

Personalized Femoral Implant Planning by Automatic Femur Segmentation

Iwan Paolucci, Severin Tobler
The Tree Nurses

Abstract—In this work we propose a fast and robust automatic femur segmentation from MRI data, in order to manufacture personal femoral implants. The automatic segmentation evolves around a Random Forest classifier, which has been designed from a data set that comprises 20 MRI volumes. Applying a 5-fold cross validation on this data set with the proposed segmentation yields a mean DICE of 0.91 with a standard deviation of 0.03. Additional methods to enhance the accuracy are discussed.

I. INTRODUCTION

PERSONAL femoral implants feature less attrition, therefore more comfortable wearing and lower implant related issues after implantation. Furthermore the operation can be carried out more swiftly, if the implant is pre-operational adjusted to the patient. The personalisation of a femoral implant requires an accurate segmentation of the femur, which is typically done with CT images. With an upcoming awareness of radiation, we investigate in an algorithm to segment the distal femur on MRI images. To be applicable in clinical routine the method should be automatic and robust. Additionally we set our goal for the accuracy with a DICE of 0.95 and for the robustness with a standard deviation of 0.05.

In this work we propose a method for fast and robust automatic femur segmentation from MRI data. The automatic segmentation is based on a Random Forest (RF) classifier [1] and does not require any manual interaction. In the following sections the segmentation pipeline, the training and testing of the segmentation and the performance of the segmentation are described and discussed.

II. MATERIALS & METHODS

The automatic segmentation evolves around a Random Forest classifier. In Figure 1 the whole segmentation pipeline is shown. The individual parts of the segmentation are covered in the following subsections. Furthermore the validation of the automatic segmentation is described.

To train the RF and evaluate the performance of the segmentation two data sets are used. The first data set comprises 20 MRI volumes and their ground truth. This data set is used to train the RF and evaluate the algorithm quantitatively. The second data set contains 10 MRI volumes without ground truth, which are used to evaluate the segmentation qualitatively.

A. Pre processing

In order to compensate for different MRI grey-scale representations, the images are normalized using z-score normalisation (Equation 1). In a second step a Wiener filter is applied

to reduce noise. This yields a normalized representation of the data, which is used for the following steps.

$$I_n = \frac{I - \mu}{\sigma} \quad (1)$$

B. Feature Extraction & Random Forest

Thirteen different features are extracted for each voxel of the normalized data, which are used to apply the RF. The features cover different aspects of the data. To describe vertical and horizontal edges we use Prewitt and Sobel filters. The derivatives are covered by Laplacian and Laplacian of Gaussian and the intensity by Gaussian and Average filters. The statistical information is represented by the entropy and the standard deviation features. To restrict the classifier to the femur we added the relative position in 3D, with the strong assumption that the femur rotates and moves only slightly between the different MRI volumes. This assumption can be made because the orientation of the patient in medical images is known according to the DICOM standard[2]. To compensate for different image resolutions the relative position is used instead of the absolute.

The RF tree model is created by training a RF with 15 trees on 2% of randomly picked voxels from the training images. Half of the samples are drawn from positive and the other half from negative samples. Applying the RF tree model on the extracted features predicts the voxels belonging to the femur.

C. Post processing

After the prediction of the Random Forest classifier, the results are post processed to remove outliers. The implemented post processing comprises following four steps, where the 2D operations are applied per slice and the 3D operation on the volume:

- 1) 2D Morphological opening
- 2) 2D Keep largest area
- 3) 2D Morphological filling
- 4) 3D Keep largest volume

Applying this four steps yields the final segmented volume. In order to present the result, the volume is converted to a mesh by applying a marching cubes algorithm [3].

D. Validation

To validate the whole segmentation, a randomized 5-fold cross validation [4] is used on the first data set. Out of the 20 volumes 16 are used to train and 4 to test the algorithm. To measure the accuracy and robustness the mean DICE coefficient[5] and the standard deviation over all 20 results is used.

