Quantitative Assessment of Registration in Thoracic CT

K. Murphy, B. van Ginneken, J.P.W. Pluim, S. Klein, and M. Staring

University Medical Center, Utrecht, The Netherlands.

Abstract. A novel method for the quantitative evaluation of registration systems in thoracic CT is utilised to examine the effects of varying system parameters on registration error. Regional analysis is implemented to determine whether registration error is more prevalent in particular areas of the lungs. Experiments on twenty-four CT scan-pairs prove that in many cases significant reductions in processing time can be achieved without much loss of registration accuracy. More difficult cases require additional steps in order to achieve maximum precision. Larger errors appear more frequently in the lower regions of the lungs close to the diaphragm.

1 Introduction

The accurate registration of intra-patient thoracic CT scans has a variety of motivating clinical applications including improved ease of visual comparison, quantitative or automatic analysis of pathology progression, and in the case of inspiration/expiration pairs, analysis of lung function. In radiotherapy planning, registration information can be used to construct pulmonary motion models in order to propagate the location of the target region [4].

Although many promising registration algorithms exist, the quantitative evaluation of these techniques poses a further challenge due to the lack of an established reference standard. Without a means for quantitative assessment the improvement and optimisation of a registration algorithm is extremely difficult. Visual analysis of registered images is a time-consuming and subjective process and particularly in 3D images it is impossible to visually quantify subtle differences between results from various systems.

In this work a registration reference standard for thoracic CT pairs is formulated in an efficient semi-automatic manner, resulting in a well-distributed mesh of corresponding landmarks throughout the lung volumes to be registered. This reference standard is used to evaluate a parametric intensity-based registration algorithm under varying conditions. Regional error analysis is implemented to determine whether registration error is more prevalent in specific areas of the lungs.

2 Materials

All scans used in this work form part of an experimental lung cancer screening programme. Twenty-four patients (22 male, 2 female, ages 54-79yrs), each with a baseline and a follow-up scan (3-9 months apart) were chosen randomly from the database. All scans were obtained at full inspiration and without contrast injection on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems). They have a per-slice resolution of 512×512, with the number of slices per scan varying from 383 to 551. Slice thickness is 1mm with slice-spacing of 0.7mm. Pixel spacing in the X and Y directions varies from 0.55mm to 0.8mm.

All registration experiments were carried out on a standard desktop PC with an Intel Core 2 Duo processor, 2.4GHz.

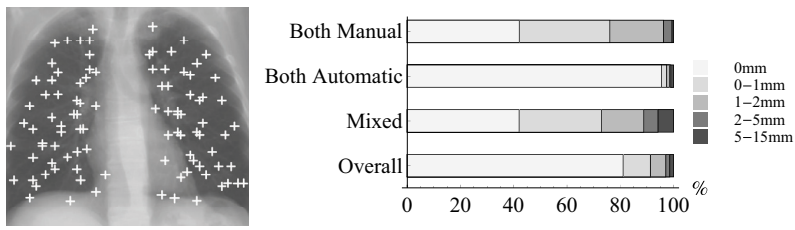
3 Methods

3.1 Reference Standard Construction

In this section an overview of the reference standard construction method will be provided. The technique used is described in detail in [5].

The first step in the construction of the reference standard is the determination of landmark locations in the baseline scan. A fully automatic system has been designed which identifies 100 well-dispersed points throughout the lungs. These points are required to be sufficiently distinctive to enable them to be matched in the corresponding follow-up scan and points on the pleural surface itself are therefore excluded. A projection view of all the landmarks selected for a scan is shown in figure 1(a).

A semi-automatic system was developed to accurately match the voxels identified as landmarks in the baseline scan with voxels at the corresponding anatomic locations in the follow-up scan. Each scan pair was processed twice by independent observers (medical students). The observers were required to match at least 20 of the 100 landmarks manually using a custom-made graphical interface. The ordering of the points presented to the users was designed such that each subsequent point was well-distanced from its predecessors. During this phase the system utilised a thin-plate-spline (TPS) [1] and the thus-far annotated point pairs to model the relationship between the two images. The TPS model was evaluated at each new point by attempting to predict the correct correspondence and comparing this prediction with the subsequent user annotation. When 20 points were manually matched the system handled the remaining points automatically, provided that the TPS model had been validated by successful predictions of the user annotated matches. The annotation procedure took 20-30 minutes per scan-pair and did not require observers with significant experience of pulmonary anatomy.



