ORIGINAL ARTICLE

Manifold-Guided Multi-Reference Registration for Left-Ventricular Motion Tracking in 3D Echocardiography

Mohammad Rashidi [1], Nader Jafarnia Dabanloo [1*], Fereidoun Nowshiravan Rahatabad [1], Maryam Shojaeifard [2]

[1] *Department of Biomedical Engineering, SR.C., Islamic Azad University, Tehran, Iran*

[2] *Echocardiography Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran*

*Corresponding Author: Nader Jafarnia Dabanloo
Email: jafarnia@srbiau.ac.ir

Received: 24 August 2026 / Accepted: 06 September 2026

Abstract

Purpose: To develop and evaluate a data-driven multi-reference registration strategy for left-ventricular motion tracking in three-dimensional echocardiography, using an intrinsic low-dimensional cardiac trajectory to determine the temporal placement of local registration references.

Materials and Methods: Twelve healthy-volunteer three-dimensional echocardiographic sequences from a publicly available myocardial-tracking benchmark were analyzed. We evaluated six low-dimensional representation methods (PCA, t-SNE, Diffusion Maps, LLE, UMAP, and Isomap) in standard and time-embedded forms using geometric and physiologically relevant criteria. Four reference frames were automatically extracted from the resulting two-dimensional trajectories. Time-embedded PCA and Diffusion Maps produced identical four-reference sets for all analyzed cases and therefore resulted in the same proposed registration configuration. We compared five primary registration configurations—single-reference, sequential frame-to-frame, the proposed manifold-guided multi-reference strategy, a partial B-Spline–Fourier (BSF) implementation, and a Fixed Four-Reference Control—using landmark-based Target Registration Error (TRE) in millimeters. We also evaluated an additional Two-Reference Control (K=2) that retained only the initial ED and principal internal references as a reference-number sensitivity analysis.

Results: For both time-embedded PCA and Diffusion Maps, the automatically selected principal internal reference coincided with the expert-defined end-systolic (ES) frame in 10 of 12 cases and differed by one frame in the remaining two cases. The proposed strategy achieved mean-cycle, ES-frame, and final-frame TRE values of 6.8 ± 1.6 mm, 11.5 ± 2.7 mm, and 4.0 ± 1.2 mm, respectively. For mean-cycle TRE, Strategy 3 showed a median paired reduction of 0.99 mm relative to the Fixed Four-Reference Control [95% CI: 0.94–1.08; |r_rb|=0.91; p_Holm=0.033]. At ES, the corresponding reduction relative to the Fixed Four-Reference Control was 2.000 mm [1.945–2.055; |r_rb|=1.000; p_Holm=0.021]. The K=2 control showed no significant mean-cycle difference from the other configurations, but Strategy 3 reduced ES-frame TRE relative to K=2 by 1.265 mm [0.304–2.445; |r_rb|=0.909; p_Holm=0.029] and final-frame TRE by 1.156 mm [0.483–2.142; |r_rb|=1.000; p_Holm=0.029]. No statistically significant final-frame difference was detected between Strategy 3 and the Fixed Four-Reference Control.

Conclusion: Within this benchmark of 12 healthy-volunteer sequences, the findings support the internal technical feasibility of using the intrinsic low-dimensional organization of a cardiac image sequence for data-driven reference-frame placement within a multi-reference registration architecture. Time-embedded PCA and Diffusion Maps produced identical reference configurations, indicating no additional reference-selection benefit from the more computationally demanding Diffusion Maps representation in the present benchmark. The K=2 sensitivity analysis showed phase-specific effects of removing the intermediate references, while the absence of a significant mean-cycle difference prevents interpretation of K=4 as uniformly superior or empirically optimal. Independent validation in larger and clinically heterogeneous datasets is required before broader generalization.

Keywords: 3 Dimentional Echocardiography; Left Ventricular Motion; Image Registration; Manifold Learning; Myocardial Tracking.

PROOF

Copyright © 2026 Tehran University of Medical Sciences.
This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International license (https://creativecommons.org/licenses/by-nc/4.0/). Noncommercial uses of the work are permitted, provided the original work is properly cited.
DOI:

1. Introduction

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide and account for a substantial proportion of global deaths [1]. In this context, assessment of left ventricular (LV) function, particularly regional wall motion, provides important information about myocardial mechanics, and abnormalities in regional wall motion are among the major indicators used in the evaluation of ischemic heart disease [2,3]. Echocardiography is one of the most widely used cardiac imaging modalities because it is noninvasive, relatively inexpensive, broadly accessible, and provides high temporal resolution [2,4]. Three-dimensional echocardiography also enables evaluation of LV geometry and motion with less dependence on the geometric assumptions inherent to two-dimensional imaging [2,4]. Nevertheless, clinical assessment of wall motion still relies predominantly on visual interpretation and qualitative classification of myocardial regions into categories such as normal, hypokinetic, akinetic, or dyskinetic [2,5]. This type of assessment is dependent on observer experience and is affected by both interobserver and intraobserver variability [6]. In addition, speckle, noise, and ultrasound artifacts can make the delineation of boundaries and the identification of subtle motion patterns more difficult [7,8]. These inherent limitations of echocardiographic imaging have also been specifically documented in the literature [7,8]. Therefore, the development of automated and quantitative methods for myocardial motion tracking may improve the objectivity, accuracy, and reproducibility of regional LV functional assessment.

Quantification of myocardial motion and deformation requires tracking cardiac tissue throughout the cardiac cycle so that parameters such as strain and strain rate can be derived from the resulting displacement field [9,10]. This task is challenging in echocardiography because cardiac motion is nonrigid, multidirectional, and time-dependent, while variable image quality, speckle, and artifacts further complicate tissue correspondence between frames [9,11,12]. Speckle-tracking methods have been widely used for this purpose, but their performance depends on image quality, frame rate, and the persistence of speckle patterns, and two-dimensional methods are additionally limited by out-of-plane motion [11–14]. The EACVI/ASE standard also specifically defines strain and strain rate within the framework of deformation imaging and speckle tracking [9]. In contrast, image registration provides a framework for estimating spatial transformations between images and thereby extracting dense displacement fields that can be used for myocardial motion and deformation analysis [15–17]. Nonrigid and spatiotemporal registration models are particularly suitable for representing complex cardiac motion, although their accuracy remains affected by noise, artifacts, intensity variation, and the magnitude of deformation between images [12,15–17].

In addition to the deformation model and similarity metric, the temporal organization of the registration process also influences motion-estimation accuracy. Two fundamental strategies are single-reference registration and sequential frame-to-frame registration [15,18,19]. In the former, end-diastole (ED) is typically selected as the reference frame, and all subsequent frames are directly registered to it. This strategy avoids stepwise error accumulation and expresses all motion fields within a common coordinate system; however, as the target frame becomes temporally and structurally farther from ED, the magnitude of deformation increases and structural similarity decreases, making registration more difficult, particularly around end-systole (ES) [18,19]. In the sequential approach, each frame is registered to an adjacent frame, and the overall displacement is obtained through composition of successive transformations [15,19]. The smaller deformation between neighboring frames may simplify pairwise registration, but errors introduced at each step can propagate to subsequent transformations and, as the transformation chain becomes longer, may lead to temporal drift [15,19]. Therefore, these two strategies create a fundamental trade-off: single-reference registration avoids error propagation but must handle large deformations, whereas sequential registration deals with smaller local deformations but is susceptible to cumulative error.

Hybrid strategies have also been proposed to combine the advantages of these two approaches. Wiputra *et al.* proposed a spatiotemporal B-spline-Fourier (BSF) framework in which information from sequential registration of adjacent frames (Eulerian)

PROOF

and direct registration to the initial frame (Lagrangian) is combined using a time-dependent weighting function [20]. The resulting trajectory is then regularized using Fourier bases in the temporal dimension and B-splines in the spatial dimension to account for continuity and periodicity of cardiac motion [20]. Although this framework combines two sources of motion information after pairwise registration, it does not directly address the problem of selecting multiple appropriate reference frames distributed across different phases of the cardiac cycle.

Related studies have demonstrated the potential of low-dimensional representations and multi-reference formulations for cardiac motion analysis and tracking. Yang *et al.* [21] employed motion-manifold learning in 3D echocardiography to generate motion priors for one-step-forward prediction of LV motion, thereby incorporating learned cardiac dynamics directly into the tracking process. Rohé *et al.* [22] introduced Barycentric Subspaces as a low-dimensional multi-reference representation of cardiac motion, using multiple reference images to overcome limitations associated with describing the entire cardiac cycle relative to a single reference frame and incorporating the resulting subspace as a prior for cardiac motion registration. Gifani *et al.* [24] used a low-dimensional echocardiographic trajectory derived using LLE to identify ED and ES in two-dimensional echocardiography, demonstrating that cardiac phase information can be represented in a compact image-derived trajectory. These studies highlight the potential of low-dimensional cardiac representations for motion analysis and cardiac-state identification. The present study extends this concept to volumetric 3D echocardiography and to a different downstream objective: the low-dimensional representation is not incorporated into the pairwise registration algorithm as a motion prior, but is instead used to determine the temporal locations of multiple local registration references.

Accordingly, the present study proposes a manifold-guided multi-reference registration strategy for LV motion tracking in 3D echocardiography. The framework uses the intrinsic low-dimensional trajectory of each cardiac sequence to identify temporally and structurally distinct reference states. The sequence is then partitioned into locally referenced temporal intervals, and the resulting displacement fields are composed into a common end-diastolic coordinate system while retaining the same underlying pairwise registration algorithm. The proposed strategy is evaluated against conventional single-reference, sequential, partial BSF, and fixed four-reference registration configurations using landmark-based tracking error on a public 3D echocardiography benchmark. The central methodological question is therefore whether intrinsic image-derived trajectory information can provide useful data-driven placement of a limited set of local registration references.

Each 3D echocardiographic cardiac cycle consists of a sequence of three-dimensional volumes and can therefore be regarded as a set of observations in a very high-dimensional space. Organized variation within such high-dimensional data can often be represented in an intrinsic lower-dimensional space [23–27]. The term low-dimensional representation is used collectively in the present study to avoid implying that the proposed framework requires a specifically nonlinear mapping. In the broader dimensionality-reduction literature, manifold-learning formulations have been described, including both linear and nonlinear approaches, with PCA representing a canonical linear case [25]. Six commonly used low-dimensional representation methods were therefore compared in the present study. Their role was to provide alternative two-dimensional trajectories from which the same deterministic reference-selection rule could identify cardiac states, rather than to establish superiority of a particular linear or nonlinear class. The specific methodological contribution is the use of the intrinsic low-dimensional trajectory of an individual 3D echocardiographic sequence to guide reference-frame placement and temporally partition the registration problem into locally referenced intervals. Accordingly, the principal contribution lies in data-driven reference-frame organization rather than in the use of a particular dimensionality-reduction algorithm.

2. Materials and Methods

2.1. Study Overview

The study framework consisted of four stages: preparation of the 3D echocardiographic data and

PROOF

landmark trajectories, extraction of low-dimensional representations, automatic identification of reference frames, and comparison of registration strategies. Data were obtained from the publicly available Benchmarking Framework for Myocardial Tracking and Deformation Algorithms [28]. This benchmark is a peer-reviewed database published in Medical Image Analysis and was developed for validation of cardiac motion-tracking algorithms [28]. Different low-dimensional representation methods were implemented in both standard and time-embedded forms and were evaluated using geometric and physiological criteria. Four reference frames were then extracted from the resulting trajectories, and the principal internal reference was evaluated against the available expert-defined ES annotation. In the final stage, five primary configurations—single-reference registration, sequential frame-to-frame registration, the proposed manifold-guided multi-reference strategy, a partial BSF implementation, and a Fixed Four-Reference Control—were compared in terms of tracking accuracy. An additional Two-Reference Control (K=2) was evaluated as a reference-number sensitivity analysis. All data preparation, dimensionality-reduction algorithms, reference-frame extraction procedures, registration strategies, and quantitative analyses were implemented and executed in MATLAB R2022b (MathWorks, Natick, MA, USA).

