INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. 49 (2004) 5347–5356 PII: S0031-9155(04)85063-2

Tracking errors in a prototype real-time tumour tracking system

Gregory C Sharp[1], Steve B Jiang[1], Shinichi Shimizu[2] and Hiroki Shirato[2]

[1] Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[2] Department of Radiation Medicine, Hokkaido University School of Medicine, Sapporo, Japan

E-mail: gcsharp@partners.org

Received 16 August 2004, in final form 18 October 2004
Published 19 November 2004
Online at stacks.iop.org/PMB/49/5347
doi:10.1088/0031-9155/49/23/011

Abstract

In motion-compensated radiation therapy, radio-opaque markers can be implanted in or near a tumour and tracked in real-time using fluoroscopic imaging. Tracking these implanted markers gives highly accurate position information, except when tracking fails due to poor or ambiguous imaging conditions. This study investigates methods for automatic detection of tracking errors, and assesses the frequency and impact of tracking errors on treatments using the prototype real-time tumour tracking system. We investigated four indicators for automatic detection of tracking errors, and found that the distance between corresponding rays was most effective. We also found that tracking errors cause a loss of gating efficiency of between 7.6 and 10.2%. The incidence of treatment beam delivery during tracking errors was estimated at between 0.8% and 1.25%.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Modern radiation therapy is capable of highly conformal treatments that deliver a lethal dose to a tumour while sparing surrounding healthy tissue. When dose conformality is reduced because of tumour motion during treatment, motion compensation methods can be used to reduce motion and restore conformality. One promising technique for motion compensation is respiratory gated treatment, which reduces tumour motion during beam-on time by limiting radiation exposure to a portion of the breathing cycle (Mageras and Yorke 2004). This can be implemented using an external respiratory signal to infer the tumour position (Ohara *et al* 1989, Kubo and Hill 1996, Ramsey *et al* 1999a, 1999b, Kubo *et al* 2000, Minohara *et al* 2000, Vedam *et al* 2001, Keall *et al* 2002), by fluoroscopic tracking of

0031-9155/04/235347+10$30.00

implanted radio-opaque markers (Shirato *et al* 1999, Shimizu *et al* 2000, 2001, Shirato *et al* 2000a, 2000b, Harada *et al* 2002) or by magnetic tracking of implanted markers (Seiler *et al* 2000, Hummel *et al* 2002). Another motion compensation technique is beam tracking, or following the target dynamically with the radiation beam (Murphy 2004). This was first implemented in a robotic radiosurgery system (Adler *et al* 1999, Ozhasoglu *et al* 2000, Schweikard *et al* 2000, Murphy 2002, Murphy *et al* 2003), and has been proposed for use with a dynamic multileaf collimator (MLC) (Keall *et al* 2001, Neicu *et al* 2003, Suh *et al* 2004).

Both gating and beam tracking require knowing the exact 3D position of the moving tumour in real time. This knowledge can be obtained through real-time x-ray imaging of fiducial markers during the treatment (Shirato *et al* 1999, 2000b, 2003, Shimizu *et al* 2001, Murphy 2002, Berbeco *et al* 2004). A radio-opaque seed is surgically implanted in or near the tumour, and the position of the seed is tracked throughout the treatment session. Tracking is performed using specialized pattern recognition software that runs in real time. While most pattern recognition methods are quite capable of isolating a high contrast target from the background, they can fail when the signal quality is poor, or when background clutter mimics the target. Therefore, a clinical tracking system must anticipate the possibility that tracking will fail, and provide tools for the detection of and response to tracking failures.

The prototype real-time tumour-tracking radiation therapy (RTRT) system at Hokkaido University implements an on-line tracking failure detection algorithm that compares tracking statistics against four user-controlled thresholds in real time. Two classes of errors are detected: *recoverable* errors and *unrecoverable* errors. A recoverable error is a temporary or intermittent error for which automatic recovery is expected. An unrecoverable error refers to an error for which tracking must be manually reinitialized. The purpose of having these two error levels is to improve treatment efficiency while maintaining treatment safety. When a recoverable error occurs, the treatment is temporarily halted by the linac beam hold-off mechanism. Treatment resumes automatically when the recoverable error condition is cleared. In contrast, when an unrecoverable error occurs, a tracking interlock is enabled and the treatment beam permanently halted. The therapist must correct the tracking failure and clear the interlock to resume treatment.

This work is an analysis of how tracking errors affect the efficiency and reliability of treatment using real-time tumour tracking. We have analysed the log files captured by the RTRT system during treatment, estimated the frequency of tracking errors, and quantified the loss of efficiency. In addition, we have examined the frequency of treatment beam delivery during tracking errors. Finally, we examine the effectiveness the tracking error detection system.

2. Methods and materials

In this section, we describe and analyse methods for tracking error detection. First, we introduce a set of indicators that can be used to detect errors. The relative merit of each indicator is evaluated against a set of examples from the RTRT log files. Next, we combine these indicators using a neural network classifier, and identify tracking errors over a large set of tracking data. This classification is used to estimate how often tracking errors occur during treatment, and their effect on treatment efficiency and accuracy. Finally, the detected tracking errors are post-processed to distinguish recoverable errors, lasting less than 1.5 s, from unrecoverable errors, lasting longer than 1.5 s. The detection rate for unrecoverable errors is thereby estimated.

2.1. *Indicators for tracking errors*

Several indicators can be used to detect the occurrence of a tracking error. We will evaluate indicators based on the quality of the pattern recognition, the quality of the triangulation and the stability of the recovered signal.

The first indicator we consider is the *pattern recognition score*, or PRS. The PRS is a measure of the certainty with which the tracking software could find the target within a single image. In the RTRT system, this score is computed using normalized cross-correlation (Shirato *et al* 2000b), but other systems may use other metrics. This score is, in essence, a measure of the quality of the match, and the pixel with the highest PRS is assumed to be the location of the target. When doing stereo tracking, the correlation score is computed for both images, but we need only consider the image with the lower score. This simplification is justified because 3D accuracy is generally limited by the imager which is tracking more poorly.

The second indicator we consider is the *distance between rays*, or DBR. The DBR is a measure of consistency of tracking between the two imagers. Once the image location of the marker has been determined, the marker is known to lie on a corresponding ray in 3D space. Therefore, when the location is known in both images, the three-dimensional position of the marker is reconstructed by intersecting two corresponding 3D rays. Usually the two rays do not intersect exactly; instead there is a gap. The DBR refers to this gap between the rays, and it is small when the marker is being tracked properly in both images, but may become large when one or both imagers have identified the wrong location for the marker. Thus, a large DBR is an indication of poor tracking.

The last indicators we consider are the *instantaneous velocity*, or IV, and the *instantaneous acceleration*, or IA. The IV and IA are measures of tracking stability over a short period of time. Poor localization of the target within an image often causes sudden jumps in the computed target location whereas we expect the motion of the markers to vary smoothly during good tracking. For off-line processing, we compute the velocity v_t and acceleration a_t at time t using positions x_{t-1} through x_{t+1}. Neglecting terms due to the sampling rate, this is given as

$$v_t = \frac{|x_{t+1} - x_t| + |x_t - x_{t-1}|}{2} \tag{1}$$

$$a_t = |x_{t+1} - 2x_t + x_{t-1}|. \tag{2}$$

Or, during on-line processing, the velocity and acceleration can be computed as

$$v_t = |x_t - x_{t-1}| \tag{3}$$

$$a_t = |x_t - 2x_{t-1} + x_{t-2}|. \tag{4}$$

An example of these indicators are shown in figure 1. This figure compares the motion of a marker in the left–right (L–R), superior–inferior (S–I) and anterior–posterior (A–P) directions, against the four indicators described above. At an approximate time $t = 92$ s, the pattern matching software loses track of the implanted marker. As a result the DBR increases, the PRS falls and the IV and IA values increase. Note that IV and IA do not reflect the gross motion of the tumour. Because they are computed over a small time interval from unfiltered measurements, they instead reflect the stability of the measurement.