(a) (b)

Fig. 1. (a)A set of automatically determined landmarks projected in the coronal direction. (b) Inter-observer differences categorised by match-types.

3.2 Registration Methods

Prior to registration the baseline and follow-up scans were down-sampled in order to reduce memory consumption. The down-sampling was by means of block-averaging such that the matrix size of 512×512 in the original images is reduced to 256×256, with the number of slices being reduced accordingly by a factor of 2. The calculated transform from the registration procedure was subsequently applied to the full resolution follow-up scan.

The registration procedure consisted of an initial affine registration step followed by an elastic registration to handle the non-rigid deformations of the lung tissue. Both registration steps involved a multi-resolution strategy using a Gaussian image pyramid. A mutual information cost function [7] was used in both cases along with a stochastic gradient descent optimizer [3]. The elastic registration deformations were modelled by a B-Spline grid [6]. The grid-size varied per resolution-level with the finest grid at the last level having a spacing of 8 voxels in each dimension.

In this work only the anatomy within the lungs is registered and all other structures are masked out. Previous experiments [5] have determined that this gives more accurate registration of the structures within the lungs. The mask used to distinguish the lungs from other anatomy was created by means of an automatic lung segmentation procedure based on the work of Hu et al. [2].

A number of experiments have been carried out in order to test the effects of tuning various parameters in the registration system. In particular the number of resolutions in the multi-resolution scheme and the number of iterations in the stochastic gradient descent optimizer are varied to determine their importance and optimal values. Registration error is analysed for each scan-pair at the various settings and regional error analysis is carried out over the entire dataset

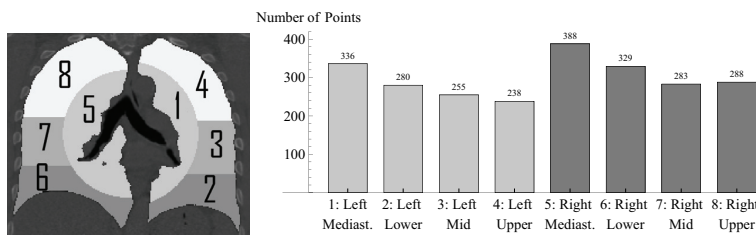
4 Results

4.1 Reference-Standard Construction Results

Inter-observer differences The inter-observer differences were analysed to verify the ability of observers and of the system to find reproducible corresponding anatomic locations for the landmarks. In figure 1(b) the inter-observer differences in mm are illustrated, categorised by match-type. The match-type indicates whether the point was marked manually by one or both observers or whether it was chosen by the system. Regardless of match-type, 97% of all points had an inter-observer difference below 2mm. As is expected, points which were marked automatically by both observers are considerably more likely to have differences of 0mm than those which were marked manually, however manual observations are within 2mm of each other in 96% of cases.

Dispersal of Reference Points In order to verify that the reference points were evenly distributed around the lung volume, and later to analyse registration error in a regional manner, each lung was divided into 4 equally sized volumetric regions as follows: (See figure 2(a)). The centre of mass, c of both lung volumes together was identified. A sphere s, centred on c, was constructed such that 25% of the left lung volume was enclosed by s. This 25% represented the portion of left lung around the mediastinum. The remainder of the left lung was divided into 3 equally sized volumes by cutting at the appropriate slices. The right lung was then divided in an analogous fashion.

The number of points in each region over all images is illustrated in figure 2(b). It is clear that the points are well distributed over all areas. The right lung has slightly more points than the left which is to be expected due to its larger size, and the mediastinal area has a slightly higher concentration of landmarks since it is generally a much more distinctive region than any of the peripheral areas.



(a) (b)

Fig. 2. (a)A slice showing a cross-section of the 3D lung partitions calculated. (b)The dispersal of points among the partitioned regions.

