A Registration-Based Propagation Framework for Automatic Whole Heart Segmentation of Cardiac MRI

Xiahai Zhuang*, Kawal S. Rhode, Reza S. Razavi, David J. Hawkes, and Sebastien Ourselin

Abstract—**Magnetic resonance (MR) imaging has become a routine modality for the determination of patient cardiac morphology. The extraction of this information can be important for the development of new clinical applications as well as the planning and guidance of cardiac interventional procedures. To avoid inter- and intra-observer variability of manual delineation, it is highly desirable to develop an automatic technique for whole heart segmentation of cardiac magnetic resonance images. However, automating this process is complicated by the limited quality of acquired images and large shape variation of the heart between subjects. In this paper, we propose a fully automatic whole heart segmentation framework based on two new image registration algorithms: the locally affine registration method (LARM) and the free-form deformations with adaptive control point status (ACPS FFDs). LARM provides the correspondence of anatomical substructures such as the four chambers and great vessels of the heart, while the registration using ACPS FFDs refines the local details using a constrained optimization scheme. We validated our proposed segmentation framework on 37 cardiac MR volumes on the end-diastolic phase, displaying a wide diversity of morphology and pathology, and achieved a mean accuracy of 2.14 $\pm$ 0.63 mm (rms surface distance) and a maximal error of 4.31 mm.**

Index Terms—**Atlas, cardiac magnetic resonance imaging (MRI), dynamic resampling and distance weighting interpolation (DRAW), free-form deformations, image registration, inverse transformation, locally affine registration, locally affine registration method (LARM), whole heart segmentation.**

I. INTRODUCTION

ACCORDING to the World Health Organization [1], an estimated 17.5 million people died from cardiovascular diseases (CVDs) in 2005, accounting for 30% of deaths around the world. Being able to provide an early diagnosis and treatment will dramatically reduce this death toll. Recent advances in novel imaging and computing technology and their introduction into clinical routine have shown tremendous potential towards achieving such an ambitious goal. Over the diverse range of imaging modalities, cardiac magnetic resonance imaging (MRI) is a unique technique which is ionizing radiation free and can provide clear anatomy of the heart. Extracting the anatomical information is the essential step for the development of clinical applications, and obtaining reproducible and unbiased quantitative measurement of the anatomy is indeed of central importance for the success of these applications. To avoid inter- and intraobserver variations of manual delineation, it is highly desirable to develop an automatic technique for whole heart segmentation of cardiac MRI.

This has been the focus of several research groups. However, only a few studies presented whole heart segmentation, while the majority investigated the segmentation of the ventricles of the heart only. In the following section, we will review the state of the art and give the motivation for developing a novel approach overcoming the current limitations.

A. Related Work

Being able to obtain an accurate segmentation of the heart on clinical cardiac MR using only the image content has been shown to be challenging, if not impossible, to achieve. We invite the reader to refer to [2] for a detailed review of this topic. For this reason, model-guided methods, incorporating prior knowledge into the segmentation procedure, have been very popular.

Boundary-based segmentation using deformable models is currently the most widely studied technique for automatically segmenting cardiac MR images [3]–[9]. In these studies, a preconstructed model is deformed towards detected boundaries in the *unseen* image ("unseen" meaning an image that requires segmentation). The deformation is constrained by prior knowledge of the overall shape. However, using these approaches, three challenges are remaining in the reported studies.

- Currently reported techniques mainly define the boundary profile using the information from a small local area of the unseen image, such as the intensity value and intensity gradient [9]. This profile can be sensitive to *intensity inconsistencies*, coming from noise, artifacts, and intensity distribution variations.
- Deformable model-based segmentation assumes that the corresponding edge point for each local node of the model is the optimal edge point from the searching direction, commonly being the normal direction from the model surface. This however does not guarantee a true anatomical

correspondence, nor even a good approximation of it if the surface of the model is not closely initialized.

- The segmentation may not naturally guarantee *diffeomorphism*, meaning a one-to-one mapping from the model to the unseen image. This segmentation may lead to the result where two or more surfaces from different substructures of the heart model intersect each other. Diffeomorphism may not be so important and could be achieved by post-processing in some applications, but a natural guarantee of this property could be useful in others.

Some of these limitations could be overcome by using a more complex model, such as a statistical shape model (SSM). This model can be built from a training set by means of principle component analysis (PCA). By adapting the modes of the PCA model, the boundaries of the SSM can be deformed to the unseen image, provided that the statistical model includes a potentially large shape variability and the model has been well initialized to the unseen data.

Another popular approach, which can overcome some of the described limitations, is to propagate the segmentation of an atlas image using image registration techniques [10]. We can classify these approaches as part of the "registration-based segmentation propagation framework." To preserve the topology of the heart, the resultant transformation of the registration should be diffeomorphic, which is nowadays achievable [11]–[14]. In addition, these methods have two potential advantages. Firstly, the boundaries with low contrast or that are not clearly discernible in the unseen images can be retrieved through the global warping of the atlas. Secondly, by using intensity-based similarity measures such as the mutual information (MI), the segmentation can be less sensitive to noise, artifacts, and different intensity distributions. The registration using MI measures, including the normalized mutual information (NMI) [15], has shown good robustness to these intensity inconsistencies in previous studies [15], [16].

The main difficulty of the segmentation-propagation frameworks is to estimate the appropriate spatial mapping, namely the resultant transformation from the registration process. Usually, the standard registration scheme uses a global affine registration to *localize* the heart, and then apply a nonrigid registration with a transformation containing high degrees of freedom (DOFs), such as free-form deformations (FFDs) [10], to *refine* the local details. However, one common issue of the segmentation-propagation and boundary-based techniques is their relative sensitivity to initialization, making the algorithms less robust to large shape variability, commonly seen when dealing with pathologies.

To tackle this initialization problem, Lorenzo-Valdes *et al.* built subject-specific atlases to segment the ventricles of multiphase MR images [17]. This personalized atlas minimized the shape variability, which improved the registration process. However, the robustness was limited to the images acquired from the same subject from which the subject-specific "atlas" was constructed. Mitchell *et al.* described an active appearance model (AAM) as the atlas image which had statistical information of both shape and texture from the training set [18]. They matched the AAM to unseen data by means of optimizing the model coefficients, for both shape and texture modeling, by minimizing the intensity root mean square (rms) difference. Lotjonen *et al.* constructed a similar AAM for the whole heart [19], adapting the model to short- and long-axis MR images using a segmentation propagation framework. However, neither of them applied constraints to guarantee that the registration result was diffeomorphic. In addition, the issue of large shape variability remains problematic in the training phase of the AAM.

B. Contribution of This Work

In this paper, we propose a registration framework able to preserve the topology and to deal with the large shape variability of the heart. The core of our framework is based on two contributions extending the current segmentation-propagation frameworks, namely a Locally Affine Registration Method (LARM) [20] and a nonrigid registration using free-form deformations with adaptive control point status (ACPS FFDs) [21].

In the proposed scheme, LARM is applied to obtain a robust initialization of the different *substructures* of the heart such as the four chambers and the major vessels. The resultant transformation globally deforms the atlas but locally maintains the shape of the predefined substructures. Such an approach is able to avoid local optima during the optimization of the global transformation. After the initialization, ACPS FFD registration is used to refine the local detail. Our scheme makes advantageous use of prior knowledge to adaptively associate each control point in the FFDs with a status, active or passive, extending the nonuniform FFDs proposed by Schnabel *et al.* [22]. This contributes to avoiding the *myocardial leaking*, meaning the epicardium of the atlas is mapped to adjacent tissues of the epicardium in the unseen image.

Furthermore, one will need an inverse transformation from LARM to propagate the segmentation. We hence propose a new algorithm for inverting the transformation based on Dynamic Resampling And distance Weighting interpolation (DRAW) [20]. DRAW is generic, i.e., widely applicable to any diffeomorphic transformations.

In the following sections, we will demonstrate that our segmentation-propagation framework is robust to different pathologies and large shape variability. Section II will present our methodological contributions in detail, while Section III will describe the experimental setup and validation results, using *in vivo* data from 37 individuals. Section IV will conclude the paper and present some insights for potential extension of the framework.

II. METHODOLOGY

The methodology is presented in this section as follows. Section II-A introduces the LARM. Section II-B presents ACPS FFD registration. Section II-C elaborates on the DRAW algorithm for inverting transformations. Section II-D describes the atlas construction and provides the automatic whole heart segmentation framework.

A. LARM: Locally Affine Registration Method

Registration using locally affine transformations is an attractive registration alternative for applications where a single global affine transformation cannot provide enough accuracy,

while a nonrigid registration would incorrectly affect the local topology due to the large shape variability of the heart anatomy [23]–[26]. An affine transformation is also challenged to provide a good initialization of substructures of the heart for a nonrigid registration with high degrees of freedom (DOFs). Locally affine registration method (LARM), which assigns a local affine transformation to each substructure such as the four chambers and great vessels, can further initialize the image pair after a global affine registration has been achieved. This initialization is crucial to achieve a robust registration of the following nonrigid registration to maintain the local topology in intersubject cases.

Arsigny *et al.* developed an elegant framework for the fusion of a set of local affine transformations and applied it to the registration of histological slices [24]. The similarity measure in the work was the sum of square differences which is not designed for the images with significantly different intensity distributions such as cardiac MRI. Based on [24], Commowick *et al.* further developed an efficient locally affine framework, where the local affine transformations can be optimized separately [26]. The separate optimization however could affect the registration robustness due to the loss of *global intensity linkage* between local regions. In this work, we propose a new method, referred to as LARM, which uses MI or NMI as the global cost function and optimize the affine transformations within a global scheme. LARM confines the driving forces of the local affine transformation parameters to local regions to improve the efficiency and effectiveness of the computation, but it also provides the function to maintain the intensity linkage based on the global joint histogram table.

