



A multimodal spatiotemporal cardiac motion atlas from MR and ultrasound data



Esther Puyol-Antón^{a,*}, Matthew Sinclair^a, Bernhard Gerber^c, Mihaela Silvia Amzulescu^c,
Hélène Langet^{b,c}, Mathieu De Craene^b, Paul Aljabar^a, Paolo Piro^b, Andrew P. King^a

^a Division of Imaging Sciences and Biomedical Engineering, Kings College London, U.K

^b Philips Research, Medisys, Paris, France

^c Division of Cardiology, Cliniques Universitaires St-Luc, Avenue Hippocrate 10-2881, Brussels B-1200, Belgium

ARTICLE INFO

Article history:

Received 14 October 2016

Revised 30 May 2017

Accepted 6 June 2017

Available online 13 June 2017

Keywords:

Cardiac motion atlas

Multi-view dimensionality reduction

ABSTRACT

Cardiac motion atlases provide a space of reference in which the motions of a cohort of subjects can be directly compared. Motion atlases can be used to learn descriptors that are linked to different pathologies and which can subsequently be used for diagnosis. To date, all such atlases have been formed and applied using data from the same modality. In this work we propose a framework to build a multimodal cardiac motion atlas from 3D magnetic resonance (MR) and 3D ultrasound (US) data. Such an atlas will benefit from the complementary motion features derived from the two modalities, and furthermore, it could be applied in clinics to detect cardiovascular disease using US data alone.

The processing pipeline for the formation of the multimodal motion atlas initially involves spatial and temporal normalisation of subjects' cardiac geometry and motion. This step was accomplished following a similar pipeline to that proposed for single modality atlas formation. The main novelty of this paper lies in the use of a multi-view algorithm to simultaneously reduce the dimensionality of both the MR and US derived motion data in order to find a common space between both modalities to model their variability. Three different dimensionality reduction algorithms were investigated: principal component analysis, canonical correlation analysis and partial least squares regression (PLS).

A leave-one-out cross validation on a multimodal data set of 50 volunteers was employed to quantify the accuracy of the three algorithms. Results show that PLS resulted in the lowest errors, with a reconstruction error of less than 2.3 mm for MR-derived motion data, and less than 2.5 mm for US-derived motion data. In addition, 1000 subjects from the UK Biobank database were used to build a large scale monomodal data set for a systematic validation of the proposed algorithms. Our results demonstrate the feasibility of using US data alone to analyse cardiac function based on a multimodal motion atlas.

© 2017 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY license. (<http://creativecommons.org/licenses/by/4.0/>)

1. Introduction

Better understanding the motion of the heart through the cardiac cycle is crucial in aiding diagnosis in a wide range of cardiovascular pathologies as well as for the optimisation of related clinical procedures. In recent years, statistical cardiac atlases have been proposed as a way of analysing cardiac motion in the context of population variation. However, to date, all such techniques have been based on motion information estimated from a single imaging modality. Several factors suggest that information from multiple modalities may be complementary and offer advantages.

* Corresponding author.

E-mail addresses: esther.puyol_anton@kcl.ac.uk, str160290@gmail.com (E. Puyol-Antón).

For instance, the spatial and/or temporal resolutions of the modalities may differ, or motion tracking algorithms may exploit different imaging characteristics such as speckle in ultrasound (US). Therefore, using the combined information may provide a richer description of cardiac function. Moreover, a multimodal atlas would open up the possibility of applying the atlas using a single, low cost, modality such as US, whilst exploiting the higher quality, but higher cost, of another modality such as magnetic resonance (MR).

In this paper we demonstrate and evaluate a framework for forming such an atlas from MR and US data. To form the atlas we first perform a spatial and temporal normalisation of subjects' cardiac geometry and motion, and then we apply a dimensionality reduction algorithm to find a common space between both modalities to extract meaningful descriptors and model the intermodality variability.

In this introduction, we first review the literature on single modality statistical cardiac motion atlases. Then, we review techniques commonly used to subsequently reduce the dimensionality of the high dimensional motion data for further analysis. Next, we review common approaches for statistical analysis. Then, we review some techniques proposed for multimodal data fusion. Finally, we briefly outline our approach in order to highlight our key contributions.

1.1. Single modality motion atlases

A statistical motion atlas provides a space in which the motions of a cohort of subjects can be directly compared. Statistical cardiac atlases have been proposed for characterising abnormal cardiac motion, as well as for detecting suspected cardiac disease as early as possible. Typically, such atlases have focused on the left ventricle (LV), as it is the chamber primarily investigated for diagnosing cardiac diseases. The shape and motion of the LV have been estimated from imaging data acquired using MR (Chandrasekara et al., 2003; Rougon et al., 2004; Perperidis et al., 2005a; Ardekani et al., 2009; Lu et al., 2009; Garcia-Barnes et al., 2010; De Craene et al., 2012; Medrano-Gracia et al., 2014a; Bai et al., 2015b; Peressutti et al., 2017), US (Duchateau et al., 2010; 2011) or Computed Tomography (Hoogendoorn et al., 2013). To date, most atlases have been built from less than 100 subjects (Chandrasekara et al., 2003; Rougon et al., 2004; Perperidis et al., 2005a; Duchateau et al., 2010; 2011; De Craene et al., 2012; Lu et al., 2009; Garcia-Barnes et al., 2010; Peressutti et al., 2017). In a few cases, larger databases have been used to construct motion atlases, such as the Cardiac Atlas Project that contains 3000 subjects, consisting of 2864 asymptomatic volunteers from the MESA cohort and 470 patients with myocardial infarction from the DETERMINE cohort (Fonseca et al., 2011; Medrano-Gracia et al., 2013b; 2014a) and the Hamersmith data set (Bai et al., 2015b) that contains 1093 healthy subjects.

In general, motion atlas construction entails the following steps: (1) geometry estimation, commonly achieved by segmenting the anatomy of interest and generating a 3D mesh that will represent it; (2) motion estimation using a motion tracking algorithm; (3) temporal normalisation that aims to ensure temporal correspondence across the cardiac cycle; (4) spatial normalisation that aims to remove bias towards subject-specific geometries; and (5) motion reorientation, which transports motion estimates from subject-specific to atlas coordinate systems.

Although many authors have broadly followed this pipeline, their works differ in the techniques used to estimate the LV geometry and motion, and to represent the motion. For example, Medrano-Gracia et al. (2015); 2013b) fitted a finite-element model to cine-MR sequences to estimate the LV geometry, whilst Sinclair et al. (2016) and Peressutti et al. (2017) employed a statistical shape model of the LV to enforce point correspondence amongst all LV geometries. Furthermore, the different motion tracking algorithms proposed can be divided into: generic methods that can be applied to any image modality (Rueckert et al., 1999; Mansi et al., 2011; Rougon et al., 2004); and modality-specific methods that try to take advantage of the nature of the data to estimate the motion, such as tagged MR-specific algorithms (Zhou et al., 2015; Arts et al., 2010), and speckle-tracking algorithms for US motion estimation (Crosby et al., 2009; Kaluzynski et al., 2001). Temporal normalisation has been achieved by landmark-based piecewise linear warping of cardiac timings (Rougon et al., 2004; Perperidis et al., 2005b; De Craene et al., 2012) or by normalising by the length of the cardiac cycle (Sinclair et al., 2016; Peressutti et al., 2017). Spatial normalisation has been previously achieved using a combination of global and local transformations (De Craene et al., 2012; Peressutti et al., 2017). First, a global transformation addresses spa-

tial differences caused by size, orientation and translation of the LV geometries, then a local transformation addresses the differences in local shape. Some authors have combined temporal normalisation with spatial normalisation to directly compute spatiotemporal deformations. For example, Perperidis et al. (2005b) used a spatiotemporal registration method that finds the temporal and spatial transformation components at the same time using image information only. Durrleman et al. (2013) extended the piecewise linear regression for shape time-series by using piecewise geodesic interpolation. Finally, reorienting motion from subject-specific coordinate systems to the atlas coordinate system is achieved by parallel transport of vector fields and tensors (Rao et al., 2004; Qiu et al., 2009; Lorenzi et al., 2011).

1.2. Motion representation and dimensionality reduction

Cardiac motion can be represented in a spatiotemporal cardiac atlas in a number of different ways, including intensity-based (Lu et al., 2009), displacement-based (Chandrasekara et al., 2003; Perperidis et al., 2005a; Garcia-Barnes et al., 2010; Bai et al., 2015b), velocity-based (Duchateau et al., 2011; 2012), and momentum-based (Ardekani et al., 2009) approaches.

In many cases, the motion representation in the atlas space is very high dimensional, and in order to be able to obtain meaningful descriptors, many authors have proposed the use of dimensionality reduction algorithms as a pre-processing step. Techniques proposed include principal component analysis (PCA) (Rougon et al., 2004; Perperidis et al., 2005a; Medrano-Gracia et al., 2010; Hoogendoorn et al., 2013; McLeod et al., 2013), independent component analysis (Suinesiaputra et al., 2009; Medrano-Gracia et al., 2010), manifold learning (Duchateau et al., 2012), and bilinear spatiotemporal models (Hoogendoorn et al., 2009; Akhter et al., 2012), in which the inter-subject variation and inter-phase variation are encoded separately in two dimensions. Usually, pattern analysis and classification follow this pre-processing step.

Alternatively, some methods focus on a regional heart motion analysis, for example by dividing the LV myocardium into 17 standard segments according to the American Heart Association (AHA) delineation, then performing an analysis of the motion within each segment (Rougon et al., 2004; Suinesiaputra et al., 2009; Punithakumar et al., 2013). Based on the AHA delineation, McLeod et al. (2012); 2013; 2015) proposed a polyaffine model to represent the 4D LV motion using a low number of parameters. An advantage of dividing the myocardium into segments in this way is that the AHA delineation is widely understood by clinicians. However, coronary perfusion territories do not follow the AHA segment delineations so they may not be the optimal representation for diagnostic purposes. Motivated by this, Bai et al. (2015a) proposed an alternative parcellation of the LV based on motion information.

1.3. Statistical analysis

Based on the motion descriptors extracted, different statistical methods have been used to analyse differences between groups. For example, Duchateau et al. (2010); 2011) introduced a framework to assess septal flash, which denotes an early septal inward/outward motion. From the motion atlas, they calculated the Hotelling's T-square distance between the velocity distributions of the atlas population and each individual. A higher value of abnormality corresponded to a lower p-value. Later, De Craene et al. (2012) used statistical parametric mapping (SPM) on the velocity field obtained from the motion atlas to quantify motion abnormality. More recently, Bai et al. (2015b) applied SPM analysis on a larger database (1000 subjects) to study the impact of gender and age on regional myocardial wall thickness.

In the multimodal domain, [Medrano-Gracia et al. \(2013a\)](#) proposed an atlas-based method to correct shape bias between different MR sequence acquisitions. They showed that after applying bias correction they can improve the detection of statistical differences in regional shape and motion between cohorts.

