How Do Registration Parameters Affect Quantitation of Lung Kinematics?

Tessa Sundaram Cook¹, Nicholas Tustison¹, Jürgen Biederer^{2,3}, Ralf Tetzlaff³,
and James Gee¹

¹ Dept. of Radiology, University of Pennsylvania, Philadelphia, PA, USA

² Dept. of Diag. Radiology, Univ. Hosp. Schleswig-Holstein, Campus Kiel, Germany

³ Dept. of Radiology, German Cancer Research Center, Heidelberg, Germany

Abstract. Assessing the quality of motion estimation in the lung remains challenging. We approach the problem by imaging isolated porcine lungs within an artificial thorax with four-dimensional computed tomography (4DCT). Respiratory kinematics are estimated via pairwise non-rigid registration using different metrics and image resolutions. Landmarks are manually identified on the images and used to assess accuracy by comparing known displacements to the registration-derived displacements. We find that motion quantitation becomes less precise as the inflation interval between images increases. In addition, its sensitivity to image resolution varies anatomically. Mutual information and cross-correlation perform similarly, while mean squares is significantly poorer. However, none of the metrics compensate for the difficulty of registering over a large inflation interval. We intend to use the results of these experiments to more effectively and efficiently quantify pulmonary kinematics in future, and to explore additional parameter combinations.

1 Introduction

The ability to quantify pulmonary deformation is useful in early detection of disease, evaluation of treatment efficacy and improved assessment of disease staging and prognosis. Structural imaging modalities can be used to capture *in vivo* deformation of the lung between sequential images by harnessing the power of non-rigid registration algorithms. Dougherty and Li have applied optical flow to computed tomography (CT) images of the lung to construct an atlas of normal pulmonary anatomy [1,2], while Hatabu et al. have developed magnetic resonance protocols for imaging the lung, [3]. In addition, Sarrut et al. have proposed an interpolation scheme for upsampling 4DCT image sequences, [4].

We have shown that non-rigid registration of serial MR images can be used to quantify lung kinematics, [5]. But validating the accuracy of these motion estimates is difficult; ground-truth motion cannot be determined without disruptive methods (such as implantation of tissue fiducials). Less invasive validation approaches for non-rigid registration in general include tracking landmarks on images, [6], comparison with known synthetic deformations, [7], tracking known

displacements of a phantom, [8], assessment of anatomic overlap, [9], and comparison to other modalities, [10]. However, these methods are often sub-optimal when validating motion quantitation of *in vivo* dynamic organs.

We have previously used landmark-based validation: identifying and tracking structures which are visible within the lungs, such as airway and vessel bifurcations, [11]. The challenges we encountered include the inability to consistently identify points in successive images as well as a limit on the total number of landmarks that can be identified. Here, we image porcine lung explants ventilated using an artificial thorax, [12]. Nearly 350 landmarks are identified within the lungs, along the artificial diaphragm, and (as controls) on the hull of the artificial thorax. The “ground-truth” motion of these points is compared with the registration-derived motion estimate to assess the accuracy of the parenchymal dynamics computation with respect to image resolution and similarity metric. We also seek to quantify the optimal inflation interval for motion analysis.

Table 1. Summary of 36 experiments performed. “Iterations” indicates the # of iterations at each of the four levels (from $\frac{1}{8}$ of full resolution to full resolution).

<u>Iterations</u>		<u>Metric</u>		<u>Intervals</u>
100/1/1/1		MI		sequential (0%-25%, 25%-50%, etc.)
100/100/1/1	×	MSQ	×	alternating (0%-50%, 50%-100%)
100/100/100/1		CC		phase (0%-100%)
100/100/100/100				

2 Materials and Methods

2.1 Data Acquisition

Porcine lung explants from a healthy animal are placed inside a dedicated chest phantom designed by Biederer and colleagues, [12]. The phantom is sealed and evacuated to make the lung expand passively. A water-filled elastic diaphragm is used to simulate tidal respiration by cyclic variation of the filling volume at a frequency of 8 cycles/min. 4DCT images are acquired in dynamic mode with a pitch of 0.1 (Siemens Somatom, slice collimation 24×1.2 mm, rotation time 1 sec, slice thickness 1.5 mm, increment 0.8 mm, 120 kV, 400 mAs, Kernel B50s). Images are reconstructed retrospectively at 0, 25, 50, 75 and 100% inspiration. The data are acquired with a matrix of 512×512 . To accommodate technical limitations, image volumes are resampled to dimensions of $256 \times 256 \times 243$ with 1.17-mm isotropic voxels.

