Robust lung nodule growth measurement by combining registration and segmentation

Sven Kabus[1], Florian Müller[2], Rafael Wiemker[1], and Bernd Fischer[3]

[1] Philips Research Europe – Hamburg, Germany
sven.kabus@philips.com
[2] Institute for Signal Processing, University of Lübeck, Germany
[3] Institute of Mathematics, University of Lübeck, Germany

Abstract. Growth assessment for lung nodules is known to be the most relevant clinical parameter to distinguish malignant nodules from benign ones. The assessment of growth is usually done by multiple segmentations in follow-up CT examinations of a patient. Each segmentation is performed separately yielding a volume number assigned to the nodule at a certain point in time. Experiments have shown that the segmented volume may change due to small variations in certain voxel values. To circumvent this ill-posed problem, we present a growth assessment scheme which combines the result of an elastic registration and of a segmentation method. Regarding the absence of publicly available ground truth data, a method for the creation of artificial nodules is developed. Experiments show that the described combined growth assessment method leads to similar or better results than a growth assessment method achieved by a segmentation method only.

1 Introduction

The differential diagnosis of lung nodules is an essential step in the early detection of lung cancer. Besides shape properties, the most important criterion for the distinction between malignant and benign nodules is the growth rate. The growth assessment is usually done by comparing follow-up CT examinations of a patient, typically with a difference of three to six months. Currently, growth assessment schemes consist of a matching of the nodule positions in a first step, followed by a (semi-)automatic segmentation of the (same) nodule in both examinations as a second step, and, finally, a comparison of the two segmented volumes from which a change in volume can be computed. Of particular importance is the fact that each segmentation of the same nodule is performed *independently* from the other one. Experiments with follow-up data have shown that the segmented volumes may vary due to small variations in certain voxel values, e.g., if a nodule is attached to a vessel in the follow-up scan but not in the baseline scan. In this case a leakage of the follow-up segmentation into the vessel is reported [1]. To circumvent unwanted leakage into adjacent structures, detection of lung wall and vessel structures can be employed [2]. However, from the view of a vessel structure, the segmentation scheme may be faced again with

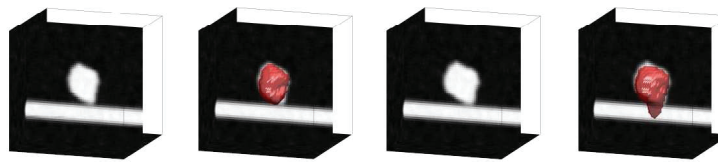

Fig. 1. Illustration of segmentation as an ill-posed problem: While the segmentation of an artificial vessel-attached nodule (left) is successful (center left), a slight modification at the nodule-vessel interface (center right) changes the segmentation result significantly (right).

a leakage problem, here from the vessel into the nodule. In particular, nodules which are heavily vascularized by vessels are affected. The leakage problem is even more prominent for small nodules, since a vessel-attached nodule with a similar or smaller diameter than the vessel diameter is likely to be characterized as protuberance of the vessel. From a mathematical point of view the segmentation of follow-up scans can therefore be expressed as an ill-posed problem, since small variations in the input data (the examinations themselves) may lead to large variations in the output data (the segmented volumes), cf. Figure 1.

In contrast to a segmentation, a regularized registration scheme is known to be a well-posed problem [3]. Its application to lung nodules enables for discrimination between expanding and shrinking lesions, for instance, by visually interpreting the displacement field and/or the according Jacobian map [4]. For a quantitative analysis we propose to combine the registration with a segmentation of the nodule which is required for the baseline scan only. The displacement field is then evaluated within the nodule, resulting in a single growth factor. Here, we compare the popular Jacobian map with two alternative criteria: the divergence of the displacement field and the quotient of the volume of the original and the deformed segmentation.

Validation should be based on ground truth data which, at first, are generated by multiple readers, at second, deal with follow-up data, and at third, are publicly available. Multiple reader delineations are, for instance, provided by the lung image database consortium (LIDC) [5]. However, no follow-up data are included in the LIDC database. Existing follow-up data known to us either originate from so-called coffee-break examinations (inter-examination time range from several minutes up to a few hours, therefore the nodule volume is assumed to not have changed between the scans) or are delineated by a single radiologist only, cf., e.g., [2]. Nevertheless, multiple reader delineations are advantageous, since inter-observer variability can be taken into account. For the LIDC data, an assessment revealed a median volumetric deviation of each delineation from the averaged delineation by 14% [6]. To our knowledge, there is no publicly available screening study including both follow-up data and delineations by multiple readers. Therefore, for validation and for comparison with a standard segmen-

tation method, artificial nodules are generated – either embedded into noisy background or into real-world data.

