Robust 5DOF Transesophageal Echo Probe Tracking at Fluoroscopic Frame Rates

Charles R. Hatt[1,2], Michael A. Speidel[1,2], and Amish N. Raval[2]

[1] Department of Medical Physics, University of Wisconsin - Madison, USA
[2] Division of Cardiovascular Medicine, University of Wisconsin - Madison, USA

Abstract. Registration between transesophageal echocardiography (TEE) and x-ray fluoroscopy (XRF) has recently been introduced as a potentially useful tool for advanced image guidance of structural heart interventions. Algorithms for registration at fluoroscopic imaging frame rates (15-30 fps) have yet to be reported, despite the fact that probe movement resulting from cardiorespiratory motion and physician manipulation can introduce non-trivial registration errors during untracked image frames. In this work, we present a novel algorithm for GPU-accelerated 2D/3D registration and apply it to the problem of TEE probe tracking in XRF sequences. Implementation in `CUDA C` resulted in an extremely fast similarity computation of $< 80\ \mu s$, which in turn enabled registration frame rates ranging from 23.6-92.3 fps. The method was validated on simulated and clinical datasets and achieved target registration errors comparable to previously reported methods but at much faster registration speeds. Our results show, for the first time, the ability to accurately register TEE and XRF coordinate systems at fluoroscopic frame rates without the need for external hardware. The algorithm is generic and can potentially be applied to other 2D/3D registration problems where real-time performance is required.

1 Introduction

Image registration between transesophageal echocardiography (TEE) and x-ray fluoroscopy (XRF) has generated interest in recent years [1] as a tool for enhancing image guidance during structural heart interventions. XRF is considered the primary imaging modality for real-time visualization of devices, while TEE can image soft-tissue structures in real-time. Combining these two modalities in a single visualization framework has the potential to provide the best of both worlds: simultaneous imaging of devices and cardiac anatomy.

Registration can be accomplished by estimating the 3D pose of the TEE probe in the XRF imaging space using a variety of methods. 2D/3D registration methods are capable of accurately estimating all pose parameters by iteratively matching a 3D model of the probe to the XRF image. In [2,3], this was accomplished by generating a digitally reconstructed radiograph (DRR) and computing a similarity metric between the DRR and the XRF image. DRR generation was significantly accelerated using a GPU-based raycasting algorithm, but the overall registration frame rate was still on the order of $0.3 - 2.0$ frames-per-second

(fps) for a single plane XRF system. During a cardiac procedure, the probe can move around quickly due to cardiorespiratory motion or physician manipulation. There is therefore a need for a registration method that performs at a typical XRF imaging frame rate (e.g. 15 fps).

In [4], accurate, 20 fps registration was accomplished by attaching radio-opaque fiducials to the TEE probe. Aside from the need for custom modification, this solution was undesirable due to its potential to increase the risk of esophageal injury. In [5], DRR generation was accelerated by modeling the TEE probe as a mesh and using `OpenGL` based rendering. DRR computation was significantly accelerated, but the overall similarity metric computation time (1 ms per iteration) is still not fast enough to perform registration at imaging frame rates based on results reported in this paper.

In this work, we report on a novel 2D/3D registration algorithm that can operate at between 23.6 and 92.3 fps. Similar to the work presented in [6], our algorithm efficiently computes an image similarity metric without explicitly generating any DRRs. In simulated and clinical datasets, our method performed similarly in terms of target registration error to standard methods combining raycasting with normalized cross-correlation (NCC) and gradient correlation (GCC) similarity metrics, but operated at much higher frame rates.

