Skull Segmentation in MRI by a Support Vector Machine Combining Local and Global Features

Jens Sjölund, Andreas Eriksson Jarlideni, Mats Andersson, Hans Knutsson and Håkan Nordström

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:

Copyright: IEEE http://ieeexplore.ieee.org/Xplore/home.jsp

Skull Segmentation in MRI by a Support Vector Machine Combining Local and Global Features

Jens Sjölund*, Andreas Eriksson Järliden†, Mats Andersson*, Hans Knutsson† and Håkan Nordström‡

* Department of biomedical engineering, Linköping university, Sweden
† Center for Medical Image Science and Visualization (CMIV), Linköping University, Sweden
‡ Elektta Instrument AB, Kungstensgatan 18, Box 7593, SE-103 93 Stockholm, Sweden

Email: {jens.sjolund, mats.x.andersson, hans.knutsson}@liu.se
Email: {andreas.eriksson, hakan.nordstrom}@elekta.com

Abstract—Magnetic resonance (MR) images lack information about radiation transport—a fact which is problematic in applications such as radiotherapy planning and attenuation correction in combined PET/MR imaging. To remedy this, a crude but common approach is to approximate all tissue properties as equivalent to those of water. We improve upon this using an algorithm that automatically identifies bone tissue in MR. More specifically, we focus on segmenting the skull prior to stereotactic neurosurgery, where it is common that only MR images are available.

In the proposed approach, a machine learning algorithm known as a support vector machine is trained on patients for which both a CT and an MR scan are available. As input, a combination of local and global information is used. The latter is needed to distinguish between bone and air as this is not possible based only on the local image intensity.

A whole skull segmentation is achievable in minutes. In a comparison with two other methods, one based on mathematical morphology and the other on deformable registration, the proposed method was found to yield consistently better segmentations.

I. INTRODUCTION

Magnetic Resonance Imaging (MRI) has two major benefits over Computed Tomography (CT): greater soft tissue contrast and absence of ionizing radiation. However, the contrast mechanism in CT—unlike MRI—is related to tissue properties such as electron density that is used to calculate absorbed dose. Consequently, calculations on radiation transport in MR images require approximations. In this work, we improve upon the common approximation that all tissue is water by also identifying bone. This is done automatically with a machine learning algorithm trained on patients for which both a CT and an MR scan were available. More specifically, we focus on segmenting the skull prior to stereotactic neurosurgery where, often, only MR images are available.

It has been shown that a clear reduction of the dose calculation error can be achieved by segmenting a few tissue types and assigning different tissue properties to them [1]. However, the clinical applicability is limited as long as a time-consuming manual segmentation is required—an automatic method is evidently needed.

There are several uses for such a method, including

• Dose calculations in radiotherapy treatment. Since the electron density is directly related to the absorbed dose it is of key importance for accurate treatment planning.

• PET/MRI attenuation correction. Density information is required to compensate for attenuation inside the body when imaging with PET. In combined PET/MRI machines it would be desirable to use the MR images to this end.

• EEG and MEG source modeling. To enable accurate modeling of the electric sources when examining the brain it is important to distinguish the skull from soft tissue due to the large differences in conductivity between the materials.

• Scatter correction in Cone-Beam CT. When using Cone-Beam CT for verification of treatment setup, image quality is degraded by scatter. If the planning MRI could be used for scatter correction it would improve image quality and facilitate image registration.

Despite the possible benefits, surprisingly little has been done in this particular area although interest appears to be increasing as combined PET/MRI machines make their way into clinical practice [2]. In contrast, great effort has been devoted to skull stripping, the process of segmenting brain from non-brain tissue without further considering the composition of the non-brain tissue [3], as well as the subsequent segmentation of the brain into different types of soft tissue [4]. What sets skull segmentation apart from soft tissue segmentation and skull stripping is the need to distinguish between bone and air, neither of which provides a signal in the MR sequences used routinely. This is particularly troublesome since bone is the most and air the least dense material in the body.

One way around this is to use a sequence with ultrashort echo time, such that signal from bone is acquired [5], [6]. However, such sequences have not reached widespread use, and so the need to automatically segment anatomical MR images remains. Attempts relying purely on intensity information have been made with some success, despite the problem of distinguishing air from bone [7]. In recent publications techniques based on deformable atlas registration are the predominant ones [8], [9]. Furthermore, a method which combines this with pattern recognition has shown promising results [10].

