Facial Nerve Image Enhancement from CBCT using Supervised Learning Technique

Ping Lu[1], *Member, IEEE*, Livia Barazzetti[1], Vimal Chandran[1], *Member, IEEE*, Kate Gavaghan[2] Stefan Weber[2], *Member, IEEE*, Nicolas Gerber[2], and Mauricio Reyes[1], *Member, IEEE*

***Abstract*— Facial nerve segmentation plays an important role in surgical planning of cochlear implantation. Clinically available CBCT images are used for surgical planning. However, its relatively low resolution renders the identification of the facial nerve difficult. In this work, we present a supervised learning approach to enhance facial nerve image information from CBCT. A supervised learning approach based on multi-output random forest was employed to learn the mapping between CBCT and micro-CT images. Evaluation was performed qualitatively and quantitatively by using the predicted image as input for a previously published dedicated facial nerve segmentation, and cochlear implantation surgical planning software, OtoPlan. Results show the potential of the proposed approach to improve facial nerve image quality as imaged by CBCT and to leverage its segmentation using OtoPlan.**

I. INTRODUCTION

Cochlear implantation is a conventional treatment for patients suffering from profound hearing loss. The surgical operation for cochlear implant requires mastoidectomy, to access the cochlea and avoid critical anatomical structures. To minimize the invasiveness of the surgical operation, a surgical robot system with an associated planning tool, OtoPlan [1], has been developed. OtoPlan assists the robotic system to perform drilling for direct cochlear access [2]. One of the main challenges of this procedure is to avoid the facial nerve with a margin of at least $0.5mm$. Any damage of the facial nerve causes temporary or permanent paralysis in the ipsilateral face. Hence, an accurate facial nerve segmentation is a critical step for an effective surgical plan.

The surgical planning is performed on cone-beam computed tomography (CBCT) images. In clinical practice, CBCT images are acquired with reduced radiation dose to patients, which may result in low image quality and less clear structure border. The diameter of the facial nerve lies in the range of $0.8-1.7mm$ [3]. Accurate segmentation of the facial nerve from the acquired CBCT images is challenging, mainly in the border region. Image enhancement is hypothesized to enhance the overall image quality and thereby to improve facial nerve segmentation.

In recent years, several CBCT image enhancement algorithms based on deterministic models have been proposed [4], [5], [6]. However, they were built from a priori knowledge of the imaged anatomy or imaging process. Alternatively, supervised learning has been proposed to learn the relationship between the acquired low-resolution and corresponding high-resolution image [7]. In this work, we propose to apply supervised learning based on multi-output random regression forest to enhance the image quality, in order to obtain a faster and more reliable facial nerve segmentation in the framework of pre-operative cochlear planning.

Below, the proposed approach is described and a detailed description of the image enhancement process for facial nerve segmentation is presented. An initial evaluation of the approach performed on CBCT images of cadaveric specimen and segmentation results, as compared to ground truth micro-CT images, is presented.

II. METHODS

In supervised learning, image features $\boldsymbol{x} = (x_1, ...x_n) \in \mathbb{R}^n$ computed on clinical CBCT are mapped to the corresponding output response $\boldsymbol{y} = (y_1, ...y_m) \in \mathbb{R}^m$ computed on micro-CT. The mapping is cast as a regression problem. Given a training set $\{\langle \boldsymbol{X_i}, \boldsymbol{Y_i} \rangle | i = 1, ..., N\}$ of CBCT and micro-CT aligned pairs of images, we extract from each i_{th} image, a feature vector $\boldsymbol{X_i} = (\boldsymbol{x}_1, ...\boldsymbol{x}_C) \in \mathbb{X}$ and responses $\boldsymbol{Y_i} = (\boldsymbol{y}_1, ...\boldsymbol{y}_C) \in \mathbb{Y}$ over a grid of C voxels. Then, a function $\hat{\boldsymbol{y}} : \mathbb{X} \mapsto \mathbb{Y}$ from a space of features $\mathbb{X}$ to the space of responses $\mathbb{Y}$ that predicts the response for any new test image feature $\boldsymbol{X}_{test} \in \mathbb{X}$ is constructed.

The complete pipeline is presented in Fig. 1, and is described below.

