

6DoF Catheter Detection, Application to Intracardiac Echocardiography*

Kristóf Ralovich^{1,2}, Matthias John³, Estelle Camus⁴,
Nassir Navab², and Tobias Heimann¹

¹ Siemens AG, Imaging and Computer Vision, Germany

² Technical University of Munich, Germany

³ Siemens AG, Healthcare Sector, Forchheim, Germany

⁴ Siemens Medical Solutions Inc., Mountain View, CA, USA

Abstract. Hybrid imaging systems, consisting of fluoroscopy and echocardiography, are increasingly selected for intra-operative support of minimally invasive cardiac interventions. Intracardiac echocardiography (ICE) is an emerging modality with the promise of removing sedation or general anesthesia associated with transesophageal echocardiography (TEE).

We introduce a novel 6 degrees of freedom (DoF) pose estimation approach for catheters (equipped with radiopaque ball markers) in single X-Ray fluoroscopy projection and investigate the method's application to a prototype ICE catheter. Machine learning based catheter detection is implemented in a Bayesian hypothesis fusion framework, followed by refinement of ball marker locations through template matching. Marker correspondence and 3D pose estimation are solved through iterative optimization.

The method registers the ICE volume to the C-arm coordinate system. Experiments are performed on synthetic and porcine *in-vivo* data. Target registration error (TRE), defined in the echo cone, is the basis of our preliminary evaluation. The method reached 8.06 ± 7.2 mm TRE on 703 cases.

Potential uses of our hybrid system include structural heart disease interventions and electrophysiological mapping or catheter ablation procedures.

1 Introduction

Image guidance is a key facilitator of percutaneous cardiac interventions. Volume intracardiac echocardiography (volume ICE) is an emerging, real-time, non-ionizing perioperative modality with the promise of removing sedation or general anesthesia (GA) associated with transesophageal Echocardiography (TEE) at comparable image quality. High-risk cardiac patients tend to have a lower tolerance towards GA and intubation, thus avoiding these factors enables earlier hospital discharge. While echocardiography is gaining a foothold in

* Disclaimer: The outlined concepts are not commercially available. Due to regulatory reasons its future availability cannot be guaranteed.

therapy due to excellent soft tissue contrast, most interventionalists are trained on fluoroscopy. Co-registration (fusion) of intra-operative echocardiography and C-arm X-Ray fluoroscopy (XRF) aims to combine the complementary benefits of the two imaging systems for use in hybrid operating rooms.

For this reason, we introduce a novel 2D detection approach for catheters (equipped with radiopaque ball markers) from fluoroscopy and investigate 3D pose recovery of a prototype ICE catheter using the extracted markers.

2 Background and Previous Works

Catheter detection in XRF is a well researched issue, automatic methods exist to extract various instruments [6,7] at interactive rates. [11] describes a system for tracking an IVUS catheter in cine X-Ray to perform image registration. In other ultrasound-angiography registration scenarios the 3D pose of the transducer is required, too: the CartoSound [12] system employs electromagnetic tracking built into the C-arm to localize the ICE catheter. Robotic self-tracking [8] has also been investigated for echo guidance. Often such external hardware setups are not feasible, and recently purely image based systems were introduced for registration [5] [9] of TEE and X-Ray.

During cardiac interventions, multiple catheters, wires and tools are commonly visible in XRF. The C-arm may have arbitrary oblique angulations, dye injection could change contrast conditions. As opposed to TEE that is always directed down the esophagus, ICE might arrive from different directions to the target area, unrestricted in 6 degrees of freedom (DoF). In summary, it is expected that fluoroscopic images will not provide global context to support the detection task.

The task of finding the 6 DoF pose of a rigid object from perspective projection of attached ball markers belongs to the Perspective-n-Point (*pnp*) problem. Even though closed form solution exists, iterative methods [2,3,4] tend to achieve more robust matches with noisy 2D marker projections. These methods operate under the assumption, that 2D-3D point correspondences are known *a priori*, while [1] solves the simultaneous pose and correspondence estimation problem. In our case, point correspondences are not known, and need to be determined.

3 Method

To facilitate our needs, we have developed a prototype ICE catheter and fitted it with radiopaque ball markers (beads, **Fig. 1**). Each marker is assigned a unique virtual identifier ($B_i, i \in \{1, \dots, 6\}$), the empirical design of five distal and single proximal ball markers is shown in **Fig. 2** together with the phased array (*PHA*) transducer. The 3D catheter pose is parameterized as (x', y', z') position and rotation: around “long axis” of *PHA* (α_{Roll}), out-of-plane (β_{Pitch}) and in-plane (γ_{Yaw}).