Several of the aforementioned studies report difficulties with the skull segmentation in particular. We propose a novel approach that uses a machine learning algorithm, known as a support vector machine, to perform the segmentation by combining local image features with prior information. A key rationale for including a priori information is to distinguish
between bone and air, which is not possible based only on image intensity. We show that a suitable prior, which has a high predictive power in its own right, can be derived from CT images.

In the suggested approach the best use of the available information is decided upon training the algorithm with examples instead of explicitly incorporating it in an ad hoc fashion. The idea is that this approach will produce a more accurate segmentation while being easily adaptable to different machines and settings.

The main reason for choosing a support vector machine (SVM) is that this method often provides considerably better classification performance than other algorithms on datasets of moderate size [11]. Moreover, SVMs have been shown to be efficient for soft tissue segmentation in MR images [12].

II. METHODS

In short, the proposed algorithm works by first registering the patient’s MR to a reference MR which has an associated skull probability atlas. For each voxel, a feature vector is constructed where this probability is used as the first entry. This is supplemented by two types of local features: intensities of a set of nearby voxels and statistical moments of the local surroundings. Each voxel’s feature vector is presented to a trained SVM that classifies it as skull or other.

In total, 40 patients were used: 20 for training, 10 for validation and 10 for evaluation. The validation set is used to tune the performance during training whereas the evaluation set is used to test the final performance.

A. Feature selection

A key aspect of any accurate machine learning algorithm is the choice of features on which to base the prediction. In particular, the mere number of features is important to consider since the necessary amount of data increases rapidly with the size of the feature space. If one can pinpoint a set of features that all have a substantial effect on the performance there is much to gain. Apart from giving a better model understanding and reducing training times, a smaller model will also be less prone to overfitting and so will generalize better.

To this end, the implemented features were determined using so called forward selection. This is a sequential procedure that starts with no features; checks which feature gives the largest improvement—as measured by cross-validation—and adds that; checks which additional feature gives the largest improvement and adds that and so forth. This is repeated until the improvement is smaller than a predetermined threshold.

In order of importance, the three features included in the final model were:

1) Probability atlas: The idea behind including a probability value is to capture the normal appearance, thereby ensuring that the prediction will not differ significantly from what can be expected. This probability distribution was estimated using CT images from 753 patients. Each patient’s CT was registered to a reference patient, after which the bone was extracted by thresholding at 300 Hounsfield units. The probability distribution was estimated by averaging over all patients. The result is illustrated in Fig. 1.

Since the probability atlas is defined in the coordinates of the reference MR, a registration is necessary to relate it to the patient’s MR (henceforth referred to as the target MR). This registration and those used to create the probability atlas and the training set are depicted in Fig. 2. Each training patient’s CT was rigidly registered and resampled to the corresponding MR. Registration of images from different patients were also allowed scale transformations. All registrations were carried out using a modified version (to enable scaling) of the normalized mutual information based registration algorithm in Leksell GammaPlan®.

2) Neighbourhood intensities: The intensities of the 7 axis-centered voxels along each coordinate axis, as shown in Fig. 3a, were included as a local feature (19 values in total).

3) Statistical moments: To reflect the local texture, statistical moments of the intensity histogram were included. The first moment \( m \) is the mean value in that region and moments \( \mu_n \) of order \( n \geq 2 \) are given by

\[
\mu_n(z) = \sum_{v \in V} (z_v - m)^n
\]

where \( z_v \) is the intensity value in voxel \( v \) which belongs to the region \( V \), taken to be a cube with a side length of 3 voxels (Fig. 3b). The first 5 moments were included as the second local feature.

The candidate features that were left out included: the same local features but computed over regions of different sizes, intensities along the gradient direction and voxel position in spherical coordinates.

B. Training procedure

The overall aim of machine learning is to make predictions on unseen data based on similar data. In particular, algorithms
have to strike a balance between learning the important trends and disregarding statistical variations present in the training data. This is not easy and no algorithm exists which is always superior to all others. Nevertheless, Support Vector Machines (SVMs) often perform well, which explains why it is one of the most popular algorithms in modern machine learning.

