Real-time registration of 3D echo to X-ray fluoroscopy based on cascading classifiers and image registration

Ma, Y., Houston, R. J., Fazili, A. & Rhode, K. S.

Original citation & hyperlink:
Ma, Y, Houston, RJ, Fazili, A & Rhode, KS 2021, 'Real-time registration of 3D echo to X-ray fluoroscopy based on cascading classifiers and image registration', Physics in Medicine and Biology, vol. 66, no. 5, 055019.
https://dx.doi.org/10.1088/1361-6560/abe420

DOI 10.1088/1361-6560/abe420
ISSN 0031-9155
ESSN 1361-6560

Publisher: IOP Publishing

Real-time registration of 3D echo to X-ray fluoroscopy based on cascading classifiers and image registration

YingLiang Ma[1], R. James Housden[2], Ansab Fazili[3], Aruna V. Arujuna[2] and Kawal S. Rhode[2]

[1] School of Computing, Electronics and Mathematics, Coventry University, CV1 5FB, UK
[2] School of Biomedical Engineering and Imaging Sciences, King's College London, SE1 7EH, UK
[3] Department of Cardiology, Lister Hospital, Stevenage, UK

E-mail: yingliang.ma@coventry.ac.uk

Received xxxxxx
Accepted for publication xxxxxx
Published xxxxxx

Abstract

3D transesophageal echocardiography (TEE) is one of the most significant advances in cardiac imaging. Although TEE provides real-time three-dimensional (3D) visualization of heart tissues and blood vessels and has no ionizing radiation, X-ray fluoroscopy still dominates in guidance of cardiac interventions due to TEE having a limited field of view and poor visualization of surgical instruments. Therefore, fusing 3D echo with live X-ray images can provide a better guidance solution. This paper proposes a novel framework for image fusion by detecting the pose of the TEE probe in X-ray images in real-time. The framework does not require any manual initialization. Instead it uses a cascade classifier to compute the position and in-plane rotation angle of the TEE probe. The remaining degrees of freedom (DOFs) are determined by fast marching against a template library. The proposed framework is validated on phantoms and patient data. The target registration error (TRE) for the phantom was 2.1 mm. In addition, 10 patient datasets, seven of which were acquired from cardiac electrophysiology procedures and three from trans-catheter aortic valve implantation procedures, were used to test the clinical feasibility as well as accuracy. A mean registration error of 2.6 mm was achieved, which is well within typical clinical requirements.

Keywords: Cardiac interventional guidance, X-ray fluoroscopy, 3D ultrasound, image fusion.

1. Introduction

Minimally-invasive cardiac interventional procedures such as the treatment of structural heart disease and cardiac electrophysiology (EP) procedures are generally guided under 2D X-ray fluoroscopy. Interventional devices are designed to be radiopaque so they are highly visible in X-ray images. However, as the soft tissues of the heart have little contrast under X-ray, contrast agents are routinely injected to visualize the anatomical structures of target areas during the key stages of procedures. The use of contrast agents is limited due to toxicity and alternative imaging methods such as echocardiography [1][2], magnetic resonance (MR) [3] and computed tomography (CT) [4], are often employed to support the image guidance of the procedure. With recent advance of transesophageal echocardiography (TEE), TEE can provide 3D and real-time visualization of heart tissues and blood vessels. Unlike X-ray fluoroscopy, TEE does not use ionizing radiation and it is easily accessible in the cardiac intervention suite. Therefore, TEE has become a popular choice for interventional imaging, particularly for the treatment of structural heart disease. However, TEE cannot be used as the solo image guidance tool as it has a limited field of view and poor visualization of surgical tools and interventional devices. In current clinical practice, X-ray images are displayed in one monitor with a 3D TEE image volume in a separated monitor. Image fusion between X-ray fluoroscopy and TEE has been proposed and clinically implemented (EchoNavigator, Philips Healthcare) [5]. EchoNavigator is an excellent tool to allow merged display of echocardiographic and fluoroscopic images in real-time and allow the interventionalist to interact with both imaging modalities simultaneously. However, in the presence of large cardiac or respiratory motions, registration between X-ray images and 3D TEE image volumes might not be accurate. The errors might be caused by the delayed image registration.

Image fusion between X-ray fluoroscopy and TEE could be also solved by other 2D/3D registration methods [6][7] or magnetic tracking sensors [8][9]. Sensor-based solutions may become inaccurate when the electromagnetic field is distorted by large metal objects inside the cardiac catheter laboratory. Furthermore, additional sensors need to be attached to the TEE probe, which make the solution less clinically translatable. On the other hand, 2D/3D registration methods will estimate the location and pose of the TEE probe in the X-ray image by registering a 3D model of the TEE probe to the live X-ray images. Therefore, 2D/3D registration methods do not require additional hardware (sensors). Most of the existing methods cannot achieve real-time performance. A maximum of only 0.5 frames per second (FPS) were achieved in [10] and 2 FPS were achieved in [7]. Although 20 FPS was achieved in [11] for TEE probe localization, this method cannot be directly used in our application as it does not calculate out-of-plane rotational parameters. The method in [6] was implemented in MATLAB; therefore, it cannot perform in real-time. Finally, 20 FPS were achieved in [12] but only with manual initialization and GPU implementation. In recent years, several solutions were developed for real-time device detection in X-ray fluoroscopic images. In [13] [14], real-time catheter and guidewire detection methods were developed. Detection methods based on machine learning algorithms have demonstrated a great potential to solve the 2D/3D registration problem for image fusion between X-ray fluoroscopy and TEE in real-time [7][11]. Additionally, Hatt et al. [15] developed a Hough forest based detection framework to localize the TEE probe in real-time (17 FPS). The solution is GPU-based implementation and can not calculate out-of-plane rotational parameters. Miao et al. [16] presented a CNN regression approach to perform real-time (10 FPS) 2D/3D registration. Although this method is able to compute all in-plane and out-of-plane parameters for 2D/3D registration, it has limited capture range (3.3 mm for the TEE probe) and requires a large memory footprint (2.39 GB).

In this paper, we propose a fully automatic 2D-3D registration framework-based probe pose estimation which is capable of registering 3D live TEE image volumes with live X-ray images in real-time. The novel framework is based on automatically detecting the location and pose of the TEE probe in live X-ray images. It does not require any manual initialization and does not have limited capture range. The estimation of 2D location and in-plane rotation of the probe is provided by a cascade image classifier. The remaining degrees of freedom (DOFs) of the TEE probe are solved by image matching between an image from a template library with binary masks and the subimage detected from the cascade classifier. The template library is constructed from 3D scans using conventional or cone beam CT (CBCT) of the TEE probe. The cascade classifier not only detects the TEE probe in real-time even without a GPU implementation, but also it can generate good initial pose estimates which could dramatically reduce the overall computational load following initialization. Therefore, the proposed framework could achieve real-time performance even without GPU implementation. Instead, multi-threaded implementation should be sufficient. Furthermore, the proposed framework does not require a high-resolution 3D model (used in [10]) which is normally created using a specialized scanner such as a nano-CT scanner. In our proposed framework, the 3D model is created using CBCT and the generated images have similar resolution as the X-ray fluoroscopic images used during the procedures because they are acquired on the same hardware. Therefore, our solution does not require specialized equipment or sensors and the light-weight algorithms reduce computational loads on the GPU and leave valuable GPU computing power for other tasks such as real-time visualization of 3D TEE image volumes.