2.2 Symmetric, Diffeomorphic Non-rigid Registration

We use an intrinsically symmetric image registration for the experiments in this paper, [13]. The method is able to capture large, diffeomorphic transformations between two images I and J , and may be guided by a variety of similarity metrics to give a dense space-time mapping. Normally, the metric derivative is

taken with respect to either I or J (whichever is the transforming image), creating a dependence on the gradients of that image. This symmetric approach resolves this bias in a fundamental way, using the transformation model, and optimizes the registration with respect to both ϕ_1 and ϕ_2 (the forward and inverse transformations). The forward and backward registrations both contribute via $I(\phi_1(\mathbf{x}, t))$ and $J(\phi_2(\mathbf{z}, 1 - t))$, respectively, where t is the path between I and J , $\mathbf{x}$ is the coordinate system of I , $\mathbf{z}$ is the coordinate system of J , ϕ_1 is the forward mapping that warps J into the space of I , and ϕ_2 is the inverse mapping that warps I into the space of J .

Performance of the algorithm has previously been evaluated for brain registration, [13]. Here, we test its effectiveness specifically for lung motion quantitation. We compare the effect of three different similarity metrics on our motion estimates: mean squares (MSQ), a $5 \times 5 \times 5$ -neighborhood cross-correlation (CC) based on the implementation in [14], and mutual information (MI). All three are adapted to the symmetric formulation of the registration. We also explore the dependence of the motion quantitation process on the resolution of the image.

Fig. 1. Sample mid-coronal vector fields (extracted from the 3-D displacements) representing expansion from (left) 0% to 25%, (middle) 0% to 50% and (right) 0% to 100%. Vector magnitudes increase nonlinearly as the degree of expansion increases; there is greater motion at the lung bases than the apices.

2.3 Estimation of Pulmonary Motion

Our image sequence is processed multiple times to study the effect of image resolution and image similarity on the accuracy of motion quantitation over progressively larger inflation intervals. First, we study the sensitivity of our motion estimates to the number of resolution levels processed. We run registrations at four increasing levels of resolution (from $\frac{1}{8}$ to 1) using MI as the similarity in all cases. Next, we examine the effect of metric choice on our motion estimates by running registrations at all four resolution levels with each of the three metrics under investigation. In both cases, three types of inflation interval are used: sequential, alternating, and phase (see table 1); the latter covers the entire inspiratory phase of respiration. Each level of the multi-resolution pyramid is used for no more than 100 iterations.

We anticipate that registration accuracy will decrease as the inflation interval increases, and that mutual information and cross-correlation will perform better

than mean squares because of their regional coverage and statistical sampling, respectively. Furthermore, we expect that most of the global motion of the lung will be captured by registration at the lowest resolution level, and that endpoint errors will generally be higher at the diaphragm than within the lung.

2.4 Assessment of Motion Quantitation Accuracy

A specially trained physician manually selects point landmarks on the 0% inspiration image and then identifies them on subsequent images (figure 2). Branching structures (both airway and vascular) are chosen on each image because of their reproducibility and distribution throughout the lungs. In addition, landmarks are placed on the diaphragm—site of greatest pulmonary deformation—and on the phantom itself. 125 landmarks identify airway and vessel bifurcations, while 142 points are selected on the diaphragm and 76 points on the hull of the phantom. This distribution of the landmarks throughout the lungs is useful because it allows us to quantify the uncertainty of our registration measurements in different anatomic regions.

Fig. 2. 3-D renderings of the landmark displacements between (left) 0% and 25% inflation and (right) 0% and 100% inflation. Vectors are rendered in 3-D space, while orthogonal cross-sections of the image volume are provided for reference. Note that the points on the phantom’s hull serve as controls and therefore lack displacement vectors.

The error between the known landmark displacement $\mathbf{dx}_l$ and the registration-derived displacement $\mathbf{dx}_r$ at each landmark location is computed as the distance between the endpoints of the two displacement vectors ($|\mathbf{dx}_l - \mathbf{dx}_r|$). In the discussion that follows, we report the mean of this endpoint error over the landmarks within the indicated anatomic region (diaphragmatic or parenchymal).

3 Results

The runtime of the 36 experiments described above depends on the number of resolutions at which an upper limit of 100 iterations is prescribed. Registration at the lower resolutions is achieved in 10-20 minutes, while registration at all four levels requires 2-3 hours for MI and 6-10 hours for CC. All registrations utilize up to 2.5 gigabytes of RAM. Figure 1 provides examples of the displacement fields generated via registration of the 0% inflation image to subsequent images in the sequence. Figure 3 summarizes the errors observed in our experiments.

