Wei Wei, Marko Rak, Julian Alpers, and Christian Hansen

Towards fully automatic 2D US to 3D CT/MR Registration: A novel segmentation-based Strategy

Pre-print version

Wei Wei, Marko Rak, Julian Alpers, Christian Hansen

Faculty of Computer Science,

Otto-von-Guericke University Magdeburg, Germany

christian.hansen@ovgu.de

TOWARDS FULLY AUTOMATIC 2D US TO 3D CT/MR REGISTRATION: A NOVEL SEGMENTATION-BASED STRATEGY

Wei Wei[a], Marko Rak[a], Julian Alpers[a], Christian Hansen[a]

[a]Faculty of Computer Science & Research Campus STIMULATE, University of Magdeburg, Germany

ABSTRACT

2D-US to 3D-CT/MR registration is a crucial module during minimally invasive ultrasound-guided liver tumor ablations. Many modern registration methods still require manual or semi-automatic slice pose initialization due to insufficient robustness of automatic methods. The state-of-the-art regression networks do not work well for liver 2D US to 3D CT/MR registration because of the tremendous inter-patient variability of the liver anatomy. To address this unsolved problem, we propose a deep learning network pipeline which – instead of a regression – starts with a classification network to recognize the coarse ultrasound transducer pose followed by a segmentation network to detect the target plane of the US image in the CT/MR volume. The rigid registration result is derived using plane regression. In contrast to the state-of-the-art regression networks, we do not estimate registration parameters from multi-modal images directly, but rather focus on segmenting the target slice plane in the volume. The experiments reveal that this novel registration strategy can identify the initial slice phase in a 3D volume more reliably than the standard regression-based techniques. The proposed method was evaluated with 1035 US images from 52 patients. We achieved angle and distance errors of $12.7 \pm 6.2°$ and $4.9 \pm 3.1\,\mathrm{mm}$, clearly outperforming state-of-the-art regression strategy which results in $37.0 \pm 15.6°$ angle error and $19.0 \pm 11.6\,\mathrm{mm}$ distance error.

Index Terms— US, CT/MR, Registration

1. INTRODUCTION

Liver tumor ablation techniques have been widely used to handle liver metastases [1]. Ultrasound (US) is the most often used device because of its real-time capability to guide the surgeon during a liver intervention. However, due to the limited image quality, US images need to be aligned with pre-operative planning data, e.g., computed tomography (CT) images or magnetic resonance (MR) images, to provide better decision support during interventions. The most important module in this context is pre- and intraoperative image registration, which can align high-quality CT/MR images with intraoperative low quality US images.

To address this registration task, [2] suggest performing registration based on 3D US reconstruction. This method requires patient breath-holding, liver swiping with a tracked US transducer, and performing 3D reconstruction from 2D US slices. Despite the high registration accuracy, this method adds inconvenient steps to the standard intervention procedure. Furthermore, once the patient breath is recovered, the obtained transformation matrix becomes invalid. In [3], a rigid slice to volume registration is proposed as a graph labeling problem achieving promising results on beating heart magnetic resonance (MR) images. Alternatively, [4] proposed a one-step deformable iterative closest point method to align a set of 2D curves in the 2D transvaginal ultrasound (TVUS) image to a set of corresponding 3D surfaces in the 3D MR volume. Despite the high accuracy, the method needs manual organ segmentation for both TVUS and MR images and therefore can not be applied in interventional procedures.

Recently, deep learning has been applied for image registration. For instance, in [5, 6] convolutional neural networks (CNNs) are proposed to learn a similarity metric from multiple modality image patches to guide the registration process. Generative adversarial networks were used by [7, 8] to predict deformation fields in the context of non-rigid image registration. Deep reinforcement learning was explored by [9, 10] to solve the registration task with state-of-the-art performance. Other researchers [11, 12] show the possibility to solve registration via a regression network. First, they applied CNNs to extract features from input patches and then use a dense network for pose estimation. This idea was extended by [13, 14] to address the 2D-3D image registration problem. The regression method learns the relation between slice pose and 3D image according to the appearance of the 2D slice. However, this does not work well as liver anatomy varies tremendously in shape, volume, vessel structures, and others.

In this work, we propose a registration pipeline, which can determine a coarse slice pose via a classification network followed by a segmentation network to extract the target slice from the 3D volume. The latter was inspired by [15], who was among the first to formulate a classical regression problem as a segmentation task. In contrast to the state-of-the-art, our method intends to solve the crucial registration task by training a deep segmentation network to circumvent the robustness issues that arise with regression networks. To the best of our knowledge, this is the first deep learning method applying a

segmentation network to solve the initial plane estimation of 2D US slice to 3D CT/MR volume registration problem.