Let $\{V_i\}$ be the set of predefined local regions which have a minimal distance between each other; let $\{G_i\}$ be the set of the assigned local affine transformations. To achieve a global non-linear transformation T from $\{G_i\}$, a direct fusion based on the distance weighting interpolation [23], [27] is used

$$T(x) = \begin{cases} G_i(x), & x \in V_i, i = 1 \ldots n \\ \sum_{i=1}^{n} w_i(x)G_i(x_i), & x \notin \bigcup_{i=1}^{n} V_i \end{cases} \quad (1)$$

where $w_i(x)$ is a normalized weighting factor related to the distance $d_i(x)$ between point x and region V_i

$$w_i(x) = \frac{1/(d_i(x))^e}{\sum_{i=1}^{n} 1/(d_i(x))^e} \quad (2)$$

where e controls the locality of affine transformations. We use 2 for e in our implementation following the recommendation in [27] for computational efficiency. In this transformation model, a global affine transformation can also be assigned to the boundary of the region of interest (ROI).

However, there are two situations that can cause nondiffeomorphism by this direct fusion. Firstly, the local regions can overlap each other after the individual transformations. To guarantee nonoverlapping, a correction of the local regions is required during the registration

$$V_i = G_i^{-1}\left(G_i(V_i) - \oplus_L (R_{ij})\right)$$
$$= V_i - G_i^{-1}\left(\oplus_L (R_{ij})\right) \quad (3)$$

where $R_{ij} = \bigcup_{i \neq j}(G_j(V_j))$ is the volume of other local regions that V_i should not overlap after transformations. $\oplus_L$ is the morphology dilation with length L, 10 mm in our implementation for the cardiac MR application, to leave enough space for the interpolation. It should be noted when $\{G_i\}$ are all identity transformations such as at the beginning of registration, the correction using (3) also guarantees that there is at least L distance between each local region.

Secondly, when the displacements of the local transformations are large, the direct fusion using (1) can produce folding [24]. A regularization step monitoring the Jacobian J_T of the deformation field is then needed

$$J_T = \sum_{i=1}^{n} \frac{\partial w_i}{\partial x} \cdot G_i + \sum_{i=1}^{n} w_i \cdot \frac{\partial G_i}{\partial x} = \nabla W \cdot G^T + \nabla G \cdot W_T. \quad (4)$$

If the determinant of the Jacobian, $\det(J_T)$, drops below a threshold (e.g., $\det(J_T) < 0.5$), a new image-pair is then generated from the original one using the current deformation field and the locally affine transformation model is reset to identity for a new optimization process. Therefore, the resultant transformation of the registration is a concatenation of the series of transformations

$$T = T_m \circ T_{m-1} \circ \ldots \circ T_1. \quad (5)$$

Since all the $\{T_j | j = 1 \ldots m\}$ are diffeomorphic, the composed T should keep this property as well.

The derivative of MI or NMI can be obtained from the operation of the entropy derivatives, which can be computed from the derivative of the probability distribution function (PDF). Let (l, k) be a joint histogram bin from the reference image I_t and floating image I_s, ω be the kernel density estimation function, the derivative of joint PDF $p(l, k) = \sum_x \omega(I_t(x)) \cdot \omega(I_s(T(x)))$ against a transformation parameter θ_i of G_i is

$$\begin{aligned} \partial p(l, k)/\partial \theta_i &= \sum_x \omega(I_t(x)) \cdot (\partial \omega(I_s(T(x)))/\partial \theta_i) \\ &= \sum_x \omega(I_t)\nabla\omega(\nabla I_s \cdot \partial G_i(x)/\partial \theta_i)w_i(x) \\ &= \sum_x W(x). \end{aligned} \quad (6)$$

If $w_i(x) \approx 0$, then $W(x) \approx 0$.

Given a volume Ω and its complementary volume $\bar{\Omega}$, we can rewrite the joint PDF as $p(l, k) = P_\Omega + P_{\bar{\Omega}}$, where $P_\Omega = \sum_{x \in \Omega} \omega(I_t)\omega(I_s)$ and $P_{\bar{\Omega}} = \sum_{x \in \bar{\Omega}} \omega(I_t)\omega(I_s)$. Therefore, the derivative of the entropy becomes

$$\begin{aligned} \frac{\partial H}{\partial \theta_i} &= -\sum_{l,k}(1 + \log(p(l,k))) \cdot \left(\frac{\partial P_\Omega}{\partial \theta_i} + \frac{\partial P_{\bar{\Omega}}}{\partial \theta_i}\right) \\ &= \mathcal{F}_\Omega^{\theta_i} + \mathcal{F}_{\bar{\Omega}}^{\theta_i} \end{aligned} \quad (7)$$

where $\mathcal{F}_\Omega^{\theta_i} = -\sum_{l,k}(1 + \log(p(l,k)))(\partial P_\Omega/\partial \theta_i)$ and $\mathcal{F}_{\bar{\Omega}}^{\theta_i} = -\sum_{l,k}(1 + \log(p(l,k)))(\partial P_{\bar{\Omega}}/\partial \theta_i)$. Let U_i be a superset volume of V_i such that when $x \in \bar{U}_i$ one or both of the following two conditions can be met: one is that $w_i(x) \approx 0$ and the other is that x is within the background which is not of

```
For each optimization step
    Global transformation, F(x) = T ∘ G, and
        T = T_m ∘ T_{m-1} ∘ ... ∘ T_1
    Optimize global affine G
    Optimize local affine G_i of T_m, compute derivatives based on (8)
    Regularization step:
        For each region V_i in {V_i}
            Overlap correction using (3)
            If V_i needs correction and is corrected, then
                If |V_i| = 0, then END registration
                else re-compute distance transformation of V_i for (2)
        End of each region V_i in {V_i}
        If MIN(det(J_{T_m})) < 0.5, then
            new a T_{m+1}, T = T_{m+1} ∘ T, and m = m + 1
    End of Regularization
End of Optimization
```

Fig. 1.   Pseudo-code of the proposed LARM.

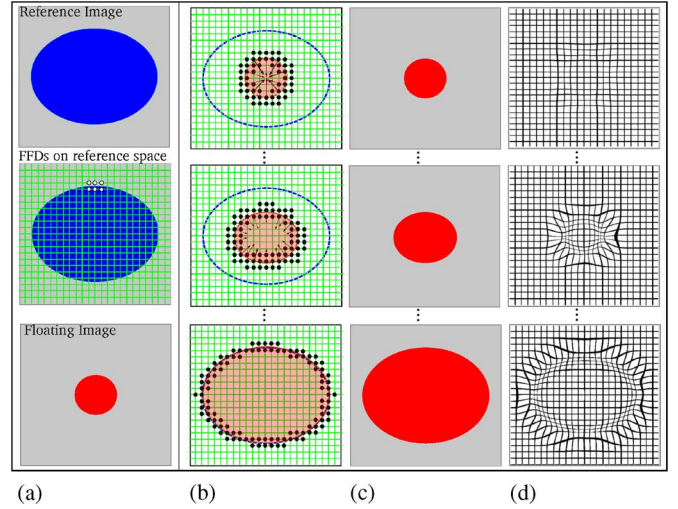(a)   (b)   (c)   (d)

Fig. 2.   Registration of an ellipse, the reference image and a circle, the floating image using adaptive control point status FFDs. The white dots in the image of middle (a) are special control points explained in the text. (b) FFD meshes, the contour of the ellipse, and the contour of the inverse-transformed floating image. The black dots are activated control points. The arrows demonstrate the registration driving forces and no arrow means a convergence of the registration. (c) Floating image inversely transformed into the reference space at different registration steps. (d) Deformed FFD meshes at different registration steps whose concatenation gives the resultant transformation.

the interest of the registration. Let $\Omega = U_i$, then $\mathcal{F}_{\bar{\Omega}}^{\theta_i}$ is summarized from the contribution of either relatively negligible small forces or less relevant forces coming from the background. Therefore, the derivative which represents the driving force is approximated as

$$\frac{\partial H}{\partial \theta_i} \approx \mathcal{F}_{\Omega}^{\theta_i} = -\sum_{l,k} \frac{\partial P_{\Omega}}{\partial \theta_i} \left(1 + \log\left(p(l,k)\right)\right). \quad (8)$$

Since $(1 + \log(p(l,k)))$, preserving the global intensity linkage, is constant for the transformation parameters, the computation complexity of (8) is only $O(|U_i \cap V|)$, which compares with $O(|V|)$ for (7), where V is the volume of the reference image. In the whole heart segmentation experiments in Section III, we empirically compute U_i as $U_i = \oplus_\rho V_i$ where $\rho = 20$ mm. (8) not only accelerates the runtime, but also blocks driving forces coming from the background such as the liver and abdomen, where the shape and intensity across subjects vary more significantly. Fig. 1 outlines the framework of the proposed LARM and describes the pseudo-code of our implementation.

B. Adaptive Control Point Status Free-Form Deformations

Since LARM provides an initialization for substructures, a refinement is needed to achieve an accurate registration. This step can be done by a nonrigid registration with high DOFs, such as using FFDs [16]. Also, FFD registration can be diffeomorphic by using a concatenation scheme introduced in [12].

Schnabel *et al.* proposed to associate a status, active or passive, to each control point of a FFD mesh to simulate the nonuniform rational *B*-splines (NURBS) [22]. Such status is priorly computed before optimizing the FFDs, using either the *reference image measures* such as local entropy, or the *joint image pair measures* such as the gradient of the associated cost function. Also, the status stays constant during the registration process. This method, embedded into a multiresolution registration scheme, improved the performance and significantly reduced the runtime compared to a standard FFD registration scheme [22].

However, there are two potential challenges in terms of applying it to cardiac MR image registration.

Firstly, in the context of cardiac MR image segmentation, the background of the image includes the adjacent tissues of the heart such as the lung and the liver which are not of interest for segmentation. The background should not deteriorate the registration process such as reducing the accuracy. However, using the status setting in [22], these areas will be taken into account, because they present the anatomical information of the background for the status determination measures.

Secondly, if the status is preset and remains constant, some control points could be inactivated, leading to significantly decreasing the modeling ability of the FFDs. For example, as shown in Fig. 2, the control points around the contour of the ellipse [white dots in middle row of (a)] will not be activated by using the proposed measures in [22]. This is because the local supports of these control points are mapped to the regions with uniform intensity values in the floating image, where the intensity gradient is low. Only the control points mapped to the boundary of the circle in the floating image will be activated, as the black dots shown in Fig. 2(b). If the status is not adaptively updated, the ellipse in the reference image will not be able to deform appropriately to match the circle in the floating image.

