TB, KR, PMB/233164, 13/12/2006

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **52** (2007) 1–15 UNCORRECTED PROOF

Robust fluoroscopic respiratory gating for lung cancer radiotherapy without implanted fiducial markers

Ying Cui[1], Jennifer G Dy[1], Greg C Sharp[2], Brian Alexander[2] and Steve B Jiang[2]

[1] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA
[2] Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

Received 20 September 2006, in final form 30 November 2006
Published DD MMM 2006
Online at stacks.iop.org/PMB/52/1

Abstract

For gated lung cancer radiotherapy, it is difficult to generate accurate gating signals due to the large uncertainties when using external surrogates and the risk of pneumothorax when using implanted fiducial markers. We have previously investigated and demonstrated the feasibility of generating gating signals using the correlation scores between the reference template image and the fluoroscopic images acquired during the treatment. In this paper, we present an in-depth study, aiming at the improvement of robustness of the algorithm and its validation using multiple sets of patient data. Three different template generating and matching methods have been developed and evaluated: (1) single template method, (2) multiple template method, and (3) template clustering method. Using the fluoroscopic data acquired during patient setup before each fraction of treatment, reference templates are built that represent the tumour position and shape in the gating window, which is assumed to be at the end-of-exhale phase. For the single template method, all the setup images within the gating window are averaged to generate a composite template. For the multiple template method, each setup image in the gating window is considered as a reference template and used to generate an ensemble of correlation scores. All the scores are then combined to generate the gating signal. For the template clustering method, clustering (grouping of similar objects together) is performed to reduce the large number of reference templates into a few representative ones. Each of these methods has been evaluated against the reference gating signal as manually determined by a radiation oncologist. Five patient datasets were used for evaluation. In each case, gated treatments were simulated at both 35% and 50% duty cycles. False positive, negative and total error rates were computed. Experiments show that the single template method is sensitive to noise; the multiple template and clustering methods are more robust to noise due to the smoothing effect of aggregation of correlation scores; and the clustering method results in the best performance in terms of computational efficiency and accuracy.

1. Introduction

Treatment errors related to respiratory organ motion may greatly degrade the effectiveness of conformal radiotherapy for the management of thoracic and abdominal lesions. This has become a pressing issue in the emerging era of image-guided radiation therapy (IGRT). Respiratory gated radiotherapy holds promise to precisely deliver a lethal dose to the tumour, while minimizing the incidence and severity of normal tissue complications, for mobile tumours in the thorax and the abdomen (Jiang 2006).

Precise target localization in real-time is particularly important for gated radiotherapy. In an idealized gated treatment, tumour position should be directly detected and the delivery of radiation is only allowed when the tumour is at the right position. However, direct detection of the tumour mass in real-time during the treatment is often difficult. Various surrogates, both external and internal, are used to identify the tumour position. Depending on the surrogates used, we categorized the respiratory gating into external (optical) gating and internal (fluoroscopic) gating. During gated treatment, the internal or external surrogate signal is continuously compared against a pre-specified range of values, called the gating window. When the surrogate signal is within the gating window, a gating signal is sent to the linac to turn on the radiation beam (Jiang 2006).

External gating techniques rely on the correlation between tumour location and the external surrogates, such as markers placed on the patient's abdomen (Chen *et al* 2001, Shirato *et al* 2003). The major weakness in external gating is the uncertainty in the correlation between the external marker position and internal target position (Jiang 2006). Current internal gating uses internal tumour motion surrogates such as implanted fiducial markers, as established by the Hokkaido group (Harada *et al* 2002, Seppenwoolde *et al* 2002, Gierga *et al* 2005). And it has been shown that internal surrogates can generate accurate gating signals. However, due to the risk of pneumothorax, the implantation of radiopaque markers in patients' lungs will unlikely become a widely accepted clinical procedure (Laurent *et al* 2000, Arslan *et al* 2002, Geraghty *et al* 2003). Therefore, it is crucial to be able to perform accurately gated treatment of lung cancer without implanted markers.

Previously, we have proposed, and shown the feasibility of, a template matching method to generate gating signals for lung radiotherapy without implanted markers (Berbeco *et al* 2005). The basic idea is (1) to generate a reference template which corresponds to the treatment position of the target in the gating window using fluoroscopic images acquired during patient setup, (2) to calculate the correlation scores between the reference template and the incoming fluoroscopic images acquired during treatment delivery, and (3) to convert the correlation score into gating signals. In this work, we will try to improve the accuracy and robustness of this technique by optimizing the way of generating the reference template as well as the way of computing the correlation scores. We will also perform a validation study by comparing the gating signals generated with the proposed techniques against those determined manually by clinicians using multiple patient datasets.

2. Methods and materials

2.1. Basic concepts of the proposed technique

Any on-board x-ray imaging systems, such as Varian OBI system (Varian Medical Systems, Palo Alto, CA, USA) and Elekta Synergy system (Elekta AB, Stockholm, Sweden), can be used as the hardware platform for the proposed technology. In this work, we plan to use an in-house developed imaging system as the platform, as shown in figure 1. This system is

Figure 1. The integrated radiotherapy imaging system (IRIS), used as the hardware platform for the proposed gating technology.