For the registration we use an elastic model based on the Navier-Lamé equations. Its principles and numerical implementation are introduced in Section 2. The second section finishes with the description of three growth assessment methods which combine the result of the registration with the result of a segmentation method. Section 3 describes the settings of the experiments. A method for generating artificial nodules as test cases is introduced here. The results are presented in Section 4 and concluded in Section 5.

2 Methods

Basic ideas. Given a reference and a template image, image registration tries to find a *displacement field* $u : \mathbb{R}^3 \to \mathbb{R}^3$ such, that the displaced template image minimizes both a certain similarity measure $\mathcal{D}$ and a regularizing term $\mathcal{S}$. Within this work, the sum of squared differences is used as similarity measure. By adding a regularizing term, the registration problem is a well-posed problem [3]. In general, a regularizer introduces assumptions about the likelihood of certain transformations. Within this context, a perfect regularizer would privilege transformations, which are likely to be the result of growth or shrinkage of a nodule. However, no general growth model for pulmonary nodules is known until today so that a more general regularizer has to be chosen. *Elastic registration* assumes that the images can be characterized as an elastic and compressible material and leads to the elastic regularizer [7]. The elastic properties are modelled by the so-called Lamé constants λ, μ. By using calculus of variations a solution for the regularized registration problem is calculated by computing a stationary point u of its Gâteaux derivative. This leads to a system of non-linear partial differential equations, known as Navier-Lamé equations,

$$\mathcal{A}[u] := \mu \Delta u + (\mu + \lambda) \nabla \cdot \nabla u = f \,, \tag{1}$$

where f is the Gâteaux derivative of $\mathcal{D}$.

Discretization & numerical solution. For the discretization of (1) finite differences in conjunction with Neumann boundary conditions have been chosen. A *time-marching iteration scheme* is employed resulting in the iterative solution of

$$(k\mathtt{A} + \mathtt{Id})\, vec(u^{(l+1)}) = k \cdot vec(f(u^{(l)})) + vec(u^{(l)}) =: \hat{f}(u^{(l)}),\ l \in \mathbb{N}_0 \,, \tag{2}$$

with k as temporal step size, $vec(\cdot)$ as lexicographically ordering operator and $\mathtt{A}\, vec(u)$ as discretization of $\mathcal{A}[u]$. For details, we refer to [3, 8]. This leads to a linear system of equations (LSE) with $3N$ unknowns, where N is the number of voxels of a single image. Since $(k\mathtt{A} + \mathtt{Id})$ does not depend on $u^{(l)}$, it has to be computed only once. Opposed to this, the term $\hat{f}$ depends on $u^{(l)}$ and needs

to be recomputed in every iteration step. For the calculation of the displaced template image a linear interpolation scheme turned out to be sufficient. The matrix $(k\mathtt{A} + \mathtt{Id})$ is sparse, symmetric [8] and highly structured with a typical size of $3 \cdot 40^3$ entries. For the solution of the LSE a conjugate gradient method has been used.

Starting with $u^{(0)} \equiv 0$ the step size k is initialized such, that the maximum displacement in the first iteration step is approximately the voxel grid width and set to be $k_{\max}$. Then, for each iteration step, the step size k is *adapted* under certain conditions. If $\mathcal{D}^{(l+1)} \geq \mathcal{D}^{(l)}$ then the step size is halved and the current iteration step repeated. If $\mathcal{D}^{(l+1)} < \mathcal{D}^{(l)}$ and $k + \Delta_k < k_{\max}$ then the step size is increased by Δ_k. This procedure works well within all our experiments. The iteration loop is stopped when the relative update of $\mathcal{D}$ is below a threshold, i.e., $(\mathcal{D}^{(l)} - \mathcal{D}^{(l+1)})/\mathcal{D}^{(l)} < 10^{-3}$. Throughout this work, the regularizing parameter λ has been set to zero and μ has been chosen as 0.01 being large enough to avoid any folding within the displacement u.

