

Depth-resolved registration of transesophageal echo to x-ray fluoroscopy using an inverse geometry fluoroscopy system

Charles R. Hatt

Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53705

Michael T. Tomkowiak, David A. P. Dunkerley, and Jordan M. Slagowski

Department of Medical Physics, University of Wisconsin-Madison, Madison, Wisconsin 53705

Tobias Funk

Triple Ring Technologies, Inc., Newark, California 94560

Amish N. Raval

Department of Medicine, University of Wisconsin-Madison, Madison, Wisconsin 53792

Michael A. Speidel^{a)}

Departments of Medical Physics and Medicine, University of Wisconsin-Madison, Madison, Wisconsin 53705

(Received 26 July 2015; revised 28 September 2015; accepted for publication 29 October 2015; published 13 November 2015)

Purpose: Image registration between standard x-ray fluoroscopy and transesophageal echocardiography (TEE) has recently been proposed. Scanning-beam digital x-ray (SBDX) is an inverse geometry fluoroscopy system designed for cardiac procedures. This study presents a method for 3D registration of SBDX and TEE images based on the tomosynthesis and 3D tracking capabilities of SBDX.

Methods: The registration algorithm utilizes the stack of tomosynthetic planes produced by the SBDX system to estimate the physical 3D coordinates of salient key-points on the TEE probe. The key-points are used to arrive at an initial estimate of the probe pose, which is then refined using a 2D/3D registration method adapted for inverse geometry fluoroscopy. A phantom study was conducted to evaluate probe pose estimation accuracy relative to the ground truth, as defined by a set of coregistered fiducial markers. This experiment was conducted with varying probe poses and levels of signal difference-to-noise ratio (SDNR). Additional phantom and *in vivo* studies were performed to evaluate the correspondence of catheter tip positions in TEE and x-ray images following registration of the two modalities.

Results: Target registration error (TRE) was used to characterize both pose estimation and registration accuracy. In the study of pose estimation accuracy, successful pose estimates (3D TRE < 5.0 mm) were obtained in 97% of cases when the SDNR was 5.9 or higher in seven out of eight poses. Under these conditions, 3D TRE was 2.32 ± 1.88 mm, and 2D (projection) TRE was 1.61 ± 1.36 mm. Probe localization error along the source-detector axis was 0.87 ± 1.31 mm. For the *in vivo* experiments, mean 3D TRE ranged from 2.6 to 4.6 mm and mean 2D TRE ranged from 1.1 to 1.6 mm. Anatomy extracted from the echo images appeared well aligned when projected onto the SBDX images.

Conclusions: Full 6 DOF image registration between SBDX and TEE is feasible and accurate to within 5 mm. Future studies will focus on real-time implementation and application-specific analysis.

© 2015 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4935534>]

Key words: inverse geometry, scanning-beam digital x-ray, image fusion, fluoroscopy, echocardiography

1. INTRODUCTION

Catheter-based cardiac interventions allow for minimally invasive treatment of structural heart disease, reducing patient trauma and opening up treatment options for patients that are too sick and/or fragile to undergo surgery. While surgeons have the luxury of direct visualization of the treatment site, this comes at the cost of increased risk to the patient, greater morbidity, and longer recovery time. In contrast, interventional cardiologists employ imaging methods such as x-ray fluoroscopy (XRF) and echocardiography (echo) to visualize their devices, identify the anatomy they wish to treat and to avoid, and to monitor the success of the therapy. Integration of these

imaging methods is desirable for optimal clinical workflow and improved therapeutic success.

Image fusion between XRF and transesophageal echocardiography (TEE) has recently been proposed¹⁻⁴ and clinically implemented (EchoNavigator, Philips Healthcare). Many structural heart interventions, such as transcatheter aortic valve replacement (TAVR), left atrial appendage closure, and the mitral clip procedure, utilize both XRF and TEE. These procedures may benefit from the enhanced guidance offered by combining information from both image modalities. For example, in TAVR, a prosthetic valve is guided and deployed using XRF, but visualization of the anatomy is poor. If XRF/TEE fusion is enabled, real-time anatomical information

from echo can be visualized continuously in the context of the devices without the need for nephrotoxic x-ray contrast.

XRF/TEE fusion is accomplished using 2D/3D registration techniques¹⁻³ or magnetic tracking sensors.^{5,6} Sensor-based methods require additional hardware and may be inaccurate due to electromagnetic field distortions in the catheterization lab.⁷ Generally, 2D/3D registration techniques will estimate the 3D location and orientation (pose) of the TEE probe by comparing the clinical XRF image to simulated XRF images of a 3D probe model (digitally reconstructed radiographs or DRRs). The model pose is iteratively adjusted until the similarity between the clinical XRF and DRR is maximized. After inferring the 3D position of the probe in the C-arm coordinate system, the 3D TEE image data can be registered to XRF data.

Using a typical monoplanar XRF C-arm system, the most challenging pose parameters to estimate are the so-called “out-of-plane” parameters, which include Euler angle rotations about the detector axes (pitch and roll) and, in particular, translations along the source-detector axis. This is because varying these parameters typically causes only subtle changes in the device appearance, which in turn do not strongly influence the similarity function maximized during pose estimation. Performing the registration using two x-ray views can help resolve this issue but the increased radiation dose to the patient is a concern.

Scanning-beam digital x-ray (SBDX) is an inverse geometry x-ray fluoroscopy technology designed for dose reduction and tomosynthesis-based 3D device tracking.⁸ The basic components of SBDX are a scanning x-ray tube, multihole collimator, high speed photon-counting detector, and a real-time reconstructor (Fig. 1). As the electron beam in the x-ray tube scans over an array of focal spot positions, small-field-of-view images of the patient are captured. After each frame, the detector data are reconstructed into a stack of full-field-of-

view tomosynthesis images (32 planes \times 15 frames/s).⁹ The tomosynthesis images are a necessary precursor to the final live image display, termed the composite image. However, the plane stacks can also be exploited for frame-by-frame 3D localization of high-contrast devices. This principle has been previously applied to localize catheter tips and electrodes,¹⁰ fiducials,¹¹ and coronary artery centerlines.¹²

In this paper, we present the first investigation of SBDX/TEE image registration. The tomosynthesis capability of SBDX is used to obtain the position of the TEE probe along the source-detector axis and an accurate initial estimate of the 3D probe pose. This is followed by refinement of the pose estimate using a 2D/3D registration procedure. A phantom study of 3D and 2D target registration error (TRE) was conducted for a variety of probe orientations and image noise levels in order to quantify the performance of the new pose estimation algorithm. To demonstrate SBDX/TEE image fusion in 3D and 2D visualizations, additional phantom and *in vivo* studies were conducted. In the 3D visualization, 3D TEE data are registered and fused with 3D catheter tip positions localized from SBDX tomosynthesis imaging.¹⁰

2. ALGORITHM

SBDX/TEE registration is achieved by estimating the 3D pose of the TEE probe based on its appearance in SBDX x-ray images. The pose estimation algorithm has two stages (Secs. 2.B and 2.C). First, an initial estimate of the probe pose is obtained by performing tomosynthesis-based 3D localization of key-points on the probe. Second, the initial pose is refined with a 2D/3D registration algorithm adapted for SBDX’s inverse geometry. Given the 3D pose and a calibration relating the echo image volume to the TEE probe, the echo