2.2. Dataset

Three-dimensional echocardiographic data were obtained from the publicly available myocardial-tracking benchmark [28]. The original database contains MR and 3D ultrasound data from a dynamic phantom and 15 healthy volunteers, comprising a total of 1,158 image volumes [28]. The 3D ultrasound data were acquired using an iE33 system (Philips Healthcare, Best, The Netherlands) equipped with an X3-1 matrix-array transducer in full-volume acquisition mode. Only the 3D ultrasound modality was used in the present study. Prior to any dimensionality-reduction analysis, reference-frame extraction, image registration, or tracking-error evaluation, all downloaded sequences were inspected for temporal completeness. In the version of the benchmark data available to us, benchmark Cases 7, 8, and 9 contained incomplete temporal data and did not provide a complete ED-to-ED cardiac cycle. This exclusion was not predefined by the original benchmark; rather, these cases were identified as non-analyzable during our initial data-completeness assessment. Because a complete cardiac cycle was required by the temporal analysis pipeline, these three cases were excluded, and the remaining 12 cases constituted the analyzable dataset from the outset. No case was excluded on the basis of low-dimensional representation characteristics, reference-frame selection, registration performance, TRE, or any other study outcome. Attempts were made to obtain the missing temporal data for the incomplete sequences, but complete versions could not be retrieved. Each included case consisted of a temporal sequence of three-dimensional LV volumes.

The motion ground truth in the benchmark was defined using 12 landmarks distributed over four LV walls at three ventricular levels. These landmarks were manually tracked throughout the entire cardiac cycle by two observers in 3D tagged MRI and were subsequently transformed into the 3D ultrasound coordinate system using a point-based similarity transformation [28]. Because the temporal sampling of the landmark trajectories differed from that of the corresponding ultrasound sequences, the landmark coordinates were temporally resampled using cubic spline interpolation to match the number of ultrasound frames. These transformed and temporally matched landmark coordinates were used as ground truth for tracking evaluation in the present study. The benchmark data covered the whole cardiac cycle [28]. In the present analysis, each sequence was therefore treated as an ED-to-ED cardiac cycle, with the first frame representing the initial end-diastolic state and the end of the sequence representing completion of the cycle and return to end-diastole. Accordingly, the final frame was physiologically interpreted as the end-diastolic state completing the analyzed cardiac cycle. This frame was defined by the temporal endpoint of the sequence and was not independently detected by the proposed manifold-based algorithm.

The reconstructed 3D ultrasound volumes were analyzed in the form provided by the benchmark. No additional spatial resampling, cropping, intensity normalization, or myocardial masking was applied to the ultrasound images before low-dimensional analysis or image registration. Apart from the

PROOF

temporal resampling of landmark coordinates described above, no temporal preprocessing was applied to the ultrasound image sequences. Spatial voxel dimensions were retained and were subsequently used to convert landmark-coordinate differences to physical TRE values in millimeters. All low-dimensional representation construction, reference selection, and registration operations were performed independently within each subject's cardiac sequence; voxel intensities were not compared across subjects, and no registration was performed between independently acquired subjects or examinations.

2.3. Standard and Time-Embedded Representations

Each three-dimensional volume was initially vectorized and treated as a high-dimensional observation, denoted by x_t for frame t. In addition to this standard representation, a two-frame time-delay embedding with a one-frame delay was constructed to explicitly incorporate local temporal information into the feature representation. For each frame, the time-embedded observation was defined as $z_t = [x_t, x_{t+1}]$. Because the analyzed sequences represented complete cardiac cycles, circular boundary handling was applied, with the final embedded observation defined as $zT = [xT, x_1]$. The idea of reconstructing the structure of a dynamical system from delayed observations originates from delay-coordinate embedding theory [29]. By incorporating the immediate temporal evolution of each cardiac state, this representation was expected to improve the organization of the cardiac-cycle trajectory and facilitate separation of morphologically similar states occurring during different phases of the cycle, particularly contraction and relaxation. Both the standard and time-embedded representations were then mapped to a two-dimensional space, and the embedded points were connected according to their temporal order to form the cardiac-cycle trajectory.

2.4. Low-Dimensional Representation Methods

Six commonly used low-dimensional representation methods were evaluated: PCA, t-SNE, Diffusion Maps, LLE, UMAP, and Isomap. These methods were treated as alternative mappings of each high-dimensional cardiac image sequence into a two-dimensional trajectory. The comparison was intended to assess the dependence of subsequent reference-frame selection on the representation method rather than to establish superiority of a particular class of dimensionality-reduction techniques. The same deterministic reference-selection procedure was subsequently applied to every representation. PCA provides a linear low-dimensional representation by projecting the observations onto orthogonal directions that account for maximal variance [30]. Because of its computational simplicity, deterministic solution, and direct geometric interpretation, PCA was included as a representative linear approach rather than solely as a baseline.

t-SNE is a nonlinear method designed to preserve local neighborhoods by modeling similarities between points as probability distributions in the original and low-dimensional spaces and minimizing their Kullback-Leibler divergence [31]. The original study by van der Maaten and Hinton specifically introduced this approach for visualization of high-dimensional data in two or three dimensions [31].

Diffusion Maps construct a similarity graph and define a Markov diffusion process to recover the intrinsic geometry of the data through the eigenvectors of the diffusion operator [32,33]. The resulting diffusion distance allows the geometric organization of the data to be characterized at multiple scales [32,33].

LLE reconstructs each sample as a linear combination of its nearest neighbors and determines a low-dimensional embedding that preserves these local relationships [34]. The original study by Roweis and Saul explicitly introduced LLE as an unsupervised, neighborhood-preserving method for nonlinear dimensionality reduction [34].

UMAP is a manifold-learning method based on concepts from Riemannian geometry and algebraic topology that generates a low-dimensional embedding by constructing a fuzzy representation of neighborhood relationships and optimizing the corresponding structure in the target space [35]. The method was designed for visualization and dimensionality reduction of high-dimensional data, and its original description emphasizes both

PROOF

scalability and preservation of meaningful data structure [35].

Isomap constructs a neighborhood graph, computes shortest paths, and applies classical multidimensional scaling to approximate geodesic distances on the manifold [36]. The original study by Tenenbaum *et al.* introduced Isomap specifically for recovering the global geometry of nonlinear low-dimensional structures [36]. Graph connectivity was evaluated prior to embedding, and the number of neighbors was increased when necessary to obtain a connected graph.

Computational execution time was additionally evaluated for all six dimensionality-reduction methods in both standard and time-embedded representations. Each method was executed 20 times for each of the 12 analyzed cases under the same computational environment, and execution time was summarized as mean ± standard deviation across the repeated runs and cases. Timing was applied to the dimensionality-reduction computation using the already prepared standard or time-embedded input representation; construction of the time-delay representation itself was not included in the measured execution time. Because each time-embedded observation concatenated two consecutive volumetric observations, its input feature dimensionality was approximately twice that of the corresponding standard representation. All timing experiments were performed in MATLAB R2022b on a Windows 11 computer equipped with an Intel Core i3 processor and 8 GB of RAM.

2.5. Evaluation of Low-Dimensional Representations

Embedding quality was evaluated using four geometric metrics and two measures related to physiological variation. Trustworthiness quantifies the extent to which neighborhoods formed in the low-dimensional space correspond to true neighborhoods in the original space, whereas Continuity measures the extent to which neighborhoods present in the original space remain preserved after dimensionality reduction; for both measures, values closer to one are preferable [37,38]. Kruskal Stress was used to quantify distortion of pairwise distances, with lower values indicating better preservation of the original distance structure [39]. Shepard-Spearman correlation was also used to evaluate preservation of the relative ordering of pairwise distances between the original and embedded spaces [37–39].

To obtain a physiologically related scalar descriptor of temporal variation in landmark geometry, a landmark-derived surrogate volume was calculated independently from the 12 three-dimensional landmark trajectories of each observer. The landmark coordinate trajectories were first temporally smoothed using a five-frame Savitzky–Golay filter. At each frame, a three-dimensional convex hull was then constructed from the landmark configuration, and the enclosed volume was calculated by decomposing the hull into tetrahedra formed between each triangular hull face and the arithmetic mean of the landmark coordinates. This quantity was interpreted as a landmark-based geometric surrogate of relative LV volumetric variation rather than as a direct or segmentation-based measurement of LV cavity volume. The five-frame Savitzky–Golay smoothing described here was applied only to the landmark trajectories used to calculate the surrogate-volume profile and was not applied to the ultrasound volumes or to the landmark trajectories used for TRE calculation.

$$V(t) = (1/6)\,\Sigma_k\,|(a_k(t) - c(t)) \cdot [(b_k(t) - c(t)) \times (d_k(t) - c(t))]|$$

where $a_k(t)$, $b_k(t)$, and $d_k(t)$ are the vertices of the k-th triangular face of the convex hull at frame t, c(t) is the arithmetic mean of the landmark coordinates, and the summation is taken over all triangular faces of the hull. To quantify correspondence between the two-dimensional low-dimensional trajectory and the landmark-derived surrogate-volume profile, Pearson correlation coefficients were calculated separately between the surrogate volume V and each low-dimensional coordinate, Y_1 and Y_2. The reported measure was $r_{max} = \max(|r(V,Y_1)|, |r(V,Y_2)|)$. The maximum absolute coordinate-wise correlation was used because the two latent coordinates were unlabeled and their signs were arbitrary; thus, volume-related variation could appear predominantly along either coordinate and with either sign. The locations of ED and ES, trajectory continuity, and representation of the contraction and relaxation trajectories were also qualitatively assessed.

PROOF

2.6. Reference-Frame Extraction

After the generation of the two-dimensional trajectory, four reference frames were extracted using a fixed, deterministic distance-based procedure. The first reference was the initial ED frame, $R_1 = 1$, and the terminal frame T represented the subsequent ED state completing the cardiac cycle. For embedded points y_t in the two-dimensional low-dimensional space, the Euclidean distance between frames i and j was defined as $d(i,j) = \|y_i - y_j\|_2$.

The principal internal reference R_3 was selected from frames between the two ED endpoints as the point with the greatest combined distance from the initial and terminal ED states: $R_3 = \arg\max [d(t,1) + d(t,T)]$, for $1 < t < T$. The intermediate contraction reference was then selected only from frames temporally between the initial ED and R_3 according to $R_2 = \arg\max [d(t, R_1) + d(t, R_3)]$, for $1 < t < R_3$. Similarly, the intermediate relaxation reference was selected only from frames between R_3 and the terminal ED according to $R_4 = \arg\max [d(t, R_3) + d(t, T)]$, for $R_3 < t < T$. This staged, phase-constrained procedure was designed to identify structurally separated reference states while preserving their temporal order within the cardiac cycle.

