Int J Comput Vis (2009) 85: 237–252
DOI 10.1007/s11263-009-0212-6

Bilinear Models for Spatio-Temporal Point Distribution Analysis

Application to Extrapolation of Left Ventricular, Biventricular and Whole Heart Cardiac Dynamics

Corné Hoogendoorn · Federico M. Sukno · Sebastián Ordás · Alejandro F. Frangi

Received: 31 January 2008 / Accepted: 13 January 2009 / Published online: 12 February 2009

Abstract In this work we describe the usage of bilinear statistical models as a means of factoring the shape variability into two components attributed to inter-subject variation and to the intrinsic dynamics of the human heart. We show that it is feasible to reconstruct the shape of the heart at discrete points in the cardiac cycle. Provided we are given a small number of shape instances representing the same heart at different points in the same cycle, we can use the bilinear model to establish this.

Using a temporal and a spatial alignment step in the preprocessing of the shapes, around half of the reconstruction errors were on the order of the axial image resolution of 2 mm, and over 90% was within 3.5 mm. From this, we conclude that the dynamics were indeed separated from the inter-subject variability in our dataset.

The work of A.F.F. was supported by the Spanish Ministry of Education and Science under a Ramon y Cajal Research Fellowship. This work was partially developed within the framework of the CENIT-CDTEAM Project funded by the Spanish CDTI-MITYC, and also partially supported by grants MEC TEC2006-03617/TCM and ISCIII FIS2004/40676.

C. Hoogendoorn · F.M. Sukno · S. Ordás · A.F. Frangi (✉)
Center for Computational Imaging and Simulation Technologies in Biomedicine (CISTIB), Department of Information and Communication Technologies, Universitat Pompeu Fabra, Carrer Roc Boronat 138, 08018 Barcelona, Spain
e-mail: alejandro.frangi@upf.edu

C. Hoogendoorn
e-mail: corne.hoogendoorn@upf.edu

F.M. Sukno
e-mail: federico.sukno@upf.edu

C. Hoogendoorn · F.M. Sukno · A.F. Frangi
Networking Biomedical Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Spain

Keywords Statistical shape modeling · Cardiac modeling · Cardiac dynamics · Bilinear models · Spatiotemporal decomposition

1 Introduction

1.1 Statistical Shape Models and Time

The use of statistical models of shape has established itself as a popular approach to image analysis problems, in the domain of both natural (Pantic and Rothkrantz 2000) and medical image analysis (McInerney and Terzopoulos 1996; Duncan and Ayache 2000). Along the way, much research has been devoted to the development of various types of shape models, as well as to solving problems arising from the construction of such models. Many of those models, however, are essentially spatial models, and extension to the spatiotemporal domain is not as trivial or trivially justifiable as the extension from d to $d + 1$ spatial dimensions.

The analysis of Point Distribution Models (PDM's) in shape space received a significant amount of attention from the mid-1980's to the early 1990's (Cootes et al. 1992; Kendall 1984; Le and Kendall 1993; Mardia and Dryden 1989). The most renowned result from this work applied in computer vision is the emergence of the Principal Component Analysis (PCA)-based statistical shape model from Cootes et al. (1995), who applied PCA to the covariance matrix of their data set in order to extract a set of orthogonal variations of the sampled points. These became well-known as the modes of variation of the shape class.

Efforts to extend linear shape models to the spatiotemporal domain have been made before, for example by

Hamarneh and Gustavsson (2004), in whose work each sample in the dataset consists of an entire sequence of observations of the same object sampled throughout the temporal exposure window. Mitchell et al. (2002) and Bosch et al. (2002) employ the same strategy, extending Active Appearance Models (Cootes et al. 2001) to Active Appearance Motion Models in order to segment the endocardium in echocardiograms and Magnetic Resonance image sequences, respectively. Perperidis (2005) constructs two linear models using PCA: one which models the variation across the mean shapes of the subjects, and another one which models the variation within the cardiac cycle. Another approach to add the element of time to 3-dimensional segmentation using statistical shape models was presented by Montagnat and Delingette (2005). After building a PCA-based model treating all subjects and all phases as separate samples, they employ a scheme which segments the cycle as one single object, rather than employing a sequential scheme to segment each frame separately. During segmentation, temporal constraints are introduced in the optimization to limit the differences between segmentations of subsequent frames to reasonable values. Lynch et al. (2008) model the temporal processes in their level-set approach for cardiac segmentation as a distance function between a set of control points and the level sets.

Statistical models of cardiac left ventricular (LV) deformation only were constructed by Chandrashekara et al. (2003), using data from one single subject. A comparable approach to modeling respiratory motion of the liver was taken by Blackall et al. (2001).

To the best of our knowledge, our previous work (Hoogendoorn et al. 2007) was the first to model individual and temporal variations of cardiac shape as two different sources of variability within the same set of data. Before, inter-subject variation and dynamics were not decoupled at all (Bosch et al. 2002; Mitchell et al. 2002; Hamarneh and Gustavsson 2004). Later, the application of the model was equipped with constraints to limit the first-order derivative of shape points over time (Montagnat and Delingette 2005) or by constraining it using a distance prior (Lynch et al. 2008). By creating a spatiotemporal model of cardiac dynamics that decouples individual and temporal variations, we can extrapolate cardiac phases from the statistical model even when they are not available from the individual measurements. The models most similar to our approach are those from Perperidis (2005). However, our method does not result in a massively higher-dimensional parameterization for intra-subject variability, resulting in a more compact model.

1.2 Bilinear Statistical Models

In biometrics, the separation of two (independent) processes that contribute to the overall pattern variability is a well-known problem, which has led to the introduction of bilinear models by Tenenbaum and Freeman (1996). The reader is referred to Tenenbaum and Freeman (2000) for a more detailed description. Dubbing the two sources of variability *style* and *content*, these names can be assigned freely depending on which is most natural given a specific problem. Aside from the examples used in Tenenbaum and Freeman (2000), the literature contains examples of the separation of

- face identity and facial expression (Abboud and Davoine 2004)
- location and content for sparse coding of natural images (Grimes and Rao 2005)
- emotion and speech content (Chuang and Bregler 2005)
- gait (walking characteristics) and viewing conditions (Hsu et al. 2005; Lee and Elgammal 2004)
- pairs from the set {identity, action, viewpoint} (Cuzzolin 2006)
- face identity and viewpoint (González-Mora et al. 2007)
- face identity and illumination (Shin et al. 2008).

While in Tenenbaum and Freeman (2000) the bilinear models and their construction are formulated quite specifically for bilinear decomposition, which we expand on in Sect. 2, they follow the same principles of the multilinear decompositions of higher order tensors presented by De Lathauwer et al. (2000), which was followed by Vasilescu and Terzopoulos (2003) for their trilinear decomposition of expression, identity and illumination.

1.3 Clinical Context

The use of statistical models in cardiac image analysis initially focused primarily on segmentation (Suri 2000; Bosch et al. 2002; Mitchell et al. 2002; Montagnat and Delingette 2005; Lynch et al. 2008). However, the analysis of cardiac function—dynamics and deformation—has since emerged as a relatively new field of application to which intensive research has been dedicated. In the context of congestive heart failure (CHF) assessment, left ventricular (LV) function in particular has been the focus of automated localization and quantification algorithms, employing various types of models. Automated methods for objective modeling and analysis of cardiac morphology and function are therefore desirable and relevant. For an overview of cardiac modeling, we refer to Frangi et al. (2005).

Most recently, deformation recovery and modeling was done by Bistoquet et al. (2007), who make use of the near non-compressibility of myocardial tissue to constrain the segmentation of the LV and simultaneously extract the underlying within-tissue deformation. Liu and Shi (2007) analysed cardiac motion under constraints obtained from biomechanical priors.

Lekadir et al. (2007) model LV deformation as interlandmark motion in a local circular coordinate system in which

anomalies present themselves as outliers. Leung and Bosch (2007) create a sparse shape variation decomposition using PCA and subsequent orthomax rotations to characterize local abnormal deformations, while Syeda-Mahmood et al. (2007) use registration and the associated deformation patterns to characterize deformation abnormalities in 2D ultrasound sequences.

While they may be robust to intersubject variability, none of the above approaches really takes this source of variation into account. By factoring from a dataset the variation attributed to inter-subject differences from that of cardiac dynamics, be it shapes, deformation fields or the images themselves, the contraction pattern of an entire population can be modeled, and then used in various clinical applications. In this paper we employ bilinear models to establish this factorization on cardiac shapes, and illustrate their power to parameterize the dynamics (or rather a set of discrete phases within the cardiac cycle) and the subject by extrapolating from a small subset of phases to the remaining phases in the cycle.

The remainder of this work is organized as follows: we introduce the concept of and techniques behind bilinear models in Sect. 2. Our extrapolation experiments are described in Sect. 3. The results are presented in Sect. 4, followed by a discussion and future research directions in Sect. 5. We conclude this work with Sect. 6.

2 Bilinear Statistical Models

A bilinear model is a two-factor model which is linear in either factor when the other one is kept constant:

$$y = \mathbf{a}^{\mathrm{T}}\mathbf{W}\mathbf{b}, \tag{1}$$

where y is a scalar observation, $\mathbf{a}$ and $\mathbf{b}$ are parameterization vectors defined by the factors, and $\mathbf{W}$ is a constant matrix governing the interaction between the factors.

Extending this to the case of multivalued observations, each element y_k^{sc} of a K-sized observation $\mathbf{y}^{sc}$ in style s and content c can be described by a bilinear model as

$$y_k^{sc} = \sum_{i=1}^{I}\sum_{j=1}^{J} w_{ijk} a_i^s b_j^c. \tag{2}$$

I and J are the sizes of the parameterization vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. $\mathcal{W}$ is now a 3-dimensional matrix, which forms a mapping from the style and content spaces into observation space and as such is of size $I \times J \times K$. Each $\mathbf{w}_{ij}$ is a K-sized base observation, much akin to the eigenface (Turk and Pentland 1991) and eigenshape (Cootes et al. 1995). In the case of speech recognition, one could think of them as voice harmonics that need to be combined to form a certain phoneme in a certain accent or intonation. The $\mathbf{a}$ and $\mathbf{b}$ vectors provide the information on how to combine those base observations. For the time being, we will adhere to the original nomenclature and call these the content and style vectors, respectively (hence the c and s superscripts). Figure 1 illustrates how the model consists of the sets of parameters $\mathbf{a}$ and $\mathbf{b}$, providing a compact representation of the style and the content, and the base observations $\mathbf{W}$, derived from the examples in the bottom-right quadrant.

