Distance Map Supervised Landmark Localization for MR-TRUS Registration

Xinrui Song¹, Xuanang Xu¹, Sheng Xu², Baris Turkbey³, Thomas Sanford⁴, Bradford J. Wood², and Pingkun Yan¹

¹Department of Biomedical Engineering and Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY, USA
²Center for Interventional Oncology, Radiology & Imaging Sciences, National Institutes of Health, Bethesda, MD, USA
³Molecular Imaging Program, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA
⁴Department of Urology, The State University of New York Upstate Medical University, Syracuse, NY, USA

1. DESCRIPTION OF PURPOSE

Image-guided interventional procedures often require registering multi-modal images to visualize and analyze complementary information. For example, prostate cancer biopsy benefits from fusing transrectal ultrasound (TRUS) imaging with magnetic resonance (MR) imaging to optimize targeted biopsy. However, cross-modal image registration is a challenging task. This is especially true when the appearance of two image modalities are vastly different.

Researchers have sought other methods to bridge the modality difference. Haskins et al.¹ designed a deep similarity metric to describe the difference between MR and US images. Some works directly use a DNN to regress the spatial relationship between images.² Aside from image similarity, Balakrishnan et al.³ also explored using segmentation DICE as a loss function to assess the registration quality. Hu et al.⁴ used soft probabilistic DICE of multiple landmark segmentations to provide smoother guidance for MR-TRUS registration. Such method, however, requires extensive labeling of not only organs, but small cysts and lesions as well. This level of requirement largely limits the feasibility of such method.

Since registration quality is most reliably evaluated with target registration error (TRE), it is sensible to directly make use of the anatomical landmark targets from images. Moreover, while image modalities and thus textures differ, anatomical landmarks are the only information shared across the moving and the fixed images. Song et al.⁵ used contrastive loss-guided pre-training methods to maximize the similarity between similar anatomical structures. Sun et al.⁶ proposed to perform pre-alignment for MR-TRUS registration with manually labeled landmarks on both images. However, such a procedure is still far from automatic due to the requirement of manual input at inference time. Heinrich et al.⁷ made use of a landmark detection method specifically designed for lung computed tomography (CT) registration, which is not generalizable to other tasks.

In this work, we propose to explicitly use the landmarks of prostate to guide the MR-TRUS image registration. We first train a deep neural network to automatically localize a set of meaningful landmarks, and then directly generate the affine registration matrix from the location of these landmarks. For landmark localization, instead of directly training a network to predict the landmark coordinates, we propose to regress a full-resolution distance map of the landmark, which is demonstrated effective in avoiding statistical bias to unsatisfactory performance and thus improving performance. We then use the predicted landmarks to generate the affine transformation matrix, which outperforms the clinicians’ manual rigid registration by a significant margin in terms of TRE.

Send correspondence to Pingkun Yan (yanp2@rpi.edu).
2. METHODS

Figure 1 gives an overview of the proposed method. We first extract the location of selected landmarks from both TRUS and MR images. Then, the two sets of corresponding landmarks are used to calculate an affine registration matrix that aligns the two images. The acquisition of landmarks is explained in Section 2.1, and the step that generates the affine registration from corresponding points is explained in Section 2.2.

2.1 Distance map-guided landmark localization

Anatomically important and stable landmarks are key to registration, as it is to our method. We referred to previous literature and consulted clinicians to finally decide on using four landmarks. Two of the landmarks are the extreme points located on the prostate boundary: the right-most and left-most extreme points observed from the axial view. The other two (shown in Figure 2) are the entrance point of the urethra into the prostate at the neck of the bladder, and the point where the urethra exits the prostate at the prostate apex. As the two extreme points of the prostate boundary can be directly extracted from prostate segmentation, which is available for pre-operative MR and can be predicted with established methods for TRUS, we will focus on the localization of the two other anatomical landmarks: the neck of the bladder, and the exit point of the urethra.

In the proposed method, a network similar to UNet is used to predict a distance map, which is then post-processed to recover the coordinate of the desired landmark. This is different from conventional landmark localization methods, which towards the end of the network, convert the feature map into a feature vector, and then directly regress the landmark coordinates with fully connected layers. Our experiments show that such an approach is prone to statistical biases when applied to 3D volumes, and could collapse to always outputting the mean value of the training samples. To overcome this issue, we designed a distance map, paired with UNet, to guide the training of the network. The skip connections and encoder-decoder structures of UNet enforce the model to make more explicit use of the image contents. More importantly, the prediction of the network is a...
full-resolution map that corresponds to the input image, resembling the human labeling process of drawing onto the image, as opposed to the abstract procedure of directly regressing the coordinates from the fully connected layers. We also added a sigmoid layer at the end of the UNet to normalize the output to a range of $[0, 1]$.

$$\text{map}_{\text{label}} = \exp(-10 \times \frac{M}{\max(M)}), M = \text{distance map}$$

(1)

The distance map is first calculated with Maurer’s algorithm. We then normalize the distance map into the label map according to Equation 1, so that the actual location of the landmark will have a value of one on the label map, and the surrounding voxels gradually fade to zero in all directions. We supervise the network training with MSE loss calculated between the label map and the predicted map.

