Published in final edited form as:

Phys Med Biol. 2013 March 21; 58(6): 1789–1808. doi:10.1088/0031-9155/58/6/1789.

Robust Fluoroscopic Tracking of Fiducial Markers: Exploiting the Spatial Constraints

Rui Li and

Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

Gregory Sharp

Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

Rui Li: bucslirui@gmail.com

Abstract

Two new fluoroscopic fiducial tracking methods that exploit the spatial relationship among the multiple implanted fiducial to achieve fast, accurate and robust tracking are proposed in this paper. The spatial relationship between multiple implanted markers are modeled as Gaussian distributions of their pairwise distances over time. The means and standard deviations of these distances are learned from training sequences, and pairwise distances that deviate from these learned distributions are assigned a low spatial matching score. The spatial constraints are incorporated in two different algorithms: a stochastic tracking method and a detection based method. In the stochastic method, hypotheses of the “true” fiducial position are sampled from a pre-trained respiration motion model. Each hypothesis is assigned an importance value based on image matching score and spatial matching score. Learning the parameters of the motion model is needed in addition to the learning the distribution parameters of the pairwise distances in the proposed stochastic tracking approach. In the detection based method, a set of possible marker locations are identified by using a template matching based fiducial detector. The best location is obtained by optimizing the image matching score and spatial matching score through non-serial dynamic programming. In this detection based approach, there is no need to learn the respiration motion model. The two proposed algorithms are compared with a recent work using multiple hypothesis tracking algorithm which is denoted by **MHT**[19]. Phantom experiments were performed using fluoroscopic videos captured with known motion relative to an anthropomorphic phantom. The patient experiments were performed using a retrospective study of 16 fluoroscopic videos of liver cancer patients with implanted fiducials. For the motion phantom data sets, the detection based approach has the smallest tracking error (μ_{err} : 0.78 – 1.74 mm, σ_{err} : 0.39 – 1.16 mm) for the images taken at low exposure (50 mAs). At higher exposure (500 mAs), the stochastic method gave the best performance (μ_{err} : ~ 0.39 mm, σ_{err} : ~ 0.27 mm). In contrast, the tracker (**MHT**) that does not model the spatial constraints only performs well when there is no occluded fiducial. With the RANDO phantom data, both of our proposed methods performed well and have the mean tracking errors around ~ 1.8 mm with the standard deviations ~ 0.93 mm at 100 mAs and ~ 0.91 mm with 0.88 mm standard deviation at 500 mAs. The **MHT** tracker has the largest tracking errors with mean ~ 4.8 mm) and standard deviation ~ 2.4 mm in both sessions with the Rondo phantom data. On the patient data sets, the detection based method gave the smallest error (μ_{err} : 0.39 mm, σ_{err} : ~ 0.19 mm). The stochastic method performed well (μ_{err} : ~ 0.58 mm, σ_{err} : ~ 0.39 mm) when the patient breathed consistently, the accuracy dropped to (μ_{err} : ~ 1.55 mm) when the patient breathed differently across sessions.

1. INTRODUCTION

Organ motion management plays an important role in the treatment of cancer using radiotherapy [1, 2, 3]. Motion management is recommended for patients with respiratory motion greater than 5mm [4], as respiration induced tumor motion before and during treatment often prevents targeted, escalated dose delivery. Additionally, tumor motion requires large margins to ensure sufficient tumor coverage, which can adversely affect normal tissue toxicity [5, 6]. Therefore, for the most precise treatments, such as stereotactic body radiotherapy, real-time monitoring of tumor position should be considered [7, 8]. When accompanied with beam gating [9] or beam tracking [10], these treatments can achieve accuracy on the order of 2 mm.

To accurately monitor the tumor position in the presence of organ motion, one common approach is to surgically implant radiopaque markers in or near the tumor mass [11, 12, 13, 14, 15, 16]. These markers are typically between 1 and 5 mm in size, and are made of gold or other high contrast metals. The Hokkaido RTRT system [11] tracks the position of a single spherical gold marker using 30Hz stereoscopic fluoroscopy, and uses beam gating to stop the beam when the target moves out of position. The IRIS system [17] developed at the Massachusetts General Hospital (MGH) is another system that can provide fluoroscopic videos during the treatment. IRIS is equipped with a tracking system that can track multiple implanted markers. Other systems, such as the Cyberknife [10] and BrainLab ExacTrac [18] are also capable of marker tracking, but operate at lower frame rates. Another common approach is to use the electromagnetic markers. The Calypso system (Calypso Medical Technologies Inc, Seattle, WA) is such a platform based on detection of electromagnetic markers for treatment setup and monitoring during radiation therapy. The disadvantages of this type of systems are: the size of the marker used is larger than the radio-opaque fiducial marker; and no image of the surrounding anatomy is available. Therefore, we focus our attention on the fluoroscopic tracking of fiducial markers.

To better characterize the position of the tumor center, multiple markers are often preferred over a single marker [19]. However, this complicates the tracking task as solving the motion correspondence problem [20, 21] is not a trivial task. In this paper, we focus on the tracking of multiple implanted fiducial markers. The common theme of the previously proposed tracking systems [19, 22, 23, 24] is the need for some form of motion model trained beforehand. The disadvantages of a trained motion model are that it requires patient-specific training data and needs periodic updates. We propose and investigate two new tracking methods that exploit the spatial constraints among the implanted markers.

The motivations for incorporating the spatial constraints are

- i. prevent loss of track caused by weak image information due to low image contrast and other objects that have similar appearance such as nearby anatomical structures and other implants;
- ii. mitigate the effect of motion model inaccuracies caused by changes in respiratory motion within or between sessions;
- iii. leverage correlated marker motion, as nearby markers do not move independently [16, 25].

