Semantic 3-D Labeling of Ear Implants using a Global Parametric Transition Prior

Alexander Zouhar, Carsten Rother, Siegfried Fuchs

University of Technology Dresden

Abstract. In this work we consider the problem of sematic part-labeling of 3-D meshes of ear implants. This is a challenging problem and automatic solutions are of high practical relevance, since they help to automate the design of hearing aids. The contribution of this work is a new framework which outperforms existing approaches for this task. To achieve the boost in performance we introduce the new concept of a global parametric transition prior. To our knowledge, this is the first time that such a generic prior is used for 3-D mesh processing, and it may be found useful for a large class of 3-D meshes. To foster more research on the important topic of ear implant labeling, we collected a large data set of 3-D meshes, with associated ground truth labels, which we will make publicly available.

1 Introduction

Semantic part-labeling of organic surfaces is an important task in computer-aided shape modeling and visualization. The problem is to partition a polygonal surface mesh into non overlapping subsurfaces each of which represents a semantic part of the underlying object. Such decomposition into parts is extremely challenging due to the anatomical variability. Moreover, it is typically not possible to consistently infer the transition boundaries between adjacent segments solely from geometric cues, and the need of strong boundary transition priors becomes immanent. Fig. 1 illustrates why this is the case in the personalized computer-aided shape modeling of ear implants. Input to the process is a polygonal surface mesh capturing a patient's outer ear geometry. The essence of the hearing aid design process is captured by a part-labeling of 6 anatomical surface regions and piecewise planar transition boundaries between the labeled segments. The ultimate goal is to minimize the user interactions and to maximize the label quality which involves the shape of the transition boundaries.

In standard 2-D labeling problems it is very common to employ shape priors to drive the segmentation towards a meaningful result. Existing region and boundary models over arbitrary shapes mainly use region and length priors. For example, in [6] the authors distinguish between regions being interior/exterior to each other along with preferred distances between their boundaries. In [7] the authors compute a globally optimal labeling for tiered scenes where the correct order between the objects and the parts is enforced.

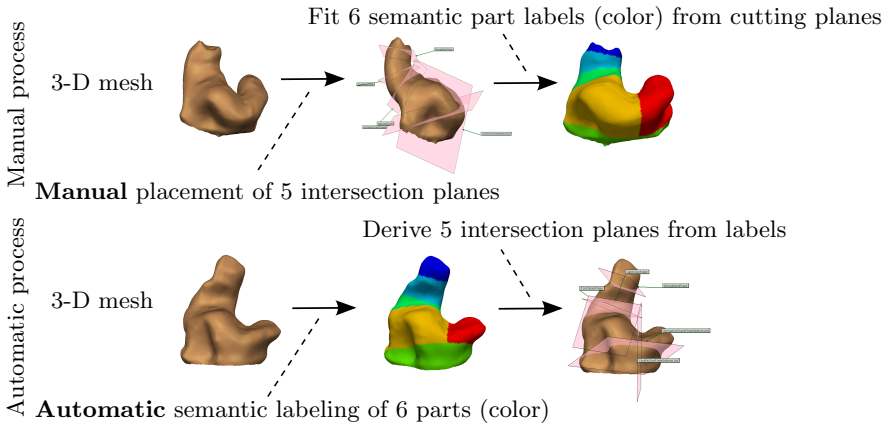


Fig. 1: Semantic 3-D labeling of ear implants in hearing aid (HA) design. **(Top row)**: A domain expert manually places 5 cutting planes along anatomical lines based on hearing aid design rules. This manual procedure is very cumbersome. Part labels (color) corresponding to 6 anatomical regions are derived from the cutting planes. Both, the cutting planes and the labeled regions play a key role in personalized hearing aid design. **(Bottom row)**: Part-labeling using our algorithm. Labels and transition boundaries between segments are optimized jointly (Bottom middle). Planes are derived from the labels (Bottom right).

