

Complete region of interest reconstruction by fusing multiview deformable three-dimensional transesophageal echocardiography images

Zhehua Mao¹ | Liang Zhao¹ | Shoudong Huang¹ | Tongxing Jin² | Yiting Fan³ | Alex Pui-Wai Lee⁴

¹Robotics Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, New South Wales, Australia

²School of Information Science and Engineering, Harbin Institute of Technology, Weihai, China

³Department of Cardiology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China

⁴Division of Cardiology, Department of Medicine and Therapeutics, Prince of Wales Hospital and Laboratory of Cardiac Imaging and 3D Printing, Li Ka Shing Institute of Health Science, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China

Correspondence

Zhehua Mao, Robotics Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia.

Email: mazhehua@gmail.com

Funding information

Australian Research Council, Grant/Award Number: DP200100982

Abstract

Background: While three-dimensional transesophageal echocardiography (3D TEE) has been increasingly used for assessing cardiac anatomy and function, it still suffers from a limited field of view (FoV) of the ultrasound transducer. Therefore, it is difficult to examine a complete region of interest without moving the transducer. Existing methods extend the FoV of 3D TEE images by mosaicing multiview static images, which requires synchronization between 3D TEE images and electrocardiogram (ECG) signal to avoid deformations in the images and can only get the widened image at a specific phase.

Purpose: This work aims to develop a novel multiview nonrigid registration and fusion method to extend the FoV of 3D TEE images at different cardiac phases, avoiding the bias toward the specifically chosen phase.

Methods: A multiview nonrigid registration and fusion method is proposed to enlarge the FoV of 3D TEE images by fusing dynamic images captured from different viewpoints sequentially. The deformation field for registering images is defined by a collection of affine transformations organized in a graph structure and is estimated by a direct (intensity-based) method. The accuracy of the proposed method is evaluated by comparing it with two B-spline-based methods, two Demons-based methods, and one learning-based method VoxelMorph. Twenty-nine sequences of *in vivo* 3D TEE images captured from four patients are used for the comparative experiments. Four performance metrics including checkerboard volumes, signed distance, mean absolute distance (MAD), and Dice similarity coefficient (DSC) are used jointly to evaluate the accuracy of the results. Additionally, paired *t*-tests are performed to examine the significance of the results.

Results: The qualitative results show that the proposed method can align images more accurately and obtain the fused images with higher quality than the other five methods. Additionally, in the evaluation of the segmented left atrium (LA) walls for the pairwise registration and sequential fusion experiments, the proposed method achieves the MAD of (0.07 ± 0.03) mm for pairwise registration and (0.19 ± 0.02) mm for sequential fusion. Paired *t*-tests indicate that the results obtained from the proposed method are more accurate than those obtained by the state-of-the-art VoxelMorph and the diffeomorphic Demons methods at the significance level of 0.05. In the evaluation of left ventricle (LV) segmentations for the sequential fusion experiments, the proposed method achieves a DSC of (0.88 ± 0.08) , which is also significantly better than diffeomorphic Demons at the 0.05 level. The FoVs of the final fused 3D TEE images

obtained by our method are enlarged around two times compared with the original images.

Conclusions: Without selecting the static (ECG-gated) images from the same cardiac phase, this work addressed the problem of limited FoV of 3D TEE images in the deformable scenario, obtaining the fused images with high accuracy and good quality. The proposed method could provide an alternative to the conventional fusion methods that are biased toward the specifically chosen phase.

KEYWORDS

3D fusion, direct method, deformation, FoV, TEE

1 | INTRODUCTION

In cardiology, transesophageal echocardiography (TEE) is a widely used diagnostic test that uses an ultrasound transducer passed into the patient's esophagus or stomach to capture images of the heart.¹ Since the esophagus and stomach are close to the heart, detailed images can be obtained by TEE to facilitate the diagnosis, management, and follow-up of patients with various cardiac diseases. In recent decades, progress in ultrasound technology has promoted the evolution of TEE from two-dimensional (2D) imaging to three-dimensional (3D) imaging modality.² 3D TEE has overcome several important limitations of 2D TEE such as the lack of anatomy and orientation information, which makes it increasingly used in the perioperative period for assessing cardiac anatomy and function. However, current 3D TEE still suffers from a narrow field of view (FoV) due to manufacturing limitations on 2D array transducers,^{2,3} which makes it difficult to inspect a complete region of interest in a single volume. To achieve a comprehensive imaging examination, the routine method requires the clinician to move the transducer to obtain multiview images¹ and combine these images through their imagination, which heavily depends on clinicians' knowledge and experience.

Image fusion (mosaicing) is an active research topic in the field of medical image computing, which aims to combine images captured from different viewpoints automatically. Registration that determines the accuracy of anatomical structures in the fused image is a critical step prior to the fusion. Many efforts have been made to enlarge the FoV of 3D TEE images by registering and fusing multiview images.^{4–7} However, these methods are based on the static images obtained by synchronizing 3D TEE images to electrocardiogram (ECG) signal at a specific phase of cardiac cycles. Since these methods are biased to the selected time point, motion information of the anatomy of interest⁸ is not available in the fused image. Registration involved in these works is generally rigid in which the transformation has a relatively small number of degrees of freedom (DoF).^{9,10}

There is another group of registration methods that consider much more DoF in the transformation than rigid ones to take the deformation in images into account, which is called nonrigid registration.⁹ Nonrigid registration has been a popular topic in recent decades due to its great potential for assisting motion correction, longitudinal studies, multimodality information fusion, and population studies. Free form deformation (FFD)¹¹ is one of the most widely used transformation models in nonrigid registration. The idea of FFD is to embed the moving image into a space and then deform it as the space is deformed. The FFD model has achieved great success in the medical image computing field since it was coupled with B-splines.¹² These registration methods use B-spline (usually cubic B-splines) curves that are controlled by a sparse set of points to define a continuous deformation field. Then, a moving image can be registered to the fixed image by the deformation field.¹³ B-splines have local supports, which means local deformation can be calculated from only a couple of surrounding control points.^{9,14} This property is beneficial for local deformation modeling and fast implementation of the B-spline-based methods.⁹ Demons-based methods are another well-known group of nonrigid registration methods. Inspired by Maxwell's Demons and optical flow equations, Thirion proposed the original Demons algorithm¹⁵ that considers image registration as a diffusion process and computes the deformation field through iterating between the computation of the Demons forces and a Gaussian smoothing regularization. Since the optical flow model is less constrained than B-splines and the deformation field is estimated at every voxel, which results in an extremely large number of DoF,¹⁶ Demons method can deal with larger local deformation in the images than B-spline-based FFD methods.^{17,18} Following Thirion's Demons method, a range of variants were proposed to improve convergence speed and precision.¹⁸ Among these, the diffeomorphic Demons algorithm¹⁹ which can preserve the topology of the objects achieves remarkable success in various computational anatomy studies, such as breast computed tomography (CT),²⁰ brain magnetic resonance (MR),²¹ and lung CT.²² In recent years,

learning-based registration methods have attracted a lot of attention because of their high efficiency after training.^{23,24} The state-of-the-art learning-based method VoxelMorph has shown comparable accuracy to the nonlearning-based methods.²⁵ While the learning-based methods have computational advantages in the registration step, they have not demonstrated superior accuracy to traditional nonlearning-based methods.^{23–25} In addition, learning-based methods usually require much time for training. When processing images from different regions of interest, different patients, or different imaging modalities, learning-based frameworks usually need to be trained separately.^{23,24}