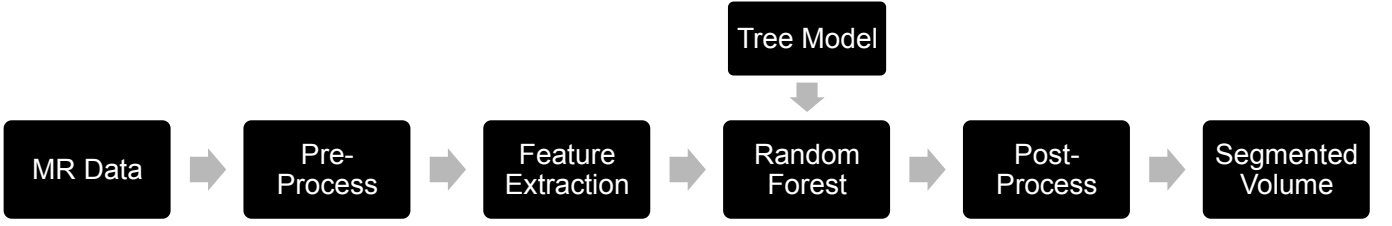


Fig. 1: Pipeline of the automatic segmentation using a Random Forest model to segment the femur from MRI data, followed by postprocessing and visualization using marching cubes.

III. RESULTS

A 5-fold cross validation yields a mean DICE of 0.91 with a standard deviation of 0.03. The range lies between 0.85 and 0.96 and has a median of 0.9. In Figure 2 the DICE after RF is compared to the DICE after the post processing. One can see, that the post processing increases the results by 0.05 on average. The average computation time for the whole proposed algorithm is 2.5 minutes for one MRI volume. The goal of 0.95 DICE is represented as the striped line in Figure 2.

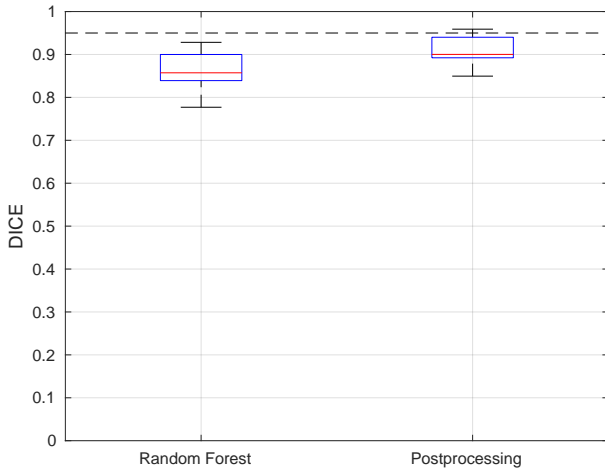


Fig. 2: DICE yielded by the 5-fold cross validation on the first data set. Once after Random Forest and once after post processing, the final segmentation.

In Figure 3 the optically best and worst result out of the second data set are shown. These are based on subjective observations and not on quantitative measurements, as the ground truth to this data is not available.

IV. DISCUSSION

The results in Figure 2 with a standard deviation of 0.03 let us conclude, that we achieved our goal of a robust algorithm, which we set with 0.05 standard deviation. We validated our algorithm with a 5 fold cross validation to ensure the algorithm is tested on MRI volumes, that have not been used for training. Therefore we can expect it to perform with the same accuracy on new data and not to have overfitting of our Random Forest model. With a mean accuracy of 0.91 DICE the goal of 0.95 has not been reached.

In order to use the algorithm for femoral implant planning, mainly the condyles have to be segmented well. With the 3D

volume rendering from the best case result in Figure 3a it is possible to measure distances between important structures to plan a femoral implant. However, the epicondyles are not fully segmented, which is something that has to be improved in the future. In the case of Figure 3b one could do the required measurements, although there are a lot of false positives. However, the accuracy is probably not high enough for clinical routine.

In Figure 3b the worst case result is shown. This particular case could have yielded better results with a more aggressive post processing approach. By applying the morphological opening with a bigger kernel the false positives could have been reduced. Because a bigger kernel would be able to split the connection from the bone to the artefacts. The separated artefacts are then discarded by keeping only the largest volume in the next post processing step. Nonetheless, this approach would also remove true positives and over all lower the segmentation performance in other images.