2 Methods

2.1 Algorithm

Image-based 2D/3D registration is accomplished by estimating the 3D location and orientation (pose) of an object, generating a DRR based on this estimate, and comparing the DRR to the XRF image using a computed image similarity metric. Using non-linear optimization, this process is repeated until the similarity converges to an optimal value. Since this is an iterative process, a key determinant of overall registration time is the time needed to generate the DRR at each step of the optimization. For our application, optimization typically requires 150-300 similarity computations, depending on how close the initial pose is to the final solution. This means that both DRR generation and similarity computation need to be completed in roughly 300 μs on average for 15 fps registration. DRRs are typically generated using raycasting techniques. These methods compute line integrals, at each DRR pixel, along simulated x-rays passing through a volume of interest. An alternative method is splatting, where a DRR is generated by spatially transforming a 3D point-cloud model, projecting the transformed points onto the detector plane, and summing up values at discrete detector pixel positions. The point model is often generated from a CT scan. Mathematically, the DRR can be expressed as:

$$D(u_i) = \sum_{j \in S_i} -V(x_j), \; S_i = \{ \; j \mid \lfloor P \cdot T \cdot x_j \rfloor = u_i \; \} \tag{1}$$

Here, S_i refers to the set of points that project onto pixel i after the model has been spatially transformed, x_j is the 3D coordinate of the jth model point, $V(x_j)$ is the intensity value associated with the jth model point, T is the spatial transformation matrix describing 6DOF translation and rotation, and P is the projection matrix modeling the imaging process. Fig. 1 demonstrates a comparison of TEE probe DRRs generated using raycasting and splatting.

When using only a CPU for computation, splatting can be more computationally efficient than raycasting, especially when the size of the point cloud is small compared to the size of the data volume it was derived from. However, splatting does not necessarily translate well to the GPU, because every projected point must perform an *indirect write* operation at a random pixel (see Fig. 2, left, line 3). However, it can be shown that the correlation similarity metric can be reformulated so that all of the write operations are replaced with extremely fast texture reads on the GPU.

Raycasting   Splatting 1,048,576 points   Splatting 65,536 points

Fig. 1. Methods for generating probe DRRs.

Consider the correlation (CC) between an XRF image $I(u)$ and a DRR $D(u)$:

$$CC = \sum_i I(u_i)D(u_i) \tag{2}$$

For a splat-generated DRR, we can substitute the $D(u)$ term to obtain:

$$CC = \sum_i I(u_i)(\sum_{j \in S_i} -V(x_j)) = -\sum_i \sum_{j \in S_i} I(P \cdot T \cdot \ x_j)V(x_j) \tag{3}$$
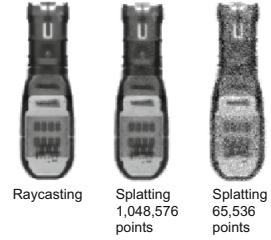
Finally, this simplifies to a sum over all 3D points:

$$CC = -\sum_j I(P \cdot T \cdot \ x_j)V(x_j) \tag{4}$$

This expression simply states that the correlation is equal to the sum, over all 3D points, of the point intensity times the value of the pixel that it projects onto. Equation 4 shows that the DRR does not need to be explicitly generated to compute the similarity, enabling more efficient computation (Fig. 2).

```
//Splat correlation w/ DRRs
1  for each point x
2    u = P*T*x
3    DRR[u]=DRR[u] + -V[x]
4  for each pixel u
5    cc+=DRR[u]*I[u]
```

```
//Direct method w/o DRRs
1  for each point x
2    u = P*T*x
3    cc+=I[u] * -V[x]
4
5
```

Fig. 2. Pseudo-code for computation of correlation for explicit DRR generation (left) and the proposed method (right).

Because this similarity is inspired by splatting and is directly computed from the image, without the explicit generation of a DRR, we refer to it as "direct splat correlation" (DSC).

2.2   Experiments

3D TEE Model A Philips X7-2t TEE probe was used in this study. The 3D model of the TEE probe was generated from a cone beam CT image (Philips FD20). A point cloud representation of the probe was created using the following procedure:

1. Features in the probe CT that were visually classified as having high-intensities were manually segmented.
2. 2^{16} points were randomly generated within the segmented region. This was the smallest number of points that fit the criteria for both computational efficiency and accuracy.
3. Linear interpolation was used to assign an intensity value to each point.

Computer Hardware and Software. All experiments were run on a Dell Precision T7500 work station running Ubuntu Linux with a 3.47 GHz Intel Xeon processor and a NVIDIA Tesla K20 GPU. VNL libraries were used for optimization, and all similarity functions were implemented in `CUDA C`. Retrospective clinical dataset processing was approved by the local institutional review board.

