# Towards Automated Ultrasound Transesophageal Echocardiography and X-Ray Fluoroscopy Fusion Using an Image-Based Co-registration Method

Shanhui Sun[1(✉)], Shun Miao[1], Tobias Heimann[1], Terrence Chen[1], Markus Kaiser[2], Matthias John[2], Erin Girard[2], and Rui Liao[1]

[1] Siemens Healthcare, Medical Imaging Technologies, Princeton, NJ 08540, USA
shanhui.sun@siemens.com

[2] Siemens Healthcare, Advanced Therapies, 91301 Forchheim, Germany

**Abstract.** Transesophageal Echocardiography (TEE) and X-Ray fluoroscopy are two routinely used real-time image guidance modalities for interventional procedures, and co-registering them into the same coordinate system enables advanced hybrid image guidance by providing augmented and complimentary information. In this paper, we present an image-based system of co-registering these two modalities through real-time tracking of the 3D position and orientation of a moving TEE probe from 2D fluoroscopy images. The 3D pose of the TEE probe is estimated fully automatically using a detection based visual tracking algorithm, followed by intensity-based 3D-to-2D registration refinement. In addition, to provide high reliability for clinical use, the proposed system can automatically recover from tracking failures. The system is validated on over 1900 fluoroscopic images from clinical trial studies, and achieves a success rate of 93.4 % at 2D target registration error (TRE) less than 2.5 mm and an average TRE of 0.86 mm, demonstrating high accuracy and robustness when dealing with poor image quality caused by low radiation dose and pose ambiguity caused by probe self-symmetry.

**Keywords:** Visual tracking based pose detection · 3D-2D registration

## 1 Introduction

There is a fast growth of catheter-based procedures for structure heart disease such as transcatheter aortic valve implantation (TAVI) and transcatheter mitral valve replacement (TMVR). These procedures are typically performed under the independent guidance of two real-time imaging modalities, i.e. fluoroscopic Xray and transesophageal echocardiaography (TEE). Both imaging modalities have their own advantages, for example, Xray is good at depicting devices, and TEE is much better at soft tissue visualization. Therefore fusion of both modalities could provide complimentary information for improved security and accuracy


S. Ourselin et al. (Eds.): MICCAI 2016, Part I, LNCS 9900, pp. 395–403, 2016.
DOI: 10.1007/978-3-319-46720-7_46

during the navigation and deployment of the devices. For example, a Xray/TEE fusion system can help the physican finding correct TAVR deployment angle on fluoroscopic image using landmarks transformed from annotations on TEE.

To enable the fusion of Xray and TEE images, several methods have been proposed to recover the 3D pose of TEE probe from the Xray image [1–3,5,6], where 3D pose recovery is accomplished by 3D-2D image registration. In [1,2,5], 3D-2D image registration is fulfilled via minimizing dissimilarity between digitally generated radiographies (DRR) and X-ray images. In [6], DRR rendering is accelerated by using mesh model instead of a computed tomography (CT) volume. In [3], registration is accelerated using a cost function which is directly computed from X-ray image and CT scan via splatting from point cloud model without the explicit generation of DRR. The main disadvantage of these methods is that they are not fully automatic and requires initialization due to small capture range. Recently, Montney et al. proposed a detection based method to recover the 3D pose of the TEE probe from an Xray image in work [7]. 3D translation is derived from probe's in-plane position detector and scale detector. 3D Rotation (illustrated in Fig. 1(a)) is derived from in-plane rotation (yaw angle) based on orientation detector and out-of-plane rotations (roll and pitch angles) based on a template matching based approach. They demonstrated feasibility on synthetic data. Motivated by the detection based method, we present a new method in this paper to handle practical challenges in a clinical setup such as low X-Ray dose, noise, clutters and probe self-symmetry in 2D image. Two self-symmetry examples are shown in Fig. 1(b). To minimize appearance ambiguity, three balls (Fig. 2(a)) and three holes (Fig. 2(b)) are manufactured on the probe. Examples of ball marker and hole marker appearing in fluoroscopic images are shown in Fig. 2(c) and (d). Our algorithm explicitly detects the markers and incorporates the marker detection results into TEE probe pose estimation for an improved robustness and accuracy.

