

# Probabilistic Sparse Matching for Robust 3D/3D Fusion in Minimally Invasive Surgery

Dominik Neumann, *Student Member, IEEE*, Saša Grbić, Matthias John, Nassir Navab, *Member, IEEE*, Joachim Hornegger, *Member, IEEE*, Razvan Ionasec

**Abstract**—Classical surgery is being overtaken by minimally invasive and transcatheter procedures. As there is no direct view or access to the affected anatomy, advanced imaging techniques such as 3D C-arm CT and C-arm fluoroscopy are routinely used in clinical practice for intraoperative guidance. However, due to constraints regarding acquisition time and device configuration, intraoperative modalities have limited soft tissue image quality and reliable assessment of the cardiac anatomy can only be made by injecting contrast agent, which is harmful to the patient and requires complex acquisition protocols. We propose a probabilistic sparse matching approach to fuse high-quality preoperative CT images and non-gated, non-contrast intraoperative C-arm CT images by utilizing robust machine learning and numerical optimization techniques. Thus, high-quality patient-specific models can be extracted from the preoperative CT and mapped to the intraoperative imaging environment to guide minimally invasive procedures. Extensive quantitative experiments on 95 clinical datasets demonstrate that our model-based fusion approach has an average execution time of 1.56 seconds, while the accuracy of 5.48 mm between the anchor anatomy in both images lies within expert user confidence intervals. In direct comparison with image-to-image registration based on an open-source state-of-the-art medical imaging library and a recently proposed quasi-global, knowledge-driven multi-modal fusion approach for thoracic-abdominal images, our model-based method exhibits superior performance in terms of registration accuracy and robustness with respect to both target anatomy and anchor anatomy alignment errors.

**Index Terms**—Model-based Cardiac Image Registration, Procedure Guidance, Anatomical Overlay, Computed Tomography.

## I. INTRODUCTION

Fluoroscopy guided cardiac interventions such as endovascular stenting, atrial ablation, closure of atrial/ventricular septal defects and transcatheter valve repair or replacement are becoming increasingly common [1]. Compared to conventional open-heart surgeries, these procedures tend to be less invasive, reduce procedural morbidity, mortality, and intervention cost, while accelerating patient recovery. For high-risk groups or

patients previously declared inoperable, minimally invasive surgery is the only treatment option [2], [3]. However, without direct access or view to the affected anatomy, advanced imaging is required to secure a safe and effective execution of the procedure.

Overlays of 3D anatomical structures based on preoperative data can provide valuable information for interventional navigation and guidance when displayed on live fluoroscopy. High-quality preoperative 3D data is routinely acquired for diagnostic and planning purposes by means of computed tomography, magnetic resonance imaging or echocardiography. However, direct 3D preoperative to 2D fluoroscopy registration [4] is difficult to solve, especially within the intraoperative setup that makes user interaction difficult and time-consuming processing undesirable. The authors of [5] propose the use of preoperative CT images and extract models which are registered with intraoperative MR and echocardiography images. Major limitations are the required optical tracking equipment and the semi-automatic delineation of the mitral annulus anatomy. Other methods rely on fiducial markers to achieve 2D/3D registration. A real-time approach to fuse images from transesophageal echocardiography (TEE) and fluoroscopy by placing markers on the TEE probe is proposed in [6].

C-arm CT is emerging as a novel imaging modality that can acquire 3D CT-like volumes directly in the operating room in the same coordinate system as 2D fluoroscopy images, which overcomes the need for 2D/3D registration. For most procedures, the patients are at an advanced age and the added radiation compared to fluoroscopy is not a major concern. Instead, a safe and effective execution of the procedure is the dominating factor [7]. Some methods work directly on the C-arm CT images [8] to extract patient-specific models and overlays for procedure guidance. However, acquiring high-quality, motion compensated and contrast-enhanced C-arm CT images is challenging and not feasible for all patients, in particular for patients with renal function deficiencies. Instead, much simpler protocols would be preferred by clinicians, where non-ECG-gated non-contrast C-arm CT volumes are acquired. Currently, clinicians align the preoperative images and the intraoperative C-arm CT image using manual registration tools. This process can be automated by using registration methods.

Surveys of registration methods for alignment of 3D medical images can be found in [9]–[13]. Multi-modal 3D/3D registration algorithms can be utilized to automate the process of aligning preoperative scans from arbitrary imaging modalities with the C-arm CT image. In [14]–[16], mutual information is used to cope with intensity inconsistencies between dif-

D. Neumann and J. Hornegger are with Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, 90158 Erlangen, Germany. E-mail: dominik.neumann@fau.de; joachim.hornegger@fau.de.

D. Neumann and S. Grbić are with Imaging and Computer Vision, Siemens Corporate Technology, Princeton, NJ 08540, USA. E-mail: dominik.neumann@siemens.com; sasa.grbic@siemens.com.

S. Grbić and N. Navab are with Computer Aided Medical Procedures, Technische Universität München, 85748 Garching, Germany.

M. John and R. Ionasec are with Siemens AG, Healthcare Sector, 91294 Forchheim, Germany.

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

ferent modalities. The authors of [17] propose an atlas-based approach to track the myocardium and ventricles from MR data. However, these methods are computationally expensive, and without appropriate guidance of a shape prior they are likely to converge into local minima. To overcome the latter limitation, the authors of [18] propose a quasi-global approach, where, considering prior knowledge on the problem domain (thoracic-abdominal CT and C-arm CT volumes), multiple starting points in parameter space are selected to approximate global optimization behavior.

We propose an extension to our previous method [19], a novel approach to fuse 3D preoperative high-quality anatomical information with live 2D intraoperative imaging using non-contrasted 3D C-arm CT as a proxy. Robust and fully-automatic learning-based methods are employed to extract patient-specific models of both target and anchor anatomies from CT. Anchor anatomies have correspondences in the pre- and intraoperative images, while target anatomies are not visible in the intraoperative image, but essential to the interventional procedure. A novel sparse matching approach is employed to align the preoperative anchor anatomies to the intraoperative setting. Sparse matching allows for high computational efficiency. In contrast to traditional image-to-image registration methods where all  $n$  voxels (or a large subset) in the images need to be sampled per iteration, our method requires only few read operations on a probabilistic 3D map. The voxels of interest are determined by a comparatively small number  $n' \ll n$  of sparsely distributed points on a geometrical pericardium model. Data and model uncertainties are learned and exploited during the matching process. Our method is able to cope with artifacts in the intraoperative images, partially visible anatomical structures, and does not require contrast agent in the intraoperative image.

The intended use case of our approach is navigation in minimally invasive surgery, including transcatheter aortic valve implantation (TAVI), which is the focus of this study. Once the preoperative data is fused to the intraoperative imaging environment, high-quality preoperative anatomical overlays can be displayed on fluoroscopy in order to facilitate safe navigation of a catheter to the aortic valve. The final valve deployment is not in the scope of this work, as the interventional cardiologists will not rely solely on static models extracted from preoperative data. Instead, live fluoroscopy is commonly used. Our previous work [19] is extended as follows:

- 1) The scheme used to incorporate prior knowledge is refined using a probabilistic sampling function instead of vertex-wise prior weights, leading to improved accuracy.
- 2) The dataset used for evaluation of our approach is extended from 88 to 95 pairs of real clinical volumes.
- 3) Quantitative and qualitative evaluation is extended by comparing to a state-of-the-art image-to-image registration technique using masked images, as well as a recently proposed quasi-global, knowledge-driven fusion approach for thoracic-abdominal images [18].
- 4) Computation time is reduced towards real-time fusion from 2.93 s to 1.56 s per registration on an off-the-shelf consumer computer through code optimization, and a detailed runtime analysis is provided.

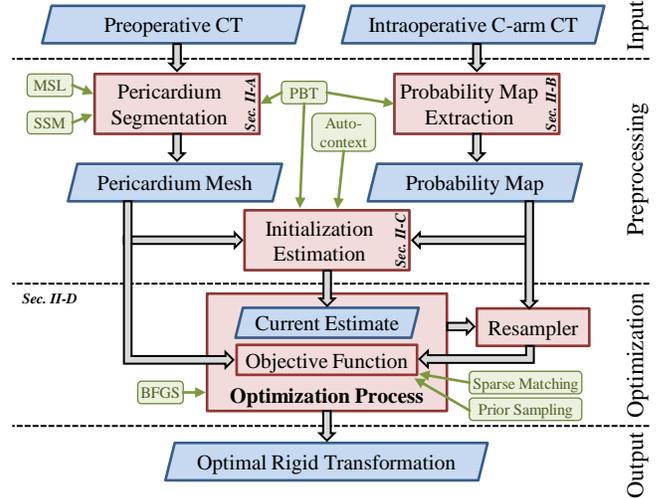


Fig. 1. Flowchart of our 3D/3D fusion method. Blue boxes represent data, while red boxes show processing steps and methods. Grey arrows indicate the general workflow, and green boxes and arrows highlight various algorithms and concepts used in specific parts of our approach.

## II. METHOD

Our fully-automatic method fuses preoperative CT (moving image  $\mathcal{M}$ ) with intraoperative C-arm CT volumes (fixed image  $\mathcal{F}$ ), such that a target anatomy, e.g. the aortic valve, is aligned. The process is based on a personalized anchor anatomical model  $A_{\mathcal{M}}$  of the pericardium extracted from  $\mathcal{M}$ , and a probability map  $\tilde{\mathcal{F}}$  derived from  $\mathcal{F}$ , see Fig. 1. A set of optimal transformation parameters  $\theta^*$  that align  $\mathcal{M}$  with  $\mathcal{F}$  is sought. The parameter vector  $\theta = (\phi_{\theta}, \mathbf{t}_{\theta})^{\top}$  represents a rigid transformation in 3D space with  $\phi_{\theta} = (\phi_x, \phi_y, \phi_z)$  denoting the Euler angles, and  $\mathbf{t}_{\theta} = (t_x, t_y, t_z)$  the translation along the axes of the coordinate system ( $x$ ,  $y$  and  $z$ ).