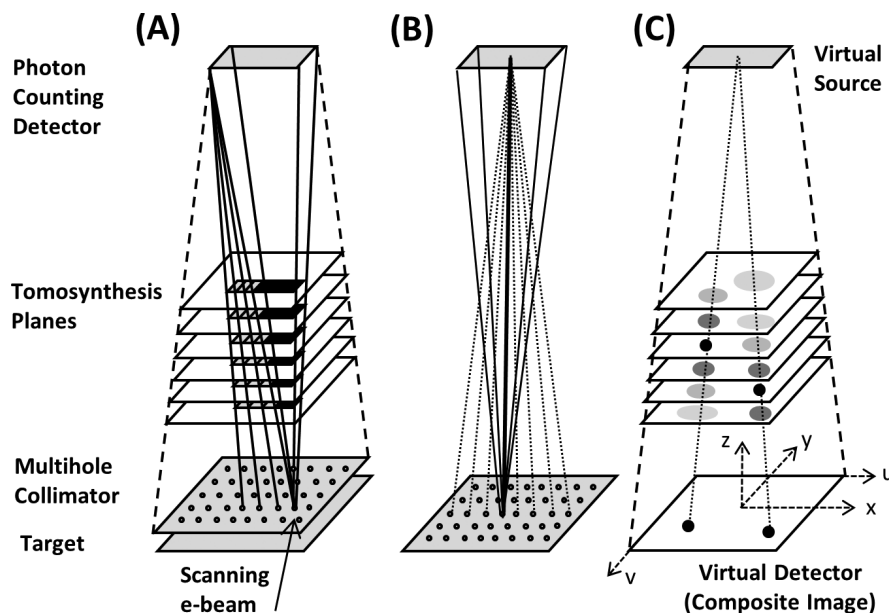


FIG. 1. (A) SBDX imaging geometry, demonstrating shift-and-add tomosynthesis at multiple planes. (B) The central rays of the individual beamlets form a cone of rays originating from the center of the detector. The tomosynthesis image pixels are formed by subdividing the lateral shift between these rays. (C) The multiplane composite can be viewed as a “virtual projection” of the in-focus features of the tomosynthesis images.

TABLE I. Glossary of symbols.

x, y, z	Coordinate axes of the SBDX system
u, v	Integer column and row indices of a pixel in a reconstructed plane or composite image
$\theta_x, \theta_y, \theta_z, t_x, t_y, t_z$	Rotation angles and translations corresponding to the x , y , and z axes
$I(u, v, z)$	Tomosynthesis plane stack
p_x, p_y	Virtual detector element pitch in the x and y directions
d_x, d_y, SDD	Distance between center of the virtual detector and the virtual source point, along x , y , and z
M	Denotes local coordinate system attached to the echo probe model
$V(x_j, y_j, z_j)$	3D point intensities defining the probe model in coordinate system M
$D(u, v)$	Digitally reconstructed radiograph of the probe model
$\text{SBDX}T_M$	Transformation from probe model coordinates to SBDX coordinates
${}^{\text{CT}}T_{\text{echo}}$	Transformation from echo image coordinates to CT coordinates used in echo calibration
${}^M T_{\text{CT}}$	Transformation from echo calibration CT coordinates to probe model coordinates
${}^M T_{\text{echo}}$	Transformation from echo image coordinates to probe model coordinates
TRE	Target registration error

image data may then be registered to XRF. Details of SBDX image reconstruction, pose estimation, and visualization of the registered images are described in Table I.

2.A. SBDX image reconstruction

During SBDX imaging, an electron beam is raster-scanned over an array of focal spot positions. A multihole collimator defines a series of narrow overlapping x-ray beamlets directed at the detector. The detector captures a small image for each collimator hole illumination, and the images are transmitted to GPU-based hardware for real-time reconstruction.⁹ A two stage reconstruction process is executed for the detector images acquired in every 1/15 s scan frame. First, digital tomosynthesis is performed in parallel at a stack of 32 planes spaced by 5 mm. As described in Ref. 8, an unfiltered backprojection technique is used (“shift-and-add” tomosynthesis). In the tomosynthesis images, in-plane objects appear sharp, and out-of-plane objects are progressively blurred as the plane-to-object distance increases. In the second stage, a 2D composite image is formed from the tomosynthesis stack in order to display all objects in focus simultaneously. The composite is generated by a plane selection algorithm, which, for each pixel position, selects the pixel value from the tomosynthesis plane with the highest local contrast and sharpness. Field-of-view and frame rate are dictated by the number of focal spots scanned and the number of electron beam dwells per focal spot.⁸ In this work, scanning was performed with 71×71 holes, 8 dwells per hole per scan frame, and 15 scan frames/s. Composite images and plane stacks were reconstructed at 15 Hz, and the isocenter plane reconstruction measured 11.4 cm wide. The source-to-detector distance (SDD) is fixed at 1500 mm.

The coordinate system of the SBDX C-arm is defined such that x corresponds to the horizontal image direction, y the vertical image direction, and z is the distance along the source-detector axis. The (x, y, z) origin is located at the center of the focal spot array. The pixel coordinates of tomosynthetic and composite images are referred to by the integers u and v . The u - and v -axes are parallel to x - and y -axes, respectively. When a plane stack $I(u, v, z)$ is described, z is assumed to take on the discrete values, in millimeters, corresponding to the plane positions.

The SBDX/TEE registration algorithm uses the 3D plane stack for initial probe pose estimation and the 2D composite image for final pose refinement. Since the SBDX image coordinate system is relevant to these tasks, a brief review is provided here. The pixel pitch in each tomosynthesis plane is defined by dividing the shift distance between adjacent back-projected images into a fixed number ($m = 10$) of pixels. Since the x-ray beamlets originate from a regularly spaced array of focal spot positions in the source and they all converge to a common point on the detector, the pixel centers for the stack of planes fall along a cone of rays originating from the center of the detector (see Fig. 1). That is, a ray corresponds to fixed (u, v) in the plane stack $I(u, v, z)$. The composite image contains the in-focus pixel value for each of these rays. Thus, the 2D composite can be viewed as an inverted “virtual projection” of the in-focus features in the patient volume, where the “virtual source” is at the center of the detector and the “virtual detector” is located at the source plane. The virtual detector pitch is the focal spot pitch (2.3 mm) divided by $m = 10$. For more details, we refer the reader to Ref. 13. The use of this virtual projection model in 2D/3D registration is described in Sec. 2.C.

The coordinate system “ M ” of the TEE probe is defined such that the probe face from which the ultrasound volume emanates points in the positive z -direction (toward the SBDX detector), and the long-axis of the probe points in the negative y -direction of the SBDX system (toward patient inferior). The rotational pose parameters for the probe angle (θ_x , θ_y , and θ_z) correspond to sequential Euler angle rotations about the SBDX coordinate system axes, in the order $y \rightarrow x \rightarrow z$. This corresponds to a rotation about the long-axis of the probe (“roll”), followed by a rotation about the short-axis of the probe (“pitch”), and then finally a rotation of the probe about the z -axis (“yaw”). Figure 2 demonstrates a TEE probe model after it has been rotated and translated to a position in the SBDX coordinate system.

2.B. Initial 3D pose estimation from tomosynthesis

An initial estimate of the 3D position and orientation of the probe in the SBDX C-arm coordinate system is obtained from the tomosynthetic plane stack $I(u, v, z)$ generated in a 1/15 s frame period. To obtain the position along the source-detector axis, the method exploits the fact that a device feature appears most in focus in the image plane closest to that feature and is progressively blurred as the plane-to-feature distance increases (Fig. 3). The z -location of the device feature is determined with finer precision than the plane-to-plane spacing by analyzing the distribution of feature sharpness versus

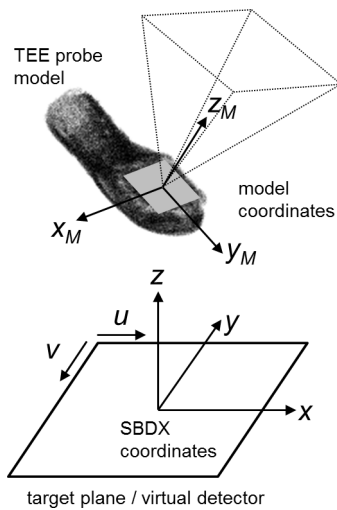


FIG. 2. The transformation $SBDX T_M$ maps 3D points in the local coordinate system of the probe model (M) to 3D points in the SBDX coordinate system.

the z -coordinate of each plane. The method has three steps: (i) detection of probe key-points in the composite image, (ii) 3D localization of key-points using the tomosynthesis planes, and (iii) principal component analysis (PCA) for orientation estimation (see Fig. 4).