2. MATERIALS AND METHODS

Multi-modal slice to volume registration is a very challenging task. Not only the different image appearances between the US and CT/MR but also the little information available in the US image can make the registration even more difficult and ambiguous. In addition, registration must keep spacial structures, such as the relative pose of voxels. Hence, we choose CNN as the baseline for our registration pipeline.

2.1. Problem formulation

The registration process can be divided into two parts consisting of a static and a dynamic registration. In the static registration stage, we determine the coarse US slice pose in the 3D volume by extracting a point-direction representation for the slice. In the dynamic registration stage, breath motion, in-plane rotation, and translation can be optimized in a local Region of Interest (ROI) as described in many publications [16, 17]. Within this work, we focus on the static stage.

Regarding the static registration stage, a deep learning registration pipeline is proposed. As shown in Figure 1, the pipeline contains two parts: pose classification and plane estimation. The output is represented by a point $\vec{p}$ inside a plane and a normal vector $\vec{n}$ perpendicular to that plane.

2.2. Pose classification

In liver US examination, there are two standard scan positions: medial and transcostal (see Figure 2). The medial position is often used to examine the left lobe of the liver, hepatic vein, left branch of the portal vein, and vena cava. On the other side, the transcostal position can examine the right lobe of the liver and the first bifurcation of the portal vein.

To analyze the distribution of the US transducer orientation of these two classes, 132 and 122 US images were randomly selected from the transcostal position class and medial position class of training data, respectively. For each image, field experts performed slice to volume registration manually. The rotation parts of registration matrices are converted to Euler angles. Figure 3 shows the distribution of slice orientations in Euler angle representation. As shown, the distribution of transducer orientations divides clearly into two classes. This separation inspired us to solve the coarse transducer orientation estimation by classifying the US image into one of those two orientation classes.

Once the classification is completed, we can take the average orientation from each class as a coarse orientation to initialize our segmentation-based plane estimation network. The pose classification is implemented in Tensorflow using a ResNet18 architecture [18]. The loss is defined by softmax cross-entropy between predictions and the true pose class. As output, the average transducer pose of the predicted pose class is passed through the resample block (see Figure 1) to initialize the orientation of the CT/MR volume correctly.

2.3. Plane estimation

After rotation resampling of the CT/MR volume, the US slice should be as parallel as possible to the XY plane of the CT/MR volume. In multi-modal image registration, images acquired by different devices may vary a lot in content and appearance. Therefore, the registration shall focus on vessel structures that are available in pre-operative 3D CT/MR images and 2D US images, as well. Vessel structures of 3D CT/MR volume can be prepared via manual or semi-automatic segmentation tools. A U-Net-based pre-processing step [17] is applied to extract vessel structures from each US slice, which is then used for registration. Afterward, the pre-processed US slice is replicated along the Z direction, i.e., slice normal vector, to match the size of the resampled CT/MR volume (see block "Replicator" in Figure 1).

To address the actual plane detection, a 3D U-Net is implemented in Tensorflow. Our network takes the vessel feature maps of the 3D CT/MR, and the replicated US slices as input and outputs a segmentation of a plane corresponding to the actual position of the US slice w.r.t. the CT/MR volume. The network is trained via a modified DICE loss between the predicted segmentation and the ground truth segmentation on the US slice position in the CT/MR volume. The DICE coefficient (see Equation 1) is based on the loss function introduced by [19], where A and B depict ground truth and prediction, respectively. In addition, we add a new false positive term on the denominator, which aims at reducing false-positive predictions. Coefficients $\alpha = 0.8$ and $\beta = 0.2$ are employed to balance the original DICE and the false-positive term.

$$DICE = \frac{2 * \sum |A * B|}{\alpha * (\sum A^2 + \sum B^2) + \beta * \sum (1 - A) * B} \quad (1)$$

Once the prediction is finished, the slice plane can be extracted robustly by the least trimmed squares plane regression method (see the last step "Plane Regression" in Figure 1) [20].

