3D motion tracking of pulmonary lesions using CT fluoroscopy images for robotically assisted lung biopsy

Sheng Xu[a], Gabor Fichtinger[a], Russell H. Taylor[a], Kevin Cleary[b]
[a]Engineering Research Center, Johns Hopkins University, Baltimore, MD, USA 21218;
[b]Imaging Science and Information Systems (ISIS) Center, Department of Radiology, Georgetown University Medical Center, Washington DC, USA 20007

ABSTRACT

We are developing a prototype system for robotically assisted lung biopsy. For directing the robot in biopsy needle placement, we propose a non-invasive algorithm to track the 3D position of the target lesion using 2D CT fluoroscopy image sequences. A small region of the CT fluoroscopy image is registered to a corresponding region in a pre-operative CT volume to infer the position of the target lesion with respect to the imaging plane. The registration is implemented in a coarse to fine fashion. The local deformation between the two regions is modeled by an affine transformation. The sum-of-squared-differences (SSD) between the two regions is minimized using the Levenberg-Marquardt method. Multi-resolution and multi-start strategies are used to avoid local minima. As a result, multiple candidate transformations between the two regions are obtained, from which the true transformation is selected by similarity voting. The true transformation of each frame of the CT fluoroscopy image is then incorporated into a Kalman filter to predict the lesion's position for the next frame. Tests were completed to evaluate the performance of the algorithm using a respiratory motion simulator and a swine animal study.

Keywords: motion estimation, real-time tracking, robust tracking, data association, textured regions, lung biopsy, robotics

1. INTRODUCTION

In lung biopsy, the position of the target lesion may vary due to intrinsic causes such as respiratory motion or extrinsic reasons such as interactions between the tissue and a surgical tool. In developing a robotic system to assist in biopsy, we need to provide some way to track the target lesion. CT fluoroscopy combines the advantages of both CT and fluoroscopy, which offers an opportunity to track the motion of the target lesion in real-time during the intervention. Our prototype system is shown in Fig. 1. The real-time CT fluoroscopy image is captured using a frame grabber (Accustream 170, Foresight Imaging, Lowell, Massachusetts, USA). The position of the target lesion is detected from the image, which serves as input for the robot controller to compensate for the motion. This paper presents a motion analysis

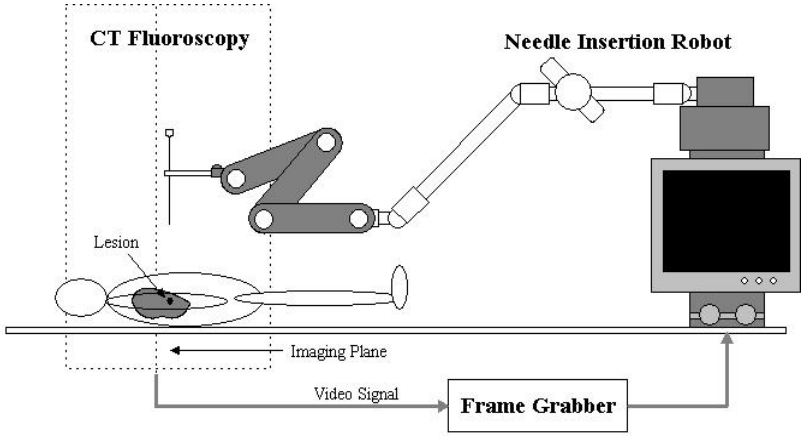

Fig. 1 CT fluoroscopy guided robotically assisted lung biopsy

method to estimate the 3D motion of a target lesion using the CT fluoroscopy image sequences. The method aims at tracking the lesion's motion no matter whether the lesion is inside or outside the imaging plane.

Motion tracking is a well-established area of research in computer vision. The tracking methods can be broadly classified into feature-based and region-based approaches. The feature-based approaches rely on the extraction of a sparse set of image primitives corresponding to distinctive scene features and the matching of such features over time using some form of search procedure. The type of features can be points, lines, edges, corners or contours. It is required for these approaches that the features being tracked stay in the scene. For the purpose of tracking pulmonary lesions using CT fluoroscopy, the available features in the CT fluoroscopy image are blood vessels and bronchi. Due to respiratory motion, the features specified in one frame may not be in another frame. Therefore, the feature-based approaches cannot be applied.