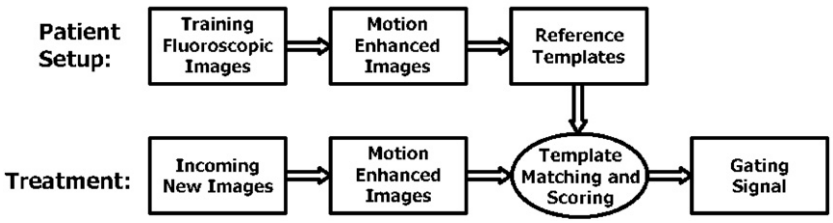

Figure 2. Block diagram for showing the process of the proposed correlation score-based method for generating the gating signal.

called the integrated radiotherapy imaging system (IRIS) (Berbeco *et al* 2004), which consists of two pairs of gantry-mounted diagnostic x-ray tubes and flat panel imagers. The system can be used to acquire a pair of real-time orthogonal fluoroscopic images for lung tumour tracking.

Figure 2 provides a block diagram to illustrate the proposed correlation score-based method for generating the gating signal. The components in the top row are the steps involved during patient set-up, while the components in the bottom row are the steps involved during the treatment session. In this section, we describe each of these components.

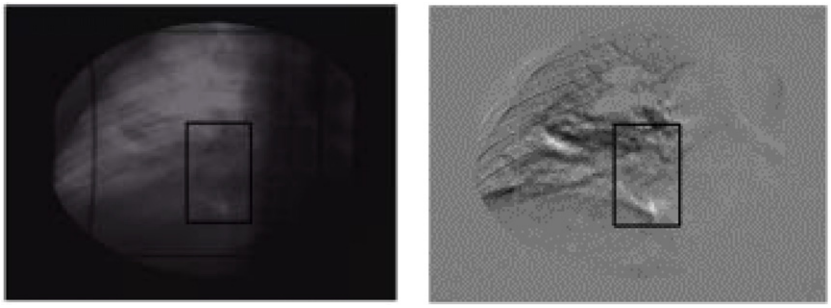

Figure 3. An illustration of the tumour image and region of interest (ROI). Left: original fluoroscopic image. Right: motion-enhanced image.

2.1.1. Training fluoroscopic images. Before each fraction of treatment, a sequence of orthogonal fluoroscopic images (e.g., with length of 10 s) are taken and used for patient setup. The tumour position in the gating window, where the treatment beam should be turned on, is identified either manually by a clinician or automatically by matching digitally reconstructed radiographs (DRRs) from the simulation 4D CT scan (Tang *et al* 2006). For the current study, fluoroscopic sequences of five patients (with the average length of 30 s) are used retrospectively to test the developed algorithms. For each patient, about 8–10 s of fluoroscopic images at the beginning of the sequence, which normally contains 2–3 breathing cycles, are used to simulate the setup or training period. These are the 'training fluoroscopic images' that we utilize to build our templates. A rectangular region of interest (ROI) is then created in those images, as shown in figure 3. This ROI is set to be large enough to contain tumour motion in the training period.

2.1.2. Motion enhancement. We would like to enhance the image of the moving tumour. We apply a pre-processing technique called motion enhancement (Jain *et al* 1995) on our training images. Given a sequence of images $I[t]$, where $t = 1, \ldots, N$ is the sequence number. We compute the average image, $\sum_{t=1}^{N} I[t]/N$, and the motion-enhanced image (MEI) is the difference between the original image and the average image, $I[t] - \sum_{t=1}^{N} I[t]/N$. The intuition behind MEI is that the average captures the static structures and smears the moving structures, thus the difference will amplify the moving structures. Figure 3 shows the original fluoroscopic image of a tumour in an ROI together with a motion-enhanced view of it.

2.1.3. Select reference templates. Typically, the gating window is set at the end-of-exhale (EOE) phase due to its longer duration and better stability. Previously, we have observed that the breathing cycle could be measured by the local relative fluoroscopic intensity fluctuation (Berbeco *et al* 2005). During normal breathing, the radiological path-length through the patients' lungs changes with the chest expansion and contraction, giving brighter and darker fluoroscopic images. The images are darker at the EOE phase and lighter at the inhale phases, which can be seen from figure 4. Here, we assume a certain gating duty cycle. Correspondingly, an intensity threshold is applied to the intensity fluctuation curve to define the gating window. All the images in the gating window, i.e., with intensities below this threshold, as shown in figure 5, are selected as EOE images for template development.

2.1.4. Template generation and matching. As shown in figure 4, the motion-enhanced images of the tumour are quite different at different breathing phases. Based on this observation, the

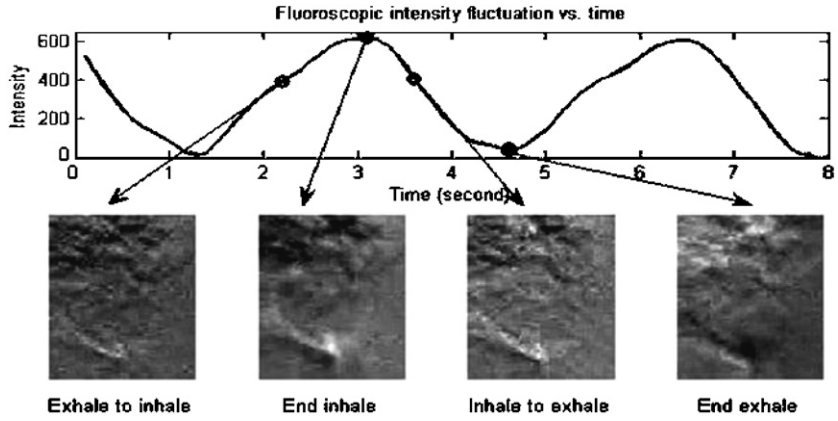

Figure 4. Motion-enhanced images at different breathing phases. Top: the fluctuation of the average image intensity in ROI showing different breathing phases. Bottom: motion-enhanced images at various breathing phases.

