

Real-Time Fusion of Endoscopic Views With Dynamic 3-D Cardiac Images: A Phantom Study

Stanislaw Szpala, Marcin Wierzbicki, Gerard Guiraudon, and Terry M. Peters*, *Senior Member, IEEE*

Abstract—Minimally invasive robotically assisted cardiac surgical systems currently do not routinely employ 3-D image guidance. However, preoperative magnetic resonance and computed tomography (CT) images have the potential to be used in this role, if appropriately registered with the patient anatomy and animated synchronously with the motion of the actual heart. This paper discusses the fusion of optical images of a beating heart phantom obtained from an optically tracked endoscope, with volumetric images of the phantom created from a dynamic CT dataset. High quality preoperative dynamic CT images are created by first extracting the motion parameters of the heart from the series of temporal frames, and then applying this information to animate a high-quality heart image acquired at end systole. Temporal synchronization of the endoscopic and CT model is achieved by selecting the appropriate CT image from the dynamic set, based on an electrocardiographic trigger signal. The spatial error between the optical and virtual images is 1.4 ± 1.1 mm, while the time discrepancy is typically 50–100 ms.

Index Terms—Image guidance, image warping, minimally invasive cardiac surgery, virtual endoscopy, virtual reality.

I. INTRODUCTION

THERE has been a progressive trend in recent years toward the use of minimally invasive surgical techniques to reduce the side effects of surgical procedures. The first laparoscopic cholecystectomy [1] was performed in 1985, but the concept of minimally invasive surgery was applied to coronary artery bypass graft (CABG) procedures almost thirteen years later. The first totally endoscopic CABG on animals was performed by Stevenson *et al.* [2] in 1998, and this was followed

shortly afterwards by Loulmet *et al.* [3] and Mohr *et al.* [4] who independently reported the procedure performed on humans. Kiaii *et al.* [5] performed the first minimally invasive robotically assisted coronary artery bypass (MIRCA) procedure through inter-costal ports. Such procedures are now being performed in a number of centers by tele-robotic systems, with the Zeus and da Vinci devices being the most commonly used in surgical practice [6]. It soon became apparent that the limited view of the optical endoscope (the primary intraoperative monitoring tool) and the problem of visually obstructed organs, would be improved with the addition of a virtual environment generated from preoperative data obtained from imaging modalities such as three-dimensional (3-D) computed tomography (CT) or magnetic resonance data. Vining *et al.* [7] had already demonstrated that 3-D reconstruction of CT images was useful in examining the tracheobronchial tree, while Gulbins *et al.* [8] showed that 3-D imaging was valuable in the planning of minimally invasive CABG.

We have developed a virtual cardiac surgical platform that can integrate endoscopy with preoperative images to support planning and guidance of cardiac intervention. For example, Chiu *et al.* [9] demonstrated an overlay of endoscopic images with a CT-reconstructed thorax phantom, Lehmann *et al.* [10], [11] discussed the employment of two-dimensional (2-D) bi-plane angiograms to animate a virtual model of the coronary arterial system, and Wierzbicki *et al.* [12] modeled the motion of the epicardial surface using elastic registration.

Recently, we superposed tracked-endoscope images of brain surface structures with a CT-based virtual model [13], while in other work, Shahidi *et al.* [14] applied a similar methodology to clinical work in patients. In the cardiac surgery context, Mourgues *et al.* [15] have developed an image-based refinement procedure for orienting a stereoscopic endoscope with respect to an animal heart.

This project is significant because minimally invasive surgery lacks adequate image guidance at both the procedure-planning and guidance phases. The work described here attempts to relate the endoscopic image to the context of the entire organ through dynamic fusion, and also serves to provide a virtual representation of aspects of the organ not seen by the endoscope. We discuss the fusion of real-time images from an optically tracked endoscope with a virtual model based on preoperative CT images of a beating heart phantom. The overlay of the endoscope images with the virtual model is robust with respect to the position and orientation of the endoscope and the phantom over the entire cardiac cycle. This is an extension of our earlier work [16] and provides an accurate method of establishing the endoscope camera calibration, the valida-

Manuscript received January 12, 2005; revised June 13, 2005. This work was supported in part the Ontario Heart and Stroke Foundation under Grant NA4755, in part by the Canadian Institute for Health Research under MOP 14735, in part by the National Science and Engineering Research Council of Canada under Grant R3146-A02, and in part by the Ontario Consortium for Image-guided Therapy and Surgery (OCITS). This paper is based on preliminary results published in: S. Szpala, G. Guiraudon, T. Peters, "Cardiac endoscopy enhanced by dynamic organ modeling for minimally invasive surgery guidance," MICCAI 2003(1): LNCS 2878, R. Ellis and T. Peters (eds), pp. 499-506. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was D. Rueckert. *Asterisk indicates corresponding author.*

S. Szpala was with the Imaging Research Laboratories, Robarts Research Institute, London, ON N6G-4M6, Canada. He is currently with the Toronto Sunnyside Regional Cancer Centre, Toronto, ON M4N 3M5, Canada.

M. Wierzbicki is with the Department of Medical Biophysics, University of Western Ontario, London, ON N6G-4M6, Canada.

G. Guiraudon is with the Lawson Health Research Institute, University of Western Ontario, and Canadian Surgical Technologies and Advanced Robotics (C-STAR), London, ON N6G-4M6, Canada.

*T. M. Peters is with the Imaging Research Laboratories, Robarts Research Institute, Imaging Labs, 100 Perth Dr., London, ON N6G-4M6, Canada, and the departments of Medical Biophysics, Radiology and Nuclear Medicine, and Biomedical Engineering, University of Western Ontario, London, ON N6G-4M6, Canada (e-mail: tpeters@imaging.robarts.ca).