The two tracking methods proposed in this paper employ the spatial constraints in a similar fashion. The first tracking method is stochastic. We formulate tracking as an inference problem in a graphical model and use an approximate inference algorithm to obtain the solution. The approximation is sampling based, and is a modified version of the well known particle filter tracking method [26]. A product of the image matching score and an

aggregated spatial constraints score is used to weigh the samples. Then, a trained motion model is used to propagate samples from time t to $t + 1$. As a result, the first method combines motion model information and spatial information of the multiple implanted markers. The second tracking method is deterministic and performs tracking through optimization. The score function we optimize is the sum of the image matching score and the spatial constraints score. In each frame of the input fluoroscopic video sequence, we use template matching to find out the candidate locations of the markers. The tracked marker locations are given by the “optimal” locations of the candidates that give the best combined score. Thus the second approach does not use motion model information at all. As observed in [27, 28], a patient can breath quite differently from session to session. Using a learned motion model could produce inaccurate predictions and in this case could potentially lead to inaccurate tracking.

2. METHODS AND MATERIALS

Image-based tracking systems use image pixel values to determine the location of a target. However, as shown in Fig. 1, there might be multiple possible target locations if image matching is used alone. Many researchers have explored using motion models or regression modeled models [19, 22, 23]. There are two main problems with these approaches. Firstly, a certain amount of training data is needed to learn the model parameters reliably. As a result, an extra imaging study is required. Secondly, the learned model parameters may not generalize well across lengthy treatment sessions, as the patient might change his breathing pattern.

Our proposed method addresses these challenges by leveraging on the spatial constraints among the implanted fiducial markers. Compared to learning the motion model or the regression model parameters, the distances among the different implanted fiducial are less likely to change from session to session. Therefore, they are more reliable and easier to learn.

2.1. Image Matching

A set of potential fiducial marker locations can be found by matching a set of predefined fiducial marker templates as shown in Fig. 1(a) with the image using normalized correlation coefficients (NCC). Template matching is widely used for object tracking [29] and NCC is a reliable similarity metric for template matching [30]. The NCC metric is given as

$$sc_{img_i} = \frac{\sum_{r,c} [I(r,c) - \bar{I}_{u,v}] [T(r-u, c-v) - \bar{T}]}{\sqrt{\sum_{r,c} [I(r,c) - \bar{I}_{u,v}]^2 \sum_{r,c} [T(r-u, c-v) - \bar{T}]^2}}, \quad (1)$$

which computes the score at location u, v in image I by comparing the image intensities against a template image with pixel indices r, c . $\bar{T}$ is the mean intensity of the template, and $\bar{I}_{u,v}$ is the mean intensity value of the image region located under the template at location u, v . Finally, sc_{img_i} is simply the NCC matching score at u, v . The NCC score ranges from -1 to 1 , which we can re-normalize to the range of 0 to 1 .

A higher NCC value indicates a more probable marker location, but using NCC alone is often insufficient for accurate marker localization. When using NCC to find the template of Fig. 1(a) within the image region of interest shown in Fig. 1(b), we see in Fig. 1(c) that there are many regions which are not the marker, but have a high NCC score. Therefore, we introduce additional information to prune away the false positives and obtain a better estimate of the marker position.

2.2. Spatial Relationship Between Markers

A spatial score term is motivated by the facts that implanted fiducial markers move in tandem and the inter-marker distances do not change significantly over time. Let D_{ij} denote the distance between the fiducial markers $\mathbf{x}_i$ and $\mathbf{x}_j$, and let $\hat{D}_{ij}$ and σ_{ij}^2 represent the mean and the standard deviation of D_{ij} over time. Our model assumes that D_{ij} follows a Gaussian distribution ψ_{ij} , where

$$\psi_{ij} = \frac{1}{2\pi\sigma_{ij}^2} \exp\left(-\frac{(D_{ij} - \hat{D}_{ij})^2}{2\sigma_{ij}^2}\right), \quad (2)$$

$\hat{D}_{ij}$ and σ_{ij}^2 can be determined either by using training data or by using expert knowledge. In our experiments, $\hat{D}_{ij}$ and σ_{ij}^2 are found from training data.

The spatial model defines how far or how close the pairwise distances of the marker positions can deviate from the learned distributions. We use this similarity score in combination with the image similarity metric to locate a configuration of fiducial markers that give the best combined score.

2.3. Stochastic Tracking: Sequential Mean Field Monte Carlo (sMFMC)

In this stochastic solution, we formulate the movement correlations of multiple markers in a graphical model as shown in Fig. 2. Let $\mathbf{x}_{i,t}$ denote the location of the i -th marker at time t . With time frozen at t , we drop the subscript t , and define a Markov network over the marker locations. The network has nodes $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_j, \dots, \mathbf{x}_n\}$ which represent the locations of the n markers. An observation $\mathbf{z}_j$ is defined as the image region centered at the true marker location $\mathbf{x}_j$. Image matching is performed using a parametrized cylindrical marker model, as shown in Fig. 1(a), which defines a marker by its length l and its in-plane angle θ . We can find estimates $\hat{l}^*$ and $\hat{\theta}^*$ from a single image using exhaustive search. We conduct the search in the image by matching the templates generated at different combinations of l and θ values in the image with the manually initialized marker locations. Once $\hat{l}^*$ and $\hat{\theta}^*$ are found, their values can be updated using limited to a subset of possible l and θ values. For location estimation, we limit our attention to the 2D coordinates $\mathbf{x}$ of the fiducial markers and they are initialized manually. However, Alg. 2 can be used to initialize the locations of the fiducial markers automatically.