Based on the FFD model, Sumner et al.²⁶ proposed an embedded deformation method in the computer graphics community, which is notable due to its simplicity, versatility, and efficiency. Embedded deformation defines the deformation field with a collection of affine transformations organized in a graph structure and estimates the deformation graph making use of the specified corresponding landmarks of the embedded object before and after the deformation. Song et al.²⁷ adopted the method to reconstruct the surface of soft tissues from 2D laparoscopy images captured with different deformations. Extracted and matched feature points between moving and fixed images are used to estimate the deformation in their work. However, such a method relies on the distinct features in images and discards all other information, which tends to fail when handling images with few distinct features such as 3D TEE images.^{9,10}

In this paper, we propose a direct dynamic fusion method based on the embedded deformation graph to address the problem of limited FoV of 3D TEE images by sequentially fusing images captured at different phases of cardiac cycles and from different viewpoints. Different from Ref.[26] and Ref.[27] that relied on explicit point constraints, we propose to use intensity information as constraints directly to estimate the deformation, thus making use of all information from the images. Since registration and fusion are performed directly on the dynamic 3D TEE images, a 3D image with extended FoV is likely to be reconstructed at different phases of cardiac cycles and therefore, avoid the bias to the specifically chosen time. The main contributions of the paper include: (1) proposing an intensity-based dynamic medical image registration method based on the embedded deformation graph such that feature extraction and matching can be avoided; (2) proposing a multiview deformable image fusion framework to enlarge the FoV of 3D TEE images so that the bias toward the specifically chosen phase in conventional rigid registration and fusion can be avoided; (3) evaluating the accuracy of the proposed method compared with the widely used B-spline-based methods, Demons-based methods, and the state-of-the-art learning-based method VoxelMorph,

qualitatively and quantitatively using in vivo 3D TEE data sets.

The remainder of this paper is organized as follows: The framework of direct dynamic fusion is presented in Section 2, followed by detailed introductions of the method and experiments. Section 3 reports the experimental results of the proposed method compared with five popularly used nonrigid registration methods, that is, two B-spline-based FFD methods implemented in Plastimatch²⁸ and ElastiX²⁹, respectively, the original Demons,¹⁵ the diffeomorphic Demons,¹⁹ and the state-of-the-art learning-based method VoxelMorph.²⁵ These five methods are all intensity based and thus comparable to the proposed method. Finally, Section 4 concludes the paper and discusses the implications and limitations of this study.

2 | METHODS

We propose a multiview registration and fusion method to enlarge the FoV of a single 3D TEE image in the deformable scenario. In this section, the details of the proposed method and the validation experiments are introduced.

2.1 | Framework overview

Our framework for enlarging the FoV of 3D TEE images with deformation is presented in Figure 1, which mainly consists of nonrigid registration and 3D image fusion. A sequential fusion strategy is used to perform the registration and fusion repeatedly so that multiview 3D TEE images captured at different phases of cardiac cycles and different viewpoints can be fused one by one. First, the global image I_g is initialized by the first image. And then, in each loop, we designate the global image as a moving image and every new coming image I_f as a fixed image and perform nonrigid registration and fusion. To deal with global motion between the moving and the fixed images, the global image I_g is rigidly transformed to an image space that is close to image I_f and denoted as \tilde{I}_g first. The rough relative pose used for the transformation is provided by SIFT3D.³⁰ Although the feature-based algorithm SIFT3D cannot precisely extract and match key features from the TEE images, it is still useful for getting an initial pose to guide our registration. After preregistration, the deformation field X is estimated by using the proposed direct nonrigid registration method. Then, the transformed image \tilde{I}_g is deformed using the deformation field and fused with image I_f . The fused image I is used as the new global image I_g for the next new coming image. The same processes are performed repeatedly until all the images are fused.

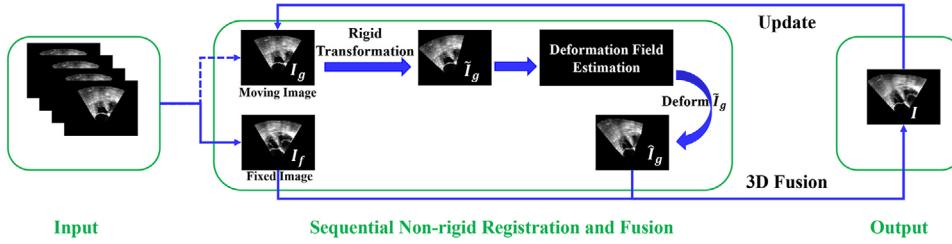


FIGURE 1 The framework of the direct dynamic fusion method. Remark: 3D TEE images are shown via 2D slices.

In order to obtain good results from our algorithm, consecutive 3D TEE images captured at different viewpoints are controlled to have more than 50% overlap. At the end of each loop, we reconstruct the global 3D TEE image at the same phase as the fixed image. Besides the area, which can be observed in the fixed image, deformation in unobserved regions can also be predicted according to the principle of as rigid as possible.³¹ Therefore, through continuously fusing the global image with the remaining images, the FoV of 3D TEE images is enlarged.

2.2 | Direct deformation

Our nonrigid registration method employs the embedded deformation graph to describe the deformation field. But due to the relatively low signal-to-noise ratio of ultrasound imaging, it is difficult to extract and match feature points from the 3D TEE images for deformation estimation as in Ref. [27]. Therefore, we propose to use a direct (intensity-based) method to compute the deformation by maximizing the similarity between the moving and the fixed images.

2.2.1 | Local deformation

The embedded deformation graph is represented by a set of nodes associated with affine matrices. Assuming that the graph nodes are selected uniformly from \tilde{I}_g by downsampling it (\tilde{I}_g is obtained by preregistering the moving image), let $\mathbf{g}_j \in \mathbb{R}^3$ denote the position of node j and an associated affine transformation consists of a matrix $\mathbf{A}_j \in \mathbb{R}^{3 \times 3}$ and a translation vector $\mathbf{t}_j \in \mathbb{R}^3$. The affine transformation applies a deformation to its nearby space. Conversely, for any voxel $\mathbf{p}_i \in \mathbb{R}^3$ in image \tilde{I}_g , its deformation is influenced by its nearby nodes. To maintain the consistency of the deformation graph, nodes that are close to one another should have similar transformations. Thus, we limit the influence of the deformation on voxel \mathbf{p}_i to its k nearest graph nodes to keep the consistency and efficiency of the deformation estimation. And the final deformation of point \mathbf{p}_i is defined as a weighted average of the effects of its nearby k

nodes by:

$$\hat{\mathbf{p}}_i = \sum_{j=1}^k \hat{w}_j(\mathbf{p}_i) [\mathbf{A}_j(\mathbf{p}_i - \mathbf{g}_j) + \mathbf{g}_j + \mathbf{t}_j], \quad (1)$$

where weight $\hat{w}_j(\mathbf{p}_i)$ denotes the influence of a nearby node on the voxel, $\sum_{j=1}^k \hat{w}_j = 1$. \hat{w}_j can be precalculated by the following formula and then normalized.

$$w_j(\mathbf{p}_i) = \left(1 - \|\mathbf{p}_i - \mathbf{g}_j\| / d_{max} \right), \quad (2)$$

where d_{max} represents the distance from the voxel to the $(k+1)$ th nearest node. We use $k = 6$ for all experiments presented in this paper.