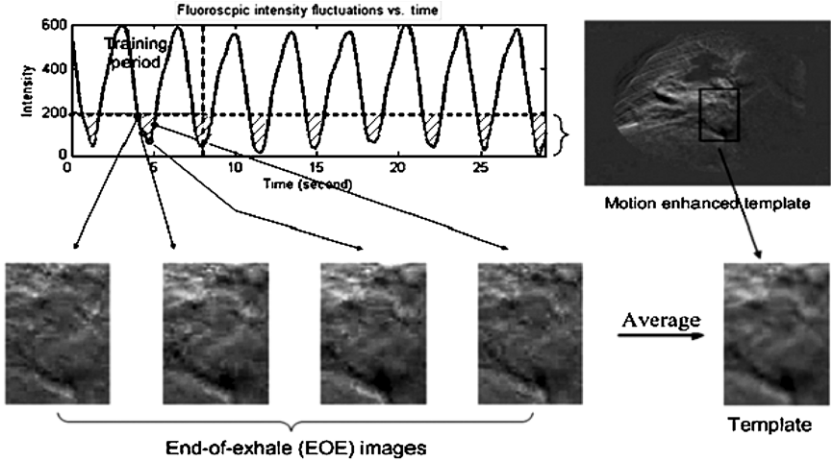

Figure 5. Single template method. Top left figure is the breathing waveform. Before the vertical dotted line is the training period. Under the horizontal dotted line is the end of exhale phase. Bottom figures showed different end of exhale images during training session, which are averaged to generate a single template.

calculated similarity between the reference template, developed using the EOE images, and the incoming images, acquired in real-time during the treatment delivery, is expected to vary significantly. Therefore, we are able to detect if the tumour in the incoming new image is at the EOE position and if we should turn on the beam. We will perform correlation score-based template matching techniques to estimate the similarity. Assuming that the image R and template T are of the same size $m \times n$, the normalized correlation coefficient (correlation score s) is defined as

$$s = \frac{\sum_m \sum_n (R_{mn} - \bar{R})(T_{mn} - \bar{T})}{\sqrt{\left(\sum_m \sum_n (R_{mn} - \bar{R})^2\right)\left(\sum_m \sum_n (T_{mn} - \bar{T})^2\right)}}. \tag{1}$$

Here $\bar{R}$ and $\bar{T}$ are the average intensity values of the image R and template T, respectively. If the score is high, that means the new image has a high similarity with the reference template, and the tumour is in the gating window. This procedure has been described in detail in a previous publication (Berbeco *et al* 2005). In this paper, we focus on how to generate robust gating signals using various template generation and matching methods. We will describe these methods in detail in section 2.2.

2.1.5. Gating based on score. Based on the correlation scores from all training images, we determine a threshold score that generates the pre-set duty cycle. This threshold is then applied to the correlation scores calculated in real-time during the treatment to generate the gating signal. When the score is above this threshold value, it indicates that the therapy beam should be enabled. Otherwise, the therapy beam should be turned off. One can think of the correlation score as a surrogate signal to generate the gating signal.

2.2. Score based gating methods

Our study evaluates the use of three different template generating and matching methods for gated treatment of lung cancer using fluoroscopy images, which will be described in this section. Before we proceed, we define our notations for clarity: R is the ROI in an incoming fluoroscopic image, T_i is the ith reference template (such as one of EOE templates in the EOE gating window), the sign $\oplus$ represents correlation, s is the score used for generating the gating signal, i.e., $g = H(s - s_0)$, where s_0 is the threshold score, $H(x)$ is the Heaviside step function. $H(x) = 0$ for $x < 0$ and $H(x) = 1$ for $x \geqslant 0$. The gating signal $g = 1$ means beam on while $g = 0$ means beam off.

2.2.1. Single template method ($s = R \oplus f(T_i)$). We build a single EOE template, $T = f(T_i)$, by combining all of the EOE motion-enhanced images. There may be many ways to combine the images. In this work, the combined single template, also called reference template, is simply the average of the motion-enhanced EOE images. During treatment, the correlation score between the reference template and each incoming motion-enhanced frame is calculated. A high correlation score indicates that the incoming image is similar to the reference and gating is enabled. The single template procedure is illustrated in figure 5, which is described in detail in our previous work (Berbeco *et al* 2005).

2.2.2. Multiple template method ($s = f(R \oplus T_i)$). We have noted that, to achieve a robust gating signal, only using one EOE template may not be enough. Our experiments showed that sometimes the gating signal resulting from a single EOE template is not satisfactory. That is because the correlation score curve may not be smooth, causing the gating signal to be unstable and noisy at the transition regions. Here, we suggest an alternative method. In this method, a pool of EOE templates is created. Instead of combining all EOE templates into one reference template, multiple templates are generated, each corresponding to an EOE frame, and a correlation score is computed for each template (i.e., $s_i = R \oplus T_i$). After a set of correlation scores are computed, it is necessary to choose an intelligent way to combine them to get a robust gating signal. We have explored three ways to combine the correlation scores: taking the maximum score, taking the average/weighted average score, and taking the moving average score (Whittaker and Robinson 1967). Among them, the best approach in terms of error rate is the average/weighted average. This procedure is explained in figure 6.