2.B.1. Key-point detection

First, the center pixel of the square transducer face of the TEE probe is located in the composite image. This was done manually, although automatic techniques¹⁴ can also be applied. A segmentation of the TEE probe is then generated by applying the Frangi vesselness filter to the composite image¹⁵ followed by thresholding and dilation with a 10-pixels wide circular structuring element. The TEE probe is typically the largest high-contrast object in the image. Therefore, the largest connected component is found and all others are removed to produce the probe segmentation mask [Fig. 4(C)]. To detect key-points within the mask, first the gradient magnitude of the composite image is computed following convolution with a Gaussian kernel ($\sigma = 1.0$ pixel). Next, a phase-symmetry filter is applied¹⁶ to enhance salient edges in the image while suppressing noisy edges. The result is multiplied by the mask to produce an edge image [Fig. 4(E)]. A local maximum filter is applied where a pixel is set to 1 if it is equal to the maximum value within a 7×7 window, and 0 otherwise. The result specifies the initial key-point locations [Fig. 4(F)].

2.B.2. 3D localization of key-points

At each key-point position (u_k, v_k) , the image gradient magnitude is sampled at all 32 planes to create a 32-vector of edge strength values [Fig. 4(G)]. The gradient magnitude in each plane of the stack $I(u, v, z)$ is computed using the finite difference method after convolution with a 2D Gaussian kernel ($\sigma = 1.0$ pixel). Since the tomosynthetic blurring behavior is locally symmetric about the true object z -position, the vector of edge strength values can be viewed as a sampled version of a function with its centroid located at the true object position. Local edge strengths about the object are obtained by applying a threshold [see Fig. 4(G)]. Denoting the original distribution of edge strengths as $C_k(z)$, the thresholded distribution is

$$\hat{C}_k(z) = \begin{cases} C_k(z) - A & \text{if } C_k(z) > A \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where $A = 0.75 \max(C_k(z))$. The z -position z_k of a key-point (u_k, v_k) is then calculated as the center-of-mass of this distribution,

$$z_k = \frac{\sum_{i=1}^{32} z_i \hat{C}_k(z_i)}{\sum_{i=1}^{32} \hat{C}_k(z_i)}. \quad (2)$$

For each vector $\hat{C}_k(z)$, the number of local peaks is found. Vectors with more than one peak are removed from consideration, as they often result in unreliable 3D localization estimates.

The localized key-point positions are converted from (u, v, z) to (x, y, z) coordinates using precalculated lookup tables. At this stage of the algorithm, most key-points belong to edges of the probe. However, some key-points have unrealistic z -coordinate values and should be labeled as outliers. To remove them, the median z -coordinate of all key-points is calculated, and any point that is greater than 15 mm away from the median value is removed. (The distance 15 mm was chosen based on the dimensions of the TEE probe.) This mechanism is also designed to reject erroneous z -coordinates caused by overlapping objects, such as a catheter.

2.B.3. Initial 3D pose

With the remaining set of 3D key-points, PCA is used to determine a rough 3D pose. PCA finds the directions of the highest variance in N -dimensional data. The first principal

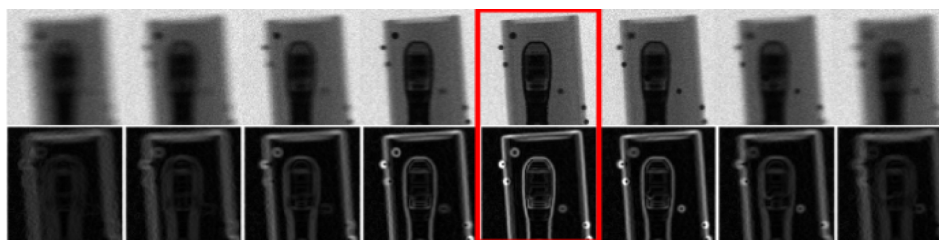


FIG. 3. Top row: Tomosynthesis images of a TEE probe head reconstructed at different planes relative to the SBDX source. Bottom row: The edge magnitudes grow weaker as the distance between the probe and the reconstructed image plane increases. The in-focus plane is indicated with the red rectangle.

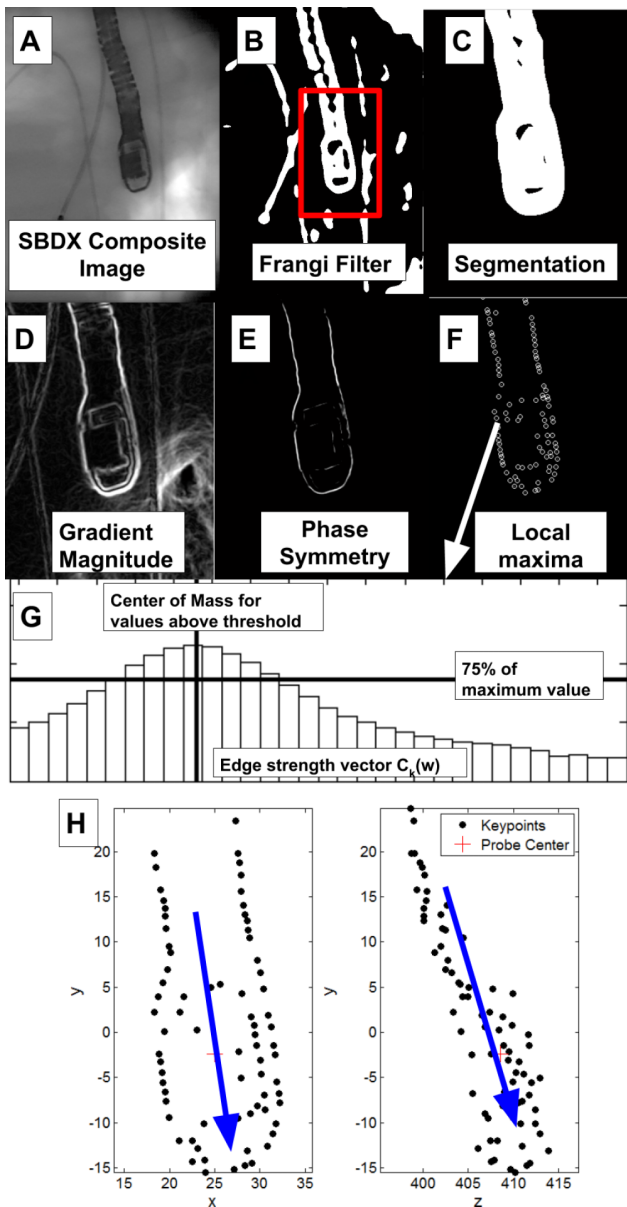


FIG. 4. (A) Original SBDX composite image. (B) Composite image after applying the Frangi filter. (C) The largest connected component of the filtered image is extracted and dilated to roughly segment the probe. (D) The gradient magnitude of a subwindow of the original image. (E) Phase-symmetry filter applied to the gradient magnitude. (F) Local maxima of the phase-symmetry image. (G) An example of the edge strength for each SBDX plane for a single 2D key-point. For the center-of-mass computation, only values greater than 0.75 of the maximal value are considered. (H) Final 3D key-points, with principal direction computed via principal component analysis.

component, therefore, defines the direction that the long-axis of the TEE probe is aligned with, which in turn is used to determine the in-plane rotation of the probe (θ_z ; yaw) and the out-of-plane pitch (θ_x). Furthermore, the average z -location of the 3D key-points is used to estimate the central z -coordinate of the probe (t_z) by finding the mean z -value of all key-points within 10 mm of the center pixel (chosen based on the size of the TEE probe). Figure 4(H) demonstrates localized probe key-points in three dimensions along with the orientation vector determined by PCA.