The tracking task in the stochastic formulation is to infer the posterior probability $p(\mathbf{X}|\mathbf{Z})$. Based on Bayes' theorem, we have:

$$p(\mathbf{X}|\mathbf{Z}) = \frac{p(\mathbf{X})p(\mathbf{Z}|\mathbf{X})}{p(\mathbf{Z})} \propto p(\mathbf{X})p(\mathbf{Z}|\mathbf{X}), \quad (3)$$

where $p(\mathbf{Z})$ is a constant scaling factor which does not affect the optimization of $p(\mathbf{X}|\mathbf{Z})$ and thus is left out in the following derivations. Based on the model in Fig. 2, $p(\mathbf{Z}|\mathbf{X})$ can be factorized as:

$$p(\mathbf{Z}|\mathbf{X}) = \prod_{i=1}^n p(\mathbf{z}_i|\mathbf{x}_i). \quad (4)$$

The configuration of locations $\mathbf{X}$ is implemented as an undirected graph, where the edges carry prior knowledge about the spatial relationship between markers. The prior $p(\mathbf{X})$ can be modeled using a Gibbs distribution:

$$p(\mathbf{X}) = \frac{1}{\mathcal{L}} \prod_{(i,j) \in E} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \prod_{i=1}^n \psi_i(\mathbf{x}_i), \quad (5)$$

where $\mathcal{L}$ is the normalization constant, ψ_{ij} is the spatial prior, and the local prior ψ_i is determined by the dynamics prior, *i.e.*, the predicted marker location based on the learned motion model which corresponds to step 2 in Alg. 1. Eqs. 3, 4 and 5 show that $p(\mathbf{x}_i|\mathbf{Z})$ is affected by three factors: the local prior ψ_i , local image observation $\mathbf{z}_i$ and the constraints enforced by its neighborhood through ψ_{ij} which is defined in Eq. 2. The exact inference of $p(\mathbf{X}|\mathbf{Z})$ is computationally expensive and not suitable for real-time tracking applications. Mean field approximation [21] is used for efficient approximation.

Mean field approximation approximates $p(\mathbf{X}|\mathbf{Z})$ with a fully factorized variational

distribution: $q(\mathbf{X}) = \prod_{i=1}^n q_i(\mathbf{x}_i)$. The optimal $q^*(\mathbf{X})$ is obtained by minimizing the Kullback-Leibler (KL) divergence:

$$\begin{aligned} KL(q(\mathbf{X}) \| p(\mathbf{X}|\mathbf{Z})) &= \int_{\mathbf{X}} q(\mathbf{X}) \log \frac{q(\mathbf{X})}{p(\mathbf{X}|\mathbf{Z})} \\ &= -\mathcal{H}(q(\mathbf{X})) = \int_{\mathbf{X}} q(\mathbf{X}) \log p(\mathbf{X}, \mathbf{Z}) + \log p(\mathbf{Z}), \end{aligned} \quad (6)$$

where $\mathcal{H}$ is the entropy function. The fully factorized form of $q(\mathbf{X})$ allows us to update each $q_i(\mathbf{x}_i)$ individually while holding the rest unchanged.

The update equation for $q_i(\mathbf{x}_{i,t})$ at time instance t is:

$$q_{i,t}(\mathbf{x}_{i,t}) = \frac{1}{\mathcal{L}_i} p(\mathbf{z}_{i,t}|\mathbf{x}_{i,t}) \cdot \int p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1}) q_{i,t-1}(\mathbf{x}_{i,t-1}) \cdot \exp \left\{ \sum_{j \in \mathcal{N}(i)} \int_{\mathbf{x}_j} q_{j,t}(\mathbf{x}_{j,t}) \log \psi_{ij}(\mathbf{x}_{i,t}, \mathbf{x}_{j,t}) \right\}, \quad (7)$$

where $\mathcal{L}_i$ is the normalization constant and $\mathcal{N}(i)$ denotes the neighborhood of node i . The derivation of this equation is found in Yu and Wu [21].

A sequential Mean field Monte Carlo (sMFMC) method is used to realize the mean field updates. For each $q_{i,t}(\mathbf{x}_{i,t})$, we use a set of samples s and their corresponding weights π to

approximate the actual distribution, thus $q_{i,t}^{(m)(k)} \sim \{s_{i,t}^{(m)(k)}, \pi_{i,t}^{(m)(k)}\}_{m=1}^M$, where M is the total number of samples and k represents the k -th iteration of the mean field update. The sMFMC algorithm is summarized in Alg. 1. The tracker starts tracking by generating a set of samples for each marker. In the first two frames, we manually initialize the locations of the fiducial markers. The samples for the third frame are then generated from the learned second order motion model. For subsequent frames, the samples are generated from the learned motion model based on the previously tracked marker locations. Once the samples are generated for the current frame, the quality of each sampled locations is assessed by combining the image matching score and “messages” passed from the samples of the neighboring markers. Each “message” is a weighted sum of the log of ψ_{ij} defined in Eq. 2. The underlying intuition of this algorithm is a model of the collaborative relationships among different markers, where the implanted fiducial markers do not move independently.

2.4. Deterministic Tracking: Nonserial Dynamic Programming (NSDP)

An alternative to stochastic tracking is to generate a discrete set of hypotheses, and then solve the tracking problem by pruning away false positive detections. This idea is illustrated

in Fig. 3. Template based matching is used to generate candidates, and a composite score of both the image matching score sc_{img} and the spatial constraints score sc_{sp} is computed. The spatial constraint score $sc_{sp_{ij}}$ is equal to ψ_{ij} defined in Eq. 2. The locations of markers are determined by finding the set of candidates that give the best overall score via Non-serial Dynamic Programming (NSDP) [31].