The region-based approaches make direct and complete use of all available image intensity information, therefore eliminating the need to identify and model a special set of features to track. The tracking algorithm of this paper falls into this category. The algorithm minimizes the sum-of-squared differences (SSD) between two regions. Variants of this method have been presented in previous work by other researchers. For example, Rehg and Witkin [1] tracked affine deformations; Hager and Belhumeur [2] reformulated the tracker for real-time performance; Shi and Tomasi [3] connected images texture, numerical issues, and tracker performance; Szeliski and Coughlan [4] used the Levenberg-Marquardt method as a problem solver to track multi-bilinear patches; and Gleicher [5] introduced difference decomposition to solve the registration problem in tracking, where the difference was a linear combination of a set of basis vectors.

The goal of this tracking algorithm is to guide a surgical-assist system (a needle placement robot). Therefore, the algorithm must be robust to be clinically acceptable. In the existing literatures on robust tracking, Toyama and Hager [6] presented the "Incremental Focus of Attention" architecture, which switches between algorithms according to the visual environment to achieve robust performance. McCane et. al [7] presented a framework for merging the results of independent motion trackers using a classification approach. Chen et. al [8] used neighborhood relaxation with multi-candidate pre-screening to robustly track image regions.

2. METHODS

Using CT fluoroscopy, the anatomy in the imaging plane can be viewed in real-time. Although the target lesion may not always stay in the imaging plane due to respiratory motion, the location of the target lesion with respect to the imaging plane can be estimated from a pre-operative CT volume. As shown in Fig. 2, a small region A is initially selected by the physician in the pre-operative image such that the region is close to the target lesion and has rich texture. If this region and the target lesion are close enough, they will have approximately the same deformation. By registering the region A to its corresponding region B of the pre-operative CT volume, the location of the target lesion with respect to the imaging plane can be estimated using the local rigidity around the lesion. The framework of the algorithm is shown in Fig. 3. The position of the current CT fluoroscopy region in the CT volume is predicted from previous frames. The CT fluoroscopy image region is then registered to the CT volume in a coarse-to-fine fashion. A multi-start strategy is used

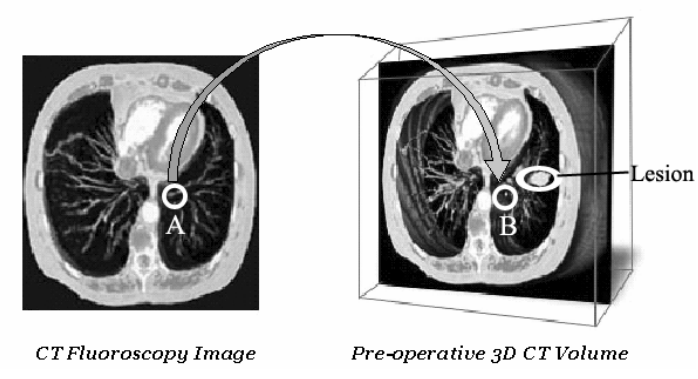

Fig.2 Local registration

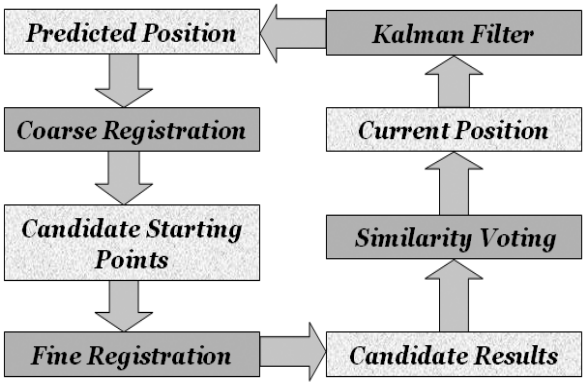

Fig.3 Algorithm framework

during the process, so multiple results are generated. The true motion of the CT fluoroscopy region is selected from these candidate results by similarity voting. The selected motion vector is then incorporated into a Kalman filter to predict the region's position for the next frame.

2.1 Coarse registration

The purpose of the coarse registration is to find good starting points for the fine registration. It is necessary that at least one of these starting points will lead to the solution of the fine registration. Since the CT fluoroscopy image has a large slice thickness (10 mm in our case), the CT volume is resampled to the same slice thickness. The original slice interval is preserved. Although the adjacent two slices have overlap in the resampled CT volume, it doesn't change the spatial position of each slice. Each slice of the resampled CT volume is then downsampled to a lower resolution. The sample rate is determined by the size of the convergence region of the fine registration so that at least one sampled pixel will fall inside the basin surrounding the minimum of the SSD residual. Cross correlation is performed at the pixel level to estimate the relationship between the CT fluoroscopy image region and the CT volume. The pixels best correlated with the center of the CT fluoroscopy image region are selected to be the starting points of the fine registration.