2.C. Pose refinement based on 2D/3D registration

After the initial pose estimation step, the final estimation of all pose parameters is achieved through 2D/3D registration. The TEE probe is modeled as a point-cloud model (Sec. 3.A.1), with its own coordinate system M . The pose parameters, applied to the probe model, refer to the three translations (t_x, t_y, t_z) and Euler angle rotations ($\theta_x, \theta_y, \theta_z$) about the axes (x, y, z) of the SBDX system. The full spatial transformation of the TEE probe is stored in the matrix ${}^{SBDX}T_M$,

$${}^{SBDX}T_M = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_z & -s_z & 0 & 0 \\ s_z & c_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c_x & -s_x & 0 \\ 0 & s_x & c_x & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_y & 0 & -s_y & 0 \\ 0 & 1 & 0 & 0 \\ s_y & 0 & c_y & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (3)$$

where $c_j = \cos(\theta_j)$ and $s_j = \sin(\theta_j)$.

As explained in Sec. 2.A, the SBDX system geometry is different than the geometry of a standard C-arm imaging system. However, when considering the displayed composite image, the imaging geometry can be viewed as a single virtual inverted cone-beam projection, where the rays originate from the center of the detector and diverge in the direction of the x-ray tube. The matrix P defines the virtual projection geometry from the SDD, distance of virtual source to the center of the virtual detector in the $x(d_x)$ and $y(d_y)$ directions, and the virtual detector element spacing in the $x(p_x)$ and $y(p_y)$ directions,

$$P = \begin{bmatrix} 1/p_x & 0 & 0 & d_x \\ 0 & 1/p_y & 0 & d_y \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/SDD & 1 \end{bmatrix}. \quad (4)$$

P was calibrated using a helix phantom (Sec. 3.A.2). The value of SDD is 1500 mm, and the nominal virtual detector element spacing is $2.3 \text{ mm}/10 = 0.23 \text{ mm}$ in both directions.

With these definitions, the 2D/3D registration proceeds as follows: (i) Given a vector of initial pose parameters, $\boldsymbol{\varphi}$, generate a DRR from a 3D model of the probe. (ii) Compute the similarity between the DRR and the SBDX composite image. (iii) Using a nonlinear optimizer, repeat with different $\boldsymbol{\varphi}$ until the similarity is maximized. DRRs were generated using a point splatting method, similar to wobbled splatting.¹⁷ Using this method, a DRR is generated by projecting point intensities, usually from a CT volume $V(x, y, z)$, onto the image. Each pixel in the DRR image takes on a value equal to the sum of the values of the voxels that project onto it,

$$D(u_i, v_i) = e^{-\alpha \sum_{j \in S_i} V(x_j, y_j, z_j)}, \quad (5)$$

where S_i is the set of all voxel indices j such that the 3D point (x_j, y_j, z_j) projects onto the 2D detector point (u_i, v_i) , i.e., $P \cdot {}^{SBDX}T_M \cdot [x_j, y_j, z_j]^T = [u_i, v_i]^T$. The voxel intensities V

are normalized to a positive value range and the parameter α controls the contrast of the DRR. To facilitate cross correlation calculations, α was set to achieve contrast approximately equal to that observed in an x-ray image.

Two similarity metrics were used for optimization: normalized cross correlation (NCC) and gradient cross correlation (GCC). Normalized cross correlation is defined as

$$\text{NCC}(I_1, I_2) = \frac{\sum_i \sum_j (I_1(u_i, v_j) - \mu_1)(I_2(u_i, v_j) - \mu_2)}{\sigma_1 \sigma_2}, \quad (6)$$

where μ is the image mean and σ is the image standard deviation. GCC is defined as

$$\text{GCC}(I_1, I_2) = 0.5 \cdot \text{NCC}(G_{x1}, G_{x2}) + 0.5 \cdot \text{NCC}(G_{y1}, G_{y2}), \quad (7)$$

where G_x is the image x -gradient and G_y is the image y -gradient. 2D/3D registration consisted of three optimization stages: (i) optimization of the in-plane parameters (t_x, t_y , and θ_z) using NCC, (ii) all parameters except t_z using NCC, and (iii) all parameters, including the DRR contrast parameter α , using GCC. The Nelder–Mead optimizer was used at every registration stage.

3. METHODS

3.A. Calibrations

3.A.1. Probe model

In order to compute splat rendered DRRs for 2D/3D registration, a point-cloud model of the TEE probe was generated from a cone-beam CT of the probe.¹⁸ This was done by manually segmenting voxels belonging to the TEE probe and then randomly sampling 2^{20} points within the segmented volume. The intensity associated with each point was obtained using linear interpolation from the CT volume.

3.A.2. SBDX C-arm calibration

The 3D/2D transformation matrix (P) describing the SBDX virtual projection geometry was calibrated using a precision manufactured phantom with steel beads arranged in a helical pattern. The helix phantom was placed at approximately the isocenter and imaged. To maximize SNR, a 64-frame average was formed. The image was then manually thresholded to segment each fiducial, and the intensity centroid of each fiducial was calculated. An initial P matrix was generated using the nominal virtual projection geometry, and the helix model pose was manually initialized. Next, the Levenburg–Marquardt algorithm was used to optimize helix model pose with fixed P . Following convergence, the helix pose was fixed, and P was optimized. This was repeated until the fiducial registration error converged to a minimal value.

3.A.3. Echo calibration

The spatial transform relating the echo image space to the TEE probe model (M), ${}^M T_{\text{echo}}$, was found using a wire phantom. The phantom consisted of a water-filled cylinder

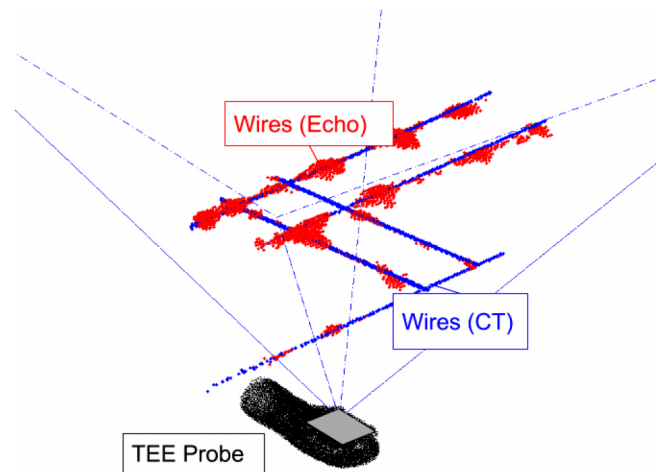


FIG. 5. Illustration of TEE probe model to echo image calibration. The probe model is registered to a CT image of the wire phantom. The wires from echo are then registered to the wires in the CT volume. This allows the spatial relationship between the TEE probe model and the echo image volume to be established.

containing metallic wires and an entrance port for the TEE probe. A CT image was acquired of the entire setup while a simultaneous 3D echo of the metallic wires was recorded. Using standard intensity-based registration, the echo image of the wires was computationally registered to the wires in the CT image to find ${}^{\text{CT}} T_{\text{echo}}$ (see Fig. 5). The probe model (generated from a previously acquired high-resolution CT) was similarly registered to the probe visible in the CT of the phantom to obtain ${}^M T_{\text{CT}}$. These two transforms were combined to obtain ${}^M T_{\text{echo}}$,

$${}^M T_{\text{echo}} = {}^M T_{\text{CT}} {}^{\text{CT}} T_{\text{echo}}. \quad (8)$$

To find the TRE of the echo volume-to-probe registration, voxels from the echo volume, p_{echo} , and the CT volume, p_{CT} , belonging to the wires were extracted by manually setting an intensity threshold for each image. For each wire voxel in echo, the distance to the nearest wire voxel in the CT image following registration was calculated. The TRE defined as the RMS distance over all of these distances was found to be 1.71 mm,

$$\text{TRE}_{\text{echo}} = \sqrt{\frac{\sum_i \|\min(p_{\text{CT}} - {}^{\text{CT}} T_{\text{echo}} \cdot p_{i,\text{echo}})\|_2^2}{N}}. \quad (9)$$

3.B. Pose estimation accuracy

A study was performed to compare the TEE pose estimation results with a ground truth reference at eight different TEE probe orientations (see Fig. 6), and a range of image signal-to-noise ratios. The ground truth was established by embedding the TEE probe within a PVC cylinder covered with spherical steel ball bearings (2.5 and 3 mm diameters). The probe was fixed in the cylinder using silicone rubber. A cone-beam CT of the entire fiducial/probe setup was used to establish the spatial relationship between the probe and fiducials. To measure the ground truth pose, the pose estimation algorithm was applied