1.4. Multimodal feature fusion

In various disciplines, information about the same phenomenon can be acquired from different views (e.g. types of detectors, different conditions, multiple experiments or subjects, etc.). Combining (or fusing) the information from the multiple views can help to exploit their natural strengths and reduce redundancies. Therefore, multimodal data fusion has gained considerable attention in the literature in recent years. When applied in a machine learning context, this concept is known as multi-view learning. In general, multi-view learning algorithms can be categorised into ([Xu et al., 2013](#)): (1) co-training methods, which try to find mutual agreement on two distinct views of the data by alternately training a set of classifiers defined in each view; (2) multiple kernel learning, which uses a predefined set of kernels for each view, and combines them either linearly or non-linearly to improve learning performance; and (3) subspace learning, which aims to obtain a low-dimensional latent subspace shared by multiple views by assuming that the input views were generated from this subspace. In the context of this paper, the main motivation for multimodal feature fusion is to find a latent subspace shared by 3D MR and 3D US data to facilitate a more robust characterisation of cardiac motion using low cost 3D US data. Therefore, we focus on multi-view subspace learning here.

Multi-view subspace learning techniques can be further subdivided into: supervised techniques that require full knowledge of the correspondences between the data from the different views ([Hotelling, 1936](#); [Wold et al., 1984](#)); semi-supervised techniques that require a subset of such correspondences to be known ([Ham et al., 2005](#); [Georg et al., 2008](#)); and lastly, unsupervised techniques that do not require such knowledge ([Wachinger et al., 2012](#); [Mateus et al., 2007](#)). Note that there is a close link between multi-view subspace learning and manifold alignment, and often the two terms can be considered as synonyms, although typically manifold alignment approaches have tended to be semi-supervised or unsupervised (e.g. [Ham et al. \(2005\)](#)).

Relevant literature in the medical field includes [Guerrero et al. \(2014\)](#), who adapted the Laplacian eigenmaps method to learn a joint low-dimensional representation of brain MR images acquired either at 1.5 and 3 Tesla, and [Lorenzi et al. \(2016\)](#), who used partial least squares to analyse the joint variation between genotype and phenotype in Alzheimer's disease patients.

1.5. Our contributions

In this paper, we present a framework for building a spatiotemporal atlas from imaging data acquired from a cohort of subjects using two different modalities (MR and US), but to apply it to a new subject using only US. The technique works by applying a multi-view subspace learning algorithm ([Sun, 2013](#); [Xu et al., 2013](#)) to simultaneously reduce the dimensionality of the two modalities into a common latent subspace. This is the first time that such a technique has been proposed in the area of atlas formation. We evaluate our approach on a large scale monomodal data set based on 1000 subjects from the UK Biobank database, and also MR and US data acquired from a cohort of 50 volunteers, and use the atlas to analyse the dominant motion patterns and the differences between motion data estimated using the two modalities.

We have previously presented a feasibility study on this work using a 'naive' single view learning approach (PCA) ([Puyol-Antón](#)

[et al., 2016](#)). We now extend this work to evaluate an improved temporal normalisation scheme that aligns specific cardiac timings and also evaluate two different multi-view subspace learning algorithms: partial least squares regression (PLS) and canonical correlation analysis (CCA). We compare these multi-view approaches to PCA.

The remainder of this paper is organised as follows. In [Section 2](#), we detail the acquisition protocols of the large scale monomodal data set and the real data set. In [Section 3](#) we describe the framework used to build the multimodal spatiotemporal cardiac motion atlas, including a description of the different multi-view learning algorithms employed. [Section 4](#) is devoted to experiments on the large scale monomodal data set and the real data set. Finally, [Section 5](#) discusses the findings of this paper in the context of the literature and suggests potential improvements for future work.

2. Materials

We now describe the data used in our experiments. These consist of a large scale monomodal data set and a real multimodal data set.

2.1. Large scale monomodal data set

For systematic validation and understanding of the robustness of the proposed dimensionality reduction algorithms, we used 1000 subjects from the UK Biobank cohort.¹ From this data set, we used the multi-slice short-axis cine MR sequence (TR/ TE = 2.7/1.16 ms, flip angle = 80°). Typical slice thickness was 8.0 mm with an in-plane resolution 1.8 mm × 1.8 mm. The temporal resolution for all subjects was 50 frames per cycle. Further acquisition parameters can be found in [Petersen et al. \(2016\)](#). Details of how we used these data to assess the dimensionality reduction algorithms can be found in [Section 4.2](#).

2.2. Real multimodal data set

Three real MR and US data sets were used for evaluation in this paper. The first is the database used for the cardiac motion analysis challenge that was held at the 2011 MICCAI workshop "Statistical Atlases and Computational Models of the Heart: Imaging and Modelling Challenges" (STACOM'11) ([Tobon-Gomez et al., 2013](#)). The STACOM'11 database includes MR and US data sets from 15 healthy volunteers acquired at the Division of Imaging Sciences and Biomedical Engineering, King's College London, United Kingdom, and the Department of Internal Medicine II - Cardiology, University of Ulm, Germany.

The second data set was recently acquired at the Division of Imaging Sciences and Biomedical Engineering, King's College London, United Kingdom, and contains 9 MR and US data sets acquired from healthy volunteers.

The third data set was acquired at the Division of Cardiology, Cliniques Universitaires St-Luc, Avenue Hippocrate 10-2881, B-1200 Brussels, Belgium. This contains MR and US data acquired from 26 healthy volunteers.

For all cases the MR acquisitions were performed using a 3T Philips Achieva System (Philips Healthcare, Best, The Netherlands), and the US data sets were acquired using an iE33 3D echocardiography system (Philips Medical Systems, Bothell, WA, United States) with a 1–5 MHz transthoracic matrix array transducer (xMATRIX X5.1). Full-volume acquisition mode was used in which several smaller imaging sectors acquired over multiple cardiac cycles are combined to form a large composite volume.

¹ <https://www.ukbiobank.ac.uk/about-biobank-uk>.

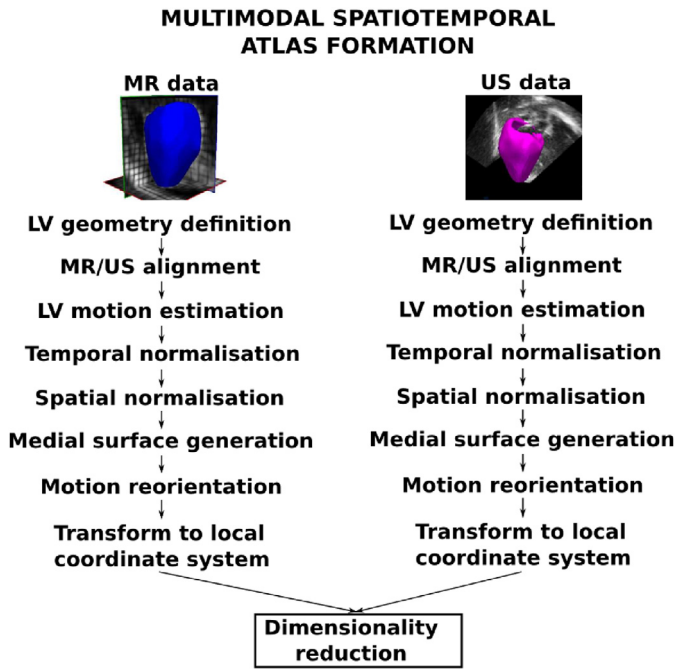


Fig. 1. Overview of the proposed framework for spatiotemporal cardiac motion atlas formation.

In particular, the three data sets contain for each subject:

- **cine SA:** a multi-slice short-axis cine-MR sequence (TR/ TE = 3.0/1.5 ms, flip angle = 60°). Typical slice thickness was between 8.0 and 10.0 mm with an in-plane resolution ≈ 1.0 mm $\times$ 1.0 mm. Typical temporal resolution was between 25 and 30 frames per cycle. In our framework, the ED frame only of the cine SA data was used for geometry estimation (see Section 3.1). The cine SA sequence was not used for motion estimation.
- **TAG:** a 3D tagged MR sequence in three orthogonal directions (TR/ TE = 7.5/3.2 ms, flip angle = 19°, tag distance between 7.7 and 8.8mm. The images have reduced field-of-view enclosing the LV, with typical isotropic 3D spatial resolution between 2.5 mm and 1.1 mm, and typical temporal resolution between 22 and 30 frames per cycle. The TAG data was used for motion estimation (see Section 3.3).
- **US:** a 3D apical ultrasound sequence. Typical slice thickness was between 0.7 mm and 1.0 mm with an in-plane resolution ≈ 0.65 mm $\times$ 0.8 mm. Typical temporal resolution was between 15 and 23 frames per cycle. The US data was used for geometry and motion estimation (see Sections 3.1 and 3.3).

The three real data sets described above were combined to form a single data set for the experiments described in this paper.

3. Methods

We now briefly outline the procedure used for multimodal spatiotemporal atlas formation, which is based on the framework described in Puyol-Antón et al. (2016) and Peressutti et al. (2017). The proposed framework is illustrated in Fig. 1.

3.1. LV geometry definition

For each subject, the LV myocardium was manually segmented in the end-diastolic (ED) cine SA and US images using ITK-SNAP (Yushkevich et al., 2006). An open-source statistical shape model (SSM) was optimised to fit to the LV binary mask following an

initial rigid registration using anatomical landmarks (Bai et al., 2015b). The use of landmarks ensured that the LV mesh was aligned to the same anatomical features for each patient, and the endocardial and epicardial surfaces of the mesh were smoothly fitted to the myocardial mask, providing surface meshes with point correspondence for each of the subjects and a 17 AHA regions segmentation. An example of a mesh fitted to cine SA and US images is illustrated in Fig. 2, visualised using Eidolon² (Kerfoot et al., 2016).

3.2. MR/US alignment

In order to apply the mesh generated from cine SA to the TAG sequence, we rigidly aligned the cine SA and TAG sequences. To achieve this, we first generated a detagged MR volume from the TAG sequence using Eidolon (Kerfoot et al., 2016), and then we used an intensity-based rigid registration (Studholme et al., 1999) to align the detagged and cine SA sequences at ED. The rigid registration algorithm was implemented in the Image Registration Toolkit (IRTK)³.

Furthermore, in order to compare US and MR-based motion estimation, the US and MR images should be represented in a common space. To achieve this, a Procrustes alignment (Gower and Dijksterhuis, 2004) was applied to align LV mesh points derived from each of the two modalities. The AHA segmentations generated in both meshes were used to ensure that the midseptum was in the same location in both modalities. An example of registered MR/US images is shown in Fig. 2.

3.3. LV motion estimation

A 3D GPU-based B-spline free-form deformation (FFD) registration was used (Rueckert et al., 1999) to estimate LV motion between consecutive frames of the MR and US sequences. As described in Sinclair et al. (2016) and Peressutti et al. (2017), the sequentially estimated FFD deformation fields were composed to obtain the FFD transformation from each cardiac phase to the reference ED phase. Furthermore, a 3D+t cyclic B-spline was fitted to the composed 3D transformations in order to estimate a full cycle 3D+t transformation (Unser, 1999) using the IRTK software. The result of the motion estimation was, for each subject n , a smooth 3D+t transformation $\psi^{3D+t}(\mathbf{x}, t)$, where $\mathbf{x}$ represents the spatial coordinates of the vertices and t is the cardiac phase.