Estimating the growth rate. The displacement field u as result of the described registration scheme together with a segmentation image allows the growth assessment of a nodule. Here, a segmentation image is defined as a binary mask indicating the affiliation of an image element to the object of interest. The set of its image elements with a value of 1 is called foreground in the following. Assuming a segmentation of the reference image, three different criteria are employed to estimate the growth rate.

The first criterion is based on a *transformation of the segmentation image* according to the displacement field u. With a given sampling rate, the growth rate is assessed by comparing the volume of the original segmentation image with the volume of the deformed segmentation image.

The second criterion calculates the determinant of the *Jacobian* of u. It represents the local volume difference between reference and template image [9] and is given by $\det(Id + \nabla u)$. Taking the mean value over the foreground of the underlying segmentation image results in the estimated growth rate.

The third criterion is inspired by vector calculus and utilizes the *divergence* of the displacement field. Again, the growth rate is estimated by taking the mean value with respect to the segmentation.

3 Experiments

A general problem for the evaluation part of this work is the lack of ground truth data. Though, common public databases as the lung image database provided by the LIDC [5] contain delineations by various radiologists, they do not contain delineations for follow-up cases. Single-reader delineations, on the other hand, do not allow for assessing the inter-observer variability. Since experiments on the LIDC data revealed that this variability is not negligible (comparing the volume of each single radiologist's delineation to the mean radiologists' delineation, a median deviation of 14% could be observed [6]), a quantitative comparison is

difficult on the basis of a single delineation. Finally, phantoms currently do not reflect any vascularization around a nodule and therefore do not provide suitable ground truth for our purpose.

Consequently, an algorithm for *generating artificial nodules* and *simulating growth or shrinkage* has been developed. The advantage of this procedure is the availability of ground truth segmentations and the possibility to influence certain characteristics of the generated nodules, e.g., their size or surrounding. Each test case consists of two images, and each image contains one artificially generated nodule. These images may be seen as follow-up images of a CT scan. Two kinds of test cases have been created. While the A-cases are embedded into noisy background, the B-cases are embedded into real-world image data of lung tissue. Category A contains 246 test cases and category B contains 68 test cases. The volumes of the artificial nodules, which are used for the evaluation of the described growth assessment methods, range from 30mm^3 to 1529mm^3 according to a mean diameter ranging from about 4 to 14mm. The simulated growth factors range from 0.7 to 1.4. The voxel resolution was set to $0.6 \times 0.6 \times 1\text{mm}^3$.

Generating test cases. The generation of artificial nodules is inspired by [10] and utilizes the parametric representation as described in [6]. Briefly, the surface of a nodule is spanned by sample vectors which depend on a spherical basis model. With their origin on the sphere the sample vectors point into the direction of the surface normals. Radial basis function interpolation is used to describe the shape profile. For further details we refer to [6].

The creation of the test cases is explained next. Starting with an artificial nodule described by a set of parameters, a second nodule with the same center is generated by changing these parameters such that growth or shrinkage is simulated. This is done by randomly choosing a sample vector as deformation center on the surface of the nodule model followed by an in- or decrease of its length randomly. Proportional to the distance to the deformation center the according neighboring vectors are modified. The two models will be denoted as P_1 and P_2 in the following and contain the parameters needed for an explicit surface description. Once the models are calculated, corresponding binary masks are created and smoothed with a Gaussian kernel with variance σ which results in image masks $Q_1^\sigma(P_1), Q_2^\sigma(P_2)$ with range $[0, 1]$. Within our experiments, σ was set to 1. Given now a foreground image F with normally distributed values ($\mu_F = 0$HU, $\sigma_F = 60$HU) and a background image B, the final test case images $R_{1,2}$ are calculated by $Q_i^\sigma(P_i) \cdot F + \overline{Q_i^\sigma(P_i)} \cdot B$, $i = 1, 2$. Note that in case of category A, the background image B consists of normally distributed intensity values ($\mu_B = -860$HU, $\sigma_B = 60$HU), whereas cases of category B each use a subvolume of real CT data. Thus, category A contains only isolated nodules, whereas category B contains mainly nodules located next to vessels and/or the pleura or other pulmonary structures. Since the parametric model is given for each of the test cases, an according ground truth segmentation can be calculated for each of these. Examples of test cases of category B created with the described method are shown in Figure 2.