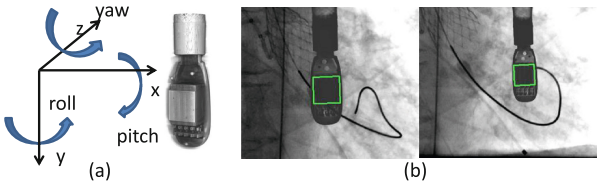


**Fig. 1.** (a) Illustration of TEE Euler angles. Yaw is an in-plane rotation. Pitch and roll are out-of-plane rotations. (b) Example of ambiguous appearance in two different poses. Green box indicates probe's transducer array. Roll angle between two poses are close to 90°. Without considering markers (Fig. 2), probe looks similar in X-ray images.

In addition, based on the fact of that physicians acquire series of frames (a video sequence) in interventional cardiac procedure, we incorporate temporal information to boost the accuracy and speed, and we formulate our 6-DOF parameter tracking inference as a sequential Bayesian inference framework. To

further remove discretization errors, Kalman filter is applied to temporal pose parameters. In addition, tracking failure is automatically detected and automated tracking initialization method is applied. For critical time points when the measurements (e.g., annotated anatomical landmarks) from the TEE image are to be transformed to the fluoroscopic image for enhanced visualization, intensity-based 3D to 2D registration of the TEE probe is performed to further refine the estimated pose to ensure a high accuracy.

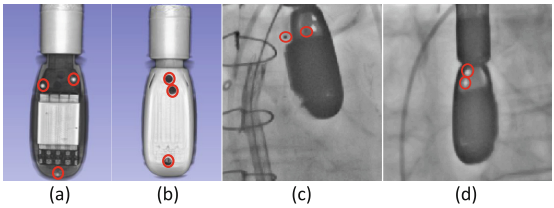


**Fig. 2.** Illustration of probe markers circled in red. (a) 3D TEE probe front side with 3 ball markers and (b) back side with 3 hole markers. (c) Ball markers and (d) hole markers appear in X-Ray images.

## 2 Methods

A 3D TEE point $Q^{TEE}$ can be projected to the 2D fluoroscopic image point $Q^{Fluoro} = P_{int}P_{ext}(R^W_{TEE}Q^{TEE} + T^W_{TEE})$, where $P_{int}$ is C-Arm's internal projection matrix. $P_{ext}$ is C-Arm's external matrix which transforms a point from TEE world coordinate to C-Arm coordinate. $R^W_{TEE}$ and $T^W_{TEE}$ are TEE probe's rotation and position in the world coordinate. The internal and external matrices are known from calibration and C-Arm rotation angles. $R^W_{TEE} = P^{-1}_{ext}R^C_{TEE}$ and $T^W_{TEE} = P^{-1}_{ext}T^C_{TEE}$, where $R^C_{TEE}$ and $T^C_{TEE}$ are the probe's rotation and position in the C-Arm coordinate system. $R^C_{TEE}$ is composed of three euler angles $(\theta_z, \theta_x, \theta_y)$, which are illustrated in Fig. 1(a), and $T^C_{TEE} = (x, y, z)$.