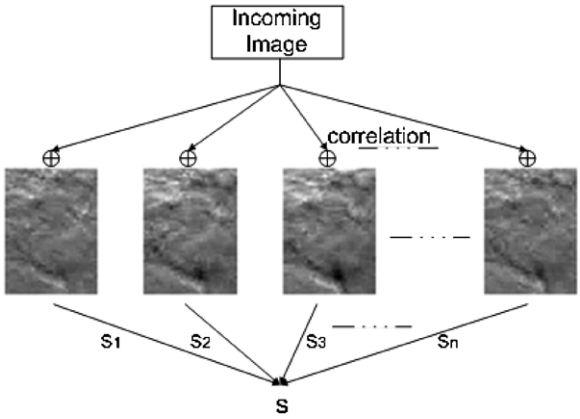

Figure 6. Multiple template method. Here, each of the end-of-exhale images is a template. We match the incoming image with them and get a set of correlation scores $s_1, s_2, \ldots, s_n$. Then we average these scores to generate the final correlation score s for gating.

We can use different weighting functions for combining the scores for EOE frames. One commonly used weighting function is the pth-order weights: $w_i = s_i^p / \sum_i s_i^p$, $s = \sum_i w_i s_i$. This means, we are giving higher weights to those correlation scores with a larger value. It is natural and reasonable to give higher weights to more similar templates. We apply p values of 1 and 2 in our experiments.

2.2.3. Template clustering method. Although we can use the entire set of motion-enhanced images in the EOE gating window as reference templates, this approach has a main drawback of heavy computational burden dealing with redundant information. Assume that we use 35% duty cycle, a typical training session including three breathing cycles will have about 30 such reference templates. Many of them are very similar, and they are not efficient in filtering out the noise. Therefore, we want to find a way somewhere in between single template method and multiple template method, which hopefully has the merits of both methods, i.e., computational efficiency and robustness. Here, instead of using all the 30 templates, we would like to find a set of representative EOE templates. Ideally, these templates should carry all of the useful information of the original frames and discard the noise. Clustering methods are ideally suited to this task. Clustering algorithms group similar objects together and summarize each group with a representative template. We will use clustering to find a small set of templates and use the averaging method to combine the scores. We need to determine the number of templates and a method to find them.

Cluster the end-of-exhale templates. To cluster the EOE templates, we apply Gaussian mixture model clustering (McLachlan and Basford 1988). We denote the parameters of this model by Θ. In this model, we assume that each cluster comes from a multivariate Gaussian distribution, and our data (the image templates) come from a finite mixture of Gaussians. A Gaussian model assumes that there is a cluster template and the members of that cluster are variations of a template. Let X denote our dataset which has d dimensions, and n denote the number of data points in X, we are trying to group the n data points into K clusters. Each of the K cluster is represented by its model parameters, $\theta_j = (\pi_j, \mu_j, \Sigma_j)$, where π_j is the prior probability (or mixture proportion) of cluster j, μ_j is the mean of cluster j and Σ_j is the

covariance matrix of cluster j. More formally, we say that the probability of the data given the model is

$$P(X|\Theta) = \sum_{j=1}^{K} \pi_j P(X|\theta_j),$$
$$P(X|\theta_j) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_j|}} \exp\left\{-\frac{1}{2}(X-\mu_j)^T \Sigma^{-1}(X-\mu_j)\right\}. \tag{2}$$

To estimate the parameters, π_j, μ_j and Σ_j for each cluster, we apply the expectation–maximization algorithm (EM) (Dempster *et al* 1977). The EM algorithm alternates through an expectation-step and a maximization-step until convergence. In the expectation-step, we estimate the cluster to which each image template belongs to given that the parameters (π_j, μ_j and Σ_j) are fixed. In the maximization-step, we estimate the parameters by maximizing the complete log-likelihood (the log-likelihood assuming we know the cluster memberships). The cluster means μ_j then become our representative templates.

Before we can cluster our data, we need to deal with two issues: (1) how can we deal with high dimensionality? (2) How can we automatically find the number of clusters?

Dealing with high dimensionality. The ROI needed to build the EOE templates is relatively large, because it needs to contain the tumour motion and sometimes the tumour is large. For example, a typical EOE template has an image size of 100 × 100 pixels, with a pixel size of 0.5 mm or so. This leads to a dimensionality of size 10 000. Our problem is that our training set only contains around 30 template data points. Our clustering algorithm requires us to estimate $\mathrm{d}K + (d(d+1)/2)K + (K-1)$ parameters. We need to reduce our dimensionality, so that we do not overgeneralize. For this reason, we project the data down to two dimensions before performing clustering.

