Deformable US/CT Image Registration with a Convolutional Neural Network for Cardiac Arrhythmia Therapy

Batoul. Dahman, Jean-Louis Dillenseger

Abstract— **Image registration represents one of the fundamental techniques in medical imaging and image-guided interventions. In this paper, we present a Convolutional Neural Network (CNN) framework for deformable transesophageal US/CT image registration, for the cardiac arrhythmias, and guidance therapy purposes. The framework consists of a CNN, a spatial transformer, and a resampler. The CNN expects concatenated pairs of moving and fixed images as its input, and estimates as output the parameters for the spatial transformer, which generates the displacement vector field that allows the resampler to wrap the moving image into the fixed image. In our method, we train the model to maximize standard image matching objective functions that are based on the image intensities. The network can be applied to perform non-rigid registration of a pair of CT/US images directly in one pass, avoiding so the time consuming computation of the classical iterative method.**

I. Introduction

Cardiac arrhythmias are related to a dysfunction of the electrical conduction pathway in the cardiac tissue [1]. When drugs become inefficient, thermal Radio-Frequency ablation by catheterization is commonly applied. Current treatments lack of motion compensation and real-time imaging and result in incomplete lesions [2].

Ultrasound-guided HIFU is an alternative to the other ablation techniques [3]. It has the potential to avoid the complications associated with endocardial therapy. HIFU energy allows focusing of the ultrasound beam from the esophagus toward the heart wall rather than from inside the cardiac cavity. It can be used to create thermal ablation of propagation path [4], [5] without damaging the intervening tissues.

Transesophageal HIFU cardiac fibrillation therapy is a mini-invasive treatment that places the HIFU transducer close to the ablation zone by navigating inside the esophagus, the probe navigation and transducer positioning is carry out using an embedded ultrasound (US) imaging system [4]. As any mini-invasive procedure, first a therapy planning (the ablation path) is defined on high-resolution anatomical preoperative 3D imaging (CT/MRI), then the ablation is performed under the 2D US guidance.

Image registration represents one of the key technique in medical imaging and image-guided interventions [6]. It can bring the pre-operative 3D data and intra-operative 2D data into the same coordinate system, to facilitate accurate diagnosis and/or to provide advanced image guidance. The pre-operative 3D data generally includes Computed Tomography (CT), Cone-beam CT (CBCT), Magnetic Resonance Imaging (MRI) and Computer Aided Design (CAD) model of medical devices, while the intra-operative 2D data is dominantly X-ray fluoroscopy or US images.

In order to guide and ablate a specific zone of the heart chosen by the experts in the 3D CT volume, Sandoval *et al.* propose a therapy guidance system by the registration of the intraoperative 2D US images to the preoperative reformatted 3D CT volume [7]. In order to reduce the number of Degree of Freedom of this registration, they simplify the 2D/3D registration problem to a 2D-2D framework by making the anatomical assumption that the 2D US images are perpendicular to the esophagus axis. They extract all the 2D CT slices perpendicular to the esophagus axis, and performed 2D US to 2D CT image registration. The aim of this 2D/2D registration is to estimate the rigid transformation matrix which best aligns the information in 2D US image to the information in the 2D slice of the reformatted CT volume using the traditional iterative registration framework, which can be time consuming in real time therapy.

Beside the classical iterative methods, deep learning-based registration methods are now under study. Jaderberg et al. [8] introduce the spatial transformer network (STN) that can be used as a building block that aligns input images in a larger network that performs a particular task, by training the entire network end-to-end. The embedded STN deduces optimal alignment for solving that specific task. De Vos *et al.* [9] propose a deep learning network for deformable image registration (DIRnet). This framework was performed with registration of images for cardiac cine MR scans. Finally in [10] a learning framework for deformable image registration is performed to pairwise medical image registration, it allows to speed up medical image analysis, and processing pipelines.

In this work, we will focus on registering a 2D CT slice to transesophageal 2D US image with unsupervised learning approach. The network can be applied to perform registration on unseen image pairs in one pass, thus non-iteratively, to prevent time consuming issue. This approach should also be an elastic registration which is more suited to handle cardiac images.