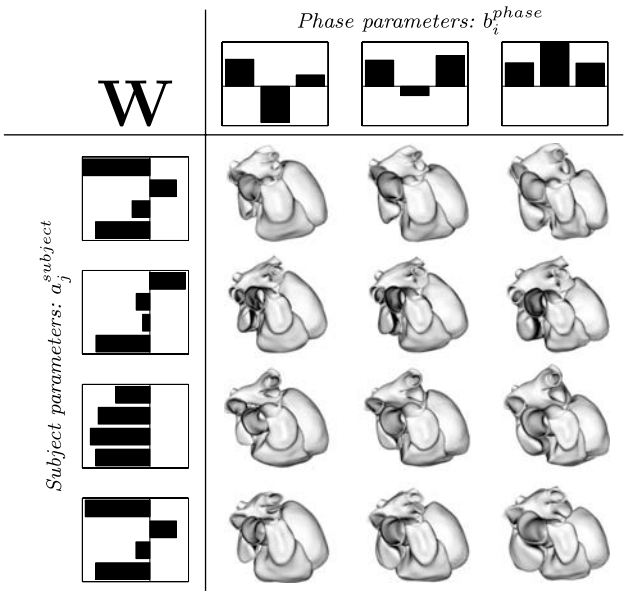

Fig. 1 The structure of a symmetric bilinear model. Along the *horizontal axis*, three phases from the cardiac cycle are shown. Along the *vertical axis* we have four different subjects. The set of basis observations $\mathbf{W}$ can not be visualized in a meaningful manner

The choice for bilinear models stems from the idea that variations in a set of observations are the consequence of the variation of two independent factors. The examples used to illustrate the usefulness of bilinear models in Tenenbaum and Freeman (2000) call upon analyzing the way we manage to recognize known characters, people or phonemes in a font or under viewing circumstances not observed before, or uttered in an accent not heard before. Somehow, we know the invariants of that character, person, or phoneme, and in an observation we can recognize those irrespective of the variations introduced by the circumstances.

2.1 Asymmetric Models

A combining matrix $\mathbf{A}$ can be the result of contracting $\mathbf{a}\mathcal{W}$ into a single matrix, leading to an asymmetric model:

$$y_k^{sc} = \sum_{j=1}^{J} a_{jk}^s b_j^c, \tag{3}$$

where

$$a^s_{jk} = \sum_{i=1}^{I} w^s_{ijk} a^s_i. \tag{4}$$

This is useful when the mapping from content to observation is dependent on style. The J observations of size K each that make up $\mathbf{A}$ are then style-specific base observations that can be mixed using the parameterization in $\mathbf{b}$. In essence, the model has then become a unilinear model. It is possible to invert the roles of style and content, and thus build a different asymmetric model from the same training data.

2.2 Symmetric Models

The symmetric model is the original bilinear model as already presented in (2):

$$y^{sc}_k = \sum_{i=1}^{I} \sum_{j=1}^{J} w_{ijk} a^s_i b^c_j,$$

where the mapping from the style and content spaces to observation space, $\mathcal{W}$, is dependent on neither style nor content. The elements of $\mathcal{W}$ are base observations that look like eigenfaces (Turk and Pentland 1991) but do not represent an orthogonal basis like eigenelements. The base observations can then be mixed using the $\mathbf{a}$ and $\mathbf{b}$ parameterizations to form any element from the training set, and other parameterizations can be used to construct observations of new style and/or new content.

2.3 Factoring Inter-Subject Variability and Cardiac Dynamics

2.3.1 Representation of Training Data

The construction of the bilinear models as presented in Tenenbaum and Freeman (2000) assumes a vector representation of the observations. This is easily achieved by employing a surface representation as is commonly used for the construction of PDM's like those in Cootes et al. (1995). In d-dimensional Euclidean space, each training shape is landmarked with n_L points—pseudolandmarks (Dryden and Mardia 1998)[1]—of anatomical correspondence throughout the training set. The set of point coordinates is then concatenated to form an $(n_\mathrm{L}d)$-dimensional shape vector, or a single point in an $(n_\mathrm{L}d)$-dimensional shape space.

After landmarking and vectorization of our observations, we have CS $(n_\mathrm{L}d)$-dimensional shape vectors: C frames per time sequence, S subjects (with one sequence each), n_L landmarks in d dimensions. As is the case in Hamarneh and Gustavsson (2004), we construct our observation matrix by 'stacking' all vectors for one subject onto each other, such that we obtain a $Cn_\mathrm{L}d \times S$ shape matrix $\mathbf{Y}$:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}^{11} & \cdots & \mathbf{y}^{1C} \\ \vdots & \ddots & \\ \mathbf{y}^{S1} & & \mathbf{y}^{SC} \end{bmatrix}. \tag{5}$$

This is the starting point for the construction of both the asymmetric and the symmetric bilinear shape models. Rather than using a 3-dimensional matrix $\mathcal{W}$, we use $\mathbf{W}$ which has a format similar to $\mathbf{Y}$, with the number of elements limited by the sizes of the parameterization vectors:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}^{11} & \cdots & \mathbf{w}^{1J} \\ \vdots & \ddots & \\ \mathbf{w}^{I1} & & \mathbf{w}^{IJ} \end{bmatrix}. \tag{6}$$

The modeling consists of minimizing the squared reconstruction error between the original observations and the approximation the model will provide. Denoting the approximation of $\mathbf{y}^{sc}$ in shorthand notation as

$$\hat{\mathbf{y}}^{sc} = \begin{cases} \mathbf{A}^s \mathbf{b}^c & \text{if the model is asymmetric} \\ \mathbf{a}^{s\mathrm{T}} \mathbf{W} \mathbf{b}^c & \text{if the model is symmetric} \end{cases}, \tag{7}$$

we minimize

$$E = \sum_{s=1}^{S} \sum_{c=1}^{C} \|\mathbf{y}^{sc} - \hat{\mathbf{y}}^{sc}\|^2. \tag{8}$$

2.3.2 Asymmetric Training

The training of an asymmetric model has a closed-form solution if the number of observations is (nearly) equally distributed over the style and content classes (Tenenbaum and Freeman 2000). As the matrix with training data $\mathbf{Y}$ is the result of the product $\mathbf{AB}$, it suffices to compute the Singular Value Decomposition (SVD) $\mathbf{Y} = \mathbf{USV}^\mathrm{T}$. Then, the matrix $\mathbf{B}$, containing the phase parameters, can be defined as the first J rows of $\mathbf{V}^\mathrm{T}$, while $\mathbf{A}$ will be defined as the first J columns of $\mathbf{US}$.

2.3.3 Symmetric Training

The training of the symmetric model requires the notion of the vector transpose (Marimont and Wandell 1992). Unlike the original application of this term, namely the conversion of a column vector into a row vector and vice versa, the vector transpose we use here is a matrix operation.

Given an $IK \times J$ matrix, where each column was constructed by stacking I K-dimensional column vectors onto

[1] In medical image analysis, 'landmarks' can refer to either anatomical landmarks or pseudolandmarks.

each other, the vector transpose of this matrix is a $JK \times I$ matrix, with the positions of the K-dimensional column vectors transposed rather than the individual elements. In the case where $K = 1$, the vector transpose is the normal transpose of the matrix. Otherwise, the vector transpose looks like

$$\mathbf{X}^{\mathrm{VT}} = \left[\begin{array}{c|c} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \hline x_{41} & x_{42} \\ x_{51} & x_{52} \\ x_{61} & x_{62} \\ \hline x_{71} & x_{72} \\ x_{81} & x_{82} \\ x_{91} & x_{92} \end{array}\right]^{\mathrm{VT}} = \left[\begin{array}{c|c|c} x_{11} & x_{41} & x_{71} \\ x_{21} & x_{51} & x_{81} \\ x_{31} & x_{61} & x_{91} \\ \hline x_{12} & x_{42} & x_{72} \\ x_{22} & x_{52} & x_{82} \\ x_{32} & x_{62} & x_{92} \end{array}\right] \tag{9}$$

for $K = 3$. It is easy to see from (9) that the vector transpose operation is invertible: $(\mathbf{X}^{\mathrm{VT}})^{\mathrm{VT}} \equiv \mathbf{X}$.

For the symmetric model, an iterative method is required to minimize E in (8). To this end, first $\mathbf{A}$ and $\mathbf{B}$ are computed. Upon convergence of these computations, $\mathbf{W}$ is computed using the results.

Analogously to the case of $K = 1$ using matrix transpositions, $\mathbf{Y} = \mathbf{AWB}$, the simplified model equation, can be rewritten as $\mathbf{Y} = (\mathbf{W}^{\mathrm{VT}}\mathbf{A})^{\mathrm{VT}}\mathbf{B}$ as well as $\mathbf{Y}^{\mathrm{VT}} = (\mathbf{WB})^{\mathrm{VT}}\mathbf{A}$. This leads to two equations that are familiar from the training of the asymmetric model:

$$(\mathbf{YB}^{\mathrm{T}})^{\mathrm{VT}} = \mathbf{W}^{\mathrm{VT}}\mathbf{A} = \mathbf{USV}^{\mathrm{T}} \tag{10}$$

and

$$(\mathbf{Y}^{\mathrm{VT}}\mathbf{A}^{\mathrm{T}})^{\mathrm{VT}} = \mathbf{WB} = \mathbf{USV}^{\mathrm{T}}. \tag{11}$$

By iterating over these equations, starting from an initial estimate of $\mathbf{B}$ using the SVD of $\mathbf{Y}$, convergence towards the real $\mathbf{A}$ and $\mathbf{B}$ is guaranteed (Magnus and Neudecker 1988). As was the case in the previous section, dimensionality reduction can be achieved by picking a fixed number of rows of $\mathbf{A}$ and $\mathbf{B}$. Unlike linear models based on PCA, this cannot be left until the end, as a change in dimensionality of one set of parameters will affect the other set of parameters. Therefore this truncation is done in each iteration. Upon convergence of the computation of $\mathbf{A}$ and $\mathbf{B}$, what is left is the computation of $\mathbf{W}$:

$$\mathbf{W} = ((\mathbf{YB}^{\mathrm{T}})^{\mathrm{VT}}\mathbf{A}^{\mathrm{T}})^{\mathrm{VT}}, \tag{12}$$

which satisfies $\mathbf{Y} = (\mathbf{W}^{\mathrm{VT}}\mathbf{A})^{\mathrm{VT}}\mathbf{B}$.

