Learning the Update Operator for 2D/3D Image Registration

Srikrishna Jaganathan[1,2], Jian Wang[2], Anja Borsdorf[2], Andreas Maier[1]

[1]Pattern Recognition Lab, FAU Erlangen-Nürmberg, Erlangen, Germany.
[2]Siemens Healthineers AG, Forchheim, Germany.
srikrishna.jaganathan@fau.de

Abstract. Image guidance in minimally invasive interventions is usually provided using live 2D X-ray imaging. To enhance the information available during the intervention, the preoperative volume can be overlaid over the 2D images using 2D/3D image registration. Recently, deep learning-based 2D/3D registration methods have shown promising results by improving computational efficiency and robustness. However, there is still a gap in terms of registration accuracy compared to traditional optimization-based methods. We aim to address this gap by incorporating traditional methods in deep neural networks using known operator learning. As an initial step in this direction, we propose to learn the update step of an iterative 2D/3D registration framework based on the Point-to-Plane Correspondence model. We embed the Point-to-Plane Correspondence model as a known operator in our deep neural network and learn the update step for the iterative registration. We show an improvement of 1.8 times in terms of registration accuracy for the update step prediction compared to learning without the known operator.

1 Introduction

In minimally invasive interventions, live 2D X-ray imaging is prominent for providing image guidance. The information available from 2D imaging alone is limited and can be augmented by overlaying the preoperative 3D volume over the 2D images. To obtain this overlay, the 3D volume needs to be accurately positioned such that, the corresponding structures are aligned between the 2D image and 3D volume. The optimal positioning of the 3D volume is accomplished using 2D/3D registration. A 2D/3D registration aims to find an optimal 3D transformation such that the misalignment between the 2D image and the 3D volume is minimized. Traditionally, to find this optimal transformation, the 2D/3D registration problem is formulated as an optimization problem. Depending on the application, 2D/3D registration can be classified into different sub-classes like modality of the images, the number of 2D views available, and constraints on the estimated transformation. A complete overview of the different traditional 2D/3D registration techniques and its different sub-class are summarized in [1].

Recently, end-to-end Deep Learning (DL)-based solutions have also been proposed for both single-view [2,3] and multi-view 2D/3D registration [4] which shows significant improvement in terms of robustness and computational efficiency but often suffer in terms of registration accuracy.

In this work, we address single-view rigid 2D/3D registration between preoperative CT volume $\mathbf{V}$ and live fluoroscopic X-ray images I_{Flr}. Generally, in single-view 2D/3D registration, the 3D CT volume is rendered using Digitally Reconstructed Radiograph (DRR) to obtain a simulated X-ray image I_{DRR}. A similarity measure is defined to find the correspondences between the I_{DRR} and I_{Flr}. Using this similarity measure, classical techniques model the 2D/3D registration as an optimization problem to find the optimal transformation such that the registration error is minimized. However, in single-view 2D/3D registration, due to the aperture problem, the 2D misalignment between the structures (obtained by finding 2D correspondences) gives only the observable 2D motion error. To find the 3D misalignment (thus the optimal 3D transformation), the unobservable motions in 2D should also be accounted for. This can be effectively constrained using the Point-to-Plane Correspondence (PPC) model [5]. With the PPC model, a dynamic 2D/3D registration framework was proposed in [5], which performs iterative 2D/3D registration by solving the PPC model at each iteration. It achieves state-of-the-art performance in terms of both registration accuracy and robustness. However, the framework relies on having accurate 2D correspondences between I_{Flr} and I_{DRR}, which makes it sensitive to outliers in the estimated 2D correspondences. To make the framework robust against outliers, a DL-based attention model was proposed [6].

In this work, we extend the PPC-based registration framework to use a learned update operator. The update operator consists of 2D matching, weighing of the matches, and estimating the 3D rigid transformation such that it satisfies the PPC constraint. Since the PPC model is differentiable, it can be directly used as a known operator [7] and can be embedded as a layer in Deep Neural Network (DNN). Learning with the PPC model as a known operator has shown promising results when it was applied to learn correspondence weights for the 2D matches [6]. We propose to learn all the three steps of the PPC update operator fused into a single DNN, contrary to the previous attempts which only learned parts of the update operator [6,8].

2   Methods

2.1   PPC-based 2D/3D registration framework

The PPC-based registration framework proposed in [5] is an iterative registration scheme and depends on the PPC constraint to estimate the optimal transformation at each iteration. Registration is performed between 3D CT volume $\mathbf{V}$ and 2D X-ray image I_{Flr} which are provided as inputs to the framework. Along with it, an initial transformation $\mathbf{T}_{\text{init}}$ is required which provides a rough initial alignment.

