Model-Based Fusion of Multi-Modal Volumetric Images: Application to Transcatheter Valve Procedures

Saša Grbić[1,3], Razvan Ionasec[1], Yang Wang[1], Tommaso Mansi[1], Bogdan Georgescu[1], Matthias John[2], Jan Boese[2], Yefeng Zheng[1], Nassir Navab[3], and Dorin Comaniciu[1]

[1] Image Analytics and Bioinformatics, Siemens Corporate Research, Princeton, USA
[2] Siemens AG, Healthcare Sector, Forchheim, Germany
[3] Technical University Munich, Germany

Abstract. Minimal invasive procedures such as transcatheter valve interventions are substituting conventional surgical techniques. Thus, novel operating rooms have been designed to augment traditional surgical equipment with advanced imaging systems to guide the procedures. We propose a novel method to fuse pre-operative and intra-operative information by jointly estimating anatomical models from multiple image modalities. Thereby high-quality patient-specific models are integrated into the imaging environment of operating rooms to guide cardiac interventions. Robust and fast machine learning techniques are utilized to guide the estimation process. Our method integrates both the redundant and complementary multimodal information to achieve a comprehensive modeling and simultaneously reduce the estimation uncertainty. Experiments performed on 28 patients with pairs of multimodal volumetric data are used to demonstrate high quality intra-operative patient-specific modeling of the aortic valve with a precision of 1.09mm in TEE and 1.73mm in 3D C-arm CT. Within a processing time of 10 seconds we additionally obtain model sensitive mapping between the pre- and intra-operative images.

1 Introduction

There has been a major trend in cardiac therapy towards minimally invasive transcatheter procedures to reduce the side effects of classical surgical techniques. Instead of full sternotomy, instruments and devices are introduced through small incisions, advanced through vessels and positioned to perform various procedures [1]. Without direct access and view to the affected structures those interventions are usually performed in so-called Hybrid ORs, operating rooms outfitted with advanced imaging equipment. Thus, procedures such as the Transcatheter Aortic Valve Replacement (TAV) are permanently guided via real-time intra-operative images provided by C-arm X-ray and Transesophageal Echocardiography systems [2].

Traditionally the field of medical image analysis has been focusing on the construction of patient-specific anatomical models from well established diagnostic modalities (e.g. CT and MR) to aid disease analysis and treatment planning [3]. In the context of valvular disease management, the authors in [4] and [5] proposed the modeling of the aortic valve from cardiac CT. Models of the mitral valve from MR have been proposed in [6]. However, these methods have not been developed to cope with the reduced quality and contrast characteristic to intra-operative images, hence their usage is strictly limited to pre-operative decision making.

Delivering high-quality models into the operating room to guide cardiac therapy will be a major focus of future clinical applications. The fusion of pre-operative models with intra-operative images for mitral valve replacement has been proposed by [7]. The pre-op model of the mitral valve was annotated from CT data and registered into the intra-op MR and echocardiography images. Major limitation is the required tracking equipment and semi-automatic delineation of the mitral annuls. Alternatively, multi-modal image registration has been proposed to fuse multi-modal data. In [8] the mutual information is used as the metric to cope with the intensity inconsistencies between CT and MR. Recently [9] proposed a method to register intracardiac echocardiography (ICE) with pre-op CT images. Hereby a recomputed simulated ultrasound reflection volume was extracted from the CT data and registered using standard techniques to the ICE volume. [10] proposed a novel similarity metric which incorporates knowledge of previously registered images. In [11] an atlas-based approach was presented to track the myocardium and left and right ventricles from MR data. The registration is used to align the cardiac atlas to the patient data. However, these methods are computationally expensive, and without the appropriate guidance of a shape prior likely to converge into local minima.

As we seek to provide both, the fusion of pre- and intra-operative imaging and patient-specific models of relevant anatomical structures, the field of simultaneous registration and segmentation is important to our work. In [12] a probabilistic framework was proposed where registration is performed jointly with segmentation. It was applied to the segmentation of brain tissues and their substructures in uni-modal MR data. [13] recently proposed a method to jointly segment the prostate and provide registration in MR. It introduced point sets to allow fast initial registration. However most of these methods suffer from long run-times, as the problem of registration and segmentation is trying to be solved simultaneously, and the necessity of manual initialization.