2.2.2 | Energy function

To estimate the parameters $\mathbf{X} = \{\mathbf{A}_1, \mathbf{t}_1, \dots, \mathbf{A}_m, \mathbf{t}_m\}$ for a deformation graph with m nodes, an energy function as the following is minimized:

$$E = w_{rot} E_{rot} + w_{reg} E_{reg} + w_{con} E_{con}, \quad (3)$$

where E_{rot} , E_{reg} , and E_{con} represent rotation term, regularization term, and constraint term, respectively, and w_{rot} , w_{reg} , and w_{con} are the corresponding weights. In this energy function, rotation and regularization terms are used to guide the optimization to a reasonable and smooth deformation.^{26,27} And the constraint term constrains the deformation by maximizing the similarity between the fixed and moving images. Additionally, the number of graph nodes m determines the expressibility of the deformation graph. The more nodes are selected from one image, the more details that the graph could express. According to our tests, we select around 1000 nodes from a moving image that contains around five million voxels for pairwise registration, which could balance the accuracy and efficiency of the algorithm.

Rotation term

To preserve the topological structure of the deformed images, the objective function (3) directly specifies a

matrix \mathbf{A}_j at each node should be close to a rotation matrix³² and local voxels are deformed as rigidly as possible. A rotation matrix in 3D space is subject to six constraints, that is, each column should be a unit length and should be orthogonal to one another. Thus, the rotation term sums the rotation error over all affine transformations of the deformation graph²⁶:

$$E_{\text{rot}} = \sum_{j=1}^m \text{Rot}(\mathbf{A}_j) \quad (\text{where } \mathbf{A}_j = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]), \quad (4)$$

$$\begin{aligned} \text{Rot}(\mathbf{A}_j) = & (\mathbf{a}_1^\top \cdot \mathbf{a}_2)^2 + (\mathbf{a}_1^\top \cdot \mathbf{a}_3)^2 + (\mathbf{a}_2^\top \cdot \mathbf{a}_3)^2 \\ & + (\mathbf{a}_1^\top \cdot \mathbf{a}_1 - 1)^2 + (\mathbf{a}_2^\top \cdot \mathbf{a}_2 - 1)^2 + (\mathbf{a}_3^\top \cdot \mathbf{a}_3 - 1)^2. \end{aligned} \quad (5)$$

Regularization term

To keep the consistency and smoothness of the deformation graph, the position of node j deformed by nearby k nodes should be conformed with the actual position calculated by node j itself. Thus, the regularization term sums the squared error between predicted positions of nodes and their actual positions²⁶:

$$E_{\text{reg}} = \sum_{j=1}^m \sum_{i \in \mathbb{N}(j)} \alpha_{ji} \left\| \mathbf{A}_j(\mathbf{g}_i - \mathbf{g}_j) + \mathbf{g}_j + \mathbf{t}_j - (\mathbf{g}_i + \mathbf{t}_i) \right\|^2, \quad (6)$$

where $\mathbb{N}(j)$ denotes the index of nearby k nodes which influences node j . And we also follow Ref. [26] to set weight α_{ji} as 1. In the energy function (3), increasing the weight of E_{reg} will make the nonrigid transformation become closer to the rigid transformation.

Constraint term

Our method estimates the deformation by directly minimizing intensity differences between the two images instead of specifying point constraints.^{26,27} Suppose the intensity value of a voxel \mathbf{p} in image I is a function w.r.t. its coordinates and denoted as $I(\mathbf{p})$, for a spatial point P_i in the heart and its projections \mathbf{p}_{gi} and $\hat{\mathbf{p}}_{fi}$ in image \tilde{I}_g and I_f , the intensity difference between \mathbf{p}_{gi} and $\hat{\mathbf{p}}_{fi}$ can be written as:

$$e_i = \tilde{I}_g(\mathbf{p}_{gi}) - I_f(\hat{\mathbf{p}}_{fi}). \quad (7)$$

Then, we can sum the squared intensity differences over n voxels in overlapping areas between the two images as the constraint term:

$$E_{\text{con}} = \sum_{i=1}^n \|e_i\|^2. \quad (8)$$

In this paper, we use iterative optimization to obtain the optimal solution of (3). In every iteration, $\hat{\mathbf{p}}_{fi}$ is computed

from \mathbf{p}_{gi} via (1). Since $\hat{\mathbf{p}}_{fi}$ may not be integers and thus not on the grid of the fixed image I_f , the intensity of $I_f(\hat{\mathbf{p}}_{fi})$ is obtained using trilinear interpolation to reduce the error for difference computation in (7).

2.2.3 | Optimization

Minimizing (3) is a routine nonlinear least-squares problem and the solution can be obtained by using Gauss–Newton (GN) method.³³ First, we rewrite the energy function (3) as a standard format as follows:

$$E = w_{\text{rot}} E_{\text{rot}} + w_{\text{reg}} E_{\text{reg}} + w_{\text{con}} E_{\text{con}} = f(\mathbf{X})^\top f(\mathbf{X}). \quad (9)$$

Let $J(\mathbf{X})$ denote the Jacobian matrix of $f(\mathbf{X})$ w.r.t. deformation parameter \mathbf{X} . Then, the optimal solution $\hat{\mathbf{X}}$ can be obtained by initializing with \mathbf{X}_0 and iterating with $\mathbf{X}_{k+1} = \mathbf{X}_k + \Delta\mathbf{X}_k$. In our algorithm, the initial value of \mathbf{X} is set as $\mathbf{X}_0 = \{\mathbf{I}_1, \mathbf{0}_1, \dots, \mathbf{I}_m, \mathbf{0}_m\}$, where $\mathbf{I} \in \mathbb{R}^{3 \times 3}$ and $\mathbf{0} \in \mathbb{R}^3$ are identity matrix and column vector of zeros, respectively. The step change $\Delta\mathbf{X}_k$ in each iteration can be obtained from the GN equation:

$$J^\top(\mathbf{X}) J(\mathbf{X}) \Delta\mathbf{X}_k = -J^\top(\mathbf{X}) f(\mathbf{X}). \quad (10)$$

After the optimal deformation parameters $\hat{\mathbf{X}}$ is obtained, the deformed image \tilde{I}_g can be obtained by deforming \tilde{I}_g using (1). And then, the deformed image \tilde{I}_g is fused with the fixed image I_f by using a 3D fusion method proposed in Ref. [7].

2.3 | Experiments

2.3.1 | Data collection

In vivo data sets for the experiments are collected from four patients using an iE33 ultrasound system (Philips Medical Systems) equipped with an X7-2 real-time 3D transducer. A 3D ultrasound transducer can generally record a sequence of 3D TEE images with different deformations at one position. By moving the transducer, multiple sequences of 3D TEE images can be acquired from one patient. In total, 29 sequences of TEE volumes are collected from four patients. The movement of the transducer between consecutive positions is controlled within $[-50, 50]$ mm and $[-10, 10]$ degrees range to ensure that the overlapping rate between images collected at consecutive positions is more than 50%. Ultrasound parameters of the imaging system are fixed during the image collection of each patient. The details of the in vivo data sets from the four patients # A, # B, # C, and # D are listed in Table 1.

