Bilinear models for the inter- and intra-patient variation of the prostate

Y Jeong[1‡], R J Radke[2] and D M Lovelock[3]

[1] School of Electrical Engineering, Pusan National University, 30 Jangjeon-dong, Geumjeong-gu, Busan 609-735, Korea.
[2] Department of Electrical, Computer, and Systems Engineering
Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[3] Department of Medical Physics, Memorial Sloan-Kettering Cancer Center
New York, NY 10021, USA

E-mail: jeongy@pusan.ac.kr

Abstract. We propose bilinear models for capturing and effectively decoupling the expected shape variations of an organ both across the patient population and within a specific patient. Bilinear models have been successfully introduced in other areas of computer vision, but they have rarely been used in medical imaging applications. Our particular interest is in modelling the shape variation of the prostate for potential use in radiation therapy treatment planning. Using a dataset of 204 prostate shapes contoured from CT imagery of 12 different patients, we build bilinear models and show that they can fit both training and testing shapes accurately. We also show how the bilinear model can adapt to a new patient using only a few example shapes, producing a patient-specific model that also reflects expected content variation learnt from a broader population. Finally, we evaluate the testing and training projection error, adaptation performance and image segmentation accuracy of the bilinear model compared to linear principal component analysis and hierarchical point distribution models with the same number of parameters.

PACS numbers: 87.57.nm

Submitted to: *Phys. Med. Biol.*

Keywords: Prostate model, Shape model, Bilinear model, Image-guided radiotherapy

‡ Current address: School of Electrical Engineering, Pusan National University, 30 Jangjeon-dong, Geumjeong-gu, Busan 609-735, Korea. Tel.: +82-51-510-1704; fax: +82-51-516-4279. A preliminary version of this paper appeared in the 2006 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (Jeong and Radke 2006).

1. Introduction

Various shape models have been proposed in the medical imaging literature. Many models are designed to be applicable across a wide population of patients, and many others are customized to a specific patient's anatomy as it varies over time. In this paper, we propose *bilinear models* as a means of capturing and effectively decoupling the expected shape variations of an organ both across the patient population and within a specific patient during a course of radiotherapy. Our particular interest is in organ shape variation as it applies to radiation therapy treatment planning. Better shape models can help localize a target organ more precisely, so the outcome of treatment can be improved. In this paper, we focus on the 3D shape modelling of the prostate, using a family of data composed of CT scans of several patients acquired over the course of several days of treatment, in which the prostate had been manually outlined by physicians.

A bilinear model is similar to a linear model using principal component analysis (PCA), but it has two vectors of control parameters: a 'style' vector and a 'content' vector. In our application, 'style' corresponds to the identity of the patient and 'content' to the relative organ volume within the patient. The resulting model is then able to generate previously unseen organ shapes for both new styles and new contents. One advantage of a bilinear model is its ability to *adapt* to a new patient, producing a patient-specific model that also reflects expected content variation learnt from a broader population. First, the base bilinear model is built from organ shapes acquired on several days from each of several training patients. To adapt to a new patient, the 'style' parameter is estimated from a few contoured scans, and subsequently fixed to produce a content-varying model. Closely-related work can be found in Söhn *et al* (2005) and Freedman *et al* (2005), where the geometric variation of the prostate, bladder and rectum was analyzed using standard PCA; our work is an extension of the method to capture both inter- and intra-patient variation.

2. Point distribution models

There have been many shape modelling methods described for medical applications; a good review can be found in McInerney and Terzopoulos (1996). In this paper, we focus on point distribution models (PDMs) (Cootes *et al* 1995). In a PDM, the shape of an object is expressed by a set of points distributed along its boundary, and the model is built based on roughly homologous landmark points associated with each of several training sets.

Our method for automatically selecting landmark points suitable for shape modelling of the prostate is based on the work by Jeong and Radke (2007), described briefly as follows. Each training dataset is composed of parallel CT slices. In each slice, a physician has indicated several points around the boundary of the prostate. The difficulty is that the numbers of slices per scan and the numbers of points per

slice differ from dataset to dataset, making it difficult to establish correspondence for a shape model. Thus, we use the fast method based on elliptic Fourier descriptors (Jeong and Radke 2007) to obtain re-sampled points on uniformly-spaced parallel slices of a smooth surface that interpolates the original points in each dataset. The number of landmark points was chosen such that the set of points can accurately capture the interpolating surface of all the training shapes with a minimal number of points. Consistently re-sampling T datasets results in the training data $\{\mathbf{x}(t)\}_{t=1}^{T}$, where $\mathbf{x}(t) = (x_1(t), y_1(t), z_1(t), \ldots, x_N(t), y_N(t), z_N(t))^T$ is a column vector representing N points at matching locations in 3D. All the training shapes were aligned using the iso-center (marked by a physician for each patient) prior to model-building, which removes the need for subsequent registration. We note that any method that roughly aligns the training shapes can be used, such as matching centers of gravity, since the subsequent model will account for minor misalignments.

The active shape model (ASM) proposed by Cootes *et al* (1994) is probably the most common PDM and has been extensively used for the modelling and segmentation of medical imagery. To build the standard ASM, PCA is applied to the covariance matrix of the training data:

$$\mathbf{C}_{\mathbf{x}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{x}(t) - \overline{\mathbf{x}})(\mathbf{x}(t) - \overline{\mathbf{x}})^T \tag{1}$$

where $\overline{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)$. A set of mode vectors $\mathbf{\Phi} = [\boldsymbol{\phi}_1\boldsymbol{\phi}_2\cdots\boldsymbol{\phi}_L]$ $(L \leq T)$ is chosen that corresponds to the L largest eigenvalues. The resulting shape model can generate shapes in terms of a weighted sum of mode vectors plus the mean shape:

$$\hat{\mathbf{x}} = \overline{\mathbf{x}} + \mathbf{\Phi}\cdot\mathbf{p} \tag{2}$$

where $\mathbf{p}$ is the vector of model control parameters. Heap and Hogg (1998) proposed the hierarchical PDM (HPDM), which constrains a standard ASM so that shapes are only considered to be 'valid' if their model parameters lie in the region covered by training data projected into the model parameter space. The valid region of parameter space is determined by clustering the projections of training data, and defining a hyper-ellipsoidal subregion for each cluster. The union of these hyper-ellipsoids gives the final valid shape region. Shapes are projected onto the model in exactly the same way as the ASM, with the extra step that the model parameters $\mathbf{p}$ are projected onto the nearest valid subregion before computing (2). In addition to the number of modes of the ASM to retain, the user must also select a reasonable number of clusters and the acceptable overlap between clusters. The advantage of the HPDM is that it explicitly ensures that generated shapes resemble the training data (perhaps at the cost of generalizability).