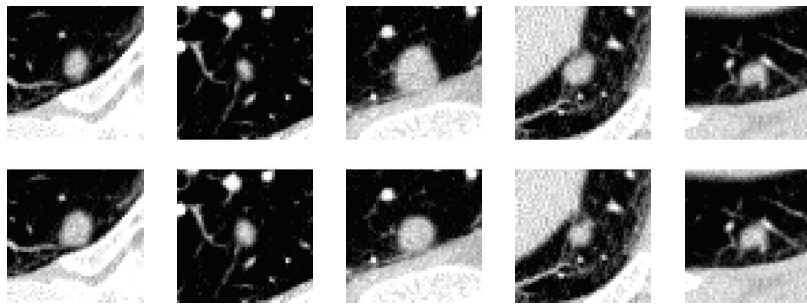

Fig. 2. Exemplary axial views (top row: baseline, bottom row: follow-up) of test cases of category B, i.e., artificial nodules embedded into real world image data.

Evaluating the methods. For the evaluation of the presented growth assessment methods the test cases of each category have been subdivided into three sets. The mean volume of both nodules of each test case has been used as leading quantity for this.

As described in Section 2 the proposed growth assessment requires a segmentation image. To distinguish the accuracy of the registration step from that of the segmentation step, we assume a correct segmentation of the baseline scan and, therefore, employ the ground-truth segmentation image for this specific scan.

To compare the three *combining* assessment methods, i.e., the methods which utilize one segmentation image and a displacement as result of an elastic registration, with a *non-combining* growth assessment scheme, a segmentation of both the baseline and the follow-up scan is computed separately with a state-of-the-art segmentation method [11]. This segmentation method has been evaluated on 33 nodules in a study at the Hematology-Oncology Clinic of Little Rock, Arkansas. Comparisons were made of manual segmentations between two radiologists, and manual vs. automatic volume segmentation. The agreement between manual and automatic volumetry proved to be equally good as the agreement between the two human readers [11].

4 Results

For the quantification of the performances of the growth assessment methods the *magnitude of the difference between the real and the assessed growth rate* (MRA) has been chosen. The results for cases from category A and B are shown in Table 1. Arranged according to the nodule volume, mean and standard deviation of MRA are given for the standard segmentation method (SEGM) as well as for the registration method evaluated by deforming the segmentation image (DEF), calculating the mean of the Jacobian map (JAC), and determining the mean of the divergence (DIV).

Table 1. Mean (standard deviation) of the MRA in terms of the percentage for categories A and B.

		SEGM	DEF	JAC	DIV
A	$30 - 70\,\mathrm{mm}^3$	4.9 (3.8)	7.1 (4.4)	7.0 (4.6)	7.0 (5.1)
	$71 - 172\,\mathrm{mm}^3$	2.3 (2.2)	4.4 (3.5)	4.5 (3.6)	5.1 (3.9)
	$173 - 1529\,\mathrm{mm}^3$	1.6 (1.3)	3.3 (2.7)	3.1 (2.6)	3.9 (3.1)
	all	2.9 (2.7)	4.6 (3.7)	4.6 (3.8)	5.1 (4.1)
B	$30 - 70\,\mathrm{mm}^3$	12.3 (12.8)	5.3 (3.6)	5.8 (3.8)	5.8 (4.3)
	$71 - 172\,\mathrm{mm}^3$	17.0 (27.7)	4.5 (3.6)	4.7 (3.8)	4.8 (4.1)
	$173 - 1529\,\mathrm{mm}^3$	8.4 (13.4)	4.1 (3.3)	4.2 (3.1)	4.1 (3.3)
	all	13.1 (18.9)	4.6 (3.5)	4.8 (3.6)	4.8 (3.9)

In category A it can be observed that all methods perform better with an increasing mean volume of the nodules. Overall, SEGM leads to a MRA of 2.9%. In comparison to this, the MRA of the combining methods is about two percentage points larger. This can be explained by the fact that segmentation of an isolated nodule is a straightforward task. Consequently, its accuracy is expected to be higher than for the combining methods which involve two algorithmic steps and thus two sources of potential errors. In contrast, growth estimation for embedded nodules as in category B is a more complicated task from a segmentation point of view since leakage into adjacent vessels may occur (cf. Figures 1,2).

