Automated temporal tracking and segmentation of lymphoma on serial CT examinations

Jiajing Xu^{a)}

Department of Electrical Engineering, Stanford University, Stanford, California 94305

Hayit Greenspan^{b)}

Department of Biomedical Engineering, Tel Aviv University, Tel Aviv, Israel 69978

Sandy Napel

Department of Electrical Engineering, Stanford University, Stanford, California 94305 and Department of Radiology, Stanford University, Stanford, California 94305

Daniel L. Rubin

Department of Radiology, Stanford University, Stanford, California 94305

(Received 18 May 2011; revised 26 August 2011; accepted for publication 30 August 2011; published 11 October 2011)

Purpose: It is challenging to reproducibly measure and compare cancer lesions on numerous follow-up studies; the process is time-consuming and error-prone. In this paper, we show a method to automatically and reproducibly identify and segment abnormal lymph nodes in serial computed tomography (CT) exams.

Methods: Our method leverages initial identification of enlarged (abnormal) lymph nodes in the baseline scan. We then identify an approximate region for the node in the follow-up scans using non-rigid image registration. The baseline scan is also used to locate regions of normal, non-nodal tissue surrounding the lymph node and to map them onto the follow-up scans, in order to reduce the search space to locate the lymph node on the follow-up scans. Adaptive region-growing and clustering algorithms are then used to obtain the final contours for segmentation. We applied our method to 24 distinct enlarged lymph nodes at multiple time points from 14 patients. The scan at the earlier time point was used as the baseline scan to be used in evaluating the follow-up scan, resulting in 70 total test cases (e.g., a series of scans obtained at 4 time points results in 3 test cases). For each of the 70 cases, a “reference standard” was obtained by manual segmentation by a radiologist. Assessment according to response evaluation criteria in solid tumors (RECIST) using our method agreed with RECIST assessments made using the reference standard segmentations in all test cases, and by calculating node overlap ratio and Hausdorff distance between the computer and radiologist-generated contours.

Results: Compared to the reference standard, our method made the correct RECIST assessment for all 70 cases. The average overlap ratio was $80.7 \pm 9.7\%$ s.d., and the average Hausdorff distance was 3.2 ± 1.8 mm s.d. The concordance correlation between automated and manual segmentations was 0.978 (95% confidence interval 0.962, 0.984). The 100% agreement in our sample between our method and the standard with regard to RECIST classification suggests that the true disagreement rate is no more than 6%.

Conclusions: Our automated lymph node segmentation method achieves excellent overall segmentation performance and provides equivalent RECIST assessment. It potentially will be useful to streamline and improve cancer lesion measurement and tracking and to improve assessment of cancer treatment response. © 2011 American Association of Physicists in Medicine. [DOI: 10.1118/1.3643027]

Key words: tracking, segmentation, region growing, computed tomography (CT), lymph node

I. INTRODUCTION

Lymphoma is the fifth most common cancer in the United States and accounts for about 8% of all cancer cases. Over the last 35 years, the annual incidence of lymphoma has nearly doubled. The American Cancer Society estimates that 74030 new cases of lymphoma will occur in 2010.¹ Lymphomas often manifest as enlarged lymph nodes and can spread throughout the lymphatic system to nearly anywhere in the body. Computed tomography (CT) has been predominately used for

detection, diagnosis, and staging. Precise evaluation of the temporal variation of lymph nodes on serial CT examinations is critical as tumor regression or progression with therapy is judged upon such changes in current clinical practice. Current practice consists of manual follow-up and measurement of the lymph nodes, which is time-consuming and imprecise.

Our goal is to develop a method to automatically identify and segment abnormal lymph nodes on serial imaging studies. Several approaches have appeared in the recent literature on tracking, segmenting, and analyzing several types of

abnormalities in medical-images. Many deal with lung nodules [e.g., (Refs. 2–6)], liver lesions [e.g., (Refs. 7–9)], and breast masses [e.g., (Refs. 10–12)]. To our knowledge, there is one main study that focused on tracking and segmentation of lymph nodes.^{13,14} In Ref. 13, Yan *et al.* utilized an automatic marker-controlled watershed approach to segment a selected lymph node. This approach requires the definition of internal and external markers. The selection of markers was shown to be important in the performance of the segmentation task. In Ref. 14, scans at two consecutive time points are used. The internal and external markers are then determined from the deformed contour from baseline scan, according to the registration of the baseline and the follow-up scans.