3.1 Effect of Image Resolution

When estimating lung motion sequentially, the final result is almost completely achieved at the lowest image resolution, while the error within the lung drops by 0.5-2 mm with processing at higher levels. Errors tend to be 1-2 mm higher at the diaphragm than within the lung, likely because the diaphragm and lung bases move more than the apical parenchyma. Upon visualization of the registration-derived displacements at the landmarks, it appears that the error at the diaphragm is angular (vectors are of appropriate magnitude but point in a different direction), while within the parenchyma it is scalar (the vectors point in an appropriate direction but are not of the correct magnitude).

Using the alternating inflation intervals, we again observe a 1 mm gain in accuracy within the lung by registering at higher levels; errors at the diaphragm remain almost constant. This is likely because the soft tissue-air interface at the diaphragm is not compromised by downsampling and remains a strong contributor to the gradient of the similarity at low image resolutions.

When registering the entire inspiratory phase, we again observe an increase in accuracy intraparenchymally with higher levels of processing. However, it is important to note that the mean endpoint error throughout the lung and diaphragm is 11 mm as a result of the large degree of inflation. By acquiring one additional image (and using alternating intervals), we are able to cut our error in half. By further doubling our sampling frequency, we gain another 25-50%.

Overall, errors are at least 1 mm higher along the diaphragm than within the lung parenchyma. In most cases, the magnitude of the resulting endpoint errors is less than 40% of the actual motion in the region. There does not appear to be much improvement between the 128^3 level and the 256^3 level. This suggests that processing at the highest data resolution does not significantly alter the final result, and is encouraging because this level is the most time-consuming.

3.2 Effect of Image Similarity Metrics

Regardless of inflation interval, mean squares is consistently the least accurate of the three metrics studied. It results in a diaphragmatic error of 0.5-4.5 mm for sequential registrations, 1.9-4 mm for alternating registrations and 1.8-6.3 mm for the inspiratory phase registrations. Within the lung however, the MSQ

Fig. 3. Mean±SD endpoint errors for the resolution (left—top legend) and similarity (right—bottom legend) experiments at each of the three intervals (rows). The middle and last bars in each group indicate (for reference) the average motion at each landmark along the diaphragm and within the lung, respectively. #Lev = # of resolution levels used, Dia = diaphragm.

error is within 0.5-1 mm of the MI and CC errors, and occasionally less than the former.

Comparison of the performance of MI and CC yields some interesting observations. For the sequential registrations, the accuracy of the two metrics is within 0.5 mm of one another. As the inflation interval increases, MI outperforms CC (by an average of 1.8 mm in the phase registrations). However, within the lung CC is 1.4-1.9 mm better in the alternating registrations (the two metrics are within 0.4 mm of one another in the other experiments). This motivates the anatomic customization of similarity computations. Matching at the diaphragm may be easier because of the clear tissue interface between air and muscle, while matching in the parenchyma may require additional regional information (such as that provided the neighborhood integration of CC).

Again, it is important to note that the phase registrations with MI and CC are only accurate to within 11-13 mm at the diaphragm and 10-11 mm within the lung. Adding just one additional image reduces the mean error to 6-7 mm at the diaphragm and only 3-6 mm within the lung.

4 Discussion

We present a quantitative evaluation of the sensitivity of pulmonary motion estimation to image resolution, similarity metric and inflation interval. As expected, accuracy drops as the inflation interval increases. It is conceivable that registering for additional iterations may reduce this error; however, the sequential and alternating registrations rarely reached the limit of 100 iterations per level.

With only one set of landmarks, it is difficult to quantify the accuracy of the points themselves. It is realistic to assume that there is an inherent error associated with identifying the same anatomic point in different images. However, this can only be quantified by performing a multi-observer experiment in which two or more physicians independently identify the same landmarks, enabling the computation of inter-observer vs. registration-observer errors.

Ideally, we would be able to analyze data at the acquired resolution, instead of halving the resolution due to memory limitations. However, it is important to note that the results at the top two resolution levels are very similar. This suggests that slightly augmenting system memory may allow us to register data at the original resolution ($0.5 \times 0.5 \times 0.8$ mm in this case), and result in even better accuracy. Furthermore, we can reduce the number of iterations at the highest level and drastically reduce the runtime required to process an image pair without sacrificing the quality of our motion estimates.