A. Feature Extraction

1) Input Features: For feature extraction, the CBCT image was rigidly registered and resampled to the micro-CT image, in order to capture the image mapping from low- to high- resolution at the same spatial locations. A uniform sampling grid with isotropic grid spacing of $0.054mm^3$ was defined over the CBCT and micro-CT images. At each node c_j of the grid $j = \{1, ...C\}$, a volume of interest (VOI) $5 \times 5 \times 5$ was extracted on which feature descriptors were computed.

We propose to use two family of features, intensity- and texture-based. The intensity-based features includes all the intensity values. The texture-based features includes first order statistical measures and the gray-level co-occurrence matrix (GLCM) [8], [9]. The list of texture features is presented in Table I. For texture-based features, a feature

[1]Ping Lu, Livia Barazzetti, Vimal Chandran and Mauricio Reyes are with the Institute for Surgical Technology & Biomechanics, University of Bern, CH-3014 Bern, Switzerland. e-mail: ping.lu@istb.unibe.ch

[2]Nicolas Gerber, Kate Gavaghan and Stefan Weber are with the ARTORG Center for Biomedical Engineering Research, University of Bern, CH-3010 Bern, Switzerland.

pooling was performed, followed by principal component analysis (PCA) to reduce dimensionality and redundancy of feature sets. This reduces the feature space from $\mathbb{R}^{214}$ to $\mathbb{R}^{34}$. Hence, the input feature set $\boldsymbol{v} = (v_1, ...v_n)$ includes all the image intensities, first order statistics and mean and variance of all the pooled GLCM features (i.e. $n \in \{\mathbb{R}^{125} + \mathbb{R}^{34} = \mathbb{R}^{159}\}$)

TABLE I

LIST OF TEXTURE - BASED FEATURES COMPUTED AT EACH GRID NODE.

Texture Features	
1st Order Statistics	**GLCM**
Mean	Energy
Std.Dev	Entropy
Skewness	Correlation
Kurtosis	Inertia
Minimum	Cluster Shade
Maximum	Cluster Prominance
	Inverse Difference Moment
	Haralick Correlation

2) Output Response: A VOI of $3 \times 3 \times 3$ was extracted at each corresponding grid node c_j from micro-CT. The output response set $\boldsymbol{y} = (y_1, ...y_m) \in \mathbb{R}^{27}$ includes all the intensity values.

B. Multi-Output Regression Model

Decision forests are a group of learning methods widely used for classification and regression tasks in machine learning and computer vision. An extension of decision forest with extra trees algorithm has been proposed to handle multi-ouput image classification [10], [11]. We adopted this technique as a regression approach for its ability to preserve local intensity patterns. During supervised learning, the algorithm randomly selects without replacement K input variables $\{v_1, ...v_k\}$ from the training data $\boldsymbol{D} := \{\langle \boldsymbol{X_i}, \boldsymbol{Y_i} \rangle | i = 1, ..., N\}$. For each selected input variable, within the interval $[v_i^{min}, v_i^{max}]$ a cutpoint s_i was randomly defined, followed by splitting $[v_i < s_i]$. Among the K candidate splits, the best split was chosen via optimizing the L2 mean square error [11].

During testing, image features were extracted and passed through the regression random forest. The computed output corresponds to the intensity of the central voxel of the designed 3x3x3 voxels window. No further post-processing was performed on the resulting image.

III. RESULTS

We report in this study preliminary results obtained on a database of right and left ears from 4 cadaver heads following an approved clinical study. Pairs of CBCT ($0.15 \times 0.15 \times 0.15 mm^3$) and micro-CT ($0.018 \times 0.018 \times 0.018 mm^3$) images were acquired from the four heads. The images were rigidly aligned using a rigid registration transform, normalized cross-correlation and gradient descent optimization. For rigid registration, we defined CBCT as the moving image and micro-CT as the fixed one. Then we resampled CBCT with the micro-CT voxel spacing. From the resulting rigidly aligned images, image patches were extracted and used for the training phase.

For evaluation, we manually segmented the facial nerve from the micro-CT image (referred hereafter as ground-truth). We used the software OtoPlan [1] to perform segmentation of the facial nerve from the original CBCT image and its corresponding enhanced version. OtoPlan is a dedicated state-of-the-art software for cochlear implantation surgical planning.