3. Bilinear models

Because linear models have often shown good performance in many computer vision applications, it is reasonable that in approaching a problem with two dominant types of

variation, one can assume a linear model for one factor when the other factor is fixed. In the literature on bilinear models, the two factors that control the model are called *style* and *content*.

Bilinear models have been proposed for several computer vision applications. For example, Tenenbaum and Freeman (2000) applied bilinear models to several situations that have two natural independent variations, such as a set of faces seen in different poses or lighting conditions and English vowels spoken by different speakers. Our description of building bilinear models follows the approach in Tenenbaum and Freeman (2000). To our knowledge, our work is the first application of bilinear models in a medical imaging context.

Mathematically, a bilinear model is expressed in terms of two control vectors for style and content and a fixed set of basis vectors. The mapping from style s and content c, (s, c), to an observation $\mathbf{x} \in \mathbb{R}^K$ is expressed as

$$x_k = \mathbf{a}_s^T \mathbf{W}_k \mathbf{b}_c \tag{3}$$

where x_k is the k-th element of the observation, $\mathbf{W}_k \in \mathbb{R}^{I \times J}$ denotes the set of basis vectors for the k-th element, $\mathbf{a} \in \mathbb{R}^I$ denotes the style control vector (presumed to be independent of content) and $\mathbf{b} \in \mathbb{R}^J$ denotes the content control vector (presumed to be independent of style). Sometimes we write $\mathbf{x} = \mathbf{a}^T \mathbf{W} \mathbf{b}$, where $\mathbf{W} \in \mathbb{R}^{IK \times J}$ collects the basis functions for all the elements of $\mathbf{x}$. The basis vectors in $\mathbf{W}$ describe the interaction between style and content in the training set. The resulting model is called *symmetric*.

On the other hand, we can fold either the style or content vectors into the basis functions to make a style- or content-specific model (which is simply linear in its parameters). For example, when the style vector is fixed and pre-multiplied with the basis vectors, the resulting model can be expressed as

$$\mathbf{x} = \mathbf{W}^s \mathbf{b} \tag{4}$$

where $\mathbf{W}^s$ expresses the style-dependent basis functions; this model can generate different contents in a single style. Such a model is called *asymmetric*.

Our general approach is to build a symmetric model from a collection of training data, i.e., many contoured organ shapes from each of many patients. Then, based on a small number of organ shapes from a new patient, we will estimate the style vector for the new patient to generate an asymmetric, style-specific model. In the next section, we describe how to build symmetric bilinear models, and in Section 5 we discuss the style adaptation problem.

4. Building symmetric bilinear models from training data

To build a bilinear model, we try to find the parameters of the model that best fit the given training data. We require that each training observation be labelled with its corresponding style and content class prior to model-building. The total squared error

to be minimized in the symmetric model is

$$\text{Error} = \sum_{t=1}^{T}\sum_{s=1}^{S}\sum_{c=1}^{C}\sum_{k=1}^{K}\chi_{sc}(t)(x_k(t) - \mathbf{a}_s^T\mathbf{W}_k\mathbf{b}_c)^2 \tag{5}$$

where S and C are the total number of styles and contents respectively. A training vector $\mathbf{x}(t) \in \mathbb{R}^K$ denotes the t-th observation in the training dataset $\{\mathbf{x}(1), \cdots, \mathbf{x}(T)\}$. $\chi_{sc}(t)$ is an indicator function; $\chi_{sc}(t) = 1$ if the t-th observation $\mathbf{x}(t)$ is in (style, content) bin (s, c), and 0 otherwise. The dimensions (I, J) of the control vectors $(\mathbf{a}, \mathbf{b})$ must be less than or equal to the numbers of style and content bins (S, C) respectively. The $\{\mathbf{a}_s\}$, $\{\mathbf{b}_c\}$ and $\{\mathbf{W}_k\}$ that minimize (5) can be found either using an iterative singular value decomposition (SVD) method or direct minimization, depending on how the training data are distributed for each style and content class, as discussed below.

4.1. Equal numbers of observations for each (style, content) class

In the case where each (style, content) class has an equal number of instances, an iterative method using the SVD can be used to solve the training problem (Tenenbaum and Freeman 2000). While this assumption is unsatisfied for many real training situations, the SVD-based method can be used to initialize the optimization problem in such cases, as discussed in the next section.

The basic idea is to iteratively alternate the role of style and content in an asymmetric formulation until convergence. We define the training matrix, $\mathbf{X}$, and its "vector transpose", denoted as $\mathbf{X}^{VT}$, as follows:

$$\mathbf{X} = \begin{bmatrix} \overline{\mathbf{x}}^{11} & \cdots & \overline{\mathbf{x}}^{1C} \\ \vdots & \cdots & \vdots \\ \overline{\mathbf{x}}^{S1} & \cdots & \overline{\mathbf{x}}^{SC} \end{bmatrix}, \qquad \mathbf{X}^{VT} = \begin{bmatrix} \overline{\mathbf{x}}^{11} & \cdots & \overline{\mathbf{x}}^{1S} \\ \vdots & \cdots & \vdots \\ \overline{\mathbf{x}}^{C1} & \cdots & \overline{\mathbf{x}}^{CS} \end{bmatrix} \tag{6}$$

where $\overline{\mathbf{x}}^{sc}$ is the mean vector for the observations in bin (s, c), i.e. let $N_{sc} = \sum_{t=1}^{T}\chi_{sc}(T)$ and $\overline{\mathbf{x}}^{sc} = \frac{1}{N_{sc}}\sum_{t=1}^{T}\chi_{sc}(T)\mathbf{x}(t)$. Note that $\mathbf{X} \in \mathbb{R}^{SK\times C}$ and $\mathbf{X}^{VT} \in \mathbb{R}^{CK\times S}$. We arrange the elements of the basis vectors into a matrix $\mathbf{W}$ as:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{1J} \\ \vdots & \cdots & \vdots \\ \mathbf{w}_{I1} & \cdots & \mathbf{w}_{IJ} \end{bmatrix}, \qquad \mathbf{W}^{VT} = \begin{bmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{I1} \\ \vdots & \cdots & \vdots \\ \mathbf{w}_{1J} & \cdots & \mathbf{w}_{IJ} \end{bmatrix}. \tag{7}$$