The most recent state of the art approaches segment and label 3-D surface meshes jointly. The authors in [8], for example, collect statistics of neighboring surface features to learn a conditional random field (CRF) with local pairwise interactions. The prior model in [8], however, is not well suited to adequately constrain the transition boundaries between adjacent segments to an a-priori known parametric form. Consider, for example, the case of subtle shape variations in form of bends which are typical for organic surfaces. In this case the geometry dependent likelihood of a difference in labels, as proposed in [8], tends to be constant (or zero) across the surface.

Not much work exists on the labeling of ear shapes. Similar to [8] the authors in [12] use data dependent pairwise terms to penalize inconsistent labels based on local feature statistics. An alternative approach is presented in [13] where the authors employ multiple shape class specific CRFs with pairwise Potts interactions to overcome the large variability of the ear. While both methods achieve reasonable recognition rates the transition boundaries between adjacent segments tend to deviate significantly from the ground-truth. By comparison, the work in [1] firstly detects a set of generic features of the ear (concavities, elbows, ridges, bumps) which are then used to derive anatomical features of the ear including points, curves, areas and cutting planes. A part labeling may readily be derived from the cutting planes as illustrated in the top row in Fig. 1. Another completely different approach would be to build a human digital ear atlas and to propagate the labels from the atlas to the new data via surface registration (see

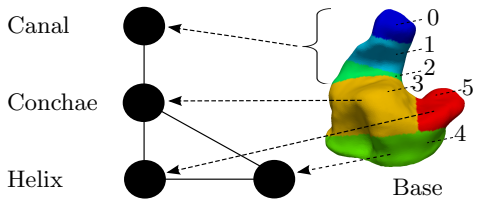


Fig. 2: Part adjacency graph of the human outer ear (courtesy [12]). Adjacent parts (black circles) in the graph are linked (solid line). The colors indicate the anatomical interpretation (dashed line) of the parts. The ear canal is composed of 3 subparts. Numbers represent anatomical part labels in the label set $\mathcal{L}$.

e.g. [2]). However, due to the variability of the ear it is extremely challenging to consistently establish anatomical correspondence across individuals.

Inspired by the practical challenge of hearing aid design, we address the surface labeling problem by jointly optimizing the part-labels and the piecewise planar transition boundaries between labeled segments as illustrated in the bottom row in Fig. 1. For this we consider 3-D surface meshes of ear implants without additional appearance information, such as color or texture. The main idea in this paper is to model the label distribution as a CRF with local pairwise interactions between labels along with a consistency term that penalizes incompatible arrangements between labels and parametric representations of the segment transition boundaries thereby emphasizing on the global consistency of a labeling. This is why we refer to the latter term as *global parametric transition prior*. Incorporating such a prior into a CRF has several advantages. Firstly, the prior encourages long range compatibility between labels giving rise to a globally consistent part layout. Secondly, the desired shape of the transition boundaries is explicitly enforced. Thirdly, the underlying energy function may be optimized jointly with respect to the labels and to the transition boundaries. Providing that the global parametric transition prior is convex the underlying energy is guaranteed to decrease monotonically during iterative optimization.

2 Model

We consider the following model. A surface mesh $\mathcal{X} = (\mathcal{V}, \mathcal{E})$ consists of vertices $\mathcal{V}$, edges $\mathcal{E}$. A labeling $h : \mathcal{V} \to \mathcal{L}$ of $\mathcal{X}$ assigns a discrete label $h_i \in \mathcal{L} = \{0, 1, 2, 3, 4, 5\}$ to each vertex $i \in \mathcal{V}$. The hearing aid design process gives rise to $|\mathcal{L}| = 6$ anatomical parts as illustrated in figure 2.