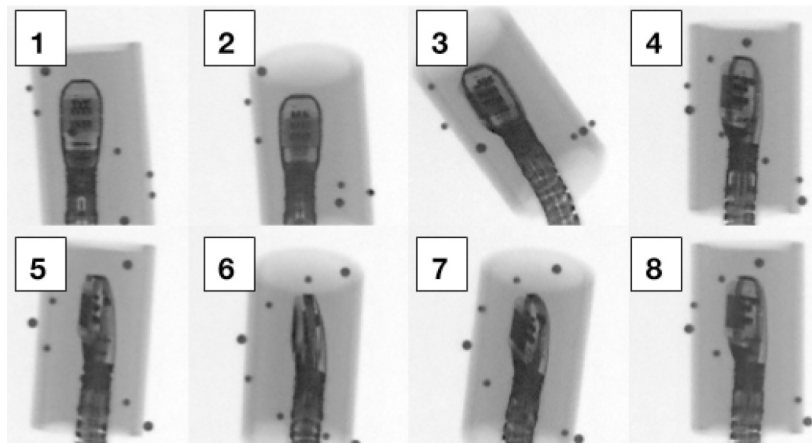


FIG. 6. The eight different TEE probe poses tested for the pose estimation experiments. The fiducials attached to the probe are used to estimate the ground truth pose.

to the fiducials only. SBDX imaging of the probe/helix was performed at 80 kV, 75 mA_{peak} (36% maximum tube current) in the 71 × 71 15 frames/s scan mode, with 23.3 cm acrylic in the x-ray beam (Fig. 7). SBDX image reconstructions were performed offline. Five different levels of signal difference-to-noise ratio (SDNR) were generated by randomly sampling and averaging 1, 2, 4, 8, and 16 frames. For each noise level and TEE probe pose, this was repeated 10 times, for a total of 400 experiments. For all experiments, the SDNR was computed as

$$\text{SDNR} = \frac{\mu_{\text{probe}} - \mu_{\text{background}}}{\sigma_{\text{background}}}. \quad (10)$$

In order to measure TEE probe and background signal statistics, ROI masks were created by manually setting two intensity thresholds, one to segment out the probe and one to sample the background near the probe. $\sigma_{\text{background}}$ was computed by subtracting two consecutive frames, finding the standard deviation of the difference image within the background, and dividing by $\sqrt{2}$. μ_{probe} and $\mu_{\text{background}}$ were computed by finding the mean within their respective masks for one image frame.

2D and 3D TREs were used to quantify pose estimation accuracy. For this experiment, the TRE was based on a set of $N = 100$ virtual points defined in the echo image space, randomly and uniformly distributed within a 50 mm wide cubic volume. The virtual points p in echo space were transformed to the C-arm coordinate system using both the ground truth pose and the estimated pose, yielding point sets p_{true} and $p_{\text{estimated}}$, respectively. The TRE_{3D} was then computed for each experiment as

$$\text{TRE}_{3\text{D}} = \sqrt{\frac{\sum \|p_{\text{true}} - p_{\text{estimated}}\|_2^2}{N}}. \quad (11)$$

The TRE_{2D} was computed the same way, but only the x and y coordinates were used in the Euclidean distance computation. TRE_{3D} is the total target registration error, while TRE_{2D} is representative of the error for points in echo in the plane parallel to the XRF detector. For this study, the overall registration error was based purely on pose estimation of the probe and did not include errors in registering the echo image space to the probe model (this additional error is considered in Sec. 3.C).

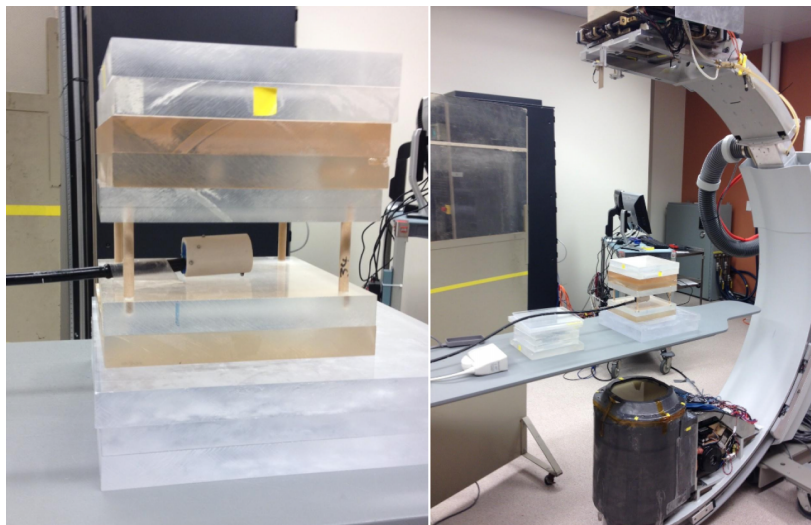


FIG. 7. The experimental setup for the phantom pose estimation experiment. Left: The TEE probe, embedded in the PVC cylinder with surrounding fiducials, was imaged between layers of acrylic to decrease the signal-to-noise. Right: A zoomed out view of the experiment showing the SBDX C-arm.

3.C. Phantom and *in vivo* studies of SBDX/TEE registration

Water tank phantom and *in vivo* experiments were conducted to evaluate two image fusion scenarios: echo-to-SBDX fusion, where features within the 3D echo image were projected onto the 2D SBDX image, and SBDX-to-echo fusion, where the 3D locations of devices from the SBDX image space were transformed into the 3D echo image space. The first scenario maintains the conventional 2D x-ray display format while adding anatomical structures rendered/segmented from echo, whereas the second scenario enables the fusion of 3D SBDX catheter tracking results with the native display format of 3D echocardiography.

3.C.1. Phantom study

For the phantom experiment, a cylindrical polyvinyl alcohol (PVA) phantom with a ventricle sized cylindrical cavity (height = 35 mm, radius = 15 mm) was fabricated. An injection catheter (MyoStar, Biosense Webster) with a metallic tip was guided through a plastic tube on the proximal side of the phantom, until it was positioned against the distal wall of the cavity. The proximal end of the catheter was attached to a translation stage, and a 5 mm/s catheter pullback was performed under simultaneous echo and SBDX imaging. The resulting trajectory of the catheter tip was a straight path mainly in the negative y -direction. Sequences from two different C-arm angles (15° LAO, 0° CC and 15° LAO, 10° CC) were performed, resulting in different appearances of the TEE probe. Imaging was performed at 80 kV, 75 mA_{peak}. Background x-ray attenuation was provided by 15 cm water, 2 cm of wood, and 1 cm of polyurethane plastic.

To evaluate the 3D TRE of SBDX-to-echo 3D image registration, first tomosynthesis-based 3D tracking of the catheter tip in SBDX space was performed using the algorithm in Ref. 10. The tip coordinate was then transformed to the echo image space using the TEE probe pose estimate and the echo-volume-to-probe calibration. The transformed coordinate was then compared to the catheter tip location as manually identified from the 3D echo images. For this task, the centroid of the reverberation artifact was located, which was presumed to correspond to the metal tip of the catheter (Fig. 8). The 3D TRE was calculated for each frame using the following equation:

$$\begin{aligned} \text{TRE}_{\text{tip}} &= \left\| p_{\text{echo}} - {}^{\text{echo}}T_{\text{SBDX}} \cdot p_{\text{SBDX}} \right\|_2, \\ {}^{\text{echo}}T_{\text{SBDX}} &= \left({}^{\text{SBDX}}T_M^M T_{\text{echo}} \right)^{-1}. \end{aligned} \quad (12)$$

As in the pose estimation accuracy experiments, TRE_{2D} was computed by considering only the x and y coordinates.