2.2 Fine registration

The lung has rich textures formed by blood vessels and bronchi, which allows the use of local information to register the CT fluoroscopy image region to the CT volume. As shown in Fig. 2, by registering the tissue inside the region to the pre-operative CT volume, the location of the target lesion with respect to the imaging plane can be determined using the local rigidity around the lesion in the CT volume. The shape of the region can be either a ring or a disk in this research, depending on which shape encloses more texture in the CT fluoroscopy image. The advantage of using a circular region is that it better approximates the motion of the region with respect to its surrounding area. Before the registration, the CT volume is smoothed by Gaussian filters. Gradient images are then calculated. The tracking algorithm is based on minimizing the SSD between the selected region in the CT fluoroscopy image and a corresponding one in the CT volume. The Levenberg-Marquardt method [9] is used to solve the SSD objective function with a good starting point. Compared to other methods for solving least square problems, the Levenberg-Marquardt method is more robust, which usually results in a larger convergence region.

The CT fluoroscopy image usually has a large slice thickness. As shown in Fig. 4, the 2D CT fluoroscopy image is actually the projection of all the tissue inside the dashed box on the imaging plane. As a result, the CT fluoroscopy image cannot be registered to the preoperative CT volume directly. What can be done is to find a subvolume of the preoperative CT volume corresponding to the dashed box, and then stack the tissue of the subvolume together to generate a synthetic CT fluoroscopy image. If the real and the synthetic CT fluoroscopy images match each other, the registration is done. The following equation (1) describes the relationship between the 2D CT fluoroscopy image $\boldsymbol{I}$ and its corresponding 3D subvolume $\boldsymbol{V}$ in preoperative CT:

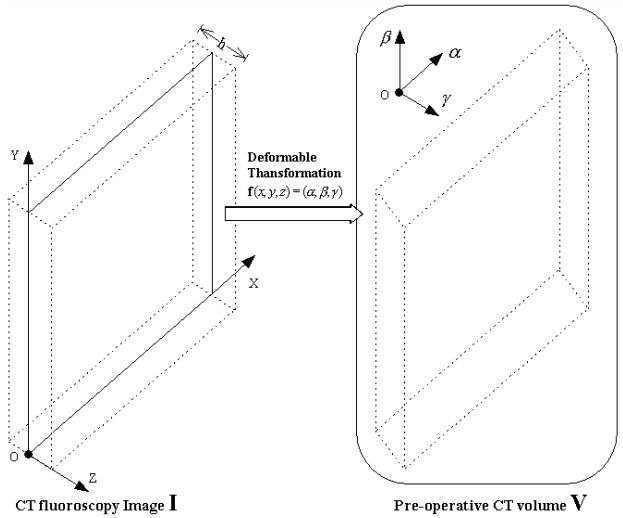

Fig.4 CT fluoroscopy image and CT volume

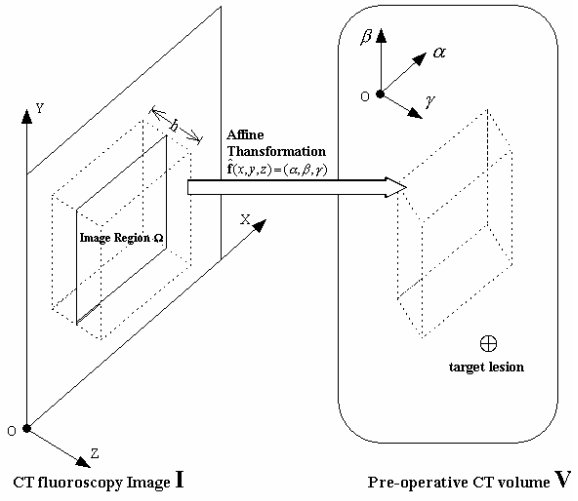

Fig.5 Local registration

$$\mathbf{I}(x,y) = \frac{1}{h}\int_{z=-h/2}^{h/2} \mathbf{V}(\,\mathbf{f}(x,y,z)\,)dz + \mathbf{n}(x,y) \tag{1}$$

where **n** represents the noise, **f** represents a deformable transformation and h is the slice thickness of the CT fluoroscopy image. The integral is to project the subvolume on a plane to simulate the CT fluoroscopy image. It can be realized by tri-linear interpolation.