We performed qualitative visual assessment of the resulting segmented facial nerve as well as a quantitative analysis of surface-to-surface distances to the ground-truth segmentation. For implementation, we use the scikit-learn package [12] and its ExtraTreesRegressor, which implements a multi output random forest regression algorithm. For model parameterization (i.e. number of estimators, number of trees, etc.) we adopted a leave-one-out strategy. Additionally, we investigated the influence of the window size on the prediction accuracy.

Figure 2 shows the results obtained after regression forest application to a CBCT image. Compared to the input CBCT image, the resulting enhanced image is much sharper and able to characterize image details not completely discernible from the CBCT image.

Following the recent findings from [7], we further explored the importance of the window size used to extract feature information during the training phase, Our results showed that using short range features (i.e. window size in our application smaller than $0.018mm$) yields suboptimal prediction results. Conversely, employing an excessively large window size yields smoother but inaccurate prediction results. This result reflects the ability of the prediction model to learn

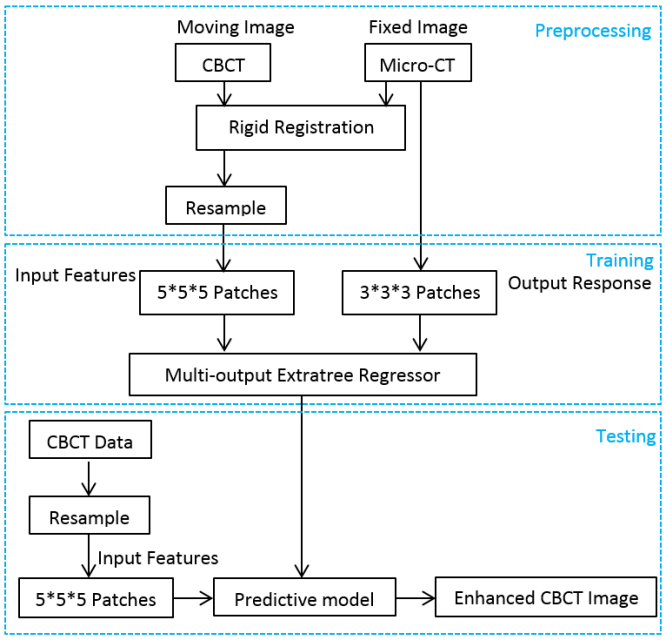
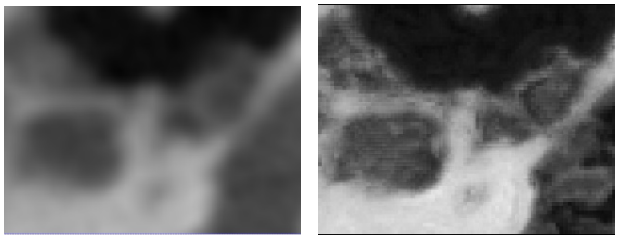


Fig. 1. The complete pipeline of the proposed approach for enhancing CBCT image



(a) original CBCT (b) enhanced CBCT

(c) original micro-CT (d) "short range" prediction

Fig. 2. Results of supervised-learning based CBCT image enhancement. Image features are extracted from (a) original CBCT image, and used to produce an enhanced version (b), that presents sharper and more clear structures, as compared to the original high-resolution micro-CT image (c). For demonstration purposes, we report results obtained using features extracted from a short range (i.e. small window size) (d), indicating the ability of the model to learn and utilize local structural information for the prediction process.

local structural information that leverages the prediction of the image content at lower resolution scales.

In the next section, we present preliminary results of segmentation of the facial nerve in the framework of minimally invasive cochlear implantation surgical planning, using the enhanced CBCT as input for the dedicated software tool OtoPlan.

A. *Facial Nerve Segmentation*

OtoPlan features a semi-automatic segmentation of the facial nerve. Based on a GUI-based tool, the user selects landmarks that approximately lie on the facial nerve's midline. A panoramic visualization is then constructed and displayed to the user, which corresponds to an "unfolding" of the facial nerve into one single view. The selected landmarks are displayed and used to cast intensity sampling lines that are perpendicular to the approximate midline of the facial nerve. A threshold based scheme is then used by OtoPlan to find the facial nerve wall as initialization prior to manual adjustments. Figure 3 illustrates this part of the segmentation process.