3.C.2. *In vivo* study

A 50 kg healthy swine with 24 cm anterior–posterior chest thickness was imaged in the 71 × 71 15 frames/s scan mode and 100 kV, 120 mA_{peak} (50% maximum tube current) x-ray technique. Procedures were approved by the local Institutional Animal Care and Use Committee. Three image

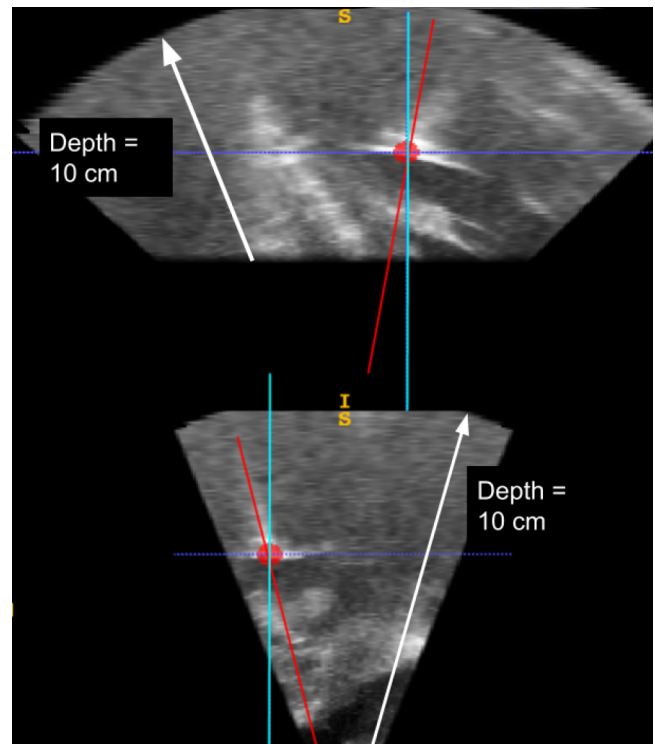


FIG. 8. Method for catheter segmentation in the *in vivo* and water tank experiments. The catheter tip was found by determining the line that passed through the 3D reverberation artifact.

sequences were performed under simultaneous SBDX and TEE guidance. For the first two sequences, an injection catheter (MyoStar) with a metallic tip was guided into the left ventricle (LV). In sequence 1, the catheter tip was manipulated throughout the left ventricle to mimic navigation toward a target site. In sequence 2, the catheter was positioned at a single location against the left ventricular wall to mimic a catheter position confirmation task. In the latter case, the catheter only underwent cardiorespiratory motion. These two sequences were used to evaluate the registration accuracy for a discrete tip.

The third sequence was used to evaluate the qualitative accuracy of anatomic echo-to-SBDX registration. Specifically, a ventriculogram was acquired under simultaneous SBDX and echo imaging in order to compare a standard x-ray ventriculogram with a proposed echo-based ventriculogram, in which the LV is segmented from the echo data and overlaid on the fluoroscopic image. Additional TRE measurements were obtained from the metallic markers of the pigtail injection catheter present in this sequence.

Since the SBDX and echo data were recorded simultaneously on separate systems, temporal synchronization of image frames was necessary. To synchronize the images, each modality was first analyzed to determine the spatial axis with the largest variation of catheter motion. Next, the 1D position of the catheter along that axis was recorded as a 1D signal. Finally, the 1D motion “signals” from both modalities were compared and the time-shift that resulted in the highest normalized cross correlation was used to temporally align the image sequences.

TRE was calculated in the same way as in the phantom study, with the exception of the ventriculogram sequence. For that sequence, the multiple metallic markers present on the pigtail catheter were indistinguishable in the echo image. Therefore, a spline, s_{echo} , was fit to a set of manually segmented points on the catheter in the echo image, and the TRE was the root mean square of the minimal distance between the markers registered from SBDX and the spline,

$$\text{TRE}_{\text{pigtail}} = \sqrt{\frac{\sum \|\min(s_{\text{echo}} - \text{echo}T_{\text{SBDX}} \cdot P_{\text{SBDX}})\|_2^2}{N}} \quad (13)$$

4. RESULTS

4.A. Pose estimation accuracy

Figures 9(a) and 9(b) show the average TRE_{3D} and TRE_{2D} for all successful registrations obtained in the pose estimation study. Figure 9(c) shows the success rates, defined as the percentage of registrations with a TRE_{3D} less than 5 mm. While this threshold is application dependent, 5 mm was chosen because it represents a registration error that would result in suboptimal placement of a prosthetic valve during TAVR (Ref. 19) or suboptimal catheter-based targeting of therapeutic injections.⁵ The eight poses tested are shown in Fig. 6. The five SDNR levels tested were 5.9 ± 0.3 , 9.4 ± 0.8 , 14.8 ± 1.2 , 22.0 ± 2.0 , and 33.4 ± 4.0 . With the exception of pose 5, the success rate was 97.1% for all experiments, with a mean TRE_{3D} of 2.32 ± 1.88 mm and TRE_{2D} of 1.61 ± 1.36 mm. Pose 5, with a rotation of $\theta_y = 76^\circ$, rarely converged, an issue which is

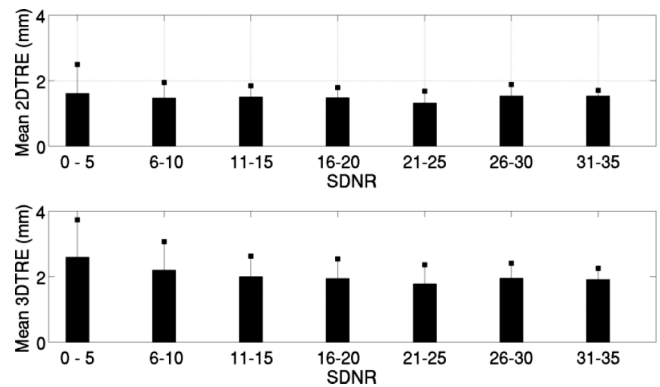


FIG. 10. Mean TRE_{2D} (top) and TRE_{3D} (bottom) for varying levels of SDNR.

addressed in Sec. 5. Considering all poses and SDNR levels, the probe localization z-error along the source-detector axis was 0.87 ± 1.31 mm.

Higher image SDNR tended to improve TRE (Fig. 10), although for SDNR in the range of 11–35 the TRE did not vary much. For experiments with probe orientations typically seen in clinical cases (TEE probe roll $< |60^\circ|$, poses 1–4 and 8), and with SDNR > 18.8 , the registration success rate was 100%, the TRE_{3D} was 1.76 ± 0.59 mm, and the TRE_{2D} was 1.40 ± 0.40 mm. For reference, the SDNRs in the *in vivo* study were 35–39.

4.B. Water tank phantom and *in vivo* studies

Table II shows the TRE results for the water tank phantom and *in vivo* studies. The full image registration pipeline