TABLE 1 Details of in vivo data sets

Patient	No. sequences	Volume size (voxel)	Resolution (mm/voxel)
# A	8	277 × 208 × 208	0.63 × 0.90 × 0.68
# B	6	240 × 160 × 208	0.64 × 0.92 × 0.68
# C	6	240 × 160 × 208	0.77 × 1.11 × 0.82
# D	9	240 × 160 × 208	0.69 × 0.98 × 0.73

2.3.2 | Experimental setup and evaluation

The proposed method is validated by comparing it with two B-spline-based methods, two Demons-based methods, and the learning-based method VoxelMorph.²⁵ For the B-spline-based and Demons-based methods, we use the implementations of them in 3D Slicer³⁴ (version 4.10) for comparisons. Since the two B-spline-based methods are implemented based on the software package Plastimatch²⁸ and ElastiX,²⁹ respectively, the packages' names are used to indicate the methods. Additionally, the original Demons and the diffeomorphic Demons methods in 3D Slicer are implemented based on Insight Segmentation and Registration Toolkit (ITK).³⁵ The codes of VoxelMorph published by the authors are used in the experiments. For the proposed method, coefficient for each term in the energy function (3) is specified as: $w_{\text{rot}} = 1$, $w_{\text{reg}} = 10000$, $w_{\text{con}} = 0.1$. The minimum step length and the maximum number of iterations are used as stop criteria for convergence, which are set as 10^{-5} and 100, respectively.

The following two groups of in vivo experiments are performed to validate the proposed direct dynamic fusion method.

(i) Pairwise registration experiments: in each experiment, two 3D TEE images collected at different cardiac phases and from different viewpoints in one patient are used for pairwise registration. Five pairs of 3D TEE images from the four patients are registered by using the proposed method and the other five nonrigid registration methods. To mitigate the influences of global motion on the registration, rigid transformation obtained from SIFT3D³⁰ is used to do the preregistration for the six methods in each experiment (all methods use the same rigid transformation to perform the preregistration in the same experiment). For VoxelMorph, we train each patient's data set individually with 3D TEE images of more than one cardiac cycle until the algorithm converges. Then, the trained models are used for registration. Since VoxelMorph cannot deal with nonoverlapping areas between images,²⁵ only overlapping areas between the moving and fixed images are used for nonrigid registration. The detailed

results of pairwise registration are reported in Section 3.1.

(ii) Sequential fusion experiments: first, 29 ECG-gated TEE volumes are extracted from the twenty-nine sequences by synchronizing acquired 3D TEE images to ECG signal at late diastole. These ECG-gated 3D TEE images extracted from the same data set are static (i.e., deformations in images are eliminated). The static images are fused together and the results are used as the references for evaluating the results obtained by proposed and competing methods. Second, we perform dynamic fusion on the selected volumes sequentially. The final volume is reconstructed at the same phase as the ECG-gated volume so that the fused volumes from both static fusion and dynamic fusion should be at the same phase and have the same images. Finally, the fused images obtained from the dynamic fusion are compared with the images from the static fusion to evaluate the proposed method. In the second step of dynamic fusion, we select one volume from each sequence in each data set (from one patient) with different deformations and therefore, four sequences containing eight, six, six, and nine 3D TEE images, respectively, are obtained. In each sequence, the final volume is selected as an ECG-gated one to make sure the final fused images are at the same phase as the referenced ones. The detailed results of sequential fusion are reported in Section 3.2.

Verifying accuracy is difficult in nonrigid registration since ground-truth deformation is usually not available in practical data.³⁶ In addition to checking the quality of the aligned images visually, there are some quantitative metrics used in many works of literature^{37,38} to assess the registration accuracy, such as computing the distance between the segmented anatomical structures or overlap of the segmented anatomical structures after registration. Since different metrics have different sensitivity to the true deformation,³⁷ the following four metrics are used jointly in this paper to evaluate the accuracy of registration and fusion:

- (i) The alignment of stitches in the checkerboard volumes.³⁹ A checkerboard volume is created by taking a part from each of the two aligned images and stitching them together. By checking alignment at the stitches, we can assess the accuracy of registration qualitatively.
- (ii) Signed distance between segmented left atrium (LA) walls. LA walls are sharp boundaries that can be easily recognized in the 3D TEE images. We segment the LA walls from both the deformed/fused images and their reference images manually and then, calculate the signed distance of each vertex on the LA walls segmented from deformed/fused

images to the closest vertex on the LA walls segmented from the reference images. The results are displayed via heatmaps in Sections 3.1 and 3.2.

- (iii) Mean absolute distance (MAD) between segmented LA walls. MAD is a statistical value that is calculated by averaging the absolute value of the metric (ii).
- (iv) Dice similarity coefficient (DSC)³⁸ between segmented anatomical structures from the fused images and those from the reference images. Generally, DSC is calculated by

$$\text{DSC}(M, N) = \frac{2|M \cap N|}{(|M| + |N|)}, \quad (11)$$

where M and N represent the target regions of the two segmentations. DSC computes the overlap of segmented anatomical structures. The maximum value of DSC is 1 and a higher DSC indicates a better accuracy of registration and fusion.

In addition, p -value approach is used to perform paired t -tests to examine the statistical significance of the results with the significance level of 0.05.

3 | RESULTS

3.1 | Pairwise nonrigid registration

We jointly use the qualitative and quantitative metrics to evaluate the accuracy of the proposed algorithm as comprehensively as possible. The qualitative evaluation of pairwise registration accuracy is performed by checking the alignment of stitches in the checkerboard volumes. The checkboard volumes based on the results from different methods are shown in Figure 2. Through checking the stitches in sharp areas such as the walls of LA in the checkerboard volumes, obvious misalignments which are indicated by red dashed cycles can be found in images obtained from the two B-spline-based methods and VoxelMorph method, especially in the images obtained from the Plastimatch method. In contrast, checkerboard volumes from the two Demons-based methods and the proposed method have smooth transitions in the stitching areas. No obvious misalignment can be found in the results from these three methods, which indicates that the Demons-based methods and the proposed method are more accurate than the compared B-spline-based methods and VoxelMorph method. Since the original Demons method cannot guarantee the topology in the deformed images, some defects are found from the images deformed by the method. One example is shown in Figure 3 (the defects are indicated by the red dashed cycles). Please note, one clear error in anatomical structure is also found

from the result obtained by the diffeomorphic Demons method in Figure 3, which is indicated by the white arrow. Such conditions are not found from the results obtained from the proposed method in our experiments.

Furthermore, to assess the registration accuracy quantitatively, we first manually segment the LA walls from both deformed and fixed images in each pair and then calculate the distance between them.⁴⁰ Nonoverlapping areas and areas prone to segmentation errors such as mitral valves are omitted to reduce computational and human errors. Both the signed distance and MAD are calculated and shown in Figure 4. The corresponding statistical values of MAD (provided by mean \pm standard deviation) are shown in Table 2. Similar to our qualitative evaluation, it is found from Figure 4 that the distribution of signed distance from the two Demons-based methods and our method is concentrated near zero and within a smaller range than that from the two B-spline-based methods and VoxelMorph, which indicates the higher accuracy of the two Demons-based methods and our method than the other three. Similarly, it is also shown from Figure 4 that Elastix method and VoxelMorph have higher accuracy than Plastimatch method. In addition, from the statistical values shown in Table 2, the MAD of the proposed method, the original Demons method, the diffeomorphic Demons method, and VoxelMorph are around (0.07 ± 0.03) mm, (0.31 ± 0.16) mm, (0.20 ± 0.09) mm, and (0.54 ± 0.22) mm, respectively, while the results from the two B-spline-based methods are greater than 0.50 mm and the results from Plastimatch method are even greater than 1.00 mm. P -values of paired t -tests are less than 0.05, which indicates that the proposed method outperforms the other five methods in terms of accuracy with statistical significance.