The Mitsubishi tracking system implements recoverable and unrecoverable tracking error detection by comparing four configuration parameters against the DBR and PRS signals. These parameters are the DBR threshold, hard PRS threshold, soft PRS threshold and the maximum allowed soft PRS errors. At each image frame, if the PRS is less than the hard PRS

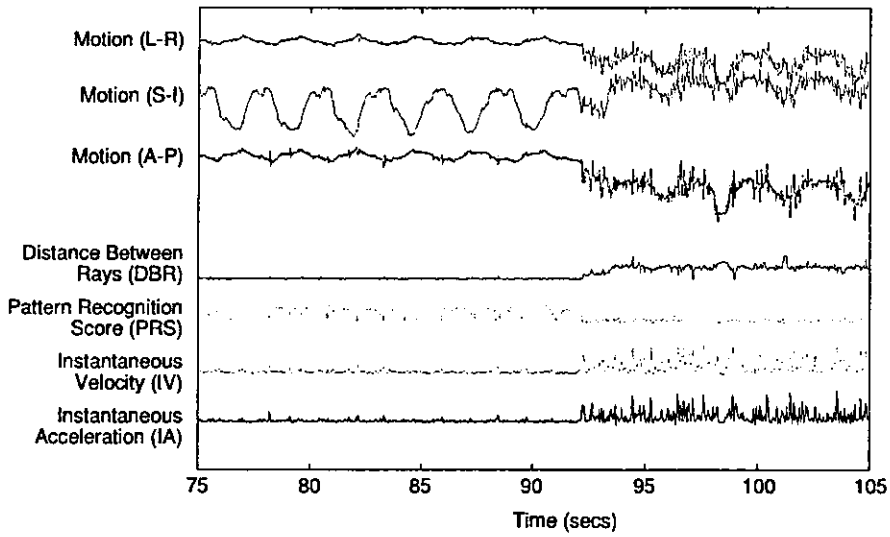

Figure 1. Four indicators that can be useful in detecting tracking errors.

threshold, or if the DBR is less than the DBR threshold, an unrecoverable error is immediately signalled. If the PRS is less than the soft PRS threshold, then a recoverable error is immediately signalled. The recoverable error signal is subsequently cleared when the PRS exceeds the soft PRS threshold. However, if the PRS remains less than the soft PRS threshold for a continuous duration greater than the maximum allowed soft PRS errors, a hard error is signalled and treatment is interrupted.

2.2. Off-line error detection

To evaluate the performance of the tracking system, and to understand the effects of tracking errors on the clinical treatment, we first need to decide what constitutes a tracking error. Our approach has four stages:

(i) identify examples of good and bad tracking to be used as training data,
(ii) train a neural network classifier using the training data,
(iii) run the neural network classifier at different detection rates to generate a set of classified traces,
(iv) identify unrecoverable error regions from the classified traces.

The first stage identifies regions within the file that have or do not have tracking errors, and is necessary for training the classifier. The regions are obtained from the tracking log files using the tracking interlock information. The tracking interlock is enabled through automatic detection by the RTRT software, or through manual detection by the therapist. Therefore, we conclude that the interval immediately preceding the interlock usually contains a tracking error. Similarly, tracking is corrected by the therapist before the interlock is cleared, and therefore we conclude that the interval immediately following the cleared interlock is usually error free. As shown in figure 2, 1 s of bad tracking is taken before each interlock enable, and 4 s of good tracking is taken after each interlock clear.

Given this set of training data, we may evaluate the effectiveness of the indicators described in section 2.1 on identifying tracking errors. A threshold is used on each indicator to classify a data point as either tracking error or good tracking. Changing the detection threshold on each indicator increases or decreases the detection rate and false alarm rate. In addition, we evaluate

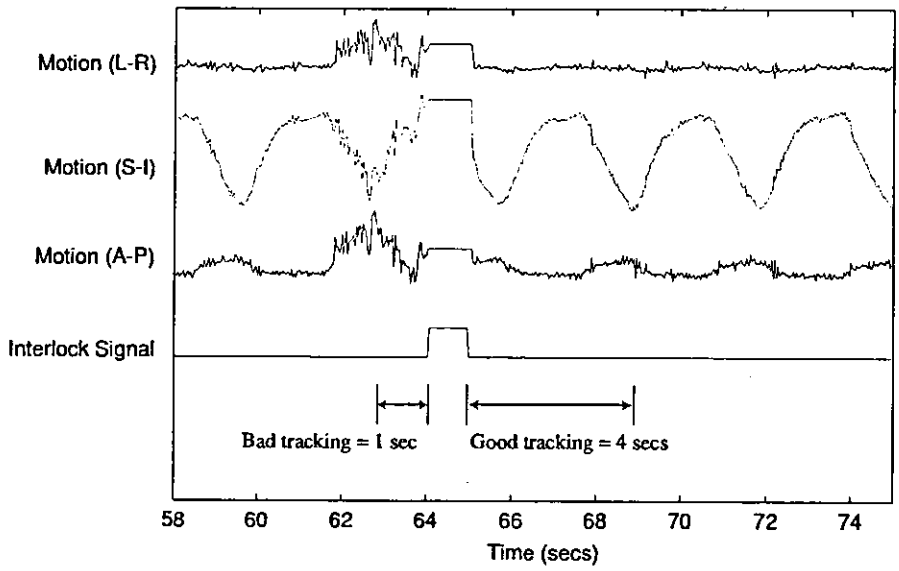

Figure 2. An interval of 1 s previous to each tracking interlock is assumed to be a tracking error, and an interval of 4 s after the interlock is cleared is assumed to be error free.

the effectiveness of using all four indicators together by building an artificial neural network classifier (Bishop 1995). The neural network uses a fully connected multilayer perceptron architecture, with four input units five hidden units and one output unit. The network is trained to produce an output of +1 when the inputs were associated with a tracking error, and produce an output of −1 when the inputs were associated with good tracking. Training was done off-line using a conjugate gradient optimizer. The relative performance of this neural network is evaluated using 10-fold cross-validation (Duda *et al* 2001). A tradeoff between detection rate and false alarm is achieved by changing the threshold on the neural network output.

The results of this comparison are shown in figure 3. It is seen that DBR is the most powerful indicator for detecting tracking errors. PRS, IV and IA are somewhat weaker cues, but detection efficiency can be boosted somewhat by using them in conjunction with DBR.

The next stage is to use the neural network classifier to identify tracking errors across the entire set of patient log files. Depending on the setting of the detection threshold, the classifier becomes more or less sensitive. In figure 4 (left), we see that by setting the threshold to a false alarm rate of 5% (FAR = 0.05), a tracking error beginning at around time $t = 43$ s is detected accurately with relatively few false alarms, while setting the threshold to a false alarm rate of 10% (FAR = 0.10) causes significant regions to be mislabelled as errors. In contrast, consider figure 4 (right), which shows the motion trace of a patient with a small tumour motion, and sudden jumps in motion caused by loss of track. We see that at FAR = 0.05 these errors are not detected, but at FAR = 0.10 these errors are correctly identified. These examples illustrate the difficulty in setting the detection sensitivity. Therefore, in the remaining analysis we have chosen to make two estimates of tracking errors, one at FAR = 0.05 and one at FAR = 0.10, with the understanding that the true classification is probably within these bounds.