Here, each $\mathbf{w}_{ij} \in \mathbb{R}^K$, so $\mathbf{W} \in \mathbb{R}^{IK\times J}$ and $\mathbf{W}^{VT} \in \mathbb{R}^{JK\times I}$. We define style and content matrices $\mathbf{A}$ and $\mathbf{B}$ as

$$\mathbf{A} = [\mathbf{a}^1 \cdots \mathbf{a}^S], \qquad \mathbf{B} = [\mathbf{b}^1 \cdots \mathbf{b}^C]. \tag{8}$$

Here, $\mathbf{A} \in \mathbb{R}^{I\times S}$ and $\mathbf{B} \in \mathbb{R}^{J\times C}$. Using the notation above, we note that the bilinear model can be expressed in either one of the following forms:

$$\mathbf{X} = (\mathbf{W}^{VT}\mathbf{A})^{VT}\mathbf{B}, \qquad \mathbf{X}^{VT} = (\mathbf{W}\mathbf{B})^{VT}\mathbf{A}. \tag{9}$$

The solution to the approximation problem

$$\min_{\mathbf{A},\mathbf{W},\mathbf{B}} \|\mathbf{X} - (\mathbf{W}^{VT}\mathbf{A})^{VT}\mathbf{B}\|^2 \text{ or } \min_{\mathbf{A},\mathbf{W},\mathbf{B}} \|\mathbf{X}^{VT} - (\mathbf{W}\mathbf{B})^{VT}\mathbf{A}\|^2 \tag{10}$$

also minimizes the approximation error in (5) when each class has the same number of observations (Tenenbaum and Freeman 2000).

It is difficult to decompose $\mathbf{X}$ into $(\mathbf{W}^{VT}\mathbf{A})^{VT}\mathbf{B}$ directly. So, to tackle the problem, $\mathbf{X}$ is expressed initially in the asymmetric form, $\mathbf{W}^s\mathbf{B}$ (where $\mathbf{W}^s = (\mathbf{W}^{VT}\mathbf{A})^{VT}$) and $\mathbf{A}$ and $\mathbf{B}$ are updated iteratively, using the following procedure:

(i) Initialization.
Initialize $\mathbf{B}$ using an SVD-based solution for the asymmetric model $\mathbf{X} = \mathbf{W}^s\mathbf{B}$. That is, decompose $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ and set $\mathbf{B}$ to be the first J rows of $\mathbf{V}^T$. $\mathbf{B}$ is orthogonal, i.e. $\mathbf{B}\mathbf{B}^T = \mathbf{I}_J$. Then ideally, $(\mathbf{X}\mathbf{B}^T)^{VT} = \mathbf{W}^{VT}\mathbf{A} (\in \mathbb{R}^{JK\times S})$ from (9).

(ii) Iteration: repeat the following until convergence.

(a) Update $\mathbf{A}$.
Compute the SVD of $(\mathbf{X}\mathbf{B}^T)^{VT} = \mathbf{W}^{VT}\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$. Update the estimate of $\mathbf{A}$ as the first I rows of $\mathbf{V}^T$. This $\mathbf{A}$ is orthogonal. So ideally, $\mathbf{X}^{VT}\mathbf{A}^T = \mathbf{W}\mathbf{B}$ $(\in \mathbb{R}^{IK\times C})$ from (9).

(b) Update $\mathbf{B}$.
Compute the SVD of $(\mathbf{X}^{VT}\mathbf{A}^T)^{VT} = (\mathbf{W}\mathbf{B})^{VT} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ $(\in \mathbb{R}^{CK\times I})$. Update the estimate of $\mathbf{B}$ as the first J rows of $\mathbf{V}^T$.

(iii) Solution.
Upon convergence (when there is negligible change in the entries of $\mathbf{A}$ and $\mathbf{B}$), solve for $\mathbf{W}$ using either one of the equations in (9), e.g. $\mathbf{W} = [(\mathbf{X}\mathbf{B}^T)^{VT}\mathbf{A}^T]^{VT}$ when using the first equation.

The convergence of the above procedure to the minimizer is guaranteed (Magnus and Neudecker 1999). We found that the procedure is not sensitive to the initial choice of $\mathbf{B}$ as long as its elements are in the typical range of the content parameter. The adequate dimensionalities of the control parameters I and J can be determined by the precentage of captured variation of the training dataset, which can be computed from the singular values of each SVD (similar to deciding the appropriate number of principal modes in PCA).

4.2. Non-equal numbers of observations for each (style, content) class

In the more general case, (5) has to be minimized using a numerical optimization algorithm such as a quasi-Newton method, with initial guesses of $\mathbf{A}$ and $\mathbf{B}$ that could be obtained from the method for the equal-observations case. The derivatives of the objective function are easy to compute.

When there are undefined entries in the training matrix $\mathbf{X}$, i.e. if there is no assigned training instance in some class (s, c), several methods can be used to fill in the missing data. One simple solution is to take the mean of the observations in the appropriate style s across all content classes or the observations in the appropriate content c across all style classes, which yielded a satisfactory result in Tenenbaum and Freeman (2000). An alternate solution would be to predict the missing data using

a model built from a complete smaller training matrix (Vlasic *et al* 2005). Also, a subspace decomposition that best describes the given data can be obtained, and this decomposition and known data act as constraints on predicting the missing data (Vlasic *et al* 2005) using probabilistic PCA (PPCA) (Tipping and Bishop 1999).

4.3. Defining the content classes

To build a bilinear shape model that can decouple inter- and intra-patient variation in our real application, we need to know beforehand the style and content class for each shape in the training data. The style can be defined as the identity of the patient, because we know which patient a specific dataset comes from. However, the content is difficult to define because even physicians cannot easily label (in a quantitative, style-independent way) the intra-patient variation for any patient scan.