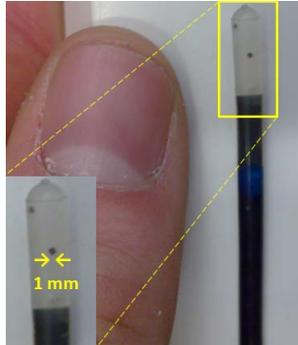


Fig. 1. Illustration of ball marker sizes in a prototype ICE catheter for reference

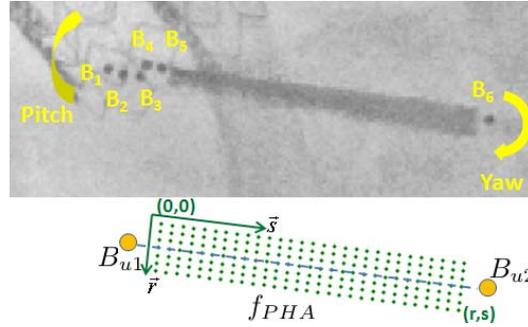


Fig. 2. (a) Typical view of ICE prototype in XRF. (b) Model used for 2D localization: steerable grid pattern defined by positions of $\{B_{u1}, B_{u2}\}$

3.1 Overview

Our automatic 2D catheter detection algorithm (**Fig. 3**) consists of (i) permissive ball marker detector and likelihood measure of *PHA* foreground/background, (ii) robust hypothesis fusion strategy and (iii) accurate hypothesis refinement.

In the last part, we discuss technical details for registration of ICE and XRF. Building on the catheter detection method, we use 2D location of ball markers from a single projection X-Ray for point correspondence determination and 3D pose estimation of ICE.

Throughout the method, we assume that the tip of the ICE catheter (including the ball markers and transducer) is rigid.

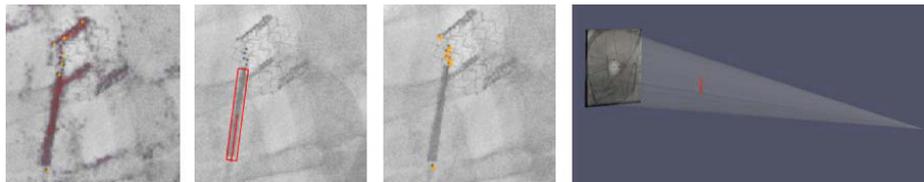


Fig. 3. (a) Initial ball marker estimates and likelihood measure of Phased Array (*PHA*) transducer foreground are integrated for (b) robust, coarse localization of catheter. (c) Ball marker locations are refined. (d) 6 DoF 3D pose is estimated together with 2D-3D point correspondences.

3.2 Approximate 2D Localization

The goal of the first step is to define the region of the XRF image *I* containing the catheter. We search for catheter parameters θ^* maximizing the posterior

probability:

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|I) \quad (1)$$

It is difficult to provide an analytical solution for Eq. (1), as $\boldsymbol{\theta}$ is defined in the space of 2D projections of the 3D catheter/markers. Thus, for coarse localization, the catheter model is approximated as a line segment (**Fig. 3(b)**):

$$\boldsymbol{\theta} \approx \boldsymbol{\theta}_B = \{(x, y), \gamma_{Y_{aw}}, s_x\} \quad (2)$$

where (x, y) , $\gamma_{Y_{aw}}$, s_x are 2D position, orientation and length of PHA . In contrast to TEE [9], the elongated shape of ICE catheter is not distinctive enough for these parameters to be estimated directly, hence we introduce a natural, parts based decomposition of $\boldsymbol{\theta}_B$. The segment is described as a pair of $\{B_{u1}, B_{u2}\}$ ball markers¹ and the likelihood $f_{PHA} = f(p(PHA|I), B_{u1}, B_{u2})$ of the phased array located in between them:

$$\{(x, y), \gamma_{Y_{aw}}, s_x\} = g(B_{u1}, B_{u2}, f_{PHA}) \quad (3)$$

with $(x, y) = \frac{1}{2}(B_{u1} + B_{u2})$, $\gamma_{Y_{aw}} = \tan\left(\frac{\partial y}{\partial x}(B_{u2} - B_{u1})\right)$ and $s_x = \|B_{u2} - B_{u1}\|$. This hierarchical decomposition allows to rewrite the argument of Eq. (1) too:

$$p(\boldsymbol{\theta}|I) \approx p(B_{u1}|I) \cdot p(B_{u2}|I) \cdot f(p(PHA|I), B_{u1}, B_{u2}) \quad (4)$$

where $p(B_i|I)$ and $p(PHA|I)$ refer to the posterior distribution of ball markers and the phased array in I , respectively.