To account for these two challenges, we propose to set the status of control points adaptively for each registration iteration, i.e., adaptive control point status (ACPS). As shown in Fig. 2(c), after i iterations, the floating image I_s is inversely transformed into I_s^i, in the coordinate of the reference image space, referred to as *reference space*. Then, the status of the control points should be updated based on the updated information from I_s^i given the shape of the floating image is priorly known, as shown in Fig. 2(b). Based on this idea, we extend the status-setting as follows:

$$\text{Status}(\varphi_j) = \begin{cases} \text{active,} & \text{if } T_i(\phi_j) \in \mathcal{M}_s \\ \text{passive,} & \text{if } T_i(\phi_j) \notin \mathcal{M}_s \end{cases} \quad (9)$$
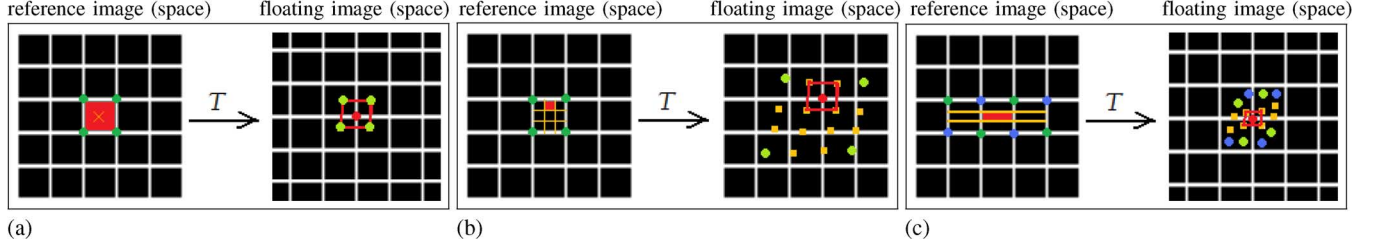
Fig. 3. DRAW using inverse distance interpolation from forwardly transformed scatter points (a); resampling more points within a pixel when the transformation field causes expansion (b); resampling more points in other directions when the transformation field causes anisotropic contraction in one direction (c).

where ϕ_j is the coordinate of control point φ_j and $\mathcal{M}_s$ is a mask image generated based on the prior knowledge. For example, to avoid the effects caused from the background, $\mathcal{M}_s$ is chosen to solely cover the region of interest of the registration.

Since the segmentation of the atlas is priorly known, this status-setting fits well into the registration task required in the refinement of the segmentation. The mask images, $\{\mathcal{M}_s\}$, are chosen to avoid activating control points in the adjacent tissues with indistinct boundaries, leading to an improvement of the registration robustness. Also, similar to LARM, the registration using ACPS FFDs can significantly reduce the computation time, because the optimization does not compute the driving forces (derivatives of cost function) for a large number of passive control points locating in background areas.

C. DRAW: Dynamic Resampling and Distance Weighting Interpolation

LARM requires a predefinition of the local regions in the reference image. However, in this segmentation propagation framework, the prior definition is only available from the floating image, the atlas. In this case, a method is required to compute an inverse transformation from the result of the LARM process where the atlas is defined as the reference image. Currently reported methods have difficulties in controlling the maximal error when the deformation field contains large displacements [28]. In this section, we propose a new algorithm for inverting general dense displacements, based on Dynamic Resampling And distance Weighting interpolation (DRAW) [20].

Let T be a diffeomorphic transformation, mapping a coordinate t in the reference image, reference space, to a coordinate s in the floating image, i.e., $s = T(t)$. To compute the inverse transformation T^{-1} of T, one can directly interpolate $t = T^{-1}(s)$ for s using the *scatter points* [28]. Given $\{s_i\}$, a set of close scatter points of s as the four green dots shown in Fig. 3(a) right, the interpolated inverse transformation $T_I(s)$ is computed as

$$T_I(s) = s + \sum w(s_i) \cdot (t_i - s_i) \tag{10}$$

where $\{s_i\}$ are transformed by T from the point set $\{t_i\}$ in the reference space and $w(s_i)$ is the normalized weighting function computed from the distance between s_i and s [refer to (2)]. It is assumed that an accurate estimation can be achieved given that the two conditions stand (for simplicity *pixel* is used for both 2-D and 3-D image elements).

- Condition 1: One closest point s_i can be found from each octant (3-D) or quadrant (2-D) of s within one pixel size.
- Condition 2: The $\{s_i\}$ should be transformed from the scatter points $\{t_i\}$ which are all within the volume of one pixel size.

As shown in Fig. 3(a), the inverse transformation of point s, the red dot in the floating image, can be interpolated from its neighboring green dots $\{s_i\}$ that are transformed from the green dots $\{t_i\}$ in the reference image. In a diffeomorphic transformation field T, the ground truth $T^{-1}(s)$ and the interpolated inverse transformation $T_I(s)$ should be both within the volume enclosed by $\{t_i\}$, the red area in Fig. 3(a). Therefore, the interpolation error $|T_I(s) - T^{-1}(s)|$ should not be larger than the diameter of the enclosed volume. However, two situations make it difficult to meet these two conditions.

Firstly, if T contains dilation then the scatter points $\{s_i\}$ can be too distant from s as shown in Fig. 3(b). When T causes dilation at a volume enclosed by vertex points $\{t_v\}$ (green dots in the reference image), $\{t_v\}$ will be sparsely transformed into a bigger area V_s in the floating image space, where a grid point s lies. In this case, one may not be able to find any of the sparse points required in Condition 1 for the interpolation of $T_I(s)$. To obtain this point set, we propose to resample more points within v_t and transform them to the floating image space, as shown in Fig. 3(b). The number of resampled points, $S_{\hat{e}}$, is estimated as

$$S_{\hat{e}} = \max(1, |J : \hat{e}|) \tag{11}$$

where J is the Jacobian matrix of T and $|J : \hat{e}| = |\hat{e} \times J \times \hat{e}|$. $|J : \hat{e}|$ represents the volume change in the direction of $\hat{e}$, where $\hat{e}$ is the standard $\Re^d$ Euclidean basis.

Secondly, if T contains anisotropic contractions, then the scatter points $\{t_i\}$ may not be within one pixel size in the reference image space, as shown in Fig. 3(c). When transformation T causes local anisotropic contractions, such that $|J : \hat{e}_j|/|J : \hat{e}_k| > 1$, the closest scatter points $\{s_i\}$ of s and their corresponding points $\{t_i\}$ would be these green dots shown in Fig. 3(c). This however violates Condition 2, because $\{t_i\}$ are not within one pixel size. To improve this situation, we propose to resample more points in the other orthogonal directions inside v_t to provide a closer scatter point set $\{s_i\}$ and a corresponding point set $\{t_i\}$ which are from the area within a pixel size
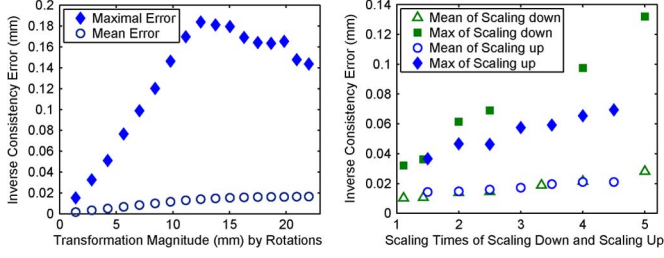
Fig. 4. Inverse-consistency errors of DRAW in generic displacement fields (left) and anisotropic scalings (right).

Fig. 5. The framework of automatic whole heart segmentation based on atlas propagation.

$$S_{\hat{e}_j} = \prod_k \max\left(1, |J:\hat{e}_j|/|J:\hat{e}_k|\right), \quad j, k = 1, 2, \ldots, d. \tag{12}$$

To combine (11) and (12), one can get the number of resample points needed from the volume of a pixel

$$S_{\hat{e}_j} = \max\left(1, |J:\hat{e}_j|\right) \cdot \prod_k \max\left(1, |J:\hat{e}_j|/|J:\hat{e}_k|\right),$$
$$j, k = 1, 2, \ldots, d. \tag{13}$$

To assess the accuracy of DRAW, we used an image with an isotropic voxel size of 0.5 mm to generate displacement fields. We first created a small local rigid box region at the center of the image which was rotated. The boundaries of the image were fixed. By rotating the box from 5° to 85° using (1) for interpolation, we generated a set of generic displacement fields. Then, we generated a set of anisotropic-scaling transformations by scaling up the image along one direction (X axis) from 1.5 to 4.5 times (the magnitudes of transformations are from 6 to 36 mm) and scale down (along X axis) the image from 1/0.9 to 1/0.2 times (3 to 8 mm). Inverse-consistency (IC) $(|T(T_I(s)) - s|)$ was computed as the error. Fig. 4 (left) shows the IC errors of inverting the first set of displacement fields with respect to their magnitudes, while Fig. 4 (right) gives the errors of inverting the scaling fields. All the maximal IC errors in the two studies are subvoxel (less than 0.5 mm).

D. Segmentation Framework and Choice of Atlas

Segmentation Framework: The segmentation framework, registering the atlas to unseen images for segmentation propagation as illustrated in Fig. 5, includes three registration steps as follows.

- Firstly, a global affine registration is applied to localize the heart. This localization is generally a challenging step to achieve full automation in cardiac MR segmentation [9], [29]. The affine registration worked well here because the MR data were acquired with similar orientations and fields-of-view. Fig. 6 (left) shows an example image. In the affine registration, we first employed the whole information of the images for a coarse localization, including the surrounding structures of the heart such as the back, chest, and liver tissues. Then, we used a mask covering the region within 20 mm to the heart of the atlas to focus on registering the heart structure.
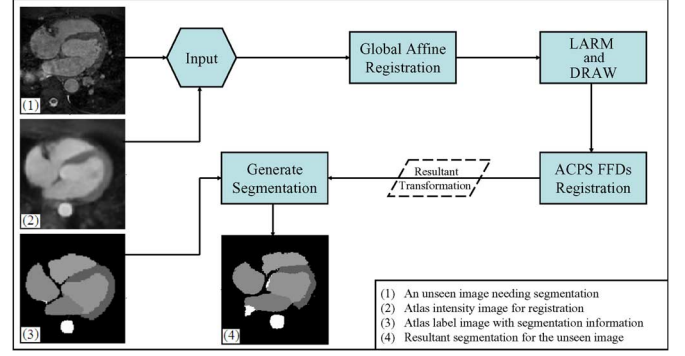
- Then, LARM is used to further initialize the substructures. Three stages of LARM registration are adopted in practice, where two, four, and seven local affine transformations are used, respectively.