A challenge to automatically segmenting lymph nodes on serial images is that they are often surrounded by soft tissues that are hard to distinguish based on intensity value alone; thus, it can be difficult to reliably differentiate them from other anatomic structures. In Ref. 13, for example, it is discussed that errors will be introduced in the detection of the lymph node boundary in cases where very high (or low) intensity structures (i.e., bones, soft tissues) are included in the region between the internal and external marker.

Our approach, similar to Ref. 14, extracts information from radiologist-circumscribed baseline images to provide information about the location of lymph nodes in the follow-up images. To address the surrounding soft tissue challenge, we leverage a sufficient amount of structure-level information from the baseline scan to enable an informed restriction on the search region for the target lymph node in the new scan. Regions containing tissue surrounding the lymph node are extracted in the baseline scan and mapped onto the follow-up scan. Thus, we restrict the search region for the target lymph node, and improve the automatic segmentation results.

Section II explains in detail each of the steps in our algorithm. Section III presents the database and algorithm validation method. Section IV reports our experimental results. Finally, Sec. V summarizes and draws conclusions.

II. METHOD

Annotation of abnormal lymph nodes is the current practice by radiologists in cancer imaging studies. In the proposed methodology, enlarged lymph nodes are first identified and circumscribed by a radiologist in a single cross sectional slice in the baseline CT scan. We then apply our algorithm to automatically locate the abnormal lymph nodes in each follow-up scan. This is done by (1) identifying tissues surrounding the enlarged lymph nodes in the baseline scan; (2) registering the baseline scan, in which the enlarged lymph nodes are circumscribed, to the images in each follow-up scan, and (3) constraining the search region in the follow-up scan based on the locations of surrounding tissues in the prior CT study and segmenting the corresponding lymph nodes. We use the registration to project the radiologist-identified contour and the surrounding tissues on the baseline scan to the follow-up scan. While the projected contour may not be accurate, mainly due to the lymph node changes across the scans, it does provide us with good initial seed points. In contrast to

the lymph node, we do not expect the surrounding tissues to change significantly across scans. Thus, the projected surrounding tissues are used to limit the segmentation of the lymph node on the follow-up scan. Figure 1 shows a block-diagram of our proposed algorithm.

II.A. Identification of surrounding tissue in the baseline scan

We assume that the tissues in the vicinity of the enlarged lymph nodes such as muscle and blood vessels do not vary substantially between each consecutive pair of studies in a series of CT studies; this is realistic, given the proximity of scans in time during temporal studies in cancer (generally separated in time by weeks).

On the baseline CT image containing a target lymph node, we apply a clustering technique based on mean shift¹⁵ to group pixels into homogeneous regions, and provide a segmentation of the entire image. Mean shift is a robust unsupervised clustering method, which does not require prior knowledge of the number of clusters. It classifies the pixels into a few clusters according to the discontinuity of the intensity distribution. It locates local density maxima, and then, each pixel is replaced by the local maxima. From this procedure, many homogeneous clusters can be produced. In practice, the mean shift algorithm takes in the following parameters: the spatial bandwidth (h_s), the range bandwidth (h_r), and the size in pixels of the smallest allowed region. The spatial bandwidth defines a spatial search window, and the range bandwidth controls the gray-level intensity resolution. Based on empirical studies we have selected the following set of parameters, $h_s = 6$ and $h_r = 4$. To prevent over-segmentation, we set the smallest allowed output region

FIG. 1. Overview of the proposed algorithm.

to 10 pixels. Sensitivity to these choices will be discussed in the Results section.