In Sect. 1 we already mentioned the relation between the bilinear models as they are formulated in Tenenbaum and Freeman (2000) and here, and tensor space decomposition (De Lathauwer et al. 2000). The two vector transpositions we use correspond to two of the three tensor unfoldings (De Lathauwer et al. 2000) of the third-order tensor we could construct along the axes of subject, phase and pseudolandmark coordinate.

3 Experiments

The assumption that shape variation introduced by inter-subject variation is independent from the dynamics is a simplification of reality. However, hearts with equal geometry at rest do not necessarily contract in the same way, while hearts with different geometry may. Factors such as myofiber orientation, local contractility defects like infarction, dilation or hypertrophy, and wall stress influenced by blood pressure and loading all influence the contraction pattern, and none of these are really independent of one another. Therefore we performed experiments to verify the suitability of bilinear modeling to capture the dynamics of the shape of the beating heart. To this end we constructed models factoring the inter-subject variations (as the style) and the dynamics (as the content), as illustrated in Fig. 1. Then, the models were used for extrapolating the learnt dynamics to subjects not present in the training set. In addition, we compared predictions based on a single phase to the difference between this shape and phases following it.

3.1 Data

3.1.1 Population

Our data consists of 80 full hearts acquired consecutively, with closed surface representations for each of five subparts, listed in Table 1 and visualized in Fig. 3. The population is distributed as 60% healthy and asymptomatic subjects, 20% subjects with Coronary Artery Disease (CAD) without a history of Myocardial Infarction (MI), and 20% subjects with CAD and a history of MI. Subject age was 58 ± 8 years, and men made up 56% of the population.

For the experiments we built models of the entire heart as well as models of only the right and left ventricle (biventricular model) and the LV only.

Table 1 The parts of the heart shape that have a closed surface representation, with the associated number of landmarks n_{L}. The parts are visualized in Fig. 3

Symbol	Description	n_{L}
LV	Left ventricular endo- and epicardium.	2677
RV	Right ventricle without septum, with trunk of pulmonary artery.	7902
LA	Left atrium with trunks of pulmonary veins.	6789
RA	Right atrium with trunks of venae cavae.	7243
AO	Trunk of the aorta.	3000

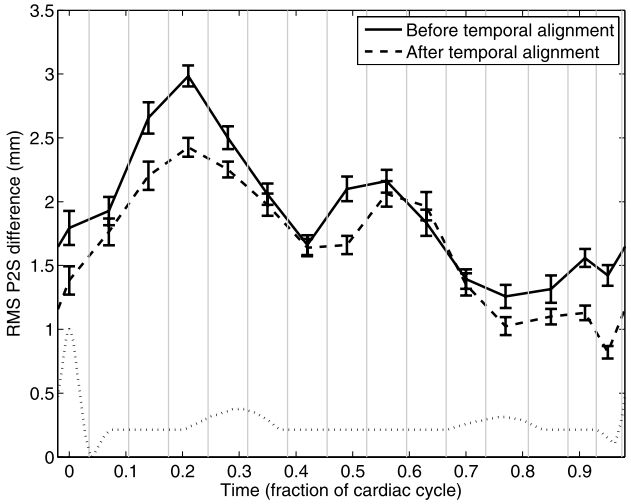

Fig. 2 RMS P2S differences between subsequent phases of the left ventricles, averaged over all 80 subjects, plus and minus one standard error. The *solid line* shows the differences before the temporal alignment step, while the *dashed line* shows the differences after this step (see Sect. 3.1.4 for details on this process). At time 0, the difference shown is that between phases 15 and 1. The *dotted curve* at the bottom represents a schematic ECG, the R-peak at the left being the reference for the ECG gating (all acquisitions are at R plus a fraction of the R-R interval). The *light grey vertical lines* indicate the interval over which each of the phases was sampled

3.1.2 Imaging Parameters

The imaging was performed using a Toshiba Aquilion 64 multi-slice Computed Tomography (CT) system (Toshiba Medical Systems, Tochigi, Japan) with a 64-row detector. Between 80 and 100 ml of contrast material (Xenetic 350) was administered at an injection rate of 5 ml/s. The rotation time of the scanner was, depending on the subject's heart rate, between 400 and 500 ms, and image reconstruction was performed on a Vitrea post-processing workstation (Vital Images, Minnetonka, MN, USA). The resulting dataset consisted of 15 image volumes (temporal phases) obtained using retrospective electrocardiographic (ECG) gating (Ohnesorge et al. 2002) with voxel dimensions of $0.4 \times 0.4 \times 2.0$ mm^3 per subject. Figure 2 shows the gating sequence over the cardiac cycle, albeit that the ECG used is schematic. Temporal relationships are only preserved with respect to the R-R interval.[2] Overlaid on this sequence are the Root Mean Square (RMS) Point-to-Surface (P2S) differences between subsequent full heart shapes, using a landmarking scheme as explained hereafter. The value at time 0 reflects the RMS P2S difference between phases 15 and 1.

[2]The R-R interval is the mean time between the R-peaks—the big peaks—in the ECG, indicating the firing of an electrical impulse to the heart muscle, which then contracts.

3.1.3 Landmarking

The data set was then used for the landmarking as presented in Ordás et al. (2007), resulting in $n_L = 27611$ points per volume. The point set sampled on the average shape was triangulated using Amira V3.0 (Visage Imaging, Carlsbad, CA, USA) to facilitate visualization of this shape, all shapes in the training set, and all shapes generated using the bilinear model. Table 1 lists the number of points sampled uniformly on each of the closed surfaces.

3.1.4 Spatial and Temporal Alignment

The subjects were imaged without sedation, which lead to heart rates of 62.2 ± 11.9 beats per minute. As the time it takes the heart to contract is largely independent of the heart rate, this causes inter-subject phase shifts due to the time points of the phases not being synchronized. Temporal alignment would therefore be desired. To this end, we explored using the method by Perperidis et al. (2005) for temporal alignment. For each phase, the normalized cross-correlation coefficient (CC) of image intensities between the phase at hand and the first phase in the cycle is computed. As the heart contracts, this scalar reduces as the images become less similar to the first frame. The phase of maximum contraction (end-systole) is identified as the phase with the minimum CC. Subsequently, the phase of maximum deceleration of relaxation (end-diastole) is identified as the phase with the minimum second derivative of the CC *after* the end-systolic phase. Together with the first and last phase, these form the phases which are to be aligned. All phases in between are to be interpolated.

We adapted this approach slightly. We did not use a reference cycle to identify at which phase end-systole and end-diastole should be placed. Instead, we used the mode of the set of identified phases for each, resulting in phase 5 for end-systole and phase 10 for end-diastole. It should be noted, however, that of the 80 subjects, the subjects for which the end-systolic phase was identified as the fifth only slightly outnumbered those for which the end-systolic phase was identified as the sixth. Figure 4 shows CC values as the temporal alignment procedure progresses.

It will become apparent, however, that the temporal alignment has its limitations. The spread of the heart rates, together with the low temporal resolution (15 phases, compared to 30 in Perperidis et al. (2005)), could result in a loss of information in some of the sequences due to linear interpolation, which may not be compensated for by the resulting correspondence. In some extreme cases, five phases used information from three original phases, while elsewhere in the same cycle, five phases were the result of interpolating between nine. Figure 5 shows how the phases identified as end-systolic and end-diastolic depended to some extent on

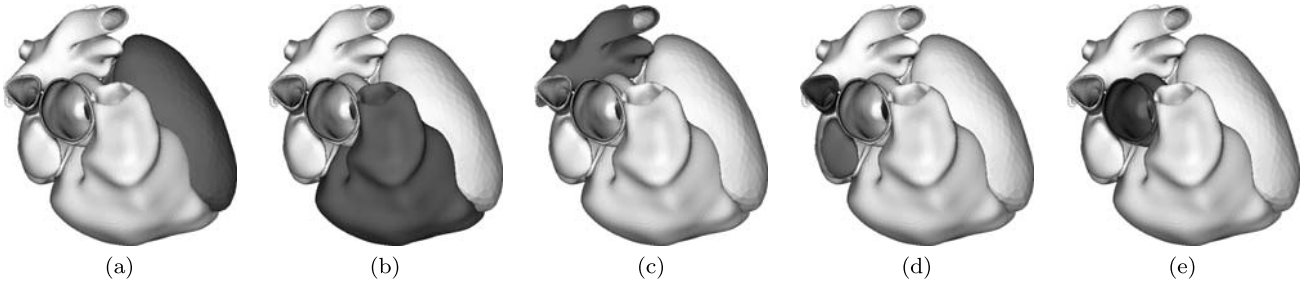

Fig. 3 Surface rendering of one of the training hearts. In dark (**a**) the left ventricle, (**b**) right ventricle, (**c**) left atrium, (**d**) right atrium, and (**e**) aorta