The framework consists of an initialization step where the surface points along with their gradients are extracted from $\mathbf{V}$ using a 3D Canny edge detector, and the gradient of X-ray image (∇I_{Flr}) is computed from I_{Flr}. The initialization step is performed only once, and the values are cached. After the initialization step, the update operation is performed iteratively until convergence. The update operation consists of the following steps. Based on the current transformation estimate, the gradient image of DRR (∇I_{DRR}) is rendered. Surface points with gradients perpendicular to the current viewing direction are selected as contour points $\mathbf{w}_i$ where $i \in 1, \ldots, N$ with N contour points. The contour points $\mathbf{w}_i$ are projected into ∇I_{DRR} to get the projected contour points $\mathbf{p}_i$. Now, 2D matching is performed to find the corresponding projected contour points $\mathbf{p}'_i$ in ∇I_{Flr}. The correspondence set $(\mathbf{p}_i, \mathbf{p}'_i)$ along with contour points $\mathbf{w}_i$, and its gradients $\mathbf{g}_i$ are used to compute the 3D motion update based on the PPC model. If the 2D matches are noisy, a weighted version of the PPC model is used to reduce the effect of the noisy matches. The weighted PPC model gives a linear constraint $\mathbf{WAdv} = diag(\mathbf{W})\mathbf{b}$, where $\mathbf{A}$ and $\mathbf{b}$ are the data terms which are computed based on the estimated 2D correspondences. $\mathbf{W}$ is a diagonal matrix which consists of weights for each correspondences. Each correspondence contributes one row of $\mathbf{A}, \mathbf{b}$ and $\mathbf{W}$. Closed form solution $\mathbf{dv} = (\mathbf{WA}^+)(diag(\mathbf{W})\mathbf{b})$ using the pseudo inverse of $\mathbf{A}$ is used for computing the 3D motion. The rigid 3D transformation matrix $\mathbf{T}_{\mathrm{est}}$ is computed from the 3D motion vector $\mathbf{dv}$, which serves as an input for the next update step.

2.2   Learnable update operator for PPC-based 2D/3D registration framework
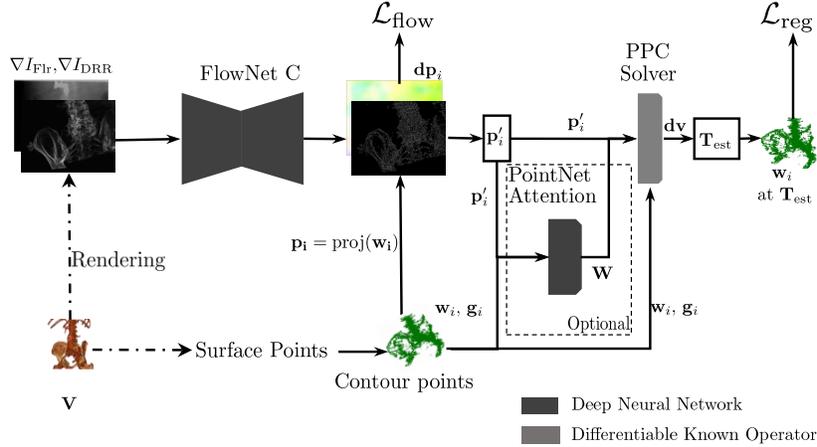
Fig. 1. Schematic of the proposed update operator for PPC based iterative registration framework.

The update operator in PPC-based registration framework finds $\mathbf{T}_{\mathrm{est}}$ given a set of inputs ($\nabla I_{\mathrm{Flr}}, \nabla I_{\mathrm{DRR}}, \mathbf{w}_i, \mathbf{g}_i, \mathbf{p}_i$). Fig. 1 shows the network architecture