The transition boundaries are induced by 5 cutting planes passing through the mesh. Let $b = (b_1, ..., b_B)$ denote a vector of transition boundaries between the adjacent segments with $b_\ell \in \mathbb{R}^4, 1 \leq \ell \leq B$ denoting a parametric representation of the ℓth boundary. Since the transition boundaries connecting red and green, and yellow and green are induced by the same cutting plane, we may think of these two transitions as one. This is why we have $B = 5$. Moreover, let

$W_i(h_i, b, \mathcal{X}) \geq 0$ denote a function that penalizes inconsistent locations of a label h_i relative to all boundaries b_ℓ. We define the behavior of $W_i(h_i, b, \mathcal{X})$ as follows. If a label h_i resides on the *correct side* of all boundaries b_ℓ then $W_i(h_i, b, \mathcal{X}) = 0$ and $W_i(h_i, b, \mathcal{X}) > 0$ otherwise. We define the following energy function:

$$U(h, b, \mathcal{X}, \theta) = \sum_{i \in \mathcal{V}} U_i(h_i, \mathcal{X}) + \sum_{\{i,j\} \in \mathcal{E}} U_{ij}(h_i, h_j, \theta_1) + \sum_{i \in \mathcal{V}} W_i(h_i, b, \mathcal{X}, \theta_2), \quad (1)$$

where

$$U_i(h_i, \mathcal{X}) = -\log(p(h_i|\mathcal{X})), \quad (2)$$

$$U_{ij}(h_i, h_j, \theta_1) = \theta_1 \delta(h_i, h_j), \quad (3)$$

$$W_i(h_i, b, \mathcal{X}, \theta_2) = \theta_2 \sum_{\ell=1}^{B} \max\left\{0; 1 - y_\ell(h_i) < b_\ell, z_i > \right\}. \quad (4)$$

Note that the weighting parameters $\theta_1, \theta_2 \geq 0$ regularize the influence of the individual terms. The variable z_i in Eqn. (4) denotes the homogeneous 3-D coordinates of a vertex $i \in \mathcal{V}$ and $y_\ell(h_i) \in \{-1, 1\}$ indicates whether a label h_i is expected to be located above b_ℓ, $(y_\ell(h_i) = 1)$ or below b_ℓ, $(y_\ell(h_i) = -1)$. The expression $< \cdot, \cdot >$ denotes the dot product between two vectors. Eqn. (4) resembles the well known *hinge loss* function (see, e.g., [10]) which in our model gives rise to a convex global parametric transition prior. The function $\delta(h_i, h_j) \in \mathbb{N}$ in Eqn. (3) returns the smallest number of links connecting two nodes in the underlying part adjacency graph

$$\mathcal{G}_A = (\mathcal{L}, \{\{0,1\}, \{1,2\}, \{2,3\}, \{3,4\}, \{4,5\}, \{3,5\}\}) \quad (5)$$

shown in figure 2, i.e., $\delta(h_i, h_j) = |h_i - h_j|$ if both labels h_i and h_j are in the set $\{4, 5\}$ or if both labels are in the set $\{0, 1, 2, 3, 4\}$, otherwise we have $\delta(h_i, h_j) = |h_i - h_j| - 1$. While the classical Potts model enforces smoothness it does not prevent incompatible labels from being adjacent. This is achieved by the layout consistency function in Eqn. (3) which also forms a metric over $\mathcal{L}$. Since the part adjacency graph $\mathcal{G}_A$ contains a loop the energy (1) is not submodular. For the unary terms in Eqn. (2) we use a randomized decision forest (see, e.g., [5]) and 3-D shape contexts as local descriptors of the vertices $\mathcal{V}$. 3-D shape contexts [9] are rich, highly discriminative local representations of global shape which we found to work well for our data.

3 Optimization

We use the energy minimization framework to jointly derive an estimate of the labeling h and of the transition boundaries b. Minimizing Eqn. (1) with respect to h and b leads to an optimization problem of the form

$$\min_{h,b} U(h, b, \mathcal{X}, \theta). \quad (6)$$

Given h, the problem (6) reduces to

$$\min_{b=(b_1,\ldots,b_B)} \sum_{\ell=1}^{B} \sum_{i \in \mathcal{V}} \max\left\{0; 1 - y_\ell(h_i) < b_\ell, z_i > \right\}, \tag{7}$$

where we can drop the weighting parameter θ_2 in Eqn. (1) since it does not depend on b. Note, that we have 20 parameters, i.e., $b \in \mathbb{R}^{20}$. As Eqn. (7) is convex and is subdifferentiable a subgradient method [3] is well suited to solve the task. Given an estimate of b optimizing the energy (1) with respect to h may be carried out using the expansion move algorithm [4] since the pairwise terms in Eqn. (3) form a metric over the label set $\mathcal{L}$.