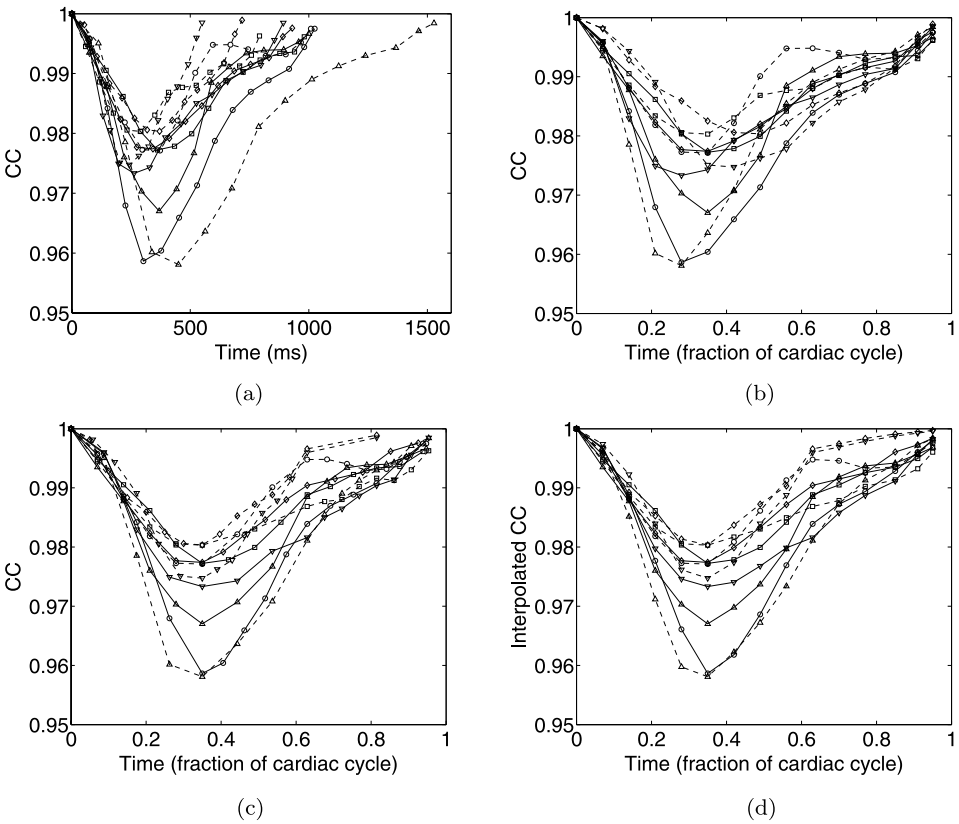

Fig. 4 The effect of the alignment on the CC of the CT images. For clarity, only values for 10 subjects are shown, including the subjects with shortest and longest R-R interval. The volume of the left ventricle follows a very similar curve (Perperidis et al. 2005). (**a**) CC against time in milliseconds before alignment. Note the spread along the temporal axis. (**b**) CC against time in percentage of the cardiac cycle before alignment. This corresponds to aligning the first phase of the cycle and the first phase of the next cycle. (**c**) CC against time in percentage of the aligned canonic cycle. Positions in time of the original acquisitions are interpolated. (**d**) Interpolated resampled CC after full alignment. In practice, the shapes were aligned by linear interpolation of corresponding landmarks

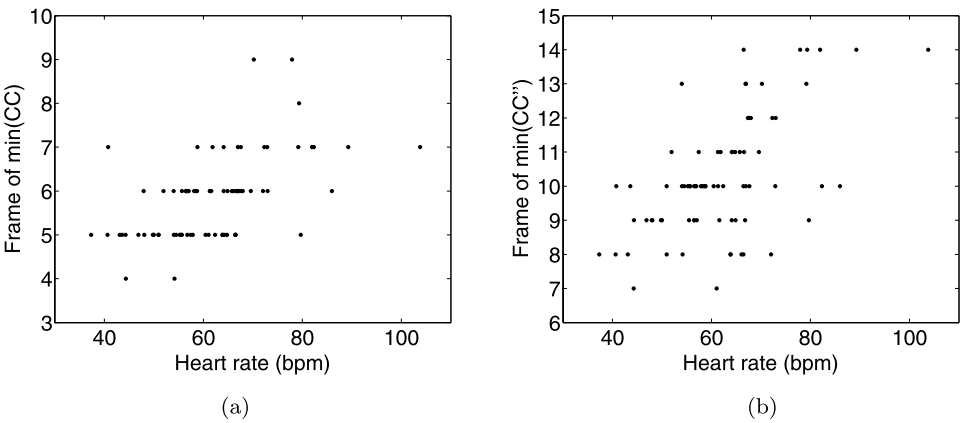

Fig. 5 Scatter plots of the relationship between heart rate and the phases in which (**a**) maximum contraction (minimum CC) and (**b**) maximum deceleration (minimum CC″) are observed. Note that a higher heart rate means a shorter R-R interval and thus the largely rate-independent duration of the beat itself taking up a larger part of this interval

the heart rate, but that even for equal heartrates a significant spread in identified phases may be observed.

Since the identification of the end-systolic and end-diastolic phases was based on the intensities of the original images, we extracted the left ventricle and the biventricular meshes at this point, as the result of temporally aligning those shapes separately would not be influenced.

After the temporal alignment, we also aligned the shapes spatially. With n_{L} landmarks describing each of our shapes, we denote the point set in $\mathbb{R}^3$ equivalent to shape vector $\mathbf{x}$ as the $n_{\mathrm{L}} \times 3$ matrix $\mathbf{X}$. The mean phase $\bar{\mathbf{X}}^{j}_{\mathrm{input}}$ of each subject j was translated to position its center of gravity at the origin. Then, it was rescaled to have unit norm. Subsequently, Procrustes alignment (Goodall 1991) was performed on these shapes, which, for each subject j, yielded a translation $\mathbf{t}_j$, rotation $\mathbf{R}_j$ and isotropic scaling s_j. Then, for each phase i,

$$\mathbf{X}^{ji}_{\mathrm{output}} = s_j \times \mathbf{X}^{ji}_{\mathrm{input}} \mathbf{R}_j + \mathbf{t}_j. \tag{13}$$

This way, the spatial relationships within the cycles were preserved. Removal of those relationships would reduce the overall variation observed in the data set, which in turn would lead to smaller reconstruction errors. By retaining the relationships instead, we possibly sacrificed some extrapolation accuracy in exchange for clinical meaningfulness of the results.

3.2 Experiments

We performed leave-n-out experiments with increasing training set sizes ($n = \{40, 20, 10, 5\}$). Thus, we divided our set of subjects into $\frac{80}{n}$ disjoint subsets, where each subset plays the role of test set once and forms part of the training set for the remainder of cases, as illustrated on the left side of Fig. 6. For each of the subjects from the test subset, five phases were used to derive the subject parameterization using the constructed model, shown at the top right in Fig. 6.

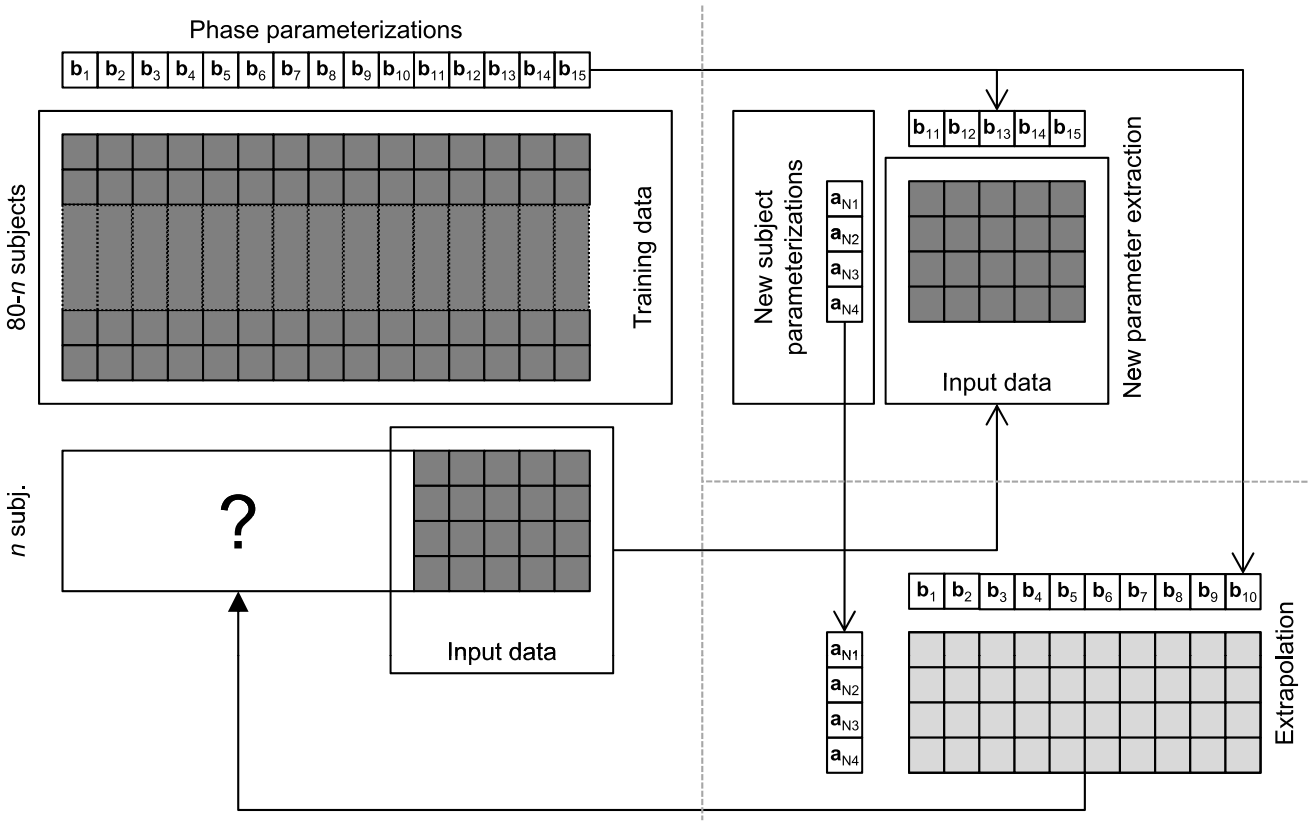

Fig. 6 Overview of the leave-n-out experiments. Dark grey blocks represent known shape data. White blocks with a $\mathbf{b}_{\#}$ and $\mathbf{a}_{\mathrm{N}\#}$ denote phase parameterizations from the model and new subject parameterizations, respectively. The subscript N is used to indicate that this is a newly derived parameterization. *Arrows* with open heads indicate that the elements on either side are the same, yet transported for clarity of the figure. Each of the $\frac{80}{n}$ disjoint sets of n subjects (*left side*) is used as a test set once, with the other sets combining to form the training data. For each of the subjects in the test set, a set of shapes, corresponding to an equally sized known set of phases (in the figure this is the set $\{11, \ldots, 15\}$), is used to derive the subject parameterizations $\mathbf{a}_{\mathrm{N}\#}$ (*top right*), using $\mathbf{W}$ (not in the figure) and the phase parameterizations from the model. The resulting subject parameterizations are then combined with $\mathbf{W}$ and the remaining phase parameterizations to extrapolate the shapes to the other phases for the test subjects (*bottom right*). The resulting shapes are represented in the figure by light grey blocks, and correspond to the block with the question mark (*left side*)

The remaining ten phases were then approximated (bottom right in Fig. 6) and the reconstruction errors recorded. Errors were grouped by phase and model dimensionality, and by phase set. The four sets of five phases covered systole, diastole, rest, and a combination of these (2 each from systole and diastole, 1 from rest), respectively.