II. Method

The proposal framework consists of three main parts, the Convolutional Neural Network (CNN), the spatial transform

* Batoul. Dahman, Jean-Louis Dillenseger "Univ Rennes, Inserm, LTSI - UMR 1099, F-35000 Rennes, France (Phone: +33 2 23 23 56 05; fax: +33 2 23 23 69 17; e-mail: batoul.dahman@univ-rennes1.fr, jean-louis.dillenseger@univ rennes1.fr).

978-1-7281-1990-8/20/$31.00 ©2020 European Union

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on November 17,2020 at 05:31:28 UTC from IEEE Xplore. Restrictions apply.

and a resampler (Figure 1). The concatenated inputs image pairs (the CT slice as moving image and the 2D US image as fixed one) is passed to the CNN. As an output, the CNN will give the parameters for the spatial transformer which generates the Displacement Vector Field (DVF) that allows the resampler to wrap the moving image into the fixed image.

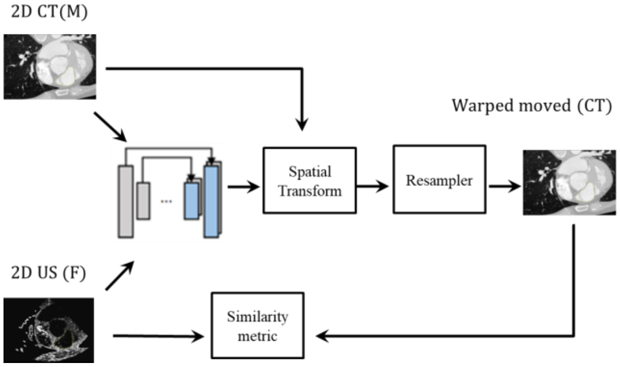

Figure 1. The general framework of the proposal approach.

A. CNN Model

The network architecture is similar to U-Net [11], [12], which consists of encoder and decoder sections with skip connections. The CNN expects concatenated pairs of moving and fixed images as its input, and applies two alternating layers of 2×2 convolutions in both the encoder and decoder stages using a kernel size of 3, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with a stride of 2 down-sampling layers in the encoder path to reduces the number of the CNN parameters.

The convolutional layers capture hierarchical features of the input image pair which are used to estimate the DVF in the decoding stage. Every step in the decoding path consists of an up-sampling, convolutions ("up-convolution") that halve the number of feature channels, and concatenating skip connections that propagate the features learned during the encoding stages directly to the layers generating the registration, that enabling image pair alignment.

The network is trained by optimizing an image similarity metric (i.e. by backpropagating dissimilarity) between pairs of moving and fixed images from a training set using mini-batch stochastic gradient descent [13]. After training, the network can be applied for registration of unseen images.

We implemented the network using Tensorflow and we trained it on a NVIDIA Corporation GPU with 1000 iterations which took approximately 6 hours.

B. Spatial transform

The spatial transformer generates the DFV that enables the resampler to wrap the moving image into the fixed image. The spatial transform is based on the spatial transformer network [8]. It compute for each pixel p, the new location in the warped moving image by adding the displacement vector (dx, dy) in that pixel.

Since the mapping from one space to the other will often require an evaluation of the intensity of the image at non-grid positions, an interpolator is required.

C. Loss function

We use mutual information (MI) as a loss function. In a previous case-based test we found that MI was one of the similarity measure the most suited to our data [14]. MI compares the information of the US images and the corresponding information extracted from the CT slices.

$$I(A,B) = \sum_{a,b} p(a,b)\, log \frac{p(a,b)}{p(a)p(b)} \; ; a \in A, b \in B \qquad (1)$$

Where A is the reference image (US), B represents the warped moving image (transformed CT slice), $p\,(A)$, $p\,(B)$ are the marginal distributions of A and B, and $p\,(A, B)$ their joint distribution.

III. Results

A. Dataset

This study has been conducted on a patient with cardiac fibrillation CT dataset, obtained from Louis Pradel University Hospital in Lyon, France. The dimensions of the reconstructed image are 512 × 512 × 323 voxels with an image spacing of 0.546875 × 0.546875 × 0.55031 mm^3.

Because we didn't have enough CT/US images couples, we chose to simulate the US images from the CT information as described in the next section.

B. Experimental protocol

In order to produce enough data for the training and testing of our framework and also to look about the robustness of the method, we arbitrary produced 250 poses in a range of ± 5 mm on translation and ± 5° in rotation around an initial pose, and we extract all the 2D CT slices using the framework described in [15]. Then we simulated the corresponding US images with the method described in [16].