used for training the update operator. The FlowNet C architecture [9] is used for estimating 2D matches between ∇I_{Flr} and ∇I_{DRR} by predicting optical flow $\mathbf{dp}_i$ at projected contour points $\mathbf{p}_i$ similar to [8]. Using $\mathbf{p}_i' = \mathbf{p}_i + \mathbf{dp}_i$ and the 3D information $(\mathbf{w}_i, \mathbf{g}_i)$, correspondence weighting matrix $\mathbf{W}$ is predicted using PointNet Attention model similar to [6]. This step is optional and we define two architectures PPC Flow and PPC Flow Attention, where we use the PointNet Attention only for the later. The 3D motion $\mathbf{dv}$ is computed using the PPC Solver which gives us the current estimated 3D transformation $\mathbf{T}_{\text{est}}$. Applying $\mathbf{T}_{\text{est}}$, the contour points $\mathbf{w}_i$ are updated . We use the mean Target Registration Error (mTRE) as registration loss $\mathcal{L}_{\text{reg}}$. It is computed using $\frac{1}{N}\sum_i^N ||\mathbf{T}_{\text{est}}(\mathbf{w}_i) - \mathbf{T}_{gt}(\mathbf{w}_i)||$ for a set of N contour points and ground truth transformation $\mathbf{T}_{\text{gt}}$. Additionally, we also use optical flow loss $\mathcal{L}_{\text{flow}}$ by computing average End Point Error (EPE) at projected contour points $\mathbf{p}_i$ as proposed in [8], to make the network training more stable.

2.3   Experimental Setup

The data set used for training and evaluation is from reconstruction data of Cone Beam CT (CBCT) for the vertebra region [6]. The data set consists of 56 acquisitions from 55 patients. The data is split into 38 patients for training, 5 patients for validation, and 12 patient for testing. The training samples are created using random initial transformation from the ground truth registration using different viewing directions from the volume similar to [8]. We generate random initial transformations with an initial mTRE in the range of [0, 45] mm. Each sample consists of $(\nabla I_{\text{Flr}}, \nabla I_{\text{DRR}}, \mathbf{w}_i, \mathbf{g}_i, \mathbf{p}_i)$ along with $\mathbf{T}_{gt}$. There are about 18,000 such samples for training and validation, and 8,000 samples for testing.

The update operator of the PPC-based registration framework performs one registration iteration. We train and evaluate the performance of the update operator using three different models namely Flow, PPC Flow, and PPC Flow Attention. The Flow and PPC Flow models are used to compare the effects of known operator learning. In the Flow model, only the 2D matching is learned by training the model with ground truth flow annotations at contour points $\mathbf{p}_i$ as proposed in [8]. Here, we use the PPC solver only for evaluating its registration performance after one update. PPC Flow uses the same optical flow architecture as the Flow model along with the PPC model embedded as a known operator. PPC Flow Attention integrates the PPC Flow with the attention model [6] to learn both correspondence estimation and weighing in an end-to-end manner. Both PPC Flow and PPC Flow Attention are trained using both registration loss (mTRE) and optical flow loss (average EPE).

We train all the networks for 200,000 iterations (100 epochs) with a batch size of 16. We use the ADAM optimizer with a learning rate of 1e-4 and weight decay is used as a regularizer with a decay rate of 1e-6.

For evaluation, we use mTRE after one registration update which measures the registration error. Lower values indicate better registration accuracy. In addition, we also use reduction factor which is computed for samples whose

$\text{mTRE}_{i+1} < \text{mTRE}_i$ as $(1 - \frac{\text{mTRE}_{i+1}}{\text{mTRE}_i})$ where mTRE_i is the error before update step and mTRE_{i+1} is the error after the application of the update step. It is set to 0.0 for samples whose error increases after the update. It takes a value between 0.0 to 1.0 where higher values are better. The reduction factor indicates how much of the error can be reduced with one update step.

3 Results

The performance of the different models evaluated on the test data set is summarized in Table 1. The Flow, PPC Flow, and PPC Flow Attention models achieve a mTRE ($\mu \pm \sigma$ computed over all samples) of 11.27 ± 11.66 mm, 6.21 ± 4.70 mm and 5.88 ± 4.50 mm respectively. The reduction factor for Flow, PPC Flow, and PPC Flow Attention is ($\mu \pm \sigma$ computed over all samples) 0.47 ± 0.38, 0.68 ± 0.17, 0.69 ± 0.17 respectively. Fig. 2 compares the change in registration error after one update for all the models used.