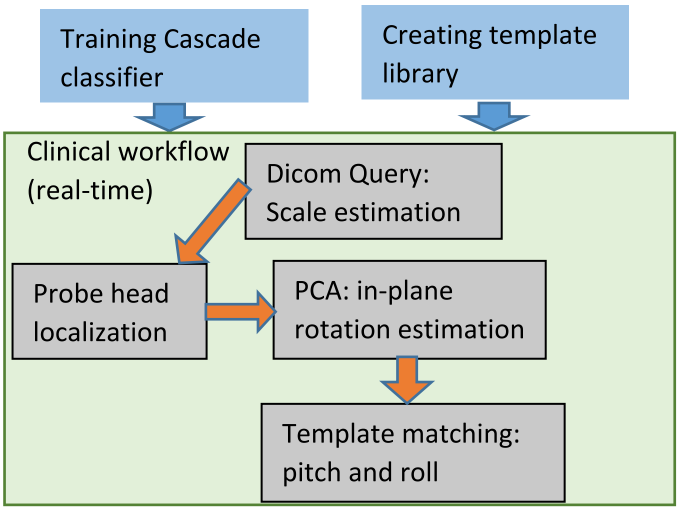


Figure 1: The full workflow of the proposed computation framework. Only the clinical workflow is computed in real-time.

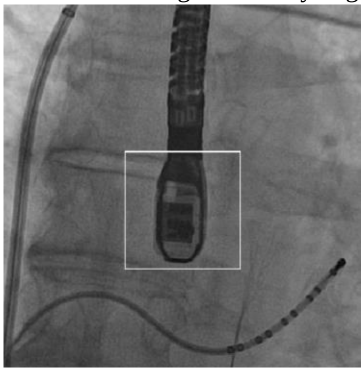
2. Method

The proposed computational framework is divided into four steps: A) Model training for the cascade classifier. B) Creating a template library from CBCT. C) Probe detection. D) Probe 2D-3D registration. (A) and (B) are offline and they only need to be done once for each type of TEE probe. (C) and (D) are computed in real-time. The full workflow is illustrated in figure 1.

2.1 Model training for the cascade classifier

A cascade classifier is a particular case of ensemble learning based on the concatenation of several classifiers, using all information collected from the output from a given classifier as additional information for the next classifier in the cascade. The cascade object classifier provides high classification accuracy and real-time performance [17]. Therefore, it has been widely used in real-time computer vision applications such as face detection [18], vehicle detection [19], pedestrian detection [20] and football detection in robotic soccer competitions [21]. The cascade classifier was also used in breast cancer detection in mammograms [22].

To create the model, the cascade classifier has to be trained with a relatively large number of sample views of a target object and arbitrary negative images which do not contain any part of the target object. In our case, 400 positive images were created using in-house software. The software allows a user to interactively select a fixed-size region of interest which is centred at the middle of the TEE probe head (figure 2(a) gives an example). Then the software down-samples the region image to the resolution 32x32. The reason for this small resolution is because high-resolution positive images require much longer training time. In our experience, if 400 64x64 positive images are used for training, it will take more than 3 weeks to train the classifier using a standard desktop computer. Even with the support of multi-threading or GPU, the training time can only be reduced to a few days. The reason for very long training time is that the cascade training algorithm searches three kinds of features: a two-rectangle feature, a three-rectangle feature and a four-rectangle feature. According to [23], there will be 45,396 sets of rectangle features for the image with a resolution of 24x24. Furthermore, there are normally 10 to 20 stages of training and, in each stage, the training algorithm will go through all rectangle features again. Also, training might take several tries to find the optimal parameters. Therefore, low-resolution training images were used. Out of a total of 400 positive images, 200 images were from X-ray images acquired during two clinical cases. The other 200 images were acquired during the phantom test experiment. There are different orientations of TEE probe in the positive images (see figure 2(b)) so that the classifier can detect the target from any angle.



(a) Creating a positive training image

(b) Positive images

(c) negative images

Figure 2. Examples of training images

The number of negative training images should be twice the number of positive images [21] to achieve optimal results. Therefore, 800 negative images were generated from the same source images as positive images used. Instead of cropping the TEE probe from the source image, negative images are sub-images which do not show any part of TEE probe head. These are background images which could contain ECG wires, interventional devices such as catheters, bone shadows and angiographic contrast agent.

To train the cascade classifier, training tools provided by OpenCV were used. The version of OpenCV is 3.4 and the training routines in our framework are

opencv_createsamplesd : Creating OpenCV vector datafile for positive and negative data.

opencv_traincascade : Using the vector datafiles created by the previous routine to train the cascade classifier.

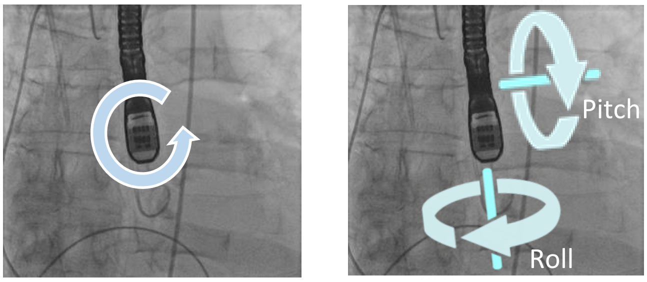
The details of training instruction can be found in [23]. The type of boosted classifier is *Gentle AdaBoost*. The minimal desired hit rate is set to 0.999 and the maximal desired false alarm rate is set to 0.5.

The process of training the cascade classifier is to create a collection of stages, where each stage is an ensemble of weak learners. The weak learners are simple classifiers called decision stumps which are machine learning models consisting of a one-level decision tree. Each stage is trained using a technique called boosting. Boosting provides the ability to train a highly accurate classifier by taking a weighted average of the decisions made by the weak learners. Therefore, the final product of training is a strong classifier which is a linear combination of weighted weak classifiers.

There are several important parameters for the training tools. They are the number of stages, the false positive rate and the feature type. The number of stages and the false positive rate are set depending on the size of the training data. As 400 positive images and 800 negative images (our dataset) are in the small dataset range, the number of stages is set to a lower number such as 10 and the false positive rate is also set to a lower value such as 0.5. There are two types of features: Haar and Local Binary Patterns (LBP). LBP have many advantages over Haar such as short training time, robustness to local illumination change and robustness to occlusion. Therefore, a cascade training tool using LBP with 10 stages was used. The time required to complete training was 2 hours 36 minutes using OpenCV 3.4 with multi-thread support. This was done in an Intel Core i7 2.9 GHz laptop.

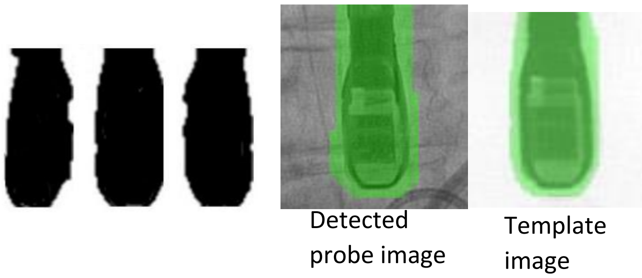
2.2 Creating a template library from CBCT

The cascade classifier is able to detect the location as well as in-plane rotation (figure 3(a)) and scale of the TEE probe. These four parameters are called in-plane parameters as they are in the plane of the X-ray image. There are two additional rotations: roll and pitch, which are out-of-plane. Roll and pitch angles could not be detected by the cascade classifier; they have to be found in a different way. Therefore a template library with binary masks is used to detect both out-of-plane rotation angles.