Principal component analysis (PCA) is a popular technique for dimensionality reduction (Jolliffe 2002). PCA finds a linear transformation, $Y = A^T X$, that projects the original high-dimensional data X with d dimensions to a lower dimensional data Y with q dimensions where $q < d$, such that the mean squared error between X and Y is as small as possible. X here is $d \times N$, where N is the number of data points, and A is a $d \times q$ matrix. The solution is the transformation matrix A whose columns correspond to the q eigenvectors with the q largest eigenvalues of the data covariance. It also projects the high-dimensional dataset to the lower dimensional subspace in which the original dataset has the largest variance (i.e., restricts attention to those directions along which the scatter of the data points is greatest).

Automatically finding the number of clusters. To automatically determine the number of clusters, we apply the Bayesian information criterion (BIC) (Schwarz 1978) score to penalize the log-likelihood function. We now maximize

$$\log P(X|\hat{\Theta}_{ML}) - (f/2)\log(n), \tag{3}$$

where f is the number of free parameters in the model. In our problem, we have

$$f = \mathrm{d}K + (d(d+1)/2)K + (K-1) \tag{4}$$

to be estimated. We run the EM algorithm from $K = 1$ to $K_{\max}$ ($K_{\max} = 4$ in our experiments). Then pick the K with the largest BIC score. The plot of BIC scores versus K is shown in figure 7. Note that if we do not add a penalty term, the log-likelihood increases as K increases. It can lead to a trivial result of picking K equal to the number of data samples (i.e., each data point will be considered as a cluster). A scatter plot of the clusters with their means and covariances is shown in figure 8.

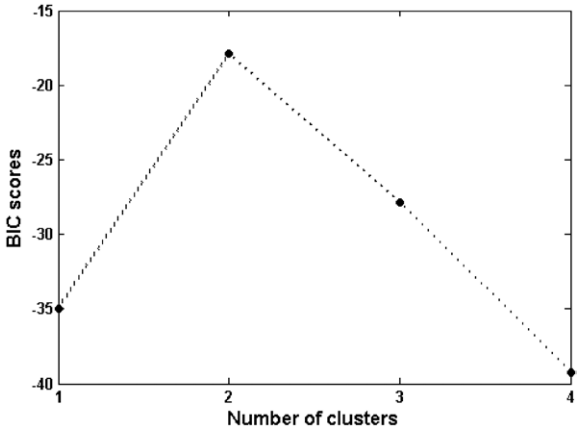

Figure 7. BIC scores versus number of clusters for patient 3 and 35% duty cycle. When the number of clusters is equal to 2, the BIC score reaches its peak, meaning 2 clusters lie in our dataset.

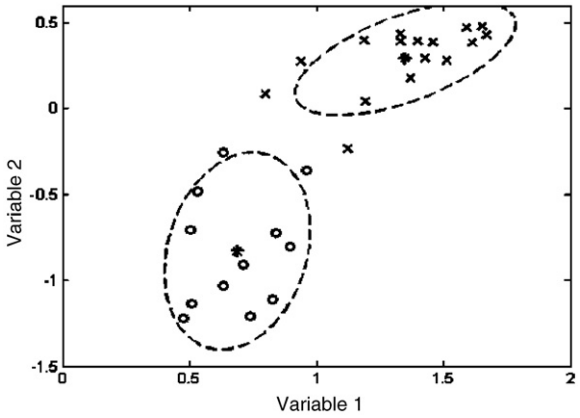

Figure 8. Scatter plot of our image data for patient 4 and 35% duty cycle in 2D with clustering result. The 'o' and 'x' each represents a cluster, with means represented by the '$*$' symbol in bold and the covariances in ellipses.

3. Results

Figure 9 shows the comparison among the three methods for an example dataset. In each subfigure, the upper curve presents the final correlation score and the lower plot shows the gating signal generated by the corresponding score. We observe that the correlation score of the single template method is noisy and leads to erratic gating signals. Compared with the single template method, the gating signal generated with the multiple template method is more stable and robust. The reason is that, in the multiple template method, we decide whether the current frame should be in the gating window by a voting process with correlation scores from all the templates. Through the voting process, the effect of errors caused by one template could be compensated by the other templates. Thus the multiple template method is less sensitive to noise, as long as the training data are representative of the incoming images. Finally, we note

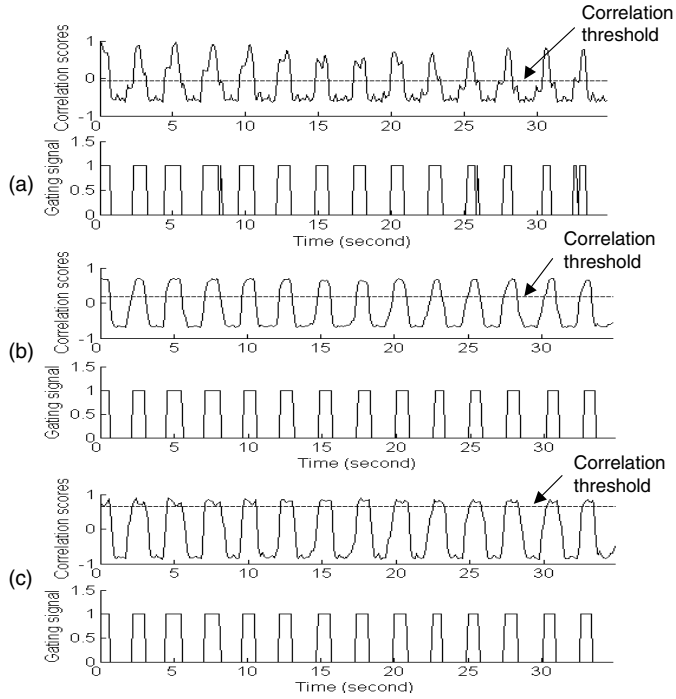

Figure 9. Results from different methods for an example patient: (a) single template method; (b) multiple templates method; (c) clustering method. For each figure, the top curve is the correlation score and the bottom plot is the gating signal generated by the correlation score. Here we use 35% duty cycle.

that the clustering method can achieve the same good performance as the multiple template method at a faster speed. Experiments on five patients show that the clustering method can also improve the accuracy.