Once the image has been segmented, we wish to extract a set of regions that surround the lymph node. These regions are characterized by the fact that they are spatially close to the lymph node and have similar intensity as the lymph node. We defined the following two criteria for identifying such regions: (1) the absolute difference between I_{mean} and the mean pixel intensity of the region is less than $w \cdot I_{\text{std}}$, where $w = 2$ is an intensity weighting factor, I_{mean} and I_{std} are the mean intensity and the standard deviation, respectively, of the pixels included in the lymph node in the baseline image; (2) the region contains a pixel that is less than d pixels away from the lymph node in the baseline images, where $d = 10$ is the minimum distance. The parameters d and w were chosen empirically. Sensitivity to these choices will be discussed in the Results section.

Figure 2 shows an example of (a) a baseline scan, (b) the segmented regions after mean-shift clustering, and (c) the regions satisfying the above two criteria. The identified regions in Fig. 2(c) would be the most difficult to distinguish without human interaction.

II.B. Alignment and registration

We adopted a two-step approach to automatically match the baseline and the follow-up images: a rigid registration in 3D to select three candidate slices, followed by a 2D nonrigid registration of the baseline slice with each of the three candidate slices. The image volumes of the baseline and follow-up scans were first registered using an affine registration, maximizing the normalized mutual information (NMI). For this we used the Image Registration Toolkit (IRTK) (Ref. 16). Various parameters, required for IRTK, were set as follows: A hierarchical coarse-to-fine approach was used with control point spacings of 12, 6, 3, and 1.5 mm. At each stage, the output transformation of the previous step was used as a starting point. The stopping condition for the optimization was either no further improvement in NMI or the reaching of a maximum number of six iterations at each level. A third stopping criterion was based on the maximum amount of displacement across all control points in the most recent iteration. This was set to 1/10th of the control point spacing and was used to stop the optimization when iterations made only small updates to the transformation. Following the 3D volume registration,

we go back to the slice of interest in the baseline scan and extract the three closest slices in the follow-up volume.

The next step is to apply the nonrigid image registration algorithm to the baseline scan and each of the three candidate follow-up scans. The registration finds the optimized geometric transformation T that aligns the baseline image (I) and follow-up image (J) under a similarity measure. We used residual complexity (RC) proposed by Myronenko as the similarity measure.¹⁷ This measure favors a registration with minimum compression complexity of the difference (residual) image between the two registered images; empirical tests show that RC outperforms other similarity measures, including sum-of-squared-differences (SSD), mutual information (MI), correlation ratio (CR), and correlation coefficients (CC) (Ref. 17). The transformation T is modeled using free form deformation (FFD) transformation with three hierarchical levels of B-spline control points¹⁶ and can be obtained by minimizing the objective function E using the gradient descent optimization method:

$$c = \text{dctn}(r); r = I - J(T);$$

$$E = \sum \log\left(\frac{c^2}{\alpha} + 1\right);$$

$$\nabla E = -\text{idctn}\left(\frac{2c/\alpha}{c^2/\alpha + 1}\right) \nabla J(T) \frac{\partial T}{\partial \theta}$$

where $\text{dctn}(\cdot)$ and $\text{idctn}(\cdot)$ are the forward and inverse multi-dimensional discrete cosine transforms, ∇J is the intensity image gradient and θ represents the transformation parameters. In our implementation, we used the default set of parameters as in Ref. 17.

In the final step, we measured the mean square error (MSE) between each of the registered candidate slices and the baseline scan. We select the slice of minimum MSE as the corresponding slice in the follow-up scan.

II.C. Segmentation of lymph node in the follow-up scan

Lymph nodes usually are so close to organs and muscle that no clear border can be drawn without anatomic knowledge. The slice selection step as well as the nonrigid alignment between the original scan and follow-up scan, enables