The score function is defined as:

$$sc = \sum_{i=1}^n \left(sc_{img_i} + \sum_{j \in \Omega} sc_{sp_{ij}} \right), \quad (8)$$

where $j \in \Omega$ refers to the j -th marker that has an edge connecting it to the marker i .

A simple rule is required when constructing the graph in Fig. 3(d), to ensure an efficient solution. The rule states that all the nodes of the graph should lay on the boundary of a simple, planar polygon. A simple polygon by definition is a polygon whose boundary does not cross itself [32]. Fig. 4 provides an example of simple polygon and a non-simple polygon.

By restricting the form of the graph, not all pairwise distances among the markers are used. However, as noted in the book by Bertele and Brioschi [31], there exists an optimal elimination order if we have a simple polygon. The order states that a node i may be eliminated, if it lies on exactly one triangle of the current triangulated polygon. In Fig. 4(a), both node 1 and node 3 can be eliminated during the first step of the optimization, and either one leads to the optimal solution by forming the dual graph (E.g., the two connected blue nodes in Fig. 4(a)) of the triangulated polygon. This elimination order always exists as long as all the nodes of the triangulated polygon lie on the contour of the polygon. Thus, a computationally efficient NSDP algorithm can be obtained and its solution is globally optimal [33].

The optimization algorithm works by sequentially eliminating nodes (marker location) following the order discussed in the previous paragraph. After eliminating nodes $1, \dots, i-1$, node i is in exactly one triangle, and denoting the other nodes in this triangle as $i.a = j$ and $i.b = k$, we compute the score of the best placement $\mathbf{x}_i$ of i as a function of the locations for markers j and k . The score for this step is stored in $V[j, k](\mathbf{x}_j, \mathbf{x}_k)$ where V is a lookup table. When we get to the last two vertices, we solve for their best locations and trace backward to find the best locations for all other nodes.

Alg. 2 runs in $O(nm^3)$ or uses $O(nm^2)$ space, where n is the number of the fiducial markers and m is the number of possible locations for each marker. During tracking, we can exploit temporal consistency and only consider locations that are within a fixed distance from the previously tracked location. As a result, we can limit m to be a small number. In our implementation, we set $m = 3$. The main computational burden of this tracking by detection algorithm is the candidates generation step, *i.e.*, the template matching step. To achieve good performance, we limit the search window size and use a fast template matching algorithm [30].

2.5. Material and Implementation Detail

The proposed tracking algorithms were implemented and evaluated retrospectively using both phantom and patient data. Retrospective research on patient images was performed under approval of the Institutional Review Board, MGH protocol #2010P-002050.

For phantom experiments, we used a 4D motion phantom from Washington University [34] and placed fifteen centimeters of solid water phantom on the motion table. The solid water phantom slabs are of size $30 \times 30 \text{ cm}^2$ each, and were imaged laterally through 30 cm to simulate fluoroscopic imaging of an adult abdomen. The gold fiducial markers we wished to track were taped onto the entrance surface of the solid water at 85 cm source to surface distance. Wires and cables of various sizes were used to simulate non-marker structures, and were affixed directly onto the cover of the imaging panel. By placing them on the panel, they act as a non-moving occluding objects that the fiducial markers must pass through. An overview of the experimental setup is shown in Fig. 5 where the magenta arrows point to the fiducial markers and the cyan arrows point to the wires used for occlusion.

Three sets of phantom data were captured, with different geometries for the occluding objects. Within each data set, the dose rate was varied to test the influence of image quality on tracking performance. The details of the data sets are shown in Table 1. Fig. 6 shows a sample sequence of the data set where two markers are being occluded. Sample pictures with varying imaging doses are shown in Fig. 7 and demonstrate the range of image quality that was evaluated.

A study using a RANDO (R) anthropomorphic phantom (The Phantom Laboratory, Salem, NY) was also conducted to demonstrate capability of the tracker. Copper wires were placed on the imaging panel as in previous experiments. Gold fiducials and implanted into a styrofoam block, which was placed between the x-ray tube and the phantom, and was subject to known motion trajectories using a computer controlled phantom. Two different imaging doses and captured two sets of data for experiments. The sample images are shown in Fig. 8.

The retrospective patient data that was evaluated in this study is summarized in Table 2. The data consists of 16 fluoroscopic videos from different imaging sessions of 3 liver patients, with 3–5 implanted fiducial markers each. The positions of the markers were manually labeled to determine ground truth. For each patient, we use 40 frames from the first session to learn the motion model parameters and the distance statistics $\hat{D}_{ij}$ and σ_{ij} for the pairwise distance distribution 2. The learned second order motion model $\mathbf{x}_t = A_0 + A_1\mathbf{x}_{t-1} + A_2\mathbf{x}_{t-2}$ is used to propagate samples in step 2 of Alg. 1.

We compare our proposed methods (stochastic: sMFMC; deterministic: NSDP) with the method MHT (multiple hypothesis tracking) described in Tang, Sharp and Jiang [19]. For both sMFMC and MHT, we train a second order linear motion model in advance to propagate the samples/hypotheses. For MHT, a local search window size of 47×47 (pixels) is used for each marker. In some cases, this size is too small and we may lose tracking of the markers, and we reverted to larger search window sizes at the expense of slowing down the tracker.

3. RESULTS AND DISCUSSION

We evaluated these three trackers on a 2.66GHz Intel PC with 3GB memory using MATLAB (MathWorks, Natick, MA). Both sMFMC and MHT took about 0.13 second per frame, while NSDP took about 0.2 second per frame. The speed of MHT deteriorates when a larger search window size is needed to keep track of the markers.