The final stage is the identification of unrecoverable errors. We define an unrecoverable error as a tracking error that lasts longer than 1.5 s. The duration of 1.5 s is based on an empirical observation that tracking errors lasting longer than this duration generally do not correct themselves. An example of the identification of unrecoverable errors is shown in figure 5. In this figure, the tracking interlock signal was triggered three times during treatment,

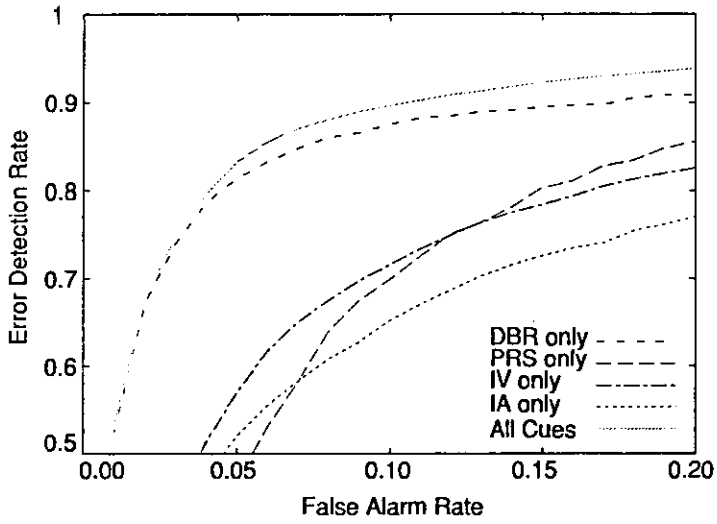

Figure 3. Estimated detection rates for finding tracking errors with different inputs.

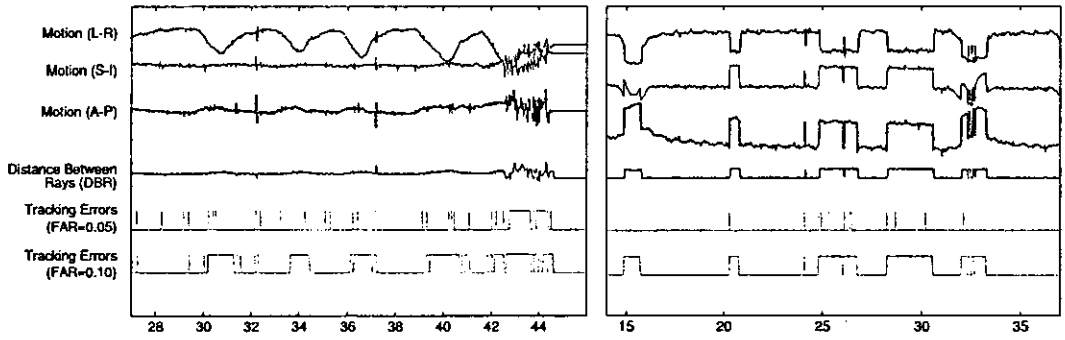

Figure 4. Detecting errors with a false alarm rate (FAR) of 0.05 and 0.10.

and preceding each interlock is relatively long duration of tracking errors. Note that the unrecoverable error at the beginning of the trace is the time required to initiate tracking, and is treated separately in our analysis. Note further that the irregular breathing motion between 75 and 95 s is the real tumour motion, not a tracking error.

2.3. On-line error detection

On-line detection of unrecoverable tracking errors is needed to stop treatment until the therapist has corrected the problem. In this section, we will evaluate how well the prototype RTRT tracking system identifies unrecoverable errors. We define a correct detection as a detection that is made between 1 s prior to the beginning of the error, and 1 s after the beginning of the error. Errors that are not detected within this interval will be considered detection failures. Detections that occur during good tracking, and more than 1 s before the start of an unrecoverable error will be considered to be a false alarm.

The detection rate for the prototype RTRT system at identifying unrecoverable errors is shown in figure 6. As noted earlier, we do not have perfect segmentation of the error regions,

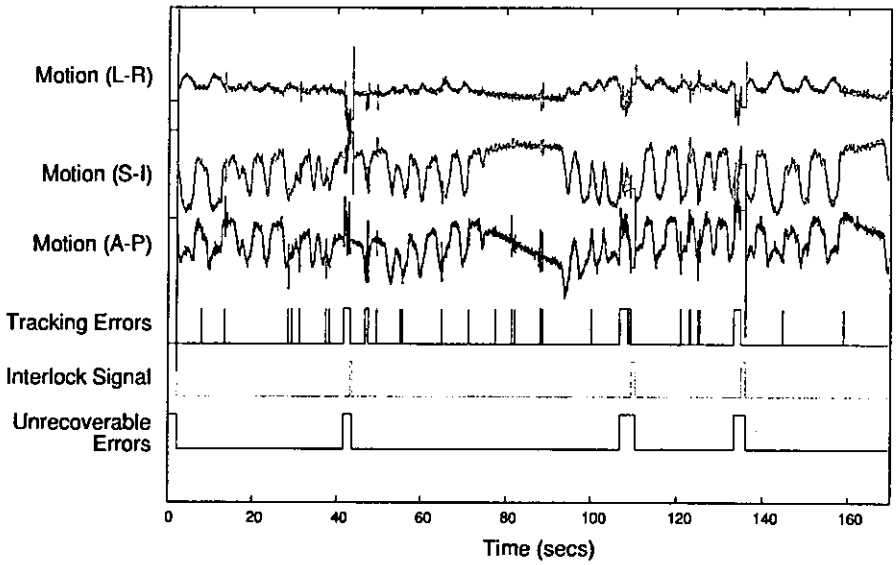

Figure 5. Identification of unrecoverable errors from tracking log files.

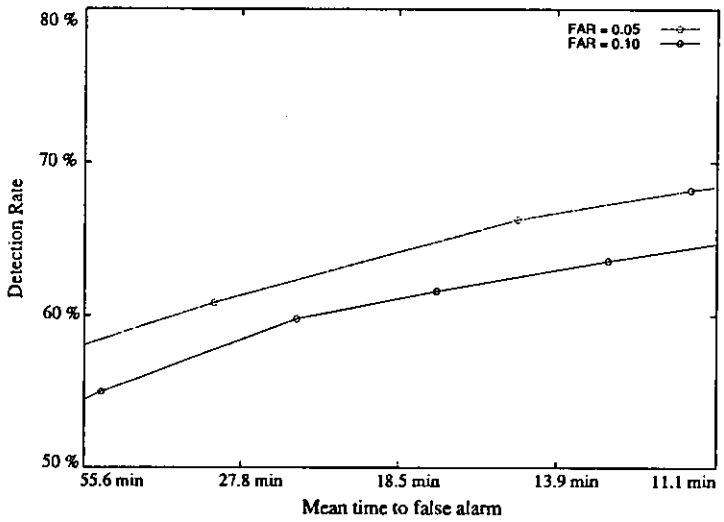

Figure 6. Estimated detection rates for unrecoverable errors on the prototype RTRT system.

and therefore we present detection rate curves for FAR values of 0.05 and 0.10. The actual detection rate is expected to fall between these two curves.

3. Results and discussion

Although it would seem that tracking errors can be identified to some degree, there is clearly a tradeoff between detection rates and false alarms. Nevertheless, from figure 3 it can be seen that DBR is the most important of the indicators. One reason for this is that a high DBR always means that tracking is incorrect, while a low PRS or high IV/IA are simply correlated with tracking errors. Unfortunately, however, a low DBR cannot guarantee good tracking.

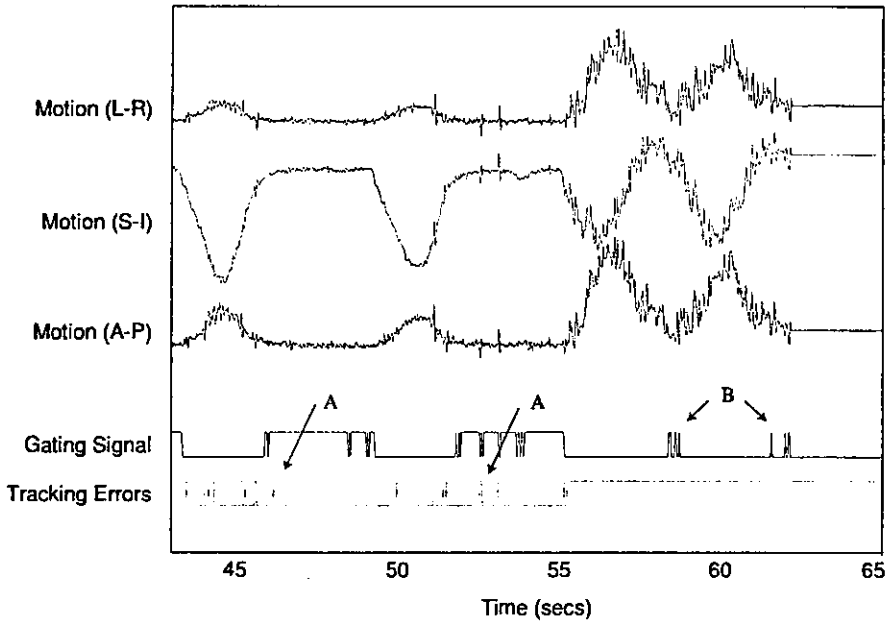

Figure 7. Tracking errors while the MV beam is delivering treatment.