FIG. 2. (a) Cropped CT image from baseline scan with lymph node outlined; (b) Regions formed by applying mean shift clustering (each region is colored with the same grayscale value); (c) Regions satisfying conditions listed in Section II A and therefore identified as containing tissues surrounding the lymph node. (All CT images are shown at window width of 400 Hounsfield units (HU) and window level of 35 HU)

us to project regions extracted in the original scan to the previously unseen scan. We focus on projecting regions that originate from anatomical structures and healthy tissue surrounding the abnormal lymph node tissue, the surrounding tissue extracted in Sec. II A. This step in the algorithm helps to disambiguate abnormal lymph nodes from normal anatomic structures. Following the nonrigid registration step, the regions that are identified in the prior CT scan (Sec. II A) are superimposed on the images in the follow-up scan. We then compute a binary mask M_r , in which the voxels with the value of 0 represent the voxels that belong to tissue surrounding the lymph node, and the voxels with value of 1 correspond to the voxels that do not.

We further restrict the spatial search for abnormal lymph nodes by using the pixel intensity distribution of the lymph node in the baseline CT images. Based on the fact that the intensities of a lymph node on the baseline and follow-up scans should not differ substantially, we constructed another binary mask M_p , where the voxels with the value of 1 represent the

voxels whose intensities fall into the range from $(I_{\text{mean}} - w \cdot I_{\text{std}})$ to $(I_{\text{mean}} + w \cdot I_{\text{std}})$, whereas the voxels with value of 0 correspond to the voxels whose intensities are beyond the range and $w = 2$ is an intensity weighting factor. Then, we combine the two masks to one mask: $M = M_r \wedge M_p$, which reduces the search region to the voxels having the value of 1 in M .

Figure 3 shows an example of (a) a baseline scan; follow-up scan (b) before and (c) after registration; (d) the tissue mask M_r ; (e) the intensity mask M_p ; (f) the final restricted search space extracted based on a combination of the two masks.

Let O represent the set of lymph node pixels on the baseline scan. After registration, O is deformed to O' on the follow-up scan. We next compute a new set of pixels S using an n -fold erosion:

$$S = O' \ominus nB$$

where $\ominus$ is a binary morphological erosion and B is a disk-shaped structuring element with a radius of 3 pixels. The scaling term n is selected so that nB is the largest that makes

FIG. 3. (a) Cropped baseline CT scan; (b) Cropped follow-up CT scan; (c) Deformed baseline CT scan following nonrigid registration; (d) Binary mask M_r (0-black, 1-white) (e) Binary intensity mask M_p (0-black, 1-white) (f) Masked version of the follow-up scan after applying mask M ; pixels in the search region are nonblack. (g) Set of pixel, S_2 , after region growing and taking the convex hull (h) S_2 after k-means clustering (All CT images are shown at window width of 400 HU and window level of 35 HU)

S contains at least 5 pixels. Then, we apply a region-growing algorithm¹⁸ with the seed pixels in set S . We define a similarity threshold t , for which only pixels whose intensity values fall into the range from $(I_{\text{mean}} - t)$ to $(I_{\text{mean}} + t)$ will be considered, where I_{mean} is the mean pixel intensity values of the growing region at each iteration. This similarity threshold is adaptively chosen after the following morphological operations. Let S_1 be the resulting set of pixels from the region-growing, and S_2 be the convex hull of S_1 , shown in Fig. 3(g).

We next apply the K-means clustering algorithm to S_2 . Here, we set K to 3, because usually the region growing will include some nonlymph-node background pixels and we would like to distinguish the core and rim part of the lymph node. The method appeared to correctly mark both the core and the surrounding tissue as the lymph node, shown in Fig. 3(h). Let L represent the resulting set of pixels marked as lymph node. Finally, a sequence of morphological operations is used to smooth the boundary along the set L since lymph node tends to have smooth boundary. An opening and closing operation is applied to remove any irregularity along the boundary.

$$L^1 = (L \ominus SE) \oplus SE$$

$$L^2 = (L^1 \oplus SE) \ominus SE$$

where $\oplus$ is a binary morphological dilation, SE is a four-connected (diamond-shaped) binary structuring element. We then trace the boundary of L^2 and compute the compactness of the boundary¹⁹ with respect to the similarity threshold t in the region growing step, defined as $C(t) = A/p^2$, where A is the area and P is the perimeter. The optimal similarity threshold T_{opt} is chosen to maximize $C(t)$ as follows,

$$T_{\text{opt}} = \arg \max_{0 \leq t \leq 0.5} C(t)$$

The final segmentation is obtained by using this optimal similarity threshold and is optimized to have a circular and smoothed boundary. The outline of the lymph nodes was then computed using the Matlab trace_boundary function. Figure 4(c) illustrates an example of the final smoothed segmentation, in which an optimal threshold of 15% is used.