	Percentile mTRE [mm]			mTRE [mm]	Reduction Factor
	50th	75th	95th	$\mu \pm \sigma$	$\mu \pm \sigma$
Initial	20.18	30.09	39.75	20.59 ± 11.76	
Flow	7.45	15.20	34.59	11.27 ± 11.66	0.47 ± 0.38
PPC Flow	5.03	8.80	15.41	6.21 ± 4.70	0.68 ± 0.17
PPC Flow Attention	**4.70**	**8.31**	**14.68**	$\mathbf{5.88 \pm 4.50}$	$\mathbf{0.69 \pm 0.17}$

Table 1. Registration error measured using mTRE in [mm] of different networks after one update operation, evaluated on test data set. The initial mTRE indicates the initial registration error. The $50^{th}, 75^{th}, 95^{th}$ percentile errors are provided to indicate percentage of samples having an error $\leq$ the indicated value. The mTRE ($\mu \pm \sigma$) and the reduction factor ($\mu \pm \sigma$) summarizes the model performance over all the samples.
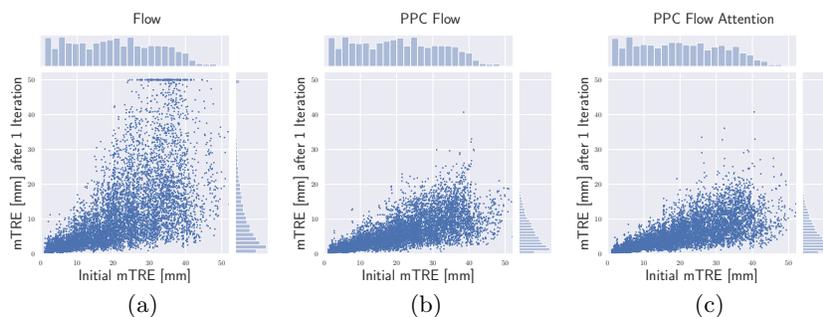
Fig. 2. Error distribution for the test data set, showing registration errors before and after one update operation for (a) Flow, (b) PPC Flow, and (c) PPC Flow Attention. For better visualization, the values of mTRE after 1 iteration are clipped at 50mm.

4 Discussion and Conclusion

We proposed a DL-based extension to learn the update operation of the PPC-based iterative 2D/3D registration framework. PPC Flow learns the 2D matching operation with PPC model as a known operator. It improves the performance of the update operation by a factor of 1.8 compared to Flow model in terms of mean reduction in registration error as shown in Table 1. Our intuition is that, the PPC model enforces constraints to learn correspondences that are relevant for the registration thus improving the performance. PPC Flow Attention learns 2D matching and weighting together in an end-to-end manner. This improves the performance over the PPC Flow model as it can effectively discard the outliers that might still be present in the correspondences predicted using PPC Flow.

The proposed update operator serves as an initial step in extending the PPC-based registration framework to a DL-based module which retains the strength of classical 2D/3D registration techniques while also providing all the benefits of DL-based methods. However, there are still some areas that need to be addressed for achieving this goal. Especially, in cases where we start with small registration errors, the learned update operator has minimal influence. We plan to address this issue in our future work either by unrolling the iterative registration or with multiple learned update operators. Additionally, one can explore how to make the update operator tailor-made for specific clinical applications.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

1. Markelj P, Tomaževič D, Likar B, et al. A review of 3D/2D registration methods for image-guided interventions. Med Image Anal. 2012;16(3):642–661.
2. Miao S, Wang ZJ, Zheng Y, et al. Real-time 2D/3D registration via CNN regression. In: Proc ISBI; 2016. p. 1430–1434.
3. Miao S, Piat S, Fischer P, et al. Dilated FCN for Multi-Agent 2D/3D Medical Image Registration. In: Proc AAAI; 2018. p. 4694–4701.
4. Liao H, Lin WA, Zhang J, et al. Multiview 2D/3D Rigid Registration via a Point-Of-Interest Network for Tracking and Triangulation. In: Proc CVPR; 2019. p. 12638–12647.
5. Wang J, Schaffert R, Borsdorf A, et al. Dynamic 2-D/3-D rigid registration framework using point-to-plane correspondence model. IEEE Trans Med Imaging. 2017;36(9):1939–1954.
6. Schaffert R, Wang J, Fischer P, et al. Learning an Attention Model for Robust 2-D/3-D Registration Using Point-To-Plane Correspondences. IEEE Trans Med Imaging. 2020;39(10):3159–3174.
7. Maier AK, Syben C, Stimpel B, et al. Learning with known operators reduces maximum error bounds. Nature machine intelligence. 2019;1(8):373–380.
8. Schaffert R, Weiß M, Wang J, et al. Learning-Based Correspondence Estimation for 2-D/3-D Registration. In: Proc BVM. Springer; 2020. p. 222–228.
9. Dosovitskiy A, Fischer P, Ilg E, et al. FlowNet: Learning Optical Flow With Convolutional Networks. In: Proc ICCV; 2015. p. 2758–2766.