The reference-selection procedure was designed to identify structurally distinct states from the geometry of the low-dimensional cardiac trajectory rather than to directly detect a predefined physiological phase. The same distance-based criterion was applied to all candidate representations without using the expert-defined ES frame or its temporal location as an input to the algorithm. Thus, the principal internal reference was determined solely from its geometric separation from the initial and terminal ED states in the corresponding low-dimensional representation.

Because independent ground-truth annotations were not available for the two intermediate contraction and relaxation references, their selection accuracy could not be directly quantified. In contrast, the benchmark provided an expert-defined ES frame for each case. ES therefore provided an independently annotated and physiologically interpretable cardiac state against which the principal internal reference selected from the low-dimensional geometry could be evaluated. For each candidate representation, the ES identification error was defined as the absolute frame difference $|R_3 - FES|$, where FES denotes the expert-defined ES frame. The expert-defined ES annotations were not supplied to the reference-selection procedure and did not affect its deterministic distance-based criterion or parameters. After reference extraction, however, agreement with expert-defined ES was used as one measure of physiological construct validity and, together with the other representation-quality measures, contributed to assessment of the candidate representations.

The six dimensionality-reduction techniques were treated as alternative within-sequence representations rather than population-level predictive models. Each representation was constructed independently for each cardiac sequence, and no parameters were learned across subjects. Registration TRE was not used to select or tune a dimensionality-reduction representation. The candidate representations and final registration experiment were nevertheless examined within the same 12-case cohort, and no independent training/validation split was used for the representation-assessment stage. Time-embedded PCA and Diffusion Maps ultimately produced identical complete four-reference configurations in all analyzed cases; consequently, the final Strategy 3 registration input, displacement-field composition, and TRE were invariant to the choice between these two representations.

The four-reference configuration was fixed as an architectural design choice before the TRE-based registration comparison, based on the physiological organization of the cardiac cycle and the qualitative geometry observed in the low-dimensional trajectories rather than on optimization of registration error. The trajectories showed an outbound contraction branch extending from the initial ED state toward a geometrically distant internal state, followed by a return relaxation branch toward the terminal ED state. Particularly in the time-embedded PCA and Diffusion Maps representations, the outbound and return paths were not simply coincident linear trajectories, indicating that intermediate states on the contraction and relaxation branches could provide distinct local reference information.

Within this two-branch architecture, K=4 was selected as the smallest symmetric configuration containing the initial ED reference, the principal internal reference, one intermediate contraction

PROOF

reference, and one intermediate relaxation reference. To assess sensitivity to reducing the number of references without changing the basic physiological interpretation of the architecture, an additional K=2 configuration was evaluated by retaining R_1 and R_3 and simultaneously removing R_2 and R_4. This provided a nested reduction in which both branch-specific intermediate references were removed together. A K=3 configuration was not included because there is no unique structurally symmetric three-reference counterpart of the present two-branch design. Once the initial ED and principal internal references are retained, a third internal reference must preferentially represent either the contraction or relaxation branch, or a different reference-selection rule must be introduced. Such a comparison would therefore alter both reference number and architectural structure. Accordingly, the K=2 analysis was used as the interpretable reduced-reference sensitivity test, and no claim is made that K=4 is empirically or universally optimal.

2.7. Registration Strategies

Five primary registration configurations were compared using the same underlying pairwise registration framework, and an additional Two-Reference Control (K=2) was evaluated as a sensitivity analysis. The principal differences among configurations were the temporal organization, reference-frame selection, and composition of the resulting displacement fields. All pairwise registrations were performed using the same intensity-based nonrigid Demons registration algorithm [40,41], such that differences in tracking performance primarily reflected the temporal organization of registration rather than differences in the pairwise registration method. For each image pair, a dense three-dimensional displacement field was estimated using a multiresolution Demons framework with three pyramid levels and 100 iterations at each resolution level. Diffusion-like regularization was applied by Gaussian smoothing of the accumulated displacement field at each iteration, using a standard deviation of 1.0, which was maintained across all pyramid levels. These registration parameters were kept identical for all evaluated configurations. All computational analyses were performed in MATLAB R2022b (MathWorks, Natick, MA, USA).

2.7.1. Single-Reference Registration — Strategy 1

The initial ED frame was selected as the only reference, and all subsequent frames were directly registered to it:

$$\Phi_{0\rightarrow t} = \text{Registration}(I_0, I_t) \quad (1)$$

This strategy does not involve stepwise propagation of registration error, although registration of frames with large deformation relative to ED can be more challenging.

2.7.2. Sequential Frame-to-Frame Registration — Strategy 2

Each frame was registered to the immediately preceding frame, and the overall transformation relative to the initial frame was obtained through composition of successive transformations:

$$\Phi_{0\rightarrow t} = \Phi_{t-1\rightarrow t} \circ \cdots \circ \Phi_{0\rightarrow 1} \quad (2)$$

Although pairwise registration in this strategy is performed between more similar adjacent frames, errors introduced at individual steps may accumulate along the transformation chain [15,19].

2.7.3. Proposed Hybrid Multi-Reference Strategy — Strategy 3

The cardiac cycle was divided into four intervals using four manifold-derived reference frames: R_1, corresponding to the initial ED; R_2, representing an intermediate contraction state; R_3, representing the principal internal reference identified from manifold geometry and subsequently evaluated against the expert-defined ES frame; and R_4, representing an intermediate relaxation state. Each frame was registered only to its corresponding local reference, and the final transformation was calculated as:

$$\Phi_{0\rightarrow t} = \Phi_Rq\rightarrow t \circ \Phi_{0\rightarrow Rq} \quad (3)$$

When necessary, the transformation of a reference frame relative to the initial ED was obtained from the short chain of transformations between consecutive reference frames:

PROOF

$$\Phi_{0\to Rq} = \Phi_{Rq-1\to Rq} \circ \cdots \circ \Phi_{R1\to R2} \quad (4)$$

Displacement fields were composed as spatial transformations rather than by simple vector addition. If:

$$\Phi(x) = x + u(x) \quad (5)$$

Then;

$$(\Phi_B \circ \Phi_A)(x) = x + u_A(x) + u_B[x + u_A(x)] \quad (6)$$

2.7.4. Partial B-Spline–Fourier (BSF) Implementation — Strategy 4

As an additional hybrid comparison strategy, a partial implementation of the B-spline–Fourier (BSF) framework proposed by Wiputra *et al.* [20] was evaluated. The implemented stages combined the Eulerian displacement field derived from registration between consecutive frames with the Lagrangian displacement field obtained from direct registration to the initial frame using the time-dependent weighting formulation described in the original framework. In the forward-marching scheme,

$$X^{t+1}_march = w(t)[X^{t}_march + dX_{t\to t+1}] + [1 - w(t)][X^{0} + dX_{0\to t+1}] \quad (7)$$

Where;

$$w(t) = ((P - t)t)/(P/2)^2 \quad (8)$$

is the time-dependent weighting function, and P denotes the length of the cardiac cycle [20]. The resulting trajectory was subsequently modeled using Fourier bases in the temporal dimension:

$$X^{t}_BSF = \Sigma^{f=N}_{f=0} F_f \cos(2\pi ft/P) + F_\{f+N+1\} \sin(2\pi ft/P) \quad (9)$$

and spatial variation of the Fourier coefficients was represented using cubic B-splines [20]. The final consistency-correction stage of the original BSF framework was intentionally not applied in the present comparison. This stage represents a post-processing refinement applied after estimation of the primary motion trajectory and, in principle, analogous consistency-refinement procedures could also be applied to displacement fields obtained from the other registration strategies evaluated here. The objective of the present comparison was to isolate differences arising from the core temporal registration architectures and reference-selection/composition strategies. Applying an additional consistency-correction stage only to the BSF configuration would therefore introduce a post-processing advantage not shared by the other methods. For this reason, the correction stage was omitted to maintain a more comparable evaluation of the underlying strategies. Accordingly, Strategy 4 is designated throughout this study as a partial BSF implementation, and its results should not be interpreted as representing the performance of the complete BSF pipeline.

2.7.5. Fixed Four-Reference Control

To isolate the contribution of manifold-guided reference selection from the effect of using four reference frames, an additional control configuration was evaluated using the same multi-reference registration architecture, the same pairwise Demons registration settings, and the same displacement-field composition procedure as Strategy 3. In this control, however, the four reference frames were selected without manifold information at approximately 0%, 25%, 50%, and 75% of the cardiac cycle. For a sequence containing T frames, the q-th fixed reference was defined as Rfixed (q) = 1 + round [(q - 1) (T - 1)/4], for q = 1, 2, 3, 4.

No expert-defined ES information or manifold-derived information was used to determine these fixed reference locations. This Fixed Four-Reference Control therefore preserved the number of reference frames and the registration architecture of the proposed strategy while removing the manifold-guided reference-selection component.

2.7.6. Two-Reference Control (K=2)

To provide a nested sensitivity analysis of reference number within the proposed two-branch architecture, an additional Two-Reference Control (K=2) was evaluated. This configuration retained the initial ED reference (R_1) and the principal internal reference (R_3), while the intermediate contraction and relaxation references (R_2 and R_4) were removed. The same pairwise Demons registration parameters, local-

PROOF

reference interval logic, and displacement-field composition procedures used for Strategy 3 were retained. The K=2 control was therefore intended as an ablation of the two branch-specific intermediate references rather than as an independently optimized registration strategy.

2.8. Tracking Evaluation

The estimated position of landmark *p* at time *t* was obtained by applying the estimated transformation to its initial position:

$$\hat{x}_{p,t} = \Phi_{0\rightarrow t}(x_{p,0}) \quad (10)$$

The estimated landmark positions and corresponding ground-truth coordinates were represented in the 3D ultrasound voxel coordinate system. Before the calculation of tracking error, coordinate differences along each spatial dimension were converted to physical distances using the corresponding voxel spacing of each ultrasound sequence. If sx, sy, and sz denote the voxel dimensions in millimeters along the x, y, and z directions, respectively, and Δx, Δy, and Δz denote the corresponding voxel-coordinate differences between estimated and ground-truth landmark positions, the physical Target Registration Error for landmark *p* at frame *t* was calculated as:

$$TRE_{p,t}(mm) = \sqrt{[(sx\Delta x_{p,t})^2 + (sy\Delta y_{p,t})^2 + (sz\Delta z_{p,t})^2]} \quad (11)$$

All TRE values reported in this study are therefore expressed in millimeters. Three metrics were extracted: mean-cycle TRE over the complete cardiac cycle, end-systolic (ES)-frame TRE evaluated at the expert-defined ES frame, and final-frame TRE. The first metric represented overall tracking performance, the second assessed tracking at the physiologically defined ES state, and the third evaluated tracking at the terminal end-diastolic frame completing the analyzed cardiac cycle.

2.9. Statistical Analysis

Because the six evaluated registration configurations were assessed on the same 12 cases and the assumption of normality was not satisfied, nonparametric procedures were used for inferential comparisons. Given the relatively small sample size, descriptive TRE results were summarized using both mean ± standard deviation and median [interquartile range, Q1–Q3]. Overall differences among configurations were assessed using the Friedman test at a significance level of α=0.05. When the Friedman test was significant, paired comparisons were performed using the Wilcoxon signed-rank test, and the resulting p-values were adjusted for multiple comparisons using the Holm procedure. All pairwise p-values reported in the manuscript and figures therefore represent Holm-adjusted values (p_Holm). To complement statistical significance with estimates of effect magnitude and uncertainty, statistically significant paired comparisons were additionally characterized using the median paired difference in TRE and the matched-pairs rank-biserial correlation (r_rb). Ninety-five percent confidence intervals for both measures were estimated using percentile bootstrap resampling of the paired cases. Paired differences were calculated as the first-listed configuration minus the second-listed configuration; consequently, positive values indicate greater TRE for the first-listed configuration and negative values indicate lower TRE.