4.2 Registration Results

For each image-pair the computed transform T which maps from locations in the deformed follow-up scan to locations in the original follow-up scan is applied to each of the landmark points l from the baseline scan. It is clear that for an accurate registration we expect $T(l) \approx lm$, where lm is the matching point marked during reference standard formulation.

For all points lm_{obs1} marked (manually or automatically) by observer1 the Euclidean distances $\delta(T(l), lm_{obs1})$ between $T(l)$ and lm_{obs1} were calculated using the appropriate T for the scan-pair. These distances δ were used as a measure of registration accuracy.

Varying number of Iterations Reducing the number of iterations performed in the stochastic gradient descent procedure is one way to considerably improve the speed of the registration system. In order to determine the importance of this parameter on the registration results the 24 scan-pairs were registered first with 512 iterations, then with 256 and finally with 128. All other parameters were kept fixed, with 4 resolution levels during the affine registration and 5 during the elastic. The registration errors $\delta(T(l), lm_{obs1})$ are shown in box-whisker plots for each scan-pair at each setting in figure 3. In most cases the median error increases slightly with fewer iterations although in a few instances, particularly those scans with the lowest error measures the number of iterations has little effect.

The time to register a down-sampled image pair was reduced from approximately 10 minutes with 512 iterations to 5 minutes with 256 iterations or 3 minutes with 128. Consideration must be given to balancing the registration accuracy against the amount of time required to complete a registration since many clinical applications demand results within a specified timeframe.

Figure 4 shows an example of a difficult case (the sixth case from figure 3). Although the images are reasonably well aligned there are clearly some errors in the vessel structure. Subtraction images shown in figures 5(a) (subtraction after registration with 128 iterations) and 5(b) (subtraction after registration with 512 iterations) illustrate the difficulty of visually assessing registration results.

Varying number of Resolutions In this experiment the number of iterations was kept fixed at 512 while the number of resolutions was varied. The registrations were carried out firstly with 4 resolutions in the affine step and 5 in the elastic step and secondly with 3 resolutions in the affine step and 4 in the elastic step. The registration errors $\delta(T(l), lm_{obs1})$ for each scan-pair are shown in figure 6. In most cases the reduction in numbers of resolutions had little effect on the registration error, however in a single case the registration result with fewer resolution levels is so poor that the box showing the interquartile range of errors cannot be seen at the scale shown in figure 6. The scans to be registered in this case were so disparate that they required extra low-resolution steps in order to overcome the large-scale differences early in the procedure. By adding the extra resolution step back into the affine phase only, the median error is reduced from approximately 18mm to just 0.5mm.

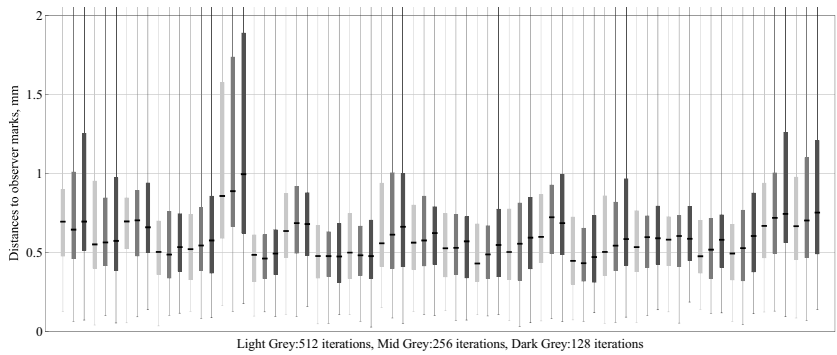


Fig. 3. Registration errors per scan-pair for varying numbers of iterations.