III. ALGORITHM EVALUATION

Patient image data were reviewed retrospectively from clinical trials conducted at the Stanford Cancer Center. The images were acquired on a LightSpeed CT scanner (GE Medical Systems, Milwaukee, Wis) with the slice thickness of 5.0mm. We obtained CT scans for 24 distinct lymph nodes from 14 patients at 3 or 4 different time points. The time interval between two adjacent scans averaged 12 weeks (range = 9 weeks to 15 weeks). For any pair of scans, we considered the scan in the earlier time point as baseline scan and the scan at the next time point as the follow-up scan, which results in 70 total test cases.

In each test case, a radiologist manually outlined the lymph node in the most representative cross-section of the

Fig. 4. (a) Cropped baseline CT scan; [scan parameters: 432 mA, 120 kVp] (b) Cropped follow-up CT scan with deformed outline O' (white outline); [scan parameters: 440 mA, 120 kVp] (c) Final segmentation result (white outline); (d) Reference standard segmentation of lymph node (white outline). (All CT images are shown at window width of 400 HU and window level of 35 HU)

lymph node in the baseline and follow-up CT scans. The marked contour in the baseline scan was used for the initial position and the one in the follow-up scan is used as the “reference standard.”

We evaluated the performance of our segmentation algorithm according to the response evaluation criteria in solid tumors (RECIST) criteria^{20,21} and several quantitative metrics. RECIST uses measurements of the sum of the longest lesion diameters (SLD), which we automatically calculated from the outline in both baseline and follow-up scans. There are two versions of RECIST; RECIST 1.0 (Ref. 20) considers lesions to be “measurable” if they exceed 10 mm in size, whereas in RECIST 1.1,²¹ measurable lesions are considered to be 15 mm or greater. A code for disease response was given according to the change in the SLD from baseline scan to follow-up scans, i.e., in RECIST 1.0, “partial remission (PR)” for a decrement of at least 30% in SLD, “progressive disease (PD)” for an increment of at least 20%, “stable disease (SD)” for neither sufficient shrinkage qualify for PR nor sufficient increase to qualify for PD, and “complete response (CR)” for disappearance of all target lesions. We compared the RECIST 1.0- and 1.1-based assessments of lymph node response computed from our segmentation to that computed from the “reference standard.” We also assessed quantitative measurements: contour distance-based measurements such as Hausdorff distance,²² and area matching-based measurements, e.g., over-estimation and under-estimation of the area our segmentation compared to the “reference standard.”^{10,14} Agreement between our segmentation and the manual “reference

TABLE I. Metrics for the performance of the proposed algorithm based on comparison of the computer segmentation and the “reference standard.”

	Hausdorff distance (mm)	Overlap ratio (%)	Over-estimated ratio (%)	Under-estimated ratio (%)
Mean	3.18	80.7	9.8	19.4
s.d.	1.82	9.7	16.7	10.2
Min	0.79	55.0	0.0	3.0
Max	12.79	99.0	74.0	40.0

standard” was assessed by concordance correlation²³ and the Bland-Altman analysis²⁴ of the change in SLD. We compared the percentage changes in SLD from baseline to “reference standard” and automated segmentation. Statistical analysis was done using R.