In summary, it is found from the comparisons that: (1) the proposed method has the best accuracy (less than 0.10 mm) in pairwise registration and can preserve the topology in the deformed images; (2) the diffeomorphic Demons is more accurate than the original Demons method and can preserve topology in the images; (3) both the proposed method and Demons-based methods are more accurate than the B-spline based methods and VoxelMorph method. Since the proposed method and the diffeomorphic Demons outperformed the other methods in terms of accuracy and VoxelMorph cannot deal with nonoverlapping areas, in the next section, the proposed method will be evaluated by comparing it with the diffeomorphic Demons method only.

3.2 | Sequential fusion

Four sequences of 3D TEE images are registered and fused, respectively, by using the proposed direct dynamic fusion method and the diffeomorphic Demons method.

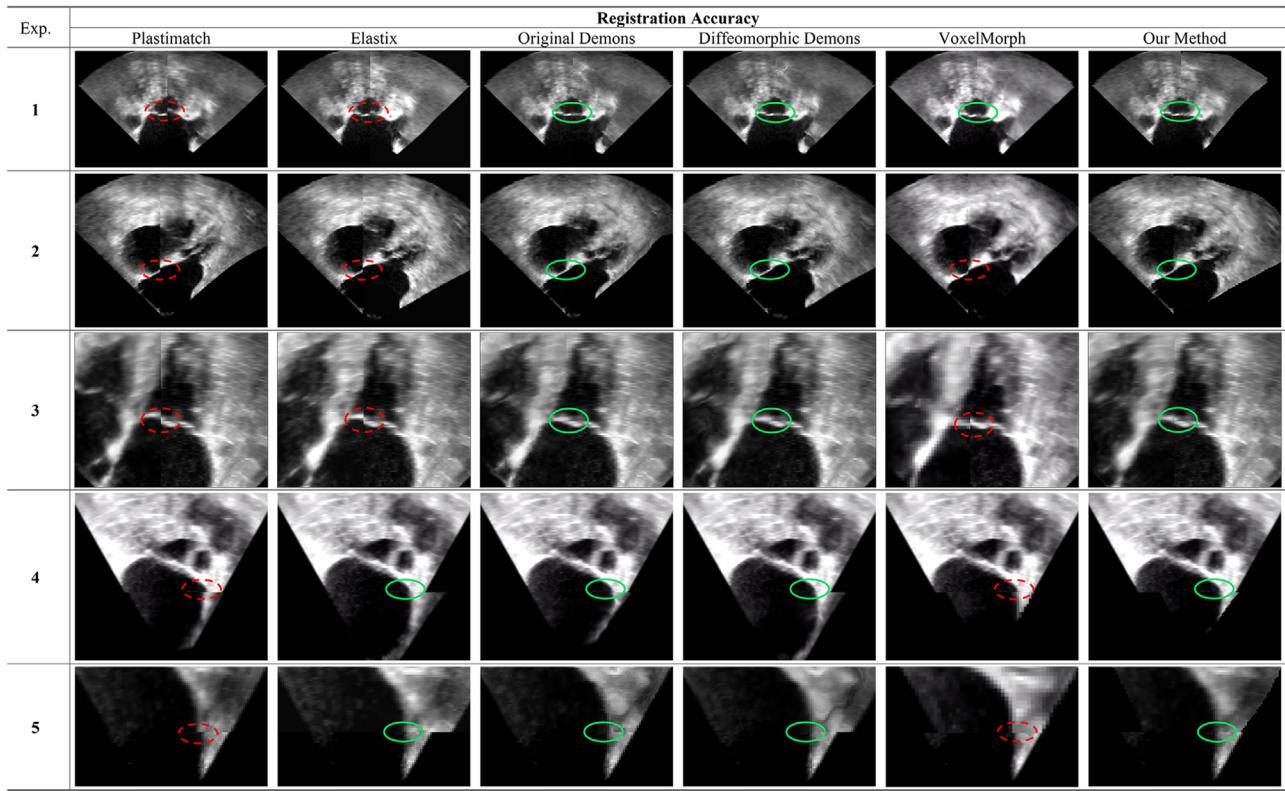


FIGURE 2 Checkerboard volumes obtained by stitching images after registration. The same positions at the stitches are marked by circles and obvious misalignments are indicated by the red dashed circles in the images.

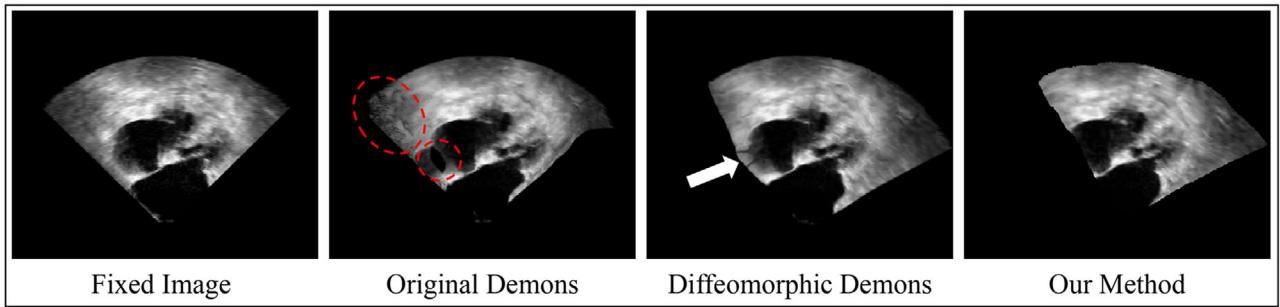


FIGURE 3 Fixed image and deformed moving images. The defects in the volume obtained from the original Demons are indicated by the red dashed cycles. And the white arrow indicates one obvious error in the volume obtained from the diffeomorphic Demons.

The results of the static fusion are used as references for evaluation. Figure 5 shows the final fused 3D volume of the sequence from Patient # A. Through the comparisons, it is found that the fused image obtained from the proposed method has almost identical anatomical structures as the image from the static fusion. However, obvious accumulating errors are found in the volume fused by the diffeomorphic Demons method, which are indicated by red boxes on images in the figure. Similar results are found from the 3D volumes fused by the other three sequences, which are shown in Figure 6.

In addition, similar to Section 3.1, distances from the segmented LA walls obtained from the two dynamic fusion methods to those obtained from the static fusion are calculated, respectively. An example of the signed distance is presented in Figure 7, which shows that the LA wall segmented from the image fused by the proposed method is more accurate than that segmented from the image fused by the diffeomorphic Demons method. The MAD for each sequence and the corresponding statistical results are summarized in Table 3. Furthermore, since the left ventricle (LV) can