Simulations. The first set of experiments tested the accuracy and speed of the algorithm *in silico* compared to standard methods for XRF/TEE registration reported in [2,3]. Simulation images (I_{sim}) were a hybrid of real background anatomy (I_{xrf}) and synthetic DRRs (I_{drr}).

Synthetic DRRs were rendered using the splatting method with a point cloud large enough to generate high quality DRRs (2^{21} points). The background anatomy was obtained using images from transcatheter aortic valve implantation (TAVI) procedures and the hybrid image was formulated as:

$$I_{sim} = I_{xrf} \cdot e^{-\alpha I_{drr}} \tag{5}$$

The parameter α controlled the probe to background contrast and was randomly varied to generate a contrast ratio ranging from 0.45 to 0.85 for each experiment. Fig. 3 shows a few examples of the simulated images.

For each experiment, the TEE probe was placed at a random location and orientation within the XRF C-arm image space. Based on our observations from an image database of TAVI cases, the TEE probe rarely has Euler angle rotations outside the range of -75° to 75° about its primary axis (y-axis), -30° to 30° about the x-axis of the image detector, and -45° to 45° about the source-detector axis (z-axis). Once an initial pose was created, a random mis-registration was applied, which the experiments attempted to recover. The random mis-registration was
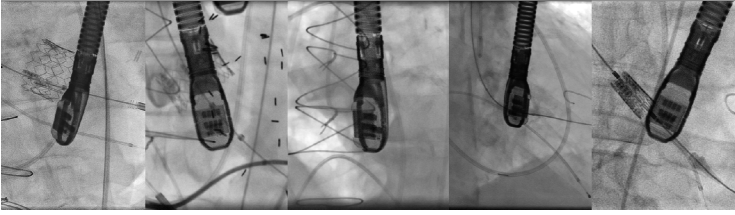
Fig. 3. Examples of simulated images

chosen from a zero mean uniform distribution over ranges of 3.0 mm, 3.0 mm, 5.0 mm, 30°, 30°, 6° for the parameters $t_x, t_y, t_z, \theta_x, \theta_y, \theta_z$, respectively.

Once the simulated image and mis-registration was generated, 3 different algorithms and 2 optimizers were tested. The first algorithm was the proposed algorithm (DSC), and the second and third were normalized cross-correlation and gradient correlation using raycasted DRRs (rcNCC, rcGCC, see [7] for details). The two optimizers tested were the Nelder-Mead (nm) and Powell (pwl) methods, both of which work well for low-dimensional optimization problems that do not have analytical cost function gradients.

The simulation experiments assumed a single x-ray projection, which typically results in inaccurate estimates of t_z. Therefore, we focused on optimization of the other five parameters, which resulted in a faster, more accurate optimization problem. Furthermore, when projecting echo data onto the XRF image, the errors in the estimation of t_z have little effect on target projection errors (for the same reason that t_z is difficult to estimate in the first place). Therefore, our accuracy metric was projection target registration error (pTRE), which was the root-mean-square error between known target points in XRF and estimated target points from echo following registration and projection to the XRF image:
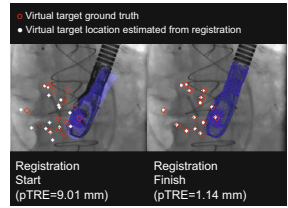
Fig. 4. Virtual targets used to compute pTRE before and after registration

$$\text{pTRE} = \sqrt{\frac{1}{N}\sum_n \left(\left\| \frac{1}{m_n}(p_n^{(xrf)} - P \cdot T \cdot p_n^{(echo)}) \right\|_2 \right)^2} \qquad (6)$$

where m_n is the projective magnification of point $T \cdot p_n^{(echo)}$.

pTRE, % of successful registrations, and frame rate were reported for each experiment. We chose to define a successful registration as a pTRE < 5.0 mm based on results from [3] where pTRE of 2.9 mm was the average error, but in reality this measure is application dependent. pTRE is only computed for successful registrations, to avoid large registration errors skewing the statistics. 5000 experiments were performed.

For all experiments, virtual target points were used to compute pTRE. The virtual target points were randomly generated from within the center of a virtual

ultrasound volume emanating from the TEE probe, at a mean distance of 50 mm from the probe face (Fig. 4). Errors from the probe model to echo volume calibration ($^{echo}T_{probe}$) were not considered in the analysis.