The proposed tracking algorithm is formulated as finding an optimal pose on the current image $t$ constrained via prior pose from image $t-1$. In our work, pose hypotheses with pose parameters $(u, v), \theta_z, s, \theta_x$ and $\theta_y$ are generated and optimal pose among these hypotheses are identified in a sequential Bayesian inference framework. Figure 3 illustrates an overview of the proposed algorithm. We defined two tracking stages: in-plane pose tracking for parameters $(u, v)$, $s$, and $\theta_z$ and out-of-plane tracking for parameters $\theta_x$ and $\theta y$. In the context of visual tracking, the searching spaces of $(u_t, v_t, \theta_{z_t}, s_t)$ and $(\theta_{x_t}, \theta_{y_t})$ are significantly reduced via generating in-plane pose hypotheses in the region of interest $(u_{t-1} \pm \delta_T, v_{t-1} \pm \delta_T, \theta_{z_{t-1}} \pm \delta_z, s_{t-1} \pm \delta_s)$, and out-of-plane pose hypotheses in the region of interest $(\theta_{x_{t-1}} \pm \delta_x, \theta_{y_{t-1}} \pm \delta_y)$, where $\delta_T$, $\delta_z$, $\delta s$, $\delta_x$ and $\delta_y$ are searching ranges. Note that we choose these searching ranges conservatively, i.e. much larger than typical frame-to-frame probe motion.

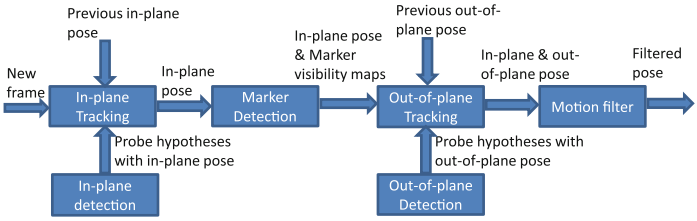


**Fig. 3.** Overview of tracking framework.

### 2.1 In-Plane Pose Tracking

To realize tracking, we use Bayesian inference network [9] as follows.

$$P(M_t|Z_t) \propto P(M_t)P(Z_t|M_t), \tag{1a}$$

$$\hat{M}_t = \underset{M_t}{\operatorname{argmax}} P(M_t|Z_t) \tag{1b}$$

where $M_t$ is in-plane pose parameters $(u, v, \theta_z, s)$. $\hat{M}_t$ is the optimal solution using maximum a posterior (MAP) probability. $P(Z_t|M_t)$ is the likelihood of an in-plane hypothesis being positive. $P(M_t)$ represents in-plane motion prior probability, which is defined as a joint Gaussian distribution with respect to the parameters $(u, v, \theta_z, s)$ with standard deviations ($\sigma_T$, $\sigma_T$, $\sigma_{\theta_z}$ and $\sigma_s$).

In-plane pose hypotheses are generated using marginal space learning method similar to the work in [10]. A series of cascaded classifiers are trained to classify probe position $(u, v)$, size $s$, and orientation $\theta_z$. These classifiers are trained sequentially: two position detectors for $(u, v)$, orientation detector for $\theta_z$ and scale detector for $s$. Each detector is a Probabilistic Boosting Tree (PBT) classifier [8] using Haar-like features [9] and rotated Haar-like features [9]. The position classifier is trained on the annotations (positive samples) and negative samples randomly sample to be away from annotations. The second position detector performs bootstrapping procedure. Negative samples are collected from both false positive of the first position detection results and random negative samples. Orientation detector is trained on the rotated images, which are rotated to 0° according to annotated probe's orientations. The Haar-like features are computed on rotated images. During orientation test stage, input image is rotated every 5° in range of $\theta_{z_{t-1}} \pm \delta_z$. Scale detector is trained on the rotated images. Haar-like feature is computed on the rotated images and the Haar feature windows are scaled based on probe's size. During scale test stage, Haar feature window is scaled and quantified in the range of $s_{t-1} \pm \delta_s$.

### 2.2 Out-of-Plane Pose Tracking

Out-of-plane pose tracking performs another Bayesian inference network derived from Eq. 1. Thus in this case $M_t$ (in Eq. 1) is out-of-plane pose parameters $(\theta_x, \theta_y)$. $\hat{M}_t$ is the optimal solution using MAP probability. $P(Z_t|M_t)$ is

likelihood of an out-of-plane hypothesis being positive. $P(M_t)$ is an out-of-plane motion prior probability, which is defined as a joint Gaussian distribution with respect to the parameters $(\theta_x, \theta_y)$ with standard deviations $(\sigma_x, \sigma_y)$.