Geometry changes or deformations of the prostate often occur for patients undergoing radiotherapy. For example, Balter *et al* (1995) reported translation/rotation of the prostate over time in 70% of the subjects studied. Melian *et al* (1997) reported that the change in position of the prostate related to bladder and rectal volumes caused the centre of mass to migrate from that of the planning scan. These changes might be candidates for defining the content classes, but they are difficult to quantify in a patient-independent way.

It has also been observed that the volumes of organs of interest change during radiation treatment. Roeske *et al* (1995) observed that the volume of the prostate can vary by approximately ±15% from its average volume. Antolak *et al* (1998) observed differences of up to 80% between minimum and maximum volumes of the prostate. This volume variation can be due to deformation by the bladder, rectum or tumor, or due to physical change of the organ. Hence, we propose to use the volume of the prostate to define the content classes for our bilinear model; content is usually defined as a global property of training data and volume can be a global property of an organ.

Due to the different sizes of prostates across the patient population, the volume is normalized for each patient so that relative volume defines the content classes. That is, for each patient dataset, the volumes of all observed prostates are measured, $\{v_i\}$, and are normalized as $\widetilde{v}_i = (v_i - v_{\min})/(v_{\max} - v_{\min})$ where $v_{\min}$ and $v_{\max}$ are the minimum and maximum observed volume. The normalized volume range [0, 1] is uniformly divided into the desired number of content bins.

An assignment of prostate training datasets using this strategy is shown in Figure 1. In this case, the number of styles is set to be the number of patients (6) and the content is partitioned into 5 bins. In Figure 1, the shape in each style and content class is the mean shape of the training shapes assigned to the corresponding class.

New shapes generated from one bilinear model built from this dataset are shown in Figure 2. The bilinear model was built following the procedure described above. A total of 99 shapes from the 6 patients' datasets were used to build the bilinear model with $I = S = 6$ and $J = C = 5$ (since the control parameter for style is found and

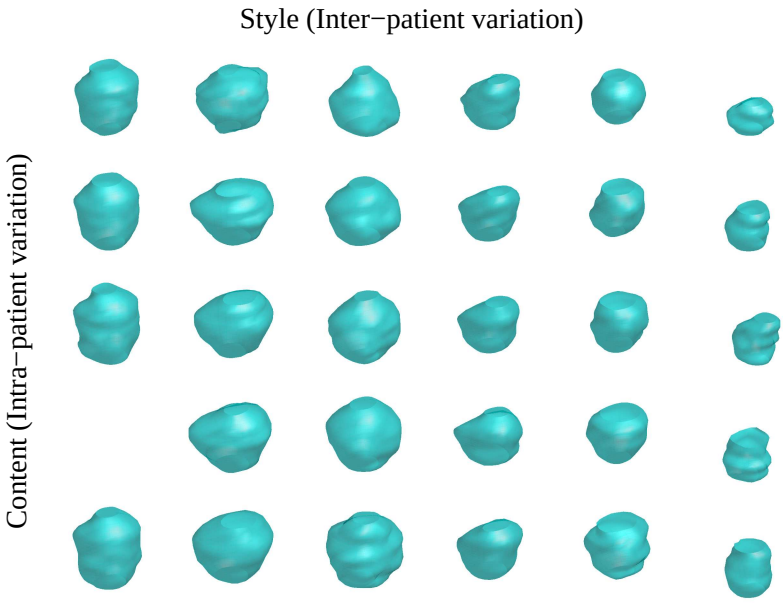

Figure 1. Mean shapes of training data assigned to each style and content class (there is one blank entry for a class that has no assigned training data).

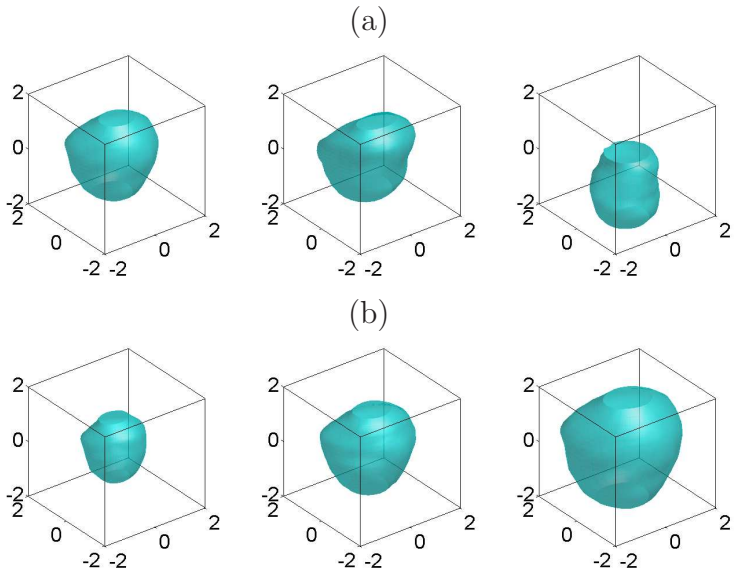

Figure 2. New observations generated using a bilinear shape model of the prostate by (a) varying the style control parameter with content control parameter fixed and (b) varying the content control parameter with style control parameter fixed (units in cm).

fixed for a new patient in our application, the dimension for the style parameter, I, is set to the number of patients for maximum expressibility). The class with no training instances was filled in with the mean observation of the same content across all styles.

5. Adapting the bilinear model to a new patient

The main rationale for using bilinear models for shape modelling of the prostate is that the available data form a natural two-factor problem, composed of shape changes due to inter- and intra-patient variation. We hypothesize that the two factors can be well-modelled using linear models individually, and that there are consistent factors governing the changes within a patient and the changes across patients. The bilinear model can be adapted easily to a new patient by estimating their style parameters from a small amount of contoured data, resulting in a patient-specific variable-content model. This is an instance of what Tenenbaum and Freeman (2000) referred to as the "translation problem".