IV. RESULTS AND DISCUSSION

In each of 70 test cases, we applied our algorithm to locate and segment the lymph nodes in the follow-up scan. Lymph node sizes ranged from 0.6 to 4 cm with a mean of 1.8 cm on baseline scan; 0.5–3.9 cm with a mean of 1.9 cm on follow-up scan. All 70 test cases were successfully detected using our method, meaning that the correct slice containing the target lymph node was located by the algorithm and the deformed lymph node after registration overlapped with the target lymph node in the follow-up scan by at least one pixel. Table I shows quantitative segmentation results for these 70 lymph nodes. Bland-Altman analysis (Fig. 5) shows consistent accuracy across varying change in SLD, with a mean difference of 1%, and 95% limits of agreement of 8%. Given that RECIST has a wide range for each category, the misassignment is unlikely to happen under the RECIST criteria. The concordance correlation between changes in SLD from baseline to “reference standard” and automated segmentation was 0.953 (95% confidence interval 0.926, 0.969). On the

conservative assumption that RECIST concordance occurs by chance with a probability of 50%, then our observation of 0 discordances in 64 cases means that the 95% exact binomial upper bound for the true discordance rate is 5.6%.

Figure 6 shows one representative result of our method. In this example, the Hausdorff distance was 2.79 mm. The overlap, overestimation, and underestimation ratios were 88.0, 1.0, and 11.0%, respectively.

Radiologist generally consider lymph nodes larger than 1 cm to be pathologically enlarged and suspicious of invasive disease.²⁵ We thus compared the segmentation performance for both small and large lymph nodes in Table II. Small lymph nodes are defined as having longest diameter less than 1 cm, while large nodes have longest diameter greater than 1 cm.

All measurable lymph nodes resulted in the same RECIST diagnosis as the “reference standard” according to both RECIST 1.0 and RECIST 1.1 criteria. In our data, 64 out of 70 were measurable by RECIST 1.0 criteria, and 61 out of 70 were considered to be measurable lesions based upon RECIST 1.1 size criteria.

For a typical test case used in this study, the procedure including registration, recognition, and segmentation required approximately 90 s on average using a PC computer (Intel Core 2 Dual 2.20 GHz with 4 GB of RAM) with the majority of the computation time spent on the registration.

IV.A. Parameter evaluation

The values of all parameters used in this work were pre-defined by studying two lymph nodes that were not included in the 70 test cases. These parameters were then kept constant for all test cases. We assessed the sensitivity of the parameters in the following experiments.

In Sec. II A, two parameters, spatial (h_s) and range (h_r) bandwidth, were used in the mean shift algorithm. Here, we

FIG. 5. Bland-Altman plot showing the mean difference in the change of SLD from baseline to manual and automated segmentation (solid middle line) and 95% limits of agreement (dashed lines) for all 64 cases measurable under RECIST 1.0.

FIG. 6. Automated segmentation example. (a) Cropped baseline scan (lymph node outlined in white by radiologist); [scan parameters: 330 mA, 140 kVp] (b) Automated segmentation in the follow-up scan (outlined in white); [scan parameters: 367 mA, 120 kVp] (c) Radiologist's "reference standard" in the follow-up scan. (All CT images are shown at window width of 400 HU and window level of 35 HU)

investigated the overlap ratio with respect to h_s and h_d . We varied h_s and h_d from 2 to 20 with an interval of 2 individually. Figure 7 show how the average overlap ratio over 70 test cases varies with different h_s 's and h_r 's. From Fig. 7(a) we can see that the mean overlap ratio is relatively constant with h_s , with only slightly worse performance at the extreme edges of scale. In Fig. 7(b), the performance curves are quite flat with respect to h_r , suggesting that the algorithm is not sensitive to this parameter within the given test range. We used $h_s = 6$ and $h_r = 4$ pixels in our experiments.

In Sec. II A, two parameters were used to locate surrounding tissues in the baseline scan (i.e., an intensity weighting factor w , and a minimum distance d from the lymph node to surrounding tissues). Similarly, we investigated the overlap ratio with respect to w and d . We varied w from 1 to 3 with an interval of 0.5, and d from 5 to 15 with an interval of 2.5 pixels. Figure 8 shows how the average overlap ratio over 70 test cases varies with different w 's and d 's. From Fig. 8(a) we can see that the mean overlap ratio is relatively constant with w , with only slightly worse performance at the extreme edges of scale. In Fig. 8(b), the performance curves are quite flat with respect to d , suggesting that the algorithm is not sensitive to this parameter within the given test range. We used $w = 2$ and $d = 10$ pixels in our experiments.