3. Results

3.1. Qualitative Analysis of the Low-Dimensional Representations

Two-dimensional representations generated by all low-dimensional representation methods in both standard and time-embedded forms were evaluated for all 12 cases, and the corresponding trajectories for the complete analyzed dataset are provided in Supplementary Figures S1–S12. According to the predefined qualitative criteria, a desirable embedding was expected to produce an organized trajectory in which the ED states at the beginning and the end of the cardiac cycle were located near each other, ES occupied a distinct region, and the intermediate frames followed a continuous path. Qualitative evaluation showed that PCA and Diffusion Maps, particularly in their time-embedded forms, generated organized and stable trajectories across the different cases. Diffusion Maps generally produced an approximately closed loop in which the two ED states were located close

PROOF

together, and ES occupied a distinct region of the loop. In contrast, PCA generated a curved outbound-and-return trajectory extending from ED to ES and subsequently returning toward the final ED state. The contraction and relaxation branches did not completely overlap, and a clear hysteresis was observed between them. This characteristic was more evident in the time-embedded representations. Thus, although the qualitative geometry of the time-embedded PCA and Diffusion Maps trajectories differed, both provided suitable trajectory organization for subsequent reference-frame extraction.

Figure 1 shows an illustrative example of the embeddings generated for Case 12 of the benchmark dataset. This case was selected for visualization because the resulting trajectories could be clearly displayed across the different dimensionality-reduction methods; it was not selected on the basis of superior quantitative embedding metrics or superior reference-frame identification performance. The first row presents the standard version of each method, whereas the second row presents the corresponding time-embedded representation, with each column representing a different dimensionality-reduction algorithm. To avoid relying on a single illustrative example, Supplementary Figures S1–S12 present the corresponding trajectories for all 12 analyzed cases and all evaluated methods in both standard and time-embedded forms, with the principal cardiac-state markers shown as defined in the main manuscript.

3.2. Quantitative Evaluation of the Low-Dimensional Representations

Figure 2A–F summarizes the results of four geometric metrics and two metrics based on the landmark-derived surrogate volume. For Trustworthiness (Figure 2A), Isomap achieved high values in both the standard and time-embedded forms, with time-embedded Isomap showing the highest overall value, whereas time-embedded PCA exhibited the lowest mean performance. For Continuity (Fig. 2B), Isomap again showed the best overall performance, particularly in its time-embedded form, whereas time-embedded PCA had the lowest mean value, although individual lower values were also observed for LLE and UMAP in some cases.

For Kruskal Stress (Figure 2C), both standard and time-embedded PCA showed distinctly lower values than the other methods. For Shepard–Spearman correlation (Figure 2D), PCA and Isomap, in both standard and time-embedded forms, showed better results with relatively limited dispersion.

PROOF

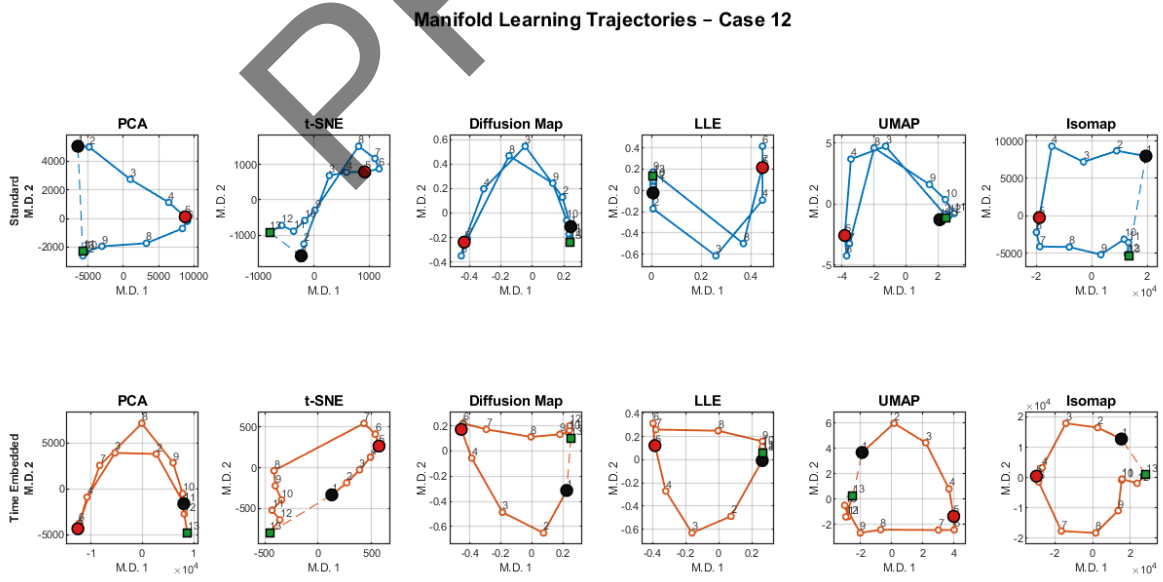


Figure 1. Illustrative two-dimensional cardiac trajectories for Case 12 of the 3D echocardiographic benchmark. The upper row shows the standard representations and the lower row shows the corresponding time-embedded representations for the six dimensionality-reduction methods. Points follow the temporal order of the cardiac cycle, and highlighted markers indicate the principal cardiac states used for interpretation and reference-frame selection. The displayed case was selected for visualization clarity rather than on the basis of superior quantitative performance

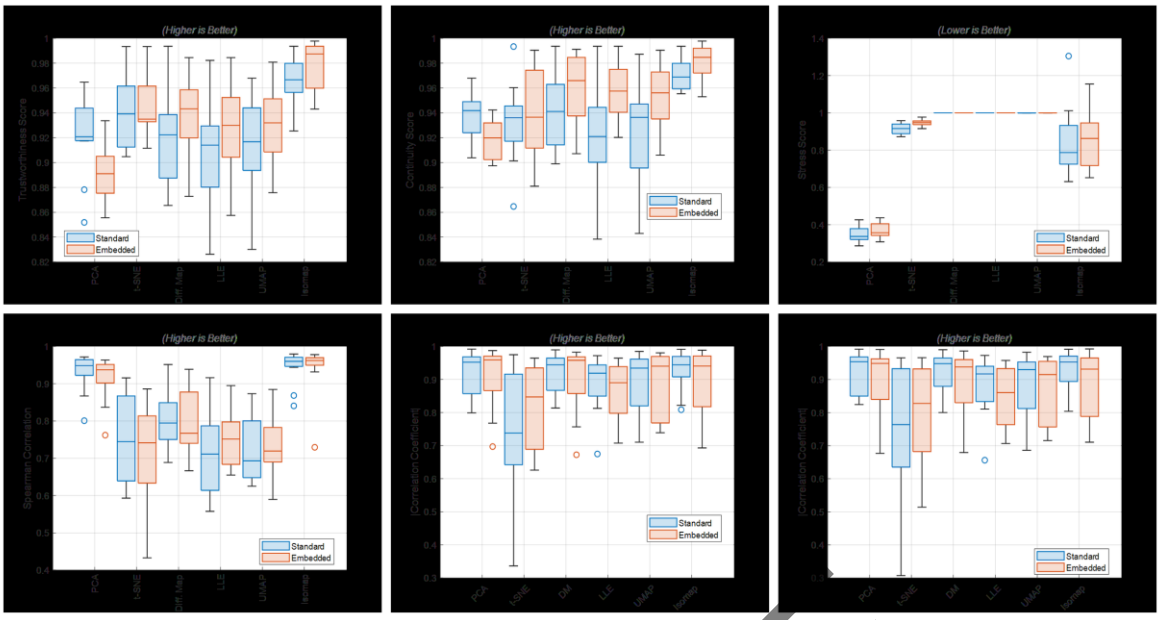


Figure 2. Quantitative evaluation of the low-dimensional representation methods in standard and time-embedded representations. (A) Trustworthiness; higher values are preferable. (B) Continuity; higher values are preferable. (C) Kruskal Stress; lower values indicate less global distance distortion. (D) Shepard–Spearman correlation; higher values indicate better preservation of the rank ordering of pairwise distances. (E) Maximum absolute correlation between low-dimensional coordinates and the landmark-derived surrogate-volume profile for Observer 1. (F) Corresponding correlation for Observer 2. Box plots summarize the 12 analyzed cases

PROOF

Figures 2E and 2F show the association between the embedding and changes in the landmark-derived surrogate volume for the annotations of the two observers. The results obtained from the two observers were highly similar. Nearly all low-dimensional representation methods showed a considerable association with this physiological surrogate, and with the exception of t-SNE, correlation coefficients were generally above approximately 0.7 for most methods. Therefore, despite differences among the methods in terms of geometric criteria, most embeddings showed a strong relationship with the physiologically relevant changes represented by this surrogate measure.

Execution-time analysis showed that time-embedded PCA required 0.460 ± 0.144 s per execution, whereas time-embedded Diffusion Maps required 1.091 ± 0.461 s. Thus, Diffusion Maps required approximately 2.37 times the execution time of PCA for the time-embedded representation, despite the two methods producing identical four-reference configurations in all analyzed cases. For the standard representations, the corresponding execution times were 0.217 ± 0.071 s for PCA and 0.518 ± 0.231 s for Diffusion Maps. Execution times for all six evaluated dimensionality-reduction methods in both representation forms are provided in Supplementary Table S2.

3.3. Reference-Frame Identification

Figure 3 shows the absolute frame-wise difference between the principal internal reference selected by the distance-based algorithm and the expert-defined ES frame. Time-embedded PCA and time-embedded Diffusion Maps produced identical reference-frame selections across the analyzed cases. For both methods, the selected principal internal reference coincided with the expert-defined ES frame in 10 of the 12 cases; in the remaining two cases, the selected reference differed from the expert annotation by one frame. Thus, the ES-based validation did not distinguish time-embedded PCA from time-embedded Diffusion Maps in terms of reference-selection performance. The expert-defined ES annotations were not inputs to the reference-selection algorithm and were not used to modify its deterministic distance criterion or parameters. After reference extraction, however, agreement with the expert-defined ES frame was used as one of the criteria for evaluating the candidate representations and, together with the other embedding-quality measures, contributed to selecting