For category B, SEGM results in a MRA of 13.1% which is about four times larger than the MRA for category A cases but still in the range of inter-observer variability [6]. In contrast, the MRA of the combining methods stays below 5% and is nearly the same as for the category A cases.

Compared to each other, the three combining methods lead in both categories to similar results. Since the numerical calculation of DEF does not include the approximation of derivatives, this method may be superior to JAC and DIV from a numerical point of view. However, our results show only minor better results for DEF.

Even though the category A cases are supposed to be less challenging, the combining methods lead to slightly worse results for the set of nodules with a volume of 30 to $70\mathrm{mm}^3$ than the according category B cases. Further investigation will clarify if this observation is significant or if it is caused by the low number of test cases.

5 Conclusion

We presented a growth assessment method for lung nodules which combines registration and segmentation. Since small variations in the examination data may lead to large variations in the segmented volumes, segmentation of follow-up

scans can be seen as an ill-posed problem. In contrast, a regularized registration scheme is a well-posed problem. It is based on an elastic model that characterizes the image as an elastic and compressible material. In combination with a segmentation image of the baseline scan, the displacement field as result of the registration is evaluated and a growth factor estimated. Three different evaluation methods have been analyzed. Motivated by the lack of ground truth, artificially generated test cases have been used for validation – either as isolated nodules or embedded into real-world data. While for isolated nodules the proposed combined method performs slightly worse than a state-of-the-art segmentation method, the situation changes for the more challenging embedded nodules: Here, the proposed combined method outperforms the non-combining method by a factor of four. Moreover, this result is independent of the choice of the evaluation method.

References

1. Reeves, A.P., Chan, A., Yankelevitz, D.F., Henschke, C.I., Kressler, B., Kostis, W.J.: On measuring the change in size of pulmonary nodules. IEEE Transactions on Medical Imaging **25**(4) (2006) 435–450
2. Ko, J., Betke, M.: Chest CT: Automated nodule detection and assessment of change over time – preliminary experience. Radiology **218**(1) (2001) 267–273
3. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press (2004)
4. Kawata, Y., Niki, N., Ohmatsu, H., Kusumoto, M., Kakinuma, R., Mori, K., Nishiyama, H., Eguchi, K., Kenko, M., Moriyama, N.: Analysis of evolving processes in pulmonary nodules using a sequence of three-dimensional thoracic images. Proceedings of SPIE: Image Processing **4322** (Apr 2003) 1890–1901
5. Armato III, S.G., McLennan, G., McNitt-Gray, M.F., Meyer, C.R., Yankelevitz, D., Aberle, D.R., Henschke, C.I., Hoffman, E.A., Kazerooni, E.A., MacMahon, H., Reeves, A.P., Croft, B.Y., Clarke, L.P.: Lung Image Database Consortium: Developing a Resource for the Medical Imaging Research Community. Radiology **232**(3) (2004) 739–748
6. Opfer, R., Wiemker, R.: A new general tumor segmentation framework based on radial basis function energy minimization with a validation study on LIDC lung nodules. Volume 6512., San Diego, CA, USA, SPIE (2007) 651217–10
7. Broit, C.: Optimal Registration of Deformed Images. PhD thesis, University of Pennsylvania (1981)
8. Kabus, S.: Multiple-Material Variational Image Registration. PhD thesis, University of Lübeck (2006)
9. Leow, A., Yanovsky, I., Chiang, M.C., Lee, A., Klunder, A., Lu, A., Becker, J., Davis, S., Toga, A., Thompson, P.: Statistical Properties of Jacobian Maps and the Realization of Unbiased Large-Deformation Nonlinear Image Registration. IEEE Transactions on Medical Imaging **26**(6) (2007) 822–832
10. Zhang, X., Chui, H., Olcott, E., Raffy, P., Yu, N., Giger, M.L., Karssemeijer, N.: Evaluation of lung nodule growth measurement for MDCT exams with different dosages using synthetic nodules. Medical Imaging 2007: Computer-Aided Diagnosis **6514**(1) (Mar 2007) 651438–8

11. Wiemker, R., Rogalla, P., Blaffert, T., Sifri, D., Hay, O., Shah, E., Truyen, R., Fleiter, T.: Aspects of computer-aided detection (CAD) and volumetry of pulmonary nodules using multislice CT. The British Journal of Radiology **78** (2005) 46–56