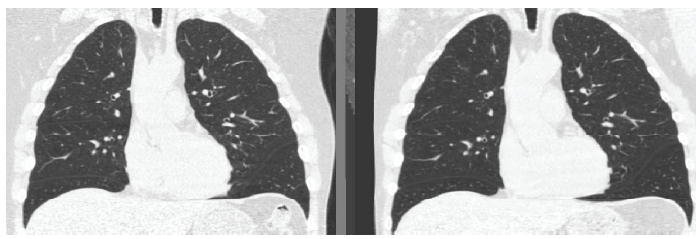
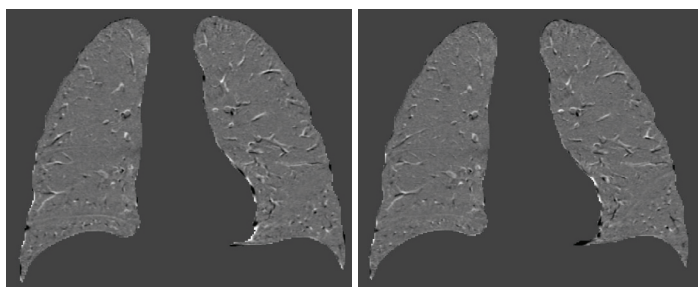


Fig. 4. A difficult case, corresponding slices from the fixed image and the deformed moving image.



(a) (b)

Fig. 5. The same difficult case as shown in figure 4. Subtraction image after registration with (a) 128 iterations and (b) 512 iterations.

The reduction of the number of resolution steps had a minimal effect on the time required to complete a registration, saving only in the order of 30 seconds of the 10 minutes.

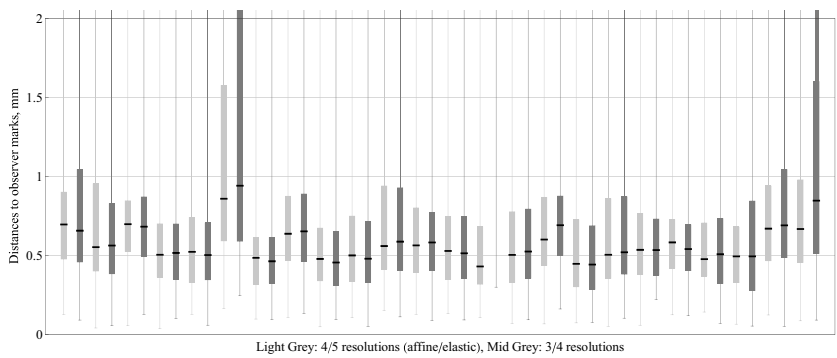


Fig. 6. Registration errors per scan-pair for varying numbers of resolutions

Registration Error per Region As described in section 4.1 and illustrated in figure 2(a) each lung was divided into 4 equal volumes to enable regional analysis. In figure 7 the error per region is shown for the system at 3 different settings. Based on the range of errors above the 0.75 quantile value it is clear that in all cases the largest errors for each lung are seen in the lower sections (region labels 2 and 6). This is to be expected since the motion of breathing affects the lower lungs much more significantly than the upper. Similarly, in all but one case, the upper section of the lung (region labels 4 and 8) has less error than any other section.

In figure 8 a closer view of the median values of the same box-plots is shown. The differences in median error values between regions are of the order of 0.1mm showing that for the majority of points there is little difference between regions. Median values for the regions close to the mediastinum and the diaphragm (region labels 1,2,5,6) tend to be slightly higher than those for the peripheral areas of the middle and upper lung.

5 Conclusion

A semi-automatic system for reference standard formulation has been used to generate a well-distributed mesh of corresponding landmark points in intra-patient thoracic CT scan pairs. The scan pairs have been non-rigidly registered using a parametric intensity based registration algorithm with various parameter