Table 1. Analysis of errors with respect to kV beam-on time.

	FAR = 0.05	FAR = 0.10
Kilovoltage beam-on time	34.0 h	34.0 h
Initial setup time	67.5 min (3.3%)	67.5 min (3.3%)
Hard errors	119.2 min (5.8%)	172.4 min (8.5%)
Error recovery time	36.4 min (1.8%)	36.4 min (1.8%)
Loss of efficiency due to errors	2.59 h (7.6%)	3.48 h (10.2%)
Soft errors	24.0 min (1.2%)	64.0 min (3.1%)

One clinical implication of tracking errors is a loss of efficiency in treatment. We have evaluated this impact using lung tumour motion data from 40 patients treated on the prototype RTRT system between 2000 and 2001 (Shirato *et al* 2000b). This analysis has been summarized in table 1. We define the loss of efficiency due to tracking errors to be the time when there were unrecoverable errors, plus the time needed to correct these errors. Loss of efficiency was estimated to be between 7 and 11% of the total kV beam-on time.

Another clinical implication is the possibility of treatment beam delivery during an undetected tracking error. If treatment persists when the position of the tumour is not known reliably, this will result in a loss of conformality. To estimate this effect, the gating signal was compared with the tracking errors detected in section 2.2. An example is shown in figure 7, where it can be seen that sometimes the gating signal is on, and also our classifier detected a tracking error. We separate these incidents into two groups: 'type A' errors, which indicate that the signal quality is poor but the tumour appears to be within the gating window, and 'type B' errors, which indicate that the tumour position is unknown, and may not be within the gating window. Although it is not possible to know with certainty whether a given incident is type A or type B, we estimated their relative frequency by manual inspection of 100 randomly selected incidents at FAR = 0.05 and FAR = 0.10. If a conclusive classification was not possible, the incident was classified as type B. The results of this analysis are shown in table 2.

Table 2. Analysis of errors with respect to MV beam-on time.

	FAR = 0.05	FAR = 0.10
Megavoltage beam-on time	13.1 h	13.1 h
MV beam-on during errors	9.1 min (1.2%)	35.2 min (4.5%)
Type A MV beam-on	2.8 min (0.36%)	25.4 min (3.21%)
Type B MV beam-on	6.3 min (0.80%)	9.9 min (1.25%)

It is seen that there is approximately a 1% occurrence of MV treatment beam delivery while the position of the tumour is unknown.

4. Summary

Because tracking errors cannot always be avoided, automatic tracking error detection is an essential component of imaging systems that perform real-time motion compensated radiotherapy. Four indicators useful in detecting tracking errors have been evaluated, and the distance between rays (DBR) was found to give the best detection accuracy. The efficiency of error detection for the prototype RTRT system was evaluated, and we found that approximately 60% of unrecoverable tracking errors could be detected within 1 s. The loss of efficiency due to tracking errors was found to be approximately 7 to 10% of kV beam-on time. The effect of tracking errors on treatment beam delivery was analysed, and it was found that the tumour position was unknown during approximately 1% of MV beam delivery.

References

Adler J R Jr, Murphy M J, Chang S D and Hancock S L 1999 Image-guided robotic radiosurgery *Neurosurgery* **44** 1299–306 (discussion 1306–7)

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

Bishop C 1995 *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press)

Duda R O, Hart P E and Stork D G 2001 *Pattern Classification* (New York: Wiley)

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

Hummel J, Figl M, Kollmann C, Bergmann H and Birkfellner W 2002 Evaluation of a miniature electromagnetic position tracker *Med. Phys.* **29** 2205–12

Keall P J, Kini V R, Vedam S S and Mohan R 2001 Motion adaptive x-ray therapy: a feasibility study. *Phys. Med. Biol.* **46** 1–10

Keall P J, Kini V R, Vedam S S and Mohan R 2002 Potential radiotherapy improvements with respiratory gating *Australas. Phys. Eng. Sci. Med.* **25** 1–6

Kubo H D and Hill B C 1996 Respiration gated radiotherapy treatment: a technical study *Phys. Med. Biol.* **41** 83–91

Kubo H D, Len P M, Minohara S and Mostafavi H 2000 Breathing-synchronized radiotherapy program at the University of California Davis Cancer Center *Med. Phys.* **27** 346–53

Mageras G S and Yorke E 2004 Deep inspiration breath hold and respiratory gating strategies for reducing organ motion in radiation treatment *Semin. Radiat. Oncol.* **14** 65–75

Minohara S, Kanai T, Endo M, Noda K and Kanazawa M 2000 Respiratory gated irradiation system for heavy-ion radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **47** 1097–103

Murphy M J 2002 Fiducial-based targeting accuracy for external-beam radiotherapy *Med. Phys.* **29** 334–44

Murphy M J 2004 Tracking moving organs in real time *Semin. Radiat. Oncol.* **14** 91–100

Murphy M J, Chang S D, Gibbs I C, Le Q T, Hai J, Kim D, Martin D P and Adler J R Jr 2003 Patterns of patient movement during frameless image-guided radiosurgery *Int. J. Radiat. Oncol. Biol. Phys.* **55** 1400–8

Neicu T, Shirato H, Seppenwoolde Y and Jiang S B 2003 Synchronized moving aperture radiation therapy (SMART): average tumour trajectory for lung patients *Phys. Med. Biol.* **48** 587–98

Ohara K, Okumura T, Akisada M, Inada T, Mori T, Yokota H and Calaguas M J 1989 Irradiation synchronized with respiration gate *Int. J. Radiat. Oncol. Biol. Phys.* **17** 853–7

Ozhasoglu C, Murphy M J, Glosser G, Bodduluri M, Schweikard A, Forster K M, Martin D P and Adler J R 2000 Real-time tracking of the tumor volume in precision radiotherapy and body radiosurgery—a novel approach to compensate for respiratory motion *Proc. 14th Int. Conf. on Computer Assisted Radiology and Surgery (CARS 2000) (San Francisco, CA)* ed H U Lemke, M W Vannier, K Inamura, A G Farman and K Doi pp 691–6

Ramsey C R, Cordrey I L and Oliver A L 1999a A comparison of beam characteristics for gated and nongated clinical x-ray beams *Med. Phys.* **26** 2086–91

Ramsey C R, Scaperoth D, Arwood D and Oliver A L 1999b Clinical efficacy of respiratory gated conformal radiation therapy *Med. Dosim.* **24** 115–9

Schweikard A, Glosser G, Bodduluri M, Murphy M J and Adler J R 2000 Robotic motion compensation for respiratory movement during radiosurgery *Comput. Aided Surg.* **5** 263–77

Seiler P G, Blattmann H, Kirsch S, Muench R K and Schilling Ch 2000 A novel tracking technique for the continuous precise measurement of tumour positions in conformal radiotherapy *Phys. Med. Biol.* **45** N103–N110

Shimizu S, Shirato H, Kitamura K, Shinohara N, Harabayashi T, Tsukamoto T, Koyanagi T and Miyasaka K 2000 Use of an implanted marker and real-time tracking of the marker for the positioning of prostate and bladder cancers *Int. J. Radiat. Oncol. Biol. Phys.* **48** 1591–7

Shimizu S, Shirato H, Ogura S, Akita-Dosaka H, Kitamura K, Nishioka T, Kagei K, Nishimura M and Miyasaka K 2001 Detection of lung tumor movement in real-time tumor-tracking radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **51** 304–10