Clinical Datasets. Validation was also performed on real images from TAVI procedures. 39 image sequences from 19 cases (3635 frames) were identified as containing significant probe movement due to physician manipulation or cardiorespiratory motion. Due to a lack of ground truth, a surrogate ground truth was created for each sequence using the following procedure:

1. Manual registration was performed for the first frame of each sequence.
2. The initial manual registration was refined using the rcGCC-pwl method.
3. rcGCC-pwl was used to define the ground truth registration at each consecutive frame. Each sequence was then visually checked, frame-by-frame, for errors. The probe could not be tracked in four sequences and they were removed from the analysis.

For the clinical datasets, we examined the tracking accuracy and robustness under real clinical conditions where both image streaming and registration delays must be considered. Assuming 15 fps image streaming, the registration lag was calculated as $n_{lag} = ceil(15 \times t_{registration})$ frames. Each skipped frame was only allowed to use the most recently finished registration result for its pTRE calculation, and every registration (T_n) was initialized with the most recently processed registration result ($T_{n-n_{lag}}$). Note that slower registration times resulted in increased pTRE due to not only more skipped frames but also less accurate frame-to-frame initialization.

3   Results

Simulations Results for the simulation studies are shown in Fig. 5. In terms of pTRE, the rcNCC and rcGCC both slightly outperformed the DSC method, with the rcGCC-pwl combination having roughly half the error of the DSC-nm method. However, the mean and standard deviation of all methods fell within the range of clinical acceptability ($< 2.0\ mm$). Furthermore, the DSC-nm combination outperformed the other methods in terms of success rate, indicating that, although the other methods are slightly more accurate, the DSC-nm method is more robust. The average frame rate for DSC was $33.0 \pm 9.4\ fps$ (mean $\pm$ std). This is over an order of magnitude faster than the other methods, which were 3.0 ± 1.2 and $2.2 \pm 0.9\ fps$ for rcNCC-nm and rcGCC-nm, respectively.

Clinical Datasets Fig. 6 shows results for all clinical sequences. It can be seen that the lowest average pTRE was found using the rcGCC method, but the meaning of this result is unclear due to the fact that rcGCC was used as a proxy measure for the ground truth. Frame-rate increased compared to simulations, averaging $73.1 \pm 19.2\ fps$ for DSC-nm, $8.9 \pm 2.2\ fps$ for rcNCC-nm, and 5.9
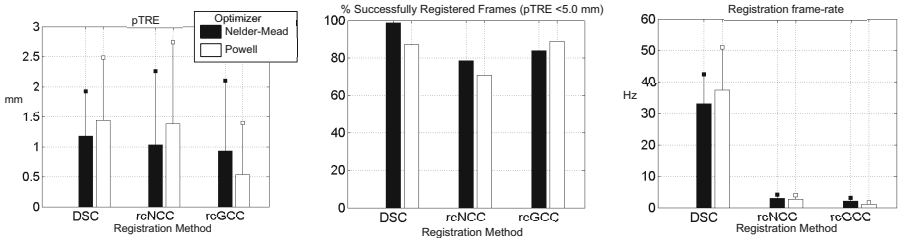
Fig. 5. Simulations. Left: Average pTRE for successful registrations. Comparisons between methods were all statistically significant ($p < 0.001$). Middle: Percentage of successfully registered frames (pTRE<5.0 mm). Right: Average frame rate.

± 1.5 fps for rcGCC-nm. This was due to the fact that the difference in pose parameters between frames was generally smaller for the clinical datasets and therefore roughly half of the similarity function computations were needed for convergence.
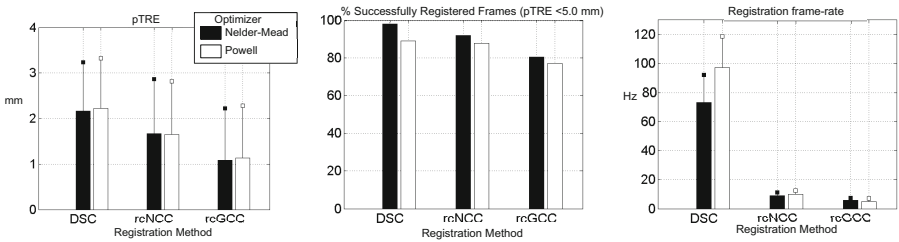
Fig. 6. Clinical datasets. Left: Average pTRE for successful registrations. Comparisons between methods were all statistically significant ($p < 0.001$). Middle: Percentage of successfully registered frames (pTRE<5.0 mm). Right: Average frame rate.