### A. Pericardium Segmentation from Preoperative CT

The method proposed by Zheng et al. [20] is used to segment the pericardium in the preoperative CT scan. It is both efficient and robust and consists of four main steps. First, the pose and scale of the heart are estimated using marginal space learning (MSL) [21], [22]. Second, a mean shape generated from a large number of annotated pericardium meshes is aligned to the estimated pose and scale. In a third step, the parameters are refined within the framework of statistical shape models (SSM) [23] using a boundary detector based on the probabilistic boosting tree (PBT) [24]. Finally, a postprocessing step is applied in order to ensure that the rib cage is not included in the segmentation. The patient-specific anchor anatomy  $A_{\mathcal{M}}$  is extracted from  $\mathcal{M}$ , yielding a geometrical model representing the pericardium in the CT scan (see Fig. 2). The model is represented by a closed triangulated surface mesh consisting of 514 vertices and 1024 triangles, where vertices correspond to the same anatomical locations among different patients. Since  $A_{\mathcal{M}}$  is independent from intraoperative information, the model segmentation can be performed prior to the intervention.

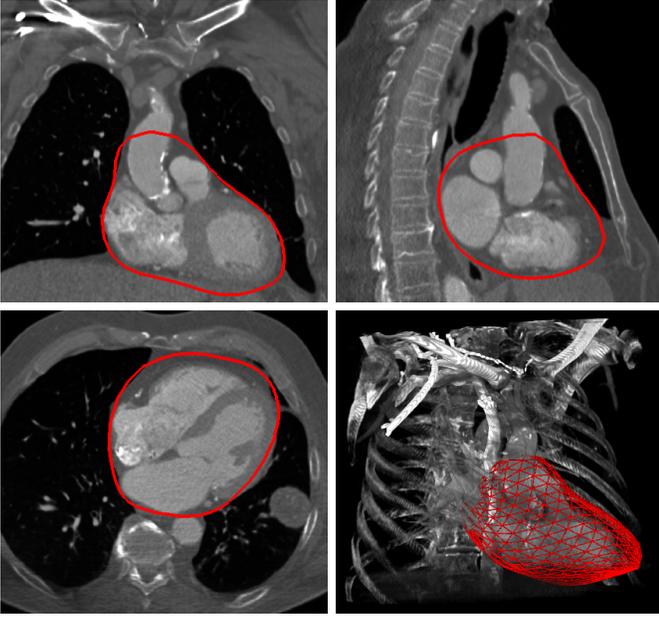


Fig. 2. Slices of a contrast-enhanced CT scan of the human torso ( $\mathcal{M}$ ) overlaid by the automatically segmented pericardium mesh  $A_{\mathcal{M}}$ , the intersection of  $A_{\mathcal{M}}$  with the plane corresponding to the visualized slice is shown in red color. Lower right: 3D rendering with 3D anatomical overlay ( $A_{\mathcal{M}}$ ).

### B. Probability Map Extraction from Intraoperative C-arm CT

Due to the lower imaging quality of intraoperative C-arm CT devices compared to preoperative CT, applying the pericardium segmentation method described in Sec. II-A to C-arm CT images can yield unsatisfactory results. Hence, in this work a different approach is proposed for the intraoperative data, where a probability map  $\tilde{\mathcal{F}} = p(A_{\mathcal{F}} | \mathcal{F})$  is derived from  $\mathcal{F}$  by evaluating a discriminative classifier on each voxel. The classifier is explicitly trained on C-arm CT images in order to delineate pericardium boundary regions  $A_{\mathcal{F}}$  robustly, while taking into account the intrinsic properties of that modality. Due to the high complexity of the detection problem (a single classifier for the entire pericardium boundary), a PBT classifier with 6 tree levels of AdaBoost [24] classifiers is used. For each AdaBoost classifier, a cascade of 40 weak learners (hyperplanes in feature space) is trained. Based on an extension of [25] to 3D space as described in [26], 3D Haar-like features are used in this work. This choice is motivated by the known computational efficiency of computing Haar-like features (through integral images), which is exploited by generating a large quantity of distinct features in little time in order to achieve a high level of robustness. In this study, a feature vector corresponding to a point (voxel) in an image consists of 97804 distinct 3D Haar-like features centered at that point.

In collaboration with medical experts, a database  $DB = \{(\mathcal{V}_i, \mathcal{P}_i) \mid i = 1 \dots n_{DB}\}$  of ground-truth pericardium meshes  $\mathcal{P}_i$  on a set of  $n_{DB} = 393$  interventionally acquired C-arm CT volumes  $\mathcal{V}_i$  was created. For training the PBT classifier, a set of positive  $\mathcal{S}^+$  and a set of negative  $\mathcal{S}^-$  samples is extracted from  $DB$ . For each  $(\mathcal{V}, \mathcal{P}) \in DB$  and each vertex in the corresponding annotation  $\mathcal{P}$ , 27 positive samples for

$\mathcal{S}^+$  are generated with regard to the position of the voxel corresponding to the given vertex and its 26 neighboring voxels. The negative samples are based on all voxels  $\mathbf{v} \in \mathcal{V}$  where the distance of  $\mathbf{v}$  to all vertices in  $\mathcal{P}$  exceeds a threshold of three times the length of the diagonal of a voxel. To keep memory requirements for training manageable and to balance both classes, subsets of 7500 samples are randomly selected for  $\mathcal{S}^+$  and  $\mathcal{S}^-$  and eventually passed to the PBT training routine. Figure 3 shows an example probability map overlaid on axial volume slices.

### C. Initialization Estimation

One of the major drawbacks in numerical optimization is the need for a reliable initial estimate in the area of attraction of the global optimum. Without this, the method is prone to converge into a local optimum. In order to find such a stable initialization  $\theta_0 = (\phi_{\theta_0}, \mathbf{t}_{\theta_0})^\top$ , our method recovers the offset  $\mathbf{t}_{\theta_0}$  between  $\mathcal{M}$  and  $\mathcal{F}$ . The rotational error ( $\phi_{\theta_0} = \mathbf{0}$ ) is neglected, since it is rather small between the CT and the C-arm CT scan due to the acquisition protocols being similar as the patients adopt similar supine positions for both scans.

Our solution is based on a concept from computer vision known as object localization, which we formulate as a classification problem. The object is a single point, the center of the pericardium in the C-arm CT scan, and we aim at locating its position. Therefore, a PBT classifier with similar settings as in Sec. II-B is trained. For each image, positive training samples are generated for all voxels that are within a radius of 10 mm of the centroid of the corresponding annotation  $\mathcal{P} \in DB$ . Negative samples correspond to voxels exceeding a Euclidean distance threshold of 20 mm to the true center. Both sets are randomly subsampled prior to the PBT training to 4000 samples for each class in order to achieve class balance and to reduce memory requirements. Due to the reduced complexity as opposed to the problem described in the previous section (in which a single position is detected instead of an entire boundary), a PBT with 3 levels of AdaBoost classifiers containing 20 weak learners per classifier is sufficient.

In the detection phase, the classifier is evaluated on each voxel and the  $n_c = 100$  candidates with highest probability are selected. The final pericardium center estimate is the centroid of the  $n_c$  candidates. Robust detections are achieved by following the idea of auto-context [27], where a classifier is trained on the output of another classifier. The probabilistic information from  $\tilde{\mathcal{F}}$  is utilized as the input for training and detection, instead of the intensities in  $\mathcal{F}$  directly, since the probability maps look similar for both contrast and non-contrast images, because the classifier for  $\tilde{\mathcal{F}}$  was trained on both types of volumes. This means that the framework is not confused by large magnitudes in intensity gradients, which most notably appear in contrast-enhanced images, for instance at the boundaries of the left ventricle or the aorta, or due to imaging artifacts caused by catheters. Thus, the method can be used with or without contrast, constituting a major benefit of our work.

To summarize, the input for both training and testing of the object localizer is the probability map  $\tilde{\mathcal{F}}$  extracted from

$\mathcal{F}$ , and the output is a pericardium center hypothesis, i.e. the estimated position of the pericardium center in  $\mathcal{F}$ , say  $\mathbf{c}_{A_{\mathcal{F}}}$ . Let  $\mathbf{c}_{A_{\mathcal{M}}}$  be the center of  $A_{\mathcal{M}}$ .  $\boldsymbol{\theta}_0$  is then calculated as

$$\boldsymbol{\theta}_0 = (\boldsymbol{\phi}_{\boldsymbol{\theta}_0}, -\mathbf{c}_{A_{\mathcal{M}}} + \mathbf{c}_{A_{\mathcal{F}}})^\top . \quad (1)$$

#### D. Optimization Strategy

Given a pericardium mesh  $A_{\mathcal{M}}$ , a probability map  $\tilde{\mathcal{F}}$  and a starting point  $\boldsymbol{\theta}_0$ , the goal is to refine  $\boldsymbol{\theta}$  to yield an optimal rigid transformation  $\boldsymbol{\theta}^*$ , such that the anchor anatomy (pericardium), and thus (see Sec. III-A3) also the target anatomy (aortic valve), is aligned in both images. This process is incorporated into a numerical quasi-Newton minimization framework utilizing the update rule of Broyden, Fletcher, Goldfarb and Shanno (BFGS) [28]. BFGS is known for its high computational efficiency and good convergence behavior. Multi-resolution optimization in a coarse-to-fine manner on three granularity levels of  $\tilde{\mathcal{F}}$  is performed.