Given M organ shapes from a new patient, $\{\mathbf{x}(1), \cdots, \mathbf{x}(M)\}$, we estimate the corresponding style vector, $\hat{\mathbf{a}}$, and content vectors, $\{\mathbf{b}_1, \cdots, \mathbf{b}_M\}$, by minimizing

$$\sum_{m=1}^{M} \sum_{k=1}^{K} (x_k(m) - \hat{\mathbf{a}}^T \mathbf{W}_k \mathbf{b}_m)^2 \tag{11}$$

where the basis matrices $\mathbf{W}_k$ are fixed from the previous training process. The initial estimates of $\hat{\mathbf{a}}$ and $\{\mathbf{b}_1, \cdots, \mathbf{b}_M\}$ are set as the mean of the style and content vectors of the training data, i.e.

$$\hat{\mathbf{a}}_{\text{initial}} = \frac{1}{S} \sum_{s=1}^{S} \mathbf{a}^s, \qquad \hat{\mathbf{b}}_{1,\text{initial}} = \cdots = \hat{\mathbf{b}}_{M,\text{initial}} = \frac{1}{C} \sum_{c=1}^{C} \mathbf{b}^c \tag{12}$$

where $\{\mathbf{a}^s\}$ and $\{\mathbf{b}^c\}$ are fixed vectors estimated during model-building. The minimization is accomplished through the same type of iterative procedure described in Section 4. This process always converged to a local minimum of (11) in our experiments. Different initial estimates of the style and content parameters may converge to slightly different sets of parameters, but we observed that they all achieved comparable approximation error for the remaining shapes.

A new shape that is specific to style $\hat{\mathbf{a}}$ can then be generated by varying the content control vector $\mathbf{b}$:

$$\mathbf{x}^{\mathbf{b}} = \hat{\mathbf{a}}^T \mathbf{W} \mathbf{b} = \mathbf{W}^a \mathbf{b} \tag{13}$$

where $\mathbf{W}^a = \hat{\mathbf{a}}^T \mathbf{W} \in \mathbb{R}^{IK \times J}$ expresses the resulting style-adapted basis matrices.

While it is desirable that the number of datasets used for adaptation should be minimal, a very limited number of organ shapes can result in a poor estimate of the style control vector. On the other hand, when a large number of shapes are used for adaptation, the model may be adapted 'too late' and can only be applied to a small number of remaining scans, limiting its usefulness. We investigate the appropriate number of adaptation datasets in our experiments below.

6. Experiments

In this section, we compare the bilinear model against a standard PCA-based linear model and a hierarchical PDM for the 3D prostates in our clinical dataset. We performed

Table 1. Summary of the datasets used for the experiments.

	Training sets						Testing sets					
Set number	1	2	3	4	5	6	7	8	9	10	11	12
Number of shapes	17	17	19	19	13	14	18	18	18	17	17	17

three types of experiments: modelling, adaptation and segmentation. In the modelling experiments, the training and testing projection error of all models are measured, which allows us to directly compare the representational ability of models with the same numbers of parameters. Next, we compare the bilinear model's performance in adapting to scans of a new patient with adaptation experiments using the other models. Finally, we run image segmentation experiments to investigate how each model performs in practice.

Each model has a notion of a number of dominant modes to be selected, and we want to investigate the performance of each model with respect to this parameter, holding other modelling choices constant. The ASMs based on linear PCA were built in a straightforward way as explained in Section 2. The bilinear models were built as described in Section 4. For the hierarchical PDMs, we fixed the number of clusters to be 3 and the overlap degree between adjacent regions to be 1.2 (Heap and Hogg 1998). These values were obtained from the experiment that produced the best segmentation results for the HPDM.

For the experiments in this section, 12 datasets were used, each containing several scans from the same patient on different days of radiotherapy, as summarized in Table 1. The training vectors for each model were composed of 400 prostate landmark points in 3D re-sampled from the clinical datasets using the method described in Jeong and Radke (2007). We used 20 slices and 20 points per slice such that the resulting set of points well approximates the interpolating surface of the training shapes as mentioned in Section 2. Sets 1–6 were used to build the various models and sets 7–12 were used for testing. The training shapes of the prostate can be all enclosed in a cube with the side length of 5 cm, and the landmark points were measured in millimeters. All the training shapes were aligned each other using the iso-center prior to model-building.

6.1. Modelling experiment

There are several dimensions to fix when building a bilinear model: S (the number of style classes), C (the number of content classes), I (the dimension of the style control vector) and J (the dimension of the content control vector). It is important to choose each dimension appropriately so the resulting model can express training data faithfully with a reasonable number of parameters. In our experiments, we always set $I = S = 6$, the number of training patients. This is primarily due to the fact that our motivating application for the bilinear model is its adaptation to a new patient. After this adaptation, the style vector is fixed (i.e., it is no longer a control parameter),

Table 2. Mean approximation error (in mm) of training (compactness C($\mathcal{M}$)) and testing shapes (generalization G($\mathcal{M}$)) using a bilinear model for several C and J values.

		C($\mathcal{M}$)						G($\mathcal{M}$)					
C		1	2	3	4	5	6	1	2	3	4	5	6
J	1	2.02	2.03	2.03	2.03	2.03	2.02	2.20	2.20	2.20	2.20	2.19	2.19
	2		1.78	1.82	1.88	1.84	1.89		1.80	1.87	1.98	1.91	2.01
	3			1.67	1.65	1.63	1.67			1.72	1.64	1.66	1.70
	4				1.51	1.48	1.56				1.43	1.48	1.50
	5					1.39	1.42					1.30	1.35
	6						1.32						1.23

so I was set to S for maximum style expressibility. The determination of J is more difficult because $J \leq C$ and C has to be determined *a priori*. Thus, we conducted two approximation experiments to choose C and J among various combinations.

The experiments are based on the error in projecting the training data onto the model (also called compactness) and the error in projecting the testing data onto the model (also called generalization) (Styner *et al* 2003). The projection error is based on a distance measure D between two shapes $\mathbf{x} = (x_1, y_1, z_1, \ldots, x_N, y_N, z_N)$ and $\hat{\mathbf{x}} = (\hat{x}_1, \hat{y}_1, \hat{z}_1, \ldots, \hat{x}_N, \hat{y}_N, \hat{z}_N)$. In our experiments, we used the average Euclidean distance between corresponding points given by

$$D(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N} \sum_{n=1}^{N} \sqrt{(x_n - \hat{x}_n)^2 + (y_n - \hat{y}_n)^2 + (z_n - \hat{z}_n)^2}. \tag{14}$$

We denote $P_{\mathcal{M}}(\mathbf{x})$ as the projection of shape $\mathbf{x}$ onto model $\mathcal{M}$. Projection of a shape onto a PCA model is performed using (2) with $\mathbf{p} = \mathbf{\Phi}^T \cdot (\mathbf{x} - \overline{\mathbf{x}})$, and projection onto a bilinear model is performed as explained in Section 5.