The deformation of the lung tissue can be very complicated. However, if the region to be registered is small, the deformation can be approximated using 3D affine transformation, which is a 3 by 4 parameter matrix represented by $\hat{\mathbf{f}}$ in Fig. 5. Equation (2) shows a basic objective function to be minimized. It is a weighted SSD between the CT fluoroscopy region and the synthetic region.

$$O(\boldsymbol{\mu}) = \sum_{(x,y)\in\Omega} w_a^2 w_b^2 \left[\frac{1}{h}\int_{z=-h/2}^{h/2} \mathbf{V}(\,\hat{\mathbf{f}}(x,y,z,\boldsymbol{\mu})\,)dz - \mathbf{I}(x,y)\right] \tag{2}$$

where $\boldsymbol{\mu}$ is a vector of twelve parameters of the affine transformation to be estimated; w_a is a Gaussian window function which gives the central pixels a larger weight. As a result, the outer pixels that can not be well represented by the affine transformation will not have a large effect on the objective function. w_b is another weighting function with gives the brighter pixels a larger weight as described in Equation 3.

$$w_b = \max\left[\frac{1}{h}\int_{z=-h/2}^{h/2} \mathbf{V}(\,\hat{\mathbf{f}}(x,y,z,\boldsymbol{\mu})\,)dz,\quad \mathbf{I}(x,y)\right] \tag{3}$$

The purpose of w_b is to emphasize the bright texture generated by the blood vessels and bronchi. If both of the pixels being compared are dark, their effect on the objective function will be minimal. The dark pixels usually correspond to the parenchyma of the lung, which can only generate noise to the SSD.

Normally, SSD is only used for the intra-modality registration. It can be shown that SSD is the optimum measure when two images only differ by Gaussian noise [10]. In this tracking problem, the image regions being compared come from the same machine but two different imaging methods – the standard CT and the CT fluoroscopy, which results in an intensity difference between the two regions. If this intensity difference is big, SSD misregistration will happen. Although there is no simple relationship to map the intensity from one imaging method to the other, the effect of the intensity difference can be reduced by subtracting the mean of the intensity from both regions. The revised objective function is then the following:

$$O(\boldsymbol{\mu}) = \sum_{(x,y)\in\Omega} w_a^2 w_b^2 \left\{\left[\frac{1}{h}\int_{z=-h/2}^{h/2} \mathbf{V}(\,\hat{\mathbf{f}}(x,y,z,\boldsymbol{\mu})\,)dz - C_V\right] - \left[\mathbf{I}(x,y) - C_I\right]\right\} \tag{4}$$

where C_v and C_I are the intensity means of the synthetic region and the CT fluoroscopy region respectively. They are given in equations 5 and 6, in which N is the number of pixels inside the region of interest Ω.

$$C_v = \frac{1}{N}\sum_{(x,y)\in\Omega} \frac{1}{h}\int_{z=-h/2}^{h/2} \mathbf{V}(\,\hat{\mathbf{f}}(x,y,z,\boldsymbol{\mu})\,)dz \tag{5}$$

$$C_I = \frac{1}{N}\sum_{(x,y)\in\Omega} \mathbf{I}(x,y) \tag{6}$$

Technically, the mean of the synthetic region is a function of the unknowns of the affine transformation. However, the SSD function is solved iteratively. Since the step size is very small between two consecutive iterations, the change of mean is tiny and has little effect on the solution. Therefore, the intensity mean of the synthetic region is treated as a

constant. It was found that the subtraction of the mean intensity from the original intensity is very effective in improving the robustness of the algorithm.

The reason to use SSD is to replace the image registration problem with non-linear minimization. Using the standard optimization methods, the tracking algorithm can be very efficient. There is no need to evaluate the entire objective function to calculate its derivatives. The derivatives can be obtained from the image gradients calculated offline from the pre-operative CT volume.

2.3 Similarity voting

Both multi-start and importance sampling strategies are used to avoid local minima, which result in multiple candidate transformations of the CT fluoroscopy image region. Many of these transformations are false results caused by local minima of the SSD residual error. Even if the global minimum is found, it may not be the true transformation due to the effect of the image noise. In response to these problems, similarity voting is used to distinguish the true transformation from the false transformations. The following measures are used to evaluate the similarity between the CT fluoroscopy image region and the synthetic region.

a) Residual error of the pixels inside the region
b) Residual error of the pixels surrounding the region
c) Residual error of the bright pixels inside the region
d) Axis orthogonality