The above experiments obtained by changing values of the parameters concluded that our segmentation algorithm is insensitive to the parameters predetermined empirically.

IV.B. Limitations

This work has several limitations. First, our algorithm relies on the assumption that the tissue surrounding the

enlarged lymph nodes does not vary substantially between consecutive pair of studies and thus can serve as localizing landmarks to detect the lymph nodes being tracked. This assumption may be violated if a patient has surgery in between the studies, if there is substantial change in size of lymph nodes that grossly distort the surrounding tissues, or other conditions where the surrounding structures change (e.g., inspiration deformation, intake of food). Based on our experience to date, however, such cases are likely infrequent. In those cases, when the assumption is not valid, the registration algorithm will fail to find the correct transformation and produce an incorrect segmentation; however, this can be readily detected and corrected by the radiologist. Similarly, if for some other reason, the deformable registration is poor, then we could fail to identify lymph nodes. But in our observation, the surrounding normal tissue changed little, whereas the lymph nodes tend to change. Thus, our strategy of locating normal tissues for deformable registration seems advantageous. Secondly, our algorithm was developed to segment lesion in 2D, though we expect it would be applicable to 3D if thin slices are acquired. We believe our method is applicable in thinner slices and therefore will be useful for estimating 3D volumes, though validation of this will require future studies. An additional limitation is that our algorithm currently cannot detect morphological changes of the lymph node, such as merging, splitting, and disappearing. Currently,

TABLE II. Comparison of performance for small and large lymph nodes.

	Hausdorff distance (mm)	Overlap ratio (%)	Over-estimated ratio (%)	Under-estimated ratio (%)
6 small lymph nodes (diameter < 1 cm)	2.04	76.8	15.2	23.2
64 large lymph nodes (diameter > 1 cm)	3.34	81.3	9.1	18.9

FIG. 7. Mean overlap ratio as a function of (a) scale bandwidth h_s and (b) range bandwidth h_r .

FIG. 8. Mean overlap ratio as a function of (a) intensity weighting factor w and (b) minimum distance d .

we used hand circumscription of the abnormal lymph nodes seen on baseline images, but our method is applicable to any type of initial segmentation in the baseline scan, such as automated segmentation.

V. CONCLUSION

In this study we presented an automated method for tracking and segmenting lymph nodes on serial CT scans. Given that the spatial relationships among structures in images obtained over time in the same patient will generally be similar, comparable regions in each follow-up scan can be useful in enabling the tracking task by identifying and registering each follow-up imaging study to the immediately-prior imaging study. Our automated lymph node segmentation method achieved good overall segmentation performance, and it potentially will be useful to streamline and improve cancer lesion measurement and tracking, and improve assessment of cancer treatment response.

ACKNOWLEDGMENT

This project has been funded from National Cancer Institute, National Institutes of Health, under grant U01CA142555-01.

^{a)} Author to whom correspondence should be addressed. Electronic mail: jiajing@stanford.edu

^{b)} Current Address: Visiting Professor, Stanford and Visiting Scientist, IBM Almaden Research Center, San Jose, California 95120

¹ "Cancer Facts and Figures 2010," American Cancer Society (2010).

² J. Pu, B. Zheng, J. K. Leader, C. Fuhrman, F. Knollmann, A. Klym, and D. Gur, "Pulmonary Lobe Segmentation in CT Examinations Using Implicit Surface Fitting," *IEEE Trans. Med. Imaging* **28**, 1986–1996 (2009).

³ B. Golosio, G. L. Masala, A. Piccioli, P. Oliva, M. Carpinelli, R. Cataldo, P. Cerello, F. De Carlo, F. Falaschi, M. E. Fantacci, G. Gargano, P. Kasae, and M. Torsello, "A novel multithreshold method for nodule detection in lung CT," *Med. Phys.* **36**, 3607–3618 (2009).