A family of bilinear models with $I = 6$ and varying C and J were built using 99 prostate shapes from the 6 training datasets. The maximum C was set to 6 in our experiment because partitioning with more than 6 classes generates many missing style/content training shapes, which makes learning the model parameters more difficult. For each model $\mathcal{M}$, the compactness is measured as the average approximation error for all the samples in the training datasets $\{\mathbf{x}(1), \ldots, \mathbf{x}(T)\}$: $\mathrm{C}(\mathcal{M}) = \frac{1}{T} \sum_{t=1}^{T} D(\mathbf{x}(t), P_{\mathcal{M}}(\mathbf{x}(t)))$. The generalization ability G($\mathcal{M}$) is similarly measured for all the samples in the testing datasets. The results are summarized in Table 2. The table indicates that the choice of C is not critical for a fixed J, which can be observed across each row. Therefore, it seems reasonable to choose $J = C$ for the bilinear models in our application. The testing error in Table 2 shows a similar trend to the training error, with slightly higher approximation error overall as expected.

For comparison, we built ASMs and HPDMs using the same 99 prostate shapes using between 7 and 12 dominant modes. The training and testing projection error are compared to the best bilinear model with the same number of parameters in Table 3. We can see that all the models have good performance in an absolute sense (i.e., between

Table 3. Mean approximation error (in mm) of ASM, HPDM and best bilinear models of the same order (N=7,...,12) over the training and testing shapes.

	Training						Testing					
order (N)	7	8	9	10	11	12	7	8	9	10	11	12
ASM	1.49	1.40	1.30	1.22	1.16	1.10	1.63	1.55	1.44	1.33	1.28	1.22
HPDM	1.66	1.56	1.46	1.37	1.27	1.21	1.84	1.76	1.68	1.58	1.56	1.51
Bilinear	2.19	1.80	1.64	1.43	1.30	1.23	2.02	1.78	1.63	1.48	1.39	1.32

1 and 2 mm error), differing by a fraction of a mm in accuracy. However, the bilinear models have an advantage that the ASMs and HPDMs do not: the ability to generate inter-patient or intra-patient variation when the other factor is fixed.

6.2. Adaptation experiment

Next, we compare the adaptation ability of the bilinear model to the other models. We investigated the adaptation performance as a function of the number of datasets used to adapt, denoted M, as well as the dimension of the style-adapted model, J. As discussed previously, it is desirable that both parameters be kept small while maintaining low projection errors. To build ASMs and HPDMs to compare against the bilinear model, we used order-J models built from the same 99 prostate shapes in the training set, plus the first M scans from the new patient used to adapt the bilinear model. For each patient, the projection error of the remaining scans not used for model-building was measured for all models. The mean projection error over all testing patients as a function of M and J=4, 5, 6 is illustrated in Figures 3(a), 3(b) and 3(c) respectively.

It can be observed that while adding the adaptation scans has relatively little effect on the ASM and HPDM, the bilinear model consistently improves with more adaptation datasets. Especially for low model orders, the bilinear model improves on the linear model after only a few examples. This is desirable, since it would be realistically unlikely that more than a couple scans could be manually contoured in a clinical setting. Figure 3(d) shows the results for testing set 12. In this view, it can be observed that "elbow points" of diminishing returns for the bilinear model occur at about M=3, and that the adapted bilinear models with 4 or 5 parameters are already as good as the ASM with 6 parameters. These results are in good agreement (though in a different context) with the results of Yan *et al* (2000), which concluded that a few initial CT scans (about 5) from a new patient over the course of fractionated radiotherapy are sufficient to construct a cost function that accurately calculates dose distributions for the patient.

6.3. On the use of inter-patient geometric information

To verify that the inter-patient geometric variation is necessary, we conducted an additional modelling experiment using a model which we call the "small PCA model";

