Automatic third molar localization from 3D MRI using random regression forests

Walter Unterpirker[1]
walter.unterpirker@student.tugraz.at
Thomas Ebner[1]
ebner@icg.tugraz.at
Darko Stern[1]
stern@icg.tugraz.at
Martin Urschler[2]
martin.urschler@cfi.lbg.ac.at

[1] Institute for Computer Graphics and Vision, BioTechMed, Graz University of Technology, Austria
[2] Ludwig Boltzmann Institute for Clinical Forensic Imaging, BioTechMed, Graz, Austria

Abstract

Radiological age estimation of living subjects from MR images has recently become very popular. Besides skeletal ossification this can be done using the mineralization status of wisdom teeth. To support potential automatic age estimation, an important preliminary step is a reliable and automatic localization of the wisdom teeth. Therefore, we propose a random regression forest framework to localize third molars, which is capable to predict landmarks up to an error of 3.55 ± 2.62 mm in mean and standard deviation in a challenging 3D MRI dataset.

1 Introduction

Radiological age estimation is currently seeing a lot of research interest not only for clinical, but increasingly for forensic applications, most prominently majority age assessment of young asylum seekers coming to Europe without valid identification documents. According to AGFAD, the study group of forensic age diagnostics, combined radiological assessment of epiphyseal plates of the hand bones and the clavicle as well as the mineralization status of the wisdom teeth, i.e. third molars, are key components of an objective, accurate age estimation [7]. While established radiological age estimation techniques rely on X-ray and CT images, recently MRI data has shown to be a promising alternative without the need for ionizing radiation. Automatic age estimation from MRI is a worthwhile goal to pursue [8], since it removes the need for subjective visual comparison to reference images, as present in the established radiological techniques. The automatic localization of third molars is an important preliminary step when designing an automatic dental age estimation method from radiological data. Therefore, in this work we present a novel automated third molar localization algorithm, taking 3D head MRIs as input, which may subsequently be used for automatic dental age estimation. We propose to employ a random forest strategy [1], and formulate localization as a regression task similar to recent work on bone localization [3, 5]. We compare different types of voxel selection methods for training the random regression forest and compare two distinct voting strategies during testing. We show on a data set of 280 3D MRIs the performance of our localization algorithm in cross-validation experiments.

1.1 Related work

State of the art automatic object and landmark localization methods rely on the use of discriminative or generative machine learning techniques, e.g. statistical models of shape (SSM) and/or appearance [6]. In [4] an SSM is used to segment the maxillary bone. By placing a predefined region below this segmentation and finding a suitable separation to split this region into multiple parts, all teeth are thus located. However, although such an SSM can handle large outliers very well, it strongly relies on a good initialization of shape and pose, which is a complex problem of its own. On the other hand, discriminative random forest (RF) models [1] have recently seen a lot of interest due to their simple adoption to diverse applications and their ability to handle large and noisy datasets very well. In [2] a classification RF was designed to locate teeth, based on a spatial assumption that all landmarks are clustered in a certain region to avoid searching the whole image. Further, a small region around the ground-truth landmarks was labeled as positive training instances and regions further away as negative ones to classify teeth. Due to multiple positive labels per landmark defined for training the classifier, this results in imprecise localizations during testing. The drawbacks of SSM and classification RF led to research where localization is formulated as a regression problem [3]. A regression RF (RRF) was trained to predict bounding boxes around anatomical structures, i.e. organs. This idea was extended in [5] by adding a weighting scheme and a multi-forest approach to very accurately localize single point landmarks between hand bones. In this work, we investigate suitability of the ideas presented in [3] to build a novel fully automated 3D MRI wisdom teeth localization algorithm based on RRF.

2 Method

For localizing third molars automatically, we train an RRF only on a small subset of voxels in each image of our training dataset of annotated MRI volumes, restricting the trained model to local appearance information around the wisdom teeth. This is contrary to [3], since they use global information from all over the image. Reasons for restricting training to local appearance information are the lower anatomical shape variability near the teeth and larger variations in intensity in more remote structures like the brain. In testing, the model predicts most likely landmark candidate positions in previously unseen images using a voting scheme.

