

Robust Model-Based 3D/3D Fusion Using Sparse Matching for Minimally Invasive Surgery

Dominik Neumann^{1,2}, Sasa Grbic^{2,3}, Matthias John⁴, Nassir Navab³,
Joachim Hornegger¹, and Razvan Ionasec²

¹ Pattern Recognition Lab, University of Erlangen-Nuremberg, Germany

² Imaging and Computer Vision, Siemens Corporate Research, Princeton, USA

³ Computer Aided Medical Procedures, Technical University Munich, Germany

⁴ Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. Classical surgery is being disrupted 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 for intra-operative guidance. However, intra-operative modalities have limited image quality of the soft tissue and a 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 novel sparse matching approach for fusing high quality pre-operative CT and non-contrasted, non-gated intra-operative C-arm CT by utilizing robust machine learning and numerical optimization techniques. Thus, high-quality patient-specific models can be extracted from the pre-operative CT and mapped to the intra-operative imaging environment to guide minimally invasive procedures. Extensive quantitative experiments demonstrate that our model-based fusion approach has an average execution time of 2.9 s, while the accuracy lies within expert user confidence intervals.

1 Introduction

Fluoroscopy guided cardiac interventions such as endovascular stenting, atrial ablation, closure of atrial/ventricular septal defects and transcatheter valve repair or replacement are a rapidly growing market. Compared to conventional open-heart surgeries, these procedures are expected to be less invasive, reduce procedural morbidity, mortality, and intervention cost, while accelerating patient recovery. For inoperable or high-risk patients, minimally invasive surgery is the only treatment option [1]. However, without direct access to the affected anatomy, advanced imaging is required to secure a safe and effective execution.

Overlays of 3D anatomical structures based on pre-operative data [2] can potentially provide valuable information for interventional guidance when displayed on live fluoroscopy. High-quality pre-operative 3D data is routinely acquired for diagnostic and planning purposes by means of Computed Tomography, Magnetic Resonance Imaging or Echocardiography. However, direct 3D pre-operative to 2D fluoroscopy registration is difficult to solve, especially within the intra-operative setup that does not allow for user interaction or time-consuming processing.

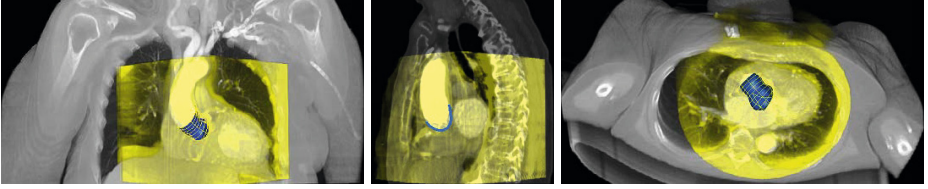


Fig. 1. Yellow: C-arm CT, gray: fused high-quality CT, blue: anatomical model

C-arm CT is emerging as a novel modality that can acquire 3D CT-like volumes directly in the OR in the same coordinate system as 2D fluoroscopy, which overcomes the need for 2D/3D registration. For most procedures, the patients are older and therefore, a safe execution of the procedure is the dominating factor [3]. Some methods work directly on the C-arm CT images [4] to extract patient-specific models for procedure guidance. However, acquiring high-quality, contrasted, and motion compensated C-arm CT images is challenging and clinicians would prefer a much simpler protocol without contrast or gating. Today, manual or semi-automatic tools are used to align the pre-operative CT and the intra-operative C-arm CT. Thus, C-arm CT serves as a bridge between 3D pre-operative data and 2D live fluoroscopy.

Multi-modal 3D/3D registration algorithms can be utilized to automate the process of aligning the pre-operative scan with the C-arm CT. [5] uses mutual information to cope with intensity inconsistencies between CT and MR and [6] presents 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 likely to converge into local minima.

We propose a method to fuse 3D pre-operative high-quality anatomical information with live 2D intra-operative imaging via non-contrasted 3D C-arm CT (Fig. 1). Robust learning-based methods are employed to automatically extract patient-specific models of target and anchor anatomies from CT. Anchor anatomies have correspondences in the pre- and intra-operative images, while target anatomies are not visible in the intra-operative image, but essential to the procedure. A novel sparse matching approach is employed to align the pre- and intra-operative anchor anatomies. Data and model uncertainties are learned and exploited during the matching process. Our method is able to cope with image artifacts, partially visible models, and does not require contrast agent.