1) *Objective Function:* A transformation  $\boldsymbol{\theta}$ , which eventually aligns the pericardium in both images, is iteratively refined. The optimization problem is defined as

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} f(\boldsymbol{\theta} | A_{\mathcal{M}}, \tilde{\mathcal{F}}) . \quad (2)$$

The objective function  $f$  is designed to maximize the combined support of all vertices of the aligned mesh  $A_{\mathcal{M}}$  within the volume of the probability map  $\tilde{\mathcal{F}}$  extracted from the C-arm CT image:

$$f(\boldsymbol{\theta} | A_{\mathcal{M}}, \tilde{\mathcal{F}}) = \frac{\sum_{\mathbf{p} \in A_{\mathcal{M}}} \mathbf{I}(\boldsymbol{\theta}(\mathbf{p}), \tilde{\mathcal{F}}) \cdot \psi(\boldsymbol{\theta}(\mathbf{p}), \tilde{\mathcal{F}})}{\sum_{\mathbf{p} \in A_{\mathcal{M}}} \mathbf{I}(\boldsymbol{\theta}(\mathbf{p}), \tilde{\mathcal{F}})} . \quad (3)$$

where  $\mathbf{p}$  denotes a vertex in  $A_{\mathcal{M}}$  (the mesh sampling strategy will be refined later on) and  $\mathbf{p}' = \boldsymbol{\theta}(\mathbf{p})$  is that vertex transformed with respect to  $\boldsymbol{\theta}$ . The indicator function  $\mathbf{I}(\mathbf{p}', \tilde{\mathcal{F}})$  evaluates to 1, if  $\mathbf{p}'$  is located inside the domain of definition of  $\tilde{\mathcal{F}}$ , otherwise it returns 0.  $\psi(\mathbf{p}', \tilde{\mathcal{F}})$  computes a value that is proportional to the inverse of the probabilistic prediction  $\tilde{\mathcal{F}}(\mathbf{v}_{\mathbf{p}'})$  at the voxel  $\mathbf{v}_{\mathbf{p}'} \in \tilde{\mathcal{F}}$  where  $\mathbf{p}'$  is located:

$$\psi(\mathbf{p}', \tilde{\mathcal{F}}) = \frac{2}{\tilde{\mathcal{F}}(\mathbf{v}_{\mathbf{p}'}) + 1} - 1 . \quad (4)$$

The output of  $\psi$  is in the range  $[0, 1]$  given that  $\tilde{\mathcal{F}}$  is probabilistic. In the rare case when the denominator in (3) evaluates to zero, or—less restrictive—when the number of vertices within the boundaries of the volume is below a certain threshold (less than a quarter of all vertices), an alternative function  $f'$  is called, which prevents  $A_{\mathcal{M}}$  and  $\tilde{\mathcal{F}}$  from diverging. Its return value is proportional to the distance of the centers of  $\tilde{\mathcal{F}}$  ( $\mathbf{c}_{\tilde{\mathcal{F}}}$ ) and  $A_{\mathcal{M}}$  transformed with respect to the current estimate:

$$f'(\boldsymbol{\theta} | A_{\mathcal{M}}, \tilde{\mathcal{F}}) \propto \|\mathbf{c}_{\tilde{\mathcal{F}}} - \boldsymbol{\theta}(\mathbf{c}_{A_{\mathcal{M}}})\|_2 . \quad (5)$$

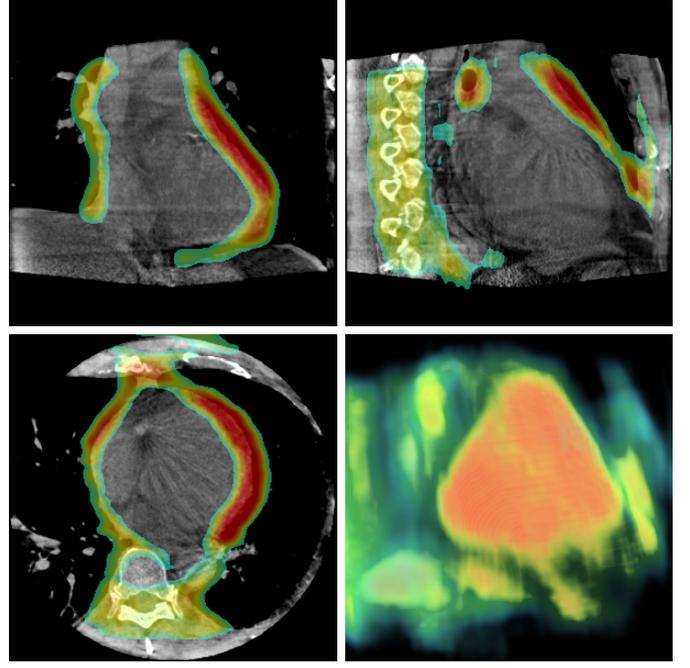


Fig. 3. Slices of a non-contrast C-arm CT volume ( $\mathcal{F}$ ) overlaid by the PBT-based probability map (thresholded), red color indicates high probability, blue colored and transparent regions are rather unlikely to contain the pericardium boundary. Lower right: Tilted frontal 3D rendering of the probability map.

2) *Gradient Computation:* In steepest-descent based minimization, the gradient  $\nabla$  of  $f$  is exploited to obtain the descent direction in each iteration. Furthermore, the BFGS method relies on the gradient in order to estimate an approximation of the inverse of the Hessian. Unfortunately,  $f$  is highly complex and therefore calculation of analytical derivatives is difficult. Hence, the gradient  $\tilde{\nabla} \approx \nabla f(\boldsymbol{\theta} | A_{\mathcal{M}}, \tilde{\mathcal{F}})$  is approximated component-wise using finite differences:

$$\tilde{\nabla}_i = \frac{f(\boldsymbol{\theta} + \boldsymbol{\delta}^i | A_{\mathcal{M}}, \tilde{\mathcal{F}}) - f(\boldsymbol{\theta} | A_{\mathcal{M}}, \tilde{\mathcal{F}})}{\delta_i^i} , \quad (6)$$

with an offset  $\delta_i^i$  (the only non-zero component of  $\boldsymbol{\delta}^i$ ) of half the resolution of  $\tilde{\mathcal{F}}$  for translation and an equivalent angle in degrees for rotation. Despite its asymmetric computation scheme, the gradient estimate is sufficiently stable in this application.

While computing the translational gradient components is straightforward, rotation in 3D poses a major problem due to its inherent non-linearity and co-dependencies. This issue is addressed by utilizing a linearization of rotation matrices  $\mathbf{R}$  using a first order approximation  $\mathbf{R}'$  as proposed by Mitra et al. [29]. With homogeneous coordinates, rigid transformation turns into a linear problem and thus can be represented by a matrix-vector multiplication with a matrix  $\mathbf{M}_{\boldsymbol{\theta}} = (\mathbf{R}_{\boldsymbol{\theta}} | \mathbf{t}_{\boldsymbol{\theta}})$ , which concatenates rotation and translation into one matrix and represents  $\boldsymbol{\theta}$ . Let  $\mathbf{R}_{\boldsymbol{\theta}}$  be the Euler 3D rotation matrix defined by  $\boldsymbol{\phi}_{\boldsymbol{\theta}} = (\phi_x, \phi_y, \phi_z)$  with  $\mathbf{R}_{\boldsymbol{\theta}}^{-1} = \mathbf{R}_{\boldsymbol{\theta}}^\top$  and

$\det(\mathbf{R}_\theta) = 1$ . Its first-order approximation [29] is given by

$$\mathbf{R}'_\theta = \begin{pmatrix} 1 & -\phi_z & \phi_y \\ \phi_z & 1 & -\phi_x \\ -\phi_y & \phi_x & 1 \end{pmatrix} \approx \mathbf{R}_\theta . \quad (7)$$

It is important to mention that  $\mathbf{R}'_\theta \approx \mathbf{R}_\theta$  only holds under small motion, i.e. when the magnitude of  $\phi_\theta$  is close to zero. Hence,  $\mathbf{R}'_\theta$  cannot be used for large angles without introducing errors. Therefore, the rotational components of  $\tilde{\nabla}$  are computed using a composite transformation. First, a point  $\mathbf{p} \in A_M$  is transformed with respect to the current estimate  $\theta$  using the exact Euler-angle representation to generate an intermediate point  $\mathbf{p}''$ . Second,  $\mathbf{p}''$  is rotated according to the minor rotation  $\delta^i$  to yield  $\mathbf{p}'$  by making use of the linearization  $\mathbf{R}'_\theta$  from (7). Altogether, we get

$$\mathbf{p}' = \delta^i(\mathbf{p}'') = \delta^i(\theta(\mathbf{p})) . \quad (8)$$

$\mathbf{p}'$  constitutes the first argument for the functions  $I$  and  $\psi$  in (3) when computing the approximate rotational gradient components of  $\tilde{\nabla}$  using (6).

3) *Multiple Resolutions*: To compensate for potential initial coarse misalignment, multi-resolution optimization is exploited. First, after convergence of a low-resolution optimizer using a low-resolution probability map  $\tilde{\mathcal{F}}|_{\text{coarse}}$ , a rough registration  $\theta^*|_{\text{coarse}}$  is determined. The next optimization step is then performed using a resampled (cf. *Resampler* in Fig. 1), finer representation of the probability map  $\tilde{\mathcal{F}}|_{\text{fine}}$ . The optimization with respect to  $\tilde{\mathcal{F}}|_{\text{fine}}$  starts at  $\theta|_{\text{coarse}}$  and eventually yields  $\theta^*|_{\text{fine}}$ . This process is repeated until a certain target resolution  $\tilde{\mathcal{F}}|_{\text{finest}}$  is reached. The resulting  $\theta^* = \theta^*|_{\text{finest}}$  then constitutes the multi-resolution-optimal set of parameters. Three different isotropic scales (4, 2, and 1 mm) are considered.