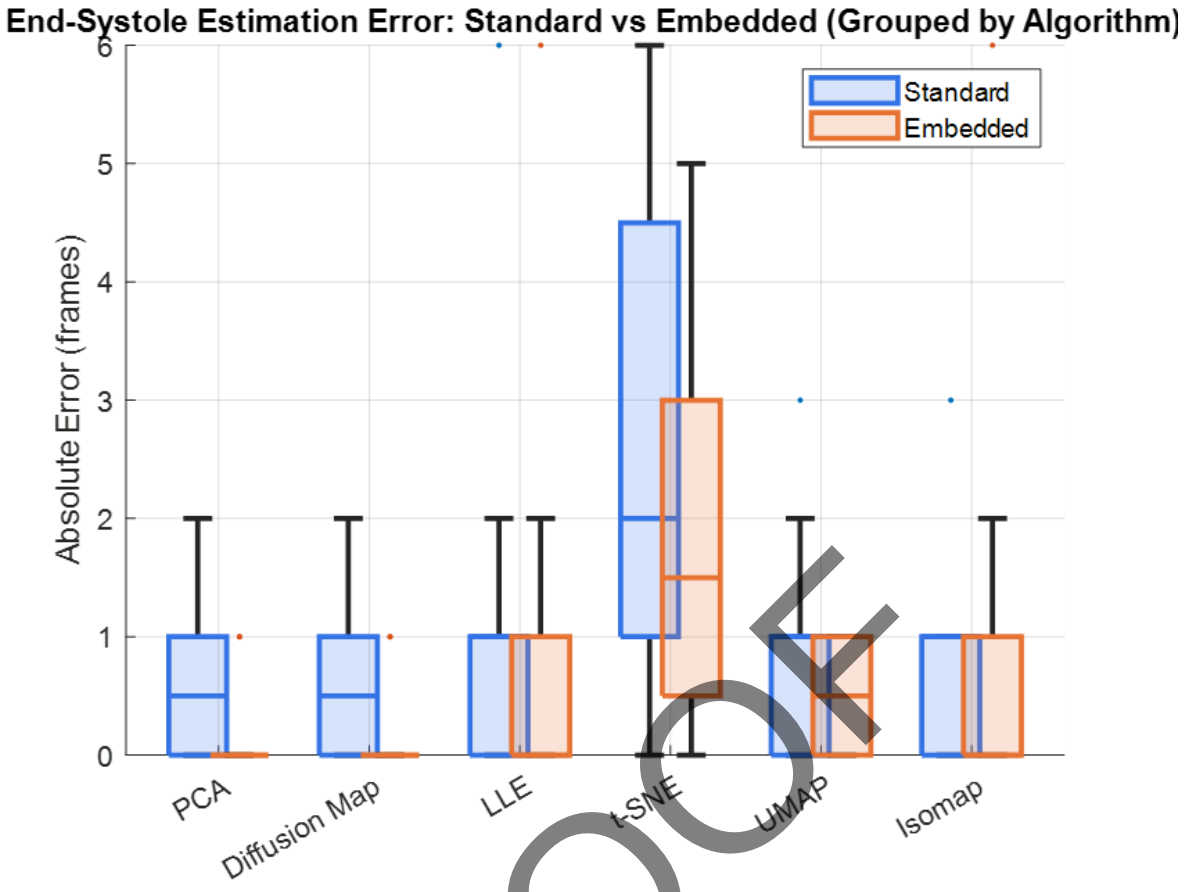


Figure 3. Absolute error, expressed in frames, between the automatically selected principal internal reference and the expert-defined end-systolic (ES) frame for each dimensionality-reduction method in standard and time-embedded forms. Lower values indicate closer agreement with the expert-defined ES annotation

PROOF

the representations carried forward to registration. Case-by-case numerical results, including the expert-defined ES frame, the automatically selected principal internal reference R_3, and the corresponding absolute ES identification error for all evaluated dimensionality-reduction methods in both standard and time-embedded forms, are provided in Supplementary Table S1.

Importantly, time-embedded PCA and time-embedded Diffusion Maps also produced identical sets of all four reference frames in every analyzed case, including the initial ED, intermediate contraction, principal internal, and intermediate relaxation references. Therefore, despite their different qualitative trajectory geometries, the two representations were indistinguishable in terms of the reference images ultimately supplied to the multi-reference registration framework. Because the complete four-reference sets were identical case by case, the downstream registration experiment contained no representation-dependent choice between time-embedded PCA and Diffusion Maps: either representation generated the same Strategy 3 input, displacement-field composition, and registration outcome. Because no independent annotations were available for the intermediate contraction and relaxation states, their selection accuracy could not be quantitatively evaluated against ground truth.

3.4. Comparison of Registration Configurations and K=2 Sensitivity Analysis

The primary quantitative comparison of the five registration configurations, together with the Two-Reference Control (K=2) sensitivity analysis, is summarized in Table 1. Corresponding case-wise distributions and pairwise statistical comparisons are shown in Figure 4, and paired effect estimates for statistically significant comparisons are reported in Table 2. Strategy 1 corresponds to single-reference registration to the initial ED frame; Strategy 2 corresponds to sequential frame-to-frame registration;

Table 1. Target Registration Error (TRE) for the evaluated registration configurations. Values are reported as mean ± standard deviation followed by median [first quartile–third quartile] across the 12 analyzed cases and are expressed in millimeters. Mean-cycle TRE represents tracking performance across the complete cardiac cycle, ES-frame TRE represents error at the expert-defined end-systolic frame, and final-frame TRE represents error at the terminal end-diastolic frame completing the analyzed cardiac cycle

Registration Configuration	Mean-cycle TRE (mm)	ES-frame TRE (mm)	Final-frame TRE (mm)
Strategy 1 – Single-reference registration	7.3 ± 1.4 7.2 [6.5–8.0]	12.5 ± 2.8 12.3 [11.0–14.4]	4.4 ± 1.4 4.1 [3.6–5.0]
Strategy 2 – Sequential frame-to-frame registration	8.1 ± 1.1 8.1 [7.5–8.6]	13.1 ± 2.5 13.1 [12.1–14.2]	6.1 ± 1.8 5.8 [5.0–7.2]
Strategy 3 – Proposed manifold-guided multi-reference	6.8 ± 1.6 7.0 [5.7–7.7]	11.5 ± 2.7 12.5 [9.5–12.8]	4.0 ± 1.2 3.9 [3.3–4.4]
Strategy 4 – Partial BSF implementation	7.3 ± 1.4 7.2 [6.7–7.9]	12.7 ± 2.6 12.4 [11.4–14.1]	5.1 ± 1.5 5.4 [4.2–5.8]
Fixed Four-Reference Control	7.7 ± 1.7 8.1 [6.7–8.7]	13.6 ± 2.8 14.5 [11.4–14.9]	4.0 ± 1.2 3.9 [3.3–4.3]
Two-Reference Control (K=2)	7.3 ± 1.4 7.2 [6.4–8.0]	12.5 ± 2.8 12.3 [11.0–14.4]	5.1 ± 1.1 5.5 [4.2–5.7]

Table 2. Paired effect-size estimates, 95% confidence intervals, and Holm-adjusted statistical comparisons for statistically significant Target Registration Error (TRE) comparisons. Paired differences are calculated as the first-listed configuration minus the second-listed configuration. Ninety-five percent confidence intervals were estimated using percentile bootstrap resampling of paired cases. r_rb, matched-pairs rank-biserial correlation; p_Holm, Holm-adjusted Wilcoxon signed-rank p-value

Metric / Comparison	Median paired difference, mm [95%CI]	r_rb [95% CI]	p_Holm
Mean-cycle TRE — Friedman p<0.001			
Strategy 1 − Strategy 2	−0.564 [−1.190, −0.221]	−1.000 [−1.000, −1.000]	0.006
Strategy 1 − Strategy 3	+0.461 [+0.014, +1.091]	+0.758 [+0.273, +1.000]	0.049
Strategy 2 − Strategy 3	+1.454 [+0.235, +2.318]	+1.000 [+1.000, +1.000]	0.016
Strategy 2 − Strategy 4	+0.535 [+0.316, +1.149]	+0.970 [+0.818, +1.000]	0.010
Strategy 3 − Fixed Four-Reference	−0.99 [−1.08, −0.94]	−0.91 [−1.00, −0.55]	0.033
ES-frame TRE — Friedman p=0.006			
Strategy 1 − Strategy 3	+1.265 [+0.304, +2.445]	+0.909 [+0.636, +1.000]	0.029
Strategy 2 − Strategy 3	+2.515 [+0.0003, +2.858]	+0.956 [+0.714, +1.000]	0.031
Strategy 3 − Fixed Four-Reference	−2.000 [−2.055, −1.945]	−1.000 [−1.000, −1.000]	0.021
Strategy 3 − Two-Reference Control (K=2)	−1.265 [−2.445, −0.304]	−0.909 [−1.000, −0.576]	0.029
Final-frame TRE — Friedman p<0.001			
Strategy 1 − Strategy 2	−1.333 [−1.879, −0.738]	−1.000 [−1.000, −1.000]	0.006
Strategy 1 − Strategy 4	−0.600 [−0.846, −0.332]	−1.000 [−1.000, −1.000]	0.006
Strategy 2 − Strategy 3	+2.025 [+0.777, +3.918]	+1.000 [+1.000, +1.000]	0.012
Strategy 2 − Strategy 4	+0.733 [+0.406, +1.033]	+1.000 [+1.000, +1.000]	0.006
Strategy 3 − Strategy 4	−1.487 [−2.019, −0.371]	−0.818 [−1.000, −0.364]	0.027
Strategy 2 − Fixed Four-Reference	+2.087 [+0.805, +3.880]	+1.000 [+1.000, +1.000]	0.012
Strategy 4 − Fixed Four-Reference	+1.390 [+0.399, +2.038]	+0.848 [+0.455, +1.000]	0.027
Strategy 1 − Two-Reference Control (K=2)	−0.666 [−1.798, −0.119]	−0.848 [−1.000, −0.455]	0.009
Strategy 3 − Two-Reference Control (K=2)	−1.156 [−2.142, −0.483]	−1.000 [−1.000, −1.000]	0.029
Fixed Four-Reference Control − Two-Reference Control (K=2)	−1.228 [−2.104, −0.511]	−1.000 [−1.000, −1.000]	0.029

PROOF

Strategy 3 corresponds to the proposed manifold-guided multi-reference strategy; Strategy 4 corresponds to the partial BSF implementation based on Wiputra *et al.* [20]; the Fixed Four-Reference Control uses predetermined temporal references; and the K=2 control retains only R_1 and R_3. Because time-embedded PCA and Diffusion Maps yielded identical four-reference sets in every analyzed case, the Strategy 3 registration configuration was identical for the two representations. The Friedman test demonstrated significant overall differences among the six evaluated configurations for mean-cycle TRE ($p<0.001$), ES-frame TRE ($p=0.006$), and final-frame TRE ($p<0.001$). All pairwise p-values shown in Figure 4 are Holm-adjusted (p_Holm).

For mean-cycle TRE (Figure 4A), Strategy 3 showed the lowest mean error (6.8 ± 1.6 mm; median 7.0 [5.7–7.7] mm). Relative to Strategy 1, Strategy 3 reduced TRE by a median paired difference of 0.461 mm [95% CI: 0.014–1.091], with r_rb=0.758 [0.273–1.000] (p_Holm=0.049). Relative to sequential registration, the corresponding reduction was 1.454 mm [0.235–2.318], with complete directional consistency of the paired ranks (r_rb=1.000 [1.000–1.000]; p_Holm=0.016). Strategy 1 also showed lower mean-cycle TRE than sequential registration by 0.564 mm [0.221–1.190] (|r_rb|=1.000; p_Holm=0.006), while the partial BSF implementation showed a 0.535-mm lower median paired TRE than sequential registration [0.316–1.149] (|r_rb|=0.970 [0.818–1.000]; p_Holm=0.010). Importantly, Strategy 3 showed a median paired mean-cycle TRE reduction of 0.99 mm relative to the Fixed Four-Reference Control [0.94–1.08], with a strong directional effect (|r_rb|=0.91 [0.55–1.00]; p_Holm=0.033). The K=2 control yielded 7.3 ± 1.4 mm, with a median of 7.2 [6.4–8.0] mm, and none of its mean-cycle pairwise comparisons reached statistical significance. This non-significant result is not interpreted as evidence of equivalence in the present 12-case cohort.