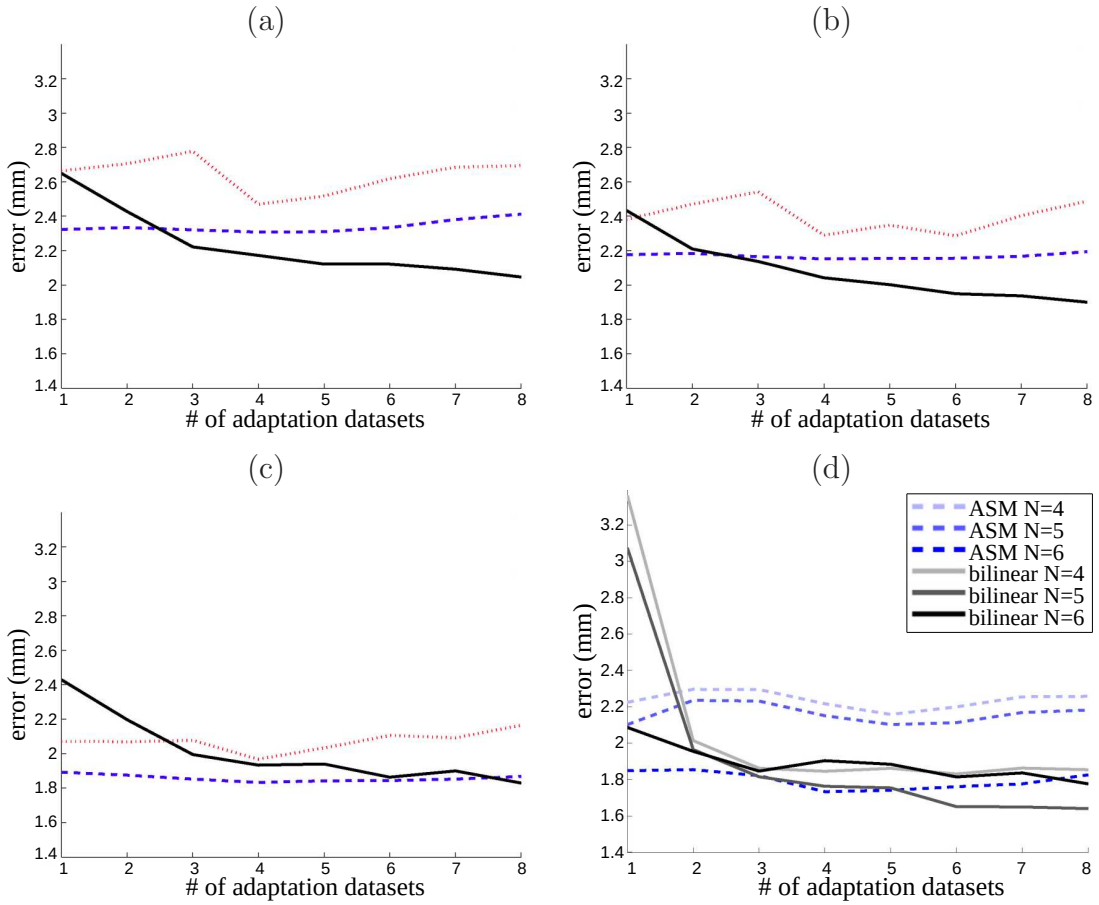

Figure 3. Mean projection error versus the number of shapes used for adaptation for models of dimension (a) 4, (b) 5 and (c) 6 (- - - -: ASM, ·····: HPDM, ——: bilinear). Panel (d) shows the projection errors of the ASMs and bilinear models with equal model orders (4, 5 and 6) for testing set 12. In this case, the bilinear model outperforms the ASM model with the same model order after only a few adaptation datasets.

the model was built using only a few initial manually-contoured shapes from a new patient. The rest of the shapes of the patient were approximated and averaged Euclidean distances of the corresponding points were measured. The result is shown in Figure 4 for testing set 12; the small PCA model has smaller approximation error than the adapted bilinear model when using more than 7 sets, but at this point there would be little clinical benefit for automatic segmentation because there remain few scans to be segmented.

6.4. Segmentation experiment

Because the shape model can be used as a constraint in segmentation, it is also useful to see how each model actually performs in segmentation problems. We note that segmentation is not the focus of this paper, and that different results would be obtained with different segmentation methods.

Each model was built using the 6 training datasets, plus 4 sets of contours from the

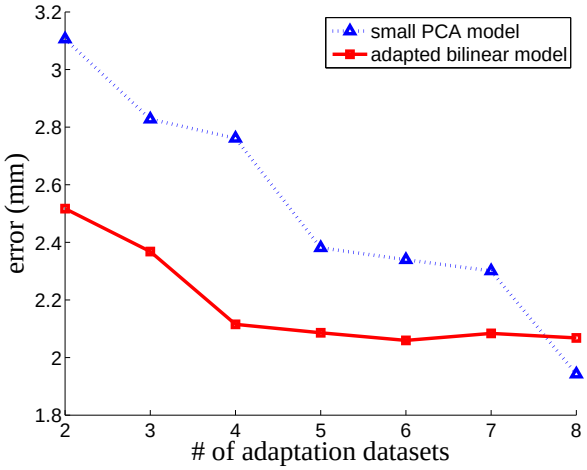

Figure 4. Approximation error for adapted bilinear model and small PCA model against the number of adaptation datasets, for testing set 12.

given testing patient. In the bilinear model, these four extra sets of contours were used for style adaptation; otherwise they were collected with the training data for model-building, just as described in the previous section. The style-adapted bilinear model, ASM and HPDM were then used to segment the remaining images of each test set. The dimensionality of all the models was set to 4 so that the segmentation was fast. For all experiments, we used the 3D model-based segmentation algorithm described in Freedman *et al* (2005). This algorithm is based on minimizing the distance between the cumulative density function (CDF) of the intensities of the pixels inside the evolving surface boundary ($\mathbf{q}^{S(\beta)}$) and the CDF of an appearance model ($\mathbf{m}(\beta)$), subject to the constraint that the evolving surface is generated by the specified parametric shape model ($S(\beta)$):

$$\min_{\beta} K\left(\mathbf{q}^{S(\beta)}, \mathbf{m}(\beta)\right) \tag{15}$$

where K is a distance measure between two CDFs. The appearance model can be obtained as the pixel-intensity histogram of the region in an image to be segmented that best matches one of the pixel-intensity histograms of the training data. Refer to Freedman *et al* (2005) for more details of the segmentation method.

One segmentation result for the bilinear model applied to the CT images from a testing dataset is shown in Figure 5. Median statistics of the segmentation results for all the scans of the testing datasets are summarized in Table 4. The segmentation is validated against manual ground-truth using the following measures:

- v_{d}, the probability of detection, calculated as the fraction of the ground-truth organ that was contained by the estimated organ. For a good segmentation, v_{d} should be close to 1.
- v_{fa}, the probability of false alarm, calculated as the fraction of the estimated organ that lies outside the ground-truth organ. For a good segmentation, v_{fa} should be close to 0.

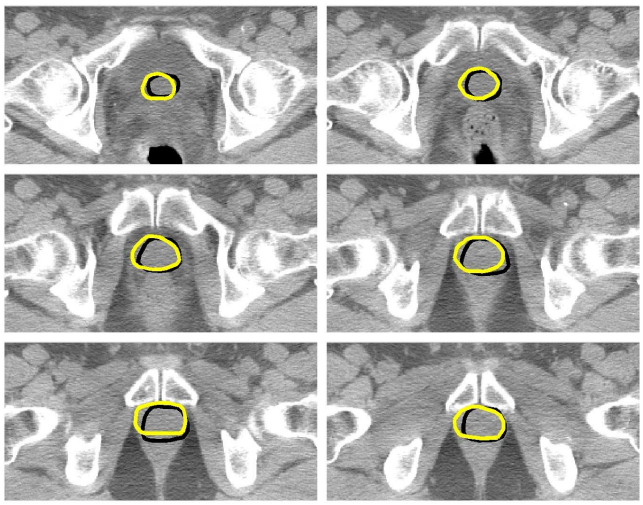

Figure 5. Cross-sections of the automatic segmentation of one 3D CT scan from testing patient 12 using the bilinear model of order 4. The dark contours are the automatic model-based segmentation results, and the light contours are the physician-labelled ground-truth.

Table 4. Image segmentation statistics over the testing datasets.