(a) In-plane rotation (b) Roll and pitch (out-of-plane)
Figure 3. Three rotations of the TEE probe

The template library is a comprehensive collection of images of the TEE probe in different roll and pitch rotation angles. It is created using CBCT, which acquires a series of X-ray images using a rotating C-arm. From CBCT, a digitally reconstructed radiography (DRR) model of the TEE probe can be constructed. The DDR model is a series of stacked 2D images, which represents the 3D model of the TEE probe. From the DDR model, simulated X-ray images can be created in any angle of roll and pitch rotation. However, searching a large template library is computationally expensive. As the TEE probe is sitting inside the oesophagus during the procedure, the probe is not free to move in all directions. Our template library covers the pitch angle from -48° to 48° and the roll angle from -90° to 90°. The angle interval is 2°. As X-rays can penetrate through the TEE probe, it makes objects appear the same under symmetrical poses. Therefore, producing template images of roll angle from -90° to 90° is sufficient to cover the full range of roll rotation (360°). Two sets of image masks are created. One is generated by binarizing images inside the template library. The other one is created from detected probe images, which are extracted by the cascade classifier. Those image masks are used to further reduce the computational cost of registering between a detected probe image and a template image. The image mask applied in the detect probe image can reduce influence from neighbourhood objects (wires or catheters) during the registration process. An example of the detect probe image and its image mask are illustrated in figure 4b. The binarization of template images was done automatically by an adaptive binarization method: Otsu's method [24]. Otsu' method is a non-parametrized and adaptive algorithm as it automatically determines the thresholding level based on minimizing the intra-class variance. The resulting binary images were verified by an imaging scientist to ensure that important features of the TEE probe are visible in the binary image. A contour-finding algorithm [25] computes all possible contours inside the binary image. The longest contour is selected and enlarged to create an image mask (figure 4a). The image mask for the detected probe image is also created from the enlarged contour of TEE probe and the method of computing the contour can be found in section 2.3.



(a) Image masks (b) Registration using masks
Figure 4. Examples of image masks and their usage in the image registration. The green areas are image masks.

2.3 Probe detection

After training the cascade classifier, it can be used to detect the TEE probe in live X-ray fluoroscopic images. The initial estimated scale is input into the classifier so that it is not necessary to search the X-ray image over a large range of search window sizes. This can substantially reduce the computational cost. To calculate the size of the search window, the size of the TEE probe in the template library is used. However, live X-ray images are often acquired under different X-ray system settings compared with settings used for image acquisition for the template library. Therefore, the size of the TEE probe measured in image pixels needs to be converted into a physical size in mm first. To convert into physical space (mm), the pixel to mm ratio (R_{dicom}) is obtained from the X-ray Dicom image header. The magnification factor M of the X-ray system is also estimated, which is based on D_{det}/D_{pat} (D_{det} is the distance from the X-ray source to the detector, and D_{pat} is the distance from the X-ray source to the patient). D_{pat} is estimated by using the distance between the X-ray source to the X-ray table surface which is provided by the real-time data streaming from the X-ray system. The real pixel to mm ratio is defined as: $R_{real} = R_{dicom}/M$. Therefore, the estimated size of search window (S_w) for the cascade classifier is computed as:

$$S_w = S_t * R_{real}/R_{real\_live} \quad (1)$$

Where S_t is the size of TEE probe in the template library which is measured in image pixels. R_{real_live} is the real pixel to mm ratio for live X-ray images and R_{real} is the real pixel to mm ratio for images in the template library.

The position of the probe is detected as the centre of the search window and the scale is computed by using Eq. (1). For detection of in-plane rotation angle, the cascade classifier has relatively low accuracy. To improve the detection accuracy, additional steps were added after applying the cascade classifier. First, the detected probe image was cropped from the live X-ray image. Then, the cropped image was down-sampled to 20% of the original image size. A multiscale vessel enhancement filter [26] was applied to the down-sampled image. It is used to enhance the visibility of wire-like structures in the X-ray images, which is based on the idea of approximating wire-like objects, such as tubular or cylindrical structures (the TEE probe head is similar to a tubular structure in the down-sampled image). Image down-sampling is not only to reduce the computation load but also to reduce or remove the vessel enhancement effect on wires, catheters or other small surgical instruments. The multiscale parameter of the vessel enhancement filter was set to the estimated size of the diameter of the TEE probe head so that only the TEE probe is enhanced after applying the filter. The enhanced image is binarized using Otsu's method [24]. Finally, a contour-finding algorithm [25] computes all possible contours inside the binary image. The longest contour is selected and principle component analysis (PCA) is applied to the contour to compute the orientation vector of the TEE probe as the first eigenvector. The workflow is presented in figure 5.

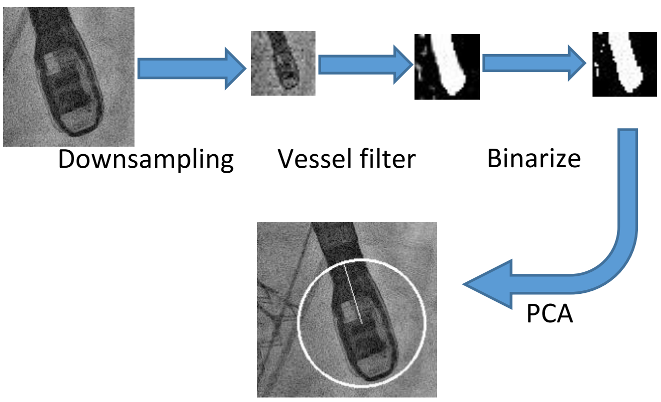


Figure 5. The workflow of computing the orientation vector. In the final picture, the circle indicates the scale of the TEE probe and the white line indicates the orientation vector.

2.4 Probe 2D-3D registration

The final step of pose estimation is to compute the out-of-plane parameters: the roll and pitch rotations. Our template library has a total of 4320 images. The number of images is calculated as following:

$$4320 = \frac{180}{2} \times \frac{96}{2} \quad (2)$$

Where 180° is the range of roll angle and 96° is the range of pitch angle. 2° is the angle interval. For multi-threaded implementation, 32 threads were used in a PC with an AMD Ryzen 9 3.5 GHz CPU (16 cores, support 32 processing threads). Therefore, each thread only computes a maximum of 135 image similarity measurements. The normalized cross correlation is used to compute the similarity between the detected probe image and an image from the template library. The normalized cross correlation can be computed by using Fast Fourier Transform (FFT) and can reduce the computational cost by up to 95% [27]. Therefore, a real-time performance could be achieved by using just multi-threaded implementation.

For initial probe pose detection, the full template library (1,080 images) is searched. As the TEE probe is sitting inside the oesophagus during the procedure, clinicians must move and adjust angles slowly to avoid damaging the oesophagus. Therefore, after obtaining the initial pose, the search range of roll and pitch rotations is halved so that only half the template images will be searched. An improved performance will then be achieved for probe pose tracking.