Shirato H, Shimizu S, Shimizu T, Nishioka T and Miyasaka K 1999 Real-time tumour-tracking radiotherapy *Lancet* **353** 1331–2

Shirato H *et al* 2000a Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor *Int. J. Radiat. Oncol. Biol. Phys.* **48** 435–42

Shirato H *et al* 2000b Physical aspects of a real-time tumor-tracking system for gated radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **48** 1187–95

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

Suh Y, Yi B, Ahn S, Kim J, Lee S, Shin S, Shin S and Choi E 2004 Aperture maneuver with compelled breath (AMC) for moving tumors: a feasibility study with a moving phantom *Med. Phys.* **31** 760–6

Vedam S S, Keall P J, Kini V R and Mohan R 2001 Determining parameters for respiration-gated radiotherapy *Med. Phys.* **28** 2139–46

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. 49 (2004) 5357–5372 PII: S0031-9155(04)85065-6

A finite state model for respiratory motion analysis in image guided radiation therapy

Huanmei Wu[1], Gregory C Sharp[2], Betty Salzberg[1], David Kaeli[3], Hiroki Shirato[4] and Steve B Jiang[2]

[1] College of Computer and Information Science, Northeastern University, Boston, MA 02115, USA
[2] Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA
[3] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA
[4] Department of Radiation Medicine, Hokkaido University School of Medicine, Sapporo, Japan

E-mail: maggiewu@ccs.neu.edu

Received 16 August 2004, in final form 18 October 2004
Published 19 November 2004
Online at stacks.iop.org/PMB/49/5357
doi:10.1088/0031-9155/49/23/012

Abstract

Effective image guided radiation treatment of a moving tumour requires adequate information on respiratory motion characteristics. For margin expansion, beam tracking and respiratory gating, the tumour motion must be quantified for pretreatment planning and monitored on-line. We propose a finite state model for respiratory motion analysis that captures our natural understanding of breathing stages. In this model, a regular breathing cycle is represented by three line segments, exhale, end-of-exhale and inhale, while abnormal breathing is represented by an irregular breathing state. In addition, we describe an on-line implementation of this model in one dimension. We found this model can accurately characterize a wide variety of patient breathing patterns. This model was used to describe the respiratory motion for 23 patients with peak-to-peak motion greater than 7 mm. The average root mean square error over all patients was less than 1 mm and no patient has an error worse than 1.5 mm. Our model provides a convenient tool to quantify respiratory motion characteristics, such as patterns of frequency changes and amplitude changes, and can be applied to internal or external motion, including internal tumour position, abdominal surface, diaphragm, spirometry and other surrogates.

1. Introduction

While modern radiation therapy treatments are capable of delivering highly conformal doses, the conformality of dose that can be delivered to thoracic and abdominal lesions is degraded

0031-9155/04/235357+16$30.00

by respiratory motion (Bortfeld *et al* 2002, Jiang *et al* 2003, Bortfeld and Jiang 2004, Goitein 2004). A variety of treatment techniques have been developed or are under development for motion compensation, including margin expansion, breath holding, respiratory gating and beam tracking (Mageras and Yorke 2004, Murphy 2004, van Herk 2004). All these motion compensation methods, except for breath holding, require an adequate understanding of the motion characteristics. In particular, if treatment is to be delivered using a beam tracking technique, the important parameters of the tumour trajectory must be quantified during simulation so they may be used in treatment planning (Neicu *et al* 2003). Furthermore, these quantities must be continually monitored during treatment to ensure proper treatment delivery.

The best parametrization for describing respiratory motion remains an open question. However, it seems clear that at least the amplitude and stability of the motion must be quantified for designing margins (van Herk 2004), and the reproducibility of position in the gating window must be quantified for designing a gated treatment (Mageras and Yorke 2004). The breathing period and waveform characteristics are useful for planning treatments that use beam tracking (Keall *et al* 2001, Neicu *et al* 2003, Murphy 2004, Suh *et al* 2004). Already there exist several methods that address different aspects of this problem. Lujan *et al* and Seppenwoolde *et al* describe a method of modelling a breathing pattern with a modified cosine function (Lujan *et al* 1999, Seppenwoolde *et al* 2002). This approach gives excellent fits for breathing patterns when inhale and exhale phases are symmetrical. Neicu *et al* have described how to capture a more detailed waveform model using a concept called the average tumour trajectory (Neicu *et al* 2003). In their work, the minima and maxima in the superior–inferior direction are matched, and the waveform is normalized and averaged to better describe the characteristic motion.

The goal of this work is to design a quantitative method for describing breathing motion that captures all of the natural characteristics, and can be used for both off-line analysis and on-line monitoring. To achieve this goal, we have designed and implemented a finite state model that characterizes a breathing cycle, as containing three breathing states: exhale (EX), end-of-exhale (EOE) and inhale (IN). These three states are expected to repeat in succession as they do in natural breathing. Any motion that does not belong to any of the above three states will be treated as irregular breathing, a fourth state (IRR).

We have implemented this model in one dimension, and evaluated its fidelity on the lung tumour motion of 23 patients with peak-to-peak tumour motion greater than 7 mm (Shirato *et al* 2000). We found the model can accurately characterize a wide variety of patient breathing patterns with an average RMS error less than 1 mm. The finite state model also provides a convenient tool for statistical analysis of motion characteristics, such as amplitude changes, frequency changes and motion stability.

2. Methods

This section discusses our method for respiratory motion analysis in image guided radiation therapy. First, we introduce the finite state model for motion analysis, and then describe an on-line implementation.

2.1. Motion modelling

The finite state model for respiratory motion analysis is based on our natural understanding of breathing motion and the requirements of motion compensated treatment. As a patient inhales and exhales, a target moves in a periodic pattern as illustrated in figure 1(a). For regular