As the image contrast from CBCT is relatively poor, this process can be daunting and prone to errors, which in turn necessitates manual corrections. We performed a preliminary evaluation by comparing segmentation results obtained using the original and the enhanced CBCT image. OtoPlan can then generate and export a mesh representation of the segmented facial nerve. We measured surface-to-surface distances between each generated mesh and the corresponding ground-truth generated mesh. Figure 4 shows a particular example result where distances are color-coded

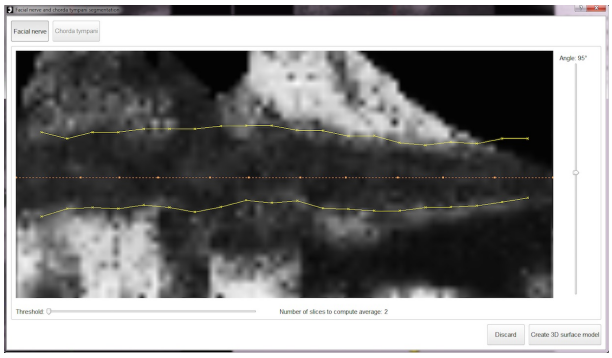
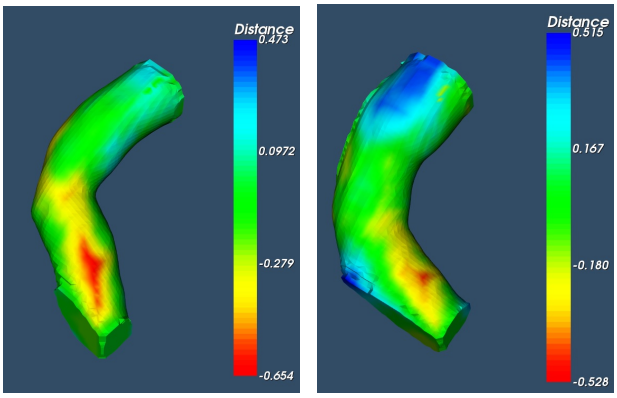


Fig. 3. Panoramic view for semi-automatic segmentation of the facial nerve. The user selects a set of landmarks that approximately correspond to the middle line of the facial nerve. A threshold based scheme is then used to cast sampling perpendicular in order to find the facial nerve wall (above and below the middle line). Due to the low contrast quality of the CBCT image, manual correction of each point is commonly required. In this example, we illustrate the enhanced CBCT image.



(a) Original CBCT (b) Enhanced CT

Fig. 4. Surface-to-surface distances from the ground-truth segmentation to OtoPlan segmentations generated using the original and enhanced CBCT image. Colormap encodes distances and best viewed in electronic version.

to visualize deviations from the ground-truth segmentation. Employing the enhanced CBCT image provides OtoPlan with a sharper and better delineated facial nerve wall that yields a more precise segmentation of the facial nerve (i.e. lower surface-to-surface distances with respect to the ground-truth).

IV. CONCLUSIONS

Good characterization of the facial nerve is of crucial importance for a safe planning of cochlear implantation interventions. Although CBCT-based imaging provides means to image the facial nerve in patients, its relatively low image contrast hinders a precisely definition of the facial nerve wall. In this work we proposed a machine-learning based approach that uses a supervised learning paradigm to learn the mapping between low-, and high-resolution imaging of the facial nerve. The approach relies on a multi-output regression forest and image features, extracted from the clinically available CBCT image, to perform voxel intensity prediction at the micro-CT image resolution level. Preliminary results on

CBCT images of the facial nerve show the ability of the proposed approach to enhance the imaging information by performing a prediction of voxel intensity information at the equivalent micro-CT resolution level. These first results also show the potential of the proposed approach to assist state-of-the-art cochlear surgical planing software, such as OtoPlan, to segment the facial nerve more precisely.

We report similar findings aligning with the literature on supervised-learning based image quality transfer [7], where local structural information from the low-resolution image was shown to convey information that can be used to predict localised voxel intensity information at the high-resolution image level. Our experiments also suggest the advantage of using a multi-output regression forest, in contrast to a single-output regression forest, in order to promote the learned local structural information.