  In the first stage, only two local affine transformations are used. The corresponding local regions are *atrial region* which includes the two atria and great vessels, and *ventricular region* which includes the two ventricles and myocardium. In the second stage, the right ventricle and the right atrium are separated from the regions defined in the first stage to include another two affine transformations. Finally, seven local regions, including two ventricles, two atria, pulmonary artery, ascending aorta and aortic arch, and descending aorta, are used to achieve a seven-local-affine LARM.

  As the atlas image needs to be defined in the reference image in LARM, the resultant transformation of LARM is inverted using DRAW to get the transformation defined from the unseen MR image to the atlas.

- Finally, registration using ACPS FFDs is employed to refine the local detail. In the status-setting, two mask images, $\mathcal{M}_s^1$ and $\mathcal{M}_s^2$, both priorly constructed from the atlas, are used.

  Firstly, the registration uses $\mathcal{M}_s^1$ to solely activate the control points which map onto the blood pool of the atlas. This registration mainly contributes to the alignment of the endocardial surfaces where the boundaries are normally distinct, and the FFDs use isotropic *B*-spline meshes with a spacing of 20 mm.

  Secondly, the registration employs $\mathcal{M}_s^2$ to activate the control points which map to a neighboring region of the endocardial and epicardial surfaces. This step fine-tunes the registration of whole heart surfaces and uses finer *B*-spline meshes with a spacing of 10 mm.

The registration implementation employs NMI [15] as the similarity measure. The optimization adopts the gradient ascent scheme with decreasing step length [30] and the multiresolution scheme [31]. The registration stops when the NMI measure has not be improved in the last five steps.

Atlas: Atlas construction is a different topic outside the scope of this work, especially when involving the statistical shape modeling [32]. In this work, we use a *simple atlas*, without statistical shape information, because it is assumed that the ex-
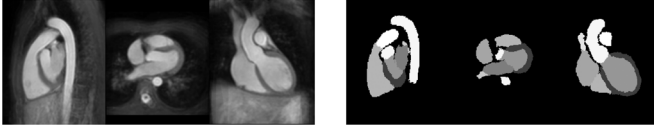
Fig. 6. Left: an atlas intensity (MR) image using a reference space of the mean shape of 10 volunteer data. Right: corresponding segmentation, atlas label image of the atlas intensity image. Images are shown in sagittal, transverse, and coronary views.

isting statistical shape models could not perform better than a simple atlas when applied to new pathologies. As the goal of this work is to develop a tool which is applicable to all pathologies, a simple atlas is employed to test the performance of the proposed approach. This atlas can be built from a selected reference space such as the mean of a group of images or a space defined by any single cardiac MR case.

Theoretically, one can use a MR image and its corresponding segmentation as the atlas for propagation. However, in practice the noise and artifacts presented in a specific case can lead to segmentation bias in all cases. To avoid this, we employ a number of MR images acquired using the same MRI sequence. They are then registered to a selected reference space. A mean intensity image, referred to as *atlas intensity image*, can be computed from this set of registered MR images. The labeling of each anatomical region of the reference space has the corresponding segmentation information of the atlas, referred to as *atlas label image*. Fig. 6 shows an atlas case that will be used in our experiments.

To achieve the registration of this set of MR images, we first label each anatomical substructure of the images, including the blood pools of each chamber and great vessel and left ventricle myocardium. Fig. 6 (right) shows a label image which is the label image of an atlas. The label images are not only the corresponding segmentation of the MR images, but also with a uniform intensity value for each substructure. The deformation fields to register the MR images can be computed from the registration results of the corresponding label images by using the following steps.

- A global affine registration to localize the heart structure.
- LARM to further initialize the seven substructures.
- A fluid registration using the sum squared difference of intensity as the similarity measure [11] to finally fine-tune the local detail.

III. EXPERIMENTS

Three steps have been done to demonstrate the performance of the proposed segmentation framework. Section III-A describes the data and the experimental setup. Section III-B compares the segmentation results using three different atlases for propagation. Section III-C demonstrates the improvements by using the proposed algorithms, compared with the alternative techniques. Section III-D analyzes the segmentation performance of the proposed segmentation framework and performs the validation in detail.

TABLE I
PARAMETERS OF THE MRI SEQUENCE

Field of view	$\sim 280 \times 240 \times 140$ mm
Acquired resolution	$\sim 2 \times 2 \times 2$ mm
Reconstructed resolution	$\sim 1 \times 1 \times 1$ mm
No. of slices	110 - 180
TR/ TE	~ 3.8 ms/ ~ 1.9 ms
TFE factor	15 - 20
Sense	2 anterior posterior 1-1.5 right left
Flip angle	~ 60 degrees
No. of phases	1 (ED phase) or 2 (dual phases)
Trigger modality	Trigger

A. Experimental Setup

Data: The cardiac MRI sequence used in the experiment was the balanced steady state free precession (b-SSFP) for whole heart imaging [33]. The imaging was either two (dual phases) or one (end diastolic phase) trigger delay(s). A 3-D triggering b-SSFP turbo field echo (TFE) sequence, with arrhythmia rejection, was modified in order to enable the acquisition of selected cardiac phases at a user defined time. The sequence was implemented on a 1.5 T clinical scanner (Philips Healthcare, Best, The Netherlands) equipped with 32 independent receive channels. For both cardiac phases a fat saturation and T2 preparation pulses (14) were used to null fat and to increase the contrast between blood and cardiac muscle. A free breathing scan was realized by enabling one navigator beam before data acquisition for each cardiac phase. The detail parameters are presented in Table I. The application of this sequence and comparisons to other routinely used MRI sequences are available in [34].

A *test dataset* of 37 cardiac MR volumes on the end-diastolic phase is used as unseen images in our experiments, among which nineteen cases have confirmed pathologies including myocardium infarction, atrial fibrillation, tricuspid regurgitation, aortic valve stenosis, Alagille syndrome, Williams syndrome, dilated cardiomyopathy, aortic coarctation, and Tetralogy of Fallot which induces a number of different forms of morphological abnormalities. It is common to see that many of these patients had symptoms of a combination of more than one of these pathologies. The subjects, aged from 5 to 80, displayed a wide diversity of heart shapes. To avoid the advantageous bias of using an atlas with similar heart shape to segment the unseen data, we acquired a *training dataset* of images from 10 separate healthy volunteers for the construction of the atlas.

All these data have manual segmentation as the gold standard. They were either done by fitting a deformable mesh model [9] with manual corrections or using the semi-automatic editing tool in the commercial product Analyze (Mayo Clinic). The manual segmentation was completed by either a clinician or a research associate with knowledge of heart anatomy. For the test dataset, the blood cavities of the four chambers and the myocardium of the left ventricle are separately segmented (examples of manual segmentation can be found in Figs. 8 and 9), while for the training dataset, in addition to these regions, the pulmonary artery, ascending aorta and aortic arch, and descending aorta are also separately delineated for the atlas construction registration. Fig. 6 (right) shows an example of these substructure labels.

Evaluation Protocols: Two different types of measures can be used to evaluate the accuracy of a segmentation result: the surface distance measures and the volume measures.

The surface measures compute the surface-to-surface distance between the propagated segmentation and the gold standard segmentation. This surface distance is calculated using each surface point s_{seg} from the propagated segmentation to the triangle defined by the three closest surface points $\{s_{gd}^1, s_{gd}^2, s_{gd}^3\}$ of s_{seg} in the gold standard segmentation. The mean surface distance, ϵ_{mean}, standard deviation, ϵ_{std}, and the rms, ϵ_{rms} of a surface will be computed. Six surfaces will be considered.

1) The endocardial surface of the left ventricle, referred to as *Left Ventricle*.
2) The endocardial surface of the left atrium, referred to as *Left Atrium*.
3) The endocardial surface of the right ventricle, referred to as *Right Ventricle*.
4) The endocardial surface of the right atrium, referred to as *Right Atrium*.
5) The epicardial surfaces of the left ventricles, referred to as *Epicardium*.
6) The inclusion of all the five surfaces above as a whole heart segmentation accuracy, referred to as *Whole Heart*.

Volume measurement is of major interest in many clinical applications such as computing stroke volume, ejection fraction and myocardium mass. Three widely used measures will be employed in our validation, including:

- dice coefficient $= 2|V_{\text{seg}} \cap V_{gd}|/(|V_{\text{seg}}| + |V_{gd}|)$ [35];
- volume overlap $= |V_{\text{seg}} \cap V_{gd}|/|V_{\text{seg}} \cup V_{gd}|$ [36];
- volume difference $= 2|V_{\text{seg}} - V_{gd}|/(|V_{gd}| + |V_{\text{seg}}|) \times 100\%$;

where V_{seg} and V_{gd} denote the volumes from the propagated segmentation and gold standard, respectively. Both Dice and volume overlap indicate how well the two volumes are overlapped, with maximum 1 meaning perfect overlap and minimum 0 indicating no overlap at all. Five substructures will be measured, as follows.

1) The blood cavity of left ventricle, referred to as *Left Ventricle*.
2) The blood cavity of left atrium, referred to as *Left Atrium*.
3) The blood cavity of right ventricle, referred to as *Right Ventricle*.
4) The blood cavity of right atrium, referred to as *Right Atrium*.
5) The myocardium of left ventricle, including septum, referred to as *Myocardium*.

Three Steps of Evaluation:

Step-one: This step uses the 19 pathological data as a test dataset to assess the sensitivity of the proposed approach to different atlases. Section II-D provides detail of the atlas construction. Three atlases are tested, whose reference shapes are based on the following.