3.1. Phantom Experiments

Fig. 9 summarizes the signal-to-noise ratio as a function of image expose for our experiments. For all algorithms and all marker configurations that we have tested, a radiographic technique of 50 mAs per frame was needed for reliable tracking. None of the

trackers were able to work reliably for the images captured with a lower signal-to-noise ratio.

To evaluate the performance of each algorithm in detail, we manually marked the ground truth position of the markers on fluoroscopic sequences captured for both phantom experiments. Fig. 10 shows the results for solid water phantom experiment, and Fig. 12 shows the results for the anthropomorphic phantom. In Fig. 12, we can see that trackers that exploit spatial constraints (sMFMC and NSDP) give similar tracking error for both the low and high technique. In contrast, the baseline method only performs well when there is no occlusion (session 1). Even for higher exposures, the performance of the baseline method starts to deteriorate with occlusions. Poor image quality was also found to adversely impact the performance of the stochastic method (sMFMC). When image noise is high, the hypotheses in sMFMC become weak and sample impoverishment occurs. The deterministic method NSDP was found to be robust with respect to changes in both image quality and occlusions.

Fig. 11 and 13 demonstrates qualitative tracking results demonstrates qualitative tracking results from solid water phantom session 3 (low imaging dose and with occlusions), and from the anthropomorphic phantom 500 mAs session. Without spatial information, the baseline method (MHT) failed to track at least one of the markers in the demonstrated frames. Even using spatial information, the stochastic sampling based method (sMFMC) had intermittent errors due to the poor image matching at the sampled image locations. However, due to its stochastic nature, the tracker was always able to recover from loss of track as demonstrated in frame 49 (lost track) and frame 69 (recovered) of Fig. 11. The deterministic tracker (NSDP) was consistently able to maintain track through these occlusions, because it maintains a balance between intensity and spatial information for all frames.

3.2. Liver Patient Experiments

The results from the three algorithms on experiments with liver patient data are summarized in Table 3. For the validation sets, there is no significant differences among the 3 methods. This is because the motion model parameters are learned from the frames in the same session. The only exceptions were lateral images of patient 2 and patient 3. For these images, we found that when 2 markers that lie very close to each other, or close to other implemented material, the markers were interchanged or confused with other patient implants. In the testing sessions, we observe larger tracking errors for sMFMC and MHT due to the fact that the learned motion models did not accurately model the respiratory motion seen in later sessions for the same patient. NSDP gives the most accurate results across different sessions and different patients because the distances among the different markers are less likely to change than the motion dynamics.

In Fig. 14, we show sample tracked frames from patient 3 session 3 to illustrate the problem with other implanted materials. The trackers (sMFMC and NDSP) that exploit spatial constraints are able to track through occlusions from other implanted materials and did not become confused with vertebrae in the background. In contrast, the MHT tracker is seen to lose track of the green marker from frame 62 and could not recover even when multiple search window sizes were considered.

In our experiments, we did not consider the issue of marker migration, which could affect spatial constraints. Detecting marker migration could be accomplished by testing against a conservative threshold on the matching score. A similar approach has been explored in [19, 35].

4. CONCLUSIONS

We propose and investigate two robust fiducial tracking algorithms that use spatial constraints to improve tracking reliability when multiple markers are being tracked. These algorithms rely on consistent inter-marker distances in two-dimensions. Experiments demonstrate that both methods improve tracking reliability, especially for cases where the markers pass through occluding objects. For low dose imaging, a detection based tracking algorithm (NSDP) was found to have the highest reliability of the three methods tested. These results are encouraging, and demonstrate that the NSDP is well suited for clinical applications of fluoroscopic tracking in image-guided radiotherapy [36].

There are several interesting extensions that can be realized from the NSDP framework: (1) we can employ tissue models to improve the penalty function used in the pairwise constraints; (2) this framework could be considered for markerless tracking [24, 37, 38] as a method for tracking multiple smaller features; (3) if we model the distance constraint in the three dimensional space, this framework could be extended to using three-dimensional spatial constraints for marker tracking in rotational therapy [39, 40]. Our current 2D tracking algorithms can also benefit from modeling the distance constraint in 3D.

Acknowledgments

This research was supported by NIH/NCI grant P01-CA21239.

References

1. Langen KM, Jones DT. Organ motion and its management. *Int J Radiat Oncol Biol Phys.* 2001; 50(1):265–278. [PubMed: 11316572]
2. Jiang SB. Radiotherapy of mobile tumors. *Semin Radiat Oncol.* 2006; 16(4):239–248. [PubMed: 17010907]
3. Bortfeld T, et al. Robust management of motion uncertainty in intensity-modulated radiation therapy. *Operation Research.* 2008; 56(6):3874–3900.
4. Keall PJ, et al. The management of respiratory motion in radiation oncology. Report of AAPM Task Group 76. *Medical Physics.* 2006; 33:3874–3900. [PubMed: 17089851]
5. Bortfeld, T., et al. *Image-guided IMRT.* Springer; 2006.
6. Leonard, L., et al. *Clinical radiation oncology.* Elsevier; 2007.
7. Guckenberger M, et al. Potential of image-guidance, gating and real-time tracking to improve accuracy in pulmonary stereotactic body radiotherapy. *Radiotherapy and Oncology.* 2009; 91(3): 288–295. [PubMed: 18835650]
8. Li G, et al. Motion monitoring for cranial frameless stereotactic radiosurgery using video-based three-dimensional optical surface imaging. *Medical Physics.* 2011; 38:3981–3995. [PubMed: 21858995]
9. Kudo H, Hill B. Respiration gated radiotherapy treatment: a technical study. *Phys Med Biol.* 1996; 49:2053–2067.
10. Adler JR, et al. The cyberknife: a frameless robotic system for radiosurgery. *Stereotactic and Functional Neurosurgery.* 1997; 69:124–128. [PubMed: 9711744]
11. Shirato H, et al. Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor. *Int J Radiat Oncol Biol Phys.* 1999; 48:435–442. [PubMed: 10974459]
12. Shirato H, et al. Physical aspects of a real-time tumor-tracking system for gated radiotherapy. *Int J Radiat Oncol Biol Phys.* 2000; 48:1187–1195. [PubMed: 11072178]
13. Shirato H, et al. Feasibility of insertion/implantation of 2.0-mm-diameter gold internal fiducial markers for precise setup and real-time tumor tracking in radiotherapy. *Int J Radiat Oncol Biol Phys.* 2003; 56:240–247. [PubMed: 12694845]