4) *Prior Sampling*: The classifier response (Sec. II-B) is more reliable in some regions of the volumes compared to others. For instance see Fig. 3, where the areas close to the left ventricle and right atrium have high responses, while classification near the spine is noisy and the right ventricle region shows low confidence. Robustness and accuracy of our method (Sec. III-B) could be improved significantly by incorporating prior knowledge into the optimization procedure. Therefore, the surface mesh of the anchor anatomy  $A_M$  is subdivided into  $n_r$  small regions  $r_j \in A_M$  with  $\bigcup_j r_j = A_M$  and  $r_j \cap r_k = \emptyset \forall j, k : j \neq k$ . Each region  $r_j$  is assigned a patient-independent weight  $w_{r_j}$  yielding a weight vector  $\mathbf{w} = \{w_{r_j} \mid r_j \in A_M\}$ . The term  $\mathbf{w}$  increases the influence of those regions that are likely to be located within a high confidence area in  $\tilde{\mathcal{F}}$ , whereas a region that is noisy or often falsely classified gets penalized. This is achieved by increasing the sampling rate of the pericardium mesh at regions  $r_j$  with high weights  $w_{r_j}$ , while reducing the sampling rate for low  $w_{r_j}$  regions. Gradient magnitude  $\|\nabla\|_2$  in the C-arm CT images was most useful for reliability predictions. Values for  $\mathbf{w}$  (cf. Fig. 4) are determined based on this observation and the C-arm CT images  $\mathcal{V}_i$  and pericardium annotations  $\mathcal{P}_i$  from DB from Sec. II-B as follows. First, vertex-wise weights  $w'_l = \{w'_l\}$  are computed for all vertices of the pericardium



Fig. 4. Visualization of an example probabilistic prior sampling of  $A_M$  (white points). Dark/red colors on the mesh depict regions of sparse sampling (small  $w_r$ ), while in bright/yellow regions points are sampled more densely during the optimization.

model described in Sec. II-A:

$$w'_l = \sum_{i=1}^{n_{DB}} \|\nabla(\mathcal{V}_i(\mathbf{p}_{il}))\|_2 , \quad (9)$$

$\mathbf{p}_{il}$  denoting the  $l^{\text{th}}$  vertex of  $\mathcal{P}_i$ . Second,  $w'$  is normalized by applying an affine transformation such that the smallest and largest values are mapped to 0 and 1, respectively. Finally, for each region  $r_j$ ,  $w'$  is mapped to the region-wise weights  $w_{r_j}$  by averaging over the vertex-wise weights  $w'_l$  of all vertices that are enclosed by  $r_j$ . In this study,  $n_r = 1024$  regions are used, each of them corresponding to a particular triangle  $\Delta$  in the pericardium mesh. Consequently, each  $w_{r_j}$  is computed as the mean of 3 vertex-based weights  $w'_l$  determined by  $\Delta$ .

Let  $\Gamma_w$  be a probabilistic function that returns three-dimensional points on the surface of a given mesh according to the sampling probabilities defined by  $\mathbf{w}$ . To incorporate the prior knowledge into our framework, the objective function  $f$  from (3) is extended to obtain:

$$f(\theta \mid A_M, \tilde{\mathcal{F}}, \Gamma_w) = \frac{\sum_{\mathbf{p} \in \Gamma_w(A_M)} I(\theta(\mathbf{p}), \tilde{\mathcal{F}}) \cdot \psi(\theta(\mathbf{p}), \tilde{\mathcal{F}})}{\sum_{\mathbf{p} \in \Gamma_w(A_M)} I(\theta(\mathbf{p}), \tilde{\mathcal{F}})} . \quad (10)$$

Motivated by a tradeoff between low computational complexity and high sampling density, 2000 points are sampled on  $A_M$ , which are drawn from  $\Gamma_w$  as follows. First, an index  $z$  corresponding to a region  $r_z$  and a threshold value  $\tau$  are drawn from a discrete uniform distribution with range  $[1, n_r]$  and a continuous uniform distribution with range  $[0.2, 1]$ , respectively. The latter range is determined experimentally to exclude regions with very low reliability. This first step is repeated until  $w_{r_z} \geq \tau$  to increase the sampling density of high-confidence regions. Finally, a random point on the surface of the triangle  $r_l$  is chosen using randomized barycentric coordinates  $(u, v, w)$ . Scalars  $u$  and  $v$  are drawn from continuous uniform distributions with range  $[0, 1]$  and  $[0, 1 - u]$ , respectively, and  $w = 1 - u - v$ . An example probabilistic prior sampling is illustrated in Fig. 4.

### III. EXPERIMENTAL RESULTS

#### A. Dataset and Error Measure

1) *Clinical Dataset*: A set of 95 corresponding clinical CT and C-arm CT volumes was compiled, each with an isotropic resolution of 1 mm. About one quarter (25) of the C-arm CT images are native, while 70 were acquired with contrast. Medical experts created a database  $DB_{\text{clinic}}$  of pericardium models  $\tilde{\mathcal{P}}$  annotated in the C-arm CT images. For 43 studies,

annotations  $\check{R}$  and  $R$  of the aortic valve are included based on the patient-specific C-arm CT and CT acquisition, respectively. Furthermore, 3 landmarks (hinges of the aortic valve) were annotated in those cases. For more information on the aortic valve model the reader is referred to [30]. In 12 C-arm CT images (9 contrast, 3 native), a large part of the anchor anatomy (pericardium) is not visible. For these cases the uncertainty in the annotation is significantly increased. However, with guidance from the corresponding CT image, the experienced experts could still create valid pericardium models.

2) *Mesh-to-Mesh Error*: Since the proposed method is based on geometrical models and the clinical dataset includes ground-truth annotations, quantitative evaluation is possible. The results are based on a symmetric mesh-to-mesh distance metric  $\varepsilon$ . Its implementation utilizes the point-to-triangle distance  $\varepsilon_{pt}(\mathbf{p}, \Delta)$  between a point  $\mathbf{p}$  and a triangle  $\Delta$ . Let  $X, Y$  be triangulated surface meshes with  $n_X$  and  $n_Y$  denoting the number of vertices in  $X$  and  $Y$ , respectively.  $\varepsilon$  is defined as:

$$\varepsilon(X, Y) = \frac{1}{2} \left( \sum_{\mathbf{p} \in X} \min_{\Delta \in Y} \frac{\varepsilon_{pt}(\mathbf{p}, \Delta)}{n_X} + \sum_{\mathbf{p} \in Y} \min_{\Delta \in X} \frac{\varepsilon_{pt}(\mathbf{p}, \Delta)}{n_Y} \right). \quad (11)$$

Note that this distance measure might underestimate misalignment tangential to the surface. However, it is a fairly natural measure closely resembling visual assessment by experts.

3) *Correlation Between Anatomies*: A good correlation between the pose of anchor and target anatomy is crucial to the applicability of our method in clinical practice. In this section, a technical justification for this assumption is provided based on the following experiment. For each dataset in DB with annotated AV landmarks, the optimal rigid registration between those landmarks in corresponding images is estimated [31]. This transformation, which is solely based on the AV, yields an average mesh-to-mesh error of 4.76 mm measured at the pericardium. Although this error seems large, it is very likely that it is mostly related to differences in the shape of the pericardium, since the distance is computed between the automatically segmented mesh on the one hand, and the manually annotated mesh on the other hand, both introducing biases and uncertainties. The error due to the differences in shape can be approximated by aligning both pericardium meshes using the vertices as landmarks, yielding an error of 4.23 mm, which is only slightly (0.53 mm) lower than the error measured using the AV-based transformation. Hence, good correlation between the pose of anchor and target anatomy can be assumed.

## B. Evaluation on Clinical Data

Our probabilistic sparse matching approach, which is based on the novel prior sampling strategy as described in Sec. II-D on  $DB_{\text{clinic}}$  is evaluated and compared against our previously published vertex-based sparse matching method [19] with prior weights both enabled and disabled. The rows in Table I show the anchor anatomy alignment error (AAE) statistics, i.e. error statistics resulting from a comparison of the optimally transformed segmented pericardium  $\theta^*(A_M)$  and the ground-truth annotation in the C-arm CT volume  $\check{P}$ . Table II shows

the target anatomy errors (TAE), i.e. a comparison of the transformed CT-based aortic valve  $\theta^*(R)$  and its C-arm CT based annotation  $\check{R}$ . From left to right, the columns contain the name of the analyzed method, followed by error measurements in mm, starting with the mean error and standard deviation, the 50<sup>th</sup> (median), 80<sup>th</sup> and 90<sup>th</sup> percentiles of the errors, as well as the maximum and minimum error. The two right-most columns show the ratio of fail-cases (a measure of robustness), where a registration accuracy with an AAE (Table I) or TAE (Table II) greater or equal than 2 cm is considered a fail-case, and an adjusted mean, where all fail-cases are excluded.

Probabilistic sparse matching achieves a mean AAE of  $5.48 \pm 1.82$  mm measured between the anchor anatomy (pericardium). As described in Sec. II-D, 2 000 points are randomly sampled on the pericardium mesh. The performance in terms of robustness and accuracy is superior compared to using only 514 points (same number as in our previous approach [19]). However, further improvements by using 3 000, 5 000 or 10 000 points are only marginal on  $DB_{\text{clinic}}$ . On the 12 cases with partially visible pericardium in the C-arm CT image, the AAE increases to  $7.12 \pm 1.63$  mm versus  $5.25 \pm 1.72$  mm for the remaining 83 cases. This discrepancy can be explained by the gain in uncertainty when annotating the pericardium with missing data, since errors in the annotation can have quite large effects on the quantitative errors computed using (11). Hence, the increase in AAE is not necessarily related to inaccuracies in the registration. Between the 70 contrasted and the 25 non-contrasted volumes, no significant discrepancy was observed with AAEs of  $5.33 \pm 1.69$  mm and  $5.91 \pm 2.09$  mm, respectively. Furthermore, with a mean TAE of  $4.67 \pm 1.94$  mm, the target anatomy (aortic valve) is aligned very well. No fail-cases are observed for both anchor and target anatomy.