2 Method

Our fully-automatic method (Fig. 2) fuses a pre-operative CT image \mathcal{M} (moving) with an intra-operative C-arm CT image \mathcal{F} (fixed), such that a target anatomy (aortic valve) is aligned. The process is based on an anchor anatomy \mathcal{A} (pericardium) extracted from \mathcal{M} , and a probability map $\tilde{\mathcal{F}}$ derived from \mathcal{F} . Optimal transformation parameters $\hat{\theta}$ are sought. $\theta = (\phi_{\theta}, \mathbf{t}_{\theta})$ represents a rigid transformation with Euler angles $\phi_{\theta} = (\phi_x, \phi_y, \phi_z)$ and translation $\mathbf{t}_{\theta} = (t_x, t_y, t_z)$.

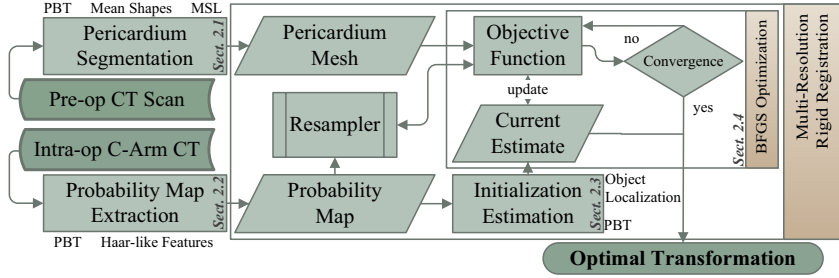


Fig. 2. Fusion workflow overview

2.1 Pericardium Segmentation

We use a recent method by Zheng et al. [7] to segment the patient-specific anchor anatomy \mathcal{A} in \mathcal{M} (Fig. 3(a)). Their technique consists of three main steps. First, the pose and scale of the heart is estimated using marginal space learning (MSL). Second, a mean shape based on manual annotations is aligned. In a third step, the parameters are refined using a boundary detector based on the probabilistic boosting tree (PBT) [8], followed by additional postprocessing.

2.2 Probability Map Extraction

A probability map $\tilde{\mathcal{F}}$ is created from \mathcal{F} by evaluating a PBT classifier on each voxel (Fig. 3(b)). The classifier was trained to robustly delineate pericardium boundary regions in C-arm CT images utilizing Haar features to achieve robustness and computational efficiency. Medical experts created a database DB of pericardium annotations $\{\mathcal{P}\}$ on 393 interventional C-arm CT scans $\{\mathcal{V}\}$. For training, positive samples were generated with regard to the position of the voxels corresponding to points in the ground-truth mesh. The negative samples for each tuple in the database $(\mathcal{V}, \mathcal{P}) \in \text{DB}$ are based on randomly selected voxels $\mathbf{v} \in \mathcal{V}$, where the distance of \mathbf{v} to all points in \mathcal{P} exceeds a certain threshold.

2.3 Initialization Estimation

In order to find a reliable initialization $\theta^0 = (\phi_{\theta^0}, \mathbf{t}_{\theta^0})$, our method recovers the offset \mathbf{t}_{θ^0} between \mathcal{M} and \mathcal{F} . We neglect the rotational error ($\phi_{\theta^0} = \mathbf{0}$), 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 almost identical positions. Our solution is based on object localization, a concept from computer vision, which we formulate as a classification problem. We evaluate a PBT classifier trained using DB from Sect. 2.2 on each voxel and choose the one that is most likely to contain the pericardium center $\mathbf{c}_{\mathcal{F}}$. Let $\mathbf{c}_{\mathcal{A}}$ be the center of \mathcal{A} , then $\theta^0 = (\phi_{\theta^0}, -\mathbf{c}_{\mathcal{A}} + \mathbf{c}_{\mathcal{F}})$. Robust detections were achieved by utilizing $\tilde{\mathcal{F}}$ as the input for training and detection, since the probability maps look similar for both contrasted and non-contrasted images, because the classifier for $\tilde{\mathcal{F}}$ was trained on both types of volumes. Thus, the method can be used with or without contrast agent injected.