Training the Model. The hierarchical decomposition defines the search spaces in more tractable terms. Our goal is to train two pixel wise classifiers, both employed in a sliding window approach: one bead detector for $p(B_i|I)$ similar to [6], and secondly a classifier $p(PHA|I)$ for the likelihood of a pixel belonging to the phased array transducer. Both detectors are implemented as a cascade of two levels, where during training the second level is trained on false positives from the first level as negative examples. Probabilistic boosting trees [10] are constructed to a depth of three, and include 50 weak learners in each node. The weak learners are image responses of discrete Haar-like features. $p(B_i|I)$ is trained on supersampled 0.25 mm resolution images, while subsampled 1 mm is used for the PHA likelihood measure to smooth out the finer, rotationally variant structure of the phased array transducer.

Detection Procedure and Hypothesis Fusion. During detection, the goal is to estimate the segment $\boldsymbol{\theta}_B$ using the trained model on unseen images. Due to scattered background, both bead and transducer classifiers produce a high

¹ At this point B_{u1} may be any of the five distal beads (ambiguities in perspective projection of various ICE poses hinder determination of exact distal marker id), also B_{u1} and B_{u2} may be swapped as the segment is invariant to 180° rotations).

rate of false alarms. In order to reliably extract the ICE catheter, we exploit local context ($g(\cdot)$ in (3)) and fuse pairs of ball markers $\{B_{u1}, B_{u2}\}_k$ and PHA candidates.

The hypothesis fusion consists of the following steps: (i) the top 40 locations indicated by $p(B_i|I)$ are clustered, (ii) $p(PHA|I)$ is evaluated on the whole image, producing a confidence map for the transducer likelihood and (iii) we select all possible pairs of marker hypotheses that satisfy the distance constraint² and finally (iv) the candidate pairs are scored by f_{PHA} a steerable filter (**Fig. 2(b)**), defined on the likelihood map: first taking the maximum response in each “column” along \overleftarrow{s} , then the 10th percentile of the remaining single row, in order to penalize hypotheses with low max. response along the $\{B_{u1}, B_{u2}\}$ axis. This score is used to sort all θ_B candidates in Eq. (1).

3.3 Search for Fiducials

The approximated coarse location is used to anchor two bounding boxes enclosing the two tails (**Fig. 4**) of the phased array: $\{V_{u1}, V_{u2}\}$. Searching within these reduced neighbourhoods, we intend to accurately localize and determine the identity of the visible fiducial beads and decide which end of the catheter is the distal one. Under various projections, the ball markers start to touch each other and overlap, making the identification difficult. In order to be able to recover partially overlapping markers, we created a circular template image (with its diameter dictated by the smallest marker in our images). The template is overlaid $\{V_{u1}, V_{u2}\}$ in a sliding window manner, the response for each pixel is calculated as the correlation coefficient of the overlapping regions. Local peaks are extracted that are further apart than one pixel. The resulting peaks are treated as marker candidates. The distal end of the catheter and $\{B_{dist_i}\}$ is indicated by more beads among the two neighborhoods. If more than one marker is present in the proximal area, we suppress the non maximum responses to keep the most likely one as B_{prox} .

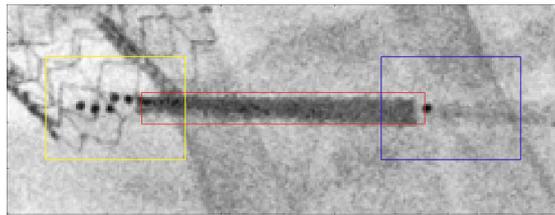


Fig. 4. Results of approximate localization (θ_B , red frame), and predicted search ranges for fiducial ball markers (V_{u1}, V_{u2} ; yellow and blue frames)

² From the C-arm we know the source-table distance that together with allowed catheter pitches (**Fig. 2(a)**) limits the 2D span of the proximal and distal markers visible in the X-Ray projection

3.4 Pose Estimation and Registration

Registration of ICE and XRF is establishing a spatial relationship between ICE voxels $P_{ICE_vox}(x_I, y_I, z_I)$ and XRF pixels $P_{Fluoro}(u, v)$:

$$P_{Fluoro}(u, v) = T_{\rightarrow Fluoro}^{C-arm} T_{\rightarrow C-arm}^{ICE_cath} T_{\rightarrow ICE_cath}^{ICE_vox} P_{ICE_vox}(x_I, y_I, z_I) \quad (5)$$

Here, the cone-beam perspective projection, intrinsic parameters (defined in the pinhole camera model) and C-arm angulation are combined in $T_{\rightarrow Fluoro}^{C-arm}$, and are considered known from the calibrated C-arm system itself. Also, the geometric relationship between the ICE image and ICE catheter $T_{\rightarrow ICE_cath}^{ICE_vox}$ is considered known from calibration, leaving the 3D pose of the ICE catheter in the 3D C-arm coordinates $T_{\rightarrow C-arm}^{ICE_cath} \equiv T$ unknown in (5).