The proposed approach presents some limitations. As for other supervised-learning based approaches, it is important to have a database of samples that characterises the expected variability of a population. Secondly, the mapping learned between low-, and high-resolution is specific to the imaging parameters used for the training database, and therefore a new model is potentially needed in case these parameters are modified. This can be circumvented, for instance, by designing imaging features that are independent of the energy parameters used for CT devices, or by designing compensation strategies based on an imaging phantom.

Further comprehensive evaluations and comparisons are needed, especially against other previously proposed super resolution approaches, as well as a more comprehensive quantitative evaluation on a larger dataset. Other future work will include an attempt to combine this approach with a fully automatic segmentation of the facial nerve that uses shape priors, as proposed by others [13], but that does not employ computationally expensive non-rigid registration techniques.

ACKNOWLEDGMENT

This research is funded by Nano-Tera.ch, scientifically evaluated by the SNSF and financed by the Swiss confederation.

REFERENCES

[1] N. Gerber, B. Bell, K. Gavaghan, C. Weisstanner, M. Caversaccio, and S. Weber, "Surgical planning tool for robotically assisted hearing aid implantation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 9, pp. 11–20, 2014.

[2] B. Bell, N. Gerber, T. Williamson, K. Gavaghan, W. Wimmer, M. Caversaccio, and S. Weber, "In vitro accuracy evaluation of image-guided robot system for direct cochlear access," *Otology Neurotology*, 2013.

[3] F. A. Reda, J. H. Noble, A. Rivas, T. R. McRackan, R. F. Labadie, and B. M. Dawant, "Automatic segmentation of the facial nerve and chorda tympani in pediatric ct scans," *Medical physics*, vol. 38, no. 10, pp. 5590–5600, 2011.

[4] Y. Lou, T. Niu, X. Jia, P. A. Vela, L. Zhu, and A. R. Tannenbaum, "Joint CT/CBCT deformable registration and CBCT enhancement for cancer radiotherapy," *Medical Image Analysis*, vol. 17, no. 3, pp. 387–400, 2013.

[5] T. E. Marchant, C. J. Moore, C. G. Rowbottom, R. I. MacKay, and P. C. Williams, "Shading correction algorithm for improvement of cone-beam CT images in radiotherapy." *Physics in medicine and biology*, vol. 53, no. 2008, pp. 5719–5733, 2008.

[6] T. Niu and L. Zhu, "Overview of x-ray scatter in cone-beam computed tomography and its correction methods," *Current Medical Imaging Reviews*, vol. 6, no. 2, pp. 82–89, 2010.

[7] D. C. Alexander, D. Zikic, J. Zhang, H. Zhang, and A. Criminisi, "Image Quality Transfer via Random Forest Regression : Applications in Diffusion MRI," *Medical Image Computing and Computer-Assisted Intervention MICCAI 2014*, vol. 8675, pp. 225–232, 2014.

[8] R. M. Haralick and K. Shanmugam, "Textural Features for Image Classification," *IEEE TSMC-3*, vol. 6, pp. 610–621, 1973.

[9] A. Ortiz, A. A. Palacio, J. M. Górriz, J. Ramírez, and D. Salas-González, "Segmentation of brain MRI using SOM-FCM-based method and 3D statistical descriptors." *Computational and mathematical methods in meindicine*, p. 638563, 2013.

[10] M. Dumont and R. Marée, "Fast multi-class image annotation with random windows and multiple output randomized trees," *Proc. International Conference on Computer Vision Theory and Applications (VISAPP) Volume*, vol. 2, pp. 196–203, 2009.

[11] R. Marée, L. Wehenkel, and P. Geurts, "Extremely randomized trees and random subwindows for image classification, annotation, and retrieval," *Decision Forests for Computer Vision and Medical Image Analysis*, pp. 125–134, 2013.

[12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[13] J. H. Noble, F. M. Warren, R. F. Labadie, and B. M. Dawant, "Automatic segmentation of the facial nerve and chorda tympani in ct images using spatially dependent feature values," *Medical physics*, vol. 35, no. 12, pp. 5375–5384, 2008.