During the **training process** of the forest, we first select voxels from regions near the mean landmark position of the given landmarks. We assume that the most stable structures in our images are around the teeth region, as illustrated in Fig. 1. For training individual trees, we push random subsamples of the selected voxel regions through each tree, which consists of split and leaf nodes, starting at its root split node. At each split node the incoming voxels are forwarded either to the left or right child node, depending on a splitting criterion, until a maximal tree depth is reached. In this case a leaf node is created. See Fig. 1 for a coarse overview of training. The decision rules that determine the splitting criterion for each split node are chosen from a pool of randomly generated feature tests. Their selection is done according to the maximization of an information gain measure, representing the change in variance of the distribution of the input voxels when splitting them into two separate sets. Each decision rule in the pool consists of feature computations, involving

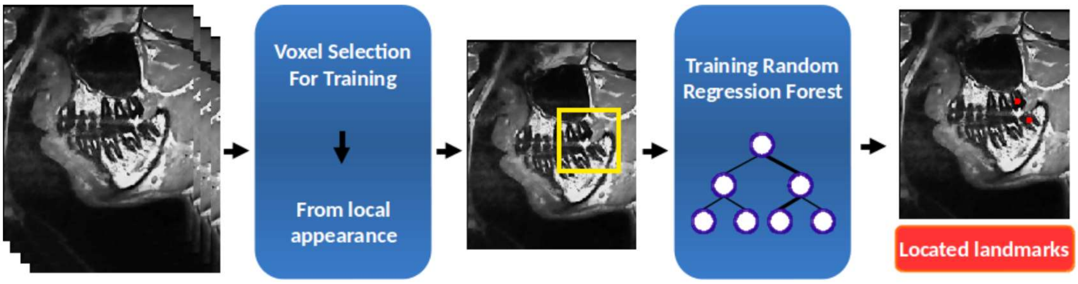

Voxel Selection For Training
From local appearance
Training Random Regression Forest
Located landmarks

Figure 1: Overview of training stage and landmark localization during testing.

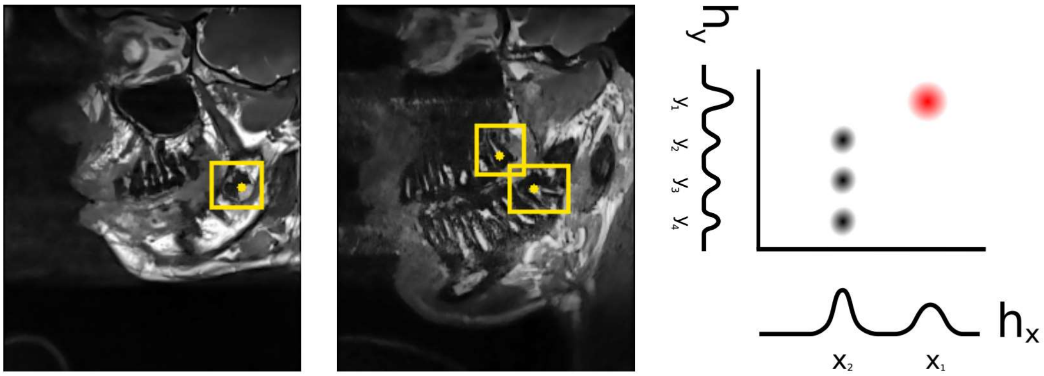
differences of mean intensity values from two cuboids, generated randomly in an arbitrary distance to a voxel position and having random sizes, and a random threshold to which the computed feature is compared to. This greedy optimization selects for each split node the best combination of feature and threshold from the pool, thus defining and storing the splitting criterion of the node. Recursively performing this operation, trees are trained by splitting until the maximum tree depth is reached, or the voxel sets reaching a node are too small for further splitting. In this case a leaf node is created, and the relative distances of the voxel to each landmark are computed. Finally, in each leaf node l_t of a tree t, the relative distances from voxels v reaching this node to all landmarks are stored using for each landmark i three separate 1D distance histograms $h_{\{x,y,z\},i}(l_t(v))$ that enable voting for landmark positions.

During **testing** an unseen image, the RRF randomly selects voxels from the whole image and pushes them through each tree. Going down a tree, voxels reach split nodes in which they are either pushed to the left or right subtree according to the stored feature/threshold combination. Eventually, voxels end up in a leaf node from which the stored relative distance histograms vote for a potential landmark position. For final landmark prediction, different methods can be applied to combine votes from individual trees to a single prediction. In our work we investigate two such methods, i.e. voting based on histogram accumulators and voting by using an image space accumulator, which can be seen as a point voting scheme.