2.5 TEE image volume registration

The 3D TEE image volume can be visualized in the 2D X-ray fluoroscopic image by aligning the TEE and X-ray system coordinate systems. The transformation matrix, $T_{TEE\_to\_Xray}$, which transforms from 3D TEE image space to 2D X-ray image space consists of a rigid body transformation matrix T_{rigid} and a projection matrix T_{proj}.

$$T_{TEE\_to\_Xray} = T_{proj}T_{rigid} \quad (3)$$

The projection matrix transforms from 3D X-ray C-arm space to 2D X-ray image space. This can be calculated by using the intrinsic parameters of the X-ray system [28]. T_{rigid} can be decomposed into two matrices.

$$T_{rigid} = T_{model\_to\_C-arm}T_{TEE\_to\_model} \quad (4)$$

Where $T_{model\_to\_C-arm}$ transforms from 3D TEE model space (Model was acquired by CBCT) to 3D X-ray C-arm space. This matrix is generated by the probe detection algorithm and probe image registration method that positions the 3D TEE model in C-arm space. $T_{TEE\_to\_model}$ relates the position of the

3D TEE images to the position of the 3D TEE model. This is the TEE probe calibration matrix and is calculated pre-procedurally using a specifically designed calibration phantom.

2.6 TEE probe calibration

$T_{TEE\_to\_model}$ is determined using a probe calibration procedure. The calibration phantom consists of a 9-L water tank and two thin metal wires. Nine metal landmarks which are visible in both X-ray and ultrasound were placed on the wires. The TEE probe was rigidly fixed beneath the wires during data acquisition. X-ray images were acquired from left anterior oblique (LAO) 45°, right anterior oblique (RAO) 45° and posterior-anterior (PA) projections using a Philips Allura Xper FD10 C-arm X-ray system, which has an internal mechanism to track the C-arm position in real-time. Simultaneously an echo volume was acquired in full volume mode, giving the maximal volume coverage possible with the TEE probe. The automatic 2D-3D registration method based on probe detection and probe image registration was then utilized to align the 3D TEE model with the X-ray images acquired from PA and LAO views. The third X-ray image, which was acquired from RAO 45° was used to confirm the accuracy of the TEE probe localization.

The 3D positions of the nine landmarks $P_{phanEcho}$ were identified manually from the TEE image data. The landmarks were also clearly visible in the X-ray images. By manually defining the 2D position of the landmarks in the PA and LAO 45° X-ray images, their 3D positions in C-arm space $P_{phanC-arm}$ could be reconstructed using back-projection [24]. The calibration procedure was repeated for three different probe positions. A classic hill-climbing optimization algorithm was employed to find $T_{TEE\_to\_model}$ by minimizing the Euclidean distance error ε given by

$$\varepsilon = \left\| T_{TEE\_to\_model} P_{phanEcho} - T^{-1}_{model\_to\_C-arm} P_{phanC-arm} \right\| \quad (5)$$

In order to validate the accuracy of the calibration, a further 2 TEE volumes were acquired of the calibration phantom along with X-ray images in the PA and LAO 45° views.

3. Experiments

The proposed real-time framework was validated on phantoms and clinical datasets which were acquired during cardiac interventional procedures.

3.1 Imaging equipment and data acquisition

For all clinical cases and phantom studies, we used an iE33 3D real-time echo system with an X7-2t 3D TEE probe (Philips Healthcare, Andover, Boston, USA) for TEE acquisition. All clinical data collections have been approved by the Local Research Ethics Committee.

We evaluated the accuracy of our method using a realistic heart phantom (Ultrasound Heart Phantom, Computerized Imaging Reference Systems, Inc., Virginia, USA). The phantom has completely anthropomorphic external and internal anatomy including left/right ventricles, left/right atria and the valves. For TEE acquisition, the TEE probe was placed on the acoustic surface of the heart phantom. Six TEE volumes were acquired to cover different sections of the phantom by varying the TEE probe position. For X-ray image acquisition, we used a Philips Allura Xper FD10 C-arm X-ray system, the same X-ray system used for the TEE probe calibration.

We collected data from 7 cardiac electrophysiological (EP) procedures. All patients had left atrial flutter and were under general anesthesia (GA) during the procedures. For five patients, TEE and X-ray data were acquired after two deca-polar catheters were inserted into the right atrium (RA), one forming a loop along the endocardial surface of the RA and the other inserted into the coronary sinus (CS). Both catheters were visible in the TEE volume. For two patients, a trans-septal puncture was performed to gain access to the left atrium (LA). TEE volumes were acquired after a lasso catheter and an ablation catheter were inserted into the LA. The movement of the C-arm was limited by other equipment such as the anaesthesia system and the ultrasound scanner. X-ray images covering 4-5 cardiac cycles were acquired from PA and either RAO 30° or LAO 30° projections.

TEE and X-ray data were also collected from 3 trans-catheter aortic valve implant (TAVI) procedures. All procedures were performed in a GE catheter laboratory equipped with a GE Innova 2100^{IQ} C-arm Xray system. Similar to the Philips Allura Xper FD10, this GE X-ray system can precisely track its C-arm position automatically. The replacement valves were either delivered using the trans-femoral approach or trans-apical approach. X-ray images and TEE volumes were acquired in the same way as the data acquisition in the EP procedures after the replacement valve reached the deployment site.

A total of 6,492 images (52 image sequences) were acquired in 10 clinical cases and phantom studies. The frame size of each sequence is 512×512 which is the native resolution of the clinical images, with pixel sizes between 0.342 mm and 0.433 mm. Among the 6,492 images, 5,520 images were acquired during clinical cases in St. Thomas hospital, London, UK and the remaining images were acquired during phantom studies. 400 of the images (200 from phantom studies and 200 from clinical cases) were used as training data for the cascade classifier. The remaining 6,092 images were used as testing images for the evaluation of the proposed detection framework.

3.2 Evaluation for probe detection

The quantitative performance evaluation of the in-plane parameters (translation T_x, T_y and rotation ϑ) detection was performed on all test images (6,092 images). The results of accuracy evaluation are summarized in Table 1.

For the clinical data the average in-plane position (T_x, T_y) error was 1.7 and 2.1 mm and the in-plane orientation error was 2.9°. Position errors were measured as the nearest distances from the contour of detected TEE probe head to the contour drawn by one clinical expert (gold standard data). The gold standard data were manually defined using a Catmull-Rom spline curve. Figure 6(a) gives an example. The orientation error was calculated as the angle difference between the detected orientation line and the line drawn by the clinical expert. The clinical expert was instructed to draw the line along the transducer array. Figure 6(b) gives an example. Errors in the position estimation are mainly caused by false detections along the shaft of the probe. Other causes of errors are caused by proximity to dense tissue. The true positive rate is 0.94 and the false positive rate is 0.08. The true positive rate is defined as TP/P, which TP is the number of objects correctly labelled by the cascade classifier as the TEE probe and P is total number of objects labelled as the TEE probe. The false positive rate is defined as FP/N, which FP is the number of objects incorrectly labelled as the TEE probe and N is total number of objects labelled as the background. Detection in low dose images has lower accuracy compared with normal dose images. Low dose screening images are of low quality and are generally acquired to aid navigation. For higher quality, high dose cine images can be acquired, which are well contrasted and less noisy than low dose screening images. Those images are classified as normal dose X-ray images in this paper.

Figure 6 illustrates detection examples and the nature of clinical images with cluttered background and low textured probe.