Digital Object Identifier 10.1109/TMI.2005.853639

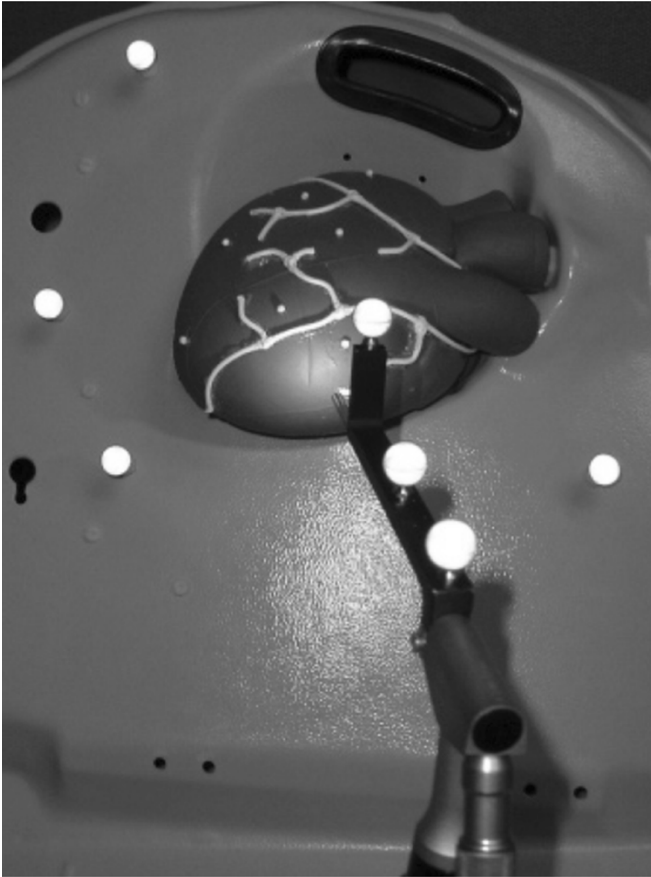


Fig. 1. The beating heart phantom in the model thorax (shown with ribs and skin removed) is visualized with an optically tracked (bright retroreflective balls) endoscope.

tion of the dynamic performance of the system in 3-D instead of 2-D, and presents a new method of animating the virtual model of the beating heart. This new approach provides results that are visually superior to those presented earlier [16], where the animated frames were the lower quality “snapshot” images acquired throughout the cardiac cycle. Our long-term objectives are similar to those of Adhami and Coste-Manière [17], who also developed an image-based infrastructure to support the planning and guidance of MIRCAB.

The paper is organized as follows: The description of the setup for overlaying optical images of a beating heart phantom with its virtual representation, and details of the system components and the software (including the algorithm for construction of the virtual model and the algorithm used in the dynamic validation), are presented in Section II. The validation of the system performance, including the intrinsic and extrinsic camera calibration, static and dynamic validation, is presented in Section III. Finally, the limitations of the procedure and the future research are described in Section IV.

II. METHODS

A. Description of the System

We overlaid endoscopic images of a beating heart phantom (simulating the environment of a human heart in situ) with a

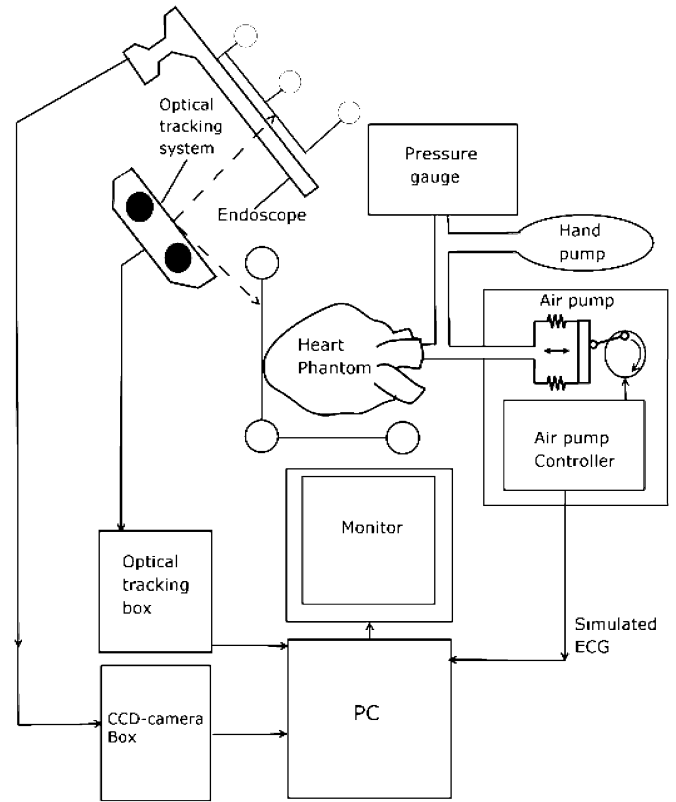


Fig. 2. The various components of the laboratory system.

virtual model of the phantom. The overlay was achieved by constructing a virtual scene placing the virtual model of the heart in the location corresponding to the actual phantom position, and with the rendering camera mimicking the real endoscope. Fig. 1 depicts the configuration of the tracked endoscope pointing at the heart phantom, and the system configuration is shown in Fig. 2. The heart phantom is continuously inflated and deflated using an air pump integrated with its controller, which has a digital pressure gauge attached to monitor the pressure. The synchronization signal simulating an electrocardiographic (ECG) trigger is interfaced with the computer through the parallel port. An optical tracking system monitors the position and orientation of both the heart phantom (casing) and the endoscope. The video endoscopic image is digitized by a frame grabber, and the endoscopic view is displayed beside the virtual scene on the monitor. The relationship between the virtual and real spaces is illustrated in Fig. 3. The virtual model of the beating heart phantom was constructed from gated 3D-CT images following the procedure described by Wierzbicki and Peters [12]. Fourteen images (representing separate phases of the cardiac cycle) were reconstructed and used to extract the dynamics of the phantom surface (see Section II-C3).

This motion information was then used to warp an image acquired during end diastole (the highest quality image of the dynamic sequence) progressively throughout the cardiac cycle. These images were sequentially displayed in the virtual scene to provide a virtual model of the beating heart phantom. The animation was synchronized with the “heartbeat” of the phantom by initiating the display cycle using the ECG “trigger” signal generated by the phantom controller. The rendering camera was

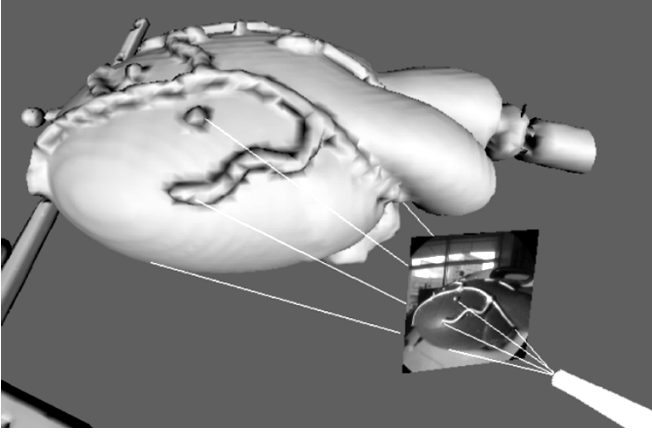


Fig. 3. The relationship of the virtual scene to the endoscopic view of the phantom.

placed in the location and orientation determined by the tracking probe attached to the endoscope, following transformation by T_R :

$$T_R = T_{EP}T_{CE} \quad (1)$$

where T_{EP} is the transform between the phantom casing (to which a reference tracking element is attached) and the tracking probe attached to the endoscope, and T_{CE} , commonly referred to as the extrinsic calibration of the camera, is the transform between the tracking probe attached to the endoscope and the coordinates associated with the endoscope lens (origin at the focal point, x axis along the optical axis of the lens, and z axis vertical on the computer screen). Unless noted otherwise, the coordinates are expressed with respect to the coordinate frame defined by the reference tracking element.