For the **histogram accumulator** voting scheme, we first create a final histogram by accumulating over all histograms $h_{\{x,y,z\},i}(l_t(v))$, over all voxels and trees, in each axis independently, similar to [5]. This results in a final histogram $H_{\{x,y,z\},i}$. The maxima in each coordinate per landmark i indicates the most probable landmark position. In contrast, the **image space accumulator**, which is also used in [6], is built by first finding the maxima in each histogram $h_{\{x,y,z\},i}(l_t(v))$ directly, which represents a potential landmark position for one voxel, ending up in one leaf node in a single tree. Then, this position is accumulated in the image space accumulator for this landmark, which is a 3D volume. After having pushed all voxels through the forest, we get as many accumulators as we have landmarks. Finally, probable landmark positions can be estimated by finding the maxima in these accumulators.

3 Experiments and results

Dataset: Our dataset consists of 280 3D MR images (PD weighted TSE sequence) with a dimension of 208 x 256 x 30 voxels and a size of 0.59 x 0.59 x 1 mm per voxel. Per image, two landmarks, either on the wisdom teeth or on the assumed location of a missing tooth, were manually annotated by a dentist. During MRI acquisition, noise and artifacts occurred, which were caused by movement or metallic parts like braces, making our dataset challenging. Figure 2(a-b) shows a selection of images with annotated wisdom teeth.



(a) (b) (c)

Figure 2: Example of MR image slices for different age groups (a) 13 and (b) 17. Difference between histogram and image space accumulators is illustrated in (c).

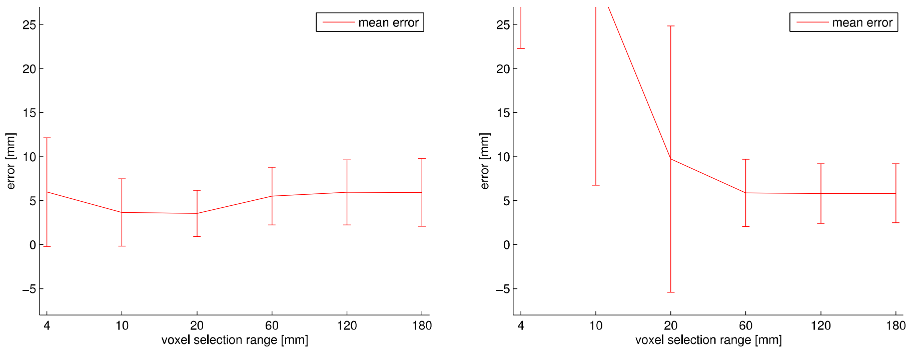
Cross validation: We randomly split our dataset into 67 % training and 33 % testing images for 5 cross validation runs and average performance measures over these runs. As a performance measure we use the Euclidean distance between a found landmark and the annotation. We train an RRF with 8 trees and a maximum tree depth of 14, since it turned out that using more or deeper trees does not improve our results. At each split node of a tree, 80 features and 20 random thresholds per each feature test are generated. The training algorithm generates 3D cuboid image features with the center at a maximum distance of ±16 *x* 16 *x* 6 mm and maximum size of ±16 *x* 16 *x* 4 mm relative to a voxel position. We have chosen this size to potentially cover a whole tooth at once.

Experimental setup: We made two experiments. With our first experiment we show the influence between using image voxels from the whole image and therefore of strongly varying structures, e.g. in the brain, and more locally selected voxels near the teeth region, which is more constrained for different subjects. Therefore, we shrink the range from where voxels can be chosen for training, with certain step sizes, starting from a global range down to a local one, around the mean position of our teeth landmarks. However, when using a small range, i.e. 4 mm, the forest is still able to cover appearance from farther away, since we use long distances features which are at most ± 16 mm away (note that neighboring teeth are located at around 10 mm distance from each other). In our second experiment, we investigate the use of different histogram accumulators and compare them to the point voting scheme.

3.1 Range for voxel selection

For the first experiment, we use the image space accumulator and compare different ranges around the mean landmark position, from where voxels can be selected for the training process. The range is defined by a sub-volume with side-length *r* for each axis. In Fig. 3a we can see, that when decreasing *r* from 180 mm to 20 mm, the mean landmark localization error and its standard deviation get smaller and reach a minimum of 3.55 ± 2.62 mm. Further decreasing the range to 4 mm, leads to imprecise localization and therefore to a larger error. The main reason is, that the forest has too few voxels from which to learn, and that global localization thus becomes very hard.