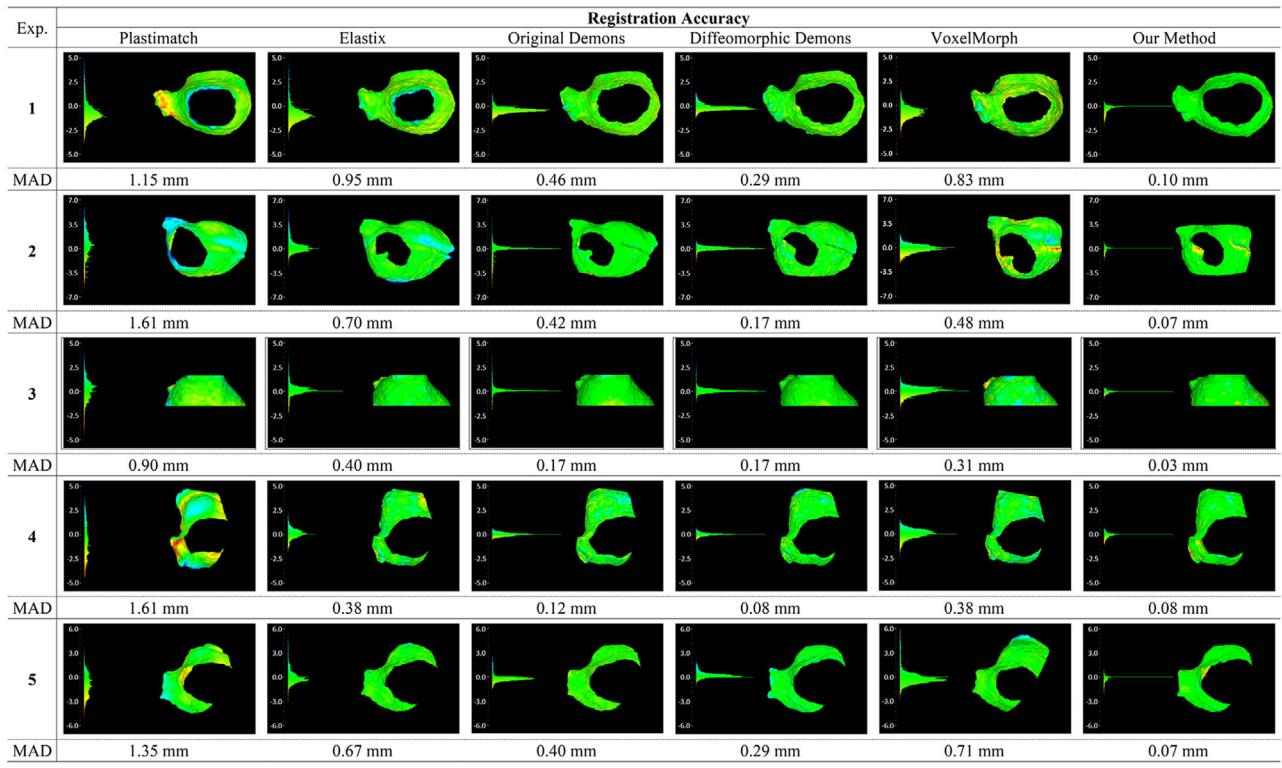


FIGURE 4 Signed distance and MAD between LA walls segmented from the deformed and fixed images

TABLE 2 Comparisons of statistical values of MAD for pairwise registration

	Mean \pm Std (mm)	p-value
Plastimatch	1.32 ± 0.31	3.41×10^{-4}
Elastix	0.62 ± 0.24	2.50×10^{-3}
Original Demons	0.31 ± 0.16	9.90×10^{-3}
Diffeomorphic Demons	0.20 ± 0.09	1.39×10^{-2}
VoxelMorph	0.54 ± 0.22	3.30×10^{-3}
Our Method	0.07 ± 0.03	-

be observed completely in the reconstructed images (see Figure 5 and Figure 6), DSC value between the segmented LV from the dynamic fusion and that from the static fusion is also calculated for each experiment and summarized in Table 3. The results show that the MAD of the results obtained from the proposed method are around (0.19 ± 0.02) mm, while the results from the diffeomorphic Demons are around (0.89 ± 0.47) mm. Moreover, DSC values calculated on the results from the proposed method (0.88 ± 0.06) are greater than those calculated on the results from the diffeomorphic Demons method (0.82 ± 0.08) . P-values of paired t-tests indicate that the proposed method outperforms diffeomorphic Demons in terms of accuracy with statistical significance.

It is found from the experiments that the proposed method can enlarge the FoV of the single volume effec-

tively. One example of comparison of the original single volume and the final fused volume is shown in Figure 8. By comparing the number of voxels in the original single volume and the fused volume, it is found that the FoVs of the fused images are enlarged to 1.83, 1.84, 2.06, and 1.97 times as compared with the original single volume of TEE image of # A, # B, # C, and # D, respectively, which are listed in the last column of Table 3.

4 | DISCUSSIONS AND CONCLUSIONS

Limited FoV remains a major challenge for examining a whole region of interest of the heart with 3D TEE images. Conventional registration and fusion methods enlarge the FoV by fusing multiview images captured at the same phase of cardiac cycles.⁴⁻⁶ However, these methods require an additional step of synchronizing 3D TEE images with ECG signal and can only reconstruct the 3D volumes at the selected phase so that motion information of the anatomy of interest is not available in the fused images. In this paper, a novel direct dynamic fusion framework is proposed to enlarge the FoV of 3D TEE images by fusing multiview images captured at different phases of cardiac cycles sequentially, avoiding the aforementioned bias. The *in vivo* experiments are performed to evaluate our method compared with the other five widely used nonrigid registration meth-

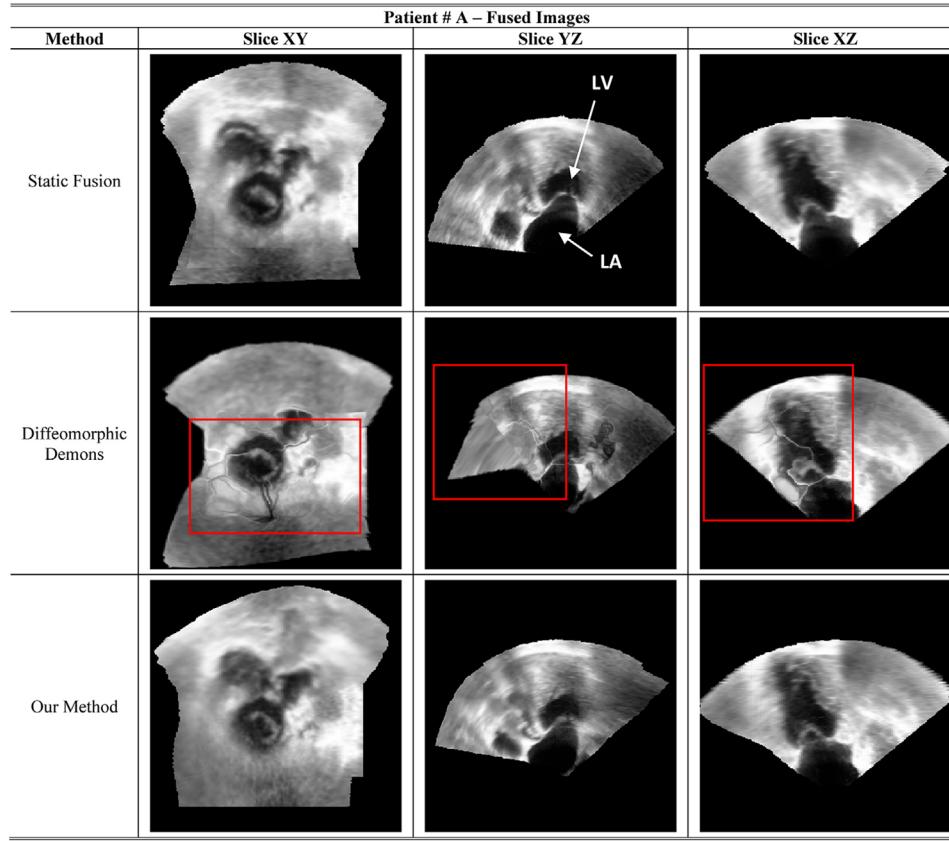


FIGURE 5 Comparisons of the fused volume of the sequence from Patient # A. Obvious accumulating errors are marked by red boxes in the images.