3.3 Extrapolation

What we wish to do amounts to extrapolation, which is only one of several applications of bilinear models, as is shown in Tenenbaum and Freeman (2000). For a set of data $\mathbf{Y}_{\text{new}}$ defined over an incomplete set of content classes associated with $\mathbf{B}_{\text{inc}} \subset \mathbf{B}$, in a new style with unknown parameterization $\mathbf{a}_{\text{new}}$, the missing elements need to be reconstructed using the characteristics of those contents learned from the training data (other styles) and the style characteristics that are to be derived from the newly presented observations.

The models we built were symmetric models so that we could reduce the dimensionality of both sets of parameterizations to eliminate noise and improve robustness. Knowing which are the phases we have for deriving the new subjects' parameters, we used the phase parameters from the model associated with these known phases:

$$\mathbf{C}_1^{\text{VT}} = \mathbf{W}^{\text{VT}} \mathbf{B}_{\text{inc}}^{\text{T}}. \tag{14}$$

What remains is a linear system of equations

$$\mathbf{Y}_{\text{new}} = \mathbf{C}_1 \mathbf{a}_{\text{new}} \tag{15}$$

which, when solved, gives us a parameterization $\mathbf{a}_{\text{new}}$ of the new subject. This system has many more equations than unknowns (the ratio running into the tens of thousands). Therefore, a least squares approximation is found.

Since computing both the subject parameters $\mathbf{a}_j$ and phase parameters $\mathbf{b}_i$ in the models involves SVD's, the last column of both $\mathbf{A}$ and $\mathbf{B}$ should be a zero vector. In practice we have found that this is not always the case, possibly due to limitations in numerical precision. Additionally, the last few columns typically contain but noise as observed in the data. We used the dimensionality reduction as described in Sect. 2.3.3 to remove these columns. In the first iteration, we use the associated singular values to determine the number of dimensions required to capture at most 95% of the variance observed, both between subjects and between phases. The resulting dimensionalities are then fixed throughout the remainder of the training procedure.

3.4 Reconstruction

Using the resulting $\mathbf{a}_{\text{new}}$ from (15), we approximate the remaining set of phases $\mathbf{Y}_{\text{rem}}$ by

$$\mathbf{C}_2^{\text{VT}} = \mathbf{W}^{\text{VT}} \mathbf{B}_{\text{rem}}^{\text{T}}, \tag{16}$$

$$\hat{\mathbf{Y}}_{\text{rem}} = \mathbf{C}_2 \mathbf{a}_{\text{new}}, \tag{17}$$

with

$$(\mathbf{B}_{\text{rem}} \subset \mathbf{B}) \wedge (\mathbf{B}_{\text{rem}} \cap \mathbf{B}_{\text{inc}} = \emptyset). \tag{18}$$

The reconstruction error is then recorded as the RMS P2S error between each predicted shape $\hat{\mathbf{y}}_{\text{rem}_i}$ in $\hat{\mathbf{Y}}_{\text{rem}}$ and its corresponding ground truth shape:

$$\text{RMS}(\hat{\mathbf{y}}) = \sqrt{\frac{\sum_i d_i^2}{n_{\text{L}}}}, \tag{19}$$

where n_{L} again is the number of landmarks used to describe each of the shapes, and d_i is the distance from the i-th landmark of $\hat{\mathbf{y}}$ to the surface defined by the ground truth shape $\mathbf{y}$ of the corresponding phase and subject.

3.5 Comparison to Direct Approximation

In image sequence segmentation, it is not uncommon that the segmentation result of a phase is used as the initialization for the next phase. We believe that the bilinear models can provide an alternative in offering an initialization for a larger part of the sequence, or even the entire sequence, based on a (preliminary) result on the first phase. Subsequently the subject parameters would be optimized to complete the segmentation.

In order to compare the initialization for multiple phases based on one phase, we used an experiment similar to the one described in the previous sections. However, this time we extrapolated five phases from one, for each of the three cardiac periods: systole, diastole, and rest. We compare the extrapolation errors—initialization errors in the segmentation setting—to the errors we would obtain should we use the one phase directly for the initialization of these phases.

4 Results

4.1 Extrapolation

The results are presented for four sets of phases $\{1 \ldots 5\}$, $\{6 \ldots 10\}$, $\{11 \ldots 15\}$ and $\{4, 5, 9, 10, 15\}$ used for the derivation of a new subject's parameters and thus for the four groups of 10 reconstructed remaining phases. The results were produced using the temporally and spatially aligned data, unless stated otherwise.

The results of the LV models, the biventricular models and the full heart model results were highly correlated. Therefore we refrain from reporting the results on the biventricular and full heart models in the same detail as those from the LV models.

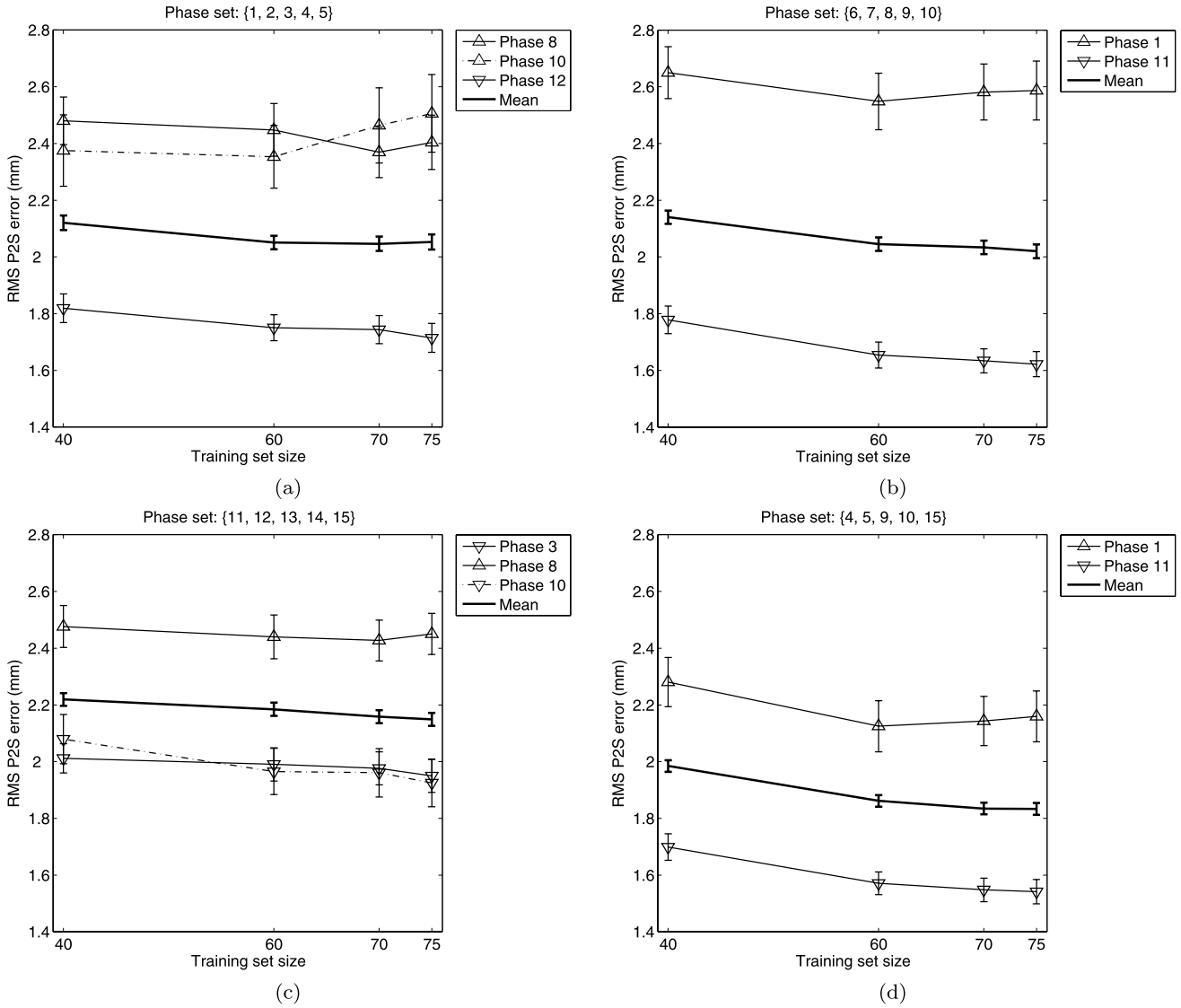

Fig. 7 The best, worst and mean RMS P2S reconstruction errors for the left ventricular model plus and minus one standard error in mm, after deriving the new subjects' parameters using (**a**) phases 1 through 5, (**b**) 6 through 10, (**c**) 11 through 15, and (**d**) phases 4, 5, 9, 10 and 15, using models built with increasing training set sizes

In Fig. 7, we present the mean RMS error of the worst and best reconstructed phases in the LV models, as well as the mean over all phases, against the size of the training sets (and consequently against the number of folds). It can be noted that larger training set sizes generally produced more accurate reconstructions. Additionally, the tendency of the curves suggests that some performance improvement could be achieved by further increasing the training set size. The errors are comparable to the differences between subsequent phases, reported in Fig. 2. Figure 8 shows the mean reconstruction errors and the standard errors for each set of phases used for extracting the subject parameters, for the biventricular and full heart models. The same tendency can be observed here, although it is more obvious in the results for the more complex structure.