We evaluate our methods with five sets of patient data to investigate if they can effectively generate correct gating signals. For each patient, we acquired fluoroscopy at 10 frames s^{-1} for a duration of 30 s, which amounts to around 300 frames. The reference tumour locations are determined by a radiation oncologist who retrospectively marked a tumour contour in the first image frame and shifted the contour in the following image frames to track the tumour. Then, we compare gating signals generated from each of the three methods against that generated from the reference tumour locations. For each case, we generated the gating signals for both 35% and 50% duty cycles. A training phase of 2–3 periods (i.e., 8–10 s) is used.

We report the false positive (FP) rate, the false negative (FN) rate and the total error rate. Specifically, false positive means that the beam is turned on when the tumour is not at the right position, while false negative means that the beam is turned off when the tumour is at the right position. Then, we divide the number of image frames with false positives/false negatives by the number of frames with reference on/off signals to get the rates. The total error rate is defined as the ratio between the number of incorrect frames and the number of total testing frames. Results for the three methods are summarized in tables 1–3.

Table 1 shows FP, FN and total error rates for each of the five patients with 35% duty cycle. Table 2 shows the same metrics with 50% duty cycle. The average FP, FN and total

Table 1. False positive (FP) rate, false negative (FN) rate and total error rates for all the patients using 35% duty cycle. The optimal number of clusters is also given for each patient.

Patient 1	Single template	Multiple template	Clustering ($K_{opt} = 2$)
FP rate (%)	9.6	4.8	4.8
FN rate (%)	7.5	8.5	8.4
Total error rate (%)	8.3	7.1	7.1
Patient 2	Single template	Multiple template	Clustering ($K_{opt} = 3$)
FP rate (%)	0	1.2	0
FN rate (%)	21.7	10.3	6.9
Total error rate (%)	15.3	7.6	4.9
Patient 3	Single template	Multiple template	Clustering ($K_{opt} = 2$)
FP rate (%)	28.3	25.0	3.3
FN rate (%)	0	0	0.7
Total error rate (%)	10.1	8.9	1.8
Patient 4	Single template	Multiple template	Clustering ($K_{opt} = 3$)
FP rate (%)	15.2	15.2	7.6
FN rate (%)	8.3	8.3	9.7
Total error rate (%)	10.5	10.5	9.0
Patient 5	Single template	Multiple template	Clustering ($K_{opt} = 2$)
FP rate (%)	2.3	1.2	1.2
FN rate (%)	13.1	14.0	9.3
Total error rate (%)	8.3	8.3	5.7

Table 2. False positive (FP) rate, false negative (FN) rate and total error rates for all the patients using 50% duty cycle. The optimal number of clusters are also given for each patient.

Patient 1	Single template	Multiple template	Clustering ($K_{opt} = 2$)
FP rate (%)	4.0	4.0	3.4
FN rate (%)	4.3	4.3	4.9
Total error rate (%)	4.2	4.2	4.2
Patient 2	Single template	Multiple template	Clustering ($K_{opt} = 3$)
FP rate (%)	4.0	0	0
FN rate (%)	7.9	6.1	6.1
Total error rate (%)	6.3	3.5	3.5
Patient 3	Single template	Multiple template	Clustering ($K_{opt} = 2$)
FP rate (%)	19.0	15.5	6.0
FN rate (%)	0	1.2	3.5
Total error rate (%)	9.5	8.3	4.7
Patient 4	Single template	Multiple template	Clustering ($K_{opt} = 3$)
FP rate (%)	15.7	11.2	8.9
FN rate (%)	6.6	9.9	9.1
Total error rate (%)	10.5	10.5	9.0
Patient 5	Single template	Multiple template	Clustering ($K_{opt} = 2$)
FP rate (%)	0	0	0
FN rate (%)	18.8	18.6	10
Total error rate (%)	7.8	7.8	4.1

error rates with their standard deviations for each of our three methods are summarized in table 3. All these tables demonstrate that the clustering approach leads to the best results. Let

Table 3. Averaged false positive (FP) rate, false negative (FN) rate and total error rate and their standard deviations for each method.