For ES-frame TRE (Figure 4B), Strategy 3 again showed the lowest mean error (11.5 ± 2.7 mm). Relative to Strategy 1, Strategy 3 reduced TRE by a median paired difference of 1.265 mm [0.304–2.445], with r_rb=0.909 [0.636–1.000] (p_Holm=0.029). Relative to sequential registration, the median paired reduction was 2.515 mm [0.0003–2.858], with r_rb=0.956 [0.714–1.000] (p_Holm=0.031). Strategy 3 also showed a 2.000-mm lower median paired TRE than the Fixed Four-Reference Control [1.945–2.055], with complete directional consistency (|r_rb|=1.000; p_Holm=0.021). The K=2 control yielded 12.5 ± 2.8 mm, with a median of 12.3 [11.0–14.4] mm. Strategy 3 reduced ES-frame TRE relative to K=2 by a median of 1.265 mm [0.304–2.445], with |r_rb|=0.909 [0.576–1.000] (p_Holm=0.029). No statistically significant difference was detected between Strategy 1 and K=2, and no significant difference was detected between Strategy 3 and the partial BSF implementation. These non-significant comparisons are not interpreted as evidence of equivalence.

For final-frame TRE (Figure 4C), sequential frame-to-frame registration showed the highest mean error (6.1 ± 1.8 mm). Relative to Strategy 1, sequential registration increased final-frame TRE by a median paired difference of 1.333 mm [0.738–1.879] (|r_rb|=1.000; p_Holm=0.006). The corresponding excess error was 2.025 mm relative to Strategy 3 [0.777–3.918] (r_rb=1.000; p_Holm=0.012), 0.733 mm relative to the partial BSF implementation [0.406–1.033] (r_rb=1.000; p_Holm=0.006), and 2.087 mm relative to the Fixed Four-Reference Control [0.805–3.880] (r_rb=1.000; p_Holm=0.012). Strategy 1 also showed lower final-frame TRE than the partial BSF implementation by a median paired 0.600 mm [0.332–0.846], with complete directional consistency (|r_rb|=1.000; p_Holm=0.006). Strategy 3 showed a 1.487-mm lower median paired TRE than the partial BSF implementation [0.371–2.019] (|r_rb|=0.818 [0.364–1.000]; p_Holm=0.027), while the Fixed Four-Reference Control showed a 1.390-mm reduction relative to partial BSF [0.399–2.038] (|r_rb|=0.848 [0.455–1.000]; p_Holm=0.027). No statistically significant difference was detected between Strategy 3 and the Fixed Four-Reference Control at the final frame; this is interpreted as absence of evidence for a detectable difference in the present cohort rather than evidence of equivalence. The K=2 control yielded 5.1 ± 1.1 mm (median 5.5 [4.2–5.7] mm) and showed greater final-frame TRE than Strategy 1 by 0.666 mm [0.119–1.798] (|r_rb|=0.848 [0.455–1.000]; p_Holm=0.009), Strategy 3 by 1.156 mm [0.483–2.142] (|r_rb|=1.000; p_Holm=0.029), and the Fixed Four-Reference Control by 1.228 mm [0.511–2.104] (|r_rb|=1.000; p_Holm=0.029). No significant final-frame difference was detected between K=2 and either

PROOF

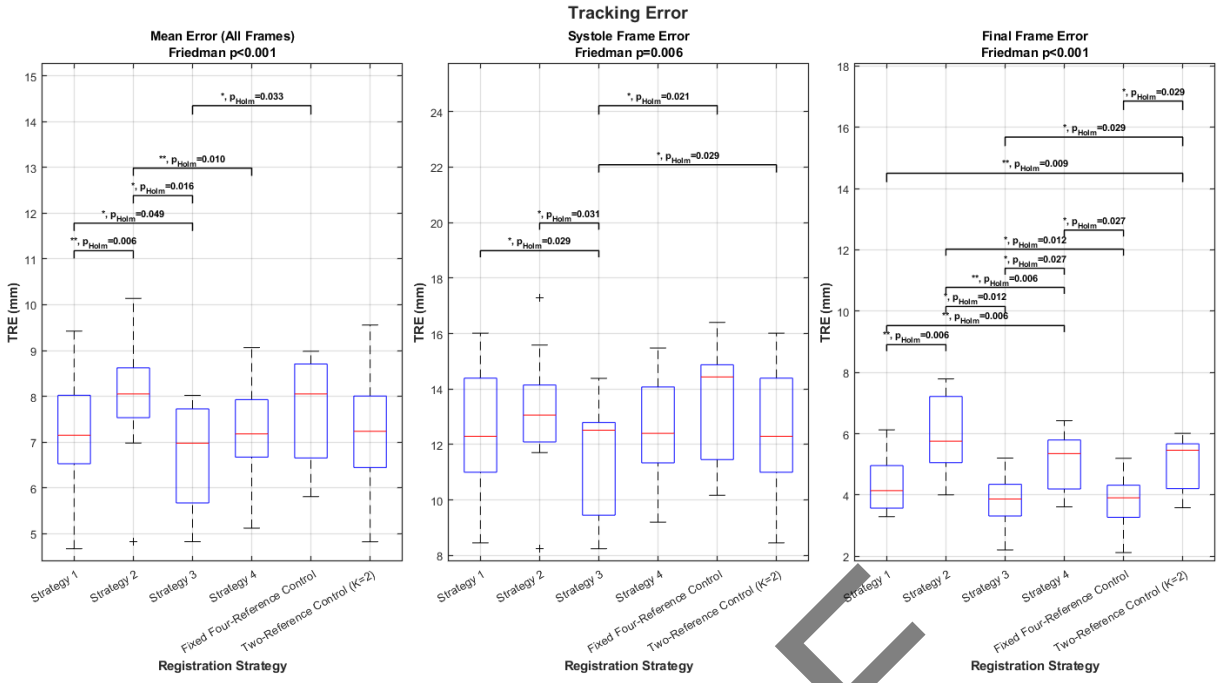


Figure 4. Comparison of Target Registration Error (TRE) among the five primary registration configurations and the Two-Reference Control (K=2) for (A) mean error across the complete cardiac cycle, (B) end-systolic (ES)-frame error, and (C) final-frame error. TRE values are expressed in millimeters. Strategy 1: single-reference registration to the initial end-diastolic frame; Strategy 2: sequential frame-to-frame registration; Strategy 3: the proposed manifold-guided multi-reference strategy; Strategy 4: partial BSF implementation; Fixed Four-Reference Control: four predetermined temporal reference locations using the same multi-reference architecture as Strategy 3; Two-Reference Control (K=2): the initial ED and principal internal references retained, with the intermediate contraction and relaxation references removed. Overall differences were assessed using the Friedman test, followed by paired Wilcoxon signed-rank comparisons with Holm correction. All pairwise p-values displayed in the figure are Holm-adjusted values (p_Holm). Statistical significance was defined as p_Holm<0.05. *p_Holm<0.05; **p_Holm<0.01

PROOF

sequential registration or the partial BSF implementation.

4. Discussion

The low-dimensional representation analysis showed that 3D echocardiographic sequences can retain temporal and physiologically related organization relevant to reference-frame selection. Time-embedded PCA and Diffusion Maps produced different qualitative trajectory geometries: Diffusion Maps generally formed a more loop-like representation of the cardiac cycle, whereas PCA produced a curved outbound-and-return trajectory with distinguishable contraction and relaxation branches. Despite these qualitative differences, the two methods produced identical sets of all four reference frames in every analyzed case. For both methods, the principal internal reference coincided with the expert-defined ES frame in 10 of 12 cases, with a one-frame discrepancy in the remaining two cases. The ES-based validation therefore did not provide evidence for superiority of either time-embedded PCA or Diffusion Maps in terms of reference-frame selection.

The identical discrete reference selections clarify the role of the low-dimensional mapping in the present framework. Although PCA and Diffusion Maps produced visibly different latent-space geometries, these differences did not translate into different reference-frame indices. Thus, within the present benchmark, no task-level reference-selection advantage was attributable to the more geometrically complex Diffusion Maps representation over time-embedded PCA. Several factors may contribute to this result. First, the present dataset consists exclusively of healthy volunteers with relatively regular ventricular motion, and the dominant cardiac variation relevant to reference placement may be represented adequately by more than one low-dimensional mapping. Second, PCA was applied to a time-embedded input in which immediate temporal evolution had already been incorporated before dimensionality reduction. Third,

the downstream objective was identification of discrete reference states using a deterministic distance-based criterion rather than faithful reconstruction of every aspect of latent geometry. Consequently, different low-dimensional geometries can produce the same reference-frame extrema. The experimentally consequential component of the present framework is therefore the organization and placement of registration references rather than the use of a particular nonlinear dimensionality-reduction method.

The execution-time analysis further demonstrated a practical computational difference. Time-embedded PCA required 0.460 ± 0.144 s per execution, compared with 1.091 ± 0.461 s for time-embedded Diffusion Maps, corresponding to approximately a 2.37-fold greater execution time for Diffusion Maps. Because construction of the time-delay representation itself was not included in the timing analysis, the longer execution times of the time-embedded variants are consistent with processing inputs having approximately twice the feature dimensionality. For the short single-cycle sequences evaluated here, the absolute computational difference remains modest. However, computational efficiency may become more consequential when processing multiple consecutive cardiac cycles, longer acquisitions, or implementing the framework for online or near-real-time processing on resource-constrained hardware. Given the identical downstream reference configurations in all 12 cases, time-embedded PCA represents the more parsimonious implementation for the present benchmark. This conclusion is restricted to the present reference-selection task and dataset and does not exclude potential advantages of other representations in more heterogeneous or pathological cardiac motion.

The reference-frame analysis provides an important intermediate physiological validation of the proposed framework. The principal internal reference was not explicitly defined as ES by the selection algorithm; rather, it was selected solely according to its geometric separation from the initial and terminal ED states in the low-dimensional trajectory. Nevertheless, for both time-embedded PCA and Diffusion Maps, this automatically selected reference coincided with the expert-defined ES frame in 10 of the 12 analyzed cases and differed by only one frame in the remaining two cases. Moreover, the complete four-reference sets obtained from the two representations were identical in every case. This finding suggests that, within the present healthy-volunteer benchmark, the low-dimensional trajectory consistently positioned a physiologically distinct state near ES at a geometrically prominent location. Importantly, this result should not be interpreted as demonstrating a general-purpose ES-detection model. Expert ES annotations were used after reference extraction as one criterion for physiological construct validation of the candidate representations and therefore are not statistically independent of the representation-assessment stage.

The quantitative representation metrics also showed that no single dimensionality-reduction method was uniformly optimal according to every criterion. Isomap performed particularly well in neighborhood-preservation measures, whereas PCA showed favorable preservation of global distance structure. The relatively high association between most representations and the landmark-derived surrogate-volume profile further indicates that physiologically related cardiac-cycle variation was represented in the low-dimensional spaces, although t-SNE generally showed weaker correspondence. Time embedding introduced immediate temporal context into methods that otherwise operate without explicit knowledge of frame order and was intended to help distinguish morphologically similar states occurring on the contraction and relaxation branches. In addition, transforming a high-dimensional volumetric cardiac cycle into a two-dimensional trajectory provides a directly visualizable representation for reference-frame selection. If future datasets containing multiple consecutive 3D cardiac cycles become available, analysis of trajectory reproducibility across cycles could provide an additional basis for assessing the stability of automatically extracted cardiac states.