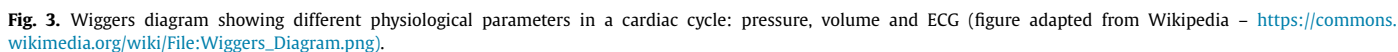
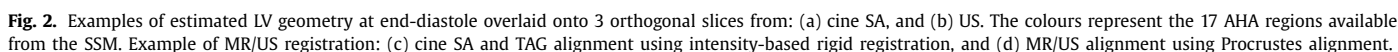
3.4. Temporal normalisation

Temporal normalisation aims to ensure temporal correspondence between subjects. In cardiology, corresponding time-points can be defined as the timings of specific cardiac physiological states, e.g. the end of diastole/systole for a given ventricle (see Fig. 3). In this paper, in contrast to Puyol-Antón et al. (2016) and Peressutti et al. (2017), we use a temporal normalisation scheme that aligns specific cardiac timings across subjects, rather than normalising only with respect to the length of the cardiac cycle. We believe that establishing correspondences between the cardiac timings of the subjects and bringing them to a normalised time-scale will reduce the inter-subject and inter-modality variability and improve the accuracy of dimensionality reduction algorithms.

We automatically identified the following cardiac timings based on volume curves computed using the estimated geometries and motions for all subjects (note that the temporal normalisation step is applied separately to the MR and US data).

² Eidolon URL: <https://github.com/ericspod/Eidolon>.

³ IRTK URL: <https://github.com/BioMedIA/IRTK>.



- **Beginning of the isovolumic contraction phase:** The point at ED at which the volume in the cavity is maximised - the end diastolic volume (EDV). This cardiac timing corresponds to the label “1” in Fig. 3. For both MR and US sequences, the ED frame corresponds to the first frame of the sequence.
- **Beginning of the isovolumetric relaxation phase:** The point at the end of systole (ES) at which the volume in the cavity is minimised - the end systolic volume (ESV). This minimum volume is the point where the first derivative of the volume of the cavity is zero. This cardiac timing corresponds to the label “2” in Fig. 3.
- **Beginning of the atrial systole phase:** The point when the atrium contracts and increases the volume in the ventricle with only a small amount of blood. Atrial contraction is complete before the ventricle begins to contract. The point of maximum atrial contraction is the minimum of the second derivative of the volume of the LV cavity. This cardiac timing corresponds to the label “3” in Fig. 3.

3.5. Spatial normalisation

Spatial normalisation aims to remove bias towards differences in subject-specific LV geometries from the motion analysis. From the previous steps, LV meshes at ED were derived from K subjects from MR and US data. First, a Procrustes alignment was performed to register each mesh to a single mesh selected randomly from the subject cohort, producing a set of affine transformations $\{\phi_{aff}^k\}$, $k = 1, \dots, K$. An unbiased LV mesh was produced by first computing an average of the aligned meshes, and then multiplying this average mesh by the inverse of the average affine transfor-

mation, which we denote by $\tilde{\phi}_{aff} = \frac{1}{K} \sum_{k=1}^K \phi_{aff}^k$. The original transformation $\{\phi_{aff}^k\}$ was normalised to enforce an average similarity transformation equal to identity $\hat{\phi}_{aff} = \phi_{aff}^k \circ (\tilde{\phi}_{aff})^{-1}$.

A Thin Plate Spline transformation $\{\phi_{TPS}^k\}$ was used to align each mesh to the unbiased LV mesh. The resulting transformation from the subject-specific coordinate system to the unbiased LV mesh is given by $\phi^k = \phi_{TPS}^k \circ \hat{\phi}_{aff}^k$ (De Craene et al., 2012).

3.6. Medial surface generation

In order to spatially reduce and regularise the number of vertices (≈ 22000) of the SSM mesh, a medial mesh with regularly sampled vertices (≈ 1000) was generated from the personalised SSM epicardial and endocardial meshes using ray-casting and homogeneous downsampling followed by cell subdivision. The same resampling strategy was employed for all personalised SSM meshes to maintain point correspondence, while reducing the dimensionality of the problem. In our case we only keep points that are more than 3 mm apart from each other, thus reducing the number of vertices of the mesh. We believe that the use of a medial surface enables a more robust motion estimation compared to the endo- and epicardial surfaces of the SSM as it is likely to be less affected by motion tracking errors which can be caused by inaccuracies in the ED segmentation.

3.7. Motion reorientation

For each subject k , for each modality, displacements $\mathbf{u}_{m,t,k} = \mathbf{x}_{m,t,k} - \mathbf{x}_{m,ED,k}$ were computed, where $\mathbf{x}_{m,t,k}$ corresponds to the spatial coordinates of vertex m at cardiac phase t for subject k . With the aim of comparing the LV motions from different subjects in the atlas coordinate system, displacement data for each subject were reoriented into the atlas coordinate system under a small deformation assumption (Ashburner, 2007) using a push-forward action (Rao et al., 2004; Perperidis et al., 2005a):

$$\mathbf{u}_{m,t,k}^{atlas} = \mathbf{J}_{\phi^k}(\mathbf{x}_{m,ED,k}) \mathbf{u}_{m,t,k} \quad (1)$$

where $\mathbf{J}_{\phi^k}$ is the Jacobian of the anatomical transformation ϕ^k mapping each subject's reference ED medial mesh to the atlas ED medial mesh, computed at location $\mathbf{x}_{m,ED,k}$.

3.8. Transform to local coordinate system

For a more intuitive understanding of the LV motion, displacements in the atlas coordinate system were projected onto a local cylindrical coordinate system, providing radial r , longitudinal l , and circumferential c information. The long axis of the LV ED atlas medial mesh was used as the longitudinal direction.

3.9. Dimensionality reduction

As a result of the previous steps, the LV displacements $\mathbf{u}_{m,t,k}^{atlas} \forall m, t, k$ are represented in a common coordinate system. With the aim of combining the motion data from US and MR and reducing the dimensionality of the data, we evaluated three different dimensionality reduction techniques: PCA (Hotelling, 1933), CCA (Hotelling, 1936), and the 2 blocks PLS regression algorithm (Wegelin et al., 2000; Wold, 1985).

All of the evaluated dimensionality reduction techniques are linear. Therefore, they try to determine optimal linear mappings from the high dimensional spaces (i.e. displacement data derived from MR and US) to a low dimensional space whilst preserving as much information as possible. All algorithms represent the multi-view input data as vectors. Consequently, for each modality, the

LV displacements at all cardiac phases were concatenated to form a single feature vector for each subject. The high dimensional MR and US feature vectors formed in this way represent the inputs to the algorithms and their corresponding low dimensional embeddings represent the outputs.

Despite the differences between the proposed algorithms, all of them can be written as a generalised eigenproblem (Wilkinson, 1965). A key difference between them is that PLS and CCA are multi-view subspace learning algorithms and try to find basis vectors for two sets of variables (MR and US derived) such that the projections onto these basis vectors are mutually maximised based on certain criteria. PCA, on the other hand, estimates an orthogonal linear transformation that transforms the data from a single set of variables (i.e. a single modality) into a low dimensional space preserving as much of the variance in the data as possible. PLS and CCA were trained on multi-view MR and US derived data, whereas PCA was trained on MR derived data only and then applied to both MR and US derived data.

We now describe PCA, PLS and CCA using a common notation (Borga et al., 1997). In the following, C_{xx} and C_{yy} correspond to the covariance matrices of the first (i.e. MR) and second (i.e. US) data sets respectively, and C_{xy} and C_{yx} are the cross-covariance matrices of the two data sets (MR/US).

PCA was proposed in Hotelling (1933) to find an orthogonal linear transformation that transforms a set of variables into a new space where they are uncorrelated. This transformation is defined such that it finds the directions of maximum data variance.

$$\begin{aligned} \max_{\mathbf{w}} \quad & \mathbf{w}^T C_{xx} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{w} = 1 \end{aligned} \quad (2)$$

It can be reformulated as a generalised eigenvector problem as follows:

$$C_{xx} \mathbf{w} = \lambda_{PCA} \mathbf{w} \quad (3)$$

where λ_{PCA} and $\mathbf{w}$ are the eigenvalues and eigenvectors respectively of the matrix C_{xx} .

PLS was proposed in Wold et al. (1984) to model the associations between blocks of observed variables. In our work, we focus on the 2 blocks PLS regression algorithm that uses the nonlinear iterative partial least squares (NIPALS) algorithm (Wegelin et al., 2000; Wold, 1985). PLS tries to find an orthogonal basis, such that the covariance between the set of projections onto these basis vectors is mutually maximised:

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} \quad & \mathbf{w}_x^T C_{xy} \mathbf{w}_y \\ \text{subject to} \quad & \mathbf{w}_x^T \mathbf{w}_x = 1, \mathbf{w}_y^T \mathbf{w}_y = 1 \end{aligned} \quad (4)$$

It can be reformulated as a generalised eigenvector problem as follows:

$$\begin{aligned} C_{xy} C_{yx} \mathbf{w}_x &= \lambda_{PLS}^2 \mathbf{w}_x \\ \mathbf{w}_y &= \frac{1}{\lambda_{PLS}} C_{yx} \mathbf{w}_x \end{aligned} \quad (5)$$

where λ_{PLS} and $\mathbf{w}_x$ are respectively the eigenvalues and eigenvectors of the matrix $C_{xy} C_{yx}$.

CCA was proposed in Hotelling (1936) and aims to find basis vectors for two sets of variables such that the correlation between the projections of the variables onto these basis vectors is mutually maximised.

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} \quad & \mathbf{w}_x^T C_{xy} \mathbf{w}_y \\ \text{subject to} \quad & \mathbf{w}_x^T C_{xx} \mathbf{w}_x = 1, \mathbf{w}_y^T C_{yy} \mathbf{w}_y = 1 \end{aligned} \quad (6)$$

It can be reformulated as a generalised eigenvector problem as follows:

$$C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx} \mathbf{w}_x = \lambda_{CCA}^2 \mathbf{w}_x$$

$$\mathbf{w}_y = \frac{1}{\lambda_{CCA}} C_{yy}^{-1} C_{yx} \mathbf{w}_x \quad (7)$$

where λ_{CCA} and $\mathbf{w}_x$ are respectively the eigenvalues and eigenvectors of the matrix $C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}$.

3.10. Atlas application

As described in Section 3.7, the LV displacement vectors for MR and US (both denoted by $\mathbf{u}_{m,t,k}^{atlas} \forall m, t, k$) were represented in a common coordinate system (see Eq. (1)). To apply the dimensionality reduction algorithms (PCA, CCA, PLS) the set of all vectors at all time points for each subject was concatenated, forming the matrices $X = [\mathbf{x}_1, \dots, \mathbf{x}_K]$ and $Y = [\mathbf{y}_1, \dots, \mathbf{y}_K]$ respectively, where X is the MR-derived motion matrix, Y is the US-derived motion matrix and K is the number of subjects. The vectors $\mathbf{x}_k$ and $\mathbf{y}_k$ represent the concatenated displacement vectors from MR and US respectively, with size $(3MT)$ where M corresponds to the total number of vertices of the medial mesh (see Section 3.6) and T the total number of frames.