From this dataset, the network was trained by randomly selecting 175 pairs of fixed and moving image slices from cardiac CT scans and simulated US images. The pairs of fixed and moving images were anatomically corresponding slices, but we added some deformations when we simulate the 2D US images, and the 75 remaining pairs were used for validation.

C. Experimental results

We compared the registration results obtained by the proposed methods to these obtained by a B-Spline free form deformation field non-rigid registration implemented in the SimpleElastix Library [17]. We used the same similarity measure MI in both methods.

1) Qualitative visual evaluation

Figure 2 shows an example of the registration of an image pair: a) The slice extracted from the CT volume; b) The simulated US image; c) The overlap between the moving CT image and the fixed US image using the classical method from SimpleElastix and d) CNN.

Visually, the results obtained by the proposed method seems to provide a better alignment than the classical free-form deformation field method, this can be seen for example at the bottom of the image on the thoracic chest.

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on November 17,2020 at 05:31:28 UTC from IEEE Xplore. Restrictions apply.

2) Quantitative evaluation

The registration comparison was carried out according two complementary metrics: the Dice similarity coefficient and the Hausdorff distance.

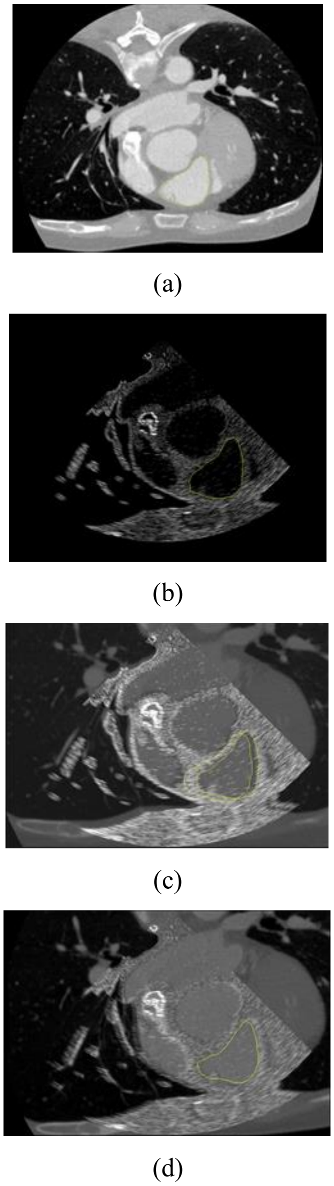

Figure 2 (a) 2D CT slices, (b) simulated 2D US slices with overlaid boundaries of the left atrium (yellow), (c) the resulted warped moving images for (c) the free-form deformation field method of SimpleElastix and (d) the proposal approach.

- Dice similarity coefficient (Dsc) :

This metric is an overlap index. It computes the number of pixels that overlap between two surfaces and normalizes it by the half of the sum of the number of non-zero pixels in the two surfaces [18]:

$$\alpha = \frac{2|A \cap B|}{|A| + |B|} \qquad (2)$$

Where A is the gold standard surface which, in our case, refers to the segmentation of the left atrium in the US fixed image (see Figure 2), and B the segmentation mapped from the warped registered CT image. $\alpha = 1$ indicates that the anatomy matches perfectly, and $\alpha=0$ indicates that there is no overlap.

- Hausdorff distance

The Hausdorff distance [19] is a Spatial distance based metrics. It is defined as the maximum of the closest distance between two objects where the closest distance is computed for each point or vertex of the two objects. A smaller Hausdorff distance indicates a closer topology between the 2 objects.

These two metrics were measured on the boundaries or surface of the segmented left atrium (see Figure 2).

Table 1 reports the mean and standard deviation of the scores measured on the validation dataset. We compared also the methods according to the computation time.

As expected, we can see in Table 1 that the computation time is highly improved using CNN (under a second) than using the classical iterative method (around one minute). More surprising, we can also notice in Table 1 that both spatial comparison metrics are improved when using CNN comparing to the traditional approach.