Table 1. Accuracy evaluation of the in-plane position (T_x, T_y) and orientation (ϑ)

	Average Error		
Test Data	T_x(mm)	T_y(mm)	θ
Calibration phantom	1.1±0.57	1.3±0.68	2.1°±1.3
Heart phantom	1.3±0.71	1.6±0.82	2.5°±1.6
Clinical Data	1.7±0.93	2.1±1.11	2.9°±1.9
Normal dose	1.5±0.77	1.9±0.93	2.6°±1.6
Low dose	2.0±1.15	2.3±1.28	3.2°±2.2

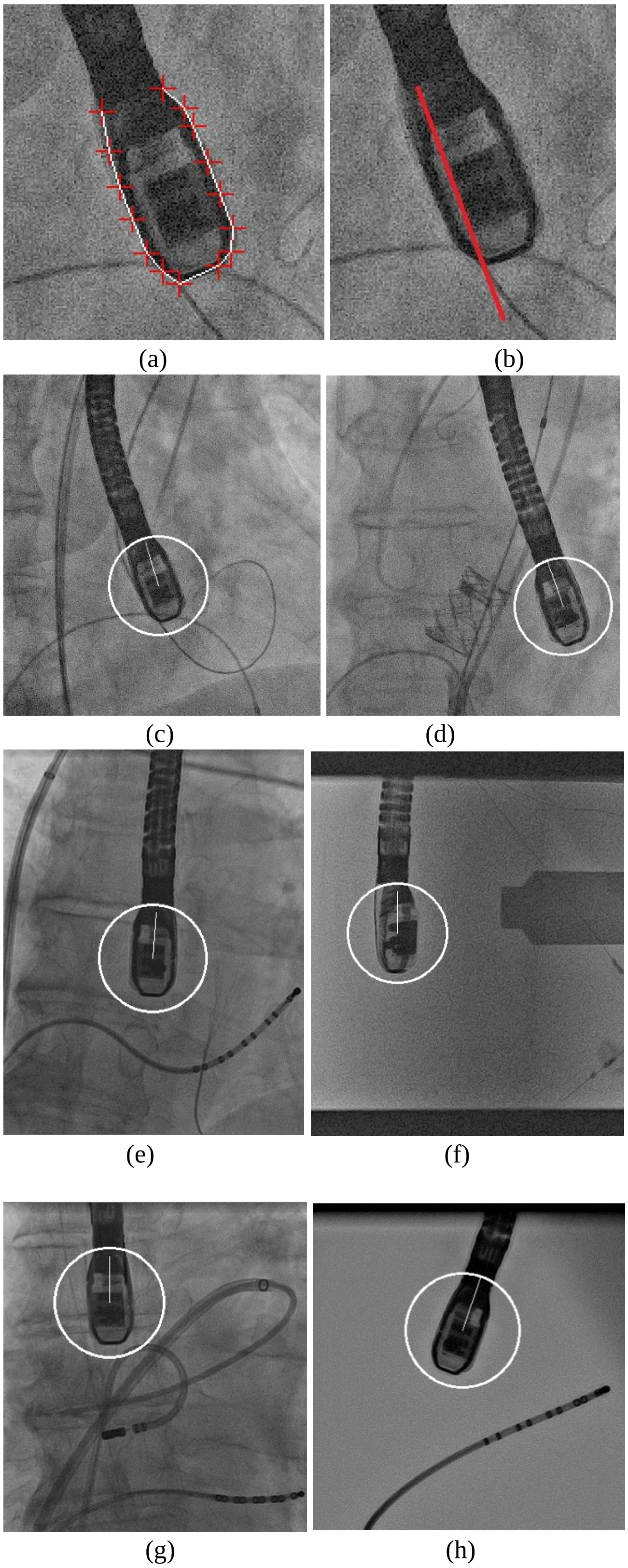


Figure 6. Examples of probe detection and estimation of in-plane parameters. The white circle indicates the size of TEE probe and the white line indicates the direction. The start point of the white line is the centre of TEE probe. (a) Manual annotation of the contour of TEE probe. (b) Manual annotation

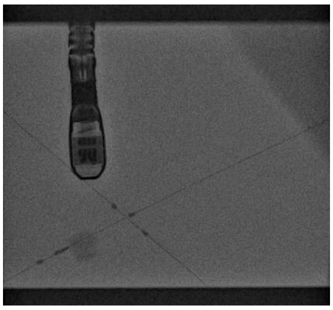
of the orientation vector of TEE probe. The red line is the orientation vector. (c)(d) Low dose images in TAVI cases. (e)(g) Normal dose images in EP cases. (f) A normal dose image in the phantom calibration study. (h) A normal dose image in the heart phantom study.

3.3 Evaluation for probe 2D-3D registration

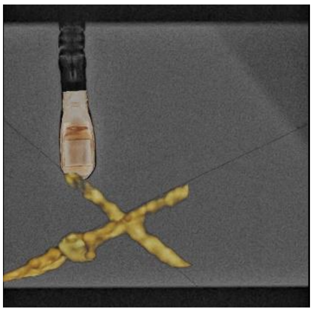
Overall accuracy of probe 2D-3D registration including accuracy evaluation of out-of-plane parameters was carried out by using target registration error (TRE). For the calibration phantom study which is described in section 2.6, TRE is defined as error between corresponding points on the X-ray and echo projection images of the crossed wires. These error measurement points were defined by manually fitting straight-line models to the projected images of the crossed wires in the two imaging modalities. The crossing points were then detected automatically from these fitted lines and corresponding measurement points were automatically defined in fixed steps along the lines from the crossing point. This provided 10 to 13 measurements per overlay view. Figure 7(a)-(c) shows an example of this. 2D projection errors were calculated between corresponding points and averaged to give an overall alignment error (e) for the overlay view as follows

$$e = \frac{1}{N}\sum_{i=1}^{N}\left|P_{i,xray} - P_{i,echo}\right|R_{dicom}\frac{D_{pat}}{D_{det}} \qquad (6)$$

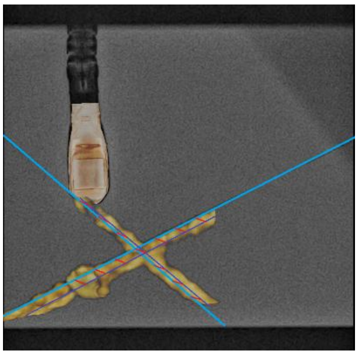
Where N is the number of measurements and $P_{i,xray}$ and $P_{i,echo}$ are the locations of corresponding points i in the overlay view. R_{dicom} is pixel to mm ratio from X-ray Dicom image header. D_{det} is the distance from the X-ray source to the detector, and D_{pat} is the distance from the X-ray source to the patient. D_{det}/D_{pat} is the magnification factor of the X-ray system. The TRE is 2.08 ± 0.61 mm and maximum error is 2.6 mm with the magnification factor ranging from 1.26 to 1.51.



(a)



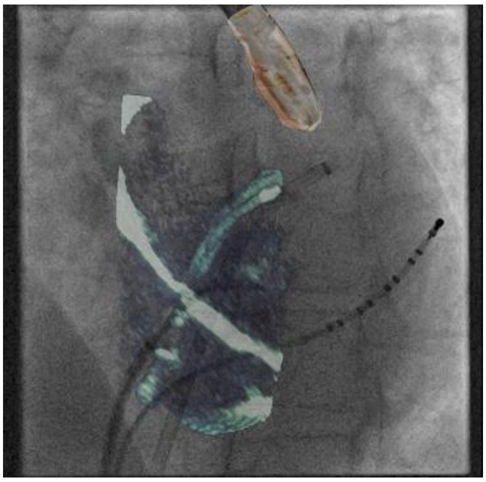
(b)



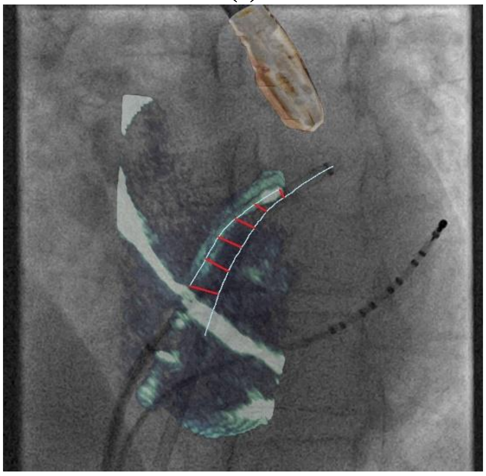
(c)

Figure 7. Phantom model experimental overlay of fluoroscopic and echocardiographic images. Errors were measured between automatically defined points on straight line models of the crossed wires. (a) Original image. (b) Echo overlay. (c) Error measurements. Blue lines are centerlines of crossed wires in both X-ray and echo images. Red lines are error distances.