Using the objective function (3) with no prior information, as implemented in our previously proposed vertex-based method [19] without patient-independent weighting, the mean errors increase significantly by more than 30% in terms of both AAE and TAE. One reason is that more outliers are generated, which leads to a fail-case that has a strong influence on the aggregated errors. When we apply the vertex-based method with vertex-wise prior weights as described in [19], the mean errors decrease by almost 1 mm AAE and also slightly with respect to TAE. However, compared to the  $5.60 \pm 1.81$  mm AAE reported in [19], the aggregated error including the newly added datasets notably increased and even an outlier is observed. By discarding vertices with low weights  $< 0.2$  in [19], similar as in the current approach where regions with weight  $w < 0.2$  are not sampled, the mean error of the vertex-based method on  $DB_{\text{clinic}}$  slightly decreases by 0.22 mm (AAE). Yet, the performance of probabilistic sampling could not be reached as other improvements are important factors, too. For instance, the vertices of the pericardium (Sec. II-A) are not uniformly distributed on the mesh, e.g. the sampling density significantly decreases from apex to septum. This can become an issue in [19] as the fixed sampling pattern solely defined by the fixed mesh topology is used, what can lead to some regions having high influence on the objective function value, although the weights of the vertices in the regions might be low, and vice versa. With our improved sampling strategy,

TABLE I  
AAE STATISTICS [MM]  $\varepsilon(\theta^*(A_M), \tilde{P})$  AFTER REGISTRATION (#STUDIES=95)

Method	Mean	SD	Median	80%	90%	Max	Min	Fail-cases	Adjusted Mean
Probabilistic Sparse Matching	<b>5.48</b>	<b>1.82</b>	<b>5.22</b>	<b>7.26</b>	<b>7.88</b>	<b>10.40</b>	<b>1.93</b>	<b>0%</b>	<b>5.48</b>
Vertex-based Sparse Matching [19]	6.27	3.33	5.43	7.81	8.61	26.18	2.61	1.1%	5.89
Vertex-based Sparse Matching (no Prior)	7.20	3.79	6.63	8.85	10.36	33.47	2.11	1.1%	6.92
Quasi-global Search	13.06	12.79	5.86	24.38	34.17	57.56	2.33	25.3%	6.41
Quasi-global Search (masked)	13.69	19.82	7.03	16.40	37.56	139.74	2.11	15.8%	6.89
ITK Registration Framework	15.13	14.95	7.77	25.42	35.70	76.65	2.50	25.3%	7.98
ITK Registration Framework (masked)	20.29	22.84	7.64	36.80	56.12	89.10	2.44	32.6%	7.12

TABLE II  
TAE STATISTICS [MM]  $\varepsilon(\theta^*(R), \tilde{R})$  AFTER REGISTRATION (#STUDIES=43)

Method	Mean	SD	Median	80%	90%	Max	Min	Fail-cases	Adjusted Mean
Probabilistic Sparse Matching	<b>4.67</b>	<b>1.94</b>	<b>4.22</b>	<b>6.54</b>	<b>7.29</b>	<b>8.83</b>	1.24	<b>0%</b>	<b>4.67</b>
Vertex-based Sparse Matching [19]	5.91	5.54	5.02	6.94	7.57	39.01	2.61	2.3%	5.12
Vertex-based Sparse Matching (no Prior)	6.14	3.54	5.56	8.47	9.44	22.35	1.18	2.3%	5.76
Quasi-global Search	21.76	22.32	8.49	49.05	60.36	77.67	1.43	37.2%	6.45
Quasi-global Search (masked)	13.72	19.73	5.25	18.29	37.50	79.80	<b>1.02</b>	18.6%	5.51
ITK Registration Framework	21.23	25.77	7.87	33.34	61.07	94.67	1.28	37.2%	5.52
ITK Registration Framework (masked)	21.81	26.15	9.59	42.31	64.16	96.20	1.08	37.2%	6.02

this problem is mitigated by sampling a larger number of points at randomized locations on the mesh surface, while still taking into account the prior knowledge on the regional confidence of the classifier. Due to the involved randomness, our sampling approach also shares some of the advantages of jittered sampling [32], which is known to improve the smoothness of the objective function, leading to better accuracy and robustness of the overall approach. To conclude, the prior knowledge improves the fusion performance significantly and probabilistic sampling further increases robustness and accuracy as opposed to vertex-based sampling. Figure 5(a-d) depicts representative qualitative fusion results from various datasets  $\in DB_{\text{clinic}}$ .

### C. On the Effect of Prior Knowledge

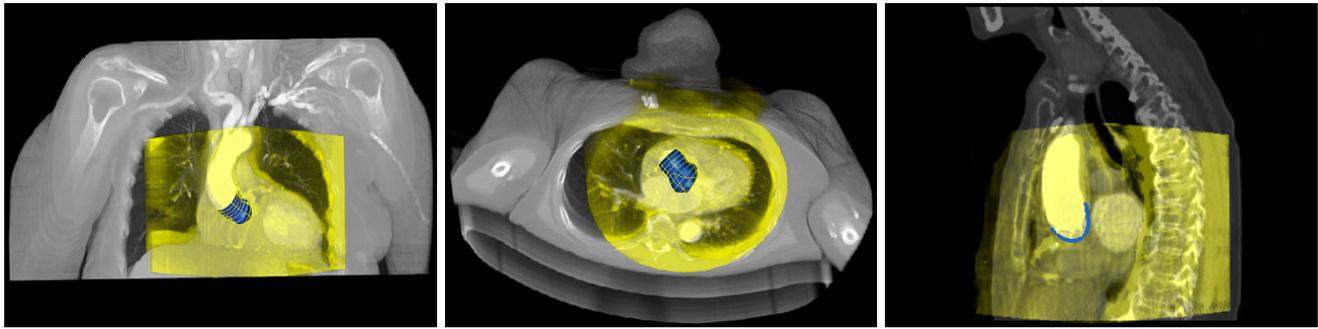
To further explain the importance of our prior sampling strategy, the objective function with standard vertex-based sampling and no prior weights is compared against the objective function from (10) with prior sampling for one example dataset. The plots in the top row of Fig. 6 are generated as follows: After performing a full registration with (blue curves) and without (red dashed curves) prior sampling, the pericardium  $A_M$  is translated along orthogonal directions (columns of the plot represent the x-, y- and z-axis). At equidistant sampling intervals of length 1 mm within the range from -50 mm to 50 mm around the estimated optimal point, the objective functions are then evaluated given the manually misaligned  $A_M$  and  $\tilde{F}$ . To facilitate visual comparison, an affine transformation is applied to the curves such that the minimum and maximum values are mapped to 0 and 1, respectively. The latter is performed for both strategies individually using the data from all three plots. The plots in the bottom row show the AAE between the manually misaligned  $A_M$  (as described above) and the fixed C-arm CT annotation  $A_F$ .

Compared to vertex-based sampling, prior sampling reduces the influence of noise, allows for smoother energy functions and exhibits better correlation with the AAE, the error metric

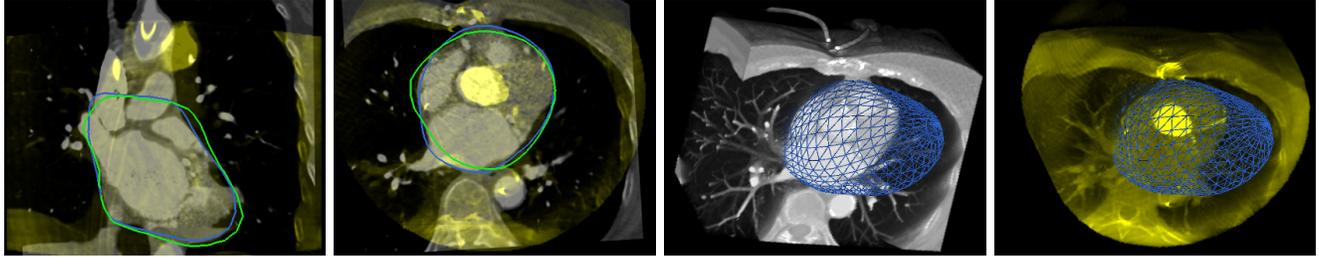
used to evaluate the registration. Throughout  $DB_{\text{clinic}}$ , this results in increased robustness and alignment with higher accuracy. Although the global optimum in terms of AAE could not be reached in this particular case using prior sampling (see for instance the plot in the center bottom row of Fig. 6, where the AAE slightly decreases until  $t_x \approx -6$  mm), the registration accuracy increased by almost 2 mm compared to vertex-based sampling. Moreover, the plots in the left column reveal that the vertex-based sampling method misses a (local) minimum of the AAE curve at  $\approx -14$  mm in the x direction. In fact, this AAE minimum is rather close to a local maximum of the marginalized objective function at  $\approx -16$  mm in x-direction. This local maximum is partially related to noisy regions in the probability map  $\tilde{F}$ . Probabilistic sampling reduces the influence of such regions, similar to filtering techniques, and thus produces nicely monotonic and smooth objective functions with large convex areas. We observed similar behavior for other datasets and also for changes in rotation.