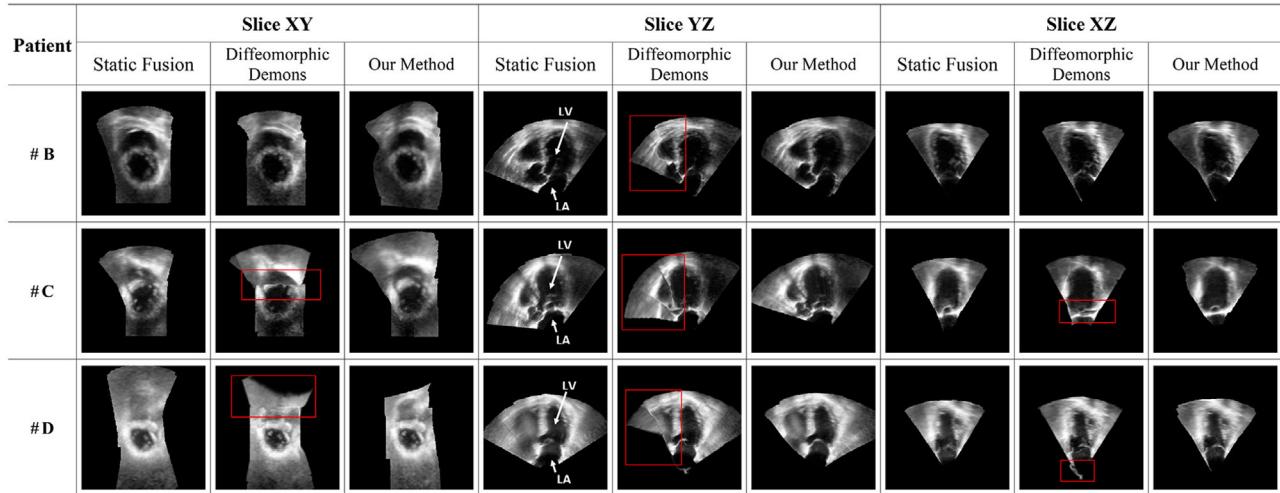


FIGURE 6 Comparisons of the fused volumes of sequences from Patients # B, # C, and # D. Obvious accumulating errors are marked by red boxes in the images.

ods, namely, two B-spline–based methods implemented in Plastimatch²⁸ and ElastiX,²⁹ the original Demons,¹⁵ the diffeomorphic Demons,¹⁹ and the state-of-the-art learning-based method VoxelMorph.²⁵ In the performed pairwise and sequential fusion experiments, qualitative

and quantitative results show that the proposed method outperforms these five methods in terms of accuracy. The final fused volumes of the four patients show that the proposed method can reconstruct the 3D TEE images with good quality and enlarge the FoV of images

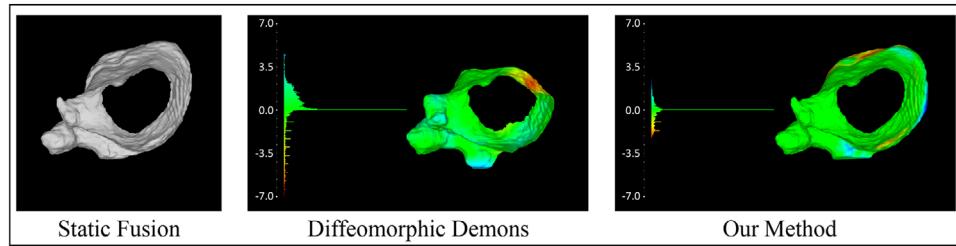


FIGURE 7 Signed distance from the segmented LA walls obtained from the diffeomorphic Demons and the proposed method to that obtained from the static fusion

TABLE 3 Comparisons of sequential fusion accuracy and the enlarged FoV after fusion

	MAD (mm)		DSC of LV		
	Diffeomorphic Demons	Our method	Diffeomorphic Demons	Our method	FoV w.r.t. original
# A	1.04	0.18	0.78	0.89	1.83
# B	0.30	0.21	0.74	0.79	1.84
# C	1.42	0.19	0.86	0.88	2.06
# D	0.81	0.17	0.91	0.94	1.97
Mean \pm Std	0.89 ± 0.47	0.19 ± 0.02	0.82 ± 0.08	0.88 ± 0.06	1.93 ± 0.11
p-value	0.03	-	0.04	-	-

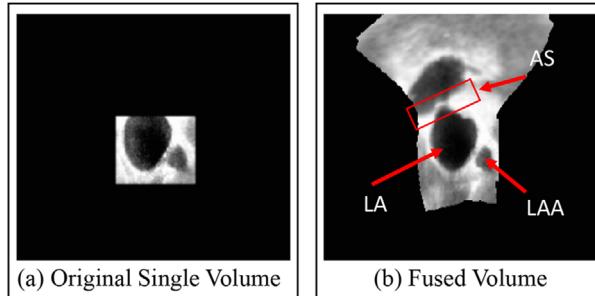


FIGURE 8 Comparison of the original single volume (a) and the fused volume (b)

in the fused volumes around two times compared with that in the original single volume.

The original embedded deformation method²⁶ based on the FFD model estimates the deformation field through specified corresponding points of an embedded object before and after the deformation. Such a constraint is in line with the feature-based methods,²⁷ which are widely used in the computer vision community. Feature-based methods compute the transformation based on the extracted features, discarding all other valuable information in the images. These methods simplify the transformation estimation but the accuracy varies. When the features are difficult to be extracted from images such as ultrasound images, the transformation estimation tends to be very inaccurate. On the other hand, direct (intensity-based) methods have dominated the field of medical image registration and

fusion because of their high accuracy and without requirements of feature extraction and matching.⁴¹ Since the proposed method estimates the deformation field based on the intensity of images directly, it is able to make use of all information in images and therefore is likely to be applied to more medical imaging modalities such as 2D/3D ultrasound, MR, and X-ray imaging.

In addition, the enlarged FoV of 3D TEE images has many implications for the clinic, such as assisting diagnosis, planning, and follow-up with 3D TEE, increasing the accuracy of measurements in 3D TEE images, and reducing the learning burden for clinicians with 3D TEE. For example, comparing the original single volume shown in Figure 8a with the enlarged FoV of the fused image shown in Figure 8b, one can see that the complete structures of both the atrial septum (AS) and left atrial appendage (LAA) are clearly visible in the fused image. Visualizing the AS and LAA in one view is very beneficial to preoperative planning of transcatheter LAA occlusion since it can facilitate transseptal puncture at a specific site of the AS in relation to the LAA position and orientation. Furthermore, since the proposed method is capable of reconstructing 3D TEE images at various phases of cardiac cycles, it makes motion estimation^{8,42} in the fused images possible. In clinical practices, it is often desirable to estimate the motion of the anatomical structures such as to build motion and deformation models for surgical or therapeutic planning and guidance and to analyze motion-related markers of diseases quantitatively.

In general, while most direct methods save computational time used for feature extraction and matching, they involve relatively high computational complexity w.r.t. similarity measurement. Therefore, our current work of direct fusion still aims at improving the accuracy for offline applications in the preoperative and postoperative periods. We currently implemented the algorithm in MATLAB on CPU and the computational cost for pairwise registration is around 1 h. In the next step, we will consider achieving the real-time performance of the algorithm to facilitate the intraoperative application such as intraoperative guidance where visualizing the latest structure of the heart is important.⁴² Since the deformation of each point in the moving image is only influenced by the limited nodes closest to it, it is beneficial to the fast implementation of the algorithm through parallel computing on GPU. Additionally, the commonly used multiresolution¹⁰ strategy in image registration and fusion could also be used to speed up the convergence of the algorithm.