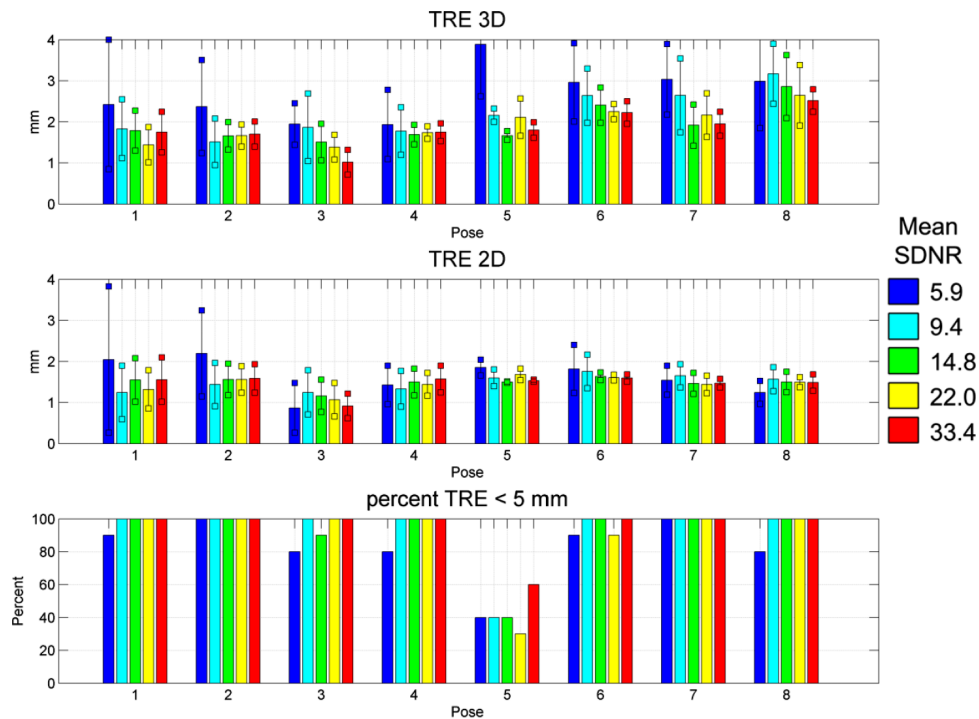


FIG. 9. Summary of pose estimation accuracy experiments for varying poses and increasing SDNR (via image averaging). Bars indicate the mean value and error bars indicate one standard deviation. Ten images were generated for each pose and SDNR level. (a) TRE_{3D}. (b) TRE_{2D}. (c) Percentage of successful registrations, defined as TRE_{3D} < 5.0 mm. For each SDNR/pose combination, $n = 10$. (Truncated result in the top panel: pose 5, 1 frame = 3.86 ± 1.26 mm.)

TABLE II. Results for the water tank phantom and *in vivo* experiments.

Sequence	Frames	TRE _{2D} (mm)	TRE _{3D} (mm)	SDNR
		mean ± std	mean ± std	
Water tank 1	26	0.95 ± 0.44	1.84 ± 0.63	63.5
Water tank 2	21	1.78 ± 0.79	2.02 ± 0.76	54.7
<i>In vivo</i> 1	53	1.57 ± 0.89	4.64 ± 1.95	35.8
<i>In vivo</i> 2	18	1.15 ± 0.49	2.62 ± 0.49	34.8
<i>In vivo</i> 3	35	1.61 ± 1.11	3.87 ± 1.13	38.8

resulted in a mean 3D error of 3.39 mm over all experiments and image frames, with errors in individual frames ranging from 0.22 to 9.1 mm. The water tank experiments showed lower TRE values than the *in vivo* experiments, which is likely due to the tighter control over experimental variables. Potential causes of errors are outlined in Sec. 5.

Figure 11 demonstrates echo-to-SBDX registration and SBDX-to-echo registration for the catheter tip in the second *in vivo* sequence. In the echo-to-SBDX registration, the catheter tip segmented from echo is registered to the 2D SBDX image (blue circle). The TRE_{2D} values in Table II represent the error in this registration. In the SBDX-to-echo registration, tomosynthesis-based catheter tip tracking is registered to two planes of the echo image volume and displayed as red circles. The error in this process is characterized by TRE_{3D}.

Figure 12 demonstrates an echo-to-SBDX registration of the endocardial surface of the left ventricle. A 3D ventricular volume was manually segmented from an end-diastolic echo image volume and then registered to SBDX using the TEE probe pose. The segmented 3D volume was then projected onto the SBDX image and the borders of the projected segmentation were displayed. For comparison, a contrast-enhanced ventriculogram was performed with SBDX. A good agreement exists between the visible borders of the x-ray contrast and the echo-based borders.

5. DISCUSSION

XRF is generally considered the primary imaging modality for guidance of devices in structural heart interventions, but soft tissue visualization is poor, the projection format

creates ambiguity, and ionizing radiation dose is an ongoing source of concern. Previous work has demonstrated the potential of SBDX to both reduce dose and provide 3D catheter tracking.^{10,20} The registration of 3D echo with SBDX could address the remaining need for real-time soft tissue anatomy in a common visualization environment. To this end, we have developed and evaluated an algorithm for SBDX/TEE registration.

The SBDX/TEE registration algorithm combines tomosynthesis-based 3D localization with a version of 2D/3D registration adapted for inverse geometry x-ray imaging. In the initial pose estimation stage, the algorithm was able to localize the correct *z*-position of the TEE probe to within 0.87 mm on average. The ability of inverse geometry fluoroscopy to resolve depth in a single image frame is a unique advantage compared to standard fluoroscopy, which generally requires either biplane imaging or multiple acquisitions at different C-arm projection angles to localize the TEE probe in three dimensions.

The study of pose estimation accuracy found TRE_{3D} < 3 mm in individual images, for all experiments conducted at SDNR levels similar to those that were encountered *in vivo*. At lower SDNR levels, the pose corresponding to a primarily lateral view of the TEE probe (pose 5, Fig. 6) resulted in poor registration convergence. Visual inspection revealed this was due to an error in the final θ_y (roll) and θ_x (pitch) parameters. Additional work is needed to address this issue, but we note that in TAVR procedures performed at our own institution, the occurrence of this pose is extremely rare since the probe is almost always facing toward the x-ray detector while imaging the heart. Future work should also validate pose estimation accuracy in the presence of overlapping high-contrast objects in the field-of-view, such as a catheter.

For the water tank and *in vivo* studies, a general increase in TRE relative to the pose accuracy study was observed. This was expected because the targets were real catheter tips rather than virtual objects. Under this scenario, additional sources of registration error included localization of the catheter tip in echo and SBDX, echo volume-to-TEE probe calibration error (TRE = 1.71 mm), and potential temporal synchronization errors. For example, the 3D localization of the catheter tip in the SBDX plane stack was expected to introduce approximately 1.0 mm error in the *z*-direction.¹⁰

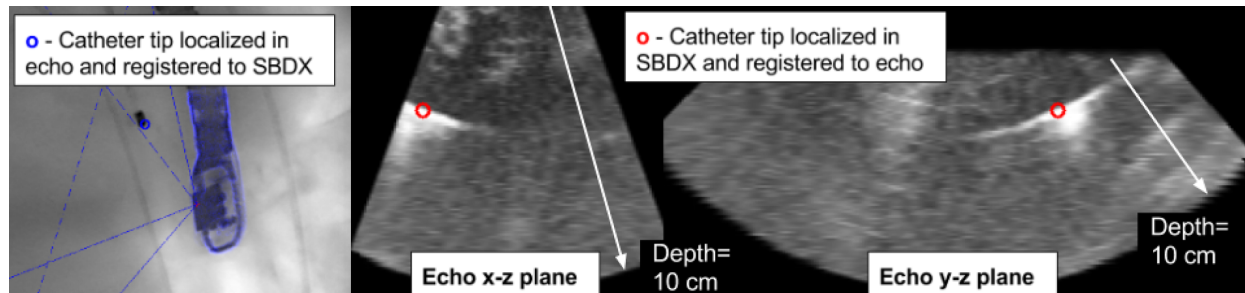


FIG. 11. Left: A SBDX composite image is shown, with the catheter tip location from echo overlaid onto the image, demonstrating TRE_{2D}. Right: Two orthogonal slices from the 3D echo corresponding to the SBDX composite image on the left. The catheter tip, localized in SBDX, is transformed and overlaid onto the echo image.