Similar to the error measurement in the phantom study, the TRE for clinical studies is defined as error distances between corresponding points in both X-ray and echo images. Although real-time synchronised visualisation of the live data stream was possible during the clinical procedures, the post-procedure analysis for this paper required that the recorded X-ray and echo data were synchronised manually, resulting in only approximately synchronised sequences. The manual synchronization was done through visual matching using landmarks such as catheters or artificial valves. Automatic registrations using our framework were performed at two separate frames in each X-ray sequence (22 overlay views generated). Corresponding catheters were manually defined in the echo and X-ray views using spline curves. Equally spaced points along the echo curve were automatically defined as measurement points. The corresponding X-ray point was defined as the closest point on the X-ray curve. The alignment error for each overlay view was again taken as the average of the 2D errors between corresponding points. An example of these error measurements is given in figure 8(b). Average errors were measured using between 4 and 6 point pairs per overlay view, depending on the length of catheter visible in the echo image. In all cases, catheters were used for the measurements, because they were the most consistently visible objects in the echo images. Figure 9 gives some examples of echo X-ray overlay views during the EP cases.

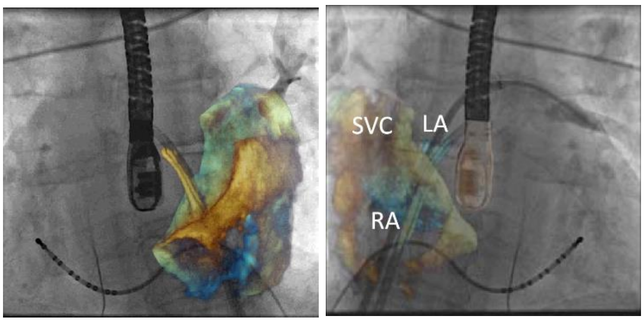


(a)

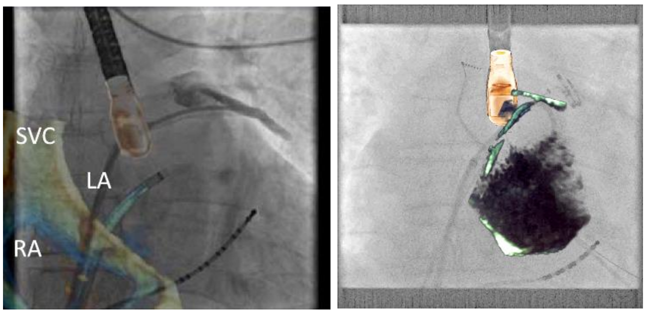


(b)

Figure 8. An example of error measurement. Errors are measured as the shortest distance from automatically defined points on the echo catheter image to a spline model of the X-ray catheter. (a) Echo X-ray overlay. (b) Error measurement. Red lines are the shortest distances.



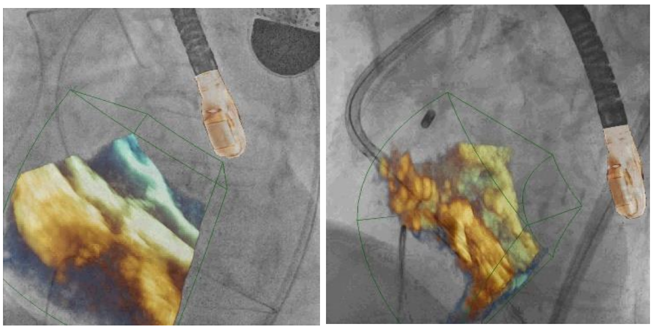
(a) (b)



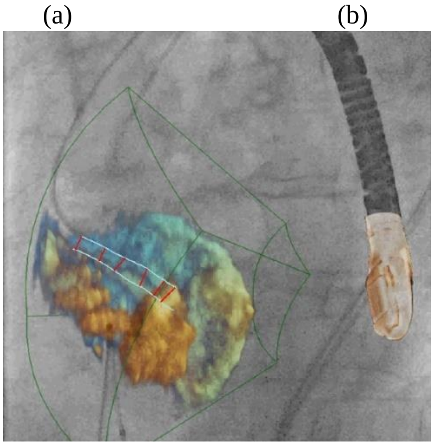
(c) (d)

Figure 9. Example of echo X-ray overlay in atrial fibrillation ablation cases. Real-time overlay views were generated during the trans-septal puncture. SVC: superior vena cava. LA: left atrium. RA: right atrium.

For TAVI cases, the procedure was performed under general anesthesia in a GE catheterisation laboratory. The TEE probe was positioned to view the left ventricular outflow tract and ascending aorta, aortic valve and left ventricle. After applying our probe 2D-3D registration method, TEE image volume can clearly visualize the native aortic valve as well as ascending aorta and register the TEE image with live X-ray fluoroscopic images. For error measurement, the guide wire was used (see figure 10 (c)). The alignment errors for both EP and TAVI cases are presented in figure 11. From a total of 120 alignment error measurements made, a median error of 2.6 mm was achieved.



(a) (b)



(c)

Figure 10. Example of echo X-ray overlay in TAVI cases. Real-time overlay views were generated during the

deployment of the artificial valve within the native aortic valve. (a) Visualization for the native aortic valve. (b) Visualization for ascending aorta. (c) Error measurements.

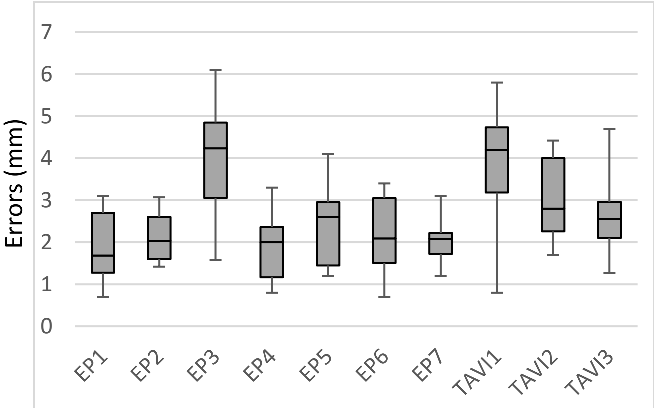


Figure 11. The box plots for TRE errors in clinical cases. The first 7 box plots were obtained from EP cases and remaining three were from TAVI cases.

3.4 Performance evaluation

The proposed real-time probe 2D-3D registration framework enables detection at 9 fps for the TEE probe in live X-ray images using only multi-threaded implementation (32 threads). Furthermore, our framework can achieve 20 fps for tracking TEE probe by reducing the search range by half. The frame rate was evaluated in a PC with an AMD Ryzen 9 3.5 GHz CPU (16 cores, support 32 processing threads). The tracking frame-rate is considered as real-time for cardiac interventional procedures as frame rates over 15 fps are rarely employed. In the low dose X-ray image setting, the frame rate for X-ray systems will drop to an average of 7.5 fps to reduce X-ray radiation doses.