From Fig. 7, it is possible to conclude that, in general, the worst and best reconstructed phases retain that status irrespective of the training set size. Thus, the quality of the reconstruction depends to some extent on the reconstructed phase, yet quite likely more on the combination of the reconstructed phase and the phases used for parameter extraction. The set of phases used for parameter extraction alone has some influence, but difference between the results for phase set 4 ({4, 5, 9, 10, 15}) and the other phase sets, presented in Figs. 7(d) and 8, would best be attributed to the spread of the phases instead of the phases themselves.

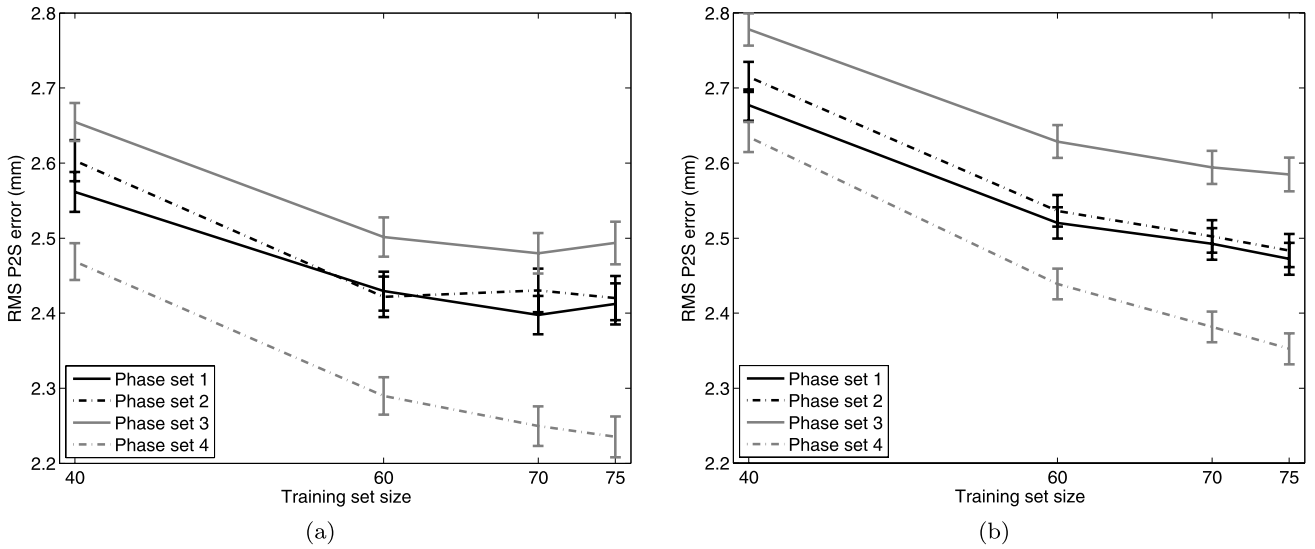

Fig. 8 The mean RMS P2S reconstruction errors for (**a**) the biventricular models and (**b**) the full heart models, plus and minus one standard error in mm, using models built with increasing training set sizes. Each line represents a set of phases used for the derivation of the new subjects' parameters, numbered in the order as they occur in Fig. 7

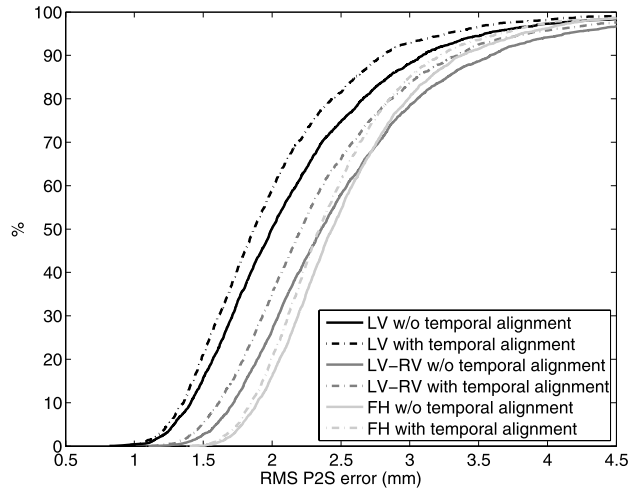

Fig. 9 The cumulative histogram of errors for the left ventricular (*black*), biventricular (*dark grey*) and full heart (*light grey*) models, using the data preprocessed without (*solid line*) and with (*dash-dot*) the temporal alignment step. Reconstructions included were all those obtained from the 16-fold experiments

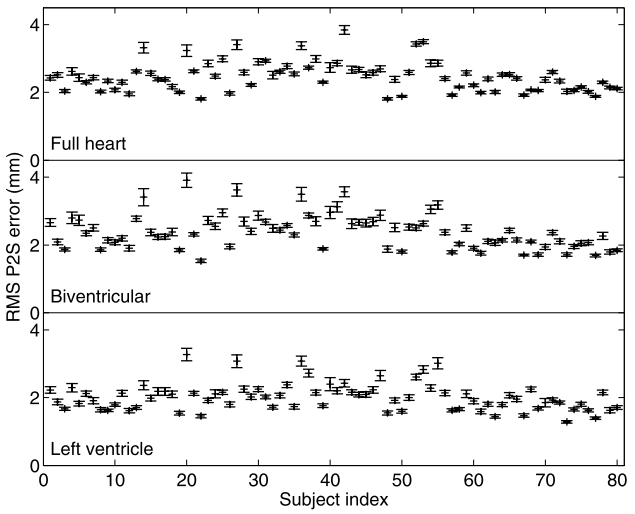

Fig. 10 The mean RMS P2S reconstruction error for each subject, plus and minus one standard error. It shows that some subjects have a worse reconstruction over the full range of structures and phases, whereas other subjects' shapes were relatively easy to approximate

Figure 9 shows the RMS P2S errors against the percentage of reconstructions, both for the data set that was aligned spatially and temporally, and for the set that was only aligned spatially. It gives an idea of the distribution of the error sizes. For each of the structures, the median error lies around 2 mm, while an error of 3 mm or more is well past the 80th percentile.

Another factor that has a certain degree of influence is which subjects' shapes are being reconstructed. Figure 10 shows the mean approximation error for each subject in the 16-fold experiments. The order of the subjects in the plot follows that of the folds. It is obvious that the shapes of certain subjects turn out to be reconstructed poorly at every phase, regardless of the phase set used for deriving the sub-

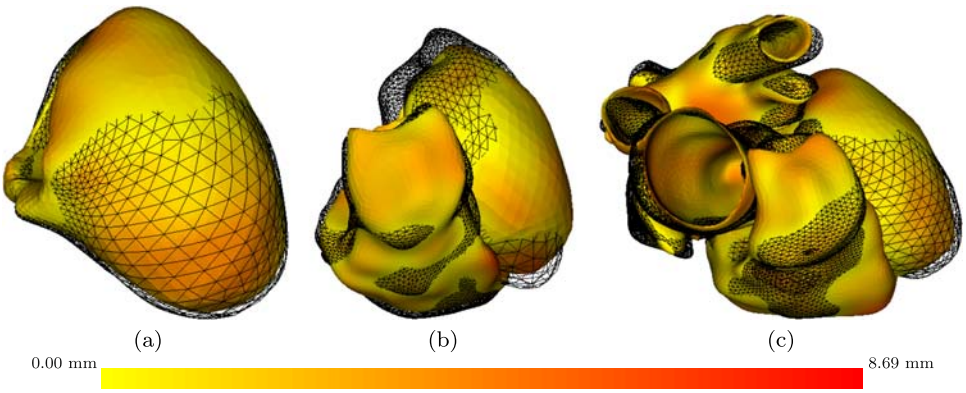

Fig. 11 Reconstructions of the shapes with the median RMS P2S errors, constructed using the highest dimensional models. (**a**) Left ventricle (1.869 mm), (**b**) biventricular (2.212 mm) and (**c**) full heart (2.344 mm). The wireframe shows the ground truth for these shapes. In (**b**) and (**c**), the difference in point density between the left ventricle and the other structures is clearly visible

ject parameters. Thus, the model did not generalize well to these subjects.

Finally, the local errors of the reconstructions are shown in Fig. 11. The errors are color-coded on the surface representations of the reconstructions with the median RMS P2S error, while the wireframe mesh shows the ground truth shape. Figure 11(c) shows that some error can accumulate in the most complex subparts, namely the atria, and especially in Fig. 11(b) it shows that the greater number of points on the right ventricle can result in lower accuracy for the left ventricle.

4.2 Comparison to Direct Approximation

The comparison between the approximation errors obtained using phases 1, 6 and 11 as approximations for the 5 following phases, and the approximation errors obtained when the subject parameterizations are derived from these same phases, each one separately, is presented in Fig. 12(a), (b) and (c) for the LV, biventricular and full heart model, respectively. In the set of phases approximated using the first phase, we can see that the bilinear approach clearly outperforms the direct approximation. This set of phases shows a dramatic change in shape as the heart goes from a state of rest to full contraction. The bilinear approach maintains the mean error within 3 mm. Less dramatic is the shape change in the second set of phases, approximated using phase 6. This is diastole, during which the heart does not return entirely to its full resting state. Recall from Sect. 3.1.4 that the end of diastole was identified as the phase of greatest deceleration; the heart is still relaxing after this point. In this group, the bilinear approach is more favored as the distance from the original phase increases. For the final set of phases, the direct use of phase 11 is favored, although the difference with the bilinear approach is not always significant.