4   Discussion

The presented DSC method is an order of magnitude faster than the rcNCC and rcGCC methods, and is the only method that is able to reliably track the TEE probe without skipping frames. The frame rate increase comes at the cost of an increase in pTRE, which may be due to the lack of normalization by the DRR standard deviation used for the NCC and GCC metric computations. However, results indicate that the increase in error is still clinically acceptable. Results from the simulations and clinical datasets both show that the DSC method results in fewer failed registrations, indicating that the DSC metric may have a wider basin of convergence than the other similarity functions and therefore is more robust to poor initialization. In the future we will explore a hybrid registration approach combining the DSC and rcGCC methods, using DSC to obtain a fast, robust initial result and then refining the registration with rcGCC

for better accuracy. Validation in bi-plane XRF is another topic for future work. We believe that the DSC method has the potential to work well as a real-time 5DOF/6DOF device tracking algorithm. In particular, the algorithm may be well suited for tracking relatively small, high contrast objects such as a recently proposed fiducial embedded intracardiac echo catheter [8]. Future work will focus on integration with echo image processing for real-time multimodal visualization during cardiac interventions.

5   Conclusion

A method for fast 2D/3D registration of devices in XRF images is presented. Average registration frame rates were $33.0 \pm 9.4 \; fps$ and $73.1 \pm 19.2 \; fps$ for simulated and clinical datasets, respectively. Results also indicate that the proposed method converges to a clinically acceptable target registration error more often than prior methods. Future work will focus on clinical implementation and application of the algorithm to other devices.

References

1. Fagan, T.E., Truong, U.T., Jone, P.N., Bracken, J., Quaife, R., Hazeem, A.A.A., Salcedo, E.E., Fonseca, B.M.: Multimodality 3-dimensional image integration for congenital cardiac catheterization. Methodist DeBakey Cardiovascular Journal 10(2), 68 (2014)
2. Gao, G., Penney, G., Ma, Y., Gogin, N., Cathier, P., Arujuna, A., Morton, G., Caulfield, D., Gill, J., Aldo Rinaldi, C., et al.: Registration of 3D trans-esophageal echocardiography to X-ray fluoroscopy using image-based probe tracking. Medical Image Analysis 16(1), 38–49 (2012)
3. 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, Part II. LNCS, vol. 7511, pp. 25–32. Springer, Heidelberg (2012)
4. Lang, P., Seslija, P., Chu, M.W., Bainbridge, D., Guiraudon, G.M., Jones, D.L., Peters, T.M.: US–fluoroscopy Registration for Transcatheter Aortic Valve Implantation. IEEE Transactions on Biomedical Engineering 59(5), 1444–1453 (2012)
5. Kaiser, M., John, M., Borsdorf, A., Mountney, P., Ionasec, R., Nöttling, A., Kiefer, P., Seeburger, J., Neumuth, T.: Significant acceleration of 2D-3D registration-based fusion of ultrasound and x-ray images by mesh-based DRR rendering. In: Medical Imaging, International Society for Optics and Photonics, p. 867111 (2013)
6. Wein, W., Röper, B., Navab, N.: 2D/3D registration based on volume gradients. In: Medical Imaging, International Society for Optics and Photonics, pp. 144–150 (2005)
7. Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L., Hawkes, D.J.: A comparison of similarity measures for use in 2D-3D medical image registration. IEEE Transactions on Medical Imaging 17(4), 586–595 (1998)
8. Ralovich, K., John, M., Camus, E., Navab, N., Heimann, T.: 6DoF catheter detection, application to intracardiac echocardiography. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part II. LNCS, vol. 8674, pp. 635–642. Springer, Heidelberg (2014)