung and upper-right corner at $(x - 20, y - 135, z - 131)$ for the right lung.

2.2 Axillary Lymph Node Annotation

All axillary lymph nodes of size at least 10mm have been annotated by placing bounding boxes around them, as shown in Figure 2. The lymph nodes are labeled as solid or non-solid depending whether they have a homogeneous interior or not. Enlarged lymph nodes with a solid interior are of particular clinical interest since they are believed to have a higher probability of being malignant than lymph nodes that for example have a fatty core.

2.3 Candidate Lymph Node Detection

Lymph node center candidates are detected in the axillary subvolumes ignoring the lymph node size, in the spirit of Marginal Space Learning [12]. Marginal Space Learning is an object detection technique where object candidates are first detected in a subspace where many object parameters (e.g. size, orientation, etc) are ignored. The candidates are refined by searching for the missing parameters using appropriate detectors.

The LN candidates are detected in two stages. The initial set of LN candidates are all voxels in the axillary subvolumes with intensity in the interval $[-100, 200]$ HU. Then, these candidates are evaluated using a fast detector based on Haar features followed by a second detector based on self-aligning gradient features. These two types of features are described below.

1. Haar-Based Features. The first stage of lymph node detection is a trained cascade of Adaboost classifiers trained using 92,000 3D Haar features [10].

2. Self-Aligning Features. The second stage uses a set of features self-aligned to high gradients. The features are computed based on rays casted in 14 directions in 3D space from each candidate location. These 14 directions are $(\pm 1, 0, 0)$, $(0, \pm 1, 0)$, $(0, 0, \pm 1)$, and also $(\pm 1, \pm 1, \pm 1)$. Of the 14 directions, 10 are shown in Fig. 3.

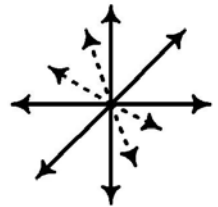


Fig. 3. Self aligning features are computed along 14 directions relative to candidate position

In each direction $d_i, 1 \leq i \leq 14$, local maxima of the gradient above each of 10 thresholds $\tau_j = 10j, 1 \leq j \leq 10$ (see Figure 4), are found at three scales $s_k = 1/2^k, 1 \leq k \leq 3$. Based on them, the following features are evaluated:

- Each of the 24 types of features types (gradient magnitude, angle, intensity value, etc) described in [12] at each of the first three local maxima for each d_i, τ_j, s_k .
- Each of the 24 types of features types computed half way between the candidate location and each of the first three local maxima, for each d_i, τ_j, s_k .
- The distance to each of the first three local maxima for each d_i, τ_j, s_k .
- The differences between distances to the corresponding first three local maxima in any combination of two different directions d_i, d_j for each τ_k, s_l .

This way about 64,000 features are obtained.

The best 1500 candidates above a threshold are kept for each axillary subvolume. An example of detected candidates is shown in Figure 8 (left). For each candidate location, a segmentation is obtained as described below.

2.4 Candidate Lymph Node Segmentation

The segmentation algorithm is specially designed for detecting clinically highly relevant solid lymph nodes. The solid lymph nodes have a blob-like shape that can be described by using a radial function $r : S^2 \rightarrow \mathbb{R}$ defined on the sphere in 3D, representing the distance from the lymph node center to the boundary in all directions. In this work, the sphere has been discretized using a triangulation with 162 vertices, 480 edges and 320 triangles, as shown in Figure 5. Example of lymph node segmentations with this sphere triangulation are shown in Fig. 8. This representation is similar to the shape representation in [3].

Each of the 162 sphere vertices represents a direction d_i . Given a candidate lymph node location C obtained by the candidate detector described in Section 2.3, a segmentation using this location as the center is determined by the radii $r_i, i = 1, \dots, N$ for all directions d_i , where $N = 162$ in our case. These radii form a vector $\mathbf{r} = (r_1, \dots, r_N)$.

The lymph nodes exhibit high shape variability, making it difficult to describe them using generative models such as PCA. To find the segmentation vector \mathbf{r} we adopt an approach similar to the Active Shape Models [2], but using a robust data cost, gradient optimization and a Gaussian MRF shape prior instead of a PCA model.