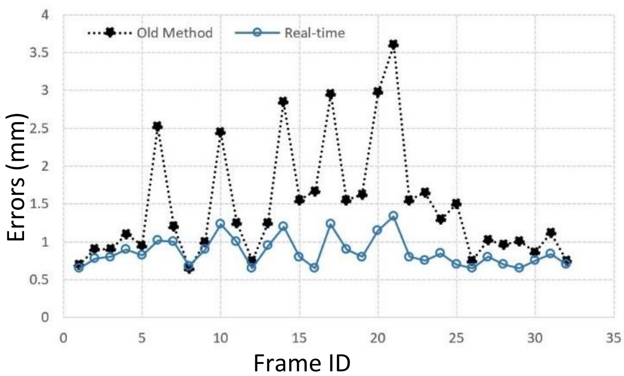
The cascade classifier is very efficient in our framework. It achieved average 182 fps on all image sequences with the native resolution (512x512) and achieved average 98 fps on images with the scale-up resolution (1024x1024). However, in order to achieve 9 fps detection speed and 20 fps tracking speed for any image sequence with the resolution of 1024x1024, the image sequence needs to be down-sampled to 512x512.

4. Discussion

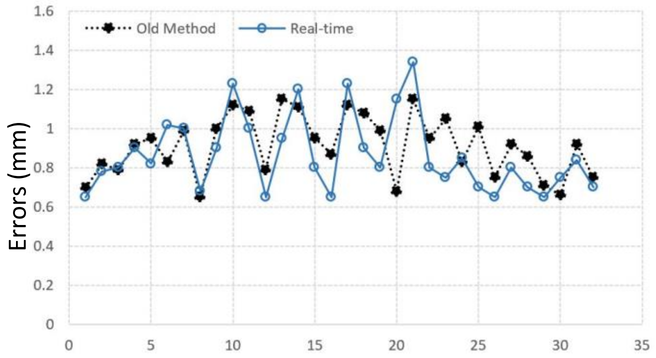
The main aim of the study was to develop an automatic and real-time TEE probe localization framework and evaluate it in the application for overlaying TEE image volumes on live X-ray fluoroscopic images, which could be used during cardiac interventional procedures. This framework does not involve the use of any additional tracking devices and therefore can be easily integrated into the current workflow of the catheter laboratory. The fundamental advantage of combining 3D echo and X-ray images is the ability to better appreciate soft tissue anatomy in relation to the catheter or guidewire during navigation and device deployment.

For example, in the EP cases, a 3D echo volume on its own can be difficult to interpret because of its limited field-of-view and lack of context for the echo coordinate system relative to the patient. In addition, catheters tend to produce artefacts in the ultrasound data reducing the clarity of the images. However, in the overlay view, the echo volume is displayed in a coordinate system that can be more easily related to the patient. Also, the highly visible catheters in the X-ray image help with identifying the catheters in the echo and so can be related to the cardiac anatomy via the echo image. The precise location of the catheter tip and electrodes on the lasso catheter in relation to atrial tissue permits better targeted ablation delivery than just visualising the catheter as a whole on fluoroscopy. Similarly during a TAVI procedure, being able to visualise both the native valve and the artificial valve on a single image will facilitate delivery and deployment of the artificial valve.

One of the accuracy requirements for a clinically useful image guidance system is 5 mm. The choice of 5 mm is motivated by the size of the smallest target structures (pulmonary veins, approximately 5 mm in radius) for cardiac interventional procedures. Based on the results reported in section 3.3, the system is sufficiently accurate to guide procedures in real-time. Our framework can achieve an accuracy of 2.4±1.1 mm for EP cases and 2.9±1.2 mm for TAVI cases. The other accuracy requirement is 3.2mm, which is the diameter of the sheath for a larger guiding catheter (8Fr) used in cardiac intervention. The majority of registration errors (83.3%) are below 3.2mm. Compared to our previous framework, the errors caused by patients' respiratory and cardiac motions are largely reduced. The majority of the probe motions caused by both respiratory and cardiac motions is translational. As the pose of TEE probe is continuously tracked throughout the whole image sequence, both translation and rotation movements are captured by our framework so that lower errors can be achieved. Figure 12 presents the comparison of translation (position) errors with in our old method [10].

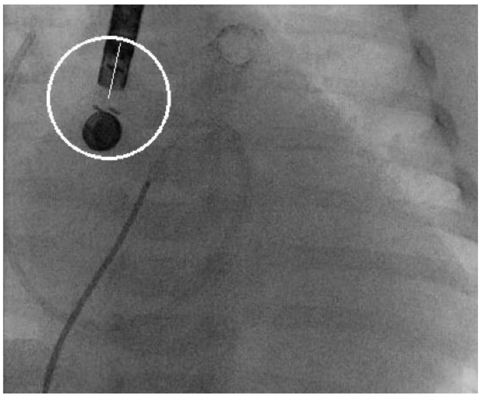


(a) Real-time comparison

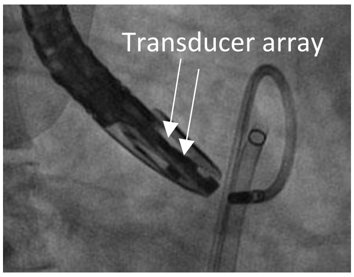


(b) Comparison (frame by frame)

Figure 12 Comparison of translation (position) errors. The black dot lines are translation errors in our old method [10]. The solid lines are translation errors generated by proposed new real-time framework. As our old method is not able to track the probe in real-time, the translate errors are mainly caused by cardiac motions. In the other hand, the current method is capable to remove those large errors by apply the real-time tracking method. (a) Real-time comparison. (b) Comparison after removing tracking delay (results are calculated frame by frame).



(a) Detecting a different TEE probe in an X-ray image. The white circle indicates the size of the TEE probe and the white line indicates the direction.



(b) In an extreme roll angle, the main feature of the TEE probe head has become a thick line.

Figure 13. The potential and drawback of our method.

The proposed framework is not limited to certain types of TEE probe. If the training data is available (even a relatively small dataset such as 200 images), the cascade image classifier could be trained to detect and localize the target TEE probe in X-ray images. Figure 13(a) gives an example of detecting a Philips S7-2T TEE probe (a 2D probe with axial rotation).

In this study, we have demonstrated real-time TEE probe detection and tracking and that it facilitates hybrid X-ray fluoroscopy and 3D echo visualisation. We anticipate that by gaining familiarity and confidence within this hybrid viewing system, there will be a lesser need for repeated X-ray fluoroscopy use to establish catheter position. With increased usage over time, we predict a greater reliance on echo views to assist catheter or guidewire navigation. In the long run, this will reduce overall fluoroscopy time and patient exposure to ionising radiation.

4.1 Limitations and further works

The position errors are higher when the TEE probe has extreme roll angles ((out-plane rotation) such as -90° and 90°. As the TEE probe in extreme angles has less distinguish features for the cascade classifier to localize the position along the shaft of the probe, higher position errors were generated by false detections along the shaft of the probe. As shown in figure 13(b), the main features of the probe head such as the ultrasound transducer array has become a thick line instead of a rectangular object. Therefore, the position of detection could slip along the shaft of the probe. If there is a failure of detection by the cascade classifier, manual initialization should be used to correct the position of the TEE probe and out-of-plane parameters still can be calculated automatically.