In Fig. 12(d) we show how the two approaches compare overall, with respect to the number of phases separating the approximated phase and the phase used for the approximation or subject parameter extraction. It is clear to see that the big differences from the first set of phases dominate.

5 Discussion and Future Work

In the previous section we reported extrapolation errors obtained using bilinear models of the heart of varying complexity. We compared the performance of the models built from spatially and temporally aligned shape data with merely spatially aligned data, and we compared the extrapolation errors with the differences between the ground truth of phases, simulating the initialization for image sequence segmentation, which is our intended future use of these models.

Errors of up to 3 mm, as observed in both experiments, should be considered very acceptable, given that the approximated shapes were not seen before. This contrasts with the generalization ability measure commonly used to evaluate PCA-based shape models (Styner et al. 2003), where the shape to be approximated is known. When providing the initialization for image segmentation, it can offer a better initialization than using a mean shape or the segmentation result of the previous phase, and it can accumulate evidence to improve the initialization as one progresses through the sequence.

We confined ourselves to test data with phases matching those that the model was built with. As is demonstrated in Tenenbaum and Freeman (2000), this is not a prerequisite. We are keen to look into the performance of the extrapolation regarding both phase and subject, and to include this in the application of the models to image sequence segmentation.

Although we find the performance of the models very acceptable, there are certainly some limitations regarding both the modeling and the temporal alignment which may have influenced the results. For one, the acquisition scheme as shown in Fig. 2 is geared towards maximizing visibility of the coronary arteries. Therefore, phases 14 and 15 both have a short time span. The bulk of cardiac motion, however, lies just behind the QRS complex, resulting in the first five to six phases to be integrated over a time interval with larger spatial changes than the rest. A different temporal sectioning might therefore influence the results. More specifically,

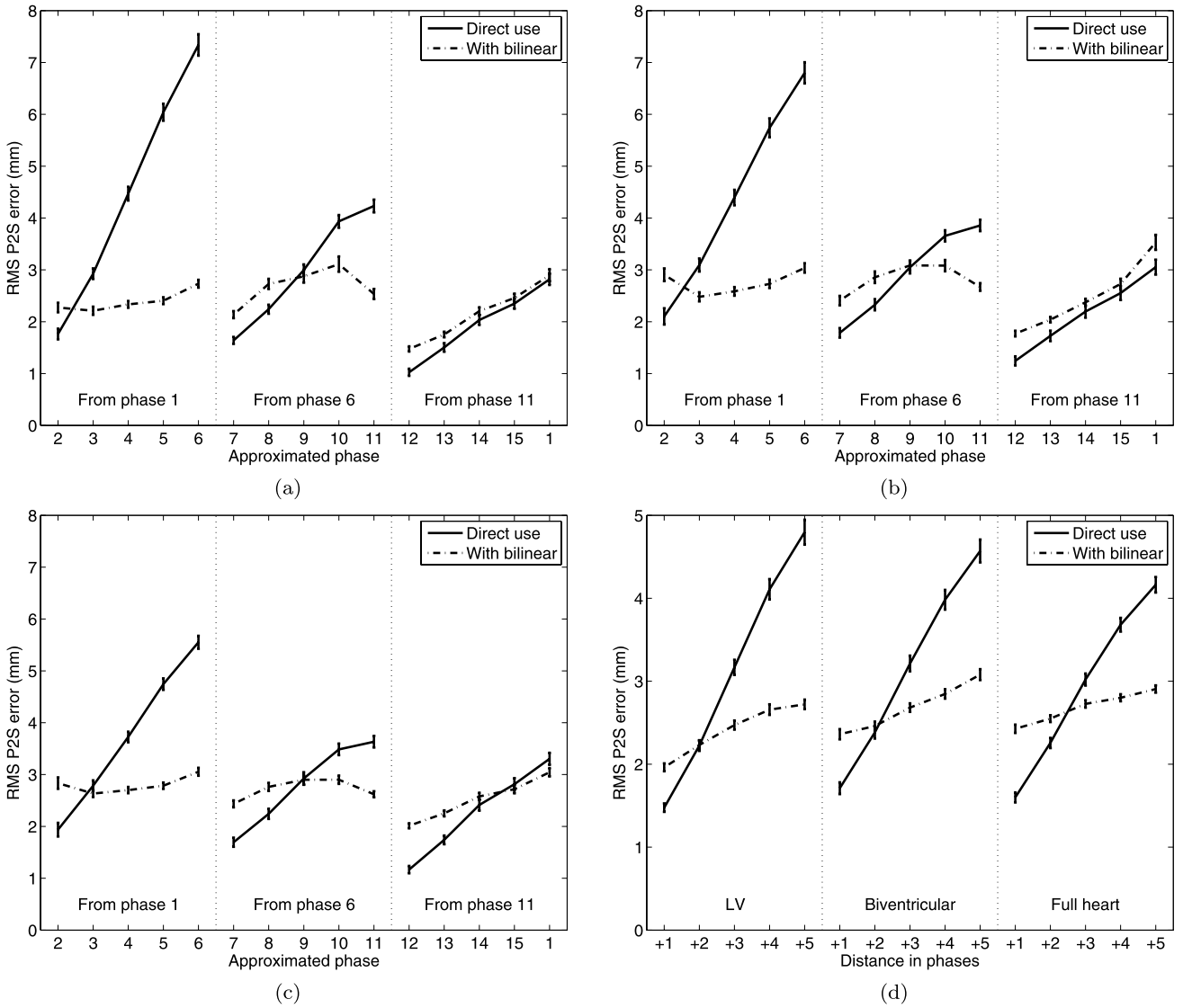

Fig. 12 The approximation errors plus and minus one standard error when deriving subject parameters from one phase only, for (**a**) the LV, (**b**) biventricular model and (**c**) the full heart model. These would be the initialization errors if used in a segmentation setting. Note that phase 1 is approximated using phase 11, as if it were the 16th phase. Plot (**d**) shows the errors per distance from the one phase that was used to derive the subject parameters, for each of the three model types: left ventricular, biventricular and full heart

higher frequency sampling should result in improved performance both without the temporal alignment as well as *of* the temporal alignment. Additionally, sedation of patients with higher heart rates would have reduced the spread in heart rate and the resulting mismatch of the phases (up to 5 phases for end-systole and up to 7 phases for end-diastole, as shown in Fig. 5) before the temporal alignment step.

Secondly, the poor approximations of certain subjects' shapes, as seen in Fig. 10, may be due to the fact that both training and testing data were mixes of healthy and pathological heart shapes. The ratios of these shapes (3:1:1) may have resulted in a bias towards the healthy hearts, resulting in poor derivations of the subjects' parameters in case of pathology. While it was out of the scope of this paper, whether the bilinear models are powerful enough to separate these groups may be a topic of further research. Also, whether the results would be better if model and test data consisted of only a single class—healthy, or one specific pathology—is an issue that warrants further looking into. When this point is resolved by patient selection, bilinear statistical shape models may provide a means to homogenize multiple acquisition protocols, which could greatly facilitate retrospective studies.

As the statistics in Table 2 and the reconstructions in Fig. 11 suggest, it is not a given that the more complex shape will always be more difficult to reconstruct. There seems to

Table 2 The main statistics for the reconstruction errors using the largest training set. Statistics were computed using all 3200 reconstructions for each model type

Stat	Median	Min	Max
Struct	With temporal alignment		
LV	1.869 mm	0.978 mm	7.224 mm
LV + RV	2.212 mm	1.084 mm	7.940 mm
FH	2.344 mm	1.380 mm	6.642 mm
	Without temporal alignment		
LV	1.999 mm	0.820 mm	8.557 mm
LV + RV	2.359 mm	1.102 mm	8.923 mm
FH	2.436 mm	1.401 mm	7.473 mm

be a tradeoff between complexity and the amount of information that can be extracted from the example shapes used for the derivation of the new subjects' parameters. However, considering the greater variability that can be observed in the atrial area, and especially around the trunks of the attached vessels and arteries, one would have expected the biventricular models to outperform the full heart model.

How to align sequences of shapes is a question that is not answered entirely. We believe to have taken the most straightforward approach by aligning the mean shapes of each cycle *after* the temporal alignment step, such that the mean would be computed over the same phases for each subject. It would deserve preference, however, to use a more unified approach to align the shapes both temporally and spatially at the same time. Nevertheless, Fig. 9 does show that employing the temporal alignment indeed improves the extrapolation performance.

We reported the RMS Point-to-Surface errors of the reconstructions. While these numbers were not unsatisfactory, they do not necessarily have a clinical meaning. Future experiments will focus on the prediction of clinical parameters such as Left Ventricular Volume and Ejection Fraction, Wall Thickening and Wall Motion, and classification of the dynamics based on the parameters acquired using the models.

Finally, the reconstruction errors could decrease further if the training set were larger, which can clearly be concluded from Figs. 7 and 8. Assuming that an appropriate number of training samples for PCA-based models of a complex shape, such as the human heart, easily runs into the triple digits (Ordás et al. 2007), there is room for improvement by increasing the number of training samples for the bilinear model as well.

6 Conclusion

We have shown in this work how to construct bilinear models of the human heart, and how to derive parameters for new subjects using these models and a limited amount of data from these new subjects. For the construction, simple SVD's and vector transpositions of the data matrix were employed to establish parameterizations for the training subjects and phases, after which a mixing matrix was computed by solving a linear system. The extraction of subject parameters was also reduced to a linear system.

Subsequently, we showed that with such parameters we could predict the shape of the heart over the previously unseen remaining two thirds of cardiac phases with a median RMS P2S error around 2 mm, and that 90% of predictions returned an error below 3.5 mm. Additionally the extraction of parameters from one phase and extrapolation over five following phases returned errors that compare favorably with the differences between subsequent phases. This suggests that a bilinear factorization of the heart shape may be appropriate to separate inter-subject variation from dynamics.