3. EVALUATION

Image data were acquired from 52 patients, approved by the ethics committee at Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University. Twenty-eight patients were examined with a US transducer on transcostal position, while the other 24 patients were examined with US transducer on medial position. In total, 1035 images were collected with visible vessel structures. Ground truth was annotated manually by field experts. For each US image, ground truth consists of the US pose class label and a transformation from a slice image to the 3D volume. The splitting of image data
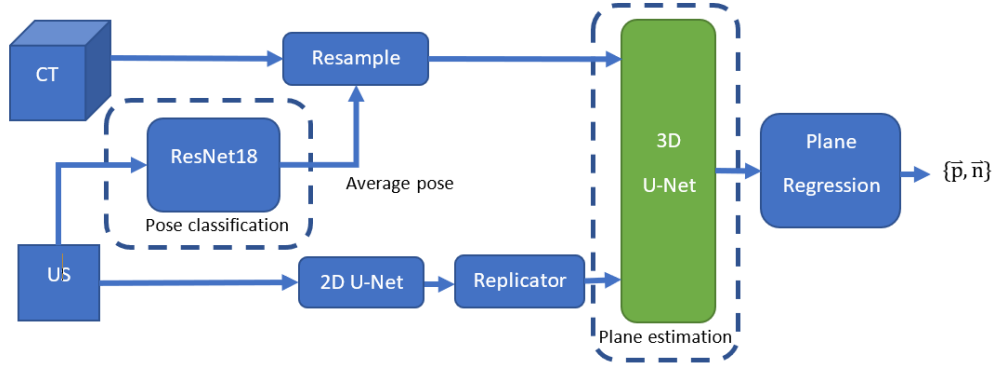
Fig. 1. Static registration: a pipeline to perform automatic slice plane estimation in 3D volume data.
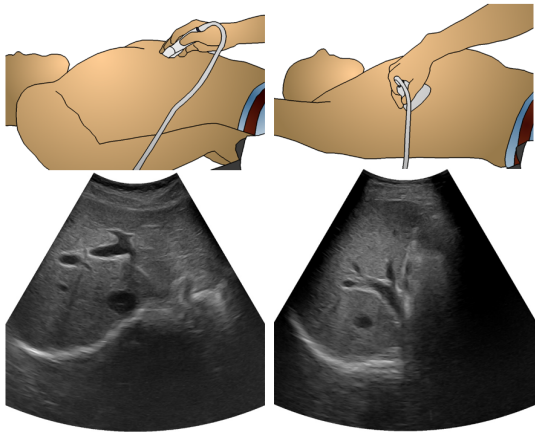

Fig. 2. Examplary US slices captured in medial (left column) and transcostal position (right column).

Fig. 3. Euler angle distribution of transcostal and medial position in the volume coordinate system. Left: projection on the pitch-yaw plane. Right: projection on the yaw-roll plane.

3.1. Results

Experimental results in Figure 4 depicts a promising classification with areas under the receiver operation characteristic (ROC) curve being equal to 0.99. The class label with the maximal probability is assigned to the US image. As a result, the experiments show a classification accuracy of 96.3 %.

is 10:2:1 for training, validation, and testing, respectively. 812 training and 164 validation images were selected to train the ResNet18 and 3D U-Net. Training and testing of the ResNet18 network were performed with US images downscaled to 128×128. The ResNet18 network was trained with a learning rate of 0.1 and a batch size of 256. During the evaluation of 3D U-Net, we downscaled all 3D images to 40×40×40 voxels with a spacing of 4 mm. Training parameters were set to a batch size of 40, batch normalization, learn rate of 0.0003, and filter size of 3×3×3. Adam optimizer was applied to minimize the loss function. In the training phase, we applied resliced 2D vessel images out of CT/MR volume data instead of using US vessel images directly. This generates more training slices and prevents overfitting during training. Data augmentation was performed to prepare those slice images by varying slice pose using a random rotation of [-20, 20]° and random translation of [-20, 20] mm. The training was done with 700 epochs, while the testing was done with the remaining 85 images. Both training and testing were performed on an NVIDIA Geforce GTX 1080Ti.
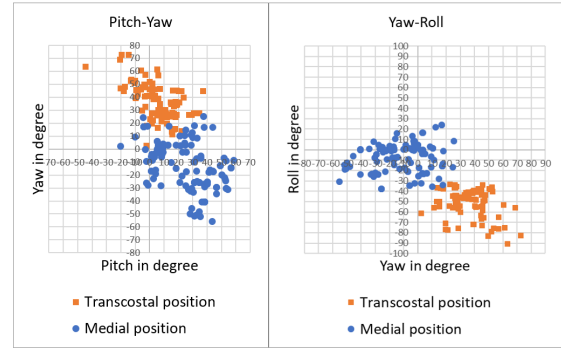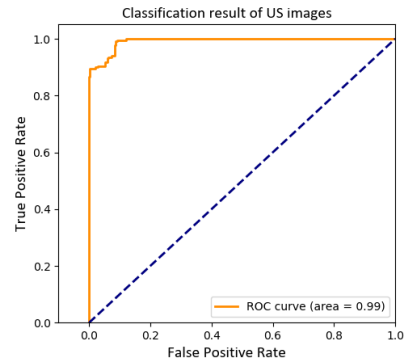

Fig. 4. ROC-curve for the classification of US images.