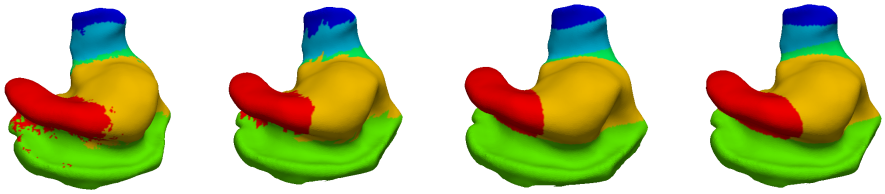
We get a very simple, iterative optimization schema. Given an estimate of h the problem (7) gives rise to a global solution since the transition prior in Eqn. (4) is convex. In turn, given an estimate of the transition boundaries b the expansion move algorithm is guaranteed to find a lower or equal energy labeling h. From this it follows that the energy in Eqn. (1) decreases monotonically. Moreover, the iterative optimization of Eqn. (6) converges more quickly if we initialize b with the least squares estimate of the cutting planes using the initial estimate of h. About 10 iterations are sufficient on our data which for a mesh with $|\mathcal{V}| \approx 20000$ vertices takes about 12s on a standard PC.

4 Learning

Given a labeled training set $\mathcal{T} = \{(\mathcal{X}, h, b)\}$ we use a supervised algorithm to learn the model parameters in Eqn. (1). The unary terms $U_i(h_i, \mathcal{X})$ comprise the decision forest structure as a parameter where we assume that the forest consists of binary trees. The parameters $\theta = (\theta_1, \theta_2)$ regularize the influence of the energy terms $U_{ij}(h_i, h_j, \theta_1)$ and $W_i(h_i, b, \mathcal{X}, \theta_2)$, respectively. Ideally, we would like to learn all model parameters jointly using a single objective function. However, whereas the weights θ_1, θ_2 are continuous variables, the random forest is a large combinatorial set. We therefore adopt a simple two-step heuristic: (1) learning of the decision forest using the labeled training data and the *information gain splitting criterion* and (2) estimation of the weights θ via cross-validation similar to [11]. To this end we allow θ to vary over a discrete possibly very large set. To keep the training process simple we follow the suggestion in [5] and proceed by growing full trees where each leaf contains only one training example.

5 Experiments

For a fair comparison of our method with prior work [1],[12],[13] we derive labels from the detected plane features in [1] as shown in the first row in Fig. 1 whereas for the methods in [12], [13] planes were fit to the inferred labels. We define a measure of label accuracy to compare the inferred labels with the ground-truth. To get a measure of label accuracy per surface we compute the *Dice* coefficient



(a) LA($\mathcal{X}$) : 4.68, (b) LA($\mathcal{X}$) : 4.71, (c) LA($\mathcal{X}$) : 4.96, (d) Ground truth

Fig. 3: Label example using (1): (a) without regularization, i.e., $\theta_1, \theta_2 = 0$, (b) without global parametric transition prior, i.e., $\theta_2 = 0$. Note, in (a–b) how the segment boundaries differ from the ground truth. In (c) we show the result after least squares fitting of planes to the labels in (b) and reassigning of misclassified labels after which the label accuracy LA($\mathcal{X}$) $\in [0, 6]$ slightly increases (6 is best).