- A healthy volunteer is randomly selected from the training dataset as the reference shape, referred to as *One Shape*.
- The mean shape of five randomly selected training data is used as the reference shape, referred to as *Five Shapes*. The mean of five shapes was computed based on the *iterative atlas construction* [32]. The registration in our construction
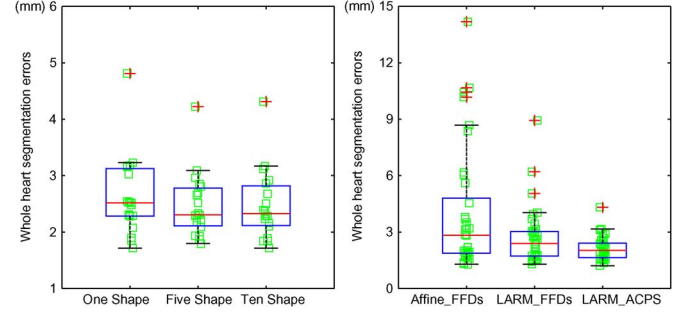

Fig. 7. The individual plots and the Box-and-Whisker diagrams of the *Whole Heart* segmentation errors using the rms surface-to-surface error measure, ϵ_{rms}. Left: the errors of the 19 pathological cases using the proposed segmentation approach combined with the three different atlases. Right: the errors of the 37 cases using the three different segmentation frameworks.

employs the five corresponding *label images* and the three transformation models introduced in the atlas construction, described in Section II-D.

- The mean shape of all the ten volunteer data is used as the reference shape, referred to as *Ten Shapes*. Fig. 6 shows the result of this atlas.

Step-two: This step assesses the improvement of the proposed two registration algorithms when applied to the segmentation framework. The atlas used is the *Ten Shapes*. Three registration frameworks, for the segmentation propagation, will be used to segment the test dataset of 37 cases, including the following.

- A single affine for global localization and a traditional FFDs [12] for refinement, referred to as *Affine_FFDs*.
- A single affine for global localization, then LARM for substructure initialization, finally traditional FFDs for refinement, referred to as *LARM_FFDs*.
- The proposed framework, a single affine for global localization, then LARM for substructure initialization, finally a registration using ACPS FFDs for refinement, referred to as *LARM_ACPS*.

All FFD registration used the multiresolution meshes [22], spacing from isotropic 20 mm to 10 mm, and the concatenation strategy [12] to maintain the diffeomorphism.

Step-three: Finally, we focus on the segmentation results of the 37 test data using the proposed segmentation framework and the atlas of *Ten Shapes*.

B. Sensitivity to Different Atlases

Fig. 7 (left) plots the rms surface distance errors of all the 19 pathological cases. The mean and standard deviation of the corresponding errors are as follows:

- using *One Shape* atlas: 2.63 ± 0.70 (mm);
- using *Five shape* atlas: 2.47 ± 0.57 (mm);
- using *Ten shape* atlas: 2.47 ± 0.61 (mm).

Both the Box-and-Whisker diagram and figures from mean and standard deviation show that the segmentation using the atlas of *Five Shapes* had the best result while using the atlas *One Shape* had the worst. However, the differences between the mean and median numbers are all less than 0.2 mm. To assess the significance of the performance difference, the two-tail, paired student t-test [37] was used: $P_{1,5} = 0.068$ (between

TABLE II
SURFACE-TO-SURFACE SEGMENTATION ERRORS IN MILLIMETERS, ϵ_{rms} OF THE THREE METHODS AND ERRORS ϵ_{mean}, ϵ_{std}, AND PERCENTAGE WITH DIFFERENT RANGES: <2 mm, 2–5 mm, AND >5 mm, OF THE 37 CASES BY THE PROPOSED APPROACH

(error in mm) Structures	Affine_FFDs ϵ_{rms}[max]	LARM_FFDs ϵ_{rms}[max]	LARM_ACPS in different measures (ranges are of unit mm)					
			ϵ_{rms}[max]	ϵ_{mean}	ϵ_{std}	$0-2$	$2-5$	>5
Left ventricle	3.79 ± 4.38[**19.6**]	1.89 ± 1.10[**7.12**]	1.47 ± 0.32 [**2.38**]	1.06	1.02	82.9%	16.5%	0.6%
Left atrium	3.26 ± 2.25[**11.5**]	2.81 ± 1.65[**9.28**]	2.38 ± 1.14 [**7.33**]	1.69	1.68	70.2%	23.5%	6.4%
Right ventricle	3.23 ± 1.79[**7.50**]	2.75 ± 1.51[**8.57**]	2.13 ± 0.70 [**4.05**]	1.50	1.51	72.2%	24.1%	3.7%
Right atrium	3.06 ± 1.74[**9.44**]	2.51 ± 1.20[**7.05**]	2.22 ± 0.75 [**5.74**]	1.51	1.62	73.0%	22.5%	4.5%
Epicardium	4.78 ± 4.88[**20.4**]	2.88 ± 2.12[**12.9**]	2.32 ± 0.82 [**5.43**]	1.69	1.60	68.6%	25.7%	5.6%
Whole heart	3.96 ± 3.23[**14.2**]	2.71 ± 1.50[**8.93**]	2.14 ± 0.63 [**4.31**]	1.47	1.55	73.6%	22.5%	3.9%

the *One Shape* and the *Five Shapes*), $P_{1,10} = 0.057$ (between the *One Shape* and *Ten Shapes*), and $P_{5,10} = 0.95$ (between the *Five Shapes* and the *Ten Shapes*). There was no strong evidence to show significant difference between each population using a significance level of 0.05. However, both $P_{1,5}$ and $P_{1,10}$ were also close to the significance level. We then further computed the 0.95 confidence interval of them, which were $CI_{1,5} = [0.01, 0.32]$ mm and $CI_{1,10} = [0.01, 0.31]$ mm respectively. This showed there were small biases of the *One Shape* population with respective to the other two, but these biases were not practically important. In addition, the Pearson correlation coefficients are $R_{1,5} = 0.94$, $R_{1,10} = 0.93$, and $R_{5,10} = 0.98$, respectively. The errors of each category are assumed to have a Gaussian distribution in the t-test.

In conclusion, the atlas constructed based on the shape of one subject could lead to a performance bias, but no evidence showed that the segmentation of pathological data could be significantly different by using an atlas from the mean shape of a training set that has a larger number of healthy volunteer data.

C. Performance Using Alternative Techniques

Table II displays the rms surface-to-surface segmentation errors of each substructure and all the surfaces as the *Whole Heart* measure of the three different registration frameworks. From the *Affine_FFDs* to the *LARM_FFDs*, the errors in all categories were improved, especially the maximal errors. This segmentation was furthermore improved by the *LARM_ACPS*. The P-values of the *Whole Heart* rms surface distance errors using the one-tail, paired t-test, assuming nonimprovement:

- test the *Affine_FFDs* against the *LARM_FFDs*: $P_{larm} = 0.002$;
- test the *LARM_FFDs* against the *LARM_ACPS*: $P_{acps} = 0.002$;
- test the *Affine_FFDs* against the *LARM_ACPS*: $P_{larm+acps} = 0.0002$.

These P-values suggested strong evidence to reject the null-hypothesis, thus the segmentation had been improved by the two proposed techniques with statistical significance. This also agrees with the mean and standard deviation of the ϵ_{rms} in Table II where the *Whole Heart* of *LARM_ACPS*, 2.14 ± 0.63 mm, had been improved about three times of the standard deviation (0.63) from that of *Affine_FFDs*, 3.96 ± 3.23 mm.

Furthermore, Fig. 7 (right) plots the errors of the *Whole Heart* measure. About two thirds of the segmentation results by the three methods had similar error figures, while the "bad" cases differed significantly: By using LARM, the four outliers in the *Affine_FFDs* were improved to less than 9 mm in the

LARM_FFDs, while by further using ACPS FFDs, the three outliers in the *LARM_FFDs* were improved to less than 4.5 mm.

Finally, Fig. 8 displays the three worst segmentation cases of the three segmentation methods and their corresponding segmentation results by these three methods and the gold standard segmentation:

- Subject-1, with a *Whole Heart* error of 14.2 mm, is the worst case of the *Affine_FFDs*. This was from a patient with severe right ventricle and atrium hypertrophy. A global affine registration did not well initialize the substructures and this induced a misalignment of the two ventricles: a large part of the right ventricle was segmented as the left ventricle. However, LARM well compensated this problem and corrected the segmentation of the two ventricles as the results of the *LARM_FFDs* and *LARM_ACPS* shown in Fig. 8.
- Subject-116, with a *Whole Heart* error of 8.93 mm, is the worst case of the *LARM_FFDs*. This was from a healthy volunteer case, but the image quality was very low. The traditional FFDs without constraints from prior knowledge produced an unrealistic deformation field and resulted in an erroneous result, while the ACPS FFDs contributed to reducing this error.
- Subject-9, with a *Whole Heart* error of 4.31 mm, is the worst case of the *LARM_ACPS*. This case was also the worst result of the *LARM_ACPS* in terms of the error measure for the left atrium (7.33 mm), right ventricle (4.05 mm), and right atrium (5.74 mm). The large error of right ventricle was mainly due to the inconsistent delineation of the cutting boundaries of chordae tendineae, papillary muscle, and trabeculae carnae between the gold standard and propagated segmentations. The error of atria was attributed to the fact that the propagated segmentation included parts of adjacent veins into the volume of atria, especially the left atrium, while by contrast, the gold standard segmentation did not include them.

D. Performance of the Proposed Framework

Table II lists the surface-to-surface errors of the proposed segmentation method for each surface category. The rms error (and its maximum in the 37 cases), the mean error, the standard deviation, and the percentage of error ranges are presented. For an average whole heart segmentation, more than 95% surface-to-surface error is within 5 mm.