After correction for barrel distortion (see Section II-C), images from the endoscope were displayed on a semi-transparent plane placed in front of the rendering camera (see Fig. 3) at a location set by transform T_G defined as

$$T_G = T_R T_{TS} \quad (2)$$

where T_{TS} represents translation from the focal point to an arbitrarily selected point along the endoscope optical axis (2 mm from the focal point) and scaling of the endoscope image plane.

In our implementation, the virtual scene may be continuously updated to reflect the beating of the phantom as well as the relative motion between the endoscope and the heart phantom.

B. Components

1) *Heart Phantom*: In our validation experiments the beating heart was represented by an inflating/deflating heart phantom (Limbs & Things Limited, Bristol, U.K., model 40503), which allows adjustment of the heart volume using an air pump and the controller. The controller was modified to provide a pulse simulating an “r-wave” trigger pulse. The 200-ms TTL pulse was triggered by a magnetic switch that closes on a preselected phase of the air pump, and is fed to the computer via the parallel port. The volume of the phantom was uniquely correlated with the phase of the air pump by using a homing procedure, which causes the pump piston to move until

the synchronization signal is detected. In this state, the pressure in the system was adjusted to 15 ± 1 mmHg.

Coronary vessels on the surface of the inflatable heart phantom were simulated using insulated flexible copper wires that were glued to the phantom with silicone. These wires, along with 2-mm-diameter white-painted ball-bearings that were also fastened to the phantom surface for validation purposes, were readily visible in the CT image volumes.

2) *Endoscope*: The endoscopic images were provided by an Aesculap (Melsungen, Germany) 0° (i.e., straight view) fixed focal length endoscope, model PE 486A.

3) *Virtual Beating Heart Phantom*: The dynamic CT images used to construct the virtual model of the beating phantom were “preoperatively” acquired using a 3D-CT scanner (GE Medical Systems LightSpeed 8 slice helical system, 1.25-mm-thick slices, 0.5-s rotational period, 120 kV/300 mA). Because of the symmetry of the inflating/deflating motion in our model, only the first half of the cardiac cycle was reconstructed (seven phases lasting 7% of the period each), while the remaining seven phases were copied from those reconstructed during the first half period, assuming a cosine time dependence. To compensate for the missing 1% of the data in the reconstructed first half of the cycle, the eighth phase was started at 51% of the total period.

4) *Tracking*: A Polaris optical tracking system (Northern Digital Inc., Waterloo, ON, Canada) was used to track the pose of both the endoscope and the heart phantom. For endoscope tracking, a custom-made tracking probe consisting of three retroreflective balls was mounted onto the endoscope, while another set of three retroreflective balls was rigidly attached to the phantom holder to establish a reference coordinate system. Since the phantom is anchored to the phantom holder, only the casing of the latter was tracked in order to establish the pose of the heart phantom relative to the endoscope.

C. Software

1) *Barrel Distortion Correction*: Radial distortion correction was achieved through the use of a polynomial

$$r' = r + a_1 r^2 + a_2 r^3 \quad (3)$$

which maps points (r) in the distorted image into the undistorted version (r'). The coefficients a_1 and a_2 of the polynomial, as well as the screen coordinates of the distortion center, were determined in a stand-alone calibration procedure based on analysis of the image distortion of a calibration pattern of a square grid of dots (viewed normally from the distance of approximately 3 cm). The dots were indexed using an automated routine recursively identifying the neighboring grid dots following color-coded guiding lines. To find the optimum values of the fitted parameters, the cumulative distance between the observed and the required (square-grid) screen positions of the dots in the calibration pattern was minimized. This procedure takes approximately one minute and is appropriate for tilts of the axis of the endoscope of $\pm 20^\circ$ from the normal of the calibration pattern.

2) *Program Design*: The program controlling the system was written in Python 2.1, and makes extensive use of open-source classes from VTK 4.1 (www.kitware.com) and Atamai 1.0 (www.atamai.com). The software was implemented

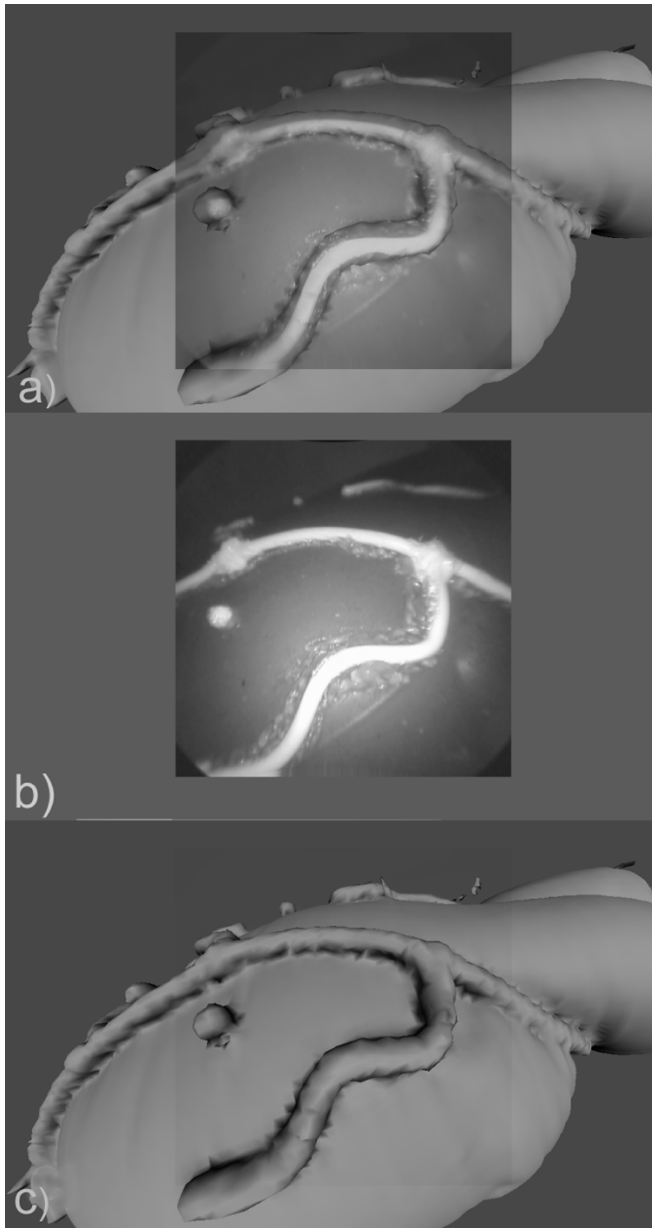


Fig. 4. The rendering pane contains a superposition of the optical-endoscope view and the virtual view. (a) an example image of the rendering pane showing the fused virtual and real endoscopic views of the heart phantom, (b) the endoscope view, and (c) the virtual view.

under Windows 2000, on an AMD Athlon MP 1800+ computer, together with ASUS V8200 Series v31.40b display adapter. A Matrox Corona video digitizer (Matrox, Dorval QC) was used to digitize the analog signal from the video camera attached to the endoscope.