Acknowledgements The authors thank Dr. R. Leta and Dr. F. Carrera from the Cardiology Service, Hospital Santa Creu i Sant Pau, Barcelona, Spain, for the acquisition of the images. We also thank Dr. G. Avegliano of the Instituto Cardiovascular de Buenos Aires, Argentina, Prof. B.H. Bijnens and the anonymous reviewers for their useful comments.

References

Abboud, B., & Davoine, F. (2004). Bilinear factorization for facial expression analysis and synthesis. *IEE Proceedings—Vision, Image and Signal Processing*, *152*(3), 327–333.

Bistoquet, A., Oshinski, J., & Škrinjar, O. (2007). Left ventricular deformation recovery from cine MRI using an incompressible model. *IEEE Transactions on Medical Imaging*, *26*(9), 1136–1153.

Blackall, J. M., King, A. P., Penney, G. P., Adam, A., & Hawkes, D. J. (2001). A statistical model of respiratory motion and deformation of the liver. In W. J. Niessen & M. A. Viergever (Eds.), *Lecture notes in computer science: Vol. 2208. Proc. 4th int. conf. medical image computing and computer assisted intervention (MICCAI)*, Utrecht, The Netherlands (pp. 1338–1340). Berlin: Springer.

Bosch, J. G., Mitchell, S. C., Lelieveldt, B. P. F., Nijland, F., Kamp, O., Sonka, M., & Reiber, J. H. C. (2002). Automatic segmentation of echocardiographic sequences by active appearance motion models. *IEEE Transactions on Medical Imaging*, *21*(11), 1374–1383.

Chandrashekara, R., Rao, A., Sanchez-Ortiz, G. I., Mohiaddin, R. H., & Rueckert, D. (2003). Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In C. J. Taylor & J. A. Noble (Eds.), *Lecture notes in computer science: Vol. 2732. Proc. 18th int. conf. information processing in medical imaging (IPMI)*, Ambleside, United Kingdom (pp. 599–610). Berlin: Springer.

Chuang, E., & Bregler, C. (2005). Mood swings: Expressive speech animation. *ACM Transactions on Graphics*, *24*(2), 331–347.

Cootes, T. F., Cooper, D. H., Taylor, C. J., & Graham, J. (1992). Trainable method of parametric shape description. *Image & Vision Computing*, *10*(5), 289–294.

Cootes, T. F., Edwards, G. J., & Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *21*(6), 681–685.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active shape models—their training and application. *Computer Vision and Image Understanding*, *61*(1), 38–59.

Cuzzolin, F. (2006). Using bilinear models for view-invariant action and identity recognition. In *Proc. IEEE int. conf. on computer vision and pattern recognition (CVPR)*, New York, NY, USA (pp. 1701–1708) 2006.

De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, *21*(4), 1253–1278.

Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. New York: Wiley.

Duncan, J. S., & Ayache, N. (2000). Medical image analysis: progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(1), 85–106.

Frangi, A. F., Niessen, W. J., Viergever, M. A., & Lelieveldt, B. P. F. (2005). A survey of three-dimensional modeling techniques for quantitative functional analysis of cardiac images. In L. Landini, V. Positano, & M.F. Santarelli (Eds.), *Advanced image processing in magnetic resonance imaging* (Chap. 9, pp. 267–342). Boca Raton: CRC Press.

González-Mora, J., De la Torre, F., Murthi, R., Guil, N., & Zapata, E. L. (2007). Bilinear active appearance models. In R. Goecke, S. Lucey, & I. Matthews (Eds.), *Proc. int. workshop on non-rigid registration and tracking through learning*, Rio de Janeiro, Brazil, 2007.

Goodall, C. (1991). Procrustes methods in shape analysis. *Journal of the Royal Statistical Society—Series B: Statistical Methodology*, *53*(2), 285–339.

Grimes, D. B., & Rao, R. P. N. (2005). Bilinear sparse coding for invariant vision. *Neural Computation*, *17*(1), 47–73.

Hamarneh, G., & Gustavsson, T. (2004). Deformable spatio-temporal shape models: extending active shape models to 2D + time. *Image & Vision Computing*, *22*(6), 461–470.

Hoogendoorn, C., Sukno, F. M., Ordás, S., & Frangi, A. F. (2007). Bilinear models for spatio-temporal point distribution analysis: application to extrapolation of whole heart cardiac dynamics. In: M. Nielsen, W. Niessen & C. F. Westin (Eds.), *Proc. 8th int. workshop on mathematical methods in biomedical image analysis (MMBIA)*, Rio de Janeiro, Brazil, 2007.

Hsu, E., Pulli, K., & Popović, J. (2005). Style translation for human motion. *ACM Transactions on Graphics*, *23*(3), 1082–1089.

Kendall, D. G. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bulletin of the London Mathematical Society*, *16*(2), 81–121.

Le, H., & Kendall, D. G. (1993). The Riemannian structure of euclidean shape spaces: A novel environment for statistics. *Annals of Statistics*, *21*(3), 1225–1271.

Lee, C. S., & Elgammal, A. (2004). Gait style and gait content: Bilinear models for gait recognition using gait resampling. In *Proc. 6th IEEE int. conf. on automatic face and gesture recognition (FGR)*, Seoul, Korea (pp. 147–152) 2007.

Lekadir, K., Keenan, N., Pennell, D., & Yang, G. Z. (2007). Shape-based myocardial contractility analysis using multivariate outlier detection. In N. Ayache, S. Ourselin, & A. Maeder (Eds.), *Lecture notes in computer science: Vol. 4792. Proc. 10th int. conf. medical image computing and computer assisted intervention (MICCAI)*, Brisbane, QLD, Australia (pp. 834–841). Berlin: Springer.

Leung, K. Y. E., & Bosch, J. G. (2007). Localized shape variations for classifying wall motion in echocardiograms. In N. Ayache, S. Ourselin, & A. Maeder (Eds.), *Lecture notes in computer science: Vol. 4791. Proc. 10th int. conf. medical image computing and computer assisted intervention (MICCAI)*, Brisbane, QLD, Australia (pp. 52–59). Berlin: Springer.

Liu, H., & Shi, P. (2007). State-space analysis of cardiac motion with biomechanical constraints. *IEEE Transactions on Image Processing*, *16*(4), 901–916.

Lynch, M., Ghita, O., & Whelan, P. (2008). Segmentation of the left ventricle of the heart in 3D + t MRI data using an optimised non-rigid temporal model. *IEEE Transactions on Medical Imaging*, *27*(2), 195–203.

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York: Wiley.

Mardia, K. V., & Dryden, I. L. (1989). Shape distributions for landmark data. *Advances in Applied Probability*, *21*(4), 742–755.

Marimont, D. H., & Wandell, B. A. (1992). Linear models of surface and illuminant spectra. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *9*(11), 1905–1913.

McInerney, T., & Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, *1*(2), 91–108.

Mitchell, S. C., Bosch, J. G., Lelieveldt, B. P. F., van der Geest, R. J., Reiber, J. H. C., & Sonka, M. (2002). 3-D active appearance models: segmentation of cardiac MR and ultrasound images. *IEEE Transactions on Medical Imaging*, *21*(9), 1167–1178.

Montagnat, J., & Delingette, H. (2005). 4D deformable models with temporal constraints: application to 4D cardiac image segmentation. *Medical Image Analysis*, *9*(1), 87–100.

Ohnesorge, B. M., Becker, C. R., Flohr, T. G., & Reiser, M. F. (2002). *Multi-slice CT in cardiac imaging: technical principles, clinical application and future developments*. Berlin: Springer.

Ordás, S., Oubel, E., Leta, R., Carrera, F., & Frangi, A. F. (2007). A statistical shape model of the heart and its application to model-based segmentation. In A. Manduca & X. P. Hu (Eds.), *Proc. SPIE medical imaging*, San Diego, CA, USA, Vol. 6511.

Pantic, M., & Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1424–1445.

Perperidis, D. (2005). *Spatio-temporal registration and modelling of the heart using cardiovascular MR imaging*. Ph.D. thesis, Imperial College London.

Perperidis, D., Mohiaddin, R. H., & Rueckert, D. (2005). Spatio-temporal free-form registration of cardiac MR image sequences. *Medical Image Analysis*, *9*(5), 441–456.

Shin, D., Lee, H. S., & Kim, D. (2008). Illumination-robust face recognition using ridge regressive bilinear models. *Pattern Recognition Letters*, *29*(1), 49–58.

Styner, M. A., Rajamani, K. T., Nolte, L. P., Zsemliye, G., Székely, G., Taylor, C. J., & Davies, R. H. (2003). Evaluation of 3D correspondence methods for model building. In *Lecture notes in computer science: Vol. 2732. Proc. 18th int. conf. information processing in medical imaging (IPMI)*, Ambleside, United Kingdom (pp. 63–75). Berlin: Springer.

Suri, J. S. (2000). Computer vision, pattern recognition and image processing in left ventricle segmentation: the last 50 years. *Pattern Analysis & Applications*, *3*(3), 209–242.

Syeda-Mahmood, T., Wang, F., Beymer, D., London, M., & Reddy, R. (2007). Characterizing spatio-temporal patterns for disease discrimination in cardiac echo videos. In N. Ayache, S. Ourselin, A. Maeder (Eds.), *Lecture notes in computer science: Vol. 4791. Proc. 10th int. conf. medical image computing and computer assisted intervention (MICCAI)*, Brisbane, QLD, Australia (pp. 261–269). Berlin: Springer.

Tenenbaum, J. B., & Freeman, W. T. (1996). Separating style and content. In M. Mozer, M. I. Jordan & T. Petsche (Eds.), *Advances in neural information processing systems*, Denver, CO, USA (pp. 662–668) 1996.

Tenenbaum, J. B., & Freeman, W. T. (2000). Separating style and content with bilinear models. *Neural Computation*, *12*(6), 1247–1283.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*(1), 71–86.

Vasilescu, M. A. O., & Terzopoulos, D. (2003). Multilinear subspace analysis of image ensembles. In *Proc. IEEE int. conf. on computer vision and pattern recognition (CVPR)*, Madison, WI, USA (pp. II: 93–99), 2006.