We propose a novel method to fuse pre-operative and intra-operative information by jointly estimating anatomical models from multiple image modalities. Thereby high-quality patient-specific models are integrated into the imaging environment of operating rooms to guide cardiac interventions. Robustness and efficiency are achieved by relying on machine learning techniques to drive the joint estimation process whereby similarities between multiple modalities are exploited. Statistical models of the anatomy are utilized within the probabilistic estimation framework to ensure physiologically compliant results. The main

benefits of our method are: 1) Completeness - by exploiting the complementary information from multiple modalities, 2) Robustness - by exploiting the redundant information from multiple modalities to reduce the estimation uncertainty, and 3) Fusion - by obtaining a model-sensitive integration of the pre-operative and intra-operative modalities.

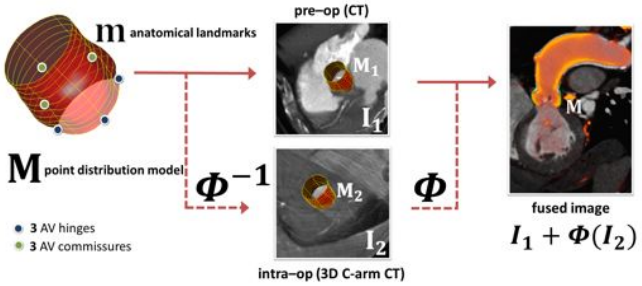

Fig. 1. Diagram of the problem formulation showing the surface model M, anatomical landmarks m, transformation ϕ to map the intra-op image I_2 to the pre-op data I_1.

2 Problem formulation

Our goal is to estimate a 3D anatomical patient-specific model M from volumetric multi-modal datasets I_1 and I_2, where I_1 is the pre-op and I_2 the intra-op image, and the transformation ϕ which maps the intra-op model M_2 to the pre-op model M_1 (see Fig 1).

$$(\hat{\phi}, \hat{M}) = \arg\max_{M,\phi} \log P(M, \phi \,|I_1, I_2\,) \tag{1}$$

ϕ is composed of an affine- A and a non-linear warping transformation D, $\phi = D\ A$. D is modeling the small deformation of M due to respiration and uncertainties in the acquisition phase between the pre- and intra-op data. The model M is represented as a point distribution model. Using the transformation ϕ the pre- M_1 and intra-op M_2 models can be computed: $M = M_1$, $M = D\ A\ M_2$ and $M_2 = A^{-1}\ D^{-1}\ M$.

3 Method

In general finding an optimal solution to Eqn. 1 is difficult and has high computational cost therefore we approximate the problem by expanding the formulation and exploiting independencies. In addition a shape constraint term is added to restrict the estimated model M in a shape space built from a database of annotations.

$$(\hat{\phi}, \hat{M}) = \arg\max_{M,\phi} \log \left(P(M\,|I_1\,) \cdot P(M\,|\phi(I_2)\,) \cdot P(M\,|I_1, \phi(I_2)\,) \cdot P(M, \phi\,|\mu, \Sigma\,)\right) \tag{2}$$

All the probabilities in our formulations are modeled using robust learning based algorithms. The first $P(M\,|I_1\,)$ and the second term $P(M\,|\phi(I_2))$ define the independent model estimations in the multi-modal images I_1 and I_2. As proposed in [4] a classifier is trained using the probabilistic boosting tree and Haar-features to estimate the posterior probability. The best model parameters for M are selected based on a joint probability term $P(M\,|I_1, \phi(I_2))$ explained in chapter 3.1. The transformation ϕ is modeled as a warping transform with Gaussian radial basis functions. The last term $P(M, \phi\,|\mu, \Sigma)$ symbolizes a regularization of the shape M and the transformation ϕ based on the learned statistical shape model defined as a Gaussian distribution with the mean μ and the covariance matrix Σ learned from manual annotations. Both the affine A and the non-linear transformation D are updated in this stage. A bias is applied towards the pre-op model $M = M_1$ as the model estimation is more robust in the pre-op images. In our case I_1 represents the CT image and I_2 the TEE and 3D C-arm CT image.

In this work we focus on the estimation of the aortic valve model. The valve is modeled hierarchically using two layers. On the coarse level it is represented as a landmark model m with 6 points (3 commissures and 3 hinges). They define the most important morphological and functional properties of the valve. The finer level is defined as a point distribution model M with 1440 points spread along a 36×20 parametric grid.

3.1 Similarity learning

The joint term $P(M\,|I_1, \phi(I_2)\,)$ in Eqn. 2 exploits the similarities between the models from the multi-modality images. Similarity functions proposed in the current literature, such as mutual information or cross correlation, could be used but as mentioned in [14] learning the similarity for a specific problem yields better performance.