### 2.2 Affine registration

The Kabsch–Umeyama algorithm is capable of finding the transformation parameters, including translation, rotation, and uniform scaling that give the least mean square error from two sets of corresponding points. In the proposed method, we first use the predicted landmarks to determine the optimal transformation parameters through the Umeyama algorithm, then apply this transformation to the ground truth landmark points to calculate the TRE for evaluation.

$$\arg\min_{\theta} \| P_{\text{TRUS}} - T_{\theta}(P_{\text{MR}}) \|_{L2}$$

(2)

Although the Kabsch–Umeyama algorithm is an established method, it is still limited to uniform scaling, which is far from real-world deformations. In this paper, we propose to use the results of the Umeyama algorithm as a starting point for an optimization-based method that allows nonuniform scaling. We first decompose the matrix produced by the Umeyama algorithm into nine transformation parameters, then used the Adam optimizer to perform gradient-descent optimization on the transformation parameters. We use the euclidean distance between the two sets of points as the loss function. This method enables fully-affine registration and leads to better results as shown in Section 3. The objective function for optimization is shown in Equation 2. This proposed extension to the Kabsch–Umeyama algorithm is described in Figure 3. Note that although we are only showing results for using the method to accommodate non-uniform scaling, it can be extended to shearing as well.

Figure 3. The proposed extension to the Umeyama algorithm.
Table 1. Registration performance of different methods. Measurements are TRE values in mm.

| landmarks     | TRUS segmentation | Umeyama | Umeyama + Adam |
|---------------|-------------------|---------|----------------|
| manual        | manual            | 4.02±0.82 | 3.71±0.77     |
| predicted     | manual            | 4.41±0.81 | 4.07±0.67     |
| predicted     | predicted         | 4.93±1.18 | 4.68±1.39     |
| Before registration |             | 7.53±2.11 |
| Manual rigid registration |       | 6.31±1.28 |

Table 2. Landmark localization performance comparison. Measurements are TRE values in mm.

| Methods                      | Neck of bladder | Urethra exit point |
|------------------------------|-----------------|-------------------|
| ResNet + coordinate label\(^{8,9}\) | 6.21            | 6.24              |
| UNet + distance map label    | 3.12            | 3.377             |

3. RESULTS

3.1 Implementation details

We divide our dataset to a training set of 21 cases, validation set of 3 cases, and test set of 5 cases. All images are resampled to 96×96×96 with isotropic spacing for network input. For Adam gradient descent optimization, we consistently used 1 × 10^4 iterations with a step size of 1 × 10^{-5}.

3.2 Experiment results

In Table 1, we present the result of the proposed registration method in comparison to no registration and manual registration by clinicians. The first row of the table shows that if the two sets of landmarks are known, the proposed method of using gradient descent optimization to finetune the result of the Umeyama algorithm achieves better results. Such improvement results from allowing non-uniform scaling, which isn’t accessible through the Umeyama algorithm.

The third row of Table 1 shows the result of using predicted TRUS segmentation to extract the extreme points, and as an auxiliary input to the landmark localization network. The second row shows a scenario in which the ground truth TRUS segmentation is available, and only the landmarks are predicted through neural networks. Although the results on the third row are generally larger in error than the second row, it is still significantly lower in comparison to the TRE before registration, or the result of manual registration (p value < 0.05 through paired T-test). Additionally, the TRUS segmentation network is trained on a small dataset. Given the better accessibility of TRUS prostate segmentation in comparison to landmark segmentations, it is perfectly feasible to improve the result of the third row with a TRUS segmentation network trained on a larger dataset. The overall performance of the pipeline can also be improved with B-spline alignments of landmarks on top of affine transformations. We will explore these ideas in the future.

In Table 2, we compare the performance of the proposed distance map-guided landmark localization method with the conventional approach of directly predicting the landmark coordinate. Our experiments show that for 3D medical images, such as MR and TRUS as in our dataset, the proposed method outperforms its conventional counterpart.