The program displays the endoscope view on a semi-transparent overlay on top of the virtual image of the heart phantom, as described above and shown in Fig. 4. The relative pose of the endoscope with respect to the phantom, T_{EP} from (1), is calculated based on the information from the tracking probes attached to the endoscope and the phantom. The 4×4 pose matrices returned by the Polaris are updated every 20 ms.

To achieve robust estimates of the endoscope pose when stationary, the position, the view-up and the focal point vectors

were averaged (up to 80 consecutive sets), followed by orthogonalization of the rotation submatrix. The averaging is performed only in the absence of detected motion between the endoscope and the phantom case, and does not compromise the dynamic tracking of the endoscope.

The virtual 3-D scene was constructed as described in Section II-A, employing the distortion correction method discussed above. The distortion correction is performed in near-real time (14 Hz) using the precomputed coefficients a_1 and a_2 (3) by warping a 32×32 grid encompassing the field of view of the endoscope according to (3), and linearly interpolating the positions of the intervening pixels. This process was implemented using the texture mapping capabilities in the graphics hardware [18]. Finally, the time-synchronized isosurfaces of the preoperative images of the beating heart phantom were added. These surfaces were generated using the marching cubes algorithm [19], and refined with the warping algorithm described below. The isosurfaces of the heart phantom were transformed from the 3-D coordinates of the CT scanner to 3-D real-space coordinates using the transform determined by registering fiducials in the CT scan with their actual positions.

3) *Animation of the Virtual Beating Heart Phantom:* The heart surface and the coronary arteries are well represented when the phantom is imaged in a stationary state. It is possible to generate a surface for each image in the four-dimensional CT dataset using the approach described for the reference image. However, increasing image artifacts from phantom motion in midcycle make these surfaces exceptionally rough, and as a result, not suitable either for direct rendering or for surface-based registration algorithms. This phenomenon is illustrated by comparing the image presented in Fig. 5(a) (made during rapid motion), with that shown in Fig. 5(b) (the heart stationary during imaging). Our new approach [12], [20], is able to extract reliable motion information even in the presence of motion-induced streaking and blurring. While the dynamic images exhibit many such artifacts, they nevertheless contain information describing the motion of the heart throughout the cardiac cycle. This information is contained in the 3-D vector fields generated by the nonrigid registration algorithm, which are computed as the quasi-static (end diastole) image is successively registered to other images throughout the cardiac cycle. Because the registration algorithm is multi-resolution, and the finest resolution cell is 10 mm^3 (approximately $10 \times 10 \times 8$ voxels) local artifact-induced perturbations in the volume are largely filtered out.

These nonrigid transformations were then applied to the static surface model generated from the end diastole CT image to deform it appropriately for the remaining time frames throughout the cardiac cycle. This registration approach ensures that there is only one mapping between the reference image and the frame of interest, to avoid the problem of potential error accumulation that could occur if the reference frame was deformed serially from one frame to the next. It also avoids the appearance of artifacts generated by rapidly moving structures during the CT scan.

The algorithm used to derive the nonrigid transformation between two image frames is an entirely automatic, intensity-based method employing a linearly interpolated free-form deformation (FFD) strategy. Within this framework, both the

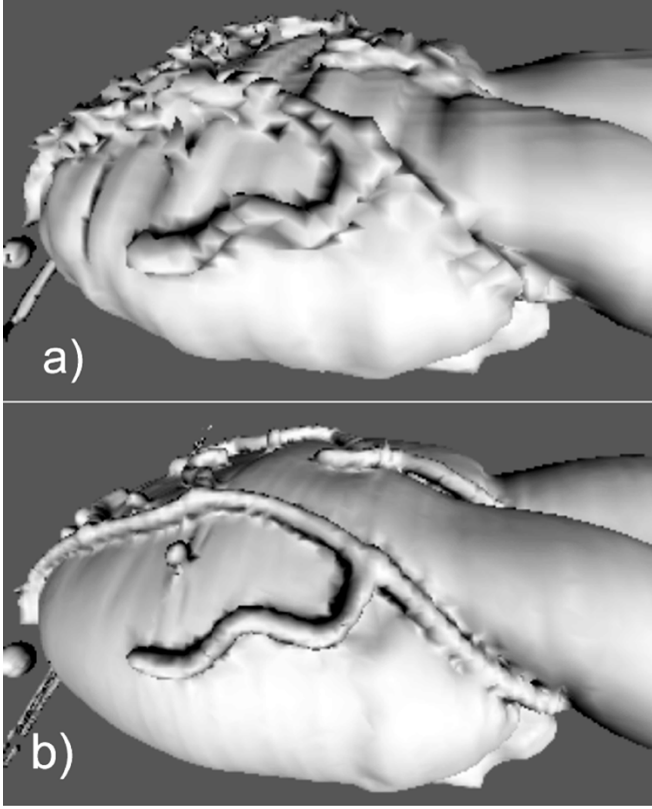


Fig. 5. The virtual model of the phantom shown at 0.54 s after the beginning of the cardiac cycle: (a) as reconstructed from the directly acquired CT volumetric images, and (b) after applying the warping algorithm to the quasi-static baseline image to map it to the same cardiac phase as shown in (a). Note the removal of motion-induced artifacts in (b).

source and target images are overlaid with a structured grid of control points (nodes), where each node is associated with a small 3-D surrounding region. Following local 3-D searches to obtain the optimum match between regions of the source and target images, a nonzero displacement vector is defined for each node to model the local deformation (deformation of points between nodes is computed via linear interpolation). The actual lengths and directions of the individual vectors are chosen using a downhill simplex optimization algorithm [21], with each node being treated independently in sequence. To find the vector \vec{r} at a given node, the optimization algorithm minimizes the following cost function:

$$\text{cost}(\vec{r}) = \sum_V |I_{\text{SO}}(\vec{r}) - I_{\text{TA}}| + \alpha \cdot \text{BE}(\vec{r}) \quad (4)$$

where $I_{\text{SO}}(\vec{r})$ are the intensities in the source image data following a translation by \vec{r} , I_{TA} are the original intensities of the target image, V is the volume of the source and target data under consideration (usually equal to grid spacing), and BE is the bending energy associated with the current choice of vector (the weight of which is controlled by the user defined parameter α). The first term of (4) is the sum of absolute difference of intensities (SAD), a simple and computationally efficient similarity measure used previously in serial image registration [22]. The second term is the 3-D equivalent of the thin-plate spline bending energy, responsible for ensuring that