As mentioned in the beginning, our main goal is to develop a robust algorithm. This has been achieved by keeping the algorithm simple, which prevents overfitting and long segmentation time. Therefore all steps that worsened the results in some cases have been removed. As an example we did not include skewness as a feature, as it worsened the results slightly in more than half of the cases.

V. CONCLUSION

We developed a fast and robust automatic segmentation algorithm. The loading of one MRI volume until the final segmentation requires in average 2.5 minutes on a commercial laptop. With a DICE of 0.91 and a standard deviation of 0.03 the segmentation has a sufficient robustness, but lacks in accuracy, which will be part of the future development.

In order to enhance the accuracy, we propose to include prior information of the patient. One could create several Random Forest models for a specific group of people, where the groups differ in age, sex, weight or height. Prior information could also be considered by supporting the segmentation with a Statistic Shape Model [6]. However, the performance of the Statistic Shape Model depends strongly on its initialization and may therefore decrease the robustness of the segmentation.

An other potential way to increase the accuracy is to exchange the 2D features of the Random Forest by 3D features. This might represent the data better, since an additional dimension is considered. The challenge is that the slice spacing is multiple times larger than the pixel spacing. The feature

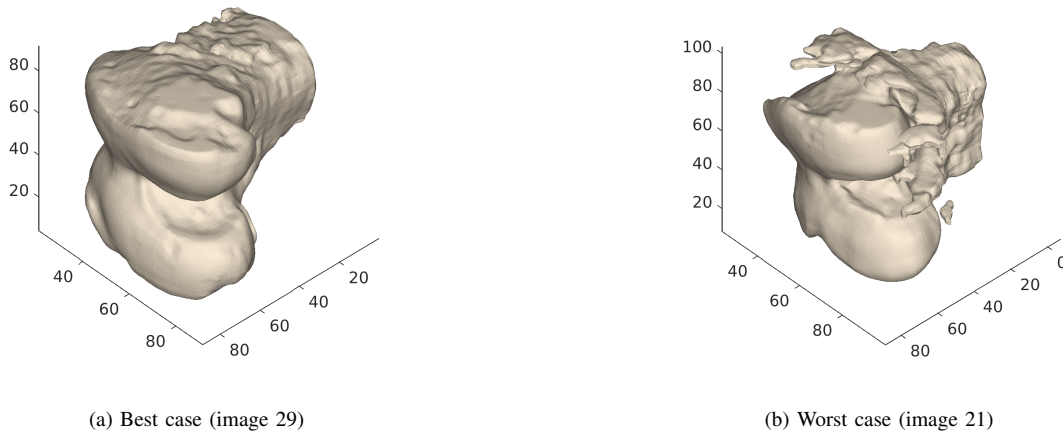


Fig. 3: Objectively the best and the worst segmentation of the second data set.

kernels must therefore be adjusted and this might lead to a loss of useful properties such as kernel separability.

REFERENCES

- [1] L. Breiman, "Random forests," English, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, ISSN: 0885-6125. DOI: 10.1023/A:1010933404324. [Online]. Available: <http://dx.doi.org/10.1023/A:1010933404324>.
- [2] M. I. . T. A. NEMA. (2015). Dicom - 7.6.2.1.1 image position and orientation, [Online]. Available: http://dicom.nema.org/medical/Dicom/2015a/output/chtml/part03/sect_C.7.6.2.html.
- [3] W. E. Lorensen and H. E. Cline, "Marching cubes: a high resolution 3d surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, Aug. 1987, ISSN: 0097-8930. DOI: 10.1145/37402.37422.
- [4] B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 316–331, 1983. DOI: 10.1080/01621459.1983.10477973. eprint: <http://amstat.tandfonline.com/doi/pdf/10.1080/01621459.1983.10477973>. [Online]. Available: <http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1983.10477973>.
- [5] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation," 2011.
- [6] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3d medical image segmentation: a review," *Medical image analysis*, vol. 13, no. 4, pp. 543–563, 2009.