Methods		35% duty cycle			50% duty cycle		
		Single	Multiple	Clustering	Single	Multiple	Clustering
FP rate (%)	Avg	11.1	9.5	3.4	8.6	6.1	3.7
	Std	11.4	10.4	3.0	8.3	6.9	3.9
FN rate (%)	Avg	10.1	8.2	7.1	7.5	8.0	6.7
	Std	8.0	5.1	3.6	7.0	6.8	2.8
Total error rate (%)	Avg	10.5	8.5	5.7	7.6	6.8	5.1
	Std	2.9	1.3	2.7	2.5	2.9	2.2

accuracy be equal to100% minus total error rate, we can see that the accuracy of the clustering method falls within the range of 91–98.2% for the 35% duty cycle and ranges between 91% and 96.5% for the 50% duty cycle. The clustering approach is also the best in terms of stability for all five patients. Low false positive rate is especially important because erratic beam-on signals will cause the extra irradiation of the surrounding healthy tissues and critical structures. We can see that the FP rates of the clustering approach are consistently below 10% for all five patients. In addition, bar plots of the three measures for our three methods are given in figure 10. In each figure, top/bottom edge of the bar represents the maximum/minimum value while the middle line represents the mean. It shows that the clustering method performs much better than the other two and its three measures (FP, FN, and error rates) are all below 10%.

4. Discussion and conclusions

In this study of correlation score-based gating, we investigate methods to improve the robustness and accuracy of respiratory gating without implanted fiducial markers for lung tumour treatment. By combining correlation scores generated by multiple templates, the multiple template and template clustering methods are insensitive to the noise caused by cardiac motion and unstable respiration. After reducing the dimension, we only keep the most distinct features in the image sequence, so that the robustness and accuracy of generating gating signals are further improved using the template clustering method. Among all three methods, the clustering approach is the best in terms of accuracy for all five patients. The FP rates, FN rates and total error rates of the clustering approach are consistently below 10% for all five patients. In addition, our experiments show that the gating signal that results from the clustering approach is smoother and less erratic (i.e., more robust and stable) compared to the single template approach.

Although the gating error can be up to 10% for some patients for the template clustering method, we found that all the errors happen near the edge of the gating window. Figure 11 shows an example of the location of the gating errors. For other patients, the situation is similar. That is to say, gating errors only happen when the tumour is very close to the correct gating position, and the beam will never turn on when the tumour is far away from the correct position. To further illustrate this point, we computed the gating signals for 32% duty cycle and compared with the reference gating signals of 35% duty cycle. We found that the false positive rate becomes 0% for four patients and 3% for one patient. The clinical implication of this observation is that, we can apply for a slightly larger margin for the residue motion to reduce the dosimetric error to a negligible level.

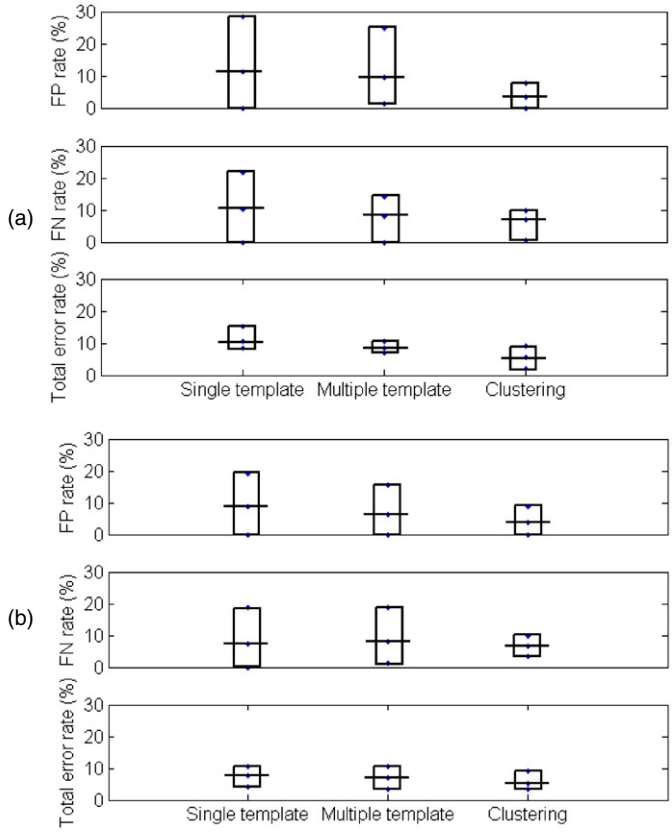

Figure 10. Calculated error rates at (a) 35% duty cycle and (b) 50% duty cycle based on the first reference. For each figure, top: FP rate; middle: FN rate; bottom: total error rate. For each bar, top is the maximum value, bottom is the minimum value and middle line represents the mean.

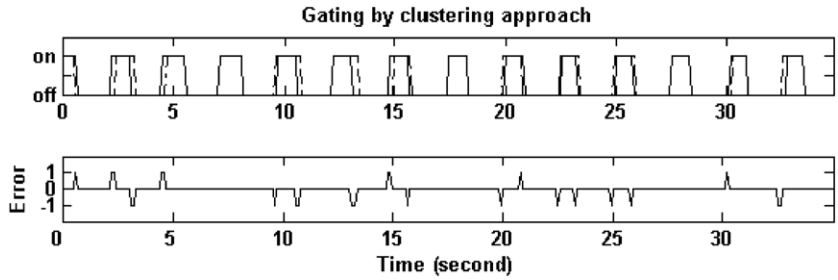

Figure 11. An example showing the location of errors for 35% duty cycle for patient 4. Upper plot: gating signals generated by template clustering method (solid line) and the reference gating signals (dashed line). Lower plot: gating errors. Positive value means false positive and negative value means false negative.