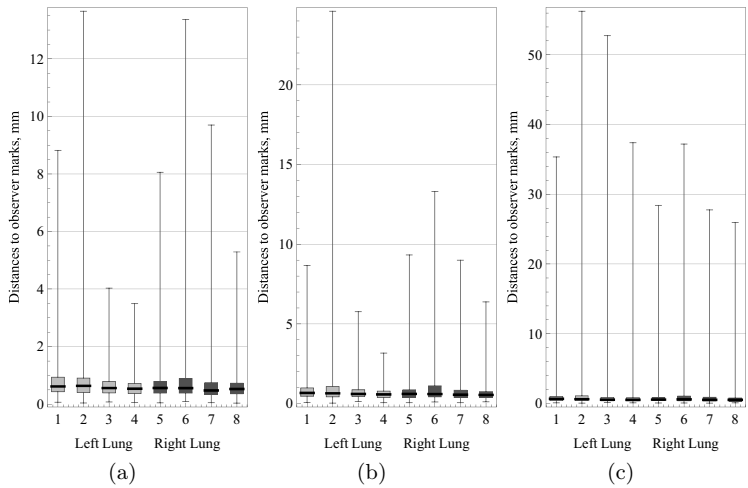


Fig. 7. Plots showing registration error per region for various system settings. The labels on the X-axes refer to the lung regions as shown in figure 2(a). (a) System with 512 iterations and 4/5 (affine/elastic) resolutions (b) System with 128 iterations and 4/5 resolutions (c) System with 512 iterations and 3/4 resolutions

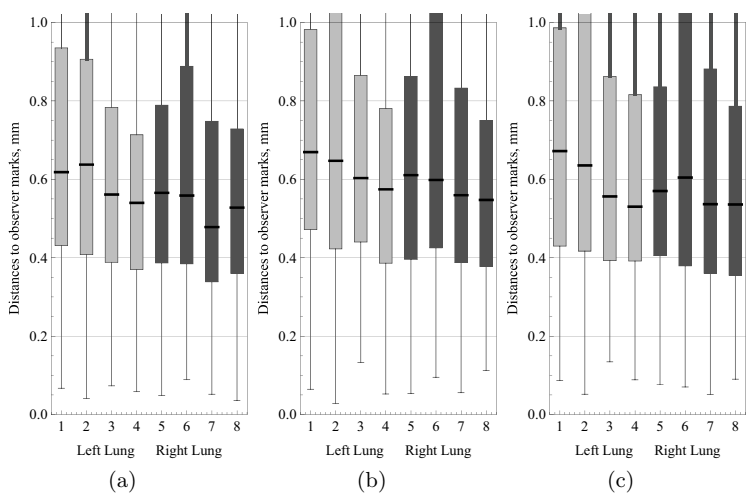


Fig. 8. A closer view of the median regions from the plots of figure 7.

settings. The constructed reference standard enabled the quantitative compari-

son of results from different system settings and the detection of subtle disparities in registration accuracies.

Regional analysis was also possible due to the regular distribution of the landmark points. It appears that larger errors are more likely to occur in the lower sections of the lung close to the diaphragm.

The results of parameter testing confirm that registration accuracy is generally negatively affected by the reduction of the number of iterations in gradient descent optimisation. However the accuracy difference in terms of millimetres is usually small or even negligible for easier cases (those where a good result is already achievable with fewer iterations, probably due to a good initial alignment of the baseline and follow-up images). Depending on the application and required accuracy the reduction in processing time may be more important than a negligibly small gain in accuracy.

The number of resolutions to be used in the multi-resolution scheme was shown to be an important factor in registration accuracy. In cases where the initial difference between scans is large the reduction of the number of resolutions proved to be detrimental to the system accuracy while providing little in terms of processing speed improvement.

These initial results suggest that optimisation of the registration algorithm may best be achieved by means of a feedback strategy whereby easier registration tasks may be completed with a minimal number of iterations while more difficult cases would be identified in the early stages and treated accordingly with additional steps.

References

1. F.L. Bookstein. Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. *IEEE Trans. PAMI*, 11:567–585, 1989.
2. S. Hu, E. A. Hoffman, and J. M. Reinhardt. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans. Med. Imaging*, 20(6):490–498, 2001.
3. S. Klein, M. Staring, and J.P.W. Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. *IEEE Trans. Image Proc.*, 16(12):2879–2890, Dec 2007.
4. J. McClelland, J. Blackall, and S. Tarte. A continuous 4D motion model from multiple respiratory cycles for use in lung radiotherapy. *Medical Physics*, 33, 2006.
5. K. Murphy, B. van Ginneken, J.P.W. Pluim, S. Klein, and M. Staring. Semi-automatic reference standard construction for quantitative evaluation of lung CT registration. In *Medical Image Computing and Computer-Assisted Intervention*, 2008.
6. 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(8):712–721, 1999.
7. P. Thévenaz and M. Unser. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Image Proc.*, 9(12):2083 – 2099, December 2000.