14. Shimizu S, et al. Use of an implanted marker for the positioning of prostate and bladder cancers. *Int J Radiat Oncol Biol Phys.* 2000; 48:1591–1597. [PubMed: 11121666]
15. Shimizu S, et al. Detection of lung tumor movement in real-time tumor-tracking radiotherapy. *Int J Radiat Oncol Biol Phys.* 2001; 51:304–310. [PubMed: 11567803]
16. Harada T, et al. Real-time tumor-tracking radiation therapy for lung carcinoma by the aid of insertion of a gold marker using bronchofiberscopy. *Cancer.* 2002; 95:1720–1727. [PubMed: 12365020]
17. Berbeco R, et al. Integrated radiotherapy imaging system (IRIS): design considerations of tumor tracking with Linac gantry-mounted KV x-ray systems. *Phys Med Biol.* 2004; 49:243–255. [PubMed: 15083669]
18. Willoughby TR, et al. Evaluation of an infrared camera and x-ray system using implanted fiducials in patients with lung tumors for gated radiation therapy. *Int J Radiat Oncol Biol Phys.* 2006; 66:568–575. [PubMed: 16887290]
19. Tang XL, Sharp GC, Jiang SB. Fluoroscopic tracking of multiple implanted fiducial markers using multiple object tracking. *Phys Med Biol.* 2007; 52:4081–4098. [PubMed: 17664596]
20. Bar-Shalom, Y.; Jaffer, AG. Tracking and Data Association. Academic Press; 1988.
21. Yu T, Wu Y. Collaborative tracking of multiple targets. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2004; 1:834–841.
22. Sharp GC, et al. Prediction of respiratory tumour motion for real-time image-guided radiotherapy. *Phys Med Biol.* 2004; 49:425–440. [PubMed: 15012011]
23. Lin T, et al. Fluoroscopic tumor tracking for image-guided lung cancer radiotherapy. *Phys Med Biol.* 2009; 54:981–992. [PubMed: 19147898]
24. Cui Y, et al. Multiple template-based fluoroscopic tracking of lung tumor mass without implanted fiducial markers. *Phys Med Biol.* 2007; 52:6229. [PubMed: 17921582]
25. Kitamura K, et al. Registration accuracy and possible migration of internal fiducial gold marker implanted in prostate and liver treated with real-time tumor-tracking radiation therapy (RTRT). *Radiat Oncol.* 2002; 62:275–281.
26. Isard M, Blake A. Contour tracking by stochastic propagation of conditional density. *European Conference on Computer Vision (ECCV).* 1996:343–356.
27. George R, et al. Investigation of patient, tumour and treatment variables affecting residual motion for respiratory-gated radiotherapy. *Phys Med Biol.* 2006; 51:5305–5320. [PubMed: 17019040]
28. Hong SM, et al. Real-time prediction of respiratory motion based on a local dynamic model in an augmented space. *Phys Med Biol.* 2011; 56:5305–5320.
29. Brunelli, R. Template Matching Techniques in Computer Vision: Theory and Practice. Wiley; 2009.
30. Lewis, JP. Vision Interface 120–123. Canadian Image Processing and Pattern Recognition Society; 1995. Fast template matching.
31. Bertele, U.; Brioschi, F. Nonserial Dynamic Programming. Academic Press; 1972.
32. de Berg, M., et al. Computational Geometry. Springer-Verlag; 2000.
33. Felzenszwalb P. Representation and detection of deformable shapes. *IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI).* 2005; 27:208–220.
34. Malinowski K, et al. Development of the 4D Phantom for patient-specific, end-to-end radiation therapy QA. *SPIE Medical Imaging.* 2007:120–123.
35. Jiang S, Kandasamy N. A dependable system architecture for safety-critical respiratory-gated radiation therapy. *Int Conf on Dependable Systems and Networks (DSN-DCCS).* 2006:141–151.
36. Jiang S, et al. Technical aspects of image-guided respiration-gated radiation therapy. *Med Dosim.* 2006; 31:141–151. [PubMed: 16690455]
37. Berbeco R, et al. Towards fluoroscopic respiratory gating for lung tumours without radiopaque markers. *Phys Med Biol.* 2005; 50:4481. [PubMed: 16177484]
38. Li R, et al. A feasibility study of markerless fluoroscopic gating for lung cancer radiotherapy using 4DCT templates. *Phys Med Biol.* 2009; 54:489–500.
39. Lewis J, et al. Markerless lung tumor tracking and trajectory reconstruction using rotational cone-beam projections: a feasibility study. *Phys Med Biol.* 2010; 55:2505–2522. [PubMed: 20393232]

40. Ali I, et al. An algorithm to extract three-dimensional motion by marker tracking in the kv projections from an on-board imager: four-dimensional cone-beam ct and tumor tracking implications. *Journal of Applied Clinical Medical Physics*. 2011; 12

Figure 1.