ACKNOWLEDGMENT

This work is supported by Australian Research Council (ARC) Discovery project “Visual Simultaneous Localisation and Mapping in Deformable Environments” (DP200100982).

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

REFERENCES

- Hahn RT, Abraham T, Adams MS, et al. Guidelines for performing a comprehensive transesophageal echocardiographic examination: recommendations from the American Society of Echocardiography and the Society of Cardiovascular Anesthesiologists. *J Am Soc Echocardiography*. 2013;26(9):921-964.
- Huang Q, Zeng Z. A review on real-time 3D ultrasound imaging technology. *Biomed Res Int*. 2017;2017.
- Fenster A, Downey DB, Cardinal HN. Three-dimensional ultrasound imaging. *Phys Med Biol*. 2001;46(5):R67.
- Housden RJ, Ma Y, Arujuna A, et al. Extended-field-of-view three-dimensional transesophageal echocardiography using image-based X-ray probe tracking. *Ultrasound Med Biol*. 2013;39(6):993-1005.
- Carminati MC, Piazze C, Weinert L, et al. Reconstruction of the descending thoracic aorta by multiview compounding of 3-D transesophageal echocardiographic aortic data sets for improved examination and quantification of atheroma burden. *Ultrasound Med Biol*. 2015;41(5):1263-1276.
- Mulder HW, Stralen vM, Ren B, et al. Atlas-based mosaicing of left atrial 3-D transesophageal echocardiography images. *Ultrasound Med Biol*. 2017;43(4):765-774.
- Mao Z, Zhao L, Huang S, Fan Y, Lee APW. Direct 3D ultrasound fusion for transesophageal echocardiography. *Comput Biol Med*. 2021;134:104502.
- Hawkes DJ, Barratt D, Blackall JM, et al. Tissue deformation and shape models in image-guided interventions: a discussion paper. *Med Image Anal*. 2005;9(2):163-175.
- Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: a survey. *IEEE Trans Med Imaging*. 2013;32(7):1153-1190.
- Oliveira FP, Tavares JMR. Medical image registration: a review. *Comput Methods Biomed Eng*. 2014;17(2):73-93.
- Sederberg TW, Parry SR. Free-Form Deformation of Solid Geometric Models. In: Evans DC, Athay RJ, eds. *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '86*. ACM Press. 1986: 151-160.
- Sdika M. A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. *IEEE Trans Med Imaging*. 2008;27(2):271-281.
- Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging*. 1999;18(8):712-721.
- Ji H, Li Y, Dong E, et al. A non-rigid image registration method based on multi-level B-spline and L2-regularization. *Signal, Image, Video Process*. 2018;12(6):1217-1225.
- Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. *Med Image Anal*. 1998;2(3):243-260.
- Hanaoka S, Masutani Y, Nemoto M, et al. Landmark-guided diffeomorphic demons algorithm and its application to automatic segmentation of the whole spine and pelvis in CT images. *Int J Comput Assisted Radiol Surg*. 2017;12(3):413-430.
- Klein A, Andersson J, Ardekani BA, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*. 2009;46(3):786-802.
- Wang M, Li P. A review of deformation models in medical image registration. *J Med Biol Eng*. 2019;39(1):1-17.
- Vercauteran T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. *NeuroImage*. 2009;45(1):S61-S72.
- Santos J, Chaudhari AJ, Joshi AA, et al. Non-rigid registration of serial dedicated breast CT, longitudinal dedicated breast CT and PET/CT images using the diffeomorphic demons method. *Physica Med*. 2014;30(6):713-717.
- Liu S, Liu S, Zhang F, et al. Longitudinal Brain MR Retrieval with Diffeomorphic Demons Registration: What Happened to Those Patients with Similar Changes? In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2015:588-591.
- Markel D, Levesque I, Larkin J, Léger P, El Naqa I. A 4D biomechanical lung phantom for joint segmentation/registration evaluation. *Phys Med Biol*. 2016;61(19):7012.
- Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. *Mach Vision Appl*. 2020;31(1):1-18.
- Chen X, Diaz-Pinto A, Ravikumar N, Frangi AF. Deep learning in medical image registration. *Prog Biomed Eng*. 2021;3(1):012003.
- Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. Vox-eMorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. 2019;38(8):1788-1800.
- Sumner RW, Schmid J, Pauly M. Embedded Deformation for Shape Manipulation. In: *ACM SIGGRAPH 2007 Papers*. ACM Press, 2007:80-es.
- Song J, Wang J, Zhao L, Huang S, Dissanayake G. Dynamic reconstruction of deformable soft-tissue with stereo scope in minimal invasive surgery. *IEEE Rob Autom Lett*. 2017;3(1):155-162.
- Zaffino P, Raudaschl P, Fritscher K, Sharp GC, Spadea MF. Plastimatch MABS, an open source tool for automatic image segmentation. *Med Phys*. 2016;43(9):5155-5160.
- Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging*. 2009;29(1):196-205.
- Rister B, Horowitz MA, Rubin DL. Volumetric image registration from invariant keypoints. *IEEE Trans Image Process*. 2017;26(10):4900-4910.
- Sorkine O, Alexa M. As-Rigid-as-Possible Surface Modeling. In: Belyaev A, Garland M, eds. *Eurographics Symposium on Geometry Processing*. The Eurographics Association, 2007:109-116.

32. Lynch KM, Park FC. *Modern Robotics*. Cambridge University Press, 2017.
33. Nocedal J, Wright S. *Numerical Optimization*. Springer Science & Business Media, 2006.
34. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging*. 2012;30(9):1323-1341.
35. McCormick MM, Liu X, Ibanez L, Jomier J, Marion C. ITK: enabling reproducible research and open science. *Front Neuroinf*. 2014;8:13.
36. Murphy K, Ginneken vB, Klein S, et al. Semi-automatic construction of reference standards for evaluation of image registration. *Med Image Anal*. 2011;15(1):71-84.
37. Ribeiro AS, Nutt DJ, McGonigle J. Which Metrics Should Be Used in Non-Linear Registration Evaluation? In: Navab N, Wells WM, Hornegger J, Frangi AF, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015:388-395.
38. Xu Z, Lee CP, Heinrich MP, et al. Evaluation of six registration methods for the human abdomen on clinically acquired CT. *IEEE Trans Biomed Eng*. 2016;63(8):1563-1572.
39. Rafii-Tari H, Lessoway VA, Kamani AA, Abolmaesumi P, Rohling R. Panorama ultrasound for navigation and guidance of epidural anesthesia. *Ultrasound Med Biol*. 2015;41(8):2220-2231.
40. Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. Meshlab: An Open-Source Mesh Processing Tool. In: Scarano V, De Chiara R, Erra U, eds. *Eurographics Italian Chapter Conference*. The Eurographics Association, 2008:129-136.
41. Viergever MA, Maintz JA, Klein S, Murphy K, Staring M, Pluim JP. A survey of medical image registration—under review. *Med Image Anal*. 2016;33:140-144.
42. Ortmaier T, Groger M, Boehm DH, Falk V, Hirzinger G. Motion estimation in beating heart surgery. *IEEE Trans Biomed Eng*. 2005;52(10):1729-1740.

How to cite this article: Mao Z, Zhao L, Huang S, Jin T, Fan Y, Lee AP-W. Complete region of interest reconstruction by fusing multiview deformable three-dimensional transesophageal echocardiography images. *Med Phys*. 2022;1-13.
<https://doi.org/10.1002/mp.15910>