Statistic	Set	ASM	HPDM	Bilinear	Set	ASM	HPDM	Bilinear
med V_{d}		0.45	0.70	0.71		0.70	0.72	0.77
med V_{fa}	7	0.05	0.11	0.09	10	0.16	0.18	0.10
med centr. dist. (mm)		12.90	6.31	5.88		9.94	8.48	6.18
med V_{d}		0.64	0.55	0.56		0.76	0.89	0.68
med V_{fa}	8	0.08	0.09	0.12	11	0.39	0.42	0.24
med centr. dist. (mm)		8.86	10.10	10.46		10.52	8.15	9.71
med V_{d}		0.69	0.65	0.64		0.71	0.70	0.78
med V_{fa}	9	0.03	0.03	0.15	12	0.54	0.43	0.36
med centr. dist. (mm)		4.97	5.72	6.56		10.80	10.10	5.95

- The centroid distance, calculated as the norm of the vector connecting the centroids of the ground-truth and estimated organs.

We can see from the table that the bilinear models have good relative and absolute performance. Although the patient-specific models built using the bilinear model can better approximate the unseen shapes of the new patient than other methods, the segmentation algorithm does not always seem to find the best possible shapes. If such an optimal segmentation is possible, the bilinear model can yield the best performance theoretically as shown in the modelling experiment in Section 6.1.

7. Conclusions

We demonstrated a novel application of bilinear models to the shape modelling of anatomical objects. The experimental results showed that the bilinear shape model performs well in fitting both training and testing shapes. In adaptation tests, the bilinear shape model performed as well as or better than a linear shape model using the same number of parameters, especially for small model orders. The results show the potential of bilinear models to compactly represent medical datasets that have two inherent types of variation. Our initial tests indicate that the adapted bilinear model can be used as an effective shape constraint in the segmentation of medical images. In future work, the bladder and rectum, the two organs surrounding the prostate, could also be modelled using the proposed method. We previously showed in Freedman *et al* (2005) that PDMs can be successfully applied to the rectal wall (basically a tubular shape), and the bladder (basically a round shape) can be represented similarly to the prostate.

We believe a better shape model using the bilinear approach is possible, since the assignment of content classes plays an important role in the bilinear model and could be improved. For example, additional information about the patients such as age or weight could be used in defining the content bins if there is correlation between the shape change of an organ and these characteristics of the patient. Although we assumed two factors in our data, the number of true inherent variation axes may be more than two. In such a case, a multilinear model (e.g. Vasilescu and Terzopoulos 2002) may express the variation better than bilinear model, although the difficulty of defining good content classes still remains.

Acknowledgments

This work was supported in part by CenSSIS, the NSF Center for Subsurface Sensing and Imaging Systems, under the award EEC-9986821.

References

Antolak J A, Rosen I I, Childress C H, Zagars G K and Pollack A 1998 Prostate target volume variations during a course of radiotherapy *Int. J. Radiat. Oncol. Biol. Phys.* **42** 661–72

Balter J M, Sandler H M, Lam K, Bree R L, Lichter A S and ten Haken R K 1995 Measurement of prostate movement over the course of routine radiotherapy using implanted markers *Int. J. Radiat. Oncol. Biol. Phys.* **31** 113–8

Cootes T F, Hill A, Taylor C J and Haslam J 1994 The use of active shape models for locating structures in medical images *Image Vis. Comput.* **12** 355–65

Cootes T F, Taylor C J, Cooper D H and Graham J 1995 Active shape models - their training and application *Comput. Vis. Image Und.* **61** 38–59

Freedman D, Radke R J, Zhang T, Jeong Y, Lovelock D M and Chen G T Y 2005 Model-based segmentation of medical imagery by matching distributions *IEEE Trans. Med. Imaging* **24** 281–92

Heap T and Hogg D 1998 Wormholes in shape space: tracking through discontinuous changes in shape *Proc. Int. Conf. on Computer Vision (Bombay, India)* 344–9

Jeong Y and Radke R J 2006 Modeling inter- and intra-patient anatomical variation using a bilinear model *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshop (New York, NY)* 76–76

Jeong Y and Radke R J 2007 Reslicing axially sampled 3D shapes using elliptic Fourier descriptors *Med. Image Anal.* **11** 197–206

Magnus J R and Neudecker H 1999 *Matrix differential calculus with applications in statistics and econometrics* rev ed (New York, NY: John Wiley & Sons)

McInerney T and Terzopoulos D 1996 Deformable models in medical images analysis: a survey *Med. Image Anal.* **1** 91–108

Melian E, Mageras G S, Fuks Z, Leibel S A, Niehaus A, Lorant H, Zelefsky M, Baldwin B and Kutcher G J 1997 Variation in prostate position quantitation and implications for three-dimensional conformal treatment planning *Int. J. Radiat. Oncol. Biol. Phys.* **38** 73–81

Roeske J C, Forman J D, Mesina C F, He T, Pelizzari C A, Fontenla E, Vijayakumar S and Chen G T Y 1995 Evaluation of changes in the size and location of the prostate, seminal vesicles, bladder, and rectum during a course of external beam radiation therapy *Int. J. Radiat. Oncol. Biol. Phys.* **33** 1321–9

Söhn M, Birkner M, Yan D and Alber M 2005 Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation *Phys. Med. Biol.* **50** 5893–908

Styner M A, Rajamani K T, Nolte L P, Zsemlye G, Szekely G, Taylor C J and Davies R H 2003 Evaluation of 3D correspondence methods for model building *Proc. Information Processing in Medical Imaging (Ambleside, UK)* vol 18 63–75

Tenenbaum J B and Freeman W T 2000 Separating style and content with bilinear models *Neural Comput.* **12** 1247–83

Tipping M E and Bishop C M 1999 Probabilistic principal component analysis *J. R. Stat. Soc.* B-*Stat. Methodol.* **61** 611–22

Vasilescu M A O and Terzopoulos D 2002 Multilinear analysis of image ensembles: TensorFaces *Proc. European Conf. on Computer Vision (Copenhagen, Denmark)* Part I ed A Heyden *et al* vol 2350 447–60

Vlasic D, Brand M, Pfister H and Popović J 2005 Face transfer with multilinear models *ACM Trans. Graph.* **24** 426–33

Yan D, Lockman D, Brabbins D, Tyburski L and Martinez A 2000 An off-line strategy for constructing a patient-specific planning target volume in adaptive treatment process for prostate cancer *Int. J. Radiat. Oncol. Biol. Phys.* **48** 289–302