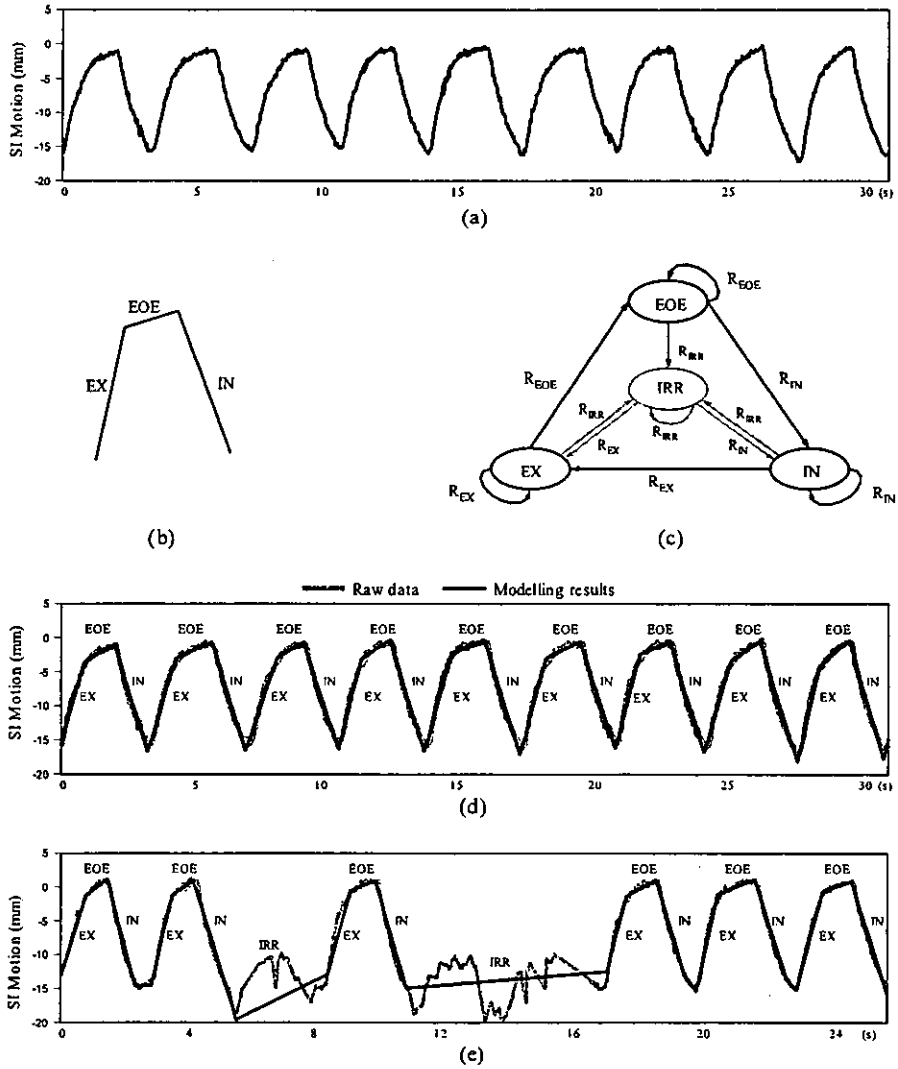

Figure 1. The finite state model for tumour respiratory motion: (a) raw data of tumour motion, (b) three states of a regular breathing cycle, (c) finite state automaton for respiratory motion, (d) regular breathing with three states per breathing cycle and (e) respiratory motion with irregular breathing.

breathing, the target motion corresponds to lung deflation, rest after lung deflation and lung expansion. Irregular breathing can occur at any time during any of the above actions. These motion patterns are modelled using three regular breathing states: exhale (EX), end-of-exhale (EOE) and inhale (IN), and one irregular breathing state (IRR). A single regular breathing cycle has exactly three breathing states, i.e., EX, EOE and IN, as shown in figure 1(b). Motion proceeds from state to state in a fixed order:

$$\ldots \Rightarrow \mathrm{EX} \Rightarrow \mathrm{EOE} \Rightarrow \mathrm{IN} \Rightarrow \mathrm{EX} \Rightarrow \ldots$$

Each of these states corresponds to a natural action: EX is the motion due to lung deflation, EOE is the motion for rest after lung deflation and IN is the motion due to lung expansion. An example of segmentation into these three states is illustrated in figure 1(d).

When a patient breathes abnormally, the respiratory motion does not fall into any of the three regular breathing states. The irregular state (IRR) is introduced to solve this problem.

The IRR is entered during irregular breathing, and is left when regular breathing resumes. An example of respiratory motion with irregular breathing is demonstrated in figure 1(e), which also shows the corresponding finite state representation by the four breathing states.

The transition from one state to another is guided by the *finite state automaton* (FSA) illustrated in figure 1(c). An FSA is an abstract machine that has a finite set of states and transitions between states. When new data arrive, the machine determines whether the tumour has transited to a next state or not. The finite state machine can begin or end in any state.

State transitions are primarily triggered by the motion velocity changes. Each state has an expected range of velocities, although the ranges for the different states overlap. Upon receipt of a new data point, a set of rules is used to decide if the tumour will stay at the current state or transit to another state. The details of the state transition rules will be described in section 2.3.

2.2. Preliminaries

This section will summarize some common notation and the fitting methods used in this paper. For generality, the corresponding notation are defined in N-dimensional space, so they can be used for 1D, 2D and 3D breathing motions. The upper-case bold italic characters, such as $\boldsymbol{V}$ or $\boldsymbol{X}$, represent vectors and directed line segments, while lower-case italic letters, such as x_0 or $v(t_0 \rightarrow t_1)$, represent scalars.

Points. An n-dimensional point at time t is represented by a vector from the origin,

$$\boldsymbol{X}(t) = \{x_0, x_1, \ldots, x_n\}.$$

Directed line segments. A directed line segment in n-D from point $\boldsymbol{X}(t_0)$ to $\boldsymbol{X}(t_1)$ is defined as

$$\boldsymbol{L} = \overrightarrow{X_0X_1} = \boldsymbol{X}(t_1) - \boldsymbol{X}(t_0) = \{\boldsymbol{X} \mid \boldsymbol{X} = \boldsymbol{X}_0 + \boldsymbol{V} \cdot t\},$$

where $t_0 \leqslant t \leqslant t_1$ and $\boldsymbol{V}$ is the velocity of $\overrightarrow{X_0X_1}$ (see below).

Line segment length. The length of a directed line segment from $\boldsymbol{X}(t_0)$ to $\boldsymbol{X}(t_1)$ is denoted as $\|\overrightarrow{X_0X_1}\|$ and is defined as

$$\|\overrightarrow{X_0X_1}\| = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{0i})^2}.$$

Velocity. The velocity from point $\boldsymbol{X}(t_0)$ to point $\boldsymbol{X}(t_1)$ is calculated from the following formula:

$$\boldsymbol{V}(t_0 \rightarrow t_1) = \frac{\overrightarrow{X_0X_1}}{t_1 - t_0} = \frac{\boldsymbol{X}(t_1) - \boldsymbol{X}(t_0)}{t_1 - t_0}.$$

Speed. The speed of tumour motion is the magnitude of the corresponding velocity,

$$v(t_0 \rightarrow t_1) = \frac{\|\overrightarrow{X_0X_1}\|}{t_1 - t_0}.$$

Amplitude. The average amplitude for two connected directed line segments $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$, where the initial point of $\boldsymbol{L}_2$ is the final point of $\boldsymbol{L}_1$, is defined as

$$\Lambda(\boldsymbol{L}_1, \boldsymbol{L}_2) = \frac{\|\boldsymbol{L}_1\| + \|\boldsymbol{L}_2\|}{2}.$$

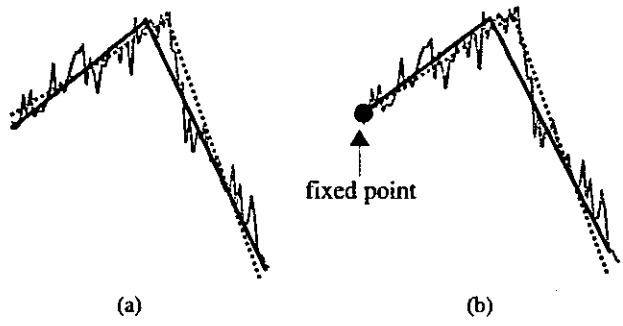

Figure 2. *Linear spline fit*: (a) free spline fit and (b) partial spline fit.

Angle of velocity change. The angle between two velocities is $\Theta(V_1, V_2)$ and calculated by

$$\Theta(V_1, V_2) = \arccos\left(\frac{V_1 \cdot V_2}{\|V_1\|\|V_2\|}\right),$$

where $V_1 \cdot V_2$ is the dot product of the two vectors.

Least square fit. Given a set of data points $\{Y(t_1), Y(t_2), \ldots, Y(t_m)\}$, we may seek a directed line segment $\overrightarrow{X_1X_m}$ that best fits the given points. We use least-squares fit which minimizes the square of differences σ^2 between the corresponding points of $\overrightarrow{X_1X_m}$ and the given points, where

$$\sigma^2 = \sum_{i=1}^{m}(X(t_i) - Y(t_i))^2.$$

Linear spline fit. A spline fit is a data analysis technique for estimating (via the least-squares criterion) the parameters in a spline polynomial model. In our method, a linear spline fit comprising two connected line segments is used for optimization. Two different situations of spline fitting are used: a free spline fit and a partial spline fit. As shown in figure 2, the free spline fit has no constraints on either of the two line segments, while the partial spline fit constrains the first knot of the first line segment to a certain prefixed point. The applications of these two spline fitting methods will be discussed in section 2.3.

Sliding window algorithm. A sliding window refers to selecting a subset of data. We use a variable-length sliding window that includes the data points from the three most recent line segments. Whenever a new knot is created, the window will move forward by one knot. New line segments can only be generated within the current sliding window. Previous line segments will not be changed by subsequent processing.