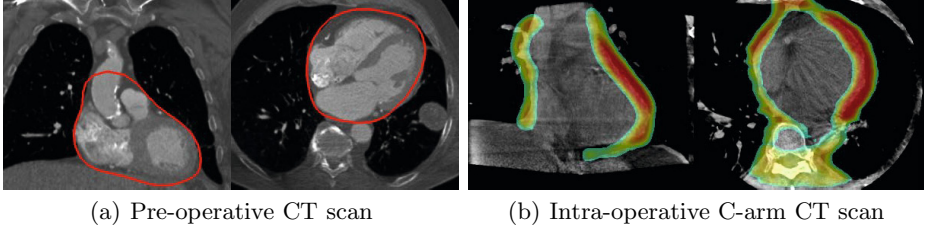


Fig. 3. Volume slices overlaid by (a) segmented pericardium, (b) probability map: red indicates high likelihood of pericardium occurrence, blue/transparent: low probability

2.4 Optimization Strategy

θ is refined within a numerical quasi-Newton optimization framework utilizing the BFGS update rule. To compensate for potential initial coarse misalignment, the concept of multi-resolution optimization is exploited by optimizing in a coarse-to-fine manner on various granularity levels of $\tilde{\mathcal{F}}$ (4, 2 and 1 mm).

Objective Function. We aim to find the transformation $\hat{\theta} = \operatorname{argmin}_{\theta} f$ that yields the minimal objective function value. f depends on \mathcal{A} and $\tilde{\mathcal{F}}$:

$$f(\theta \mid \mathcal{A}, \tilde{\mathcal{F}}) = \frac{\sum_{\mathbf{p} \in \mathcal{A}} \text{is_inside}(\theta(\mathbf{p}), \tilde{\mathcal{F}}) \cdot \psi(\theta(\mathbf{p}), \tilde{\mathcal{F}})}{\sum_{\mathbf{p} \in \mathcal{A}} \text{is_inside}(\theta(\mathbf{p}), \tilde{\mathcal{F}})}. \quad (1)$$

\mathbf{p} denotes a point in \mathcal{A} and $\mathbf{p}' = \theta(\mathbf{p})$ is that point transformed w.r.t. θ . The indicator function $\text{is_inside}(\mathbf{p}', \tilde{\mathcal{F}})$ evaluates to 1, if \mathbf{p}' is inside the physical boundaries of the volume $\tilde{\mathcal{F}}$, otherwise 0. $\psi(\mathbf{p}', \tilde{\mathcal{F}})$ returns a value inversely proportional to the probabilistic prediction at the voxel $\in \tilde{\mathcal{F}}$ where \mathbf{p}' is located. When the number of points within the boundaries of the volume is below a certain threshold, an alternative objective function prevents \mathcal{A} and $\tilde{\mathcal{F}}$ from diverging.

Gradient Computation. 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 does not allow for analytical derivations. Hence, we approximate $\tilde{\nabla} \approx \nabla f(\theta \mid \mathcal{A}, \tilde{\mathcal{F}})$ component-wise with finite differences:

$$\tilde{\nabla}_i = f(\theta + \delta^i \mid \mathcal{A}, \tilde{\mathcal{F}}) - f(\theta \mid \mathcal{A}, \tilde{\mathcal{F}}). \quad (2)$$

$\tilde{\nabla}_i$ denotes the i^{th} component of $\tilde{\nabla}$ and δ^i is a 6D offset vector where all components are zero, except for the i^{th} component δ_i^i , which is set to a particular step size. Despite its asymmetric computation scheme, the gradient is sufficiently stable in this application. For the translational components, δ_i^i equals the resolution of $\tilde{\mathcal{F}}$. This choice asserts that (2) does not evaluate to zero, since the majority of points in \mathcal{A} transformed w.r.t. θ will correspond to a different $\mathbf{v} \in \tilde{\mathcal{F}}$ than their corresponding points transformed w.r.t. $\theta + \delta^i$. Regarding the rotational components, we experimentally determined that a spacing proportional to the resolution of $\tilde{\mathcal{F}}$ (e.g. $\delta_i^i = 1^\circ$ when resolution is 1 mm) works properly.

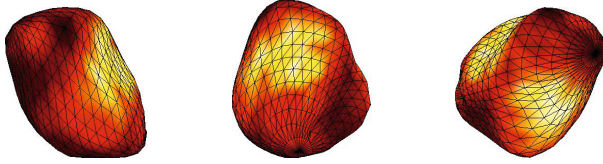


Fig. 4. Prior weights, dark colors: small weights, bright colors: high influence

While computing the translational gradient components is straightforward, rotation in 3D poses a major problem due to its inherent non-linearity and co-dependencies. We address these issues by utilizing a linearization of rotation matrices \mathbf{R} using a first order approximation $\tilde{\mathbf{R}}$ as proposed by Mitra et al. [9]. Let \mathbf{R}_θ be the 3D rotation matrix defined by $\phi_\theta = (\phi_x, \phi_y, \phi_z)$ with $\mathbf{R}_\theta^{-1} = \mathbf{R}_\theta^\top$ and $\det(\mathbf{R}_\theta) = 1$. Its first-order approximation is given by

$$\tilde{\mathbf{R}}_\theta = \begin{pmatrix} 1 & -\phi_z & \phi_y \\ \phi_z & 1 & -\phi_x \\ -\phi_y & \phi_x & 1 \end{pmatrix} \approx \mathbf{R}_\theta . \quad (3)$$