Though we process only one image sequence in this experiment, the importance of small inflation intervals is clear. Attempting to quantify lung motion over an entire phase (i.e., end-inspiration to end-expiration), whether we use tidal limits or respiratory extremes, is inaccurate. If a 1-cm error is acceptable, then it is appropriate to acquire only two images, reducing scan time and radiation exposure. However, most clinical applications of this work would not tolerate such a large inaccuracy. Hence, this work motivates the need for finer sampling of the respiratory cycle.

In the future, we hope to further explore the issues in this paper with additional image sequences and multi-observer validation. Evaluation of different registration algorithms would also be interesting to determine if our current motion quantitation approach can be improved. Another avenue of exploration is the sensitivity of our methods to different degrees of image noise. It would also be interesting to investigate the sampling frequency problem in greater detail, as well as to compare motion quantitation in CT and MR. The ultimate goal of our work is to develop robust methods for motion quantitation that allow the development of a thorough spatio-temporal analysis of pulmonary motion and the subsequent construction of dynamic atlases of human lung motion.

References

1. Dougherty, L., Asmuth, J.C., Gefter, W.B.: Alignment of CT lung volumes with an optical flow method. *Acad. Rad.* 10(3), 249–254 (2003)
2. Li, B., Christensen, G.E., Hoffman, E.A., McLennan, G., Reinhardt, J.M.: Establishing a normative atlas of the human lung: intersubject warping and registration of volumetric CT images. *Acad. Rad.* 10, 255–265 (2003)
3. Hatabu, H., Ohno, Y., Uematsu, H., Oshio, K., Gefter, W.B., Gee, J.C.: Lung biomechanics via non-rigid reg. of serial MR images. *Rad.* 221P, 630 (2001)
4. Sarrut, D., Boldea, V., Miguët, S., Ginestet, C.: Simulation of four-dimensional CT images from deformable registration between inhale and exhale breath-hold CT scans. *Med. Phys.* 33(3), 605–617 (2006)
5. Sundaram, T., Gee, J.: Towards a model of lung biomechanics: pulmonary kinematics via registration of serial lung images. *Med. Image Anal.* 9, 524–537 (2005)
6. Betke, M., Hong, H., Thoams, D., Prince, C., Ko, J.: Landmark detection in the chest and registration of lung surfaces with an application to nodule registration. *Med. Image Anal.* 7, 265–281 (2003)
7. Schnabel, J.A., Tanner, C., Castellano-Smith, A.D., Degenhard, A., Leach, M.O., Hose, D.R., Hill, D.L.G., Hawkes, D.J.: Validation of nonrigid image registration using finite-element methods: application to breast MR images. *IEEE TMI* 22(2), 238–247 (2003)
8. Dougherty, L., Asmuth, J., Blom, A., Axel, L., Kumar, R.: Validation of an optical flow method for tag displacement estimation. *IEEE TMI* 18(4), 359–363 (1999)
9. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., Mazziotta, J.: Automated Image Registration: I General methods and intra-subject intra-modality validation. *J. Comp. Asst. Tomo.* 22, 141–154 (1998)
10. Ledesma-Carbayo, M.J., Mahia-Casado, P., Santos, A., Perez-David, E., Garcia-Fernandez, M.A., Desco, M.: Cardiac motion analysis from ultrasound sequences using nonrigid registration: validation against doppler tissue velocity. *Ultrasound in Med. and Biol.* 32(4), 483–490 (2006)
11. Sundaram, T.A., Gee, J.C.: Quantitative comparison of registration-based lung motion estimates from whole-lung MR images and corresponding two-dimensional slices. In: *Proc. ISMRM 15th Mtg*, p. 3039 (2007)
12. Biederer, J., Plathow, C., Schoebinger, M., Tetzlaff, R., Puderbach, M., Bolte, H., Zaporozhan, J., Meinzer, H.P., Heller, M., Kauczor, H.U.: Reproducible simulation of respiratory motion in porcine lung explants. *Röntgenstr* 178(11), 1067–1072 (2006)
13. Avants, B.B., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration: Evaluating automated labeling of elderly and neurodegenerative cortex and frontal lobe. In: *Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS*, vol. 4057, pp. 50–57. Springer, Heidelberg (2006)
14. Hermosillo, G., Chéfd'Hotel, C., Faugeras, O.: A variational approach to multi-modal image matching. *Intl. J. Comp. Vis.* 50(3), 329–343 (2002)