the calculated transformation is smooth. The definition of the “bending energy” in 3-D is [23]

$$\begin{aligned} \text{BE}(\vec{r}) &\propto \int_0^X \int_0^Y \int_0^Z \left[\left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial^2 x} \right)^2 + \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial^2 y} \right)^2 \right. \\ &\quad + \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial^2 z} \right)^2 + 2 \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial x \partial y} \right)^2 \\ &\quad + 2 \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial x \partial z} \right)^2 + 2 \left(\frac{\partial^2 \text{FFD}(\vec{r})}{\partial y \partial z} \right)^2 \left. \right] \\ &\quad \times dx dy dz \end{aligned} \quad (5)$$

for a FFD grid defined from 0 to X , 0 to Y , and 0 to Z . According to the above formulation, the BE for a vector \vec{r} at a given node is calculated by inserting that vector into the FFD grid from the *previous* iteration (for the first iteration the FFD grid is composed of zero-length vectors). In our implementation, however, we always assume that the previous FFD grid was all zeros. This not only speeds up the calculation (only nearest neighbors of the current node must be considered), but also ensures that the solution is not affected by possible errors in previous iterations. While this change makes the FFD grid globally less smooth, visual inspection of results shows that the BE term can be properly balanced with the SAD term by setting α to 19.0. To avoid local minima, we employ a multi-resolution grid approach when registering two images. The optimization takes place at grid resolutions of approximately 30^3 mm^3 , 20^3 mm^3 , and finally 10^3 mm^3 .

The result is a high-quality dynamic image volume [Fig. 5(b)] that does not display the artifacts [shown in Fig. 5(a)] that would be present if the originally acquired volumes were displayed in sequence.

4) Camera Calibration: Here, we describe the procedure used to establish transform T_{CE} (extrinsic camera calibration) from (1) and the camera focal length (intrinsic camera calibration). Although transform T_{CE} may be established based on geometrical considerations, it was refined using the following stand-alone procedure (software written in Microsoft Visual C++ 6.0, utilizing VTK 4.1 classes). The procedure consists of nonlinear least-squares minimization of an error function dist , defined as the cumulative screen distance between the endoscopic view and the calculated screen positions of markers of known real-space location (retroreflective tracking balls, tracked by the Polaris tracking system)

$$\text{dist} = \sum_i \left| (u_i^{\text{screen}} - u_i^{\text{calc}})^2 + (v_i^{\text{screen}} - v_i^{\text{calc}})^2 \right|^{1/2}. \quad (6)$$

The summation over the markers i is based on multiple snapshots of four markers collected from various endoscope poses. In (6), u and v are the screen coordinates of a marker in a snapshot, and the superscripts screen and calc denote the endoscopic view and the calculated coordinates, respectively. The screen coordinates were calculated from the known real-space coordinates using

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = A_{3 \times 4} (T_{4 \times 4}^R)^{-1} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (7)$$

where (x, y, z) are the real-space coordinates of the marker; λ is the scaling factor, the endoscope pose transform $T_{4 \times 4}^R$ was defined in (1) and was saved with each snapshot, and $A_{3 \times 4}$ is the perspective projection transform matrix

$$A_{3 \times 4} = \begin{pmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (8)$$

Here, f is the camera focal length, while u_0 and v_0 denote the screen coordinates of the center of the rendering pane. Minimization of the error function $dist$ yields optimum values of the translation components and the Euler angles defining transform T_{CE} in (1), as well as the camera focal length f .

To ensure both the accuracy of this procedure and its independence on the pose of the endoscope with respect to the markers (and, consequently, the heart phantom), a large number of these snapshots were collected from a range of distances and viewing angles.

5) *Three-Dimensional Validation Software*: In the validation published in [16], we used a single snapshot of the rendering pane to find the error between the endoscopic and the virtual location of a marker. Due to the 2-D nature of the endoscopic view displayed in the snapshot, only the view-plane projection of the error could be calculated. As the magnitude of a projection depends on its direction, these errors did not necessarily provide an accurate estimate of the errors for other viewing directions. Here, we use approximately orthogonal pairs of endoscope views of the rendering pane to extract the true 3-D coordinates of a marker. This was possible upon utilizing the pose transform of the endoscope obtained with the tracking system, using the algorithm described below.

The unknown real-space 3-D coordinates (x, y, z) of a marker are related to the screen coordinates (u, v) of the marker in each endoscopic view through (7) and (8). Writing (7) for two different endoscope viewing directions, combined with the knowledge of matrices $A_{3 \times 4}$ and $T_{4 \times 4}^R$, allows the 3-D coordinates (x, y, z) to be calculated in the following manner: elimination of λ leaves two first-order equations of type: $ax + by + cz + d = 0$ for each snapshot of the investigated marker, while combining two views generates the total of four equations, which can be rewritten in a matrix form

$$C_{4 \times 1} = D_{4 \times 3} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (9)$$

The matrices $C_{4 \times 1}$ and $D_{4 \times 3}$ are constructed from the coefficients a, b, c, d in the four equations described above. This over-determined system can be solved for (x, y, z) using linear least squares minimization.

Although the 3-D coordinates of a marker can be established using (9) constructed out of only two snapshots, the procedure can be extended by combining N snapshots and writing (7) N -times to generate an over-determined matrix equation similar to (9), with $C_{4 \times 1}$ and $D_{4 \times 3}$ replaced by $C_{2N \times 1}$ and $D_{2N \times 3}$

$$C_{2N \times 1} = D_{2N \times 3} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (10)$$

Similarly, (x, y, z) can be calculated through linear least squares minimization of (10). Use of multiple snapshots improves the localization accuracy of the marker coordinates.

III. RESULTS

This section contains analysis of the performance of selected components of the system, and validation of the overall system under static and dynamic conditions.

A. Endoscope Barrel Distortion

The endoscope image of the calibration grid after removal of the barrel distortion was compared to the original rectilinear array of points. The average error (expressed as a percentage of the field of view) was 0.34%, which is comparable to the value of 0.24% reported by Asari [24] in his similar procedure.

B. Camera Calibration

In order to establish transform T_{CE} (1) using the minimization routine described in Section II-C4), we collected 17 snapshots of 6 markers for the total of 54 data points. The screen coordinates of the markers were determined manually. The choice of poses of the endoscope was arbitrary, but was designed to cover situations anticipated during the experiment, i.e., the endoscope positioned close to and far from the heart phantom, with a wide range of endoscope viewing directions. Because establishing the transform T_{CE} constitutes the last step required in the construction of the system in the case of a stationary heart phantom, the errors found after completion of this procedure are equivalent to the errors found in the static validation, as described below.