It is important to mention that $\tilde{\mathbf{R}}_\theta \approx \mathbf{R}_\theta$ only holds under small motion ($\|\phi\|_2 \rightarrow 0$). Hence, we cannot use $\tilde{\mathbf{R}}_\theta$ for large angles without introducing errors. Therefore, we compute the rotational components of $\tilde{\nabla}$ using a composite transformation. First, a point $\mathbf{p} \in \mathcal{A}$ is transformed w.r.t. the current θ using the exact Euler-angle representation to generate an intermediate point \mathbf{p}'' . Second, \mathbf{p}'' is rotated according to the minor rotation δ^i to yield \mathbf{p}' using $\tilde{\mathbf{R}}_\theta$. Altogether, we get $\mathbf{p}' = \delta^i(\mathbf{p}'') = \delta^i(\theta(\mathbf{p}))$. \mathbf{p}' constitutes the first argument for `is_inside` and ψ in (1) when computing the rotational components of $\tilde{\nabla}$ using (2).

Prior Weights. The classifier response (Sect. 2.2) is more reliable in some regions of the volumes compared to others. For instance see Fig. 3(b), 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 could be improved significantly (cf. Sect. 3.1) by incorporating prior weights $\mathbf{w} = \{w_{\mathbf{p}} \mid \mathbf{p} \in \mathcal{A}\}$. Each point \mathbf{p} in the pericardium model \mathcal{A} is assigned a patient-independent weight $w_{\mathbf{p}}$. \mathbf{w} increases the influence of those points that are likely to be located within a region of high confidence in $\tilde{\mathcal{F}}$, whereas a point that lies in a noisy or often falsely classified region gets penalized. Gradient magnitude, as well as the distance of a point to the upper and lower boundary of the pericardium were most useful for reliability predictions. Based on these observations and DB from Sect. 2.2, we computed values for \mathbf{w} (Fig. 4). Incorporation of \mathbf{w} into (1) yields

$$f(\theta \mid \mathcal{A}, \tilde{\mathcal{F}}, \mathbf{w}) = \frac{\sum_{\mathbf{p} \in \mathcal{A}} \text{is_inside}(\theta(\mathbf{p}), \tilde{\mathcal{F}}) \cdot \psi(\theta(\mathbf{p}), \tilde{\mathcal{F}}) \cdot w_{\mathbf{p}}}{\sum_{\mathbf{p} \in \mathcal{A}} \text{is_inside}(\theta(\mathbf{p}), \tilde{\mathcal{F}})} . \quad (4)$$

3 Experimental Results

We compiled a set of 88 corresponding clinical CT and C-arm CT volumes, each with an isotropic resolution of 1 mm. 18 image pairs are native, while 70 were acquired with contrast agent injected. Medical experts annotated the pericardium in each volume, 43 studies include annotations of the aortic valve (AV). The data is organized in a database $DB_{cl} = \{(\mathcal{V}_i, \check{\mathcal{V}}_i, \mathcal{P}_i, \check{\mathcal{P}}_i, \mathcal{R}_i, \check{\mathcal{R}}_i) \mid i = 1 \dots 88\}$. \mathcal{V}_i , \mathcal{P}_i and \mathcal{R}_i denote the i^{th} CT volume, its pericardium and AV annotation. Analogously, $\check{\mathcal{V}}_i$, $\check{\mathcal{P}}_i$ and $\check{\mathcal{R}}_i$ denote these structures for the C-arm CT acquisition.

The results are based on a symmetric mesh-to-mesh distance metric ε :

$$\varepsilon(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \left(\sum_{\mathbf{p} \in \mathcal{X}} \min_{\Delta \in \mathcal{Y}} \varepsilon_{p2t}(\mathbf{p}, \Delta) + \sum_{\mathbf{p} \in \mathcal{Y}} \min_{\Delta \in \mathcal{X}} \varepsilon_{p2t}(\mathbf{p}, \Delta) \right), \quad (5)$$

where \mathcal{X} , \mathcal{Y} are triangulated meshes and $\varepsilon_{p2t}(\mathbf{p}, \Delta)$ denotes the point-to-triangle distance. Note that (5) might underestimate misalignment tangential to the surface. However, it is a fairly natural measure closely resembling visual assessment.