In the above similarity measures, (a) is the value of the SSD objective function and (b) gives the "global" information of the region. The bright pixels in (c) are segmented using automatic thresholding [11]. They are emphasized because the blood vessels and bronchi appear bright in the image, which form most of the useful texture. The assumption for (d) is that the deformation is small, so the principle axes are nearly perpendicular to each other after the affine transformation. Although none of the four similarity measures is sufficient to independently determine the true transformation, the true transformation should generally have larger values of the similarity measures compared to the other candidate transformations. If the candidate transformations are ranked in terms of overall evaluations from all the similarity measures, the true transformation is very likely to have the best overall ranking. As shown in Fig. 6, the candidate transformations are evaluated and ranked for each similarity measure. The four rankings for each candidate are summed. The candidate transformation with the smallest sum is selected. With enough similarity measures, the similarity voting can be robust in picking out the true motion vector. In addition, it avoids using arbitrarily set thresholds, which is problem specific and prone to failure.

2.4 Kalman filtering

The Kalman filter is a classical tool that produces estimates optimal in the least-squares sense of the state of a dynamic system from noisy measurements and an uncertain model of the system dynamics. The use of predictors like Kalman filter can reduce computational cost by only processing a certain area of the image, and the noise outside that area does not influence the processing [12]. The price paid for the efficiency is the increased possibility of dropping the true motion vector. Using the multi-start strategy mentioned above, the dropping problem can be alleviated. In this research, the dynamic system equation of the Kalman filter is chosen to be a constant velocity update. Although it is possible to

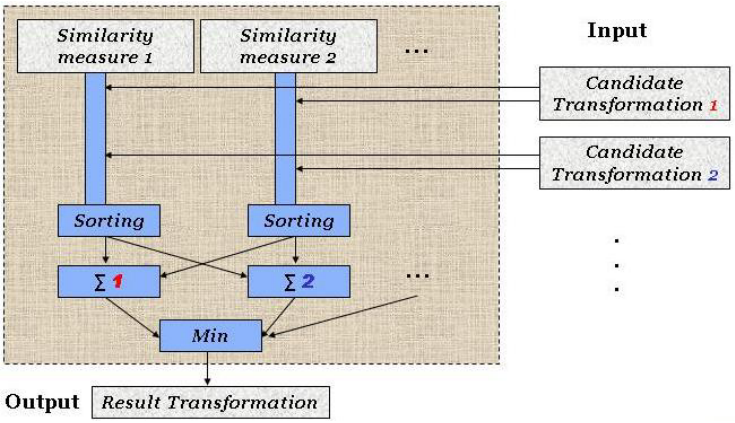

Fig.6 Similarity voting

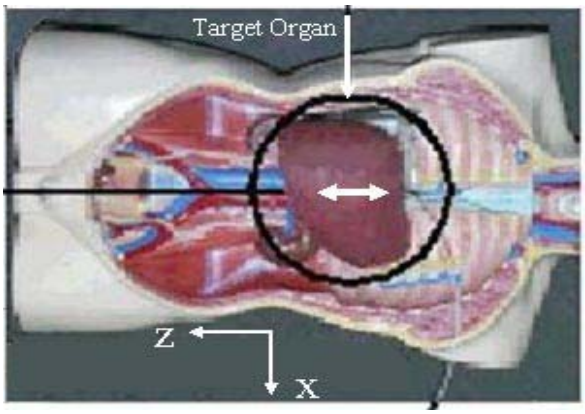

Fig.7 Respiratory motion simulator


Fig.8 Motion tracking test of respiratory motion simulator

use a more complicated motion model to boost the Kalman filter's performance, the disadvantage is that the tuned Kalman filter is effective only for a narrow class of motions. Given that the motion of the target lesion is not limited to respiratory motion, and the patient can hold the breath during the needle insertion, it may not be worthwhile to employ a complicated motion model.

3. EXPERIMENTAL RESULTS

Three experiments were carried out using both synthetic and real CT fluoroscopy images (Siemens Somatom Volume Zoom CT). The synthetic CT fluoroscopy image was generated from a pre-operative CT volume of the lung region with 2.0 mm slice thickness and 0.72 mm pixel size. As a result, the exact match of the CT fluoroscopy image region is known. Clustering analysis of the SSD objective function was performed. It was observed that, for the objective function to converge, the starting point had to be within ±2mm for translation and within ±20° for each of the three rotation angles. A similar experiment was made using two different CT volumes of the same patient. With one volume as the pre-operative CT volume, and the other one to generate synthetic CT fluoroscopy images, similar results were obtained.