(a) The marker template used for matching is parameterized by length l , location $\mathbf{x}$ and the angle of rotation θ . Only small deviations in l and θ are tracked. (b) A sample input frame. (c) The score image from matching template (a) with image (b) using the normalized correlation coefficient. Bright pixels are candidate locations for the fiducial markers. It is evident that there can be many false positive matches, suggesting that image matching alone is insufficient for tracking the fiducial markers.

Figure 2.

Proposed graphical model for incorporating spatial constraints in sequential mean field Monte Carlo (SMFMC). $\mathbf{x}_{i,t}$ is the location of the i -th marker at time t and $\mathbf{z}_{i,t}$ is the corresponding image observation. In the model, a set of K image observations $\mathbf{Z}_t = \{\mathbf{z}_{1,t}, \dots, \mathbf{z}_{K,t}\}$ are generated from a hidden state containing the locations of K markers $\mathbf{X}_t = \{\mathbf{x}_{1,t}, \dots, \mathbf{x}_{K,t}\}$. Undirected edges connect the markers, and all marker positions are tracked jointly. Tracking is formulated as the inference problem $p(\mathbf{X}_t | \mathbf{Z}_t)$, where marker positions are estimated from image observations. We sometimes drop the subscript t for clarity, and describe the solution to the posterior $p(\mathbf{X} | \mathbf{Z})$.

Figure 3.

Deterministic tracking using nonserial dynamic programming (NSDP). For input **(a)**, we perform marker detection with predefined marker templates. This yields multiple detections as shown in **(b)** where candidate locations are selected from pixels with high matching scores. The optimal locations for markers **(c)** are obtained based on the triangulations shown in **(d)** through NSDP. The white edges encode the spatial relationship between different markers and they are obtained through triangulation. The purple edge connecting the two blue nodes is the dual graph of the triangulated polygon.

Figure 4.

The set of pairwise constraints in (a) is allowed for NSDP tracking, because all of the vertices lie on the boundary of a simple polygon. In contrast, configuration (b) is not allowed because there is a vertex that lies inside the polygon.

Figure 5.
Experimental setup for capturing phantom data. The fiducial markers are shown with magenta arrows, while static occlusions are shown with cyan arrows.

Figure 6. Sample cropped images from session 1 (top row), session 2 (middle row), and session 3 (bottom row) of the motion phantom. For clarity, we only show the sample images taken with the highest imaging technique (125kVp, 500mAs).

Figure 7. Sample cropped images taken with varying imaging doses from 50 mAs (left column), and 500 mAs (right column) of the motion phantom. The first row shows all three fiducials, while the second row is zoomed to show the region near the occluding objects.

Figure 8. Sample cropped images of the RANDO® phantom with inserted gold fiducial and wires to simulate occlusions.

Figure 9.

Image signal to noise ratio plotted against exposure for the phantom experiment (125 kVp). The signal to noise ratio is computed as $SNR = (\mu_{fg} - \mu_{bg}) / \sigma_{bg}$, where μ_{fg} and μ_{bg} refer to the average pixel values of the object foreground and background, and σ_{bg} refers to the noise standard deviation of the background. All of the tracking algorithms failed to track reliably at an imaging technique lower than 30 mAs, which corresponds to a minimum SNR of approximately 1.4.

Figure 10.

Evaluation of average tracking error (in mm) for three tracking methods and two different imaging doses. As described in Table 1, session 1 is without occluding structures, session 2 has a single occluding structure, and session 3 has two occluding structures. The baseline method (MHT) only performs well when there is no occlusion, whereas the methods that exploit spatial constraints (sMFMC and NSDP) can tolerate occlusions. Reducing the imaging dose adversely affects the performance of the stochastic method (sMFMC), because the set of hypothesized marker locations becomes weak and sample impoverishment occurs. The deterministic method NSDP is more robust with respect to both varying imaging doses and occlusions.

Figure 11.

Sample tracked frames from session 3 at low exposure (50 mAs). Detected markers are shown with colored rectangles, and incorrect tracking results are highlighted with the colored arrows pointing to the actual marker locations. The baseline method (MHT) failed to track at least one marker in all frames shown. The stochastic sampling based method (sMFMC) has intermittent problems due to the poor image matching at the sampled image locations. However, due to its stochastic nature, the tracker recovers from loss of track as demonstrated in frame 49 (lost track) and frame 69 (automatic recovery). The deterministic tracker (NSDP) was able to track through occlusions because it performs an exhaustive search to optimize a balance between imaging matching and spatial information.

Figure 12.

Evaluation of average tracking error (in mm) for three tracking methods on the Rando® phantom. It is shown that only the methods that exploit spatial constraints (sMFMC and NSDP) can tolerate occlusions. In these experiments, there is no observable performance difference between sMFMC and NSDP. However, the dynamical model learnt for sMFMC does not change between the training and testing session. In real situation, there is always difference of the patients' respiration patterns between the simulation session and the treatment session.

Figure 13. Sample tracked frames of the Rando® phantom at the 500mAs. Detected markers are shown with colored rectangles, and incorrect tracking results are highlighted with the colored arrows pointing to the actual marker locations.

Figure 14.

Sample tracked frames. Detected markers are shown with colored rectangles, and incorrect tracking results are highlighted with the colored arrows pointing to the actual marker locations. The first row shows the tracking results of MHT when applied to patient 3 session 3 in lateral view, while the second row and third row demonstrate the results of applying sMFMC and NDSP to the same sequence. By incorporating spatial constraints among the implanted fiducial markers, both sMFMC and NDSP are able to maintain the tracking through occlusion with other implants and vertebrae.