3.1 Quantitative Evaluation on Clinical Data

We evaluated our method on DB_{cl} with both, prior weights (Sect. 2.4) enabled and disabled. In Table 1 (left), the left and right columns for each scenario show error statistics resulting from a comparison of the optimally transformed segmented pericardium $\hat{\theta}(\mathcal{A})$ and the ground-truth annotation in the C-arm CT volume $\check{\mathcal{P}}$, and results for a comparison of the transformed CT-based aortic valve $\hat{\theta}(\mathcal{R})$ and its C-arm based annotation $\check{\mathcal{R}}$, respectively. With prior weights incorporated, a registration accuracy of 5.60 ± 1.81 mm measured between the anchor anatomy (pericardium) is achieved. Furthermore, with an error of 4.63 ± 1.90 mm, the target anatomy (AV) is aligned very well. When the patient-independent weighting is ignored, the mean errors increase significantly by more than 35% regarding the pericardium and almost 55% for the AV. One reason is that more outliers are generated, having a strong influence on the overall errors. Hence, the prior information improves the fusion performance significantly.

3.2 Comparison to State-of-the-Art Fusion

We compared our model-to-image registration to an image-to-image fusion approach utilizing ITK, a state-of-the-art medical imaging library. Multi-resolution optimal parameters $\hat{\theta}_*$ were obtained by an optimizer for rigid versor transformations. We customized the maximum and minimum step lengths adaptively for each resolution, the maximum number of iterations was set to 200, and we initialized the procedure by aligning the center of both volumes. The similarity metric is based on mutual information. Quantitative results are presented in Table 1 (right). We excluded cases where the method failed ($\varepsilon > 20$ mm). Using this error-threshold rule, we disregarded almost half of all measurements. However, since this framework is not specifically designed to align the pericardium, the large number of fail cases is understandable. Many of the failures occurred for

Table 1. Errors [mm] with/without prior weights (left) and for state-of-the-art (right)

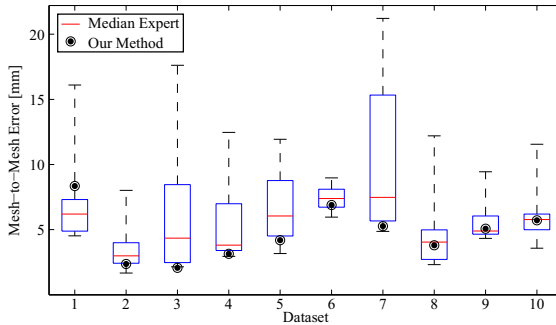
Method	Prior		No Prior		State-of-the-Art		Prior	
$\varepsilon(\cdot, \cdot)$	$\hat{\theta}(\mathcal{A}), \check{\mathcal{P}}$	$\hat{\theta}(\mathcal{R}), \check{\mathcal{R}}$	$\hat{\theta}(\mathcal{A}), \check{\mathcal{P}}$	$\hat{\theta}(\mathcal{R}), \check{\mathcal{R}}$	$\hat{\theta}_*(\mathcal{P}), \check{\mathcal{P}}$	$\hat{\theta}_*(\mathcal{R}), \check{\mathcal{R}}$	$\hat{\theta}(\mathcal{P}), \check{\mathcal{P}}$	$\hat{\theta}(\mathcal{R}), \check{\mathcal{R}}$
#Studies	88	43	88	43	47	21	47	21
Mean	5.60	4.63	7.57	7.17	7.19	6.33	5.03	4.45
Std	1.81	1.90	4.38	7.10	4.86	4.71	1.80	2.14
Median	5.29	4.64	6.91	5.52	4.96	4.56	5.02	4.32

images with significant differences in the size of the field of view between the CT and the C-arm CT scan. Still, a mean error of 7.19mm is substantially larger than the error of our method, with a mean of 5.03mm on the same dataset.

3.3 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, we compared our method to individual performances of a group of 10 experts on 10 datasets.