The second experiment was performed with CT fluoroscopy images from a respiratory motion simulator. As shown in Fig. 7, the target organ had one degree of freedom and was moved by a motor in the cranio-caudal direction. After a pre-operative volumetric scan was obtained, the data was saved in DICOM format with 512 x 512 resolution, 1mm slice thickness and 0.74 mm pixel size. The motor of the simulator was aligned with the CT gantry such that the motion of the phantom was perpendicular to the imaging plane (Fig. 8). Since the organ phantom was rigid, its motion curve was the same as that of the motor. A sequence of CT fluoroscopy images were taken at about 6Hz and saved in the DICOM format. The resolution of the CT fluoroscopy image was 256 x 256, with 10mm slice thickness and 1.48 mm pixel size. A 10mm radius region was selected on the first frame of the CT fluoroscopy image. The motion tracking was initialized. Fig. 9 shows the motion tracking results compared to the known motion curve of the motor at two different motion modes. The major motion in the Z direction (perpendicular to the imaging plane) was very consistent with the motion curve of the motor. The maximum displacement in X and Y directions was under 0.8 mm and treated as noise. The average position error was under 1mm, if the delay in time is ignored. As might be expected, the tracking algorithm tended to have a small position error when the target's motion was slow or close to zero.

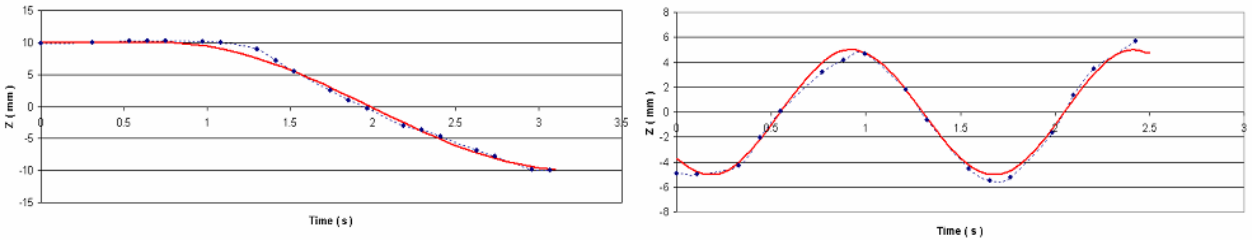

Fig.9 Experimental results of simulator test. The left and right curves show the displacement of the target organ at two different motions. The displacement in Z direction is compared to the ground truth of the motion curve of the phantom.

Fig.10 Swine study

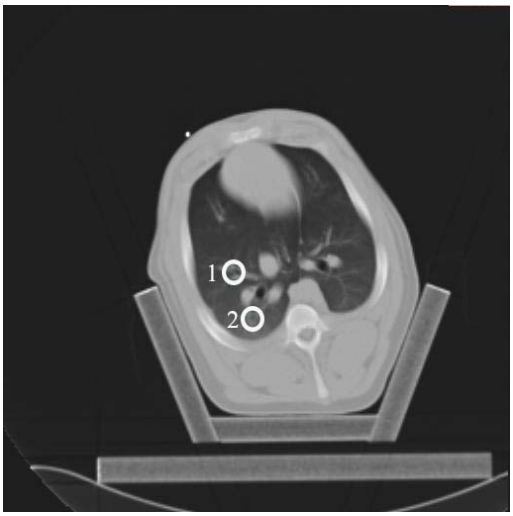

Fig.11 CT fluoroscopy image of Swine study

The third experiment was a swine study (Fig. 10). This study was done under an approved animal care protocol. Most of the parameters were the same as the second experiment except that the pixel size was 0.47 mm for the preoperative CT image and 1.48 mm for the CT fluoroscopy image. The respiratory rate was set at 11 cycles per minute using a ventilator. Instead of using the frame grabber, the CT fluoroscopy images were obtained from the DICOM files saved by the CT. Since the CT machine can only save about the last twenty frames of CT fluoroscopy images, only a portion of the respiratory cycle was recorded. Fig. 12 shows the trajectories of two different locations of the CT fluoroscopy image during the respiratory cycle. The curves on the left and right correspond to the regions 1 and 2 in Fig. 11 respectively. While the motion patterns are similar in some respects, there are enough differences to note that different regions in the lung (even regions that are close by) will move with different patterns. With a 2G Hz Pentium 4 CPU, the frequency of the tracking algorithm was between 2Hz and 3Hz. The algorithm worked for most of the lung region. For some regions with little texture, the SSD algorithm failed, most likely because it did not have enough information to uniquely determine the match of the CT fluoroscopy image region in the CT volume.