between the area of the estimated labels of the mth part ($m \in \mathcal{L}$) $\hat{\mathcal{A}}_m = \{i \in \mathcal{V} | \hat{h}_i = m\}$ and the area of the ground-truth labels $\mathcal{A}_m = \{i \in \mathcal{V} | h_i = m\}$:

$$\mathrm{LA}(\mathcal{X}, m) = \frac{2|\hat{\mathcal{A}}_m \cap \mathcal{A}_m|}{|\hat{\mathcal{A}}_m \cup \mathcal{A}_m|}, \tag{8}$$

where $\hat{h}_i$ and h_i denote the estimated label and the ground-truth label of the ith vertex of $\mathcal{X}$, respectively. For a surface $\mathcal{X}$ the label accuracy LA($\mathcal{X}$) amounts to

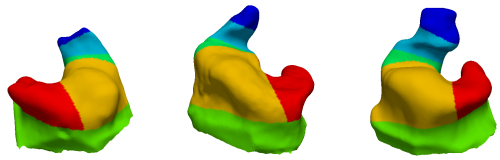
$$\mathrm{LA}(\mathcal{X}) = \frac{1}{|\mathcal{V}|} \sum_{m=0}^{|\mathcal{L}|-1} \mathrm{LA}(\mathcal{X}, m). \tag{9}$$

The value LA($\mathcal{X}$) $\in [0, 6]$ ranges between 0 and 6, where 6 is best. Note that in contrast to the classical Hamming loss the *Dice* coefficient avoids overemphasizing large area parts over small area parts. This is important, since the human ear involves regions with both large and small area segments.

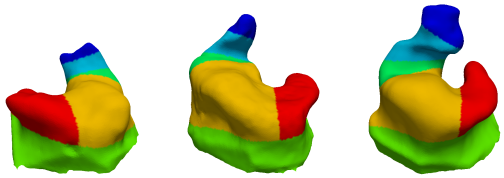
We have a novel data set of 427 human outer ear impressions at our disposal which in turn were laser scanned to reconstruct 3-D triangular surface meshes. A typical 3-D mesh of the ear is composed of roughly 20000 vertices with an average resolution of 0.22 mm. Topologically, a reconstructed outer ear surface constitutes a compact, orientable 2-manifold with boundary.

We randomly pick 90% of the surfaces for training while setting the other 10% aside for testing. For model learning the training set was divided in two halves, and the weighting parameters θ_1, θ_2 were optimized against one half via cross-validation. The randomized decision forrest was then retrained using the entire training set. For the parameters we obtain $\theta_1 = 10, \theta_2 = 20$. Inference was carried out using the algorithm in section 3.

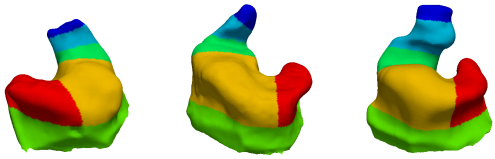
Fig. 3(a) shows a test surface labeled by our model (1) without regularization ($\theta_1, \theta_2 = 0$) and in Fig. 3(b) without global parametric transition prior ($\theta_1 = 10, \theta_2 = 0$). A visual comparison of the two results with the ground truth in



(a) LA($\mathcal{X}$) : 4.91, (b) LA($\mathcal{X}$) : 4.31, (c) LA($\mathcal{X}$) : 4.32



(d) LA($\mathcal{X}$) : 4.95, (e) LA($\mathcal{X}$) : 4.45, (f) LA($\mathcal{X}$) : 4.28



(g) Ground truth, (h) Ground truth, (i) Ground truth

Fig. 4: Most accurate label examples on our test data using (a): [1], (b): [12], (c): [13]. The 2nd row depicts the result using our model (1) with global parametric transition prior (d)-(f). In terms of label accuracy LA($\mathcal{X}$) $\in [0, 6]$ (6 is best) our model performs slightly better except for (f) when compared with (g)-(i).

Fig. 3(d) reveals several inaccuracies despite the overall consistent layout. Note, how the transition boundaries deviate from the ground-truth in Fig. 3(d). The label accuracy according to Eqn. (9) is depicted below the surfaces. In Fig. 3(c) we show the result after estimating planes from the label output in Fig. 3(a) via least squares fitting which slightly improves the label accuracy.