The purpose of pose estimation is to infer the rigid transform T consisting of a rotation $R(\alpha_{Roll}, \beta_{Pitch}, \gamma_{Yaw})$ and translation $\vec{t} \equiv (x', y', z')$. For this we need to determine the correspondence between ball markers on the catheter $\{B'_j\}$ and the extracted 2D marker projections $\{B_i\}$.

We address the correspondence problem within $\{B_{dist_i}\}$ with an exhaustive search: for each combinations of markers, we execute a version of POSIT [2] and iterative optimization of pose. The estimated pose of beads is reprojected to the X-Ray plane. Using the sum of 2D point-to-point ($\sum P2P$) distance, we pick the marker combination that yields the smallest residual.

4 Experiments and Results

Our learning based models require large amounts of examples to train a classifier. Manually labeling markers in real X-Ray sequences is a laborious task, and reliable 3D Ground Truth (GT) location of the markers would require a co-registered C-arm CT. As such data was not readily available, we synthesised Digitally Reconstructed Radiographs (DRR) from a 3D image of the catheter and background angiographs without ICE. DRRs allow the 3D GT to be generated at the same time. We produced 11 000 ($0 \leq \alpha_{Roll} \leq 360, -30 \leq \beta_{Pitch} \leq 30, -180 \leq \gamma_{Yaw} \leq 180$) randomized images for training and 1 000 images for testing. In addition, we captured 641 XRF frames from an *in-vivo* porcine study and manually annotated the beads³. 400 of these frames were used for training, the remaining 241 for testing. Line segment detection accuracy is shown in Table 1a. More than 97% of the cases are successfully ($\|(\delta x, \delta y)\| < 5 \text{ mm}, \gamma_{Yaw} < 5^\circ$) detected which is comparable to the state of the art [7,6].

For the rest of the evaluation, we have selected those cases, where hypothesis fusion and template matching was able to localize B_{Prox} and $\|\{B_{Dist_i}\}\| = 5$ beads. These represent those cases where direct correspondence can be established, assuming beads do not overlap. 715 and 193 such cases were detected for DRR and *in-vivo* data, respectively. Native resolution of the tested X-Ray images is in the range of 0.3 to 0.4 mm, sub-pixel accuracy is achieved by our

³ We would like to acknowledge Zhivko Dimitrov for annotation of *in-vivo* data sets.

Table 1. (a) 2D catheter θ_B localization performance, translation error in mm . (b) Point-to-point 2D marker localization performance, error in mm , displayed as mean \pm std. dev. Considered in images where θ_B succeeded and $\|\{B_{Dist_i}\}\| = 5$.

(a)			(b)		
	synthetic	in-vivo		synthetic	in-vivo
No. of Images	1000	241	No. of Images	715 (71.5%)	193 (80.0%)
Success/Rate	979/97.9%	235/97.5%	Symmetric	0.31 \pm 1.55	0.22 \pm 0.19
Avg \pm Std	1.20 \pm 0.55	1.10 \pm 0.58	Hausdorff	0.67 \pm 3.73	0.36 \pm 0.51

marker detection shown in Table 1b. To evaluate the accuracy for the envisioned image fusion application, we employ the target registration error (TRE). The target points are defined as the four corners of the sound cone at the depth of 50 mm . The effective value of TRE (ETRE) is computed by keeping only the TRE components which are parallel to the detector plane. Due to lacking 3D GT, we could only evaluate the 6 DoF pose estimation on the 715 DRR images for which all markers could be detected. 12 cases were excluded, where the sum of 2D point-to-point distance was above 1 mm . In the remaining 703 images the correspondence and 3D pose estimation performed well, yielding an average TRE of 8.06 ± 7.2 mm and ETRE of 2.81 ± 1.5 mm (Table 2).

Table 2. 3D pose estimation accuracy on DRR cases where $\|\{B_{Dist_i}\}\| = 5$. From 715, we excluded 12 cases where \sum P2P reprojection error is above 1 mm .