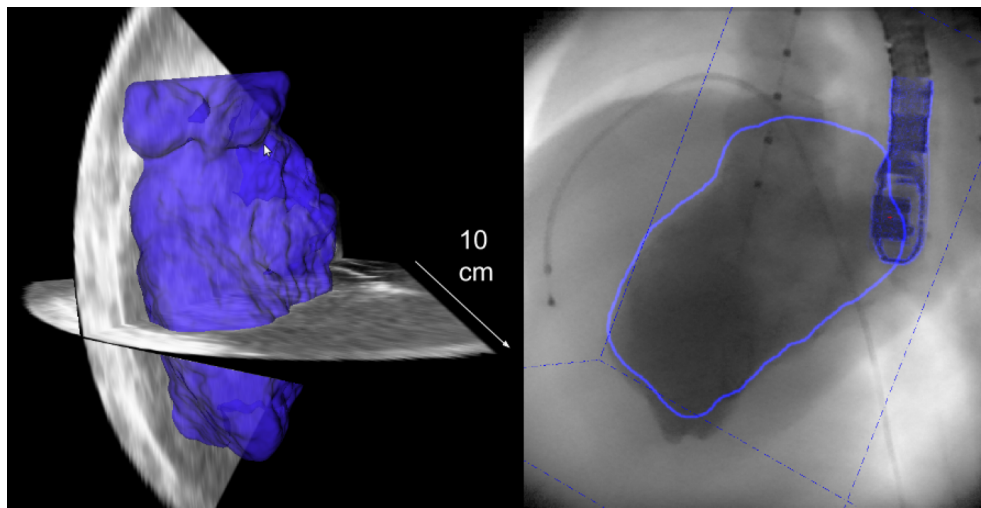


FIG. 12. Left: Two orthogonal slices through a 3D echo volumetric frame taken during the *in vivo* experiment, with a semiautomatically generated segmentation of the left ventricle. Right: The corresponding SBDX composite image, with the left ventricle segmentation borders registered and projected onto the SBDX image.

This study demonstrates two potential approaches to SBDX/echo visualization. In an echo-centric display, 3D echo images could be augmented with 3D representations of the catheter device derived from SBDX device localization. Alternatively, a live 2D fluoroscopic image could be combined with soft tissue anatomy segmented from simultaneous 3D echo. Future work will investigate the utility of these approaches in different structural heart interventional tasks. Note that in this initial study, SBDX/TEE image fusion was implemented in MATLAB. For real-time guidance experiments, implementation on GPU-based hardware will be required. Additionally, future work should include automated procedures for initialization of the registration.

6. CONCLUSIONS

Image registration between a low dose SBDX system and TEE has been demonstrated. A novel 6 degree-of-freedom localization algorithm was presented, and the registration feasibility and accuracy were evaluated in phantoms and *in vivo*. Future technical work will focus on real-time implementation and fully automatic registration initialization.

ACKNOWLEDGMENT

Partial financial support for this work was provided by NIH Grant No. R01 HL084022.

^{a1}Author to whom correspondence should be addressed. Electronic mail: speidel@wisc.edu

¹G. Gao, G. Penney, Y. Ma, N. Gogin, P. Cathier, A. Arujuna, G. Morton, D. Caulfield, J. Gill, and C. A. Rinaldi, "Registration of 3D trans-esophageal echocardiography to x-ray fluoroscopy using image-based probe tracking," *Med. Image Anal.* **16**, 38–49 (2012).

²P. Lang, P. Sestija, M. W. Chu, D. Bainbridge, G. M. Guiraudon, D. L. Jones, and T. M. Peters, "US-fluoroscopy registration for transcatheter aortic valve implantation," *IEEE Trans. Biomed. Eng.* **59**, 1444–1453 (2012).

³C. Hatt, M. Speidel, and A. Raval, "Robust 5DOF transesophageal echo probe tracking at fluoroscopic frame rates," in *International Conference on Medical Image Computing and Computer Assisted Interventions*, edited by J. H. Nassir Navab, S. Wells, and A. F. Frangi (Springer, Munich, Germany, 2015).

⁴P. Mountney, R. Ionasec, M. Kaizer, S. Mamaghani, W. Wu, T. Chen, M. John, J. Boese, and D. Comaniciu, "Ultrasound and fluoroscopic images fusion by autonomous ultrasound probe detection," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2012* (Springer, Berlin Heidelberg, 2012), pp. 544–551.

⁵C. R. Hatt, A. K. Jain, V. Parthasarathy, A. Lang, and A. N. Raval, "MRI—3D ultrasound—X-ray image fusion with electromagnetic tracking for transcatheter therapeutic injections: *In-vitro* validation and *in-vivo* feasibility," *Comput. Med. Imaging Graphics* **37**, 162–173 (2013).

⁶A. Jain, L. Gutierrez, and D. Stanton, "3D TEE registration with x-ray fluoroscopy for interventional cardiac applications," in *Functional Imaging and Modeling of the Heart* (Springer, Berlin Heidelberg, 2009), pp. 321–329.

⁷L. E. Bø, H. O. Leira, G. A. Tangen, E. F. Hofstad, T. Amundsen, and T. Langø, "Accuracy of electromagnetic tracking with a prototype field generator in an interventional OR setting," *Med. Phys.* **39**, 399–406 (2012).

⁸M. A. Speidel, B. P. Wilfley, J. M. Star-Lack, J. A. Heanue, and M. S. Van Lysel, "Scanning-beam digital x-ray (SBDX) technology for interventional and diagnostic cardiac angiography," *Med. Phys.* **33**, 2714–2727 (2006).

⁹M. A. Speidel, M. T. Tomkowiak, A. N. Raval, D. A. Dunkerley, J. M. Slagowski, P. Kahn, J. Ku, and T. Funk, "Detector, collimator and real-time reconstructor for a new scanning-beam digital x-ray (SBDX) prototype," *Proc. SPIE* **9412**, 94121W (2015).

¹⁰M. A. Speidel, M. T. Tomkowiak, A. N. Raval, and M. S. Van Lysel, "Three-dimensional tracking of cardiac catheters using an inverse geometry x-ray fluoroscopy system," *Med. Phys.* **37**, 6377–6389 (2010).

¹¹M. A. Speidel, B. P. Wilfley, A. Hsu, and D. Hristov, "Feasibility of low-dose single-view 3D fiducial tracking concurrent with external beam delivery," *Med. Phys.* **39**, 2163–2169 (2012).

¹²M. T. Tomkowiak, A. N. Raval, M. S. Van Lysel, T. Funk, and M. A. Speidel, "Calibration-free coronary artery measurements for interventional device sizing using inverse geometry x-ray fluoroscopy: *In vivo* validation," *J. Med. Imaging* **1**, 033504 (2014).

¹³M. T. Tomkowiak, M. S. Van Lysel, and M. A. Speidel, "Monoplane stereoscopic imaging method for inverse geometry x-ray fluoroscopy," *Proc. SPIE* **8669**, 86692W (2013).

¹⁴C. Hatt, M. Speidel, and A. Raval, "Hough Forests for real-time, automatic device localization in fluoroscopic images: Application to TAVR," edited by N. Navab, J. Hornegger, S. Wells, and A. F. Frangi, *International Conference on Medical Image Computing and Computer Assisted Interventions* (Springer, Munich, Germany, 2015).

- ¹⁵A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98* (Springer, Berlin Heidelberg, 1998), pp. 130–137.
- ¹⁶P. Kovese, "Symmetry and asymmetry from local phase," *Presented at the Tenth Australian Joint Conference on Artificial Intelligence, Perth, Australia, 1997*.
- ¹⁷W. Birkfellner, R. Seemann, M. Figl, J. Hummel, C. Ede, P. Homolka, X. Yang, P. Niederer, and H. Bergmann, "Wobbled splatting—A fast perspective volume rendering method for simulation of x-ray images from CT," *Phys. Med. Biol.* **50**, N73–N84 (2005).
- ¹⁸C. R. Hatt, M. A. Speidel, and A. N. Raval, "Efficient feature-based 2D/3D registration of transesophageal echocardiography to x-ray fluoroscopy for cardiac interventions," *Proc. SPIE* **9036**, 90361J (2014).
- ¹⁹N. Piazza, P. de Jaegere, C. Schultz, A. E. Becker, P. W. Serruys, and R. H. Anderson, "Anatomy of the aortic valvar complex and its implications for transcatheter implantation of the aortic valve," *Circ.: Cardiovasc. Interventions* **1**, 74–81 (2008).
- ²⁰M. A. Speidel, B. P. Wilfley, J. M. Star-Lack, J. A. Heanue, T. D. Betts, and M. S. Van Lysel, "Comparison of entrance exposure and signal-to-noise ratio between an SBDX prototype and a wide-beam cardiac angiographic system," *Med. Phys.* **33**, 2728–2743 (2006).