Fig. 10 plots all the errors and Box-and-Whisker diagrams of the 37 cases. It agrees with the results in Table II that the
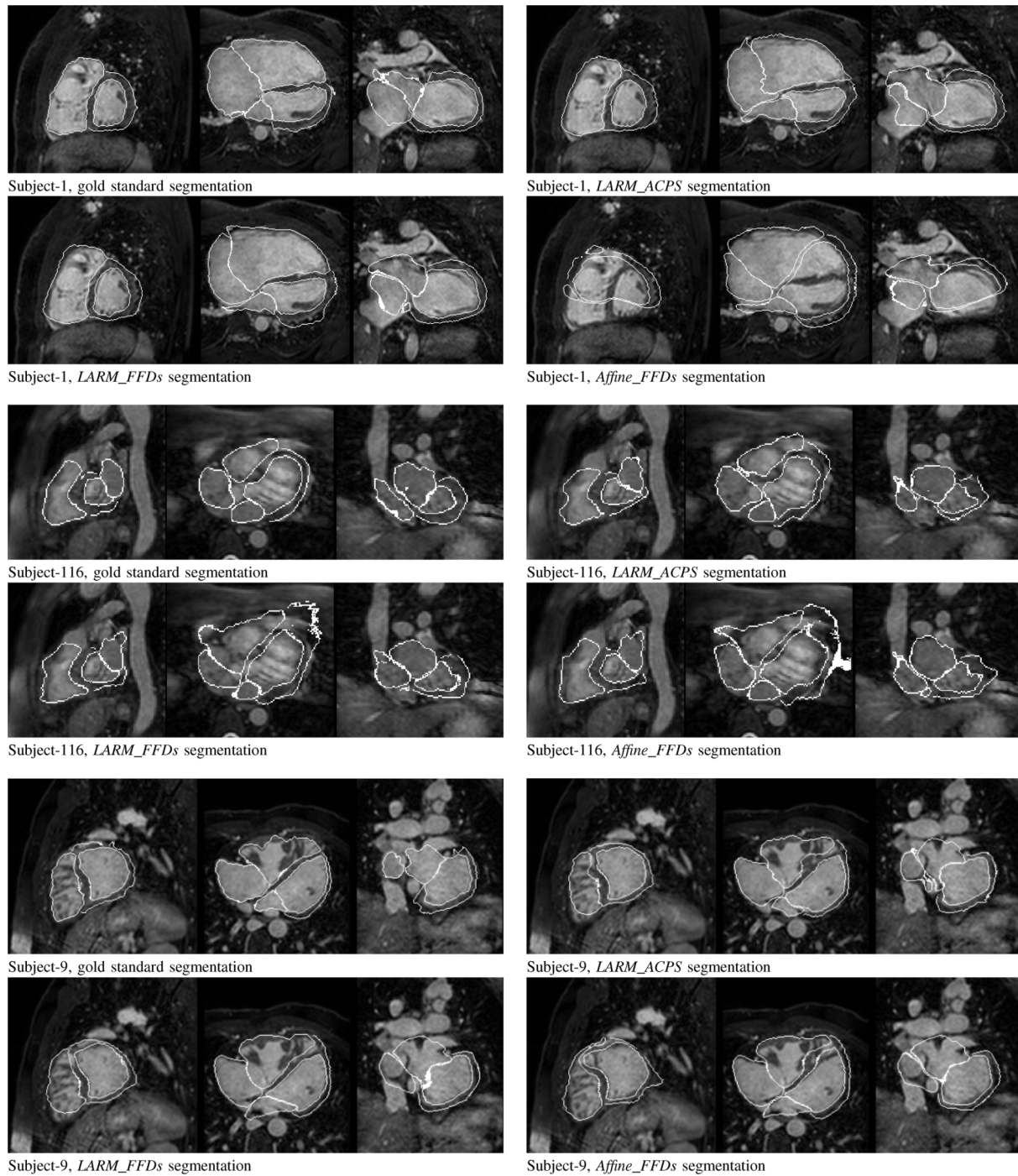
Fig. 8. The three worst cases by the three segmentation methods. Subject-1 is the worst case of the *Affine_FFDs*, subject-116 is the worst case of *LARM_FFDs*, and subject-9 is the worst case of *LARM_ACPS*. Images are displayed with delineated contour superimposing on the MR images, in sagittal, transverse, and coronary views. Subject-1 and subject-9 are pathological cases while subject-116 is a healthy case.

segmentation of the *Left Ventricle* was better than other surface categories. The worst case of the *Left Ventricle* was subject-119, from a healthy volunteer, with a rms surface error of 2.38 mm; the worst case of the *Epicardium* was subject-43, from a Tetralogy of Fallot patient, with a rms surface error of 5.43 mm. The segmentation by the proposed method and the gold standard segmentation are shown in Fig. 9. The worst segmentation of other surface categories was from one case, subject-9, shown in Fig. 8.

Finally, Fig. 11 shows the color map of surface-to-surface errors of the whole heart segmentation. One can realize from this map that the big errors were mainly distributed to the area of connections between substructures. For example, the area between the right ventricle and pulmonary artery was one of the worst regions. Similar situation was found in the areas between the left atrium and pulmonary veins and left auricula. The proposed method also had a moderate error range in a small part of the endocardium of the right ventricle. This was because there

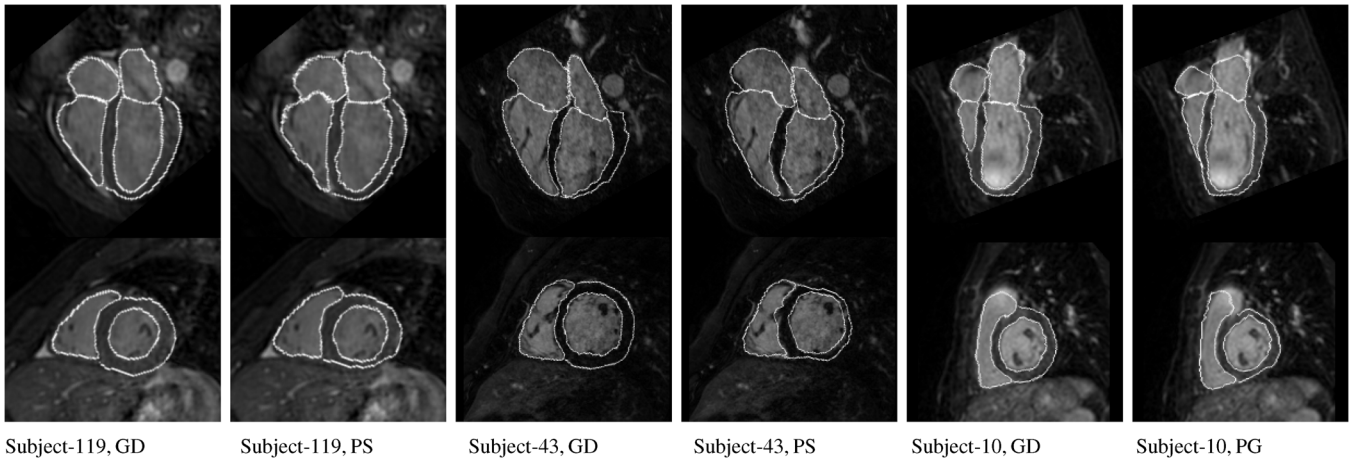| Subject-119, GD | Subject-119, PS | Subject-43, GD | Subject-43, PS | Subject-10, GD | Subject-10, PG |

Fig. 9. Three segmentation cases. Images are displayed with delineated contour superimposing on the MR images, in four-chamber (top) and short-axis (bottom) views. Subject-119 is a healthy case while subject-10 and subject-43 are pathological cases. The *Whole Heart* segmentation errors (ϵ_{rms}) for each subject are 2.39 mm, 3.17 mm, and 2.86 mm, respectively. GD: gold standard segmentation; PS: propagated segmentation.
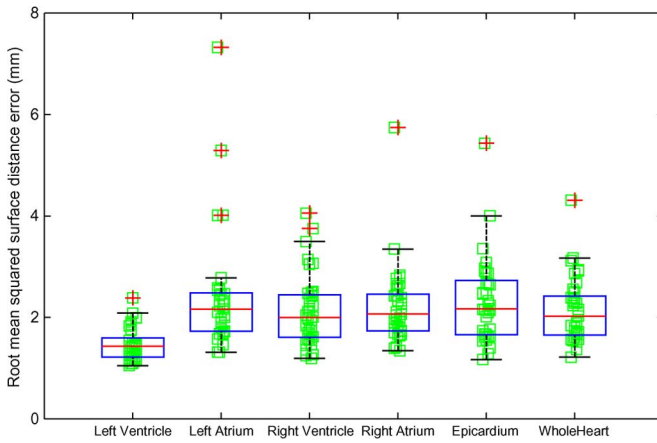

Fig. 10. The individual plots and the Box-and-Whisker diagrams of the segmentation errors using the rms surface distance measure, ϵ_{rms}. This figure gives the errors of each substructures as well as the *Whole Heart* from the 37 cases, by the proposed segmentation framework.
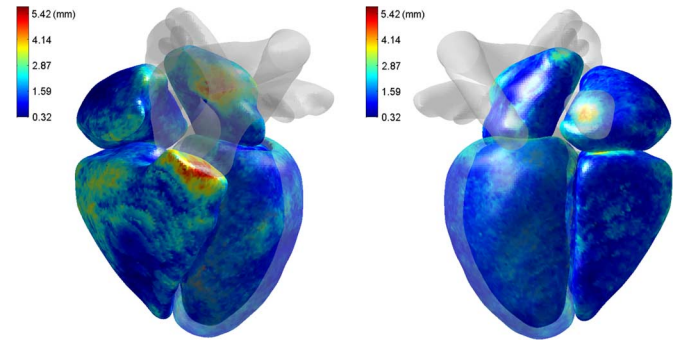

Fig. 11. Two views showing the error map of surface-to-surface distance for the whole heart segmentation by the proposed method. This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. Please refer to the web version of this article for interpretation of the color map.

were many pathological cases which developed thick right ventricle myocardium, while the atlas image did not include such myocardium. The segmentation therefore could wrongly segment part of the myocardium into the right ventricle blood cavity due to this ambiguity. Examples of this error can be found in cases of subject-1 and subject-9 in Fig. 8.

Table III presents the mean segmentation errors and their worst cases in volume measures. The average of all the substructures were 0.84 ± 0.05 (Dice measure), 0.73 ± 0.07 (volume overlap), and $(9.1 \pm 7.2)\%$ (volume difference), respectively. The worst cases for each substructure using Dice measure were subject-116 (*Left Ventricle*), subject-9 (*Left Atrium*), subject-10 (*Right Ventricle* and *Right Atrium*), and subject-43 (*Myocardium*). The segmentation of subject-116 and subject-9 are displayed in Fig. 8 and that of subject-10 and subject-43 are shown in Fig. 9.