Out-of-plane pose hypothesis generation is based on a K nearest neighbour search using library-based template matching. At training stage, we generate 2D synthetic X-Ray images at different out-of-plane poses and keeping the same in-plane pose. Roll angle ranges from $-180°$ to $180°$. Pitch angle ranges from $-52°$ to $52°$, and angles out of this range are not considered since they are not clinically relevant. Both step sizes are $4°$. All in-plane poses of these synthetic images are set to the same canonical space: probe positioned at image center, $0°$ yaw angle and normalized size. Global image representation of each image is computed representing out-of-plane pose and saved in a database. The image representation is derived based on method presented in [4]. At test stage, $L$ in-plane pose perturbations (small translations, rotations and scales) about the computed in-plane pose (Sect. 2.1) are produced. $L$ in-plane poses are utilized to define $L$ probe ROIs in the same canonical space. Image representation of each ROI is computed and is used to search (e.g. KD-Tree) in the database and resulting K nearest neighbors. Unfortunately, only using global representation is not able to differentiate symmetric poses. For example, a response map of an exemplar pose to all the synthetic images shown in Fig. 4. Note that there are two dominant symmetrical modes and thus out-of-plane hypotheses are generated around these two regions. We utilize markers (Fig. 2) to address this problem. For each synthetic image, we thus save the marker positions in the database. The idea is that we perform a visibility test at each marker position in $L * K$ searching results. The updated searching score $\hat{T}_{score} = \frac{T_{score}}{N} \sum_{i=1}^{N} \alpha + P_i(x_i, y_i)$, where $T_{score}$ is a searching score. $P_i$ is $i^{th}$ marker's visibility ($[0.0, 1.0]$) at marker position $(x_i, y_i)$ in the corresponding synthetic image template. $N$ is the number of markers. $\alpha$ is a constant value 0.5. Marker visibility test is fulfilled using two marker detectors: ball marker detector and hole marker detector. Both detectors are two cascaded position classifiers (PBT classifier with Haar-like features), and visibility maps are computed based on the detected marker locations.

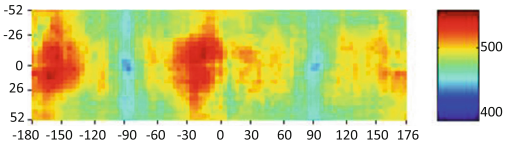


**Fig. 4.** An example of template matching score map for one probe pose. X-axis is roll angle and Y-axis is pitch angle. Each pixel represents one template pose. Dark red color indicates a high matching score and dark blue indicates a small matching score.

### 2.3 Tracking Initialization and Failure Detection

Initial probe pose in the sequence is derived from detection results without considering temporal information. We detect the in-plane position, orientation and scale, and out-of-plane roll and pitch hypotheses in the whole required searching

space. We get a final in-plane pose via Non-maximal suppression and weighted average to the pose with the largest detection probability. The hypothesis with largest searching score is used as out-of-plane pose. For initializing tracking: (1) we save poses of $N_i$ (e.g. $N_i = 5$) consecutive image frames. (2) A median pose is computed from $N_i$ detection results. (3) Weighted mean pose is computed based on distance to the median pose. (4) Standard deviation $\sigma_p$ to the mean pose is computed. Once $\sigma_p < \sigma_{threshod}$, tracking starts with initial pose (i.e. the mean pose). During tracking, we identify tracking failure through: (1) we save $N_f$ (e.g. $N_f = 5$) consecutive tracking results. (2) The average searching score $m_{score}$ is computed. If $m_{score} < m_{threshold}$, we stop tracking and re-start tracking initialization procedure.