### D. Comparison to State-of-the-Art Fusion

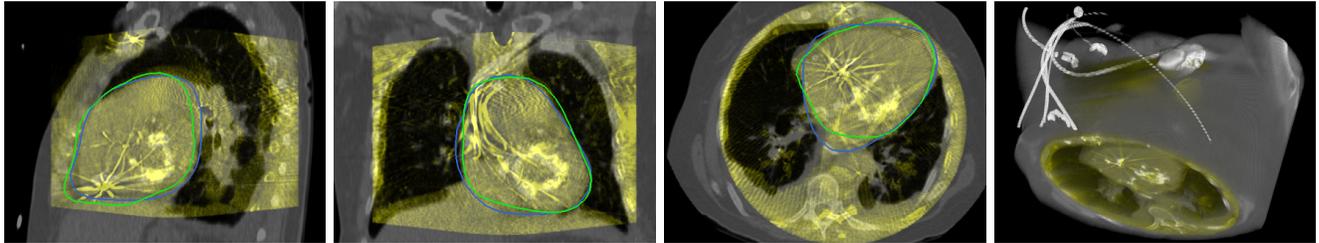
Our model-to-image registration is quantitatively compared to two state-of-the-art image-to-image registration approaches. Intuitively, one might expect that the registration results will improve when the optimizer focuses on the anchor anatomy (pericardium) only as opposed to using the entire image. Therefore, an option to mask a region of interest (ROI) in the preoperative CT image (denoted as "Name-of-Method (masked)" in Table I and II) was implemented. The mask image is created by (i) automatic segmentation of the pericardium as described in Sec. II-A, (ii) converting the pericardium mesh into a binary image of the same size, spacing and pose as the CT image, where voxels inside the mesh are set to enabled and all the other voxels are set to disabled, and (iii) dilating the binary image (increasing the size of the ROI) utilizing a spherical structuring element with a radius of 5 mm. Step (iii) ensures that the entire heart as well as a small area around it is included in the ROI. In our experiments, this last step led



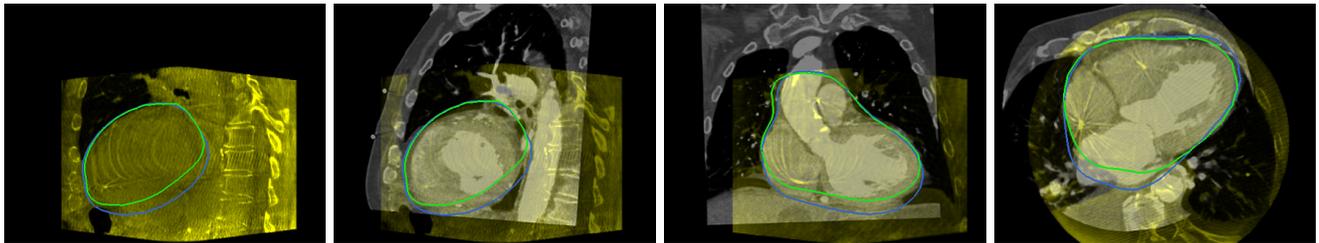
(a) Study A: Contrast CT and C-arm CT. Blue: mapped aortic valve from CT scan. AAE = 1.93 mm, TAE = 1.24 mm.



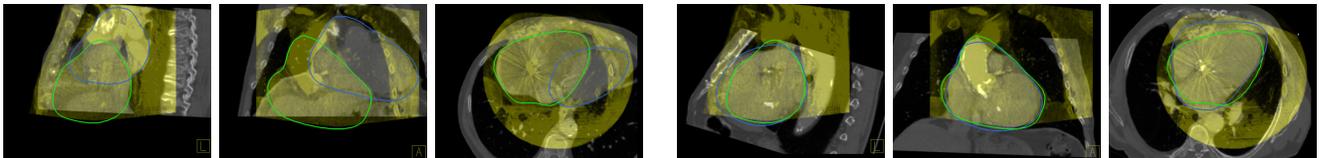
(b) Study B: Contrast CT and C-arm CT. Green: ground-truth C-arm CT pericardium annotation, blue: fused CT pericardium. AAE = 4.07 mm.



(c) Study C: Non-contrast CT and C-arm CT. Coloring as in Fig. 5(b). AAE = 4.94 mm.



(d) Study D: Contrast CT and non-contrast C-arm CT. Coloring as in Fig. 5(b). AAE = 6.11 mm.



(e) Study E: Contrast CT and C-arm CT. Left: Quasi-global search, AAE = 40.17 mm. Right: Quasi-global search (masked), AAE = 5.24 mm.

Fig. 5. (a-d) Representative qualitative results from several automatically registered datasets. Yellow: intraoperative C-arm CT, gray: aligned high-quality preoperative CT overlay, others: anatomical models automatically extracted (from CT scan) or annotated (based on C-arm CT), mapped into the joint coordinate system. The left image in the row (d) illustrates that especially in non-contrasted C-arm CT images, there are uncertainties involved in annotating the pericardium. It is not clear whether the fused or the annotated pericardium fits better. Non-ideal annotations usually increase the measured quantitative error. (e) Comparison method (Quasi-global Search) without and with pericardium masking. In this case, masking improves the result significantly.

to significant improvements in robustness, since the bordering area between the pericardium and the lungs provides crucial information that can improve the outcome of the similarity metrics (large gradient magnitudes and homogeneous regions).

1) *ITK Registration Framework*: The first method utilizes the Insight Segmentation and Registration Toolkit (ITK), an

open source medical imaging library [33]. Since intensities in the CT image and the C-arm CT image do not necessarily correlate (Sec. I), the similarity metric is based on mutual information [34]. All voxels in the image and 50 bins for the histogram as proposed by Mattes et al. [34] are used.

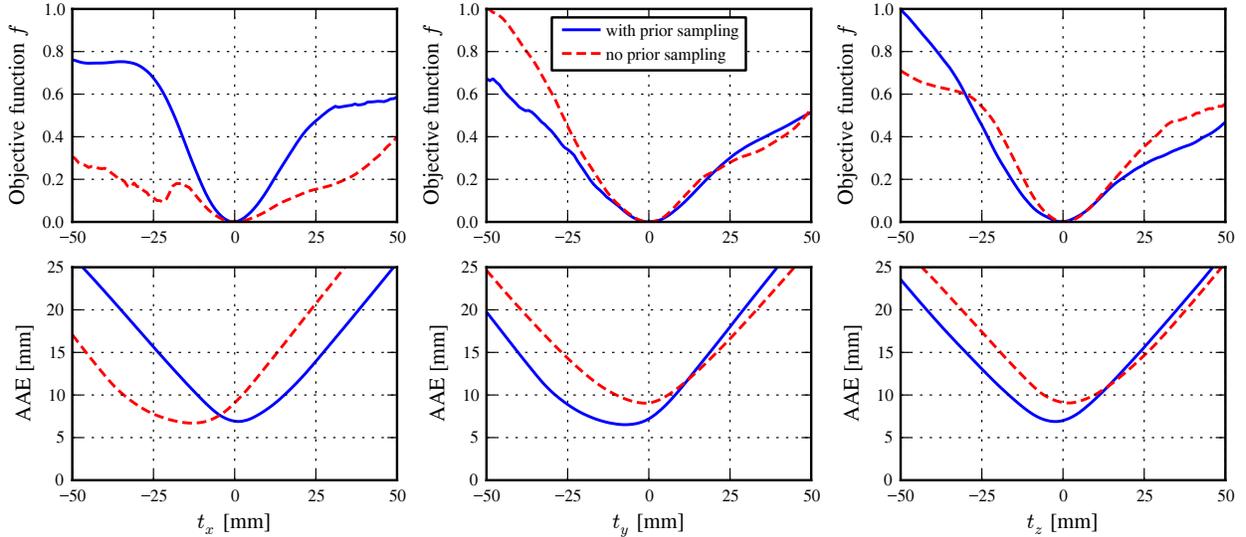


Fig. 6. Top row: Objective functions for one dataset marginalized over the x, y and z direction (left, middle and right). The red (dashed) curves were generated using vertex-based sampling, while the more sophisticated objective function leveraging prior knowledge by performing a probabilistic prior sampling of  $A_M$  was used to create the blue (solid) curves. Bottom row: Plots of corresponding (to the plots above) anchor anatomy alignment error (AAE). Plots in the same column share the same axis. For more information we refer the reader to Sec. III-C.

Optimal transformation parameters  $\theta_{\text{ITK}}^*$  are obtained by a multi-resolution optimizer for rigid versor transformations. The scales for the components of the versor were adjusted according to [33]. Maximum and minimum step lengths are customized adaptively for each resolution, the maximum number of iterations is set to 200, and the procedure is initialized using the method presented in Sec. II-C of this paper. Although intermediate results from our proposed method are used for initialization, which adds bias to the ITK method, it makes comparison more fair as opposed to simple strategies like aligning the centroid of both volumes, because they are unlikely to produce good initializations for datasets where the heart is not centered in both volumes. Results obtained running this framework on  $\text{DB}_{\text{clinic}}$  are presented in the last two rows of Table I (anchor anatomy) and Table II (target anatomy). The method fails for approximately 25% of the studies. However, since this framework is not specifically designed to align the pericardia and no preprocessing is performed, the large number is understandable. Many of the failures occur for images with significant differences in the size of the field of view between the CT and the C-arm CT acquisition. The mean error of 15.13 mm yielded by this framework is substantially larger than the error of our method, with a mean of 5.48 mm on the same datasets. A rigid registration is computed in  $8.3 \pm 7.1$  minutes on average. Please note that no optimizations with respect to runtime were implemented.

To increase the focus on aligning the pericardia, the experiment are repeated using an image mask derived from the CT pericardium mesh as described above. Indeed, for some images, the alignment can be improved significantly. For instance, for one dataset the AAE is reduced from 13.23 mm to 3.56 mm when utilizing the mask image. However, the number of cases where the results get worse due to the masking prevail.

2) *Quasi-global Search*: The second method is a quasi-global knowledge-driven registration approach for thoracic-

abdominal CT and C-arm CT images designed for image-guided interventions [18]. Given an intraoperative C-arm CT image, in the first step, three surrogate 2D images, so-called Anatomy Targeted Projections (ATP), are created. An ATP is a maximum-intensity-like 2D projection, which focuses on a specific anatomy or tissue type (e.g. bone, soft tissue, etc.). This is achieved by projecting the intensity of the voxel with the maximum likelihood of belonging to the targeted tissue type along each projection ray. The use of 2D ATPs instead of the 3D volume can reduce computational costs significantly and thus allows for a large number of similarity metric evaluations within a reasonable time frame. The authors chose an adaption of normalized mutual information (NMI) as similarity metric. Second, multiple starting points in registration parameters space are chosen to approximate a global search, motivated by the assumption that most CT volumes have a larger field of view compared to C-arm CT images. In a third step, the globally optimal candidate is selected by analyzing similarity values and gradients at all starting points. Last, a local multi-resolution optimization is performed, yielding an optimal set of rigid transformation parameters. The authors claim that their method is fast and robust with low target registration and maximum registration errors on 20 datasets.