Hence, a subject can be embedded in the low dimensional space as:

$$\mathbf{z}_k = \mathbf{w}_x^T \mathbf{x}_k \quad \text{and} \quad \mathbf{q}_k = \mathbf{w}_y^T \mathbf{y}_k \quad (8)$$

where $\mathbf{w}_x$ and $\mathbf{w}_y$ are the projection matrices computed from X and Y using a dimensionality reduction algorithm (see Section 3.9, Eqs. (3), (5), and (7)).

Furthermore, the low dimensional embeddings can be reconstructed in the original space as:

$$\hat{\mathbf{x}}_k = \mathbf{w}_x \mathbf{z}_k \quad \text{and} \quad \hat{\mathbf{y}}_k = \mathbf{w}_y \mathbf{q}_k \quad (9)$$

4. Experiments and results

We now describe our experiments on the large scale monomodal and real data sets. All experiments were carried out using the Python programming language, using standard Python libraries (Numpy, SciPy, etc.), VTK libraries, and the scikit-learn Python toolkit (Scikit, 2011).

4.1. Error measures

For all experiments we employed a common set of error measures. The three different evaluation metrics used are illustrated in Fig. 4 and explained below.

Embedding distance:

The embedding distance corresponds to the normalised distance between the low dimensional coordinates obtained using the MR motion vector and those obtained using the US motion vector, and is defined as:

$$E_d^k(MR - US) = \sqrt{\sum_{l=1}^L \frac{(z_{l,k} - q_{l,k})^2}{\sigma_l}} \quad (10)$$

where L is the number of components retained by the dimensionality reduction algorithm, and σ_l is the standard deviation of the combined MR/US cloud of points formed by the embedding in dimension l for subject k .

Reconstruction error:

The reconstruction error is defined as the norm between the original motion vector and the reconstructed motion vector. It is computed separately for both $\mathbf{x}_k$ (MR) and $\mathbf{y}_k$ (US),

$$E_D^k(MR - MR) = \|\mathbf{x}_k - \hat{\mathbf{x}}_k\| \quad (11)$$

$$E_D^k(US - US) = \|\mathbf{y}_k - \hat{\mathbf{y}}_k\| \quad (12)$$

where $\hat{\mathbf{x}}_k$ and $\hat{\mathbf{y}}_k$ are the reconstructed vectors from the embedded coordinates $\mathbf{z}_k$ and $\mathbf{q}_k$ for subject k (see Eq. (9)).

Prediction error:

The prediction error is defined as the norm between the original MR motion vector and the predicted MR motion vector given the US motion vector.

$$E_D^k(MR - US) = \|\mathbf{x}_k - \tilde{\mathbf{x}}_k(\mathbf{y}_k)\| \quad (13)$$

where $\tilde{\mathbf{x}}_k(\mathbf{y}_k) = \mathbf{w}_x \mathbf{q}_k$ is the predicted vector for subject k .

The reconstruction errors are intended to measure how well the dimensionality reduction algorithm can find a common space in which correspondences between the two data sources are preserved whilst avoiding overfitting. The embedding and prediction errors are intended to measure how closely the MR and US data match. Therefore, these two measures quantify how well US-based motion data alone can be used to analyse a new subject using a multimodal atlas. In other words, how similar are the atlas embeddings based on MR and US data.

All metrics were computed for each subject in a leave-one-out cross validation. That is, the atlas was formed from $K-1$ subjects, then the left-out subject was used for embedding into the atlas. The mean of each metric was computed to form the final $E_d(MR - US)$, $E_D(MR - MR)$, $E_D(US - US)$ and $E_D(MR - US)$.

4.2. Large scale monomodal data set experiments

For systematic validation and understanding of the results, a large scale monomodal data set was used for multimodal atlas formation as described below.

First, the atlas formation pipeline as described in Sections 3.1–3.8 was applied to the cine MR images from the UK Biobank cohort, resulting in a set of MR derived displacements for each subject in an atlas coordinate system. Next, to produce synthetic US derived displacements, Gaussian noise was added to the real MR derived displacements. These synthetic US derived displacements were stored and used later as a ground truth for the validation of the different algorithms. The mean and standard deviation of the Gaussian distribution determine the similarity between the motions of the two modalities.

Two experiments were carried out on this data set. In both cases, we first estimated the intrinsic dimensionality d of the MR data using PCA. The estimated value of d (which was typically between 15 and 35) was used for the PCA, CCA and PLS algorithms.

4.2.1. Assessment of reconstruction errors whilst varying the numbers of subjects and vertices

The first experiment aimed to assess the variability of the reconstruction error with regard to the numbers of vertices and subjects. For this experiment we fixed the amount of Gaussian noise added to produce the US displacements to realistic values estimated from our real data set ($\mu_R = -0.25$ and $\sigma_R = 4$). Then, we varied the number of vertices from 150 to 950 by decimating the atlas mesh using the algorithm described in Schroeder et al. (1992), and varied the number of subjects from 50 to 1000. For each combination, using a leave-one-out cross validation we computed the $E_D(MR - MR)$ and $E_D(US - US)$. This experiment assesses the robustness of the three different dimensionality reduction algorithms with regard to number of features and samples.

Fig. 5 shows intensity plots for the reconstruction errors for each dimensionality reduction algorithm used. The x-axis corresponds to the number of subjects varying from 50 to 1000, and the y-axis corresponds to the number of vertices varying from 150 to 950. Note that the intensity plots for $E_D(MR - MR)$ share the same colour bar, whilst those for $E_D(US - US)$ have different colour bars for CCA, PLS and PCA.

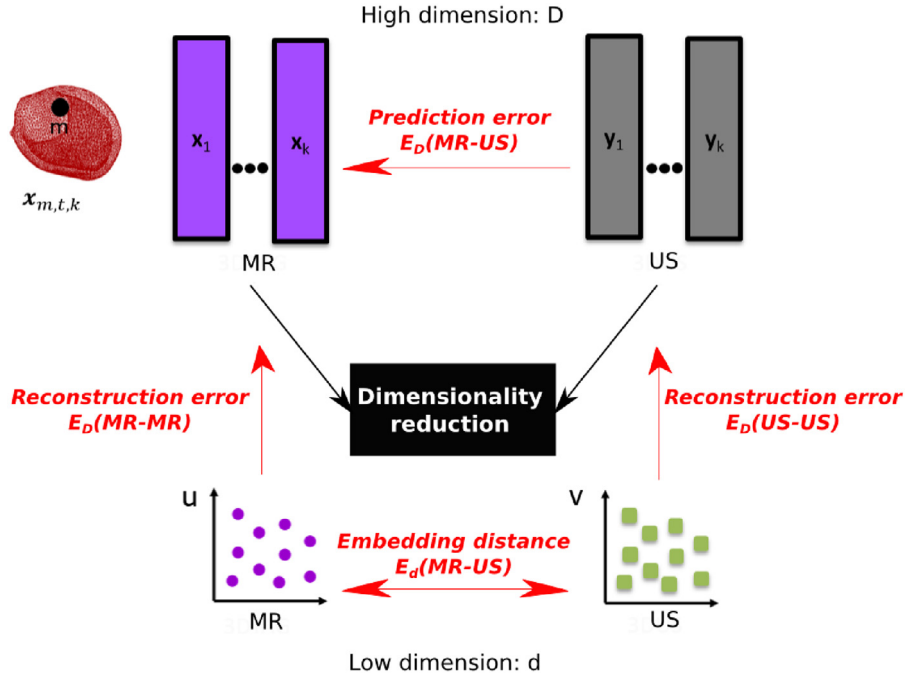


Fig. 4. Illustration of the metrics used for validation.

In general $E_D(MR - MR)$ errors were between 1.2 mm and 2.1 mm, whilst $E_D(US - US)$ for CCA and PLS were between 2.3 mm and 4.7 mm and for PCA between 4.5 mm and 7.2 mm. In all cases, errors increased as the number of subjects decreased and the number of vertices increased. If we look in more detail at the intensity plots for $E_D(MR - MR)$ we notice that the CCA errors were higher for fewer subjects compared to PLS and PCA. The reconstruction error $E_D(US - US)$ was significantly lower for CCA and PLS compared to PCA.

4.2.2. Assessment of prediction errors

In the second experiment we computed the prediction error for varying amounts of Gaussian noise added to the MR displacements to create the synthetic US displacements. For this experiment we fixed the number of vertices to 650 and the number of subjects to 50, which correspond to the same values used in the real database. The aim of this experiment was to assess how the variability between the MR and US data affects the evaluated algorithms.

Fig. 6 a shows the variation of the prediction error whilst varying the Gaussian noise level. The x axis corresponds to the amount of Gaussian noise used (σ_M from 5 mm to 50 mm). For a small σ_M (i.e. more correlated data) the $E_D(MR - US)$ is better for PCA, but then it has a steep slope, whilst the errors for PLS and CCA increase less quickly.

For each σ_M value we also computed the Pearson's correlation coefficient between the MR and US displacements, to assess the correlation between the two input variables (see Fig. 6b). In the following section, we will use these Pearson's correlation coefficient values to relate the results of this experiment to the real data set.

4.3. Real data set experiments

On the real data set, we first present the validation of the three dimensionality reduction algorithms proposed, and we show how the LV volume, LV mass and the ejection fraction are affected by the use of the different algorithms. Then, we compare the use of our new temporal normalisation step to that proposed in Puyol-Antón et al. (2016) and Peressutti et al. (2017) by computing the

LV volume and the dimensionality reduction errors using the two approaches. Finally, we focus on the PLS algorithm, which outperforms the other two methods, and we use our framework to analyse the similarity between motions derived from MR and US.

4.3.1. Dimensionality reduction

We present here the validation of the three proposed dimensionality reduction algorithms: PCA, CCA and PLS. Before applying any of the three algorithms, for each subject k , the vertices of the reference ED medial mesh that fell outside of the field-of-view (FoV) of the MR or US images were removed from the analysis as they are not correctly estimated by the motion tracking algorithm. The intersection of all subjects' FoVs was computed to only retain the points that fell inside the FoV for all subjects. Most of the removed vertices belonged to the LV apex. After removing the vertices that are outside the FoV, the medial mesh had approximately 650 vertices. Note also that, similar to the results presented on the large scale monomodal data set, the intrinsic dimensionality of the real data was computed using PCA (retaining 90% of the variance, $d = 25$), and then used for the PCA, CCA and PLS experiments.

For each dimensionality reduction algorithm, we computed the mean embedding distance $E_d(MR - US)$, the mean reconstruction errors $E_D(MR - MR)$ and $E_D(US - US)$, and the mean prediction error $E_D(MR - US)$ using a leave-one-out cross validation. Fig. 7 shows these errors. We can see that CCA and PLS have much lower errors than PCA for all measures apart from $E_D(MR - MR)$. This is to be expected since PCA is trained on MR data only. However, an advantage of CCA and PLS is that they offer separate mappings for the two modalities, allowing the prediction of one given the other. Consequently, CCA and PLS have lower $E_D(US - US)$, $E_d(MR - US)$ and $E_D(MR - US)$ than PCA.