⁴ J. G. Goldin, M. S. Brown, and I. Petkovska, "Computer-aided Diagnosis in Lung Nodule Assessment," *J. Thorac. Imaging* **23**, 97–104 (2008).

⁵ E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," *Med. Phys.* **36**, 2934–2947 (2009).

⁶ B. Zhao, L. H. Schwartz, C. S. Moskowitz, M. S. Ginsberg, N. A. Rizvi, and M. G. Kris, "Lung cancer: Computerized quantification of tumor response—Initial results," *Radiology* **241**, 892–898 (2006).

⁷ H.-c. Cho, L. Hadjiiski, B. Sahiner, H.-P. Chan, M. Helvie, C. Paramagul, and A. V. Nees, "Similarity evaluation in a content-based image retrieval (CBIR) CADx system for characterization of breast masses on ultrasound images," *Med. Phys.* **38**, 1820–1831 (2011).

⁸ C. Platero, J. M. Poncela, P. Gonzalez, M. C. Tobar, J. Sanguino, G. Asensio, and E. Santos, "Liver segmentation for hepatic lesions detection and characterisation," in *ISBI 2008, 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (Paris, France, 2008), pp. 13–16.

⁹ C. Hansen, S. Zidowitz, M. Hindennach, A. Schenk, H. Hahn, and H.-O. Peitgen, "Interactive determination of robust safety margins for oncologic liver surgery," *Int. J. Comput. Assist. Radiol. Surg.* **4**, 469–474 (2009).

¹⁰ E. Song, L. Jiang, R. Jin, L. Zhang, Y. Yuan, and Q. Li, "Breast mass segmentation in mammography using plane fitting and dynamic programming," *Acad. Radiol.* **16**, 826–835 (2009).

¹¹ A. Mencattini, G. Rabottino, M. Salmeri, R. Lojacono, and E. Colini, "Breast mass segmentation in mammographic images by an effective region growing algorithm," in *Advanced Concepts for Intelligent Vision Systems*, Vol. 5259, J. Blanc-Talon, et al., eds. (Springer Berlin/Heidelberg, 2008), pp. 948–957.

¹² J.-H. Chung, V. Rajagopal, T. A. Laursen, P. M. F. Nielsen, and M. P. Nash, "Frictional contact mechanics methods for soft materials: Application to tracking breast cancers," *J. Biomech.* **41**, 69–77 (2008).

¹³ J. Yan, B. Zhao, L. Wang, A. Zelenetz, and L. H. Schwartz, "Marker-controlled watershed for lymphoma segmentation in sequential CT images," *Med. Phys.* **33**, 2452–2460 (2006).

¹⁴ J. Yan, B. Zhao, S. Curran, A. Zelenetz, and L. H. Schwartz, "Automated matching and segmentation of lymphoma on serial CT examinations," *Med. Phys.* **34**, 55–62 (2007).

¹⁵ D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 603–619 (2002).

¹⁶ D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans. Med. Imaging* **18**, 712–721 (1999).

¹⁷ A. Myronenko and X. Song, "Intensity-based Image Registration by Minimizing Residual Complexity," *IEEE Trans. Med. Imaging* **29**, 1882–1891 (2010).

¹⁸ R. Adams and L. Bischof, "Seeded Region Growing," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 641–647 (1994).

¹⁹ R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, New York, 2001).

²⁰ P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, M. C. Christian, and S. G. Gwyther, "New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada," *J. Natl. Cancer Inst.* **92**, 205–216 (2000).

²¹ E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancy, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij, "New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)," *Eur. J. Cancer* **45**, 228–247 (2009).

²² V. Chalana and Y. Kim, "A methodology for evaluation of boundary detection algorithms on medical images," *IEEE Trans. Med. Imaging* **16**, 642–652 (1997).

²³ L. I. K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics* **45**, 255–268 (1989).

²⁴ J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet* **1**, pp. 307–310 (1986).

²⁵ H. Kramer and H. J. Groen, "Current concepts in the mediastinal lymph node staging of non-small cell lung cancer," *Ann. Surg.* **238**, 180–188 (2003).