The 3D U-Net achieved an average DICE score of 0.61 between ground truth and prediction volume. To evaluate the

performance of plane estimation, we calculate the accuracy of the plane estimation by considering the angle between the estimated and ground-truth plane (out-plane rotation error). In addition, we calculate the average distance between vessel voxels in the ground truth and prediction plane (distance error). The plane estimation results in Figure 5 show a significant error reduction comparing angle errors directly after pose classification and after plane estimation. Specifically, the angle error is reduced from $20.8 \pm 8.5°$ to $12.7 \pm 6.2°$. As shown in the left column of Figure 5, the angle error of the estimated plane is positively proportional to the initial pose give by the pose classification network. The dashed line in Figure 5 divides the results into two groups: successful registration and failed registration. The points under the dashed line show successful registrations where angle errors are reduced, and estimated plane poses are improved. While the points above the dash line show failed registrations, yielding worse angle errors. Those failed registrations were caused by little visibility of vessel structures available in the US images. The experiments show a successful registration rate of 82.8%.
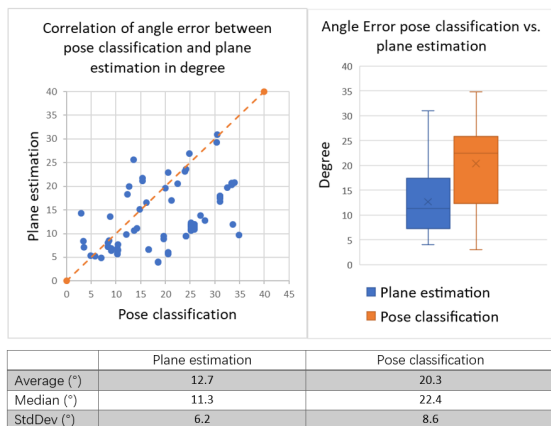

	Plane estimation	Pose classification
Average (°)	12.7	20.3
Median (°)	11.3	22.4
StdDev (°)	6.2	8.6

Fig. 5. Comparison of angle error after pose classification and after plane estimation.

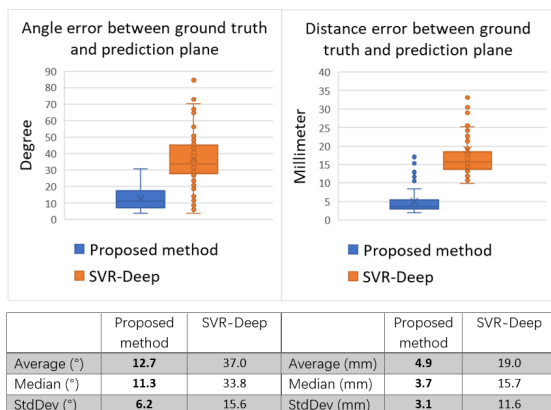|	Proposed method	SVR-Deep		Proposed method	SVR-Deep
Average (°)	**12.7**	37.0	Average (mm)	**4.9**	19.0
Median (°)	**11.3**	33.8	Median (mm)	**3.7**	15.7
StdDev (°)	**6.2**	15.6	StdDev (mm)	**3.1**	11.6

Fig. 6. Comparison of the proposed method and the SVR-Deep method. Left: Angle error; Right: Distance error.

For comparison, we implemented the state-of-the-art deep regression SVR-Deep [14] and evaluated with the same image data. As shown in Figure 6, the proposed method achieves an orientation accuracy of $12.7 \pm 6.2°$, outperforming the SVR-Deep method resulting in $37.0 \pm 15.6°$. Also, the proposed method achieves a distance error of 4.9 ± 3.1 mm to the ground truth plane, significantly better than 19.0 ± 11.6 mm of the SVR-Deep method. The best and the worst visual results of the registration are shown in Figure 7. The slice to volume transformation is calculated using the proposed method to determine out-plane parameters followed by manual corrections on in-plane parameters.
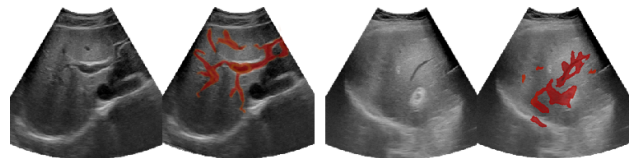

Fig. 7. Visualization of registration results with overlay of resliced vessel tree on the US image. Left image pair: The best case. Right image pair: The worst case.