For the real data set, the Pearson's correlation coefficient between the MR and US displacement data is 0.71, which corresponds to a σ_M in the large scale monomodal experiments of just under 25 mm (see Fig. 6b). We can see from Figs. 5, 6 a and 7 that for this σ_M the reconstructed error and prediction error predicted by the large scale monomodal data set and those found using the real data set are similar. Furthermore, if we compare the results

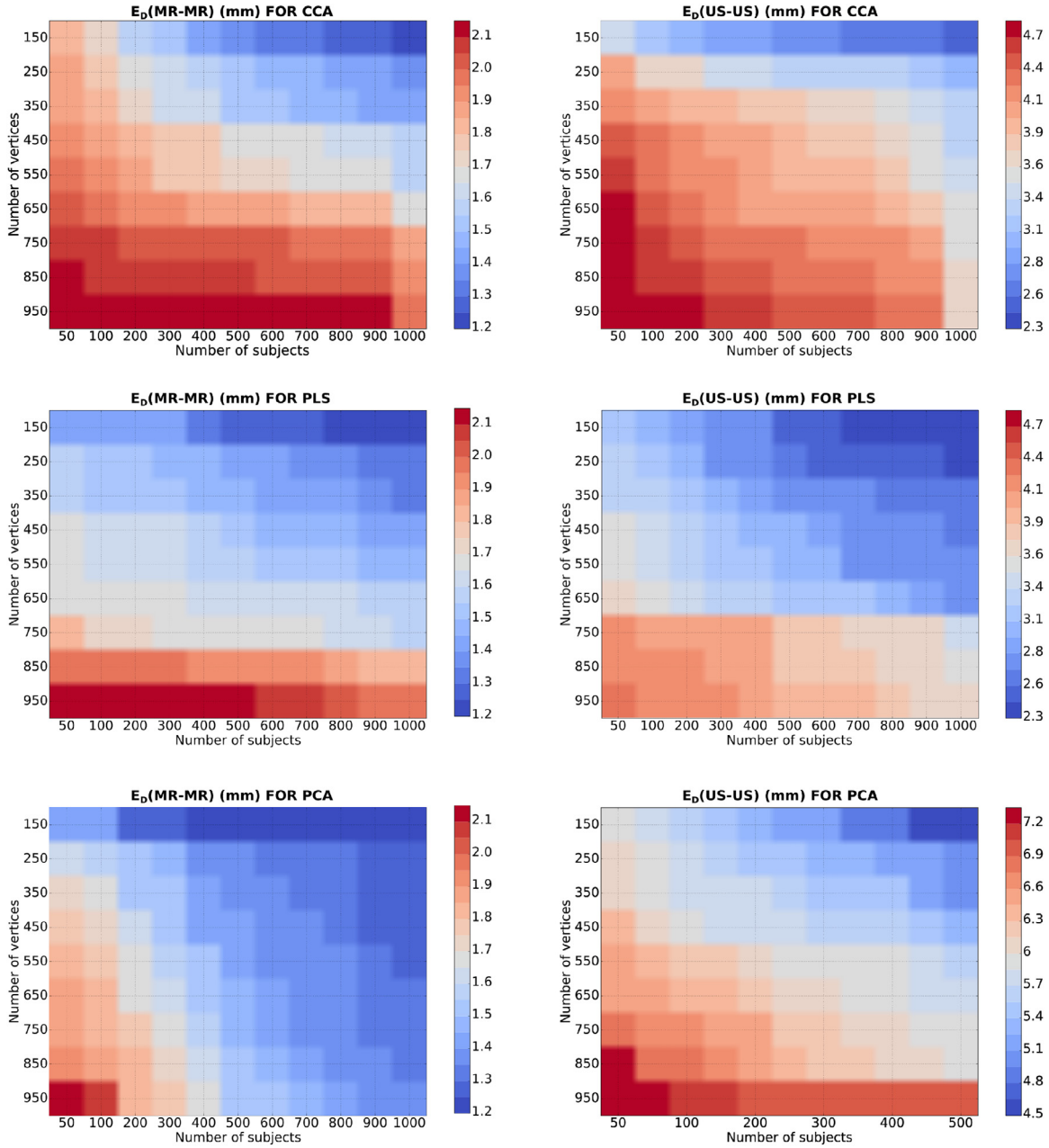


Fig. 5. Reconstruction error in mm $E_D(MR-MR)$ and $E_D(US-US)$ for large scale monomodal data set for CCA, PLS and PCA.

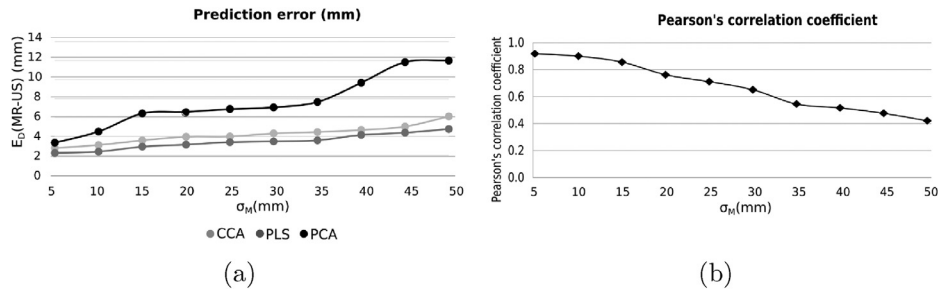
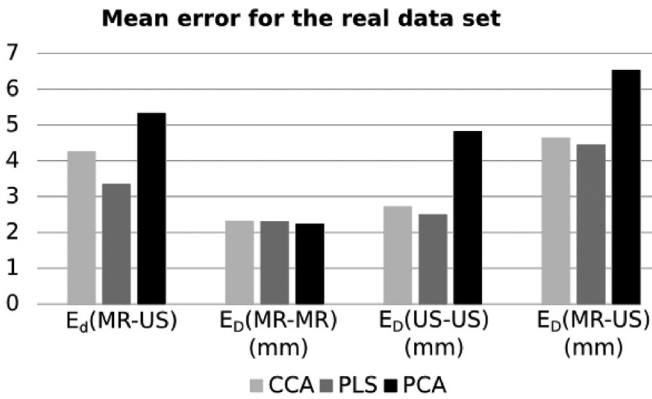


Fig. 6. (a) Prediction errors in mm for large scale monomodal data set for CCA, PLS and PCA. (b) Pearson's correlation coefficient between the MR and US displacement data for the different Gaussian noise levels used.

Table 1

Mean and standard deviation of the errors with respect to the ground truth (GT) for ESV, LV mass and ejection fraction for the different dimensionality reduction algorithms. The first row corresponds to the MR reconstructed data, the second row to the US reconstructed data, and the last one corresponds to the predicted MR from US data. A * indicates a significant difference in mean between the data and the GT, which is shown as the first entry in each row for reference.

		ESV (mL)	E _f %	LV mass (g)
MR only	GT	72.54 ± 31.06	56.98 ± 9.80	103.98 ± 28.20
	PCA	6.13 ± 14.73*	−3.48 ± 5.22*	−6.77 ± 17.85*
	CCA	7.32 ± 16.21*	−3.18 ± 6.40*	−5.09 ± 14.73*
	PLS	7.11 ± 16.79*	−3.15 ± 6.10*	−2.78 ± 14.38
US only	GT	47.09 ± 35.03	52.07 ± 11.94	83.11 ± 34.76
	PCA	7.44 ± 18.04*	−8.56 ± 7.67*	−7.83 ± 15.95*
	CCA	2.57 ± 15.57*	−3.52 ± 5.97*	−2.41 ± 13.87*
	PLS	0.45 ± 14.08	−1.36 ± 4.88	−2.51 ± 14.16*
US-MR predicted	GT	72.54 ± 31.06	56.98 ± 9.80	103.98 ± 28.20
	PCA	−18.02 ± 18.22*	−13.52 ± 8.22*	−26.94 ± 22.56*
	CCA	−3.18 ± 6.68	−5.23 ± 4.86*	−28.01 ± 20.65*
	PLS	−4.50 ± 2.19	−3.12 ± 3.71	−20.77 ± 19.95*

**Fig. 7.** Mean errors for the real data set.

obtained on the real data set with the closest results of the large scale monomodal data set (number of vertices = 650 and number of subjects = 50), one can see that the patterns are similar, i.e. for the MR data the three algorithms have similar errors, while for the US data PLS has the best results and PCA the worst.

4.3.2. LV volume, LV mass and ejection fraction

For a further analysis of the results obtained, we compared the input medial mesh and the reconstructed medial mesh over the cardiac cycle. To quantify the similarity between both medial meshes we computed LV volumes before and after applying the dimensionality reduction algorithms. Ideally, the algorithms should preserve the main physiological characteristics. Fig. 8 shows an example of LV volume comparison for a sample subject from the real data set for the different algorithms proposed in this work.

From the LV volume for each subject, we computed the LV mass as the mean of the ED and ES LV volume multiplied by 1.05 g/mL (Lang et al., 2015), and the ejection fraction, which measures the fraction of outbound blood pumped from the heart with each heartbeat, as follows:

$$E_f(\%) = 100 \cdot \frac{\text{EDV} - \text{ESV}}{\text{EDV}} \quad (14)$$

where EDV corresponds to the LV volume at ED, and ESV corresponds to the LV volume at ES.

Table 1 summarises the ESV, LV mass and ejection fraction errors with respect to the ground truth over all real data set subjects for the reconstructed MR and US, and the predicted MR from

US using the different dimensionality reduction algorithms investigated. To test the accuracy of the different methods we used a Student's *t*-test (99% confidence) between the ground truth values and those reconstructed using each method (a significant difference in mean is shown with a * in the table).

From Fig. 8 and Table 1 we can see that PLS generally reconstructs the clinical parameters more accurately compared to CCA and PCA. Specifically, if we focus on the US-MR predicted row we can see that there was no significant difference in mean between the PLS reconstructed ESV and E_f and the ground truth. However, from Table 1 it seems that all algorithms underestimated the contraction of the LV, leading to a lower volume at ES. There are several possible reasons for this underestimation. First, we employ linear dimensionality reduction techniques and it is possible that the underlying manifold that represents the variation of cardiac cycle motions is nonlinear. Second, we utilise enough dimensions to ensure that 90% of the variance of the data is retained. We justify these simplifications in the Discussion section but both could lead to a slight underestimation of the displacement magnitude and hence ejection fraction. Finally, the use of a medial mesh rather than the epicardial boundary to compute volumetric measures may also cause a slight underestimation.

From a clinical point of view, healthy individuals typically have ejection fractions between 50% and 65% (Kumar et al., 2014) depending on the modality, and outside this range is considered abnormal. From our results, we can see that for MR data all the algorithms reconstructed a mean ejection fraction inside the normal range. For US, the difference between the ground truth and the reconstructed ejection fraction for PCA is high enough to consider a subject pathological. This may have an impact on diagnosis and treatment.

Overall, based on the large scale monomodal data set and real data set results, we can see that PLS and CCA have similar errors and both have lower errors than PCA. However, PLS seems to have the most faithful reconstructed LV volume, when compared to CCA. Additionally, in our experiments we found that CCA was sometimes unstable when the number of features was higher than the number of subjects. Furthermore, CCA is known to be unstable when *X* and *Y* are nearly collinear (Wegelin et al., 2000; Wold, 1985).