Given \( n \) points \( x_i \) in feature space, each belonging to either of two classes denoted by a class indicator \( y_i \in \{-1, 1\} \), the main idea behind SVMs [13] is to maximize the separation between examples in the two classes. This leads to the convex optimization problem

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\
\text{subject to} & \quad y_i^T \alpha = 0 \quad i = 1, \ldots, n \\
& \quad 0 \leq \alpha_i \leq C, \quad i = 1, \ldots, n
\end{align*}
\]

(2)

where \( e = [1, \ldots, 1]^T \) is the vector of all ones, \( C > 0 \) is a regularization parameter and \( Q \) is an \( n \times n \) matrix, defined as \( Q_{ij} = y_i y_j K(x_i, x_j) \), where \( K(x_i, x_j) \) is a kernel function. In this work, a Gaussian radial basis function,

\[
K(x_i, x_j) = \exp \left(-\gamma \|x_i - x_j\|^2\right)
\]

(3)

with shape parameter \( \gamma \), was used as kernel. Suitable choices of the parameters \( C \) and \( \gamma \) were determined by cross-validation to be \( 2^{-9} \) and \( 2^{-1} \), respectively.

In practice, libraries tailored to solve problem (2), such as LIBSVM [14], greatly facilitate the use of SVMs.

Since a single CT or MRI head study contains a vast number of voxels (~10^7) and we use data from several patients, we found it necessary to sample training data. A completely random sampling would contain relatively few examples of bone and many of the air outside the head. A trained SVM only makes use of the training examples on the boundary between the two classes. So, to increase the fraction of relevant cases in the training data, half of the samples were drawn randomly from voxels known to be bone whereas the other half was sampled from voxels which were not. Moreover, the latter samples were required to be situated where the estimated probability of being bone was greater than one percent according to the probability atlas.

C. Performance measures

A full description of a binary classifier’s performance is given by a confusion matrix, schematically illustrated in Table I. Although the confusion matrix can be used to analyze trade-offs in performance, it is necessary for the feature selection to condense this information into a single measure. One measure recommended in [15] is the Matthews Correlation Coefficient (MCC). It can be calculated, using the definitions from Table I, as

\[
\text{MCC} = \frac{P_{11}P_{00} - P_{10}P_{01}}{\sqrt{(P_{11} + P_{10})(P_{11} + P_{10})(P_{00} + P_{01})(P_{00} + P_{01})}}
\]

(4)

From this it can be seen that the MCC assumes values between -1 and +1, where +1 corresponds to perfect agreement and -1 to total disagreement between predicted and actual values. A value of 0 means that the classifier is no better than a random prediction.

To facilitate comparisons with related work, we compute the Dice coefficient \( D \) given by

\[
D(S_{\text{MR}}, S_{\text{CT}}) = \frac{2|S_{\text{MR}} \cap S_{\text{CT}}|}{|S_{\text{MR}}| + |S_{\text{CT}}|}
\]

(5)

where \( S_{\text{MR}} \) and \( S_{\text{CT}} \) are the sets of voxels segmented as skull using MR and CT, respectively.

To compare with [9] we will also consider the distribution of distances from voxels segmented as bone in the CT to the closest voxel segmented as bone in the MR. However, this
TABLE II. RESULTS FOR THE PROPOSED METHOD AND [7] (IN PARENTHESES)

<table>
<thead>
<tr>
<th>Patient</th>
<th>MCC</th>
<th>Dice</th>
<th>$d$ [mm]</th>
<th>Vol(S_{MR})/Vol(S_{CT})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.69 (0.51)</td>
<td>0.70 (0.52)</td>
<td>1.44 (2.97)</td>
<td>1.23 (0.80)</td>
</tr>
<tr>
<td>2</td>
<td>0.66 (0.41)</td>
<td>0.68 (0.42)</td>
<td>2.07 (4.69)</td>
<td>1.11 (0.64)</td>
</tr>
<tr>
<td>3</td>
<td>0.67 (0.56)</td>
<td>0.69 (0.57)</td>
<td>1.93 (5.11)</td>
<td>0.86 (0.60)</td>
</tr>
<tr>
<td>4</td>
<td>0.76 (0.55)</td>
<td>0.77 (0.54)</td>
<td>1.28 (3.69)</td>
<td>1.12 (0.52)</td>
</tr>
<tr>
<td>5</td>
<td>0.71 (0.40)</td>
<td>0.73 (0.41)</td>
<td>1.53 (3.56)</td>
<td>1.09 (0.58)</td>
</tr>
<tr>
<td>6</td>
<td>0.75 (0.38)</td>
<td>0.76 (0.37)</td>
<td>1.49 (4.90)</td>
<td>0.81 (0.40)</td>
</tr>
<tr>
<td>7</td>
<td>0.76 (0.36)</td>
<td>0.77 (0.35)</td>
<td>1.45 (4.50)</td>
<td>0.81 (0.40)</td>
</tr>
<tr>
<td>8</td>
<td>0.80 (0.47)</td>
<td>0.81 (0.45)</td>
<td>0.95 (3.37)</td>
<td>1.02 (0.45)</td>
</tr>
<tr>
<td>9</td>
<td>0.74 (0.52)</td>
<td>0.76 (0.53)</td>
<td>2.40 (4.64)</td>
<td>1.01 (0.62)</td>
</tr>
<tr>
<td>10</td>
<td>0.79 (0.49)</td>
<td>0.80 (0.49)</td>
<td>3.23 (4.73)</td>
<td>1.00 (0.56)</td>
</tr>
<tr>
<td>Mean</td>
<td>0.73 (0.46)</td>
<td>0.75 (0.46)</td>
<td>1.78 (4.22)</td>
<td>1.01 (0.56)</td>
</tr>
</tbody>
</table>