Fig. 4. In each direction, local gradient maxima above different thresholds τ_j are found

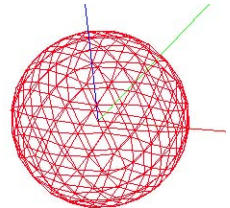


Fig. 5. Sphere triangulation with 162 vertices and 320 triangles

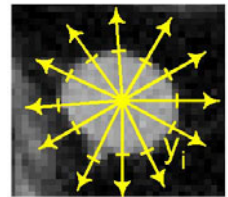


Fig. 6. Measurements y_i are found for each direction d_i as the most probable boundary location

Given the candidate location C , the most likely lymph node boundary locations $y_i, i = 1, \dots, 162$ are found in each direction d_i as

$$y_i = \arg \min_{r \in (0, R_{max})} |I(C) - I(C + (r+1)d_i)| > 50 \quad (1)$$

From the measurement vector $\mathbf{y} = (y_1, \dots, y_{162})$, the segmentation \mathbf{r} is obtained by minimizing the following energy function

$$E(\mathbf{r}) = \alpha \sum_i \rho(r_i - y_i) + \sum_i \frac{1}{2|\partial i|} \sum_{j \in \partial i} (r_i - r_j)^2 \quad (2)$$

where $\rho(x) = \ln(1 + x^2/2)$ and ∂i are the neighbors of i on the sphere mesh.

The first term is the data term, while the second term is the Gaussian MRF prior. If a measurement y_i does not exist, its corresponding term is removed from the data term of eq. (2).

Using the robust cost function ensures that any sporadic outliers in the measurements y_i are ignored. This is illustrated in Figure 7.

Minimization is done by gradient descent, starting with $\mathbf{r} = \mathbf{y}$ as initialization. The energy gradient can be computed analytically, obtaining the update iteration:

$$r_i \leftarrow r_i - \eta \left(\alpha \frac{r_i - y_i}{1 + (r_i - y_i)^2/2} + r_i - \frac{\sum_{j \in \partial i} r_j}{|\partial i|} \right) \quad (3)$$

In practice, we use $\eta = 0.1$ and 1000 gradient update iterations, while $\alpha = 1.6$.

Other segmentation methods such as [3,6,11] could possibly be used, but they lack a robust data term, making them more prone to oversegmentations in low gradient locations.

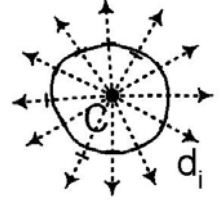


Fig. 7. The segmentation uses a robust cost and a Gaussian MRF to obtain a smooth result that fits most of the y_i

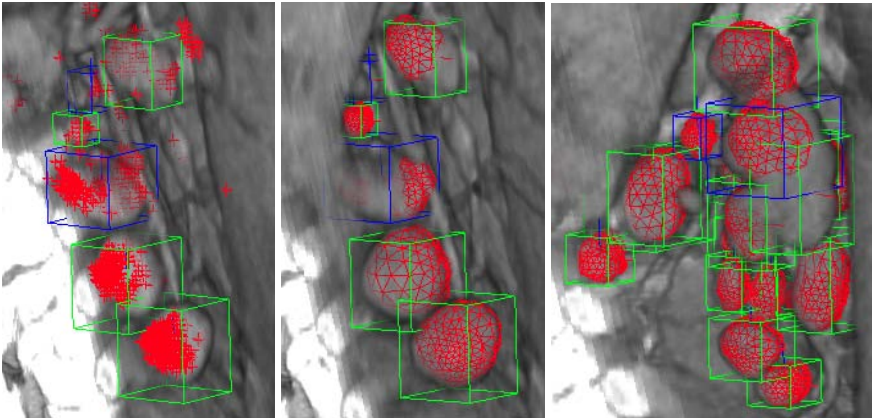


Fig. 8. Left: Detected candidates. middle, right: Detected lymph nodes. right: The method can also handle lymph node conglomerates.

2.5 Final Lymph Node Verification

For each of the candidate lymph node centers obtained using the Candidate detector from Section 2.3, a segmentation with 162 vertices is obtained as described in Section 2.4. The segmentation is used to obtain more informative features for the final evaluation of the lymph node candidates. From the segmentation, a bounding box is extracted for each lymph node. Candidates whose second largest bounding box size is less than 9mm are automatically rejected.