Comparison of the registration configurations showed that temporal organization and reference placement were associated with differences in landmark-based TRE. The statistical comparisons were interpreted using paired effect magnitude and uncertainty in addition to Holm-adjusted statistical significance. This distinction is particularly important in the present 12-case cohort because a p-value alone does not characterize either the physical magnitude of a TRE difference or the consistency of its direction

PROOF

across cases. Median paired differences provide effect estimates in millimeters, matched-pairs rank-biserial correlations describe directional consistency, and percentile-bootstrap confidence intervals indicate the uncertainty of these estimates. Conversely, non-significant comparisons are not interpreted as evidence of equivalence.

The three TRE outcomes used in the registration comparison characterize complementary aspects of tracking performance rather than redundant measurements of the same behavior. Mean-cycle TRE reflects the overall accuracy of the estimated motion trajectory across the complete cardiac cycle and therefore provides the broadest measure of tracking performance. ES-frame TRE evaluates tracking at a physiologically important state that is substantially displaced from the initial ED configuration and therefore represents a demanding point of the cardiac cycle for methods that rely on an initial reference. Final-frame TRE, in contrast, evaluates the accuracy with which the tracking trajectory returns to the terminal ED state at completion of the cardiac cycle. This metric is particularly informative for strategies involving successive displacement-field compositions, because errors introduced throughout the cycle can influence the final estimated position. The three outcomes should therefore be interpreted jointly: mean-cycle TRE describes overall performance, ES-frame TRE emphasizes performance around a highly deformed cardiac state, and final-frame TRE reflects end-of-cycle tracking behavior after propagation through the temporal registration architecture.

For mean-cycle tracking, Strategy 3 achieved the lowest mean-cycle TRE (6.8 ± 1.6 mm). Its advantage over the conventional baselines was supported by the magnitude and direction of the paired effects rather than statistical significance alone. Relative to Strategy 1, Strategy 3 reduced TRE by a median of 0.461 mm [95% CI: 0.014–1.091], with r_rb=0.758, whereas the reduction relative to sequential registration was 1.454 mm [0.235–2.318], with r_rb=1.000. The comparison with the Fixed Four-Reference Control is particularly informative because the two configurations used the same number of reference frames, the same pairwise registration algorithm, and the same displacement-field composition architecture; their principal difference was the temporal placement of the references. Strategy 3 showed a median paired mean-cycle TRE reduction of 0.99 mm relative to the Fixed Four-Reference Control [0.94–1.08], with a strong rank-biserial effect (|r_rb|=0.91). Thus, within this benchmark, the observed difference is supported by both effect magnitude and directional consistency and supports a contribution of data-driven reference placement beyond the use of four references alone. No statistically significant difference was detected between Strategy 3 and the partial BSF implementation for mean-cycle TRE; this absence of significance is not interpreted as evidence of equivalent performance.

The ES-frame analysis provides particularly clear evidence of a phase-specific reference-placement effect at a cardiac state characterized by substantial structural change relative to ED. Strategy 3 reduced TRE relative to Strategy 1 by a median 1.265 mm [0.304–2.445] and relative to sequential registration by 2.515 mm [0.0003–2.858], with rank-biserial correlations of 0.909 and 0.956, respectively. Relative to the Fixed Four-Reference Control, Strategy 3 reduced ES-frame TRE by 2.000 mm [1.945–2.055], with complete directional consistency of the paired ranks. Thus, the ES findings are supported not merely by Holm-adjusted significance tests but by measurable differences in physical TRE and highly consistent paired effects. The larger ES difference relative to the Fixed Four-Reference Control indicates that predetermined temporal spacing did not reproduce the performance obtained when the reference locations were derived from the intrinsic cardiac trajectory.

The final-frame results showed a different pattern and provide additional information about phase dependence of the registration architectures. Sequential frame-to-frame registration exhibited the largest final-frame TRE (6.1 ± 1.8 mm) and showed higher error than Strategy 1 by a median of 1.333 mm, Strategy 3 by 2.025 mm, the partial BSF implementation by 0.733 mm, and the Fixed Four-Reference Control by 2.087 mm. This pattern is consistent with the susceptibility of a long sequential transformation chain to propagation of registration inaccuracies toward the end of the cycle. The favorable final-frame behavior of Strategy 1 is also informative: Strategy 1 showed a median paired advantage of 0.600 mm over the partial BSF implementation, with complete directional

PROOF

consistency. Direct registration to the initial ED avoids a long composition chain and, at cycle completion, can additionally benefit from the renewed structural similarity between the terminal and initial ED states. Strategy 3 and the Fixed Four-Reference Control both retained low final-frame TRE, and no statistically significant difference was detected between them. This result is interpreted as absence of evidence for a detectable difference in the present cohort rather than evidence that the two configurations are equivalent.

Taken together, the three outcomes show that the relative behavior of the registration architectures is phase dependent rather than uniformly ordered across the cardiac cycle. Strategy 3 combined the lowest mean-cycle and ES-frame TRE with low final-frame TRE, while the Fixed Four-Reference Control reproduced the favorable final-frame behavior but not the mean-cycle or ES performance. The effect estimates relative to the fixed control were 0.99 mm for mean-cycle TRE and 2.000 mm for ES-frame TRE in favor of Strategy 3. These comparisons strengthen the interpretation that, within the present benchmark, using intrinsic image-derived trajectory information to determine where references are placed can influence tracking performance beyond simply increasing the number of references. At the same time, the absence of significant differences from the partial BSF implementation for mean-cycle and ES-frame TRE means that the proposed framework should not be interpreted as uniformly superior to all hybrid approaches.

The reference-selection and registration results are also mutually informative. Figure 3 shows that the low-dimensional trajectory identified reference configurations associated with physiologically distinct positions along the cardiac cycle, including a principal internal reference closely corresponding to expert-defined ES. Figure 4 shows that changing the temporal reference architecture can materially alter tracking performance. This point is reinforced by the K=2 sensitivity experiment: removing the two branch-specific intermediate references produced measurable phase-specific TRE changes, whereas replacing time-embedded PCA with Diffusion Maps did not alter any of the four selected reference frames. The combined findings therefore support the role of the low-dimensional trajectory primarily as a mechanism for data-driven organization of the temporal reference structure rather than as a motion model incorporated into the registration algorithm itself.

Comparison with Strategy 4 should be interpreted specifically as a comparison with the partial BSF implementation evaluated in this study. This implementation combined the Eulerian and Lagrangian displacement estimates using temporal weighting and represented the resulting trajectory using the B-spline–Fourier formulation [20], but it did not include the final consistency-correction stage of the complete BSF pipeline. No statistically significant difference was detected between Strategy 3 and the partial BSF implementation for mean-cycle or ES-frame TRE, whereas Strategy 3 showed a median 1.487-mm final-frame advantage [0.371–2.019]. Strategy 1 and the Fixed Four-Reference Control also showed lower final-frame TRE than the partial BSF implementation. These findings should not be interpreted as establishing superiority over the complete BSF framework. The proposed approach differs conceptually in that the low-dimensional representation is used to determine temporal reference placement before registration rather than as a motion prior or post-registration regularization model.

The K=2 sensitivity analysis provides phase-specific evidence regarding the contribution of the intermediate references. Removing R_2 and R_4 did not produce a statistically significant mean-cycle difference relative to any of the evaluated configurations; given the small sample size, this result should not be interpreted as evidence of equivalence or as proof that reference number is unimportant. At ES, Strategy 3 reduced TRE relative to K=2 by a median 1.265 mm [0.304–2.445], with $|r_{rb}|=0.909$, supporting a contribution of the intermediate contraction reference R_2. At the final frame, Strategy 3 reduced TRE relative to K=2 by 1.156 mm [0.483–2.142], and the Fixed Four-Reference Control showed a 1.228-mm advantage over K=2 [0.511–2.104], with complete directional consistency in both comparisons. These estimates support a useful role for intermediate reference support at specific portions of the cardiac cycle, particularly around ES and during return toward terminal ED.

The present sensitivity analysis should not be interpreted as identifying K=4 as the optimal reference number. The K=2 experiment was selected as a nested and physiologically symmetric reduction of the

PROOF

proposed two-branch architecture because it removes the contraction- and relaxation-branch intermediate references together. A K=3 configuration was not evaluated because it is not a structurally symmetric ablation of this specific architecture. Once R_1 and R_3 are retained, a third internal reference must preferentially represent either the contraction or relaxation branch or be selected using a newly defined rule, thereby changing both reference number and reference-selection structure. Formal optimization of reference number would therefore require a generalized K-reference formulation with prespecified selection rules, matched fixed-K controls, and preferably independent or nested validation. The present results support phase-specific contributions of the intermediate references but do not establish universal superiority or an empirically optimal value of K.

The omission of the final consistency-correction stage was deliberate rather than an attempt to reproduce the complete BSF pipeline. The purpose of the present comparison was to evaluate the core temporal organization of the competing registration strategies under a common framework. Because the consistency-correction stage is a post-processing refinement applied after estimation of the primary motion trajectory, an analogous consistency-refinement procedure could in principle also be applied to displacement fields generated by the other registration strategies. Applying this additional refinement exclusively to Strategy 4 would therefore confound the comparison between the underlying registration architectures with the additional effect of post-processing. The present BSF results should consequently be interpreted as those of the temporal weighting and B-spline–Fourier components without the final consistency-refinement stage, and not as the performance of the complete BSF method. A future study could separately evaluate the incremental effect of applying an equivalent consistency-refinement procedure across the different registration strategies.

Several limitations should be considered when interpreting these findings. First, the present evaluation was performed on a relatively small cohort of 12 healthy-volunteer 3D echocardiographic sequences. The benchmark used in this study is a well-established reference dataset that was developed through an open challenge and has been used for comparative evaluation of multiple cardiac motion-tracking algorithms [28]. Its standardized imaging data and landmark-based reference information therefore provide a valuable basis for controlled methodological comparison. Nevertheless, the cardiac motion patterns represented in the present analysis primarily reflect normal ventricular contraction and relaxation. Pathological ventricular motion, including regional hypokinesia, akinesia, dyskinesia, or altered ventricular geometry, may produce different low-dimensional trajectories and may affect both reference-frame selection and registration performance. The effect-size estimates and their confidence intervals also emphasize the uncertainty associated with the small sample size. Several comparisons showed large and highly consistent rank-biserial effects, but the precision of the estimated physical TRE differences varied across comparisons. Statistical significance and strong directional effects in this 12-case cohort should therefore not be interpreted as precise population-level estimates, and non-significant comparisons should not be interpreted as evidence of equivalence. Validation in larger and more heterogeneous datasets, particularly cohorts including pathological LV motion, is required to establish the precision, robustness, and clinical generalizability of the observed effects.

A further limitation concerns the availability of complete temporal data. Although the original benchmark included 15 healthy volunteers, three benchmark cases (Cases 7, 8, and 9) were temporally incomplete in the version of the dataset available to us and could not provide the complete ED-to-ED cardiac cycle required by the analysis pipeline. These cases were identified during the initial data-completeness assessment before any dimensionality-reduction, registration, or TRE analysis and were therefore not excluded according to study outcome or algorithmic performance. Nevertheless, because the analysis was necessarily restricted to the 12 cases with complete available temporal data, a potential complete-case or availability-related selection bias cannot be excluded. The findings should therefore be interpreted specifically for the analyzable 12-case subset.