The reason for less accurate localization by the cascade classifier is that the classifier is the sum of several weak classifiers and the transducer array is the dominate feature. Therefore, the weak classifier for detecting transducer array carries a higher weight than other weak classifiers. When the transducer array becomes a thick line instead of a rectangular object, the cascade classifier will become less accurate and generate false detection along the shaft of the probe. The potential solution for improving accuracy in the extreme angle is to train a second cascade classifier dedicated to the extreme angles.

5. Conclusion

This paper presents a novel and real-time TEE probe detection and tracking framework, which is based on a cascade image classifier and a 2D-3D image registration method. The proposed framework works robustly in both normal dose and low dose X-ray fluoroscopic images and it can be applied to overlay real-time TEE image volumes with live X-ray fluoroscopic images. The proposed framework does not require any user interaction and the TEE probe could be continuously tracked throughout the whole image sequence in real-time. As the proposed framework requires a relatively small number of images for training the cascade classifier, it could potentially be applied to all types of TEE probes used in cardiac interventional procedures. This could enable wider usage of real-time TEE echo in cardiac interventional procedures and combine the strength of both echo and X-ray image modalities.

Acknowledgements

This research was supported by the National Institute for Health Research Biomedical Research Centre at Guy's and St. Thomas' NHS Foundation Trust and King's College London and the Wellcome/EPSRC Centre for Medical Engineering [WT 203148/Z/16/Z].

References

[1] Kwan J. 2014 Three-dimensional echocardiography: a new paradigm shift *Journal of echocardiography* **12(1)** 1-11.

[2] Regoli F, Faletra FF, Scaglione M, et al. 2012 Pulmonary vein isolation guided by real-time three-dimensional transesophageal echocardiography *Pacing and Clinical Electrophysiology: PACE* **35(3)** e76-9.

[3] George AK, Sonmez M, Lederman RJ, Faranesh AZ 2011 Robust automatic rigid registration of MRI and X-ray using external fiducial markers for XFM-guided interventional procedures *Medical Physics* **38(1)** 125-141

[4] Vernikouskaya I, Rottbauer W, Gonska B, et al 2017 Image-guidance for transcatheter aortic valve implantation (TAVI) and cerebral embolic protection *International journal of cardiology* **249** 90-95.

[5] Housden, R. J., Arujuna, A., Ma, Y., et al., 2012. Evaluation of a real-time hybrid three-dimensional echo and x-ray imaging system for guidance of cardiac catheterisation procedures *MICCAI 2012* **15** 25-32.

[6] Hatt, C. R., Tomkowiak, M. T., Dunkerley, D. A., Slagowski, J. M., Funk, T. , Raval, A. N. and Speidel, M. A. 2015 Depth-resolved registration of transesophageal echo to x-ray fluoroscopy using an inverse geometry fluoroscopy system *Medical Physics* **42** 7022-7033

[7] Mountney P., Ionasec R., Kaizer M., Mamaghani S., Wu W., Chen T., John M., Boese J. and Comaniciu D. 2012 Ultrasound and fluoroscopic images fusion by autonomous ultrasound probe detection *MICCAI 2012* **15** 544-551.

[8] Hatt C. R., Jain A. K., Parthasarathy V., Lang A., Raval A. N. 2013 MRI-3D ultrasound X-ray image fusion with electromagnetic tracking for transendocardial therapeutic injections: In-vitro validation and in-vivo feasibility *Computerized Medical Imaging and Graphics* **37** 162–173

[9] Jain A., Gutierrez L., and Stanton D. 2009 3D TEE registration with x-ray fluoroscopy for interventional cardiac applications *FIMH 2009, Lecture Notes in Computer Science* **5528** 321–329.

[10] Gao, G., Penney, G., Ma, Y., et al. 2012 Registration of 3D trans-esophageal echocardiography to X-ray fluoroscopy using image-based probe tracking *Medical Image Analysis* **16** 38–49.

[11] Heimann T., Mountney P., John M. and Ionasec R. 2013 Learning without Labeling: Domain Adaptation for Ultrasound Transducer Localization MICCAI 2013 **16** 49-56.

[12] Hatt C. R., Speidel M. A., Raval, A. N. 2016 Real-Time Pose Estimation of Devices from X-ray Images: Application to X-ray/Echo Registration for Cardiac Interventions *Medical Image Analysis* **34** 101-108

[13] Ma Y., Gogin N., Cathier P., et al. 2013 Real-time x-ray fluoroscopy-based catheter detection and tracking for cardiac electrophysiology interventions *Medical Physics* **40(7)** 071902-1- 071902-13

[14] Ma Y., Alhrishy M., Narayan S. A., Mountney P., Rhode K. S. 2018 A novel real-time computational framework for detecting catheters and rigid guidewires in cardiac catheterization procedures *Medical Physics* **45(11)** 5066-5079

[15] Hatt C. R., Speidel M. A., Raval, A. N. 2015 Hough Forests for Real-Time, Automatic Device Localization in Fluoroscopic Images: Application to TAVR *MICCAI 2015* **18** 307-314

[16] Miao, S., Wang Z, and Liao R 2016 A CNN regression approach for real-time 2D/3D registration *IEEE transactions on medical imaging* **35(5)** 1352-1363

[17] Nogueira, D., Abreu, P. and Restivo, M.T. 2019 Comparing classification techniques for identification of grasped objects. *BioMed Eng OnLine* **18(21)**

[18] Liao S., Jain A. K. and Li S. Z. 2016 A Fast and Accurate Unconstrained Face Detector *IEEE Transactions on Pattern Analysis and Machine Intelligence* **38(2)** 211-223

[19] Zhuang X, Kang W, Wu Q 2016 Real-time vehicle detection with foreground-based cascade classifier *IET Image Processing* **10(4)** 289–296

[20] Cheng W. and Jhan D. 2011 A cascade classifier using Adaboost algorithm and support vector machine for pedestrian detection *IEEE International Conference on Systems, Man, and Cybernetics, Anchorage 2011* 1430-1435

[21] Albani D., Youssef A., Suriani V., Nardi D., Bloisi D. 2017 A Deep Learning Approach for Object Recognition with NAO Soccer Robots *RoboCup 2016* 392-403

[22] Krishnamoorthy N., Ramasamy A., Jones I. 2016 Classification of Malignant and Benign Micro Calcifications from Mammogram Using Optimized Cascading Classifier *Current Signal Transduction Therapy* **11** 98-104

[23] Viola P., Jones M.J. Robust Real-Time Face Detection 2004 *International Journal of Computer Vision* **57** 137–154

[24] Otsu N. 1979 A Threshold Selection Method from Gray-Level Histograms *IEEE Transactions on Systems, Man, and Cybernetics* **9(1)** 62-66

[25] Suzuki, S. and Abe, K. 1985 Topological Structural Analysis of Digitized Binary Images by Border Following *Computer Vision, Graphics, and Image Processing* **30(1)** 32-46

[26] Frangi A. F., Niessen W. J., Vincken K. L. and Viergever M. 1998 A Multiscale vessel enhancement filtering MICCAI 1998 130-137

[27] Kaso A. 2018 Computation of the normalized cross-correlation by fast Fourier transform *PLOS ONE* **13**

[28] Hawkes, D.J., Colchester, A.C.F., Mol, C.B. 1987 The accurate 3-D reconstruction of the geometric configuration of vascular trees from X-ray recordings *Physics and Engineering of Medical Imaging, NATO ASI Series* **119** 250-256