Quantitative results (Tables I and II) show that on average, the AAE of the quasi-global search is more than twice as large as the AAE of our method, and the TAE increases from 4.67 mm to 21.76 mm. This indicates that finding a solution to the problem of aligning the pericardium in such heterogeneous images (varying field of view, contrast and non-contrast images, etc.) from different modalities is hard and working with image intensities directly might not be sufficient. However, while our learning-based method works particularly well for the application described in this paper, it is not well suited for general-purpose registration without major efforts

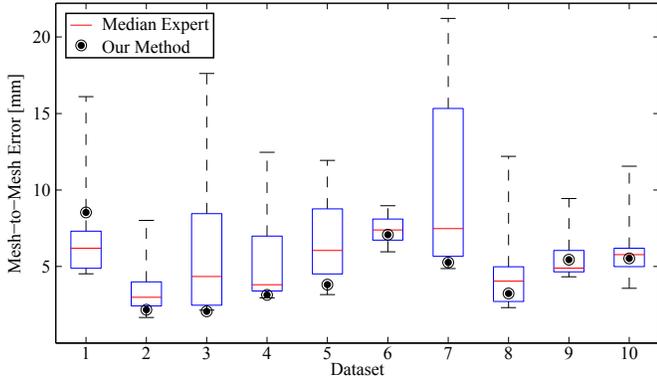


Fig. 7. Inter-user variability compared to our method. The edges of the boxes indicate 25<sup>th</sup> and 75<sup>th</sup> percentiles of the expert errors (AAE).

in terms of training data acquisition and manual annotation. In contrast, the quasi-global search is not specialized for a particular application, it is rather an approach designed for handling various registration tasks from a larger problem domain. The observations regarding the quasi-global search when using the pericardium mask are similar to the ITK registration framework when using the mask. On the one hand, there are datasets, where the alignment improves significantly (for an example see Fig. 5(e)). On the other hand, in many cases performance decreases. Besides, masking can lead to individual excessive errors as high as 139.74 mm (see Table I), whereas the maximum error of the standard quasi-global search is below 60 mm. Although a smaller number of fail-cases is observed, the details discussed above lead to a slightly higher mean AAE of 13.69 mm for the method with pericardium masking, compared to 13.06 mm without masking. Surprisingly, the TAE (see Table II) in the 43 datasets where annotations of the aortic valve are available decreases from 21.76 mm without masking to 13.72 mm with masking. This is due to the images in this specific subset working better with pericardium masking (AAE =  $11.17 \pm 11.25$  mm) compared to the method without masking (AAE =  $16.35 \pm 13.26$  mm). Incorporating our advanced PBT-based initialization (Sec. II-C) into the algorithm by Zhang et al. without masking yields mean quantitative errors of  $12.61 \pm 12.15$  mm AAE and  $20.28 \pm 21.36$  mm TAE, which correspond to only a marginal improvement over the version without custom initialization. The same holds when using image masks, yielding  $12.10 \pm 17.84$  mm AAE and  $12.63 \pm 18.95$  mm TAE. These results suggest that intensity-driven methods are prone to be inaccurate and unstable for such heterogeneous images, and that they are likely to diverge from a good initialization. Furthermore, on our testing machine (see Sec. III-F), the mean runtime until a rigid registration is computed is  $2.48 \pm 1.0$  s.

### E. Inter-user Variability Study

Ascribing a rational meaning to quantitative results is challenging. In most cases, the true performance of a system would not only be measured in absolute terms but rather relative to the manual performance of experts. Thus, our method is compared to the individual performances of a group of

$n_{\text{user}} = 10$  technical experts, who work on such clinical data on a daily basis. The users were asked to align  $n_{\text{data}} = 10$  pairs of volumes (a subset of  $\text{DB}_{\text{clinic}}$ , see Sec. III-A1) manually. An intuitive, custom software tool that allows for adjusting rigid transformations in a convenient manner was provided. The progress is visualized in real-time using 2D MPR planes and simultaneous 3D rendering of the two volumes in colored semi-opaque mode. In order to register two images manually, only mouse and keyboard are required. The latter is used to toggle between two modes via simple keystrokes. The user either drags the moving volume along an axis (translation mode) or rotates it around an axis (rotation mode). In both cases, the axis is defined by the active (mouse focus) MPR plane.

Let  $\theta_{ji}$  be the manually estimated transformation parameters of the  $i^{\text{th}}$  user for the  $j^{\text{th}}$  pair of volumes. The fit to the ground-truth C-arm CT annotation  $\check{P}_j$  is compared, i.e. the error  $\varepsilon(\check{P}_j, \theta_{ji}(\check{P}_j))$  between the manually annotated fixed C-arm CT pericardium and the moving CT pericardium transformed with respect to the user’s manual transformation is computed. Results are shown in Fig. 7. Our automated method exhibits lower errors than the median user in 80% of all cases and shows high robustness with no outliers. There exists only one pair of volumes where the automatic fusion is inferior to more than 75% of the users. Moreover, the users’ manual fusion time per data pair ranges from two to five minutes, while our method takes less than two seconds on average (see Sec. III-F), which means a speedup of up to 99%. To conclude, with no fail-case and reliable performance, our fully-automatic approach outperforms manual registration in terms of robustness, accuracy and runtime.

### F. Runtime Performance

Our method is designed to be used interventionally, resulting in a need for low computational costs associated with the registration. A runtime analysis was conducted on an off-the-shelf consumer laptop with an Intel® Core™ i7-3720QM CPU @ 2.60 GHz with 4 physical cores and 8 GB of main memory. Our prototype is implemented in C++ utilizing OpenMP [35] for efficient parallel programming.

The average runtime for the entire process of estimating rigid transformation parameters  $\theta^*$  measured on  $\text{DB}_{\text{clinic}}$  is  $1.562 \pm 0.286$  seconds. The speed-up compared to the previously reported runtime of  $2.9 \pm 0.4$  s [19] for the vertex-based method is mainly due to code optimizations. Most of these improvements also benefit the previous method: the runtime of [19] is reduced to  $1.702 \pm 0.279$  seconds, however, the performance of the current method could not be reached, partially due to [19] requiring about 20% more cost function evaluations. While registration within a few seconds as opposed to several minutes (unoptimized ITK method) is important, reducing speed from 2.9 s to 1.6 s is desirable, but not crucial to the procedure. Below, the runtime behavior of the four major components of the registration framework is discussed briefly. For a detailed overview see Fig. 8.

The runtime for the automatic model segmentation (Sec. II-A) from the preoperative CT volume by Zheng et

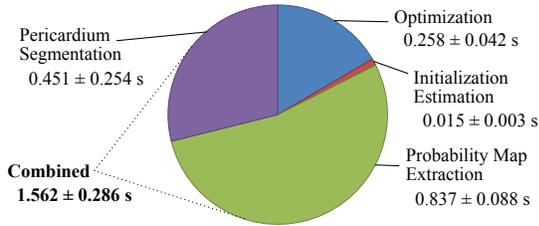


Fig. 8. Relative / absolute runtime of the four major framework components.

al. [20] does not rely on intraoperative data and thus it can be performed offline, prior to the intervention. It scales linearly with the number of voxels in the volume, ranging from 0.181 s for a scan consisting of  $206 \times 206 \times 103$  voxels to 1.70 s for a volume with  $497 \times 497 \times 278$  voxels. On average over all CT volumes  $\in \text{DB}_{\text{clinic}}$ , the mean runtime is  $0.451 \pm 0.254$  s.

The cost for generating a probability map (Sec. II-B) is strongly correlated with the size and resolution of the C-arm CT volume, since the classifier has to be evaluated on each voxel. This is done in parallel on the CPU. Please note that this is a task that can be outsourced to the graphics processing unit (GPU), potentially resulting in a massive gain in performance. For a typically sized volume of  $240 \times 240 \times 180 \text{ mm}^3$ , probability map generation takes approximately 0.81, 1.94, 5.73 or 39.9 seconds for resolutions of 4, 3, 2 and 1 mm, respectively. In our standard approach, the 4 mm classifier is utilized, i.e. on average, probability map generation accounts for  $0.837 \pm 0.088$  seconds of the overall runtime.

The use of statistical object localization to estimate the initialization  $\theta_0$  for the optimizer (Sec. II-C) adds very little to the complexity. The maximum runtime is 24 milliseconds.

The last component is the iterative optimization (Sec. II-D) with an average runtime of  $0.258 \pm 0.042$  s, which is mainly influenced by the number of objective function evaluations and by the cost for resampling the probability map (multiple resolutions). The latter needs to be done twice (4 mm  $\rightarrow$  2 mm and 4 mm  $\rightarrow$  1 mm resolution), consuming approximately 0.141 s. The evaluation of  $f$  and the approximate gradient computation  $\bar{\nabla}$  is combined in one function  $g$ . Averaged over  $\text{DB}_{\text{clinic}}$ ,  $g$  was called  $144.7 \pm 22.1$  times (accumulated over all granularity levels). Independent of the current resolution, the mean runtime of one call is 0.81 milliseconds. Thus, the average cost of successive calls to  $g$  for one registration is 0.117 s. Since the computational complexity of  $g$  scales linearly with the number of sampling points, the moderate increase from 514 [19] to 2000 points has no significant impact on the runtime.

#### IV. CONCLUSIONS

In this paper, a fast and fully-automatic method to fuse preoperative CT and intraoperative 3D C-arm CT data is presented. A novel sparse matching approach is employed to align the preoperative anchor anatomy to the intraoperative setting. Data and model uncertainties are learned and exploited for the matching process. Quantitative and qualitative evaluation demonstrate a fast and accurate mapping of the anchor and target anatomy to the intraoperative modality. In direct

comparison with a state-of-the-art registration framework and a recently proposed quasi-global, knowledge-driven fusion approach, our method outperforms both significantly in terms of robustness and accuracy regarding the targeted application on a database of 95 clinical CT and C-arm CT volumes. Furthermore, an inter-user variability study with ten users confirms that the accuracy of our method lies within the confidence interval of the expert group. While computation times of our method (1.6 s) and the quasi-global approach (2.5 s) are comparable, the ITK registration framework is significantly slower, typically consuming five to ten minutes per registration, which is similar to manual alignment performed by experts.