Table 1 Average Dice similarity coefficient, Hausdorff distance and computation time results for SimpleElastix and CNN (U-Net) *acros the segmented left atrium in the US fixed image, and the segmentation mapped from the warped registered CT (see Figure 2).*

method			
	Dice sim. Coef.	***Hausdorff distance (mm)***	***Comp. time (sec)***
SimpleElastix	0.7 (0.01)	1.7 (0.02)	65 (0.1)
CNN (U-Net)	0.8 (0.02)	1.2 (0.05)	0.7 (0.02)

These results were obtained from simulated US images. We are well aware that there are differences between simulated US images and real US images. However, we found in a previous study that a method tuned on simulated data obtained a good results on real data [7]. So we are hopeful that our method works on real data.

IV. Conclusion

In this study, we present an unsupervised learning-based-approach for transesophageal US/CT cardiac image registration. The results indicate a high improvement in terms of computation time without any loss (even with some improvements) in terms of registration accuracy.

In our future work, we will integrate the unsupervised learning approach to a minimally-invasive HIFU procedure to improve the planning and the guidance of the therapy. We are going to perform our approach on 2D/3D image-based registration to refine the estimation of the transesophageal probe pose in the 3D preoperative volume.

Finally we will include physical phantom and real-patients data to evaluate the contribution of our registration scheme for the therapy guidance.

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on November 17,2020 at 05:31:28 UTC from IEEE Xplore. Restrictions apply.

ACKNOWLEDGMENT

This work was part of the CHORUS (ANR- 17-CE19-0017) project which have been supported by the French National Research Agency (ANR).

REFERENCES

[1] B. C. Sinclair-Smith, “Electrical reversion of cardiac arrhythmias,” *South. Med. J.*, vol. 65, no. 3, pp. 289–293, 1972.

[2] J. Huang, S. & Miller, “Catheter Ablation of Cardiac Arrhythmias,” *3rd Ed. edn Elsevier*, 2014.

[3] S. Pichardo and K. Hynynen, “New design for an endoesophageal sector-based array for the treatment of atrial fibrillation: A parametric simulation study,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 56, no. 3, pp. 600–612, 2009.

[4] F. Bessiere *et al.*, “Ultrasound-Guided Transesophageal High-Intensity Focused Ultrasound Cardiac Ablation in a Beating Heart: A Pilot Feasibility Study in Pigs,” *Ultrasound Med. Biol.*, vol. 42, no. 8, pp. 1848–1861, 2016.

[5] E. Constanciel *et al.*, “Design and evaluation of a transesophageal HIFU probe for ultrasound-guided cardiac ablation: Simulation of a HIFU mini-maze procedure and preliminary ex vivo trials,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 60, no. 9, pp. 1868–1883, 2013.

[6] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, “A review of 3D/2D registration methods for image-guided interventions,” *Med. Image Anal.*, vol. 16, no. 3, pp. 642–661, 2012.

[7] Z. Sandoval, M. Castro, J. Alirezaie, F. Bessière, C. Lafon, and J.-L. Dillenseger, “Transesophageal 2D Ultrasound to 3D Computed Tomography registration for the guidance of a cardiac arrhythmia therapy,” *Phys. Med. Biol., vol. 63, no 15*, p. 155007, 2018.

[8] Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial Trans-former Networks,” *28th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, pp. 2017–2025, 2015.

[9] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, “End-to-end unsupervised deformable image registration with a convolutional neural network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10553 LNCS, pp. 204–212, 2017.

[10] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[12] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5967–5976, 2017.

[13] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.

[14] Z. L. Sandoval and J. L. Dillenseger, “Evaluation of computed tomography to ultrasound 2D image registration for atrial fibrillation treatment,” *Comput. Cardiol. (2010).*, vol. 40, pp. 245–248, 2013.

[15] B. Dahman, “High Intensity Focused Ultrasound Therapy Guidance System by Image-based Registration for Patients with Cardiac Fibrillation,” *CinC*, vol. 46, 2019.

[16] J. L. Dillenseger, S. Laguitton, and É. Delabrousse, “Fast simulation of ultrasound images from a CT volume,” *Comput. Biol. Med.*, vol. 39, no. 2, pp. 180–186, 2009.

[17] K. Marstal, F. Berendsen, M. Staring, and S. Klein, “SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 574–582, 2016.

[18] L. R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[19] D. P. Huttenlocher, W. J. Rucklidge, and G. A. Klanderman, “Comparing images using the Hausdorff distance under translation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1992-June, pp. 654–656, 1992.

Authorized licensed use limited to: SUN YAT-SEN UNIVERSITY. Downloaded on November 17,2020 at 05:31:28 UTC from IEEE Xplore. Restrictions apply.