4.3.3. Temporal normalisation

To evaluate the temporal normalisation method proposed in Section 3.4, we compared it to performing a global rescaling with respect to the length of the cardiac cycle. To this end, we computed the volume of the LV cavity for each time point. In theory,

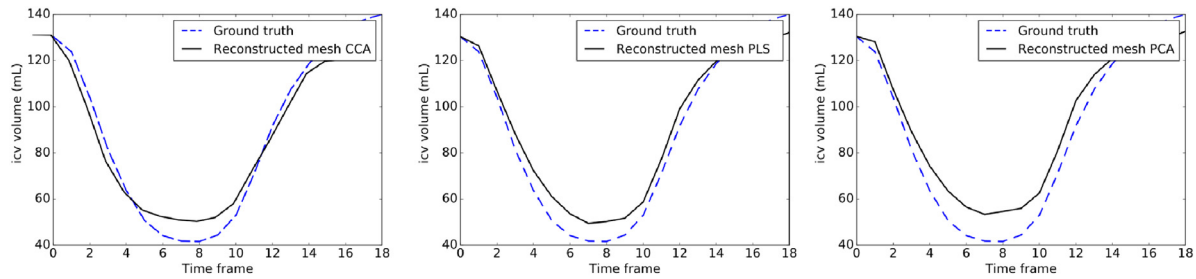


Fig. 8. US LV volume before (blue-dashed) and after (black) applying dimensionality reduction algorithms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the volume of the cavity is greater in the ED frame, and is lower in the ES frame, where the heart is more contracted. We wanted to validate that after the temporal normalisation the volume changes looked physiologically plausible, and that the three cardiac timings described in Section 3.4 were correctly aligned.

In Figs. 9a–c the original and temporally normalised volume curves of the LV medial mesh for each subject at the different time points are presented. To enable easier visualisation of the alignment of the cardiac timings all volumes were rescaled to between 0 and 1. We can see that for all subjects the three considered cardiac timings – beginning of isovolumic contraction phase, beginning of isovolumetric relaxation phase and beginning of atrial systole phase – are correctly aligned for the two modalities MR and US after our proposed normalisation technique but not after the global rescaling approach.

Fig. 10 compares the results of the different dimensionality reduction algorithms using the proposed temporal normalisation and only rescaling by cardiac cycle length. We can see that the proposed temporal normalisation reduces the errors for all three algorithms. We believe that the higher errors for the global rescaling are because frames may not correspond to the same position in the cardiac cycle and therefore displacements from different cardiac timings will be compared.

4.3.4. Similarity between MR and US data

Based on the results presented so far, we considered that PLS has the better performance in terms of errors and stability. Therefore, for the remaining experiment we only use PLS. This experiment aims to find the areas where MR and US derived motion are more similar. To this end, we computed the mean point in the low dimensional space (i.e. the average motion), and then back projected this to the atlas space using the MR and US projection matrices. We then computed the Euclidean distances between the MR and US displacements on a vertex-by-vertex basis. The average such distance for each AHA segment was then computed. This helped us to visualise the areas where PLS finds greater similarities between MR and US derived motion.

Fig. 11 shows the Euclidean distances between the mean MR and US motions in each AHA segment for each time point of the cardiac cycle. We can see that the AHA regions with higher distances are segments 1,2,7,8, which correspond to the base and mid anterior and antero-septal wall. In diagnostic US, it is common that the anterior wall of the myocardium is masked by noise and by limited acoustic windows through the subject's rib cage and lungs (Armstrong and Ryan, 2012). In Fig. 12 we show an example where the anterior wall is missing. This area corresponds to segments 1,2,7,8.

We also see in Fig. 11 that the higher distances occur in frames 6–11, which correspond to the ejection phases (see Fig. 3), where the myocardium has higher contractility.

5. Discussion and conclusion

We have described a complete framework for the formation of a multimodal statistical atlas of cardiac motion, from its construction steps to the impact of different dimensionality reduction algorithms. The three algorithms proposed to build the multimodal cardiac motion atlas were: (1) PCA; (2) CCA; and (3) PLS. This is the first time that such a framework has been described, and we believe that our validation and analysis offer valuable insights into the nature and utility of multimodal MR and US data for analysing cardiac function. Each modality offers the possibility to observe the heart's motion from a distinct perspective, and sometimes offers complementary information. In clinical practice, MR is increasingly accepted as the gold standard for characterising cardiac function and anatomy, as it enables the acquisition of high contrast and high resolution sequences. However, MR is contraindicated for patients with cardiac pacemakers or metallic cardiovascular electronic devices, and furthermore it is a high cost technology. In contrast, US is commonly used in the clinic for assessing cardiac function due to its low cost, portability and good spatial resolution. However, it suffers from image artefacts and limited field of view. A multimodal atlas will benefit from the synergies between the motion features derived from the two modalities, and furthermore, it could be used in clinics to detect cardiovascular disease using only US data.

Our results show that, using a multimodal atlas based on PLS, US-derived motion alone can be used to analyse a new subject's motion whilst exploiting the multimodal (MR+US) data of a population sample. Fig. 7 shows that the accuracy of this process (i.e. $E_D(MR - US)$) is approximately 2.5 mm, which is similar to the magnitude of motion tracking errors (Tobon-Gomez et al., 2013). Note that $E_D(MR - US)$ effectively quantifies how accurate the US processing pipeline is for analysing motion compared to the gold standard of MR.

A possible limitation of the proposed framework to build the multimodal atlas is the temporal normalisation step, which relies on the correct identification of cardiac timings. For instance, the beginning of the atrial systole phase may vary between healthy and diseased subjects, and might be more difficult to identify in patient data. Ideally, we would like to use valve cardiac timings as it would be more accurate. However, the spatiotemporal quality of both TAG and CINE MR data (as obtained in the clinical setting) is currently not sufficient to allow accurate detection of cardiac valve timings. 3D ultrasound imaging suffers from the same problem.

Another limitation of the current approach is that all of the dimensionality reduction algorithms are linear, and therefore they cannot capture the nonlinear structure of high-dimensional data. However, the advantages of a linear model are that it is simpler, and the results are more easily interpreted. In addition, CCA suffers from numerical problems in its eigen-decompositions, as the CCA optimisation procedure involves computation of two inverses. Thus, it is only stable in problems in which the number of fea-

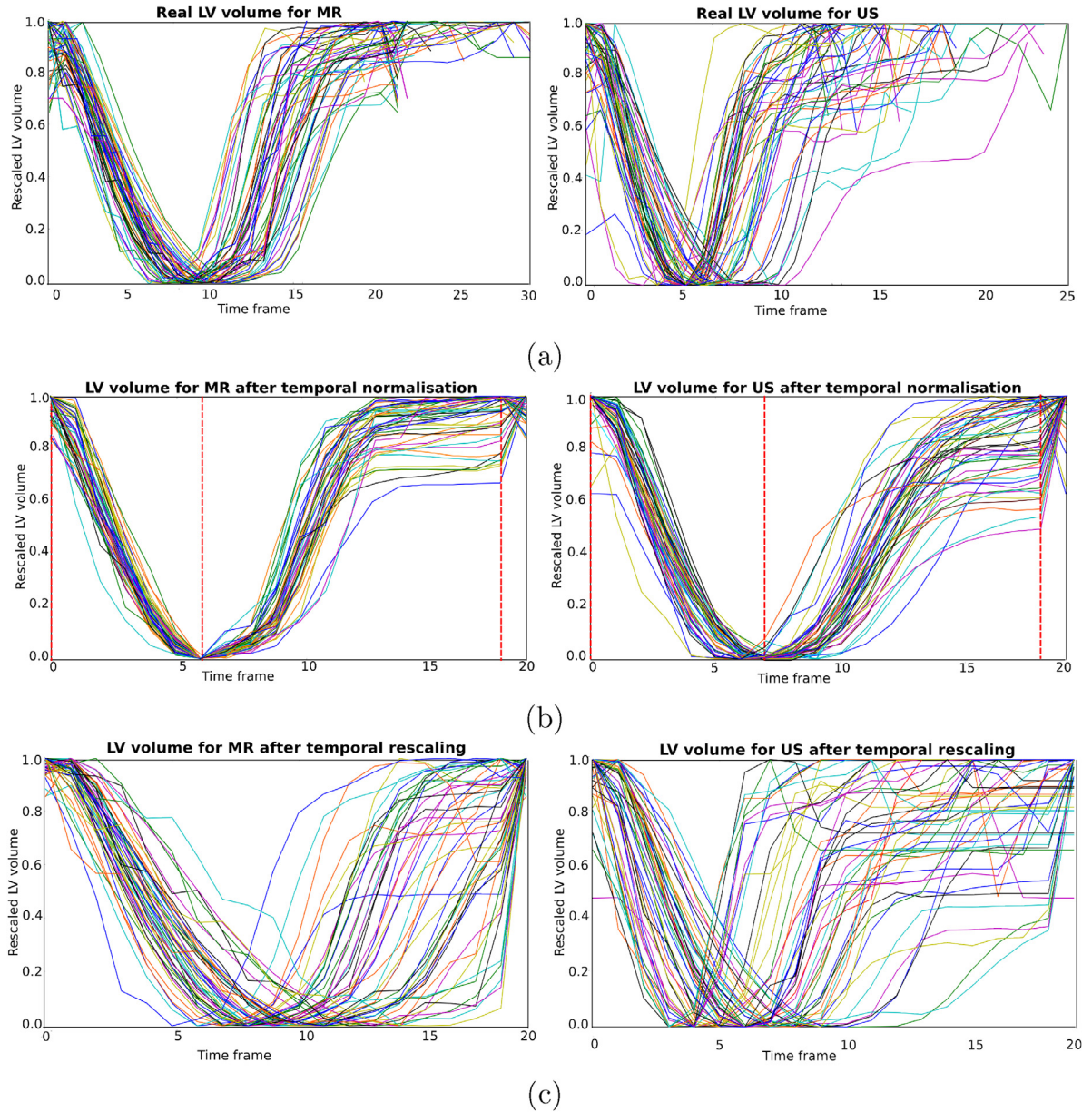


Fig. 9. Rescaled LV volume for each subject over the cardiac cycle (a) before applying temporal normalisation, (b) after applying temporal normalisation, (c) after rescaling with respect to the length of the cardiac cycle. The red dotted lines correspond to the three cardiac timings defined in Section 3.4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

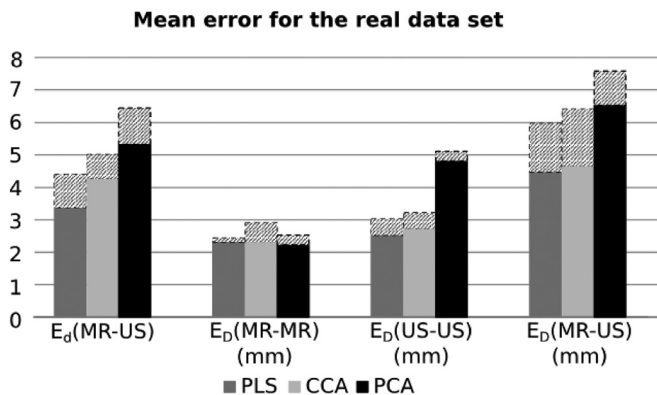


Fig. 10. Mean errors for the real data set using the proposed temporal normalisation, and (stacked) using global rescaling by cardiac cycle length.

tures does not exceed the number of observations. Future work will focus on a nonlinear version of the proposed multi-view algorithms, and on applying the multimodal atlas to clinical problems to learn the representation of pathological motion patterns from certain cardiac diseases in order to detect them in early stages. The results presented in Section 4.3.4 showed that the large variation of MR and US-derived motion is focused on the basal free wall segments, which correspond to the wall with higher variation from a clinical point of view. In future work we aim to use this information as prior knowledge to weight the contributions of features in the dimensionality reduction process.