We employ a boosting framework in order to train a cascade of strong classifiers. Each strong classifier $\boldsymbol{F}_{strong}$ consists of k weak classifiers $\boldsymbol{F}_{weak}$ which learn the similarity between pairs of image patches $I_{S1} \in I_1$ and $I_{S2} \in I_2$, $\boldsymbol{F}_{weak}(I_{S1}, I_{S2})$. The weak learners are constructed based on Haar-like features extracted locally from rectangular patches I_{S1} and I_{S2} from image slices sampled perpendicular to the tubular aortic root surfaces M_1 and M_2. The patch size is fixed for both modalities.

The weak learner is modeled as a 2D piecewise constant function defined on a 2D feature space by the feature responses of $h(I_{S1})$ and $h(I_{S1})$. The 2D feature space is separated in equal rectangular non-overlapping regions. Therefore we quantize the feature responses from both modalities in 64×64 bins whereby the values are scaled between the minimum and maximum feature responses $h(I_{S1})$ and $h(I_{S1})$.

$$\boldsymbol{F}_{weak}(I_{S1}, I_{S2}) = \sum_{b=1}^{B} \sum_{c=1}^{C} \beta_{b,c}\; R_{b,c}\,[h(I_{S1}) \times h(I_{S2})] \tag{3}$$

where B and C are the bin numbers for the feature responses in each modality and $\beta_{b,c}$ symbolizes the constant associated with the region $R_{b,c}$ representing

a bin in the 2D feature space. As in [15] the optimal weights $\beta_{b,c}$ would be determined by fitting a least-squares regression function. During detection a probability for each weak classifier is evaluated by extracting Haar-features from pairs of image patches. The features are assigned to a bin $R_{b,c}$ based on the feature response and multiplied with the corresponding weight $\beta_{b,c}$.

A cascade of l strong classifiers $\boldsymbol{F}_{strong}$ is trained in order to determine the posterior probability $P(M\,|I_1, \phi(I_2)\,) = \boldsymbol{S}(I_{S1}, I_{S2})$ of the similarity function.

3.2 Model-Based Fusion Approach

The first stage in our hierarchical model estimation algorithm consists of pre-aligning the multi-modal images using the anatomical landmarks. The affine transformation A is estimated by obtaining a least-squares solution based on the independently detected landmarks m_1 from the image I_1 and m_2 from the image I_2. The landmark detectors are trained using the probabilistic boosting tree classifier and Haar-like features. The surface M is initialized by learning a correlation model between measurements extracted from the landmarks m_1 and the point distribution model M, as described in [16]. The nonlinear warping transformation D is set to identity. Based on A the model M can be projected to the image I_2.

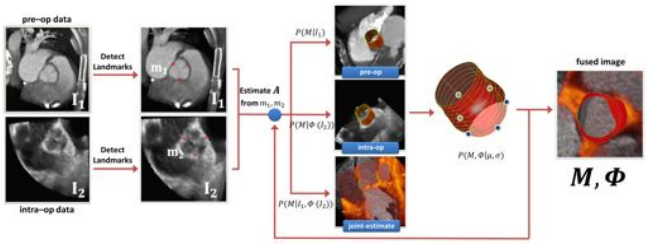

Fig. 2. Diagram showing the model based fusion approach for the estimation of the model M and the transformation ϕ.

In the optimization phase we apply an iterative approach. We sample candidates N_1 and N_2 along the surfaces normals of M_1 and M_2, and evaluate the probability $P(M\,|I_1\,)$ for each candidate $n_1 \in N_1$ and $P(M\,|\phi(I_2)\,)$ for each point $n_2 \in N_2$ using the learned detectors. The joint probability $P(M\,|I_1, \phi(I_2)\,)$ is determined by training a boosting classifier, as mentioned in chapter 3.1, to evaluate pairs of candidates. A cross product of the candidates $N_1 \times N_2$ is constructed and the highest probable candidate pair (n_i, n_j) is selected by multiplying the single modality probabilities with the joint term.

$$(n_i, n_j) = \arg\max_{n_i, n_j} \log\left(P(n_i\,|I_1\,) \cdot P(n_j\,|\phi(I_2)\,) \cdot P(n_i, n_j\,|I_1, \phi(I_2)\,)\right) \quad (4)$$

The estimated candidate pairs are used to update the models M_1 and M_2. The second step of the iteration involves calculating the posterior probability

$P(M, \phi \,|\mu, \Sigma)$ of M and ϕ based on the learned statistical shape models. This could be perceived as a regularization to the shape M. Thereby M_1 is projected to the PCA shape space using the largest 40 eigenvectors. ϕ is updated by computing the rigid transformation R based on the posterior probability of the pairs (n_i, n_j). D is updated by obtaining a least-squares solution to the warping transformation $\hat{D} = \arg\min \left\| T\ M_2 - D^{-1} M_1 \right\|^2$ using radial basis functions. Thereby the number of control points is much smaller than the number of shape points M. The algorithm converges in a small number of iterations. Figure 2 demonstrates the complete estimation approach.