The following features are computed from the segmentation result

- Each of the 24 features types (gradient magnitude, angle, intensity value, etc) from [12] are computed at the 162 segmentation vertices. For each feature type, the 162 values are sorted in decreasing order.
- For each of the 24 feature types, the 81 sums of feature values at the pairs of opposite vertices are computed and sorted in decreasing order.
- The 81 diameters (distances between opposite vertices relative to the segmentation center) are sorted in decreasing order. For each diameter the following features are computed:
 1. The size of each diameter.
 2. Asymmetry of each diameter, i.e. the ratio of the larger radius over the smaller radius.
 3. The ratio of the i -th sorted diameter and the j -th diameter for all $1 \leq i < j \leq 81$.
 4. For each of the 24 feature types, the max or min of the feature values at the two diameter ends.
 5. For each of the 24 feature types, the max or min of the feature values half way to the diameter ends.

In total there are about 17,000 features.

The classifier trained with these features assigns a score p_i to each candidate i , a higher score meaning a higher likelihood to be a lymph node. All candidates with the score below a threshold τ are automatically removed. On the remaining candidates, a non-maximum suppression scheme is implemented as follows: The algorithm repeats adding the remaining candidate of highest score and removing all candidates close to it. In practice, we chose $N_{max} = 25$ so on each axillary region, a maximum of 25 lymph nodes are detected.

Algorithm 1. Non-maximal Suppression

Input: Candidates $c_i = (x_i, y_i, z_i)$ with scores $p_i > \tau$ and bounding boxes b_i .

Output: Set D of detected lymph nodes.

- 1: Find the candidate c_i with highest score p_i .
 - 2: **if** c_i exists **then** initialize $D = \{i\}$ **else** $D = \emptyset$, **stop**.
 - 3: **for** $n = 2$ **to** N_{max} **do**
 - 4: Remove candidates c_j inside any box $b_i, i \in D$.
 - 5: Find remaining candidate c_j of highest score p .
 - 6: **if** c_j exists **then** add j to detected set: $D \leftarrow D \cup \{j\}$ **else stop**.
 - 7: **end for**
-

Two examples of detected and segmented lymph nodes are shown in red in Figure 8, middle and right. The method can detect parts of lymph node conglomerates as shown in Figure 8, right.

Training details. The training parameters of the three classifiers are given in Table 1. For comparison, a Random Forest [1] classifier with 50 trees was also trained on the segmentation-based features for the verification step.

Table 1. Training details for the three classifiers

Classifier	Features	Type	# Weak	TPR	FPR
1	Haar	AdaBoost cascade	20,50	98%	1%
2	self-aligning	AdaBoost cascade	30,90,270	94%	0.7 – 1%
3	segmentation-based	AdaBoost	27		

3 Experimental Validation

The experiments are performed on a dataset containing 101 CT volumes. All volumes have been converted to 1.5mm isotropic voxel size. In the 101 volumes, a total of 362 solid lymph nodes and 323 non-solid lymph nodes have been found and annotated.

Out of the 101 cases, the region extraction failed only on the left side of one patient that actually had the left lung removed.

The experimental results below are based on a six-fold cross-validation. The CT volumes were divided into six disjoint sets. For each fold, the union of five of the sets was used to train the three classifiers and the remaining set was used for evaluation. Training all three classifiers for each fold took about five hours.

Evaluation Methodology. Since the solid lymph nodes are often very similar and difficult to distinguish from the non-solid ones, we adopted the following evaluation measure for the detection results. A solid lymph node is considered detected if there exists a detection with the center inside the lymph node bounding box. A detection is considered false positive if its center is not inside any annotated solid or non-solid lymph node. Thus any detection on a non-solid lymph node is neither considered a positive nor a false alarm.

Results. Using this evaluation measure, we obtained the ROC curves

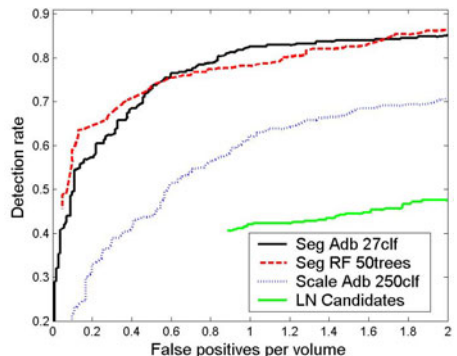


Fig. 9. Detection results with six-fold cross-validation on 101 volumes containing 362 lymph nodes

shown in Figure 9. The solid black curve represents the ROC curve of the system with the verification step having 27 Adaboost classifiers while the interrupted red curve being the ROC of the system with the verification step trained as a Random Forest with 50 trees. For comparison, we show in green color the ROC curve of the LN candidate detector. Thus, the verification step based on segmentation has a great impact on the overall performance of the system.