C. Static Validation

To establish a baseline against which to compare the errors measured under dynamic conditions, we determined the static case intrinsic accuracy of the registration between the virtual model and the model based on optical images from the endoscope. To ensure a robust registration result we acquired 43 randomly posed images of the rendering pane containing a Polaris marker placed rigidly on the tracked phantom casing. (A rigid model was employed for this test to ensure that the results were not affected by the elastic nature of the beating-heart model.) Each of these images contained both the endoscopic view of the marker and its virtual representation (similar to Fig. 4(a) with the heart phantom replaced by the marker). The screen coordinates of the markers were determined separately for the endoscopic and the virtual views by identifying their centroids manually on each of the 43 images. The 3-D locations of the marker were calculated by solving (10) independently for the endoscopic and the virtual views, resulting in a 3-D displacement of 0.24 mm.

D. Dynamic Validation

In this step we validated the precision of the overlap between the dynamic optical and virtual images of the beating heart phantom. The snapshots of the rendering window were saved and analyzed for all 14 phases of the cardiac cycle. The snapshots of selected phases of the cardiac cycle are shown in Fig. 6.

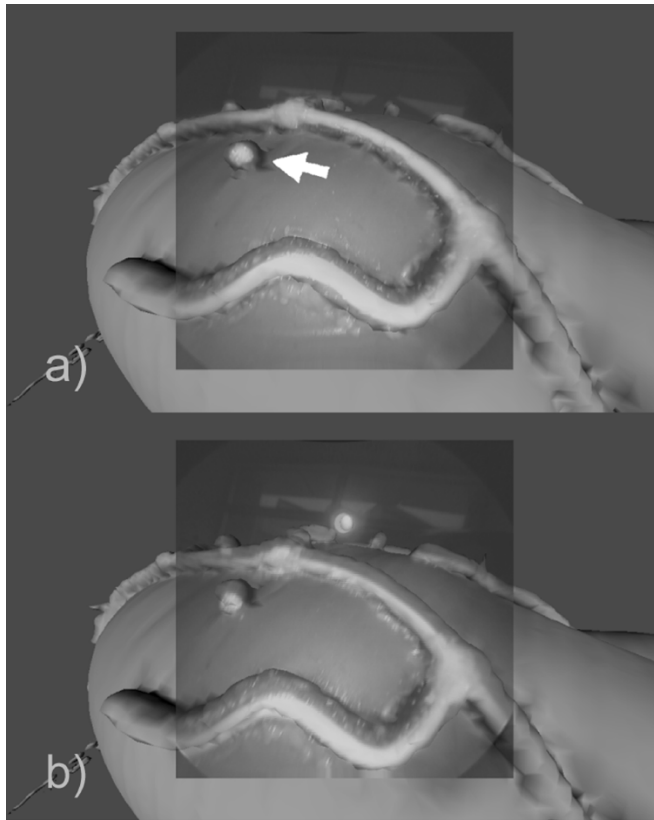


Fig. 6. Selected snapshots from the cardiac cycle: at the beginning of the cardiac cycle (a) and 0.48 s later (b). The (darker) rectangular area at the center of each snapshot marks the view of the optical endoscope. The white arrow points at the marker used for validation.

The opacity of the endoscopic view (displayed as a rectangle in the images below) was set to 65% to allow simultaneous viewing of the optical and virtual images. The overlap of the optical and the virtual images is achievable at all phases of the cardiac cycle with the coronary vessels and the fiducial markers being represented appropriately in all phases. The high accuracy of the overlay is partially attributable to the use of the warping algorithm to create the sequence of images from the clean end-diastole image, rather than using each of the raw dynamic images in sequence.

The errors in time synchronization and the spatial shift between the optical and virtual images were quantified based on the real-space 3-D coordinates of a fiducial marker placed on the surface of the phantom (marked with an arrow in Fig. 6(a)). The time-dependent 3-D coordinates of the marker were calculated from dual-view (i.e., collected from two different endoscope viewing directions) snapshot pairs collected for 14 phases of the cycle using terminology outlined in Section II-C5 (using tracked Polaris balls in the calibration stage). Since the endoscope was not intrinsically stereoscopic, these snapshot pairs were taken from two separate cardiac cycles. The time dependent overlap of the optical and the virtual 3-D positions of the marker were characterized along the principal direction of marker motion established by principal component analysis (PCA), for each of the optically derived 3-D positions of the marker calculated at each of the 14 phases of the cardiac cycle. The optical and the virtual position of the marker along the principal direction of

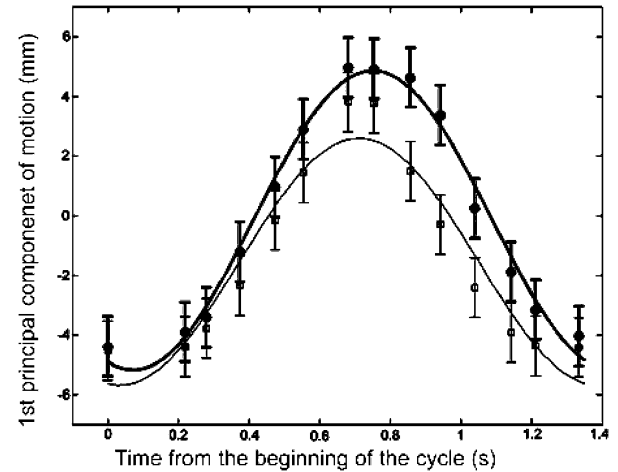


Fig. 7. The position of a reference marker on the cardiac surface throughout the cardiac cycle: ●—optical image, □—virtual model, along with the analytically fitted continuous trajectories.

TABLE I
FITTED PARAMETERS OF THE COSINE FUNCTION (11) MODELING THE TIME OSCILLATION OF THE OPTICAL AND THE VIRTUAL MARKER ALONG THE PRINCIPAL DIRECTION OF MOTION

	Optical	Virtual
Time lag t_0 (s)	0.07 ± 0.02	0.03 ± 0.04
Amplitude A (mm)	-5.0 ± 0.4	-4.1 ± 0.8
Spatial shift B (mm)	-0.2 ± 0.3	-1.5 ± 0.5

motion versus time elapsed from the beginning of the cardiac cycle are plotted in Fig. 7. The vertical error bars correspond to the errors of 1.5 mm in the calculation of the 3-D marker position from a pair of endoscopic snapshots (determined in a separate validation routine applied to additional markers of coordinates known *a priori*). It should be noted that the optical positions of the marker carry a temporal error associated with the lack of synchronization between switching the frames by the endoscope camera and the cardiac cycle. Since the endoscope snapshot displays the marker with up to a 0.07 s delay (at 14 frames per second), the plot of the optical points may be shifted by a similar amount. To partially compensate for this delay, the program displays the virtual image calculated at the time instant delayed by the mean delay expected (0.035 s).