We also estimated the intra-observer error. The expert observer has marked the tumour position twice for most of the images. We found that the difference between two marked

tumour positions is within two pixels on average. Using the second set of manually marked data as the reference, we re-calculated the error rates for all the patients and for two duty cycles. We found that no matter which set of reference data is used, the corresponding error rates are within 2% and the performance of each method is the same.

A major requirement for clinical implementation is the ability to compute the correlation scores and to generate gating signals in real-time. In this work, all the computation has been done using MatLab (The MathWorks, Inc., Natick, MA, USA). However, we do not expect any problems when applying the methods to clinical practice. In our methods, all the templates used are acquired during the patient setup session. When a new frame comes in during the treatment delivery, we only need to match it with the pre-acquired templates, process the correlation scores, and then decide whether to turn the beam on or off. The template matching and correlation score processing are quite efficient in practice. For an image of size $m \times n$, the time we need to get the gating signal at a certain time point scales as O (mn).

Although not seen in the current datasets, aperiodic intra-fraction motion does happen for some patients. However, this should not be a problem for the proposed gating methods. The beam is only turned on when the correlation score is high, meaning the ROI of incoming image is similar to the reference template and the tumour is at the right position. If tumour drifts away or if shifts/rotations/deformations of the anatomy between patient setup and treatment happen, the correlation score should be always low and beam will not be turned on. We then need to re-position the patient.

The image datasets used in this work are limited to one fraction. In the future we should investigate the impact of the inter-fraction variability on the proposed technique, although the impact is likely minor since the reference templates are built from the same fraction. In addition, we will apply our methods to inhale. Treatment in the inhale phase would prove useful for heart/lung sparing in sites like breast or Hodgkins lymphoma.

For our next step, we will (1) test the algorithms using more and longer patient data, (2) find a better way to get reference gating signal for validation, and (3) evaluate the dosimetric consequence of current error level to see if there is a need to further lower the error rates. Then, we will consider clinical implementation of our methods.

Acknowledgments

The project is partially supported by an NCI grant (1 R21 CA110177 A 01A1) and by NSF-career IIS-0347532, as well as by Varian Medical Systems, Inc.

References

Arslan S, Yilmaz A, Bayramgurler B, Uzman O, Nver E and Akkaya E 2002 CT-guided transthoracic fine needle aspiration of pulmonary lesions: accuracy and complications in 294 patients *Med. Sci. Monit.* **8** CR493–7

Berbeco R I, Jiang S B, Sharp G C, Chen G T, Mostafavi H and Shirato H 2004 Integrated radiotherapy imaging system (IRIS): design considerations of tumour tracking with linac gantry-mounted diagnostic x-ray systems with flat-panel detectors *Phys. Med. Biol.* **49** 243–55

Berbeco R I, Mostafavi H, Sharp G C and Jiang S B 2005 Towards fluoroscopic respiratory gating for lung tumours without radiopaque markers *Phys. Med. Biol.* **50** 4481–90

Chen Q S, Weinhous M S, Deibel F C, Ciezki J P and Macklis R M 2001 Fluoroscopic study of tumor motion due to breathing: facilitating precise radiation therapy for lung cancer patients *Med. Phys.* **28** 1850–6

Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc. B* **39** 1–38

Geraghty P R, Kee S T, McFarlane G, Razavi M K, Sze D Y and Dake M D 2003 CT-guided transthoracic needle aspiration biopsy of pulmonary nodules: needle size and pneumothorax rate *Radiology* **229** 475–81

Gierga D P, Brewer J, Sharp G C, Betke M, Willett C G and Chen G T 2005 The correlation between internal and external markers for abdominal tumors: implications for respiratory gating *Int. J. Radiat. Oncol. Biol. Phys.* **61** 1551–8

Harada T, Shirato H, Ogura S, Oizumi S, Yamazaki K, Shimizu S, Onimaru R, Miyasaka K, Nishimura M and Dosaka-Akita H 2002 Real-time tumor-tracking radiation therapy for lung carcinoma by the aid of insertion of a gold marker using bronchofiberscopy *Cancer* **95** 1720–7

Jain R, Kasturi R and Schunck B G 1995 *Machine Vision* (New York: McGraw-Hill)

Jiang S B 2006 Radiotherapy of mobile tumors *Semin. Radiat. Oncol.* **16** 239–48

Jolliffe I T 2002 *Principal Component Analysis* (Berlin: Springer)

Laurent F, Latrabe V, Vergier B and Michel P 2000 Percutaneous CT-guided biopsy of the lung: comparison between aspiration and automated cutting needles using a coaxial technique *Cardiovasc. Intervent. Radiol.* **23** 266–72

McLachlan G J and Basford K E 1988 *Mixture Models, Inference and Applications to Clustering* (New York: Marcel Dekker)

Schwarz G 1978 Estimating the dimension of a model *Ann. Stat.* **6** 461–4

Seppenwoolde Y, Shirato H, Kitamura K, Shimizu S, van Herk M, Lebesque J V and Miyasaka K 2002 Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **53** 822–34

Shirato H *et al* 2003 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.* **56** 240–7

Tang X, Sharp G and Jiang S 2006 Patient setup based on lung tumor mass for gated radiotherapy *Med. Phys.* **33** 2244 (abstract)

Whittaker E T and Robinson G 1967 *The Calculus of Observations: A Treatise on Numerical Mathematics* (New York: Dover)