When we look at the error at a range *r* of 180 mm, we can see that using voxels from



(a) (b)

Figure 3: Localization results for different voxel selection ranges r using (a) image space accumulator and (b) histogram accumulator.

the whole image is unfavourable. The algorithm seems to fail for images, in which artifacts change the appearance of the shape. Also by the occurence of strongly varying structures, like the brain, the forest may vote with a large uncertainty for the landmarks. Another challenge occurs due to strong translations leading to occlusions, which only shows in a few images. The algorithm is not able to precisely localize landmarks, in case the shape (e.g. the nose, mouth or chin), is partly occluded.

Overall, we can conclude, that using voxels from regions with lower variability for the training process, helps to improve the localization results, as expected. On the other hand, we obtain a large standard deviation error, which is due to mislocalization of teeth.

3.2 Histogram accumulator and image space accumulator

In the second experiment, we make a comparison between **image space accumulator** and **histogram accumulator** voting schemes. Figures 3a and 3b depict the difference in results between these two accumulators. The histogram accumulator yields to much larger errors for ranges r, smaller than 60 mm. The best result we can achieve, using the histogram accumulator, is at a range r of 180 mm with an error of 5.82 ± 3.35 mm, compared to the best error of 3.55 ± 2.62 mm using the image space accumulator.

This happens because we treat each coordinate in the histograms independently. For example, assume the 2D case, in which two points vote to the position (x_1, y_1) as illustrated as a red circle in Fig. 2c. Now, three different points vote to the same x, but different y coordinates $(x_2, y_{2,3,4})$. The histogram accumulator sums up over all axes independently and develops the highest peak at position (x_2, y_1), although more points are voting to the position (x_1, y_1). This happens especially, when the forest localizes multiple possible landmark candidates at different positions in the image when choosing voxels only from small regions. However, using larger regions, i.e. from the global shape, the forest is still able to predict the landmarks quite well with these histograms. By using the image space accumulator we circumvent this drawback, since we directly vote into an image, as shown in Fig. 2c.

4 Conclusion

We have shown a fully automatic third molar localization framework, which is based on random regression forests. By investigating the role of areas from where voxels are selected for training and using a point voting scheme, we achieved localization results with a mean localization error of 3.55 ± 2.62 mm, which is well below the distance of individual teeth in our datasets. Future work will concentrate on employing subsequent steps for a more precise localization, e.g. a multi-forest approach [5], comparing different voting accumulation types, e.g. as done in [6], as well as to classify whether a teeth is present or not and to determine its orientation. Finally, enabled by our low localization error, located wisdom teeth will be used in an age estimation system, combined with other body parts, i.e. skeletal bones from hand and clavicle, to achieve more robust age estimation results.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF): P 28078-N33.

References

[1] L. Breiman. Random forests. *Machine Learning*, pages 5–32, 2001.

[2] E. Cheng, J. Chen, J. Yang, H. Deng, Y. Wu, V. Megalooikonomou, B. Gable, and H. Ling. Automatic dent-landmark detection in 3D CBCT dental volumes. In *Engineering in Medicine and Biology Society, EMBC, 2011, Annual International Conference of the IEEE*, pages 6204–6207, 2011.

[3] A. Criminisi, D. Robertson, E. Konukoglu, J. Shotton, S. Pathak, S. White, and K. Siddiqui. Regression forests for efficient anatomy detection and localization in CT scans. *Medical Image Analysis*, 17(8):1293 – 1303, 2013.

[4] N. T. Duy, H. Lamecker, D. Kainmueller, and S. Zachow. Automatic detection and classification of teeth in CT data. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2012)*, volume 7510 of *Lecture Notes in Computer Science*, pages 609–616. Springer Berlin Heidelberg, 2012.

[5] T. Ebner, D. Stern, R. Donner, H. Bischof, and M. Urschler. Towards automatic bone age estimation from MRI: Localization of 3D anatomical landmarks. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014)*, volume 8674 of *Lecture Notes in Computer Science*, pages 421–428. Springer International Publishing, 2014.

[6] C. Lindner, P. Bromiley, M. Ionita, and T. Cootes. Robust and accurate shape model matching using random forest regression-voting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2014.

[7] A. Schmeling, A. Olze, W. Reisinger, and G. Geserick. Forensic age diagnostics of living people undergoing criminal proceedings. *Forensic Science International*, 144(2-3):243–245, 2004.

[8] D. Stern, T. Ebner, H. Bischof, S. Grassegger, T. Ehammer, and M. Urschler. Fully automatic bone age estimation from left hand MR images. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2014)*, volume 8674 of *Lecture Notes in Computer Science*, pages 220–227. Springer International Publishing, 2014.