is not an entirely satisfactory measure since a perfect score could be achieved by, trivially, classifying every voxel as bone. Instead, we propose to use the symmetric root mean square distance. To define this, let

$$d(v, S_{MR}) = \min_{u \in S_{MR}} \| v - u \|_2$$ (6)

denote the shortest distance from voxel $v$ to the set of voxels $S_{MR}$. This is easily calculated using a distance transform. The root mean square distance from voxels in the set $S_{CT}$ to the set $S_{MR}$ is

$$d_{rms}(S_{CT}, S_{MR}) = \sqrt{\frac{1}{|S_{CT}|} \sum_{v \in S_{CT}} d^2(v, S_{MR})}$$ (7)

which allows us to define the symmetric root mean square distance, $d$, as

$$d = \sqrt{\frac{1}{2} d_{rms}^2(S_{CT}, S_{MR}) + \frac{1}{2} d_{rms}^2(S_{MR}, S_{CT})}$$ (8)

that is 0 if and only if $S_{MR} = S_{CT}$.

D. Imaging modalities and settings

The MRI images used in this work were acquired with a 1.5 T Siemens Avanto using a T1-weighted 3D spoiled gradient recalled echo sequence (echo time 4.6 ms, repetition time 11 ms, flip angle 20°, voxel size 1x1x1 mm$^3$) without contrast agent.

The corresponding CT images were acquired with a Siemens Sensation 16 with tube voltage 120 kV, exposure 300 mAs, in-plane resolution 0.5x0.5 mm$^2$ and slice thickness 1 mm. For the CT images used to make the probability atlas, the majority have been taken with the exact same settings except that 292 out of the total 753 image stacks had a slice thickness of 2 mm instead. The ground truth was extracted from the CT images by classifying all voxels with a Hounsfield value greater than 300 as bone. This threshold value was deemed a reasonable trade-off between including trabecular bone and excluding soft tissue, although partial volume effects sometimes make such a distinction impossible.

III. RESULTS

In the feature selection, we required an MCC improvement of at least 0.01 to include further features. The forward selection then came to a halt after three iterations. In order of importance, the three features included were: the probability atlas, the neighbourhood intensities for ±3 voxels and statistical moments computed over 3$^3$ voxels.

A SVM using these features was trained on a set containing 100 samples from each of the 20 training patients, for a total of 2000 training samples. Each of the 10 test patients were registered to the reference MR. In these images, each voxel at a location with a value of the probability atlas greater than 0.01 was classified by the SVM. This restriction rules out a large amount of voxels that are highly unlikely to be bone, thereby allowing the whole procedure to take only a few minutes, whereof about half is needed to compute the statistical moments.

For comparison, we performed skull segmentations of the same patients using an algorithm based on mathematical morphology [7] that is part of the BrainSuite software [16]. In their skull stripping step, which is necessary prior to the skull segmentation, we used the following set of parameters: diffusion iterations 5, diffusion constant 50, edge constant 0.7 and erosion size 2. The skull segmentation was done with the default settings.