We used a large-scale monomodal database to compare the performances of the different algorithms under different circumstances, i.e. varying number of samples/features and similarity between modalities. Based on the results we have an approximative idea of the lower bound of the errors of the algorithms as well as how many subjects would be necessary to achieve a good

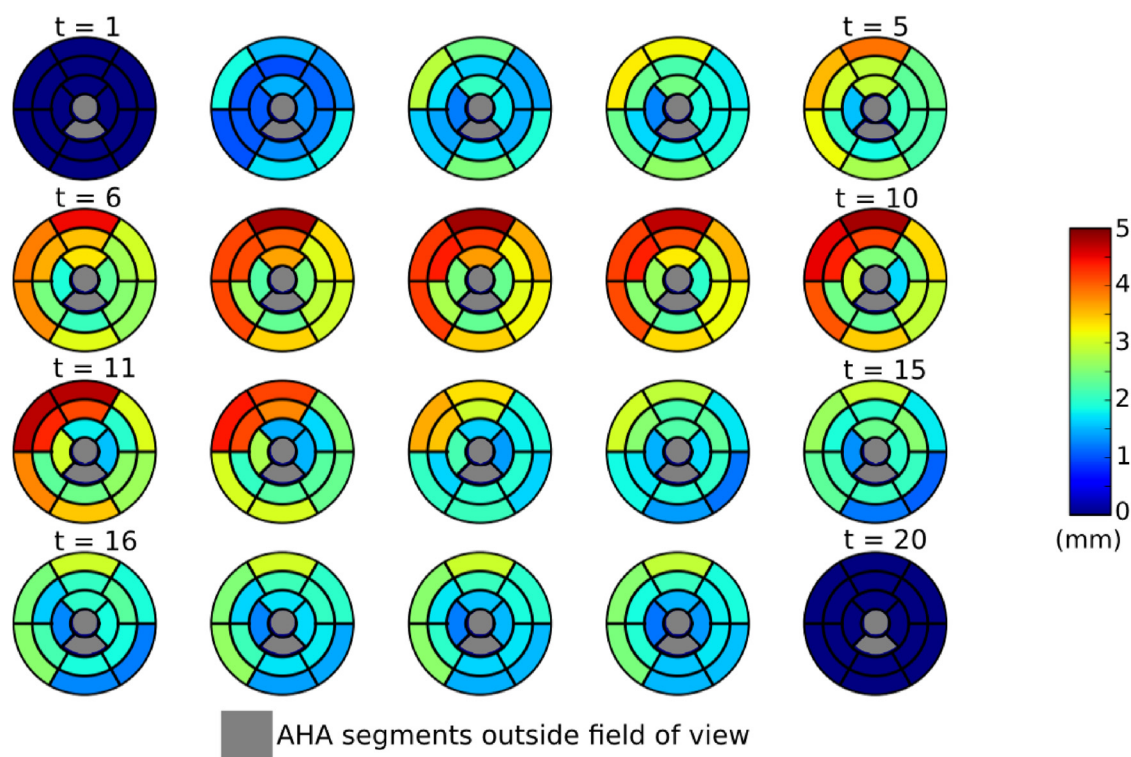


Fig. 11. Euclidean distance in mm between mean MR and US motion for PLS.

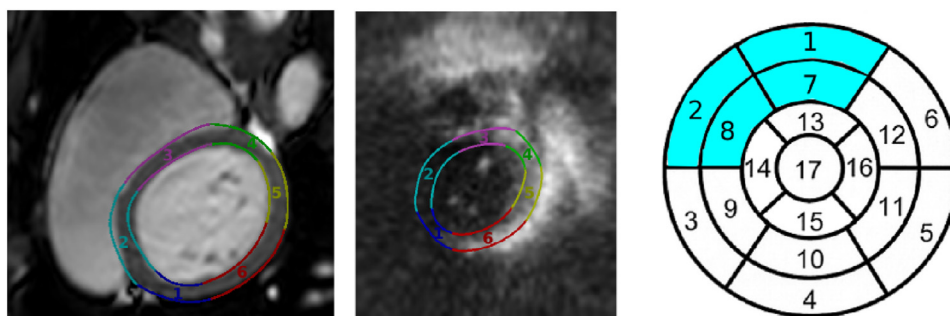


Fig. 12. AHA segments on the base for the ED frame on MR and US, and the 17 standardised AHA segments.

level of accuracy. However, a limitation of the proposed method to construct the synthetic US displacements is the use of a global Gaussian noise model to model the variability between modalities. Greater realism could be achieved by using a locally varying model, but at the expense of the clarity and conciseness of the experimental results.

Our current clinical data set is composed of data from two different centres, which might introduce some bias into the model. In order to minimise this bias, we acquired the data using identical machines (both MR and US). However, we did observe some variation in spatial and temporal resolution, which we presume is due to the lack of standardisation in clinical imaging protocols. This variation was noticeable both between images acquired at the same centre as well between images acquired at different centres. Nevertheless, we found no systematic differences between centres in terms of spatial and temporal resolution.

In the current framework (see Fig. 1) the MR and US geometry/motion estimation pipelines are performed independently, and the results are combined using a multi-view subspace learning algorithm. This strategy was used as we wanted to preserve the potentially complementary motion information provided by the two

modalities. However, alternative pipelines in which the processing of the two modalities is more closely coupled are possible and we plan to investigate these in future work, e.g. joint segmentation and/or joint motion tracking approaches (Liu et al., 2017).

In conclusion, we have presented a new framework to build a multimodal cardiac atlas from MR and US. We aimed to simulate a scenario in which a high quality atlas was formed using MR and US data, and a new subject is related to this atlas using only low cost US data, which is widely available in clinics and can be acquired for almost every subject. We have demonstrated the feasibility of this clinical workflow and we believe that the establishment and validation of this framework could greatly widen the applicability and use of statistical cardiac motion atlases in clinics.

Data access statement

The first real data set is freely available at: <http://stacom.cardiacatlas.org/motion-tracking-challenge/>. The second and third real data sets cannot be made publicly available because informed consent from participants did not cover public deposition of data. Finally, the large scale monomodal UK Biobank data set is publicly available for approved research projects.

Acknowledgements

This work is funded by the [Kings College London & Imperial College London EPSRC Centre for Doctoral Training in Medical Imaging \(EP/L015226/1\)](#). We acknowledge financial support from the Department of Health via the [NIHR](#) comprehensive Biomedical Research Centre award to Guys & St Thomas NHS Foundation Trust with KCL and Kings College Hospital NHS Foundation Trust. The research leading to these results has received funding from the [EU FP7](#) for research, technological development and demonstration under grant agreement VP2HF (no 611823). This research has been conducted using the UK Biobank Resource under Application Number 17806.