Table 1

Experimental technique for the motion phantom data. We consider in detail two different imaging techniques, and three different imaging scenarios (sessions). Power cables and coils are used to create occlusions with the tracking of gold fiducial markers during the captured fluoroscopic video sequences.

session	kVp	mA	ms	# of frames	description
1	125	10	5	84	2cm peak to peak motion along Y-axis. 12 cycles per minute. All clips are visible.
		50	10	84	
2	125	10	5	320	Same motion table parameters. Power cables and coils are used. One clip is occluded in some frames.
		50	10	320	
3	125	10	5	320	Same motion table parameters. Power cables are adjusted so there are two clips become occluded in some frames.
		50	10	320	

Table 2

Data used in retrospective patient study. 16 fluoroscopic videos from 3 liver patients were used. Ground truth positions of 3 – 5 implanted markers were manually labeled to provide ground truth. For each patient, we use 40 frames from the first session to learn the dynamical model parameters and the distance statistics $\hat{D}_{ij}$ and σ_{ij} for the pairwise distance distribution 2 (these are denoted T for training, and V for validation). For step 2 of Alg. 1, $\mathbf{x}_t = A_0 + A_1\mathbf{x}_{t-1} + A_2\mathbf{x}_{t-2}$ is used to propagate samples, while for MHT the same model is used to propagate hypotheses.

patient	session	note	# of frames	
			AP	LAT
1	1	T: 40 V: 40	100	100
	2		80	80
	3		140	140
	4		140	140
2	1	T: 40 V: 40	100	100
3	1	T: 40 V: 40	100	100
	2		100	100
	3		140	140

Table 3

Comparison of the three tracking methods on the validation and test datasets. The error statistics reported are in mm.

patient	session	LAT				AP			
		MHT	sMFCM	Mean	Std. Dev.	MHT	sMFCM	Mean	Std. Dev.
1	1: V	0.34	0.42	0.35	0.19	0.35	0.40	0.31	0.34
	2	0.51	0.64	0.34	0.33	0.39	0.58	0.29	0.30
	3	0.49	0.67	0.33	0.29	0.40	1.15	0.32	0.38
	4	9.45	8.56	1.82	6.88	2.01*	8.22	0.16	1.00
2	1: V	1.44	0.52	0.50	0.69	0.42	0.47	0.35	0.26
	1: V	4.32*	0.62	0.57	3.85*	0.42	0.48	0.41	0.32
	2	12.37*	0.72	0.40	10.62*	0.43	0.59	0.41	0.30
3	3	7.62*	0.94	0.45	5.57*	0.49	0.97	0.41	0.44
									0.22

Numbers with a "*" indicate that multiple window sizes have been tried and reported numbers are the best results.

Algorithm 1

Sequential Mean Field Monte Carlo

-
- 1: Set $k = 0$, sample $q_{i,t-1}(\mathbf{x}_{i,t-1})$ and $q_{i,t-2}(\mathbf{x}_{i,t-2})$ for $\{\tilde{s}_{i,t-1}^{(m)(k)}\}_{m=1}^M$ and $\{\tilde{s}_{i,t-2}^{(m)(k)}\}_{m=1}^M$ correspondingly.
 - 2: $\forall \tilde{s}_{i,t-1}^{(m)(k)}$ and $\tilde{s}_{i,t-2}^{(m)(k)}$, sample from $p(\mathbf{x}_{i,t}|\mathbf{x}_{i,t-1}, \mathbf{x}_{i,t-2})$ for $\{s_{i,t}^{(m)(k)}\}_{m=1}^M$ using $\mathbf{x}_t = A_0 + A_1\mathbf{x}_{t-1} + A_2\mathbf{x}_{t-2}$.
 - 3: **for** $k=1:\text{max_iter}$ **do**
 - 4: Calculate the “message” from neighbors:

$$M_{i,t}^{(m)} = \prod_{j \in N(i)} \pi_{j,t}^{(m)} \log \psi_{ij}(s_{i,t}^{(m)(k)}, s_{j,t}^{(m)(k)}).$$

- 5: Compute image matching likelihood for each $s_{i,t}^{(m)(k)}: w_{i,t}^{(m)} = p(\mathbf{z}_{i,t} | s_{i,t}^{(m)(k)})$ using Eq. 1.
- 6: Re-weight the samples by:

$$\pi_{i,t}^{(m)(k)} = w_{i,t}^{(m)} \times \exp(M_{i,t}^{(m)}).$$

- 7: Normalize to obtain:

$$q_{i,t}^{(m)(k)}(\mathbf{x}_{i,t}) \sim \{s_{i,t}^{(m)(k)}, \pi_{i,t}^{(m)(k)}\}_{m=1}^M.$$

- 8: **end for**
-

Algorithm 2

Nonserial Dynamic Programming

-
- 1: **Input:** marker candidate locations $\mathbf{x}_i$'s, candidate image matching score sc_{img_i} 's, spatial scores for neighboring markers sc_{spij}
 - 2: **for** $i=1:n-2$ **do**
 - 3: Eliminate the i -th node, set $j=i.a$ and $k=i.b$.
 - 4: **for** each pair of locations $\mathbf{x}_j$ and $\mathbf{x}_k$ **do**
 - 5: Compute:

$$V[j, k](\mathbf{x}_j, \mathbf{x}_k) \leftarrow \max_{\mathbf{x}_i} sc(\mathbf{x}_i, (\mathbf{x}_j, \mathbf{x}_k)) + V[i, j](\mathbf{x}_i, \mathbf{x}_j) + V[i, k](\mathbf{x}_i, \mathbf{x}_k)$$

- 6: **end for**
 - 7: **end for**
 - 8: Find $\mathbf{x}_{n-1}$ and $\mathbf{x}_n$ that maximize $V[n-1, n]$ and trace back to obtain the optimal locations for other markers.
-