Table II summarizes the results on the set of test patients for the proposed method and the one of [7]. The mean and standard deviation of the performance measures explained in Section II-C were: for MCC 0.73 ± 0.043 with the proposed method and 0.46 ± 0.073 with the other; for Dice 0.75 ± 0.045 and 0.46 ± 0.076 and for symmetric root mean square distance 1.78 ± 0.66 and 4.22 ± 0.74. Compared with an individually optimized thresholding of the probability atlas, the proposed method resulted in a mean improvement in MCC of 0.045 ± 0.020.

In addition to these measures we also considered the total volume of skull voxels since this is likely to be important for the intended applications. Consequently, Table II includes the quotient of Vol(S_{MR}), the volume of voxels segmented from MR, to Vol(S_{CT}), the volume of voxels segmented from CT. Taken over all the test patients, the volume quotients were 1.01 ± 0.14 and 0.56 ± 0.12 for the respective methods.

Histograms of the distribution of distances from voxels segmented as bone in the CT to the closest voxel segmented as bone in the MR are shown in Fig. 4 (cf. Fig. 5(a) in [9]). In the mathematical terms of Eq. (6) the set in question is \{d(x, S_{MR}) | x \in S_{CT}\}.

To visualize how the accuracy varies spatially Fig. 5 shows 3D renderings of the patients with the best (patient 8) and the worse (patient 2) results. The streaks in the renderings correspond to fragments of the stereotactic frame that the preprocessing failed to remove.

IV. DISCUSSION

The proposed method consistently outperforms the one based on mathematical morphology [7]. In their defense they do point out that their method is unreliable in the lower portion of the head because of its complex morphology—an observation that is partly true also for the proposed method. A factor contributing to this is that the MR and CT are acquired at different times and so there may be differences in spine, teeth and jaw bone positions.
The proposed method also competes successfully with the algorithm based on deformable registration. For example, it can be seen from Fig. 4 that for more than 90% of the voxels segmented as bone in the CT the distance to the closest voxel classified as bone in the MR is less than 3 mm. This can be compared with [9] where the corresponding fraction was about 80%

An advantage of the SVM algorithm is that it is robust to a certain degree of misclassification in the training data— in our case remaining parts of the stereotactic frame labeled as bone (the Hounsfield values are in the same range). With the intended applications in mind, it is thus encouraging that the volume of the segmented skull is centered around the correct volume.

The forward selection showed that the probability atlas had the highest predictive power of all features. It relies, however, on an accurate registration, which therefore deserves special attention. For example, spatial distortions adversely affect the registration and should be counteracted. Because they depend on the magnetic field strength and the image acquisition protocol (including distortion correction) it is, however, hard to make a general statement of their severity.

In an application, the training time is not an issue since it only has to be done once. On the other hand, the number of support vectors directly affect the time for classification. Thankfully, the SVM was found to generalize well based on a limited data set. Moreover, by performing part of the computation on a graphics processing unit (GPU), it has been shown possible to reduce computation time by a factor of around 120-150 [17].

It seems plausible that the method presented could be further improved by combining information from different MR sequences, e.g. T2- or proton density weighted. Another possibility would be to use a more general deformation model, which would add to the computational complexity but might improve the segmentation.

V. CONCLUSION

In this work we have demonstrated the feasibility of using a support vector machine classifier to segment the skull bone in T1-weighted MR images. Since T1-weighted images are more or less always acquired when performing radiosurgery planning this could potentially allow for a smooth implementation into current workflows; both saving time for the clinic and sparing the patient from additional radiation.

Despite the few degrees of freedom included, the method resulted in accurate segmentations which compared favorably with both a method based on mathematical morphology [7] and a method based on deformable registration [9]. The proposed method is however highly dependent on an accurate registration between the patient’s MR and the MR used as reference.

The time required to perform a segmentation of a patient’s MR was on the order of a few minutes which is promising considering that we believe that there is still room for large time-savings. Future work might include pursuing these time-savings, including information from several MR sequences and investigating more advanced MR to MR registrations that possibly make use of more degrees of freedom.

ACKNOWLEDGMENT

The authors would like to thank Stephen Towe and Dr Tommy Hinks for helpful comments in the preparation of the manuscript and Marcus Hennix for help with the visualization of the results. The research was supported by the Swedish Research Council (grant 2012-4281) and the Linneaus center CADICS.

REFERENCES

Fig. 5. Illustration of the segmentation for the best (top) and the worst (bottom) test patients. The images to the left show the bone extracted from CT in white and from MR in orange. The images to the right show the distances (in millimeters) to the bone segmented in MR overlaid in color on the bone extracted from CT.