### 2.4 3D-2D Registration Based Pose Refinement

In addition, we perform 3D-2D registration of the probe at critical time points when measurements are to be transformed from TEE images to fluoroscopic images. With known perspective geometry of the C-Arm system, a DRR can be rendered for any given pose parameters. In 3D-2D registration, the pose parameters are iteratively optimized to maximize a similarity metric calculated between the DRR and the fluoroscopic image. In the proposed method, we use Spatially Weighted Gradient Correlation (SWGC) as the similarity metric, where areas around the markers in the DRR are assigned higher weights as they are more distinct and reliable features indicating the alignment of the two images. SWGC is calculated as Gradient Correlation (GC) of two weighted images: $SWGC = GC(I_f \cdot W, I_d \cdot W)$, where $I_f$ and $I_d$ denote the fluoroscopic image and the DRR, respectively, $W$ is a dense weight map calculated based on the projection of the markers, and $GC(\cdot, \cdot)$ denotes the GC of the two input images. Using SWGC as the similarity metric, the pose parameters are optimized using Nelder-Mead optimizer to maximize SWGC.

## 3 Experiment Setup, Results and Discussions

For our study, we trained machine learning based detectors on $\sim 10,000$ fluoroscopic images ($\sim 90\,\%$ images are synthetically generated images and $\sim 10\,\%$ images are clinical images). We validated our methods on 34 X-Ray fluoroscopic videos (1933 images) acquired from clinical experiments, and 13 videos (2232 images) from synthetic generation. The synthetic images were generated by blending DRRs of the TEE probe (including tube) with real fluoroscopic images containing no TEE probe. Particularly for the test synthetic sequences, we simulate realistic probe motions (e.g., insertion, retraction, roll etc.) in the fluoroscopic sequences. Ground truth poses for synthetic images are derived from 3D probe geometry and rendering parameters. Clinical images are manually annotated using our developed interactive tool by 4 experts. Image size is $1024 \times 1024$ pixels. Computations were performed on a workstation with Intel Xeon (E5-1620) CPU 3.7 GHz and 8.00 GB Memory. On average, our tracking

algorithm performs at 10 fps. We performed our proposed detection algorithm (discussed in Sect. 2.3, tracking is not enabled), proposed automated tracking algorithm and registration refinement after tracking on all test images. Algorithm accuracy was evaluated by calculating the standard target registration error (TRE) in 2D. The targets are defined at the four corners of the TEE imaging cone at 60 mm depth and the reported TRE is the average TRE over the four targets. 2D TRE is a target registration error that z axis (depth) of the projected target point is not considered when computing distance error. Table 1 shows success rate, average TRE and median TRE at 2D TRE $< 4$ mm and $< 2.5$ mm respectively. Figure 5 shows success rate vs 2D TRE on all validated clinical and synthetic images.

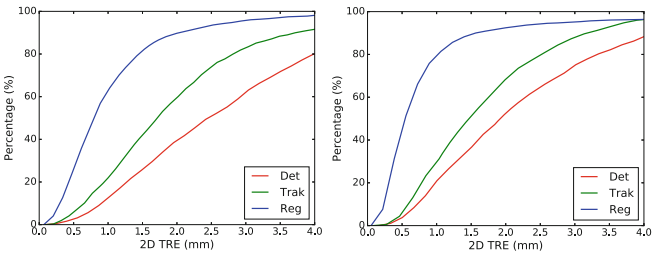


**Fig. 5.** Result of success rate vs 2D TRE on clinical (a) and synthetic (b) validations of the proposed detection, tracking and 3D-2D registration refinement algorithms.