Table III shows no significant difference between the volumes of the gold standard and the propagated segmentation for the *Left Atrium* ($P = 0.667$), the *Right Atrium* ($P = 0.217$),

the *Myocardium* ($P = 0.940$). However, the difference was found for the *Left Ventricle* ($P = 0.028$) and the *Right Ventricle* ($P = 0.048$). This was due to biases of the segmented volumes, shown in their confidence intervals. Fig. 12 shows the Bland–Altman plots of the volumes, which confirms that both the *Left Ventricle* and the *Right Ventricle* had segmentation biases of 4.3 mL and 6.2 mL, respectively. The larger bias of *Right Ventricle* was partly due to the inconsistent definition of the valves between the gold standard and the propagated segmentation, as the *Right Atrium* had -2.6 mL segmentation bias. These biases are practically negligible in many applications, considering the ranges of their volumes were 67–204 mL and 65–488 mL. The bias for all the segmentation of substructures was 1.70 mL. Finally, their Pearson correlation coefficients are also given in Table III.

IV. DISCUSSION AND CONCLUSION

The majority of previous studies in the literature reported mean surface distance (ϵ_{mean}) around 1.5–3.0 mm for ventricle segmentation. In recent works, Lynch *et al.* [38] achieved 1.25 ± 1.34 mm for ventricle segmentation of cine MR data using level-set. Koikkalainen *et al.* [8] investigated several artificial

TABLE III
SEGMENTATION ERRORS BY THE PROPOSED METHOD USING VOLUME MEASURES: DICE COEFFICIENT, VOLUME OVERLAP, PERCENTAGE OF VOLUME DIFFERENCE (DIFF), P VALUE AND 0.95 CONFIDENCE INTERVAL (CI), OF UNIT mL, AND PEARSON CORRELATION (R)

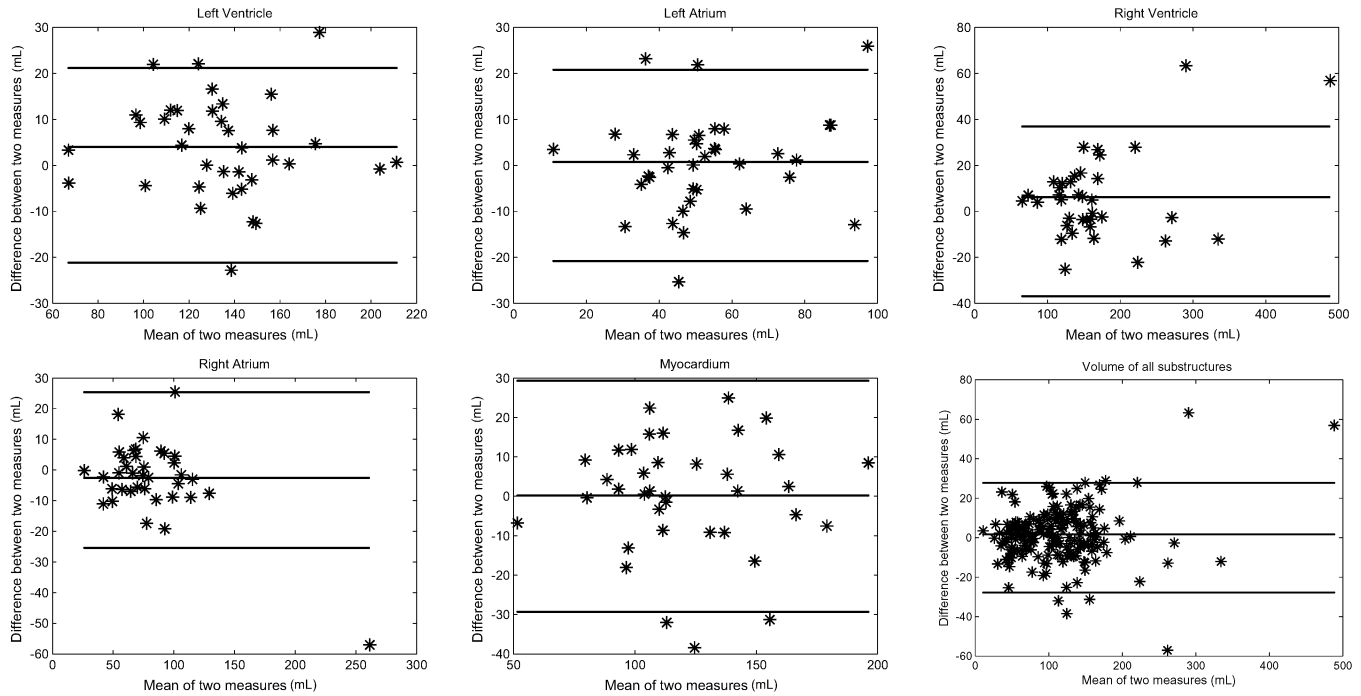| Structures	Dice [min]	Overlap [min]	Diff (%) [max]	P Value (CI mL)	Pearson R
Left ventricle	0.92 ± 0.02 [**0.87**]	0.85 ± 0.04 [**0.78**]	6.5 ± 4.9 [**19.0**]	0.028 $(0.5, 7.5)$	0.942
Left atrium	0.81 ± 0.10 [**0.47**]	0.69 ± 0.12 [**0.30**]	14.1 ± 12.0 [**48.4**]	0.667 $(-2.7, 4.2)$	0.861
Right ventricle	0.87 ± 0.04 [**0.77**]	0.77 ± 0.06 [**0.63**]	7.5 ± 4.9 [**19.7**]	0.048 $(0.1, 12.3)$	0.977
Right atrium	0.84 ± 0.05 [**0.71**]	0.73 ± 0.07 [**0.55**]	9.0 ± 7.2 [**28.7**]	0.217 $(-6.8, 1.6)$	0.965
Myocardium	0.77 ± 0.06 [**0.52**]	0.63 ± 0.07 [**0.35**]	8.6 ± 6.9 [**26.7**]	0.940 $(-4.7, 5.1)$	0.892

Fig. 12. Bland–Altman plots of the propagated segmented volumes by the proposed method and the gold standard segmentation volumes. Middle line is the bias (mean), upper and lower lines are 2 standard deviations.

enlargement methods of training sets for statistical shape modeling and compared their performance in whole heart segmentation of MR data. The best result reported was 2.06 ± 0.58 mm (mean surface distance). Cocosco *et al.* [39] validated their segmentation results using regression of segmented volumes and reported $R = 0.968$ (left ventricle) and $R = 0.969$ (right ventricle). Kurkure *et al.* [29] reported a mean Dice number of 0.855 ± 0.123 for ventricle segmentation from cine cardiac MR images. van Assen *et al.* [7] presented a semiautomatic method using a 3-D active shape model driven by fuzzy inference. They reported mean surface distance errors of 1.72 mm (endocardium) and 1.55 mm (epicardium) for ventricle segmentation of short-axis 3-D MR data from 15 healthy subjects. Finally, Peters *et al.* [9] achieved 0.76 ± 0.30 mean surface distance error for the whole heart segmentation, which is the best result reported to our knowledge. Their method was based on a deformable model trained with statistical shape information. However, the validation was not independent, as the meshes of gold standard segmentation came from the same mesh model of their segmentation tool with manual correction. Furthermore, the test data they used had the same or similar pathology. Therefore, an objective interwork comparison would be difficult.

We have presented a registration-based segmentation propagation framework and tested the performance using a dataset involving nine different pathologies. We showed that no significant difference was found by using the three atlases constructed from training sets of healthy subjects to segment the pathological cases. Additionally, we showed that the segmentation errors had been significantly reduced by the two proposed registration algorithms: the proposed locally affine registration method (LARM) for substructure initialization and the proposed free-form deformations registration with adaptive control point status (ACPS FFDs) for refinement of local details. The proposed segmentation framework chieved a rms surface-to-surface error of 2.14 ± 0.63 mm and a Dice measure of 0.84 ± 0.05 between the propagated segmentation and the gold standard segmentation.

There are three limitations of this work. Firstly, using the intensity-based registration it is difficult to consider the thin regions which can be critical in separating substructures in some cases. For example, the thin membrane between epicardium and liver or the thin atrial wall between the two atria can be displayed as a thin region whose intensity significantly differs from its neighboring regions. However, the intensity-based registration may fail to incorporate this information due to their relatively small sizes. Secondly, the propagation can have a big variation in the delineation of the valves. Finally, this approach takes relatively longer computation time to finish, 2–4 h per volume,
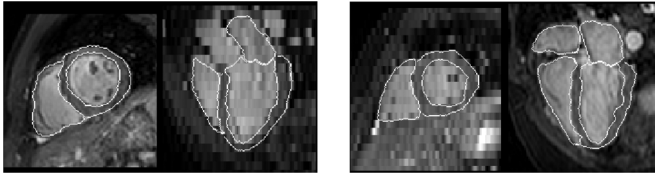
Fig. 13. The short-axis and long-axis views of two segmentation examples from a short-axis multislice image (left) and a long-axis multislice image (right).
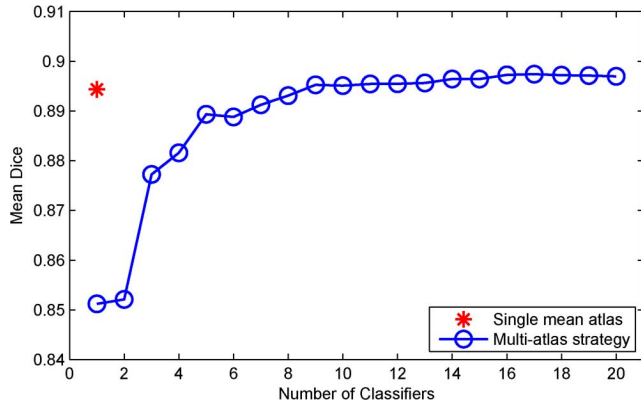

Fig. 14. Segmentation using single mean atlas and multiatlas propagation.

compared to the deformable model-based, boundary-searching techniques.

In our future work, compensative techniques will be considered to deal with these three problems. The boundary-searching technique will be incorporated to improve the first limitation. A flat plane positioned by some landmarks will be used to achieve the consistent definition of valves. For example, the plane defined by the basal myocardium of the left ventricle can define the mitral valve. For the run time, the parallel computation such as using graphic processing units will be used to accelerate the segmentation process.