Let θ_{ji} be the i^{th} expert’s transform for the j^{th} pair of volumes. We compare the fit of the θ_{ji} -transformed CT pericardium to the ground-truth C-arm CT annotation, i.e. we evaluate $\varepsilon(\check{\mathcal{P}}_j, \theta_{ji}(\mathcal{P}_j))$. Our automated method exhibits lower errors than the median expert in 80% of all cases and shows high robustness with no outliers (see Fig. 5). There exists only one dataset, where the automatic fusion is inferior to more than 75% of the experts. Moreover, the experts’ manual fusion time per data pair ranged from two to five minutes, while our method takes only 2.9 ± 0.4 s (in DB_{cl}) on average, which means a speedup of up to 99%.

**Fig. 5.** Inter-user variability, box edges indicate 25th and 75th percentiles of user errors

4 Conclusion

We presented a method to fuse pre-operative CT and intra-operative 3D C-arm CT data. A novel sparse matching approach is employed to align the

pre-operative anchor anatomy to the intra-operative setting. Data and model uncertainties are learned and exploited. Quantitative and qualitative evaluation demonstrate an accurate mapping of the target anatomy to the intra-operative modality. In direct comparison with a state-of-the-art registration framework, our method outperforms it significantly in robustness and accuracy. Furthermore, an inter-user variability study confirms that the accuracy of our fully-automatic method lies within the confidence interval of the expert group while reducing the duration of the alignment process from five minutes to three seconds. Thus, comprehensive patient-specific models can be estimated from high-contrast CT and fused into the imaging environment of operating rooms to facilitate guidance in cardiac interventions without tedious and time-consuming manual interactions.

References

1. Leon, M.B., Smith, C.R., Mack, M., Miller, D.C., Moses, J.W., Svensson, L.G., Tuzcu, E.M., Webb, J.G., Fontana, G.P., Makkar, R.R., Brown, D.L., Block, P.C., Guyton, R.A., Pichard, A.D., Bavaria, J.E., Herrmann, H.C., Douglas, P.S., Petersen, J.L., Akin, J.J., Anderson, W.N., Wang, D., Pocock, S.: Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N. Engl. J. Med.* 363, 1597–1607 (2010)
2. Grbic, S., Ionasec, R., Vitanovski, D., Voigt, I., Wang, Y., Georgescu, B., Navab, N., Comaniciu, D.: Complete valvular heart apparatus model from 4D cardiac CT. *Med. Image Anal.* 16(5), 1003–1014 (2012)
3. Maddux, J.T., Wink, O., Messenger, J.C., Groves, B.M., Liao, R., Strzelczyk, J., Chen, S.Y., Carroll, J.D.: Randomized study of the safety and clinical utility of rotational angiography versus standard angiography in the diagnosis of coronary artery disease. *Catheter. Cardiovasc. Interv.* 62(2), 167–174 (2004)
4. Zheng, Y., John, M., Liao, R., Boese, J., Kirschstein, U., Georgescu, B., Zhou, S.K., Kempfert, J., Walther, T., Brockmann, G., Comaniciu, D.: Automatic aorta segmentation and valve landmark detection in C-arm CT: Application to aortic valve implantation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) *MICCAI 2010, Part I. LNCS*, vol. 6361, pp. 476–483. Springer, Heidelberg (2010)
5. Wells, W., Viola, P., Nakajima, S., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. *Med. Image Anal.* 1(1), 35–51 (1996)
6. Lorenzo-Valdés, M., Sanchez-Ortiz, G.I., Mohiaddin, R.H., Rueckert, D.: Atlas-based segmentation and tracking of 3D cardiac MR images using non-rigid registration. In: Dohi, T., Kikinis, R. (eds.) *MICCAI 2002, Part I. LNCS*, vol. 2488, pp. 642–650. Springer, Heidelberg (2002)
7. Zheng, Y., Vega-Higuera, F., Zhou, S.K., Comaniciu, D.: Fast and automatic heart isolation in 3D CT volumes: Optimal shape initialization. In: Wang, F., Yan, P., Suzuki, K., Shen, D. (eds.) *MLMI 2010. LNCS*, vol. 6357, pp. 84–91. Springer, Heidelberg (2010)
8. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: *Tenth IEEE Int'l Conf. on Comp. Vision*, vol. 2, pp. 1589–1596. IEEE (2005)
9. Mitra, N.J., Gelfand, N., Pottmann, H., Guibas, L.: Registration of point cloud data from a geometric optimization perspective. In: *Proc. of the 2004 Eurographics/ACM SIGGRAPH Symp. on Geom. Proc.*, pp. 22–31. ACM, New York (2004)