The average spatial error along the first principal component between the optical and the virtual data points is 1.4 ± 1.1 mm. To compute the phase lag between the optical and the virtual data points, cosine functions with the period $T = 1.357$ s (corresponding to a heart rate of 44.2 bpm) were fitted independently to the optical and the virtual data as follows:

$$y = A \cos(2\pi(t - t_0)/T) + B. \quad (11)$$

The fitted values of the parameters used in (11) are listed in Table I. The time lag t_0 (measured from the ECG synchronization signal) of displaying the optical (0.07 ± 0.02 s) or the virtual sequence (0.03 ± 0.04 s) is small compared to the interval between consecutive heart phases of 0.11 s. The phase difference between the two cosines (0.04 ± 0.05 s that constitutes 0.7% of

the motion period) represents the error in the synchronization of the display of the optical and the virtual images.

The amplitude A of the virtual cosine is smaller by 18% with respect to the amplitude of the optical cosine. This relatively large difference is partially caused by limitations in the reconstruction of the virtual images from the dynamic CT scan, since each image is integrated over 7% of the cycle, and the resulting averaging of the marker position “flattens” the peaks. The inaccuracies of the CT images described above translate into a 5% reduction of the amplitude A in the virtual cosine. This can be attributed to several sources, including the limited number of the data points (14 frames per cycle), errors caused by smearing of features in the CT images of the moving fiducial, and the smoothing effect of the second term of (4).

The absolute value of the difference in the spatial shift B between the optical and the virtual cosines is small compared to the amplitude A ($1.3 \text{ mm} \ll 5.0 \text{ mm}$). Nevertheless, the optical and the virtual curves are clearly shifted. This spatial shift B is a consequence several potential error sources: phantom deformability leading to small variations in positioning of the phantom in its casing, the effect of tracking errors (0.15–0.3 mm), non-linear distortions in the CT images, errors in the extraction of the position of the tracking marker at the heart phantom casing, and accumulation of error due to the many steps required to construct the virtual scene.

IV. DISCUSSION

The dynamic validation results presented in this paper are more general than those described earlier [16]. Here, we analyze the 3-D errors, while [16] considered 2-D errors. As projections, the 2-D results strongly depend on the selected viewing direction of the endoscope, and do not measure the maximum errors. Furthermore, in spite of using the same equation (11) in this paper to model the changes of the optical and the virtual position of the marker, the fitted coefficients differ, since they pertain to the 2-D or 3-D analysis of motion respectively. Also the fitted values of B differ between the two papers, since the origin along the principal direction of motion in [16] was arbitrary, while here it is set to the mean value of the optical data points. Only the difference of the values of B between the virtual and the optical data is meaningful.

The accuracy of the overlay achieved with the phantom system (0.24 mm in the static case and $1.4 \pm 1.1 \text{ mm}$ in the dynamic case) is close to our target value of 1 mm. Nevertheless, to reduce the small spatial discrepancies in the display of the virtual image with respect to the endoscopic image, we plan to improve the virtual model of the beating heart (phantom). The images currently used in the animation were constructed based on time-integrated CT images, and suffer from “flattening” of the peak motions. Instead, we plan to extrapolate the vector fields used in the warping algorithm (employed to construct the virtual model) based on preoperative and intraoperative data relating to the heart motion obtained from trans-esophageal ultrasound.

This research represents only a preliminary stage of a project that aims to dynamically register a virtual model of the beating heart to that of a patient, and provides a “proof of concept” for

fusing dynamic endoscopic images and the virtual model of the beating heart phantom. The next phase of this project involves validation using an *in vivo* porcine model, as part of an ongoing study by our colleagues to develop minimally invasive cardiac procedures.

While we have demonstrated the applicability of this approach in a laboratory setting, its adaptation to the clinical environment will present a number of challenges. In a clinical situation involving closed-chest surgical procedures, optical tracking of fixed landmarks cannot be used and an alternative means of registering the heart to thoracic landmarks must be employed. Preliminary experience with an ultrasound-based registration of the rib cage to its virtual model suggests that we can employ this approach to register the patient to the model. We also plan to use intraoperative angiogram data to establish the position of the heart within the thoracic cavity, in conjunction with image-based tracking as proposed by Mori *et al.* [25].

The algorithm used to construct the virtual model of the beating heart phantom can be applied to the heart *in vivo* with minor modifications. In the experiments reported here we warped the image of the stationary phantom, but in clinical practice the quasi-stationary image obtained during end diastole would be used.

The phantom motion (inflating/deflating movement following a cosine time dependence) is determined by the characteristics of the commercially available heart/thorax phantom we are using, and does not accurately reflect the dynamics of a real heart. The more complex time dependence of the motion of the human heart [26] will necessitate more reconstructed phases of the heart motion, ideally defined at nonuniformly spaced intervals.

Our prototype system operates at 14 fps due to the limitation of the graphics subsystem employed in this project. However, current graphics systems will allow processing and display of acceptable images at 20–30 fps.

The procedure for removing the lens barrel distortion was calibrated prior to finding the intrinsic and extrinsic camera calibration. In fact, it would be possible to combine the two procedures into one step, by the addition of the lens barrel distortion parameters to the minimization routine used to establish the camera calibration. We chose to perform these two steps separately to improve robustness of the procedure by reduction of the number of fitted parameters within a single step.

V. CONCLUSION

This paper comprises a component of a larger project aimed at building a virtual representation of the thoracic cavity and the beating heart, and registering it to a patient under operating room conditions. In addition to enhancing planning and guidance for coronary artery bypass procedures, the availability of such a tool would enable many other minimally invasive cardiac interventions, which are currently performed using open-chest or open-heart surgery with a heart-lung bypass machine. Within such an environment, preoperative planning can be based on actual anatomy rather than on estimates of the positions of

anatomical targets, permitting both the anticipation and prevention of potential complications. By mapping function (acquired from tracked electrophysiological probes) to the anatomy, interventional accuracy and precision for electro-ablative treatments would increase. We believe the use of such an image-based planning and guidance system will allow treatment to be delivered with lower risk, and safer and more effective interventions may be offered to a larger patient population.

In general, accurate anatomical representation of cardiac anatomy and physiology will allow new targets to be approached and treated, particularly intracardial sites that currently are beyond the reach of surgical intervention on the beating heart.

ACKNOWLEDGMENT

The authors would like to thank Dr. D. Gobbi and Dr. M. Drangova, R. Gupta, C. Wedlake, A. So, C. Norley, and C. Nikolov for assistance in various phases of this work.