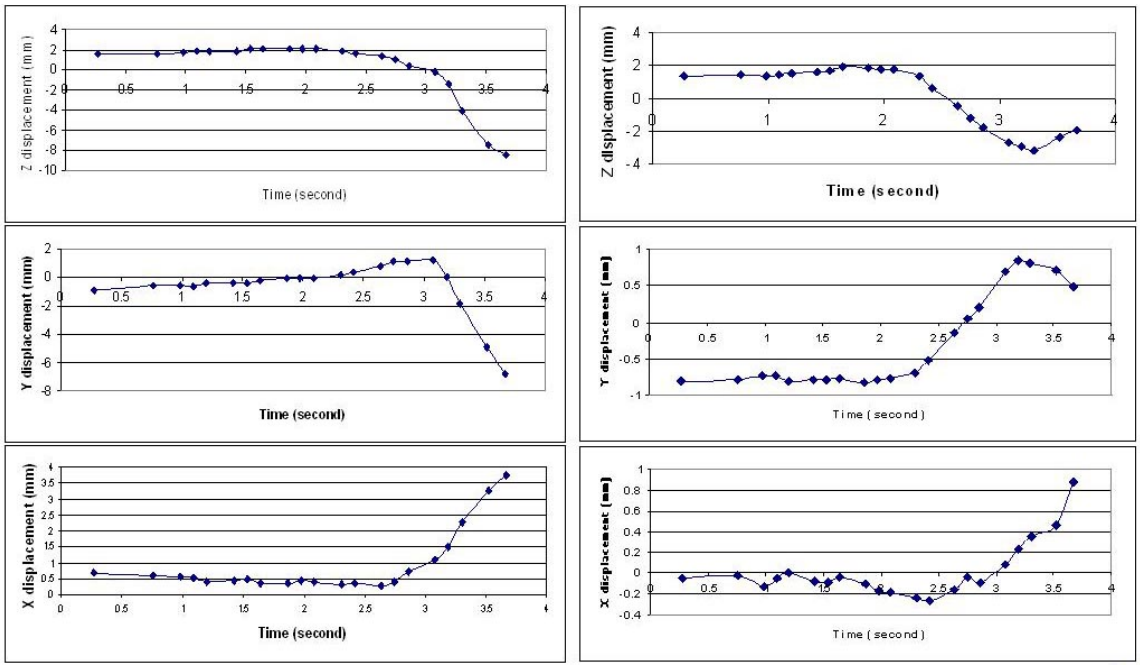

Fig.12 Motion estimation. The pictures on the left / right shows the motion curves of region 1 / 2 in Fig. 11.

4. DISCUSSION

The basic assumption of region-based tracking methods is that the image pixels inside the tracking region undergo a single affine transformation. This assumption is often violated by the lung region. The disadvantage of such an approach is the lack of resolution – a fixed region size means that the approach cannot adapt to the underlying data [13]. One way to improve the algorithm is to use more sophisticated deformable model such as spline deformation rather than the affine transformation. As a result, the analysis region can be bigger, which should also make the algorithm more robust. Another problem that may affect the accuracy is that, instead of tracking the target lesion directly, the algorithm tracks the region inside the CT fluoroscopy imaging plane. If the local rigidity assumption of tracking region and the target lesion is violated, more error will be introduced.

Compared to the conventional motion tracking problem in computer vision, the tracking problem of this paper has the following properties:

(1) since the CT fluoroscopy image region changes over time, the matching template is not fixed, and the efficient tracking algorithm presented in [2] cannot apply;
(2) there is no illumination change in the problem. Instead, the difference between CT images and CT fluoroscopy images needs to be adjusted;
(3) the algorithm does not need to deal with the object occlusion. The lack of texture and large deformation are the major reasons to lose the tracking.

Experimental results showed that the algorithm worked reasonably well in most of the lung area. At some locations where there was not much texture, the algorithm could not track the target. This could be mitigated by selecting a region close to the target lesion that the algorithm is able to track, and use that region to infer the location of the target. In addition, the algorithm can be combined with other methods to increase its robustness. For instance, we can track the motion of the chest wall and try to correlate this motion with the motion of the target lesion. In case the SSD algorithm fails, the motion of the chest wall may be able to be used to estimate the motion of the target lesion.