References

- Akhter, I., Simon, T., Khan, S., Matthews, I., Sheikh, Y., 2012. Bilinear spatiotemporal basis models. *ACM Trans. Graphics (TOG)* 31 (2), 17.
- Ardekani, S., Weiss, R., Lardo, A., George, R., Lima, J., Wu, K., Miller, M., Winslow, R., Younes, L., 2009. Cardiac motion analysis in ischemic and non-ischemic cardiomyopathy using parallel transport. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 899–902.
- Armstrong, W., Ryan, T., 2012. *Feigenbaum's Echocardiography*. Lippincott Williams & Wilkins.
- Arts, T., Prinzen, F., Delhaas, T., Milles, J., Rossi, A., Clarysse, P., 2010. Mapping displacement and deformation of the heart with local sine-wave modeling. *Med. Imaging IEEE Trans.* 29 (5), 1114–1123.
- Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. *Neuroimage* 38 (1), 95–113.
- Bai, W., Peressutti, D., Parisot, S., Oktay, O., Rajchl, M., O'Regan, D., Cook, S., King, A., Rueckert, D., 2015. Beyond the AHA 17-segment model: motion-driven parcellation of the left ventricle. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 13–20.
- Bai, W., Shi, W., de Marvao, A., Dawes, T., O'Regan, D., Cook, S., Rueckert, D., 2015. A bi-ventricular cardiac atlas built from 1000+ high resolution MR images of healthy subjects and an analysis of shape and motion. *Med. Image Anal.* 26 (1), 133–145.
- Borga, M., Landelius, T., Knutsson, H., 1997. A unified approach to PCA, PLS, MLR and CCA.
- Chandrasekara, R., Rao, A., Sanchez-Ortiz, G., Mohiaddin, R., Rueckert, D., 2003. Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In: *Biennial International Conference on Information Processing in Medical Imaging*, pp. 599–610.
- Crosby, J., Amundsen, B.H., Hergum, T., Remme, E.W., Langeland, S., Torp, H., 2009. 3-D speckle tracking for assessment of regional left ventricular function. *Ultrasound Med. Biol.* 35 (3), 458–471.
- De Craene, M., Duchateau, N., Tobon-Gomez, C., Ghafaryasl, B., Piella, G., Rhode, K., Frangi, A., 2012. SPM to the heart: mapping of 4D continuous velocities for motion abnormality quantification. In: *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 454–457.
- Duchateau, N., De Craene, M., Piella, G., Frangi, A., 2012. Constrained manifold learning for the characterization of pathological deviations from normality. *Med. Image Anal.* 16 (8), 1532–1549.
- Duchateau, N., De Craene, M., Piella, G., Hoogendoorn, C., Silva, E., Doltra, A., Mont, L., Castel, M., Brugada, J., Sitges, M., et al., 2010. Atlas-based quantification of myocardial motion abnormalities: added-value for the understanding of CRT outcome? In: *Statistical Atlases and Computational Models of the Heart*. Springer, pp. 65–74.
- Duchateau, N., De Craene, M., Piella, G., Silva, E., Doltra, A., Sitges, M., Bijmens, B., Frangi, A., 2011. A spatiotemporal statistical atlas of motion for the quantification of abnormal myocardial tissue velocities. *Med. Image Anal.* 15 (3), 316–328.
- Durrleman, S., Pennec, X., Trounev, A., Braga, J., Gerig, G., Ayache, N., 2013. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *Int. J. Comput. Vision* 103 (1), 22–59.
- Fonseca, C., Backhaus, M., Bluemke, D., Britten, R., Do Chung, J., Cowan, B., Dinov, I., Finn, J., Hunter, P., Kadish, A., et al., 2011. The cardiac atlas project: imaging database for computational modeling and statistical atlases of the heart. *Bioinformatics* 27 (16), 2288–2295.
- Garcia-Barnes, J., Gil, D., Badiella, L., Carreras, F., Pujades, S., Martí, E., et al., 2010. A normalized framework for the design of feature spaces assessing the left ventricular function. *Med. Imaging IEEE Trans.* 29 (3), 733–745.
- Georg, M., Souvenir, R., Hope, A., Pless, R., 2008. Manifold learning for 4D CT reconstruction of the lung. In: *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pp. 1–8.
- Gower, J., Dijkstra, G., 2004. *Procrustes Problems*, 3. Oxford University Press Oxford.
- Guerrero, R., Ledig, C., Rueckert, D., 2014. Manifold alignment and transfer learning for classification of Alzheimer's disease. In: *International Workshop on Machine Learning in Medical Imaging*, pp. 77–84.
- Ham, J., Lee, D., Saul, L., 2005. Semisupervised alignment of manifolds. In: *AISTATS 2005*, p. 120.
- Hoogendoorn, C., Duchateau, N., Sanchez-Quintana, D., Whitmarsh, T., Sukno, F., De Craene, M., Lekadir, K., Frangi, A., 2013. A high-resolution atlas and statistical model of the human heart from multislice CT. *Med. Imaging IEEE Trans.* 32 (1), 28–44.
- Hoogendoorn, C., Sukno, F., Ordás, S., Frangi, A., 2009. Bilinear models for spatio-temporal point distribution analysis. *Int. J. Comput. Vision* 85 (3), 237–252.
- Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* 24 (6), 417.
- Hotelling, H., 1936. Relations between two sets of variates. *Biometrika* 28 (3/4), 321–377.
- Kaluzynski, K., Chen, X., Emelianov, S., Skovoroda, A., O'Donnell, M., 2001. Strain rate imaging using two-dimensional speckle tracking. *Ultrason. Ferroelectr. Freq. Control IEEE Trans.* 48 (4), 1111–1123.
- Kerfoot, E., Fovargue, L., Rivolo, S., Shi, W., Rueckert, D., Nordsletten, D., Lee, J., Chabiniok, R., Razavi, R., 2016. Eidolon: visualization and computational framework for multi-modal biomedical data analysis. *LNCS 9805, Medical Imaging and Augmented Reality 2016 (MIAR 2016)*, 9805. Springer.
- Kumar, V., Abbas, A., Fausto, N., Aster, J., 2014. *Robbins and Cotran Pathologic Basis of Disease*. Elsevier Health Sciences.
- Lang, R., Badano, L., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F., Foster, E., Goldstein, S., Kuznetsova, T., et al., 2015. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the american society of echocardiography and the european association of cardiovascular imaging. *J. Am. Soc. Echocardiography* 28 (1), 1–39.
- Liu, H., Wang, T., Lei, X., et al., 2017. Spatiotemporal strategies for joint segmentation and motion tracking from cardiac image sequences. *IEEE J. Transl. Eng. Health Med.* 5.
- Lorenzi, M., Ayache, N., Pennec, X., 2011. Schild's ladder for the parallel transport of deformations in time series of images. In: *Biennial International Conference on Information Processing in Medical Imaging*, pp. 463–474.
- Lorenzi, M., Gutman, B., Hibar, D., Altmann, A., Jahanshad, N., Thompson, P., Ourselin, S., 2016. Partial least squares modelling for imaging-genetics in alzheimer's disease: plausibility and generalization. In: *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*, pp. 838–841.
- Lu, Y., Radau, P., Connelly, K., Dick, A., Wright, G., 2009. Pattern recognition of abnormal left ventricle wall motion in cardiac MR. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 750–758.
- Mansi, T., Pennec, X., Sermesant, M., Delingette, H., Ayache, N., 2011. iLogDemos: a demons-based registration algorithm for tracking incompressible elastic biological tissues. *Int. J. Comput. Vision* 92 (1), 92–111.
- Mateus, D., Cuzzolin, F., Horaud, R., Boyer, E., 2007. Articulated shape matching using locally linear embedding and orthogonal alignment. In: *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8.
- McLeod, K., Seiler, C., Sermesant, M., Pennec, X., 2012. A near-incompressible poly-affine motion model for cardiac function analysis. In: *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 288–297.
- McLeod, K., Seiler, C., Sermesant, M., Pennec, X., 2013. Spatio-temporal dimension reduction of cardiac motion for group-wise analysis and statistical testing. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013*. Springer, pp. 501–508.
- McLeod, K., Sermesant, M., Beerbaum, P., Pennec, X., 2015. Descriptive and intuitive population-based cardiac motion analysis via sparsity constrained tensor decomposition. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 419–426.
- Medrano-Gracia, P., Cowan, B., Bluemke, D., Finn, J., Kadish, A., Lee, D., Lima, J., Suinesiaputra, A., Young, A., 2014. Continuous spatio-temporal atlases of the asymptomatic and infarcted hearts. In: *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*. Springer, pp. 143–151.
- Medrano-Gracia, P., Cowan, B., Bluemke, D.A., Finn, J., Kadish, A., Lee, D., Lima, J., Suinesiaputra, A., Young, A., 2013. Atlas-based analysis of cardiac shape and function: correction of regional shape bias due to imaging protocol for population studies. *J. Cardiovasc. Magn. Reson.* 15 (1), 80.
- Medrano-Gracia, P., Cowan, B.R., Suinesiaputra, A., Young, A.A., 2015. Atlas-based anatomical modeling and analysis of heart disease. *Drug Discovery Today: Disease Models* 14, 33–39.
- Medrano-Gracia, P., Cowan, B.R., Finn, J., Fonseca, C., Kadish, A., Lee, D., Tao, W., Young, A., 2010. The cardiac atlas project: preliminary description of heart shape in patients with myocardial infarction. In: *STACOM/CESC*, pp. 46–53.
- Medrano-Gracia, P., Suinesiaputra, A., Cowan, B., Bluemke, D., Frangi, A., Lee, D., Lima, J., Young, A., 2013. An atlas for cardiac MRI regional wall motion and infarct scoring. In: *Statistical atlases and computational models of the heart. Imaging and modelling challenges*. Springer, pp. 188–197.
- Peressutti, D., Sinclair, M., Bai, W., Jackson, T., Ruijsink, J., Nordsletten, D., Asner, L., Hadjicharalambous, M., Rinaldi, C., Rueckert, D., et al., 2017. A framework for combining a motion atlas with non-motion information to learn clinically useful biomarkers: application to cardiac resynchronisation therapy response prediction. *Med. Image Anal.* 35, 669–684.
- Perperidis, D., Mohiaddin, R., Rueckert, D., 2005. Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2005*. Springer, pp. 402–410.
- Perperidis, D., Mohiaddin, R., Rueckert, D., 2005. Spatio-temporal free-form registration of cardiac MR image sequences. *Med. Image Anal.* 9 (5), 441–456.
- Petersen, S., Matthews, P., Francis, J., Robson, M., Zemrak, F., Boubertakh, R., Young, A., Hudson, S., Weale, P., Garratt, S., et al., 2016. UK biobanks cardiovascular magnetic resonance protocol. *J. Cardiovasc. Magn. Reson.* 18 (1), 8.

- Punithakumar, K., Ayed, I., Islam, A., Goela, A., Ross, I., Chong, J., Li, S., 2013. Regional heart motion abnormality detection: an information theoretic approach. *Med. Image Anal.* 17 (3), 311–324.
- Puyol-Antón, E., Peressutti, D., Aljabar, P., De Craene, M., Piro, P., King, A., 2016. Towards a multimodal cardiac motion atlas. *Biomedical Imaging (ISBI)*, 2016 13th IEEE International Symposium on.
- Qiu, A., Albert, M., Younes, L., Miller, M., 2009. Time sequence diffeomorphic metric mapping and parallel transport track time-dependent shape changes. *Neuroimage* 45 (1), S51–S60.
- Rao, A., Chandrashekar, R., Sanchez-Ortiz, G., Mohiaddin, R., Aljabar, P., Hajnal, J., Puri, B.K., Rueckert, D., 2004. Spatial transformation of motion and deformation fields using nonrigid registration. *Med. Imaging IEEE Trans.* 23 (9), 1065–1076.
- Rougon, N., Petitjean, C., Preteux, F., 2004. Building and using a statistical 3D motion atlas for analyzing myocardial contraction in MRI. In: *Medical Imaging 2004*, pp. 253–264.
- Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *Med. Imaging IEEE Trans.* 18 (8), 712–721.
- Schroeder, W., Zarge, J., Lorensen, W., 1992. Decimation of triangle meshes. In: *ACM Siggraph Computer Graphics*, 26, pp. 65–70.
- Scikit, S.-I.D., 2011. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Sinclair, M., Peressutti, D., Puyol-Antón, E., Bai, W., Nordsletten, D., Hadjicharalambous, M., Kerfoot, E., Jackson, T., Claridge, S., Rinaldi, C., Rueckert, D., King, A., 2016. Learning optimal spatial scales for cardiac strain analysis using a motion atlas. *International Workshop on Statistical Atlases and Computational Models of the Heart*.
- Studholme, C., Hill, D., Hawkes, D., 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* 32 (1), 71–86.
- Suinesiaputra, A., Frangi, A.F., Kaandorp, T., Lamb, H., Bax, J., Reiber, J., Lelieveldt, B., 2009. Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. *Med. Imaging IEEE Trans.* 28 (4), 595–607.
- Sun, S., 2013. A survey of multi-view machine learning. *Neural Comput. Appl.* 23 (7–8), 2031–2038.
- Tobon-Gomez, C., De Craene, M., Mcleod, K., Tautz, L., Shi, W., Hennemuth, A., Prakosa, A., Wang, H., Carr-White, G., Kapetanakis, S., et al., 2013. Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. *Med. Image Anal.* 17 (6), 632–648.
- Unser, M., 1999. Splines: a perfect fit for signal and image processing. *IEEE Signal Process. Mag.* 16 (6), 22–38.
- Wachinger, C., Yigitsoy, M., Rijkhorst, E., Navab, N., 2012. Manifold learning for image-based breathing gating in ultrasound and MRI. *Med. Image Anal.* 16 (4), 806–818.
- Wegelin, J.A., et al., 2000. A survey of partial least squares (PLS) methods, with emphasis on the two-block case. *Tech. Rep.* University of Washington, Department of Statistics.
- Wilkinson, J., 1965. *The Algebraic Eigenvalue Problem*, 87. Clarendon Press Oxford.
- Wold, H., 1985. *Partial Least Squares*. Encyclopedia of statistical sciences.
- Wold, S., Albano, C., Dunn III, W., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W., Sjöström, M., 1984. *Multivariate data analysis in chemistry*. In: *Chemometrics*. Springer, pp. 17–95.
- Xu, C., Tao, D., Xu, C., 2013. A survey on multi-view learning. *CoRR abs/1304.5634*.
- Yushkevich, P., Piven, J., Hazlett, H., Smith, R., Ho, S., Gee, J., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures. *Neuroimage* 31 (3), 1116–1128.
- Zhou, Y., Bernard, O., Saloux, E., Manrique, A., Allain, P., Makram-Ebeid, S., De Craene, M., 2015. 3D harmonic phase tracking with anatomical regularization. *Med. Image Anal.* 26 (1), 70–81.