We also evaluated a system in which the segmentation and verification steps are removed and replaced with a lymph node detector that searches the scale of the lymph node bounding box. This detector is trained using steerable features, as described in [12]. For each of the lymph node candidates, this detector searches 890 combinations of the three lymph node sizes and reports the bounding box of highest score above a threshold. Non-maximal suppression as described in Algorithm 1 is used to further reduce the number of detections. The results using this “scale detector” are shown as a dotted blue curve in Figure 9. This shows that the segmentation-based detector has a much better performance than the scale detector. An added benefit is the fact that the segmentation based detector is about 5 times faster than the scale detector.

Table 2. Detection results and comparison with other methods

Method	Target Area	# cases	Size of nodes	TPR	FP/vol	PPV	Time/vol
Ours	Axillary	101	> 10.0mm	82.3%	1.0	74.9%	5-20sec
Feuerstein [5]	Mediastinum	5	> 1.5mm	82.1%	113.4	13.3%	1-6min
Kitasaka [7]	Abdomen	5	> 5.0mm	57.0%	58	30.3%	2-3h
Dornheim [4]	Neck	1	> 8.0mm	100%	-	76.3%	17min

A comparison with other lymph node detection methods present in the literature is shown in Table 2. Our method achieves a detection rate of 82.3% at 1 false positive per volume, i.e. a 74.9% Positive Predictive Value. This compares favorably with the previous work [5,7]. Dornheim[4] obtains a better detection rate but is evaluated on a single volume, which cannot be considered anywhere close to a thorough evaluation. Moreover, our method is also the fastest, because of the use of the lymph node center detector that ignores the lymph node size and shape and potentially eliminating millions of expensive verifications.

4 Conclusion and Future Work

In this paper, we presented a novel method for automated lymph node analysis based on integrating segmentation with a learning-based detector.

While we address the more restricted problem of solid axillary lymph node detection, we obtain better results with a more thorough evaluation (101 cases compared to 5 cases in [5,7]). At the same time, the proposed method is faster than any of the other existing lymph node detection methods.

In the axillary region, there are no airways or intestines to segment. Nevertheless, a vessel segmentation could further improve the accuracy of our approach.

In the future, we plan to study the improvement brought by using more than one segmentation at each candidate location. We also plan to use the proposed method for segmenting abdominal and mediastinal lymph nodes and also different types of lesions.

References

1. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
2. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., et al.: Active shape models—their training and application. *CVIU* 61(1), 38–59 (1995)
3. Dornheim, J., Seim, H., Preim, B., Hertel, I., Strauss, G.: Segmentation of Neck Lymph Nodes in CT Datasets with Stable 3D Mass-Spring Models Segmentation of Neck Lymph Nodes. *Academic Radiology* 14(11), 1389–1399 (2007)
4. Dornheim, L., Dornheim, J.: Automatische Detektion von Lymphknoten in CT-Datensätzen des Halses. In: *BVM* (2008)
5. Feuerstein, M., Deguchi, D., Kitasaka, T., Iwano, S., Imaizumi, K., Hasegawa, Y., Suenaga, Y., Mori, K.: Automatic mediastinal lymph node detection in chest CT. In: *SPIE*, vol. 7260, p. 30 (2009)
6. Kiraly, A.P., Naidich, D.P., Guendel, L., Zhang, L., Novak, C.L.: Novel method and applications for labeling and identifying lymph nodes. In: *SPIE* (2007)
7. Kitasaka, T., Tsujimura, Y., Nakamura, Y., Mori, K., Suenaga, Y., Ito, M., Nawano, S.: Automated extraction of lymph nodes from 3-d abdominal ct images using 3-d minimum directional difference filter. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) *MICCAI 2007, Part II. LNCS*, vol. 4792, p. 336. Springer, Heidelberg (2007)
8. Leibe, B., Leonardis, A., Schiele, B.: Robust object detection with interleaved categorization and segmentation. *IJCV* 77, 259–289 (2008)
9. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: *SPIE Medical Imaging* (2009)
10. Tu, Z., Zhou, X.S., Barbu, A., Bogoni, L., Comaniciu, D.: Probabilistic 3D polyp detection in CT images: The role of sample alignment. In: *CVPR* (2006)
11. Yan, J., Zhuang, T., Zhao, B., Schwartz, L.H.: Lymph node segmentation from CT images using fast marching method. *Computerized Medical Imaging and Graphics* 28(1-2), 33–38 (2004)
12. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE TMI* 27(11) (2008)