In order to evaluate the algorithm, it is desirable to use a second tracking method to obtain the ground truth of the target motion. However, this is often impractical. For example, although the magnetic tracker can track the motion of the internal organ, its performance is very poor in a CT scanner. Currently, we are trying to get a 4D CT dataset [14] to evaluate the algorithm.

Our future work will focus on improving the robustness of the algorithm. A post-failure recovery strategy should be established. Chest wall motion tracking and its correspondence to the target lesion will be investigated. Additional animal studies are also planned.

5. CONCLUSION

This paper presented a new algorithm to track the motion of pulmonary lesions using CT fluoroscopy. Initial experimental results showed that the algorithm worked well with a respiratory motion simulator and has reasonable performance in a swine study. Future experiments are planned to incorporate the algorithm with the robotically assisted lung biopsy system.

To the best of the authors' knowledge, there are no other published methods that are non-invasive and can track pulmonary lesions at any location of the lung. The algorithm is automated and can run in nearly real-time. Although the current speed of the algorithm (2-3Hz) is not fast enough to keep up with the CT fluoroscopy update rate (6 Hz), with a faster CPU and the optimization of the algorithm, it is hoped that the algorithm will work in real-time.

ACKNOWLEDGMENT

The authors gratefully acknowledge the longstanding support and advice of Charles White, MD at University of Maryland Medical Center and Gregory D. Hager, PhD at the Johns Hopkins University. We also thank David Lindisch, RT, for his assistance with the experiments at Georgetown University. The robot was designed and built by the Urology

Robotics Laboratory at Johns Hopkins Medical Institutions under the direction of Dan Stoianovici, PhD. This work was primarily supported by U.S. Army grant DAMD17-99-1-9022 and National Cancer Institute (NIH) grant 1 R21 CA094274-01A1. Research infrastructure was also provided by the National Science Foundation under ERC cooperative agreement EEC9731478.

REFERENCES

1. J. M. Rehg, A.P. Witkin, "Visual tracking with deformation models", Proc. IEEE intl. Conf. On *Robotics and Automation,* vol.1, 844-850, IEEE, Sacramento, CA, USA, 1991.
2. G. Hager and P. Belhumeur, "Efficient Region Tracking With Parametric Models of Geometry and Illumination", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. **20**(10): 1205-1039, 1998.
3. J. Shi and C. Tomasi, "Good features to track", IEEE Conf. On *Computer Vision and Pattern Recognition*, 593-600, IEEE, Seattle, WA, USA, 1994.
4. R. Szeliski and J. Coughlan, "Spline-based image registration", *Intl. J. of Computer Vision*, vol. **22**, 199-218, 1997.
5. M. Gleicher, "Projective Registration with Difference Decomposition", IEEE Conf. On *Computer Vision and Pattern Recognition*, 331-337, IEEE, San Juan, Puerto Rico, 1997.
6. K. Toyama and G. Hager, "Incremental Focus of Attention for Robust Vision-Based Tracking", *Int. J. Computer Vision*, **35**(1), 45-63, 1999
7. B. McCane, B. Galvin, K. Novins, "Algorithmic Fusion for More Robust Feature Tracking", *Intl. J. of Computer Vision*, vol. **49**(1), 79-89, 2002.
8. Y. Chen, Y. Lin and S. Y. Kung, "A Feature Tracking Algorithm Using Neighborhood Relaxation With Multi-Candidate Pre-Screening", Proc. of IEEE Int'l Conf. on *Image Processing*, vol. II, 513--516, IEEE, Lausanne, Switzerland ,1996.
9. J. J. More, "The Levenberg-Marquardt Algorithm: Implementation and Theory, in Numerical Analysis", *Lecture Notes in Mathematics*, vol. 630, G. A. Watson, ed., 105-116, Springer-Verlag, Berlin, 1977.
10. J. V. Hajnal, D. L.G. Hill, D. J. Hawkes, *Medical Image Registration*, CRC Press LLC, 2001.
11. S Hu and E. A. Hoffman, "Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images", *IEEE Trans. Medical Imaging*, vol. **20**(6), 2001.
12. M. Kohler, *Using the Kalman Filter to track Human Interactive Motion Modeling and Initialization of the Kalman Filter for Translational Motion*, Technical Report, 1997.
13. S. Krüger, *Motion Analysis and Estimation using Multiresolution Affine Models, Ph.D. thesis*, 1998.
14. D. A. Low, M. Nystrom, E. Kalinin, and P. Parikh, "A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing", Medical physics, **30**(6), 1254-63, 2003.