Due to limited availability of clinical data, we enlarged our training data set using synthetic images. Table 1 and Fig. 5 show our approach performs well on real clinical data utilizing hybrid training data. We expect increased robustness and accuracy after larger number of real clinical cases become available. Tracking algorithm improved robustness and accuracy comparing to detection alone approach. One limitation of our tracking algorithm is not able to compensate all discretization errors although temporal smoothing is applied using Kalman filter. This is a limitation of any detection based approach. To further enhance accuracy, refinement is applied when physicians perform the measurements. To

**Table 1.** Quantitative results on validations of the proposed detection (Det), tracking (Trak) and 3D-2D registration refinement (Reg) algorithms. Numbers in the table show success rate, mean TRE (mm), median TRE (mm) under different TRE error ranges.

| | Clinical data | | Synthetic data | |
|---|---|---|---|---|
| Method | TRE $< 4$ mm | TRE $< 2.5$ mm | TRE $< 4$ mm | TRE $< 2.5$ mm |
| Det | (80.0 %, 2.09, 2.02) | (50.9 %, 1.47, 1.48) | (88.4 %, 1.86, 1.73) | (64.7 %, 1.38, 1.35) |
| Trak | (91.6 %, 1.71, 1.61) | (73.7 %, 1.38, 1.36) | (96.4 %, 1.59, 1.42) | (79.7 %, 1.28, 1.22) |
| Reg | (98.0 %, 0.97, 0.79) | (93.4 %, 0.86, 0.75) | (96.4 %, 0.69, 0.52) | (94.3 %, 0.63, 0.51) |

better understand the performance from registration refinement, in our study we applied the refinement step on all images after tracking. Note that the refinement algorithm did not bring more robustness but improved the accuracy.

## 4 Conclusion

In this work, we presented a fully automated method of recovering the 3D pose of TEE probe from the Xray image. Tracking is very important to give physicians the confidence that the probe pose recovery is working robustly and continuously. Abrupt failed probe detection is not good especially when the probe does not move. Detection alone based approach is not able to address abrupt failures due to disturbance, noise and appearance ambiguities of the probe. Our proposed visual tracking algorithm avoids abrupt failure and improves detection robustness as shown in our experiment. In addition, our approach is a near real-time approach (about 10 FPS) and a fully automated approach without any user interaction, e.g. manual pose initialization as required by many state-of-the-art methods. Our proposed complete solution addressing TEE and X-Ray fusion problem is applicable to clinical practice due to high robustness and accuracy.

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

## References

1. Gao, G., et al.: Rapid image registration of three-dimensional transesophageal echocardiography and X-ray fluoroscopy for the guidance of cardiac interventions. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 124–134. Springer, Heidelberg (2010)
2. Gao, G., et al.: Registration of 3D transesophageal echocardiography to X-ray fluoroscopy using image-based probe tracking. Med. Image Anal. **16**(1), 38–49 (2012)
3. Hatt, C.R., Speidel, M.A., Raval, A.N.: Robust 5DOF transesophageal echo probe tracking at fluoroscopic frame rates. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 290–297. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24553-9_36
4. Hinterstoisser, S., et al.: Gradient response maps for real-time detection of textureless objects. IEEE Trans. Pattern Anal. Mach. Intell. **34**(5), 876–888 (2012)
5. Housden, R.J., et al.: Evaluation of a real-time hybrid three-dimensional echo and X-ray imaging system for guidance of cardiac catheterisation procedures. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 25–32. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33418-4_4
6. Kaiser, M., et al.: Significant acceleration of 2D–3D registraion-based fusion of ultrasound and X-ray images by mesh-based DRR rendering. In: SPIE, p. 867111 (2013)
7. Mountney, P., et al.: Ultrasound and fluoroscopic images fusion by autonomous ultrasound probe detection. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 544–551. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33418-4_67

8. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1589–1596 (2005)
9. Wang, P., et al.: Image-based co-registration of angiography and intravascular ultrasound images. IEEE TMI **32**(12), 2238–2249 (2013)
10. Zheng, Y., et al.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable featurs. IEEE Trans. Med. Imaging **27**(11), 1668–1681 (2008)