Next, in Fig. 4 (first row) we show the best test examples obtained by our competitors in [1], [12], [13] together with the result using our model (1) with global parametric transition prior, i.e., $\theta_2 > 0$ (Fig. 4 (second row)). In terms of label accuracy LA($\mathcal{X}$) our algorithm performs slightly better when compared with the ground-truth in Fig. 4 (third row).

For a quantitative comparison of the methods several statistics were computed over the test data which we summarize in table 1. From the table one can see that on average our model (1) with global parametric transition prior performs best. While the methods [12], [13] perform equally well they were outperformed by [1]. Also note, that our model (1) without global parametric transition prior ($\theta_2 = 0$) achieves a higher label accuracy than [1] after least squares fitting of the planes as illustrated in Fig. 3.

Table 1: Various statistics computed over 43 test examples: $\overline{\mathrm{LA}(\mathcal{X})}$ (average label accuracy), $\widetilde{\mathrm{LA}(\mathcal{X})}$ (median label accuracy), $\sigma_{\mathrm{LA}(\mathcal{X})}$ (standard deviation of label accuracy). The methods used for comparison were: our model (1) w/o global parametric transition prior ($\theta_2 = 0$), our model (1), the canonical ear signature (CES) [1], the layout CRF with spatial ordering constraints in [12], the joint shape classification and labeling (JSCL) model in [13].

	Our (Eqn. 1,$\theta_2 = 0$)	Our (Eqn. 1)	CES [1]	CRF [12]	JSCL [13]
$\overline{\mathrm{LA}(\mathcal{X})}$	3.98	4.10	3.80	3.28	3.35
$\widetilde{\mathrm{LA}(\mathcal{X})}$	4.00	4.10	3.79	3.27	3.31
$\sigma_{\mathrm{LA}(\mathcal{X})}$	0.47	0.48	0.49	0.60	0.60

6 Conclusions and future work

We have proposed a new framework for the semantic labeling of 3-D meshes of ear implants. Specifically, we incorporated a parametric representation of the transition boundaries into the labeling model which in the case of ear implants was used to enforce planar transitions between the parts. This combined model lead to a performance boost in label accuracy and outperformed previous methods for ear implant labeling. We will make our data set publicly available.

References

1. Baloch, S., et al.: Automatic Detection of Anatomical Features on 3D Ear Impressions for Canonical Representation. MICCAI (2010)
2. Baloch, S., et al.: Deformable Registration of Organic Shapes via Surface Intrinsic Integrals: Application to Outer Ear Surfaces. MICCAI (2010)
3. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge Un. Press (2004)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Analysis, and Machine Intelligence (2001)
5. Criminisi, A., Shotton, J.: Decision Forests for Computer Vision and Medical Image Analysis. Springer Publishing Company, Incorporated (2013)
6. Delong, A., Boykov, Y.: Globally optimal segmentation of multi-region objects. In: ICCV. pp. 285–292. IEEE (2009)
7. Felzenszwalb, P.F., Veksler, O.: Tiered scene labeling with dynamic programming. In: CVPR. pp. 3097–3104. IEEE (2010)
8. Kalogerakis, E., et al.: Learning 3D mesh segmentation and labeling. SIGG. (2010)
9. Koertgen, M., Park, G.J., Novotni, M., Klein, R.: 3D shape matching with 3d shape contexts. Proc. 7th Central European Seminar on Comp. Graphics (2003)
10. Nello, C., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 1 edn. (2000)
11. Winn, J., Shotton, J.: The layout consistent random field for recognizing and segmenting partially occluded objects. CVPR (2006)
12. Zouhar, A., et al.: Layout Consistent Segmentation of 3-D meshes via Conditional Random Fields and Spatial Ordering Constraints. MICCAI 2010 (2010)
13. Zouhar, A., et al.: Joint shape classification and labeling of 3-d objects using the energy minimization framework. GCPR (2013)