An additional limitation concerns image-intensity variability and the absence of additional intensity normalization or myocardial masking. All low-dimensional representation and registration operations

PROOF

in the present study were performed within each subject's reconstructed 3D echocardiographic sequence rather than between separate examinations or different subjects. Consequently, variability specifically associated with comparing independent acquisitions—such as between-study differences in probe placement, scanner setup, global gain choices, or spatial sampling—was not introduced into the within-case reference-selection distances. Nevertheless, acquisition within a common sequence does not eliminate intrinsic frame-to-frame ultrasound variability. Motion-dependent speckle decorrelation, acoustic shadowing, attenuation, reverberation, and other image-appearance changes may remain across the cardiac cycle and may affect both the high-dimensional image distances used for low-dimensional representation construction and intensity-based Demons registration [7,8,12,15–17]. We did not perform a dedicated analysis quantifying the effect of such frame-to-frame intensity fluctuations on reference selection and therefore cannot exclude their contribution. The temporally organized trajectories, identical four-reference selections obtained with time-embedded PCA and Diffusion Maps, and close correspondence of the principal internal reference with expert-defined ES provide indirect evidence that reference selection was not grossly destabilized by these variations; however, these observations should not be interpreted as a dedicated robustness analysis. Future work should explicitly evaluate the effect of intensity normalization, myocardial masking, and other ultrasound-specific preprocessing procedures on low-dimensional trajectory organization and registration performance.

Another limitation is that the candidate low-dimensional representations and the final registration experiment were examined within the same 12-case cohort, and this should be distinguished from independent external validation. However, the dependence differs from conventional outcome-driven model selection. The dimensionality-reduction algorithms were applied independently within each sequence rather than trained across subjects, registration TRE was not used to select a representation, and neither the deterministic reference-selection criterion nor the Demons registration parameters were fitted to the expert annotations or final tracking outcomes. Moreover, time-embedded PCA and Diffusion Maps produced identical complete four-reference configurations in every analyzed case, making the final Strategy 3 registration and TRE invariant to the choice between these two representations. The expert-defined ES annotation was used after reference extraction as a physiological construct-validation measure and contributed to assessment of the candidate representations. We therefore acknowledge a residual within-sample representation-assessment dependence, particularly for interpretation of ES concordance, rather than claiming statistically independent validation. Independent external validation in a larger dataset, preferably including pathological ventricular motion and greater variability in image characteristics, remains necessary before claims regarding robustness or generalizability can be made.

A further limitation concerns uncertainty in the benchmark reference trajectories themselves. The ground-truth landmarks were manually tracked by human observers in 3D tagged MRI and were subsequently transformed into the 3D ultrasound coordinate system using a point-based similarity transformation [28]. Consequently, the reference landmark positions are subject both to observer-dependent localization variability and to residual uncertainty introduced by the cross-modality transformation. The measured TRE therefore reflects not only error associated with the evaluated registration strategy but also uncertainty in the available reference standard. Because the same benchmark landmarks were used to evaluate all registration configurations, this source of uncertainty is common to the comparative analysis and does not selectively favor one strategy; nevertheless, it limits interpretation of the absolute TRE magnitude. Future validation using complementary ground-truth approaches or datasets with independently characterized reference uncertainty would further strengthen evaluation of the proposed framework.

5. Conclusion

Within the established 3D echocardiographic benchmark evaluated in this study, comprising 12 healthy-volunteer sequences, the proposed manifold-guided multi-reference strategy demonstrated favorable landmark-based tracking performance across complementary phases of the cardiac cycle.

PROOF

Time-embedded PCA and Diffusion Maps produced identical four-reference configurations in all analyzed cases, and their principal internal reference coincided with the expert-defined ES frame in 10 of 12 cases, with a one-frame discrepancy in the remaining two cases. The paired effect estimates complemented the significance tests by demonstrating measurable and directionally consistent TRE differences for several key comparisons. Strategy 3 showed a median mean-cycle TRE reduction of 0.99 mm relative to the Fixed Four-Reference Control and a 2.000-mm reduction at ES, supporting a contribution of data-driven reference placement beyond use of multiple references alone for these outcomes. At the final frame, no statistically significant difference was detected between Strategy 3 and the Fixed Four-Reference Control, and this result is not interpreted as evidence of equivalence. The K=2 sensitivity analysis showed phase-specific effects of removing the intermediate references: Strategy 3 reduced ES-frame TRE relative to K=2 by 1.265 mm and final-frame TRE by 1.156 mm. These findings support a functional contribution of the intermediate contraction and relaxation references within the evaluated architecture while not establishing K=4 as universally superior or empirically optimal. No statistically significant difference between Strategy 3 and the partial BSF implementation was detected for mean-cycle or ES-frame TRE, and the comparison should not be interpreted as an evaluation of the complete BSF framework because its final consistency-correction stage was not implemented. Overall, the results support the internal technical feasibility of using the intrinsic low-dimensional organization of the cardiac cycle for data-driven reference-frame placement rather than demonstrating superiority of a particular dimensionality-reduction algorithm. The findings should not be interpreted as an unbiased estimate of performance on unseen data. Independent validation in larger and clinically heterogeneous datasets, particularly those containing pathological LV motion, is required before broader generalization can be established.

PROOF

References

1. Roth GA, Mensah GA, Johnson CO, et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 Study. J Am Coll Cardiol. 2020;76:2982–3021.

2. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr. 2015;28(1):1–39.e14.

3. Pellikka PA, Arruda-Olson A, Chaudhry FA, et al. Guidelines for performance, interpretation, and application of stress echocardiography in ischemic heart disease. J Am Soc Echocardiogr. 2020;33(1):1–41.e8.

4. Lang RM, Badano LP, Tsang W, et al. EAE/ASE recommendations for image acquisition and display using three-dimensional echocardiography. J Am Soc Echocardiogr. 2012;25(1):3–46.

5. Cerqueira MD, Weissman NJ, Dilsizian V, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. Circulation. 2002;105(4):539–542.

6. Ciampi Q, Picano E, Paterni M, et al. Quality control of regional wall motion analysis in stress Echo 2020. Int J Cardiol. 2017;249:479–485.

7. Perperidis A. Postprocessing approaches for the improvement of cardiac ultrasound B-mode images: a review. IEEE Trans Ultrason Ferroelectr Freq Control. 2016;63(3):470–485.

8. American Society of Echocardiography. Recommendations for the Identification and Mitigation of Cardiac Ultrasound Artifacts. J Am Soc Echocardiogr. 2026.

9. Voigt JU, Pedrizzetti G, Lysyansky P, et al. Definitions for a common standard for 2D speckle tracking echocardiography. J Am Soc Echocardiogr. 2015;28(2):183–193.

10. Smiseth OA, Torp H, Opdahl A, Haugaa KH, Urheim S. Myocardial strain imaging: how useful is it in clinical decision making? Eur Heart J. 2016;37(15):1196–1207.

11. Dandel M, Lehmkuhl H, Knosalla C, et al. Strain and strain rate imaging by echocardiography. Int J Cardiol. 2009;132(1):11–24.

12. Crosby J, Amundsen BH, Hergum T, et al. 3-D speckle tracking for assessment of regional left ventricular function. Ultrasound Med Biol. 2009;35(3):458–471.

13. Blessberger H, Binder T. Two-dimensional speckle tracking echocardiography: basic principles. Heart. 2010;96(9):716–722.

14. Seo Y, Ishizu T, Atsumi A, et al. Three-dimensional speckle tracking echocardiography. Circ J. 2014;78(6):1290–1301.

15. Ledesma-Carbayo MJ, Kybic J, Desco M, et al. Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. IEEE Trans Med Imaging. 2005;24(9):1113–1126.

16. De Craene M, Piella G, Camara O, et al. Temporal diffeomorphic free-form deformation: application to motion and strain estimation from 3D echocardiography. Med Image Anal. 2012;16(2):427–450.

17. Zhang Z, Ashraf M, Sahn DJ, Song X. Temporally diffeomorphic cardiac motion estimation from three-dimensional echocardiography by minimization of intensity consistency error. Med Phys. 2014;41(5):052902.

18. Sundar H, Litt H, Shen D. Estimating myocardial motion by 4D image warping. Pattern Recognit. 2009;42(11):2514–2526.

19. Tascón-Vidarte JD, et al. Accuracy and consistency of intensity-based deformable image registration in 4DCT. PLoS One. 2022;17:e0271064.

20. Wiputra H, Chan WX, Foo YY, et al. Cardiac motion estimation from medical images: a regularisation framework applied to pairwise image registration displacement fields. Sci Rep. 2020;10:18510. doi:10.1038/s41598-020-75525-4.

21. Yang L, Georgescu B, Zheng Y, Meer P, Comaniciu D. 3D ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. 2008. doi:10.1109/CVPR.2008.4587518.

22. Rohé MM, Sermesant M, Pennec X. Low-dimensional representation of cardiac motion using Barycentric Subspaces: a new group-wise paradigm for estimation, analysis, and reconstruction. Med Image Anal. 2018;45:1–12. doi:10.1016/j.media.2017.12.008.

23. Wachinger C, Yigitsoy M, Rijkhorst EJ, Navab N. Manifold learning for image-based breathing gating in ultrasound and MRI. Med Image Anal. 2012;16(4):806–818.

24. Gifani P, Behnam H, Shalbaf A, Alizadeh Sani Z. Automatic detection of end-diastole and end-systole from echocardiography images using manifold learning. Physiol Meas. 2010;31(9):1091–1103.

25. Izenman AJ. Introduction to manifold learning. WIREs Comput Stat. 2012;4(5):439–446. doi:10.1002/wics.1222.

26. Lee JA, Verleysen M. Nonlinear Dimensionality Reduction. New York: Springer; 2007.

27. van der Maaten LJP, Postma EO, van den Herik HJ. Dimensionality Reduction: A Comparative Review. Tilburg University Technical Report TiCC-TR 2009-005; 2009.

28. Tobon-Gomez C, De Craene M, McLeod K, et al. Benchmarking framework for myocardial tracking and deformation algorithms: an open access database. Med Image Anal. 2013;17(6):632–648. doi:10.1016/j.media.2013.03.008.

29. Takens F. Detecting strange attractors in turbulence. In: Rand D, Young LS, editors. Dynamical Systems and Turbulence, Warwick 1980. Lecture Notes in Mathematics. Vol 898. Berlin: Springer; 1981. p. 366–381.

30. Jolliffe IT. Principal Component Analysis. 2nd ed. New York: Springer; 2002.

31. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579–2605.

32. Coifman RR, Lafon S. Diffusion maps. Appl Comput Harmon Anal. 2006;21(1):5–30.

33. Nadler B, Lafon S, Coifman RR, Kevrekidis IG. Diffusion maps, spectral clustering and reaction coordinates of dynamical systems. Appl Comput Harmon Anal. 2006;21(1):113–127.

34. Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science. 2000;290(5500):2323–2326.

35. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426. 2018.

36. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science. 2000;290(5500):2319–2323.

37. Venna J, Kaski S. Neighborhood preservation in nonlinear projection methods: an experimental study. In: Artificial Neural Networks — ICANN 2001. Berlin: Springer; 2001. p. 485–491.

38. Lee JA, Verleysen M. Quality assessment of dimensionality reduction: rank-based criteria. Neurocomputing. 2009;72(7–9):1431–1443.

39. Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 1964;29:1–27.

40. Thirion JP. Image matching as a diffusion process: an analogy with Maxwell's demons. Med Image Anal. 1998;2(3):243–260. doi:10.1016/S1361-8415(98)80022-4.

41. Vercauteren T, Pennec X, Perchant A, Ayache N. Diffeomorphic demons: efficient non-parametric image registration. NeuroImage. 2009;45(1 Suppl): S61–S72. doi:10.1016/j.neuroimage.2008.10.040.

PROOF