The main limitation of our approach to be applicable in other domains like liver or lung fusion for instance, or with data from other modalities, is the need for a significant number of training datasets from a broad spectrum of potential scanners and acquisition protocols. Furthermore, an anchor anatomy needs to be defined, and annotated manually in each image in order to train a classifier to create reliable probability maps. In our application, a large database of manual annotations of the pericardium was necessary. It contains non-gated contrast and non-contrast images acquired from different detectors (size and resolution), various fields of view and hundreds of distinct patient geometries.

To conclude, comprehensive patient-specific models can be estimated from high-contrast CT and fused into the imaging environment of operating rooms to facilitate guidance in minimally-invasive cardiac surgery, while meeting interventionally necessary constraints such as low computation time, high accuracy and robustness against noisy data, only partially visible models and imaging artifacts.

#### REFERENCES

- [1] T. Walther, M. Chu, and F. Mohr, "Transcatheter aortic valve implantation: Time to expand?" *Current Opinion in Cardiology*, vol. 23, no. 2, pp. 111–116, 2008.
- [2] M. B. Leon, C. R. Smith, M. Mack, D. C. Miller, J. W. Moses, L. G. Svensson, E. M. Tuzcu, J. G. Webb, G. P. Fontana, R. R. Makkar, D. L. Brown, P. C. Block, R. A. Guyton, A. D. Pichard, J. E. Bavaria, H. C. Herrmann, P. S. Douglas, J. L. Petersen, J. J. Akin, W. N. Anderson, D. Wang, and S. Pocock, "Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery," *New England Journal of Medicine*, vol. 363, pp. 1597–1607, 2010.
- [3] S. K. Kodali, M. R. Williams, C. R. Smith, L. G. Svensson, J. G. Webb, R. R. Makkar, G. P. Fontana, T. M. Dewey, V. H. Thourani, A. D. Pichard, M. Fischbein, W. Y. Szeto, S. Lim, P. S. Greason, S. C. Malaisrie, P. S. Douglas, R. T. Hahn, D. Whisenant, J. J. Akin, W. N. Anderson, and M. B. Leon, "Two-year outcomes after transcatheter or surgical aortic-valve replacement," *New England Journal of Medicine*, vol. 366, no. 18, p. 1686, 2012.
- [4] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, "A review of 3D/2D registration methods for image-guided interventions," *Medical Image Analysis*, vol. 16, no. 3, pp. 642–661, 2012.
- [5] C. A. Linte, M. Wierzbicki, J. Moore, S. H. Little, G. M. Guiraudon, and T. M. Peters, "Towards subject-specific models of the dynamic heart for image-guided mitral valve surgery," in *Medical Image Computing and Computer Assisted Intervention*, ser. Lecture Notes in Computer Science, N. Ayache, S. Ourselin, and A. Maeder, Eds. Springer Berlin Heidelberg, 2007, vol. 4792, pp. 94–101.
- [6] P. Lang, P. Seslija, M. Chu, D. Bainbridge, G. Guiraudon, D. Jones, and T. Peters, "US-fluoroscopy registration for transcatheter aortic valve implantation," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1444–1453, 2012.

- [7] J. T. Maddux, O. Wink, J. C. Messenger, B. M. Groves, R. Liao, J. Strzelczyk, S.-Y. Chen, and J. D. Carroll, "Randomized study of the safety and clinical utility of rotational angiography versus standard angiography in the diagnosis of coronary artery disease," *Catheterization and Cardiovascular Interventions*, vol. 62, no. 2, pp. 167–174, 2004.
- [8] Y. Zheng, M. John, R. Liao, J. Boese, U. Kirschstein, B. Georgescu, K. S. Zhou, J. Kempfert, T. Walther, G. Brockmann, and D. Comaniciu, "Automatic aorta segmentation and valve landmark detection in C-arm CT: Application to aortic valve implantation," in *Medical Image Computing and Computer Assisted Intervention*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6361, pp. 476–483.
- [9] J. Maintz and M. A. Viergever, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1–36, 1998.
- [10] T. Makela, P. Clarysse, O. Sipila, N. Pauna, Q. C. Pham, T. Katila, and I. E. Magnin, "A review of cardiac image registration methods," *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1011–1021, 2002.
- [11] J. V. Hajnal and D. L. Hill, *Medical Image Registration*. CRC press, 2010.
- [12] K. Murphy, B. van Ginneken, J. M. Reinhardt, S. Kabus, K. Ding, X. Deng, K. Cao, K. Du, G. E. Christensen, V. Garcia, T. Vercauteren, N. Ayache, O. Comaniciu, G. Malandain, B. Glocker, N. Paragios, N. Navab, V. Gorbunova, J. Sporrang, M. de Bruijne, X. Han, M. P. Heinrich, J. A. Schnabel, M. Jenkinson, C. Lorenz, M. Modat, J. McClelland, S. Ourselin, S. E. A. Muenzing, M. A. Viergever, D. D. Nigris, D. L. Collins, T. Arbel, M. Peroni, R. Li, G. C. Sharp, A. Schmidt-Richberg, J. Ehrhardt, R. Werner, D. Smeets, D. Loeckx, G. Song, N. J. Tustison, B. B. Avants, J. C. Gee, M. Staring, S. Klein, B. C. Stoel, M. Urschler, M. Werlberger, J. Vandemeulebroucke, S. Rit, D. Sarrau, and J. P. W. Pluim, "Evaluation of registration methods on thoracic CT: The EMPIRE10 challenge," *IEEE Transactions on Medical Imaging*, vol. 30, no. 11, pp. 1901–1920, 2011.
- [13] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, 2001.
- [14] W. Wells, P. Viola, S. Nakajima, and R. Kikinis, "Multi-modal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.
- [15] J. P. Pluim, J. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
- [16] Y. Lu, Y. Sun, R. Liao, and S. Ong, "Registration of pre-operative CT and non-contrast-enhanced C-arm CT: An application to trans-catheter aortic valve implantation (tavi)," in *Proceedings of Asian Conference on Computer Vision*, ser. Lecture Notes in Computer Science, K. Lee, Y. Matsushita, J. Rehg, and Z. Hu, Eds. Springer Berlin Heidelberg, 2013, vol. 7725, pp. 268–280.
- [17] M. Lorenzo-Valds, G. Sanchez-Ortiz, and D. Rueckert, "Atlas-based segmentation and tracking of 3D cardiac mr images using non-rigid registration," in *Medical Image Computing and Computer Assisted Intervention*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2002, vol. 2488, pp. 642–650.
- [18] L. Zhang, C. Chefd'hotel, V. Ordy, J. Zheng, X. Deng, and B. Odry, "A knowledge-driven quasi-global registration of thoracic-abdominal CT and CBCT for image-guided interventions," in *Proceedings of SPIE Medical Imaging 2013: Image-Guided Procedures, Robotic Interventions, and Modeling*, vol. 8671, Mar. 2013.
- [19] D. Neumann, S. Grbic, M. John, N. Navab, J. Hornegger, and R. Ionasec, "Robust model-based 3D/3D fusion using sparse matching for minimally invasive surgery," in *Medical Image Computing and Computer Assisted Intervention*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2013, vol. 8149, pp. 171–178.
- [20] Y. Zheng, F. Vega-Higuera, S. K. Zhou, and D. Comaniciu, "Fast and automatic heart isolation in 3D CT volumes: Optimal shape initialization," in *Machine Learning in Medical Imaging*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6357, pp. 84–91.
- [21] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuring, and D. Comaniciu, "Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and Steerable features," *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, pp. 1668–1681, 2008.
- [22] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuring, and D. Comaniciu, "Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features," in *Proceedings of International Conference on Computer Vision*. IEEE Computer Society, October 2007, pp. 1–8.
- [23] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Medical Image Analysis*, vol. 13, no. 4, p. 543, 2009.
- [24] Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," in *Tenth IEEE International Conference on Computer Vision*, vol. 2, 2005, pp. 1589–1596.
- [25] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, December 2001, pp. 1511–1518.
- [26] Z. Tu, X. S. Zhou, L. Bogoni, A. Barbu, and D. Comaniciu, "Probabilistic 3D polyp detection in CT images: The role of sample alignment," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2006, pp. 1544–1551.
- [27] Z. Tu, "Auto-context and its application to high-level vision tasks," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 1–8.
- [28] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1-3, pp. 503–528, 1989.
- [29] N. J. Mitra, N. Gelfand, H. Pottmann, and L. Guibas, "Registration of point cloud data from a geometric optimization perspective," in *Proceedings of Symposium on Geometry Processing*. ACM New York, 2004, pp. 22–31.
- [30] S. Grbic, R. Ionasec, D. Vitanovski, I. Voigt, Y. Wang, B. Georgescu, N. Navab, and D. Comaniciu, "Complete valvular heart apparatus model from 4D cardiac CT," *Medical image analysis*, vol. 16, no. 5, pp. 1003–1014, 2012.
- [31] B. K. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987.
- [32] J. Tsao, "Interpolation artifacts in multimodality image registration based on maximization of mutual information," *IEEE Transactions on Medical Imaging*, vol. 22, no. 7, pp. 854–864, 2003.
- [33] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *The ITK Software Guide*, 2nd ed., Kitware, Inc., 2005.
- [34] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "Nonrigid multimodality image registration," in *Proceedings of SPIE Medical Imaging: Image Processing*, vol. 4322, Jul. 2001, pp. 1609–1620.
- [35] B. Chapman, G. Jost, and R. v. d. Pas, *Using OpenMP: Portable Shared Memory Parallel Programming (Scientific and Engineering Computation)*, 1st ed. Cambridge, MA (US): The MIT Press, 2007. [Online]. Available: <http://mitpress.mit.edu/books/chapters/0262533022chap1.png>