A large number of cardiac MR segmentation studies in the literature were performed using the multislice images which are more routinely used for cardiology studies. Fig. 13 shows two examples of segmentation using the same tool in Section III-D on short-axis and long-axis data. In the future, we will investigate the segmentation of cine multislice MR data and validate the performance on cine data.

Multiatlas propagation and segmentation (MAPS) is a useful strategy [40] to improve the atlas-based segmentation. However, standard MAPS employs the original intensity images as the atlases for propagation, often with a selection strategy to use the best ranked atlases [40]. Artifacts in cardiac MR may create large errors to dramatically deteriorate the propagation registration performance. For example, Fig. 14 shows the mean Dice score of the five substructure segmentation using the proposed method and the standard MAPS. The test data were 21 cardiac MR images with all local regions segmented [Fig. 6 (right)] such that each of them can be used as an atlas for MAPS using LARM_ACPS registration for propagation. We employed the leave-one-out strategy where 20 MR images were used as the training data for computing the atlas intensity image in the proposed method or as the atlas set in MAPS. We

ranked the segmentation results using the NMI similarity of the region of interest after the propagation registration and fused the multiple labels using the "vote rule" [41]. Fig. 14 shows that MAPS achieved similar Dice scores as the proposed method after fusing more than 8 classifiers and the best result of MAPS was 0.8974, a very small gain of accuracy compared to 0.8943 of the proposed method. In the future, we will investigate techniques to maximize the gain of accuracy using multiple classifier strategy.

REFERENCES

[1] Cardiovascular diseases World Health Organization, Fact Sheet No. 317, Feb. 2007 [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/

[2] J. S. Suri, "Computer vision, pattern recognition and image processing in left ventricle segmentation: The last 50 years," *Pattern Anal. Appl.*, vol. 3, no. 3, pp. 209–242, 2000.

[3] J. Lötjönen, V. Jarvinen, B. Cheong, E. Wu, S. Kivistö, J. Koikkalainen, J. Mattila, H. Kervinen, R. Muthupillai, F. Sheehan, and K. Lauerma, "Evaluation of cardiac biventricular segmentation from multiaxis MRI data: A multicenter study," *J. Magn. Reson. Imag.*, vol. 28, pp. 626–636, 2008.

[4] M. R. Kaus, J. von Berg, J. Weese, W. Niessen, and V. Pekar, "Automated segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 8, no. 3, pp. 245–254, 2004.

[5] G. Hautvast, S. Lobregt, M. Breeuwer, and F. Gerritsen, "Automatic contour propagation in cine cardiac magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1472–1482, Nov. 2006.

[6] H. C. van Assen, M. G. Danilouchkine, A. F. Frangi, S. Ords, J. J. M. Westenberg, J. H. C. Reiber, and B. P. F. Lelieveldt, "SPASM: A 3-D-ASM for segmentation of sparse and arbitrarily oriented cardiac MRI data," *Med. Image Anal.*, vol. 10, pp. 286–303, 2006.

[7] H. C. van Assen, M. G. Danilouchkine, M. S. Dirksen, J. H. C. Reiber, and B. P. F. Lelieveldt, "A 3-D active shape model driven by fuzzy inference: Application to cardiac CT and MR," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 5, pp. 595–605, Sep. 2008.

[8] J. Koikkalainen, T. Tolli, K. Lauerma, K. Antila, E. Mattila, M. Lilja, and J. Lotjonen, "Methods of artificial enlargement of the training set for statistical shape models," *IEEE Trans. Med. Imag.*, vol. 27, no. 11, pp. 1643–1654, Nov. 2008.

[9] J. Peters, O. Ecabert, C. Meyer, R. Kneser, and J. Weese, "Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation," *Med. Image Anal.*, vol. 14, pp. 70–84, 2009.

[10] D. Rueckert, M. Lorenzo-Valdeés, R. Chandrashekara, G. I. Sanchez-Ortiz, and R. Mohiaddin, "Non-rigid registration of cardiac MR: Application to motion modelling and atlas-based segmentation," in *Proc. IEEE Int. Symp. Biomed. Imag. (ISBI)*, 2002, pp. 481–484.

[11] W. R. Crum, C. Tanner, and D. J. Hawkes, "Anisotropic multi-scale fluid registration: Evaluation in magnetic resonance breast imaging," *Phy. Med. Biol.*, vol. 50, no. 21, pp. 5153–5174, 2005.

[12] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers, "Diffeomorphic registration using B-splines," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2006, Proceedings*. New York: Springer, 2006, vol. 4191, Lecture Notes in Computer Science, pp. 702–709.

[13] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, pp. S61–S72, 2009.

[14] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, pp. 95–113, 2007.

[15] C. Studholme, D. L. G. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3-D medical image alignment," *Pattern Recognit.*, vol. 32, no. 1, pp. 71–86, Jan. 1999.

[16] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.

[17] M. Lorenzo-Valdes, G. I. Sanchez-Ortiz, R. Mohiaddin, and D. Rueckert, "Atlas-based segmentation and tracking of 3-D cardiac MR images using non-rigid registration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2002, Proceedings*, T. Dohi and R. Kikinis, Eds.   New York: Springer, 2002, vol. 2488, Lecture Notes in Computer Science, pp. 642–650.

[18] S. C. Mitchell, J. G. Bosch, B. P. F. Lelieveldt, R. J. van der Geest, J. H. C. Reiber, and M. Sonka, "Active appearance models: Segmentation of cardiac MR and ultra sound images," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 1167–1178, Sep. 2002.

[19] J. Lotjonen, S. Kivisto, J. Koikkalainen, D. Smutek, and K. Lauerma, "Statistical shape model of atria, ventricles and epicardium from short- and long-axis MR images," *Med. Image Anal.*, vol. 8, pp. 371–386, 2004.

[20] X. Zhuang, K. Rhode, S. Arridge, R. Razavi, D. Hill, D. Hawkes, and S. Ourselin, "An atlas-based segmentation propagation framework using locally affine registration—Application to automatic whole heart segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2008*, D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, Eds.   New York: Springer, 2008, 2008, vol. 5242, Lecture Notes Computer Science, pp. 425–433.

[21] X. Zhuang, K. S. Rhode, R. Razavi, D. J. Hawkes, and S. Ourselin, "Free-form deformations using adaptive control point status for whole heart MR segmentation," in *Functional Imaging and Modeling of the Heart, 5th International Conference, FIMH 2009*, N. Ayache, H. Delingette, and M. Sermesant, Eds.   New York: Springer, 2009, vol. 5528, Lecture Notes in Computer Science, pp. 303–311.

[22] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2001, Proceedings*.   New York: Springer, 2001, vol. 2208, Lecture Notes in Computer Science, pp. 573–581.

[23] J. A. Little, D. L. G. Hill, and D. J. Hawkes, "Deformation incorporating rigid structures," *Comput. Vis. Image Understand.*, vol. 66, no. 2, pp. 223–232, 1997.

[24] V. Arsigny, X. Pennec, and N. Ayache, "Polyrigid and polyaffine transformations: A novel geometrical tool to deal with non-rigid deformations—Application to the registration of histological slices," *Med. Image Anal.*, vol. 9, pp. 507–523, 2005.

[25] X. Zhuang, D. J. Hawkes, W. R. Crum, R. Boubertakh, S. Uribe, D. Atkinson, P. Batchelor, T. Schaeffter, R. Razavi, and D. Hill, "Robust registration between cardiac MRI images and atlas for segmentation propagation," *Proc. SPIE Med. Imag.*, vol. 6914, p. 7, 2008.

[26] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P. Y. Bondiau, N. Ayache, and G. Malandain, "An efficient locally affine framework for the smooth registration of anatomical structures," *Med. Image Anal.*, vol. 12, no. 4, pp. 427–441, 2008.

[27] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proc. 23rd Nat. Conf. ACM*, 1968, pp. 517–527.

[28] W. R. Crum, O. Camara, and D. J. Hawkes, "Methods for inverting dense displacement fields: Evaluation in brain image registration," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2007, 10th International Conference, Brisbane, Australia, October 29–November 2, 2007, Proceedings, Part I*, N. Ayache, S. Ourselin, and A. J. Maeder, Eds.   New York: Springer, 2007, vol. 4791, Lecture Notes in Computer Science, pp. 900–907.

[29] U. Kurkure, A. Pednekar, R. Muthupillai, S. D. Flamm, and I. A. Kakadiaris, "Localization and segmentation of left ventricle in cardiac cine-MR images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1360–1370, May 2009.

[30] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2879–2890, Dec. 2007.

[31] P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," *IEEE Trans. Image Process.*, vol. 9, no. 12, pp. 2083–2099, Dec. 2000.

[32] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen, "Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling," *IEEE Trans. Med. Imag.*, vol. 21, no. 9, pp. 1151–1166, Sep. 2002.

[33] S. Uribe, V. Muthurangu, R. Boubertakh, T. Schaeffter, R. Razavi, D. L. Hill, and M. S. Hansen, "Whole-heart cine MRI using real-time respiratory self-gating," *Magn. Reson. Med.*, vol. 57, no. 3, pp. 606–613, 2007.

[34] S. Uribe, T. Tangchaoren, V. Parish, I. Wolf, R. Razavi, G. Greil, and T. Schaeffter, "Volumetric cardiac quantification by using 3-D dual-phase whole-heart MR imaging," *Radiology*, vol. 248, pp. 606–614, 2008.

[35] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[36] W. R. Crum, O. Camara, and D. L. G. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1451–1461, Nov. 2006.

[37] S. A. Glantz, *Primer of Biostatistics*, 6th ed.   New York: McGraw-Hill, 2005.

[38] M. Lynch, O. Ghita, and P. F. Whelan, "Segmentation of the left ventricle of the heart in 3-D+t MRI data using an optimized nonrigid temporal model," *IEEE Trans. Med. Imag.*, vol. 27, no. 2, pp. 195–203, Feb. 2008.

[39] C. A. Cocosco, W. J. Niessen, T. Netsch, E.-J. P. Vonken, G. Lund, A. Stork, and M. A. Viergever, "Automatic image-driven segmentation of the ventricles in cardiac cine MRI," *J. Magn. Reson. Imag.*, vol. 28, pp. 366–374, 2008.

[40] P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *Neuroimage*, vol. 46, no. 3, pp. 726–738, 2009.

[41] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.