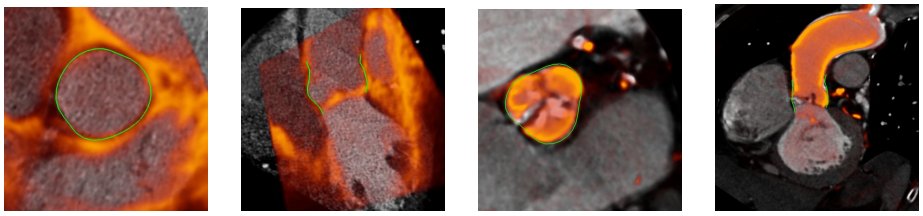

Fig. 3. Example of the joint aortic valve model estimation from pre- and intra-op volumetric data. The left 2 images show fused CT-TEE data sets and the right 2 images show fused CT-3D C-arm CT data. The mapping of the intra-op image I_2 to the pre-op image I_1 is done by the estimated non-linear transform ϕ.

4 Experimental Results

The most relevant intra-operative modalities with 3D capabilities (3D C-arm CT and TEE) in the OR environment were incorporated for evaluation. In total 56 volumes, 13 pairs of CT-TEE data sets and 15 pairs of CT-3D C-arm CT data pairs were selected to demonstrate the effectiveness of our method. This dataset was solely used for evaluation and not included in training. The ground-truth annotations were obtained from clinical experts by manually placing the anatomical landmarks in the pre- (m_1) and intra-op (m_2) images and finally delineating the aortic valve surfaces M_1 and M_2.

As our algorithm depends on the automatic detection of the anatomical landmarks m_1 and m_2 during the initialization step in order to estimate the affine transform A we evaluate their detection performance on the test dataset. For training 160 separate landmarks annotations in CT, 320 in TEE and 192 in 3D C-arm CT were used to train the landmark detectors. The error is computed as the Euclidian distance between the automatic estimation and the expert annotation. For the hinges we obtain an error of $2.40 \pm 0.81mm$ in CT, $2.56 \pm 0.71mm$ in TEE and $2.30 \pm 1.56mm$ in 3D C-arm CT and for the commissures $2.74 \pm 1.01mm$ in CT, $3.31 \pm 1.55mm$ in TEE and $2.98 \pm 1.44mm$ in 3D C-arm CT.

The mesh-to-mesh error was computed between the ground-truth annotations and the detected models in order to obtain quantitative results for the automatic

surface estimation. Results shown in table 1 confirm that our model-based fusion estimation approach yields the best results.

Table 1. System precision for aortic valve surface model estimation in CT, TEE and 3D C-arm CT. Comparison between our novel model-based fusion approach and single modality estimations.

	single modality estimation			fusion approach		
	Mean	STD	Median	Mean	STD	Median
CT-TEE [mm]	1.22	0.23	1.13	1.09	0.22	1.10
CT-3D C-arm CT [mm]	1.96	0.54	1.99	1.73	0.49	1.79

In transcatheter aortic valve procedures both the selection of the appropriate stent size but also the positioning of it in the intra-op data has clinical significance. However in 3D C-arm CT the aortic valve annulus is not visible as the contrast is injected at the cusp area. Fusing the 3D C-arm image with pre-op CT data would allow the physician to properly examine the annulus area and enable accurate positioning of the stent during the procedure. We evaluate the error for the aortic valve annulus ring circumference, extracted from the estimated aortic valve model M, by comparing the result of the independent detection in 3D C-arm CT image and our model-based fusion approach. Quantitative and qualitative results are shown in figure 4.

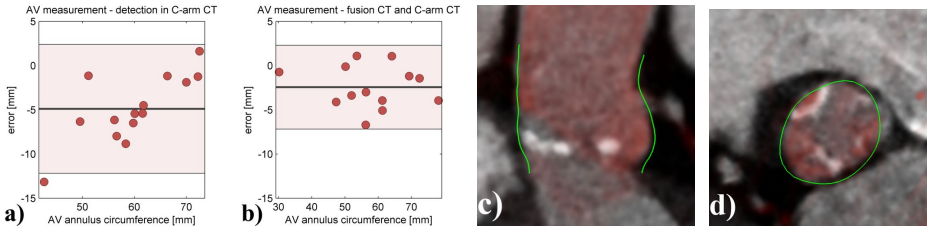

Fig. 4. Bland-Altman plots for the aortic valve annulus circumference measurement extracted from the model M with (a) independent detection in 3D C-arm CT and (b) fusion of pre-op CT and 3D C-arm CT. (c) and (d) are showing short and long axis views of the model M and the fused pre- and intra-op images $I_1 + \phi(I_2)$.