Decisive line segments. There are two line segments which are important in the state transitions. One is the most recent line segment L_f, as shown in figure 3(a). L_f is a least-squares fit of all the data points after last state transition point, and represents the tumour's current moving state. The other is the decision line segment. When a new data point arrives, a short directed line segment is fit to the K most recent points. This line segment is called the decision line segment and is denoted by L_d. These two line segments and their velocities are used to determine whether or not a new state has occurred.

Variables and symbols. To describe our algorithm, we introduce the following variables and symbols. Θ_c and Λ_c are two adjustable thresholds used for testing velocity change and

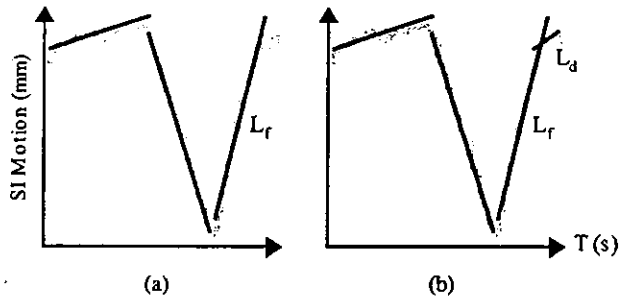

Figure 3. The most recent line segment L_f and the decision line segment L_d.

Table 1. Symbols and variables.

Symbols	Descriptions	Symbols	Descriptions
Θ_c	Velocity threshold	Λ_c	Amplitude threshold
L_f	The most recent line segment	L_d	The decisive line segment
V_f	The velocity of L_f	V_d	The velocity of L_d
S_f	The breathing state of L_f	L_{f-i}	The ith line segment before L_f
R_S	The velocity range of state S	next(S)	The next regular state after S

amplitude. L_f and L_d are the decisive line segments described above. V_f and V_d are the corresponding velocities of L_f and L_d. S_f is the breathing state of L_f. The ith line segment before L_f is denoted as L_{f-i}. R_S is the velocity range of state S (i.e., R_{EX}, R_{EOE}, R_{IN} or R_{IRR}). When S is a regular breathing state, the next regular breathing state after state S is denoted as next(S). These symbols are summarized in table 1.

2.3. Algorithm

The overall flow and decision making of our on-line algorithm is illustrated in figure 4. When each data point arrives, a decision line is generated from the data in the sliding window. Various properties of the decision line are tested against a set of rules to identify if a state transition has occurred, as summarized in table 2. Nine different actions, marked A through I, are possible. These actions control state transitions, including a possible retrospective modification of the three most recent line segments. We will now give a more detailed explanation for each action in figure 4 and table 2.

(A) *Decision line generation.* When a new data point arrives, a new decision line segment L_d is generated, and its velocity V_d is computed. By comparing L_d and V_d with the last line segment, L_f and V_f, different decisions for state transitions will be taken.

The decision rules follow the flow in figure 4. For a tumour in the irregular state (S_f = IRR), we only test whether the tumour has resumed regular breathing or not. If it has, we update the current state; otherwise, we continue in the IRR state. For a tumour in one of the regular breathing states, we test if the velocity difference, $\Theta(V_f, V_d)$, is less than Θ_c. If so, there will be no state transition. Otherwise, V_d is compared with the range of velocities allowed by the next regular state of the finite state model. If it is in range, the amplitudes of the last two line segments are checked to see if the previous greedy transition should be retracted. Only if the decision line L_d has passed these tests for state and amplitude will a new line segment be generated.

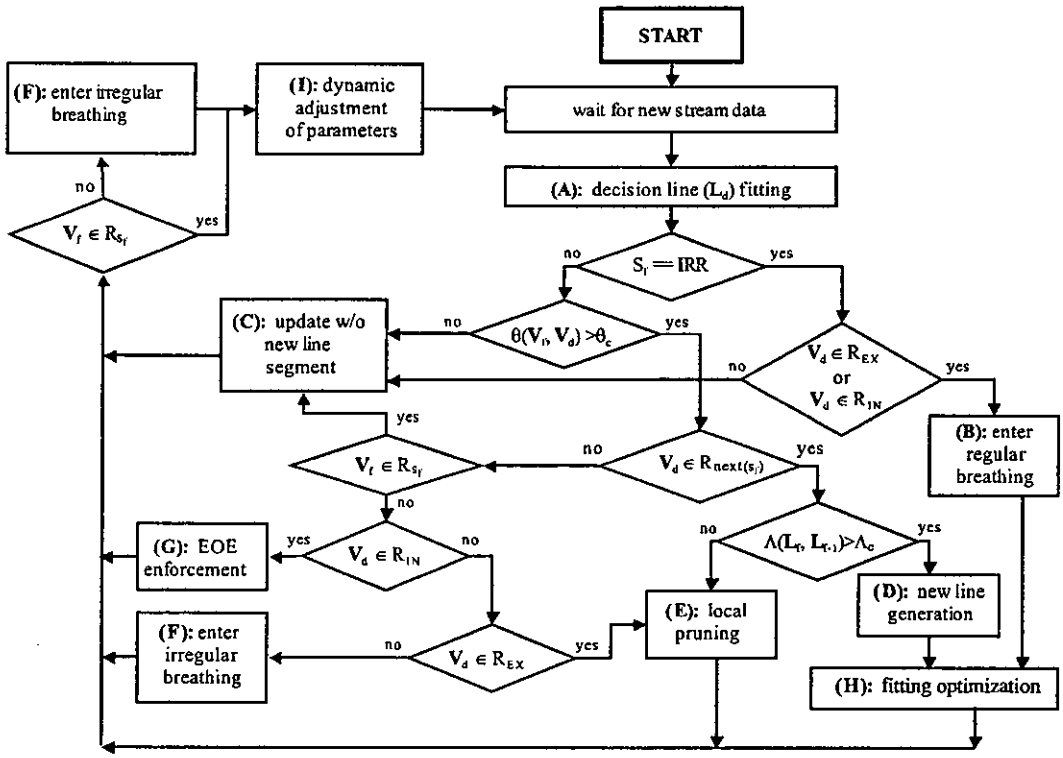

Figure 4. Algorithmic flow of on-line segmentation process.

Table 2. Decision rules.

Rule	Conditions	Decisions
1	S_f = IRR and ($V_d \in R_{EX}$ or $V_d \in R_{IN}$)	B: Enter regular breathing
2	$S_f \neq$ IRR and $\Theta(V_f, V_d) \leqslant \Theta_c$	
3	S_f = IRR and $V_d \notin R_{EX}$ and $V_d \notin R_{IN}$	C: Update without new
4	$S_f \neq$ IRR and $\Theta(V_f, V_d) > \Theta_c$ and $V_d \notin R_{next(S_f)}$ and $V_d \in R_{S_f}$	line segment
5	$S_f \neq$ IRR and $\Theta(V_f, V_d) > \Theta_c$ and $V_d \in R_{next(S_f)}$ and $\Lambda(L_f, L_{f-1}) > \Lambda_c$	D: New line generation
6	$S_f \neq$ IRR and $\Theta(V_f, V_d) > \Theta_c$ and $V_d \notin R_{next(S_f)}$ and $V_d \notin R_{S_f}$ and $V_d \in R_{EX}$	
7	$S_f \neq$ IRR and $\Theta(V_f, V_d) > \Theta_c$ and $V_d \in R_{next(S_f)}$ and $\Lambda(L_f, L_{f-1}) \leqslant \Lambda_c$	E: Local pruning
8	$S_f \neq$ IRR and $\Theta(V_f, V_d) > \Theta_c$ and $V_d \notin R_{next(S_f)}$ and $V_d \notin R_{S_f}$ and $V_d \notin R_{EX}$ and $V_d \notin R_{IN}$	F: Enter irregular breathing
9	$S_f \notin R_{S_f}$	
10	$S_f \neq$ IRR and $\Theta(V_f, V_d) > \Theta_c$ and $V_d \notin R_{next(S_f)}$ and $V_d \notin R_{S_f}$ and $V_d \in R_{IN}$	G: EOE enforcement