In Table 3, we compare the TRE calculated with landmarks not included in the process of affine registration prediction. In this case, we are using the center points of cysts as the evaluation landmark. This experiment demonstrates the generalizability of the proposed registration pipeline. Even if a landmark is left outside the process of registration prediction, the registration based on other landmarks ultimately leads to a universally better landmark alignment. The proposed method in this table refers to the same method from the third row of Table 1, with both the landmarks and the TRUS segmentation predicted through neural networks. The registered TRE outperforms manual registration and is on par with that calculated from manually labeled landmarks. This evaluation result is most relevant to clinical accuracy, because abnormal tissues, such as cysts, could be a desirable biopsy sample.
Table 3. Registration error of landmarks not used for registration. Measurements are TRE values in mm.

| Methods                        | No registration | Manual rigid registration | Proposed       | Manual landmarks + Adam |
|-------------------------------|-----------------|---------------------------|----------------|-------------------------|
| TRE of cyst                   | 4.92±1.12       | 4.40±0.59                 | **2.42 ± 1.19** | **2.34 ± 1.01**         |

Figure 4. Automatic registration result vs manual registration without the aid of landmark points. Yellow contour marks the boundary of ultrasound segmentation and cyst landmark; green contour marks its MR counterpart.

In Figure 4, we qualitatively compare the registration result of the proposed method and manual registration that ignores landmarks for guidance. In each figure, the ultrasound image is overlayed with colormap onto MR images. The larger contour shows the prostate segmentation, and the smaller contour shows a sphere drawn around the center of the cyst. The proposed method achieves much better alignment for both the cyst landmark and the overall prostate contour.

4. CONCLUSION

In this paper, we present a novel pipeline to guide multi-modal registration, which utilizes a novel landmark localization method that outperforms conventional coordinate-regression methods. The proposed registration pipeline itself also shows great potential to be improved even further in future studies.

5. ACKNOWLEDGEMENT

This work was partially supported by National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH), USA under awards R21EB028001 and R01EB027898, and through an NIH Bench-to-Bedside award made possible by the National Cancer Institute.
REFERENCES

[1] Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P. A., Wood, B. J., and Yan, P., “Learning deep similarity metric for 3d mr–trus image registration,” *International journal of computer assisted radiology and surgery* **14**, 417–425 (2019).

[2] Guo, H., Kruger, M., Xu, S., Wood, B. J., and Yan, P., “Deep adaptive registration of multi-modal prostate images,” *Computerized Medical Imaging and Graphics* **84**, 101769 (2020).

[3] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V., “Voxelmorph: a learning framework for deformable medical image registration,” *IEEE transactions on medical imaging* **38**(8), 1788–1800 (2019).

[4] Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C. M., Emberton, M., et al., “Weakly-supervised convolutional neural networks for multimodal image registration,” *Medical image analysis* **49**, 1–13 (2018).

[5] Song, X., Chao, H., Xu, X., Guo, H., Xu, S., Turkbey, B., Wood, B. J., Sanford, T., Wang, G., and Yan, P., “Cross-modal attention for multi-modal image registration,” *Medical Image Analysis* **82**, 102612 (2022).

[6] Sun, Y., Yuan, J., Qiu, W., Rajchl, M., Romagnoli, C., and Fenster, A., “Three-dimensional nonrigid mr–trus registration using dual optimization,” *IEEE transactions on medical imaging* **34**(5), 1085–1095 (2014).

[7] Heinrich, M. P. and Hansen, L., “Voxelmorph++ going beyond the cranial vault with keypoint supervision and multi-channel instance optimisation,” *arXiv preprint arXiv:2203.00046* (2022).

[8] Wu, Y., Hassner, T., Kim, K., Medioni, G., and Natarajan, P., “Facial landmark detection with tweaked convolutional neural networks,” *IEEE transactions on pattern analysis and machine intelligence* **40**(12), 3067–3074 (2017).

[9] Alansary, A., Oktay, O., Li, Y., Le Folgoc, L., Hou, B., Vaillant, G., Kamnitsas, K., Vlontzos, A., Glocker, B., Kainz, B., et al., “Evaluating reinforcement learning agents for anatomical landmark detection,” *Medical image analysis* **53**, 156–164 (2019).

[10] Maurer, C. R., Qi, R., and Raghavan, V., “A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(2), 265–270 (2003).

[11] Umeyama, S., “Least-squares estimation of transformation parameters between two point patterns,” *IEEE Transactions on Pattern Analysis & Machine Intelligence* **13**(4), 376–380 (1991).

[12] Kingma, D. P. and Ba, J., “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980* (2014).