5 Conclusion

In this paper, we propose a novel approach to estimate comprehensive patient specific models of the aortic valve by model-sensitive fusing of multimodal pre- and intra-operative data. Fast and robust machine learning techniques are employed during the estimation exploiting redundant and complementary informa-

tion from the multimodal images. Thereby high-quality patient-specific models are integrated into the imaging environment of operating rooms to guide cardiac interventions. Comprehensive quantitative and qualitative experiments on the aortic valve modeling demonstrate the effectiveness of our approach with an accuracy of 1.09mm in CT-TEE and 1.73mm in CT-3D C-arm CT.

References

1. Walther T, Chu MW, M.F.: Transcatheter aortic valve implantation: time to expand? In: Curr Opin Cardiology. Volume 23. (2008) 111–6
2. Vahanian, A., Alfieri, O., Al-Attar, N.: Transcatheter valve implantation for patients with aortic stenosis. In: Eur J Cardiothorac Surg. Volume 34. (2008) 1–8
3. Frangi, A.F., Rueckert, D., Duncan, J.S.: Three-dimensional cardiovascular image analysis. IEEE Trans. Med. Imaging **21**(9) (2002) 1005–1010
4. Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Higuera, F., Navab, N., Comaniciu, D.: Patient-specific modeling and quantification of the aortic and mitral valves from 4-D cardiac CT and TEE. TMI **29**(9) (2010) 1636–51
5. Waechter, I., Kneser, R., Korosoglou, G., Peters, J., Bakker, N.H., Van Der Boomen, R., Weese, J.: Patient specific models for planning and guidance of minimally invasive aortic valve implantation. In: MICCAI. (2010) 526–533
6. Conti, C.A., Stevanella, M., Maffessanti, F., Trunfio, S., Votta, E., Roghi, A., Parodi, O., Caiani, E.G., Redaell, A.: Mitral valve modelling in ischemic patients: Finite element analysis from cardiac magnetic resonance imaging. In: Computing in Cardiology. (2010) 1059–1062
7. Linte, C., Wierzbicki, M., Moore, J., Little, S.H., Guiraudon, G.M., Peters, T.M.: Towards subject-specific models of the dynamic heart for image-guided mitral valve surgery. In: MICCAI. (2007) 94–101
8. Wells, W.M., Viola, P., Kikinis, R.: Multi-modal volume registration by maximization of mutual information. MedIA **1**(1) (1996) 35–51
9. Wein, W., Camus, E., John, M., Diallo, M., Fahrig, R., Khamene, A., Xu, C.: Towards Guidance of Electrophysiological Procedures with Real-Time 3D Intracardiac Echocardiography Fusion to C-arm CT. MICCAI (2009) 9–16
10. Zollei, L., Wells, W.: Multi-modal image registration using dirichlet-encoded prior information. In: Int. Workshop on Biomed. Image Registration. (2006) 34–42
11. Lorenzo-Valds, M., Sanchez-Ortiz, G.I., Mohiaddin, R., Rueckert, D.: Atlas-based segmentation and tracking of 3d cardiac mr images using non-rigid registration. In: MICCAI. Volume 2488. (2002) 642–650
12. Pohl, K.M., Fisher, J., Grimson, L., Kikinis, R., Wells, W.M.: A Bayesian model for joint segmentation and registration. NeuroImage **31**(1) (May 2006) 228–39
13. Gao, Y., R, R.S., Fichtinger, G., Tannenbaum, A.: A Coupled Global Registration and Segmentation Framework With Application to Magnetic Resonance Prostate Imagery. TMI **29**(10) (2010) 1781–1794
14. Bronstein, Michael, A.M., Science, C., Israel, T., Michel, F., Paragios, N., Galen, E., Ile-de france, I.S.: Data Fusion through Cross-modality Metric Learning using Similarity-Sensitive Hashing. CVPR (2010) 3594 – 3601
15. Zhou, S., Shao, J., Georgescu, B., Comaniciu, D.: Boostmotion: Boosting a discriminative similarity function for motion estimation. In: CVPR. (2006) 1761–1768
16. Grbić, S., Ionasec, R., Vitanovski, D., Ingmar, V., Wang, Y., Georgescu, B., Navab, N., Comaniciu, D.: Complete valvular heart apparatus model from 4D cardiac CT. MICCAI 2010 **13**(Pt 1) (2010) 218–26