(B) *Resumption of regular breathing.* This action defines the transition from IRR state to normal breathing. Because of large variations of velocity changes in the EOE state, we only transition from IRR to regular breathing in the EX or IN state. Resumption of regular

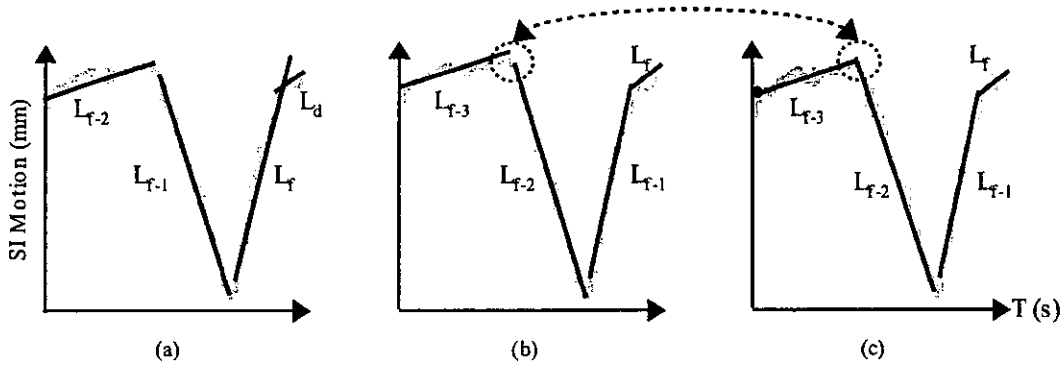

Figure 5. New line generation and fitting optimization: (a) rule checking, (b) new line generation and (c) local optimization.

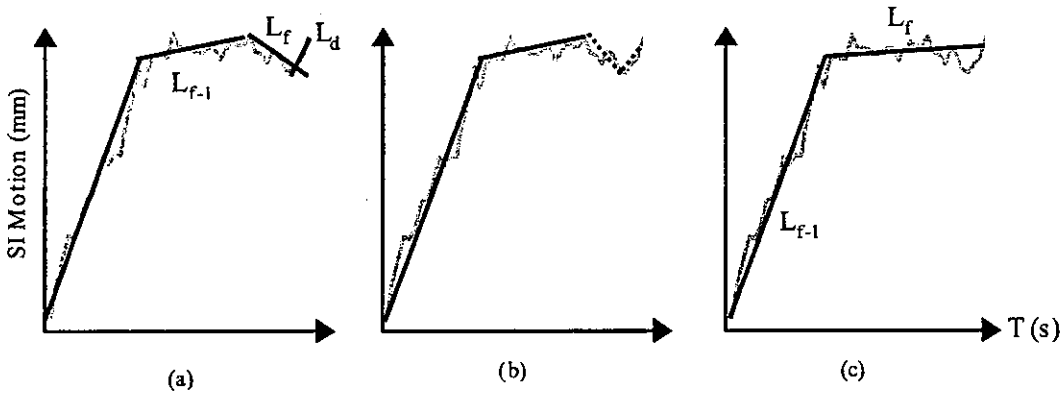

Figure 6. Amplitude checking and local pruning: (a) rule checking, (b) without amplitude checking and (c) with amplitude checking and local pruning.

breathing follows the same rules as initial state identification. A free spline fit is used to generate a new line segment, which becomes the first regular breathing state.

(C) *Update without new line segment.* If there is no need to generate a new line segment or to modify other existing line segments, new data point is added to the most recent existing line segment L_f. There are three situations that require only this simple update. The first case is that the tumour is in irregular breathing and it has not resumed regular breathing. The second case is that the tumour is in a regular breathing state and the velocity has no significant changes. The third case occurs when there is significant velocity change, but the decision velocity V_d does not fall within the velocity range of the next regular state of S_f and still lies within the velocity range of the current state.

(D) *New line generation.* When a motion transition from one state to another has occurred, a new line segment will be generated to represent the new motion state. This process is illustrated in figure 5. First, a free spline fit is applied to the points covered by the previous line segment L_f and the decision line segment L_d. As shown in figure 5(b), the new line segment becomes the new L_f, and the previous line segment becomes L_{f-1}. Finally, a local optimization over line segments L_{f-2} and L_{f-3} is performed, as will be discussed in action (H).

(E) *Local pruning.* Local pruning is an action that combines the last two line segments, L_f and L_{f-1}, when velocity change is more than Θ_c, but the amplitude change of the previous two line segments is less than Λ_c. This process is illustrated in figure 6. The

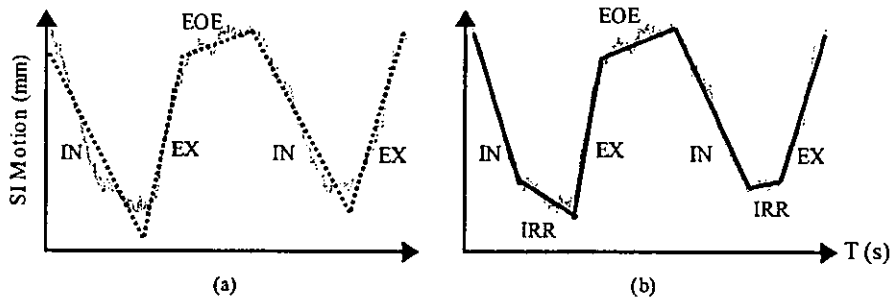

Figure 7. Irregular breathing handling: (a) without IRR state and (b) with IRR state.

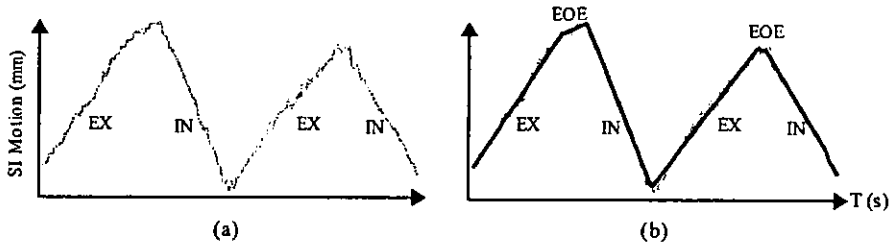

Figure 8. Enforcement of EOE state: (a) alternate exhale and inhale and (b) with EOE enforcement.

dotted line shows that without amplitude checking, an undesired new line segment can be generated. Local pruning allows backtracking when this situation occurs. The segments L_{f-1}, L_f and L_d are replaced with a single-line segment, which becomes the new L_f. The state of the new L_f is the same as that of the original L_{f-1}.

(F) *Entering irregular breathing.* There are two cases when a tumour enters irregular breathing. One case is that tumour has extended breathing hold after the IN state. An example is shown in figure 7. The dotted line segments in figure 7(a) show the line segments that would be generated if we use only the three regular breathing states. The fitting quality is much worse than the fitting quality with an irregular breathing state (figure 7(b)).

The other irregular breathing case occurs when the velocity of the final line segment does not match the allowed velocities of its corresponding state. In this case, the last line segment will be marked as IRR state.

(G) *EOE enforcement.* Another characteristic breathing pattern exists where the patient breathing seems to alternate between inhale and exhale states, and lacks a true EOE state. An example is shown in figure 8. The EOE enforcement action is used to ensure that an EOE line segment is generated at the transition from EX to IN. The minimum length of an EOE state is a tunable parameter, which we have set to 133 ms (4 data points for 30 Hz imaging rate).

(H) *Fitting optimization.* Using a free spline fit for new line segment generation, it is possible that two adjacent line segments are not well connected at the given separating point. This situation is clearly observed in figure 5(b). Modification of the line fitting parameters of L_{f-2} and L_{f-3} is required to make sure that the line segments are well connected. The optimization process is performed using a partial spline fit, with the beginning point of L_{f-3} fixed during fitting. The result of this optimization process is demonstrated in figure 5(c). After local optimization, line segment L_{f-3} is frozen, and will not be modified again by any later process. Note that because the transition point may change, L_{f-2} changes during the optimization of L_{f-1}.