4. CONCLUSION

In this work, we propose a novel deep learning-based registration pipeline, which estimates the initial pose of a US slice before the fine registration and motion compensation. The evaluation results show an orientation and distance errors of $12.7 \pm 6.2°$, and 4.9 ± 3.1 mm, respectively. This result is promising as an initial pose for the fine registration methods. Moreover, the experiments show that our method outperforms state-of-the-art regression-based deep learning methods.

The performance of our method depends on the visibility of the vessels. All the US images involved in this evaluation are selected automatically, such that least 1 % of a typical liver area is covered. This restriction is reasonable, not only because vessel structures can help to increase the registration accuracy, but more importantly, according to the clinical feedback, only complex cases where tumors are surrounded by vessels need guidance in the intervention procedure.

The prediction results of our pipeline can be used further as initial poses in the dynamic registration step, where the result is adapted to in-plane motion from breathing. In addition, further experiments shall be made on more challenge data and even with other organs.

Acknowledgement

5. REFERENCES

[1] Wells S. A. et al., "Liver ablation: best practice," *Radiol Clinics*, vol. 53, no. 5, pp. 933–971, 2015.

[2] Fusaglia M. et al., "A novel ultrasound-based registration for image-guided laparoscopic liver ablation," *Surg Innov*, vol. 23, no. 4, pp. 397–406, 2016.

[3] Porchetto R. et al., "Rigid slice-to-volume medical image registration through markov random fields," in *Med Comp Vis Bayes Graph Models Biomed Imag*, pp. 172–185. Springer, 2016.

[4] Yavariabdi A. et al., "Mapping and characterizing endometrial implants by registering 2d transvaginal ultrasound to 3d pelvic magnetic resonance images," *Comput Med Imag Graph*, vol. 45, pp. 11–25, 2015.

[5] Simonovsky M. et al., "A deep metric for multimodal registration," in *Int Conf Med Imag Comp Comp-Assis Interv*. Springer, 2016, pp. 10–18.

[6] Haskins G. et al., "Learning deep similarity metric for 3d mr-trus registration," *arXiv preprint arXiv:1806.04548*, 2018.

[7] Fan J. et al., "Adversarial similarity network for evaluating image alignment in deep learning based registration," in *Int Conf Med Imag Comp Comp-Assis Interv*. Springer, 2018, pp. 739–746.

[8] Yan P. et al., "Adversarial image registration with application for mr and trus image fusion," in *Int Workshop Mach Learn Med Imag*. Springer, 2018, pp. 197–204.

[9] Krebs J. et al., "Robust non-rigid registration through agent-based action learning," in *Int Conf Med Imag Comp Comp-Assis Interv*. Springer, 2017, pp. 344–352.

[10] Ma K. et al., "Multimodal image registration with deep context reinforcement learning," in *Int Conf Med Imag Comp Comp-Assis Interv*. Springer, 2017, pp. 240–248.

[11] Sloan J. M., Goatman K. A., and Siebert J. P., "Learning rigid image registration-utilizing convolutional neural networks for medical image registration," 2018.

[12] Hou B. et al., "3-d reconstruction in canonical coordinate space from arbitrarily oriented 2-d images," *IEEE Trans Med Imag*, vol. 37, no. 8, pp. 1737–1750, 2018.

[13] Hou B. et al., "Predicting slice-to-volume transformation in presence of arbitrary subject motion," in *Int Conf Med Imag Comp Comp-Assis Interv*. Springer, 2017, pp. 296–304.

[14] Salehi S. S. M. et al., "Real-time deep registration with geodesic loss.," *arXiv preprint arXiv:1803.05982*, 2018.

[15] Ernst P. et al., "A CNN-based framework for statistical assessment of spinal shape and curvature in whole-body MRI images of large populations," in *Inf Conf Med Imag Comp Comp-Assis Interv*. 2019, pp. 3–11, Springer.

[16] Weon C. et al., "Position tracking of moving liver lesion based on real-time registration between 2d ultrasound and 3d preoperative images," *Med Phys*, vol. 42, no. 1, pp. 335–347, 2015.

[17] Wei W. et al., "Fast registration for liver motion compensation in ultrasound-guided navigation," in *IEEE Int Symposium Biomed Imag*, 2019, pp. 1132–1136.

[18] He K. et al., "Deep residual learning for image recognition," in *IEEE Conf Comp Vis Pattern Recog*, 2016, pp. 770–778.

[19] Milletari F. et al, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Int Conf 3D Vis*. IEEE, 2016, pp. 565–571.

[20] Rousseeuw P. J. and Van Driessen K., "Computing lts regression for large data sets," *Data Mining Knowl Disc*, vol. 12, no. 1, pp. 29–45, 2006.