	Roll $^\circ$	Pitch $^\circ$	Yaw $^\circ$	Depth mm	TRE mm	ETRE mm
Avg. \pm Std.	2.03 \pm 1.5	2.07 \pm 1.6	0.13 \pm 0.2	7.28 \pm 7.8	8.06 \pm 7.2	2.81 \pm 1.5
Median/ P_{90}	1.76/4.21	1.68/4.37	0.10/0.25	4.74/17.93	5.54/17.58	2.59/4.84

5 Discussion

The overall 2D marker detection error is lower for *in-vivo* data compared to synthetic. This is due to the fact, that the X-Ray sequences used as background for DRR generation are from actual cardiac interventions, and contain various other wires and tools that appear similar to the ICE catheter, while the porcine data featured less clutter.

The 3D TRE is non-isotropic: the depth direction contributes the major part of the error, as small inaccuracies of detected beads in the X-Ray detector plane are magnified along its normal. When targets are displayed as overlays on the 2D XRF images, the depth error “disappears”. To reflect this clinical scenario, we introduced the ETRE. The 2.81 ± 1.5 mm ETRE is comparable to the 5 mm range envisaged by our interventional collaborators as clinically useful.

Future work should be aimed at techniques to recover pose from only 3 or 4 visible distal markers, and to exclude extraneous marker candidate outliers. Moreover, we plan on investigation of temporal information to increase detection rate in challenging backgrounds.

6 Conclusions

To the best of our knowledge, this is the first published method to automatically extract and register an ICE catheter with XRF. The extraction method is agnostic to catheters as long as both ends are beads marked, and connected with a visible part in between, hence, general enough to be used for other devices. We provided quantitative evaluation on a number of real and synthetic images and arrived at convincing preliminary results.

References

1. David, P., Dementhon, D., Duraiswami, R., Samet, H.: Softposit: Simultaneous pose and correspondence determination. *IJCV* 59, 259–284 (2004)
2. Dementhon, D.F., Davis, L.S.: Model-based object pose in 25 lines of code. *Int. J. Comput. Vision* 15(1-2), 123–141 (1995)
3. Garro, V., Crosilla, F., Fusiello, A.: Solving the pnp problem with anisotropic orthogonal procrustes analysis. In: 2011 3DIMPVT, pp. 262–269 (2012)
4. Hoffmann, K.R., Esthappan, J.: Determination of three-dimensional positions of known sparse objects from a single projection. *Medical Physics* (1997)
5. Lang, P., Seslija, P., Chu, M., Bainbridge, D., Guiraudon, G., Jones, D., Peters, T.: Us-fluoroscopy registration for transcatheter aortic valve implantation. *IEEE Transactions on Biomedical Engineering* 59(5), 1444–1453 (2012)
6. Lu, X., Chen, T., Comaniciu, D.: Robust discriminative wire structure modeling with application to stent enhancement in fluoroscopy. In: *CVPR 2011*, p. 1121 (2011)
7. Ma, Y., Gogin, N., Cathier, P., Housden, R., Gijsbers, G., Cooklin, M., O’Neill, M., Gill, J., Rinaldi, C., Razavi, R., Rhode, K.: Real-time x-ray fluoroscopy-based catheter detection and tracking for cardiac electrophysiology interventions. *Medical Physics* 40(7) (2013)
8. Ma, Y., Penney, G.P., Bos, D., Frissen, P., Rinaldi, C.A., Razavi, R., Rhode, K.S.: Hybrid echo and x-ray image guidance for cardiac catheterization procedures by using a robotic arm: a feasibility study. *Physics in Medicine and Biology* (2010)
9. Mountney, P., Ionasec, R., Kaizer, M., Mamaghani, S., Wu, W., Chen, T., John, M., Boese, J., Comaniciu, D.: Ultrasound and fluoroscopic images fusion by autonomous ultrasound probe detection. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) *MICCAI 2012, Part II*. LNCS, vol. 7511, pp. 544–551. Springer, Heidelberg (2012)
10. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: *ICCV 2005*, vol. 2, pp. 1589–1596. IEEE (2005)
11. Wang, P., Chen, T., Ecabert, O., Prummer, S., Ostermeier, M., Comaniciu, D.: Image-based device tracking for the co-registration of angiography and intravascular ultrasound images. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part I*. LNCS, vol. 6891, pp. 161–168. Springer, Heidelberg (2011)
12. Webster, B.: Cartosound module soundstar catheter (February 2014), <http://www.biosensewebster.com/cartosound.php>