REFERENCES

- [1] W. Reynolds Jr, "The first laparoscopic cholecystectomy," *J. Soc. Laparoendosc. Surg.*, vol. 5, pp. 89–94, 2001.
- [2] E. R. Stephenson Jr, S. Sankholkar, C. T. Ducko, and R. J. Damiano Jr, "Successful endoscopic coronary artery bypass grafting: An acute large animal trial," *J. Thorac. Cardiovasc. Surg.*, vol. 116, pp. 1071–1073, 1998.
- [3] D. Loulmet, A. Carpentier, N. d'Attellis, A. Berrebi, C. Cardon, O. Ponzio, B. Aupècle, and J. Y. M. Relland, "Endoscopic coronary artery bypass grafting with the aid of robotic assisted instruments," *J. Thorac. Cardiovasc. Surg.*, vol. 118, pp. 4–10, 1999.
- [4] F. W. Mohr, V. Falk, A. Diegeler, and R. Autschbach, "Computer-enhanced coronary artery bypass surgery," *J. Thorac Cardiovasc Surg*, vol. 117, pp. 1212–1215, 1999.
- [5] B. Kiaii, W. D. Boyd, R. Rayman, W. B. Dobkowski, S. Ganapathy, G. Jablonsky, and R. J. Novick, "Robot-assisted computer enhanced closed-chest coronary surgery: Preliminary experience using a harmonic scalpel and Zeus," *Heart Surg. Forum*, vol. 3, no. 3, pp. 194–197, 2000.
- [6] G. Czibik, G. D'Ancona, H. W. Donias, and H. L. Karamanoukian, "Robotic cardiac surgery: Present and future applications," *J. Cardiothorac. Vasc. Anesth.*, vol. 16, no. 4, pp. 495–501, 2002.
- [7] D. J. Vining, K. Liu, R. H. Choplin, and E. F. Haponik, "Virtual bronchoscopy: Relationships of virtual reality endobronchial simulations to actual bronchoscopic findings," *Chest*, vol. 109, pp. 549–553, 1996.
- [8] H. Gulbins, H. Reichensperner, C. Becker, D. H. Boehm, A. Knez, C. Schmitz, R. Bruening, R. Haberl, and B. Reichart, "Preoperative 3D-reconstructions of ultrafast-CT images for the planning of minimally invasive direct coronary artery bypass operation (MIDCAB)," *Heart Surg. Forum*, vol. 1, pp. 111–115, 1998.
- [9] A. M. Chiu, D. Dey, M. Drangova, W. D. Boyd, and T. M. Peters, "3-D image guidance for minimally invasive robotic coronary artery bypass," *Heart Surg. Forum*, vol. 3, no. 3, pp. 224–231, 2000.
- [10] G. Lehmann, A. Chiu, D. Gobbi, Y. Starreveld, D. Boyd, M. Drangova, and T. Peters, "Toward dynamic planning and guidance of minimally invasive robotic cardiac bypass surgical procedures," in *Lecture Notes in Computer Science*, W. Niessen and M. Viergever, Eds. Berlin, Germany: Springer-Verlag, vol. 2208, MICCAI 2001, pp. 368–375.
- [11] G. Lehmann, D. Habets, D. W. Holdsworth, T. M. Peters, and M. Drangova, "Simulation of intra-operative 3-D coronary angiography for enhanced minimally invasive robotic cardiac intervention," in *Lecture Notes in Computer Science*, T. Dohi and R. Kikinis, Eds. Berlin, Germany: Springer-Verlag, vol. 2489, MICCAI 2002(2), pp. 268–275.
- [12] M. Wierzbicki and T. Peters, "Determining epicardial surface motion using elastic registration: Toward virtual reality guidance of minimally invasive cardiac intervention," in R. Ellis and T. Peters, Eds. Berlin, Germany: Springer-Verlag, vol. 2878, MICCAI 2003(1), pp. 722–729.
- [13] D. Dey, D. G. Gobbi, P. J. Slomka, K. J. Surry, and T. M. Peters, "Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas," *IEEE Trans. Med. Imag.*, vol. 21, no. 1, pp. 23–30, Jan. 2002.
- [14] R. Shahidi, M. R. Bax, C. R. Maurer, J. A. Johnson, E. P. Wilkinson, B. Wang, J. B. West, M. J. Citardi, K. H. Manwaring, and R. Khadem, "Implementation, calibration and accuracy testing of an image-enhanced endoscopy system," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1524–1535, Dec. 2002.
- [15] F. Mourgues, T. Vieville, V. Falk, and È. Coste-Manière, "Interactive guidance by image overlay in robot assisted coronary artery bypass," in *Lecture Notes in Computer Science*, R. Ellis and T. Peters, Eds. Berlin, Germany: Springer-Verlag, vol. 2878, MICCAI 2003(1), pp. 173–181.
- [16] S. Szpala, G. Guiraudon, and T. Peters, "Cardiac endoscopy enhanced by dynamic organ modeling for minimally invasive surgery guidance," in *Lecture Notes in Computer Science*, R. Ellis and T. Peters, Eds. Berlin, Germany: Springer-Verlag, vol. 2878, MICCAI 2003(1), pp. 499–506.
- [17] L. Adhami and È Coste-Manière, "A versatile system for computer integrated mini-invasive robotic surgery," in T. Dohi and R. Kikinis, Eds. Berlin, Germany: Springer-Verlag, vol. 2488, MICCAI 2002, pp. 272–281.
- [18] J. Neider, T. Davis, and M. Woo, *OpenGL Programming Guide*. Menlo Park, CA: Addison-Wesley, 1993.
- [19] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3-D surface construction algorithm," *ACM Comp. Graph.*, vol. 21, no. 4, pp. 163–169, 1987.
- [20] M. Wierzbicki, M. Drangova, G. Guiraudon, and T. Peters, "Validation of dynamic heart models obtained using nonlinear registration for virtual reality training, planning, and guidance of minimally invasive cardiac surgeries," *Med. Image Anal.*, vol. 8, pp. 387–401, 2004.
- [21] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–313, 1965.
- [22] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," *Phys. Med. Biol.*, vol. 46, pp. R1–R45, 2001.
- [23] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Trans. Med. Imag.*, vol. 18, no. 8, pp. 712–721, Aug. 1999.
- [24] K. V. Asari, S. Kumar, and D. Radhakrishnan, "A new approach for non-linear distortion correction in endoscopic images based on least squares estimation," *IEEE Trans. Med. Imag.*, vol. 18, no. 4, pp. 345–354, Apr. 1999.
- [25] K. Mori, D. Deguchi, J. Sugiyama, Y. Suenaga, J. Toriwaki, C. R. Maurer Jr, H. Takabatake, and H. Natori, "Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images," *Med. Image Anal.*, vol. 6, no. 3, pp. 321–336, 2002.
- [26] G. D. Buckberg, C. Clemente, J. L. Cox, H. C. Coghlan, M. Castella, F. Torrent-Guas, and M. Gharib, "The structure and function of the helical heart and its buttress wrapping. IV. Concepts of dynamic function from the normal macroscopic helical structure," *Sem. Thorac. Cardiovasc. Surg.*, vol. 13, no. 4, pp. 342–357, 2001.