Part-based recognition of 3-D objects with application to shape modeling in hearing aid manufacturing

Dissertation — Abbreviated Version

ALEXANDER ZOUHAR

Technische Universität Dresden

March/31/2015

1 Introduction

Recent advances in 3-D scanning technology have fostered the research in many application areas related to 3-D mesh processing including 3-D content based retrieval, reverse engineering and more recently the digital manufacturing of implants and prosthetic devices. Particular examples of the latter include dental implants and hearing aids. Existing computer aided design (CAD) software systems often provide various mesh processing algorithms to support a wide range of customized design workflows. These digitized workflows typically involve cumbersome manual work carried out by experienced domain experts.

In this work we consider the problem of semantic part-labeling of 3-D meshes of ear implants since this helps to automate the design of hearing aids. In particular, we present a new conditional random field (CRF) framework that outperforms previous approaches for this task. Many problems in computer vision including semantic segmentation can be cast as labeling problems for which random field models provide a principled way to accommodate higher-level constraints of an object class. In particular, the semantic part-labeling of ear implants requires an understanding of the anatomy and of the hearing aid (HA) design rules.

Our work involves three major topics related to 3-D mesh processing of ear implants: representation of 3-D objects using local descriptors, registration and shape clustering of 3-D objects, semantic part-labeling of 3-D objects.

2 Problem statement

HA design involves interpreting the anatomy of the human outer ear. A domain expert interactively transforms a 3-D surface mesh into a hearing aid shape based on hearing aid design rules. A large number of CAD operations in HA design depend on *cuts*. A cut is defined by a plane passing through the mesh. Five *principal cuts* (see Fig. 1(left)) drive the design of nearly all types of hearing aids (see e.g. Fig. 1(right)). The spatial arrangement of the principal cuts induces a compositional and hierarchical part structure of the ear (described below).

The manual placement of the cutting planes (orientation, location, normal direction) is (1) cumbersome, (2) not reproducible, (3) dependent on the skills of a human operator. Therefore, the goal is to automate the process, aiming at reproducibility, faster HA design and better fitting devices.

Existing solutions for this task rely on anatomical features of the ear [BMF+10, SBB+09]. Firstly, a comprehensive set of characteristic features (e.g., peaks, concavities, elbows and bumps) is detected and represented through points, simple contours and areas on the surface. Secondly, the detected features are used to automatically initialize various types of CAD tools, such as the cuts shown in Fig. 1.

Automatic shape analysis of the human outer ear is extremely challenging: (1) existing algorithms often fail to recognize the anatomical features consistently across individuals due to the anatomical variability and due to ambiguities arising from subtle shape variations, (2) the design of hearing aids involves numerous free parameters with complex dependencies and constraints that must be satisfied.

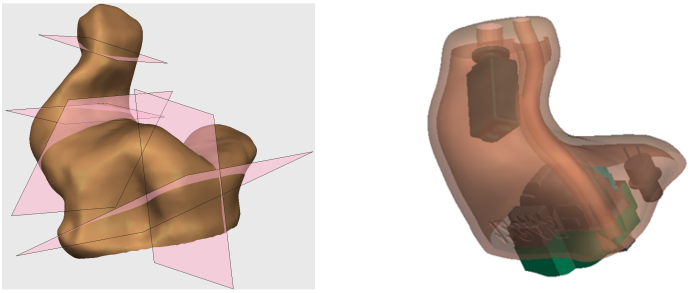


Abbildung 1: (Left) Principal cuts in hearing aid design. (Right) Final mesh (transparent) with electronic components inserted (source: [BMF+10]).

Part-based recognition of 3-D ear implant objects

The spatial arrangement of the principal cuts induces a compositional structure of the ear (see Fig. 2). The recognition of this part structure may be cast as semantic part-labeling problem, i.e., given a surface $\mathcal{X} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ the objective is to assign discrete part-labels to the vertices in $\mathcal{V}$ or to the faces in $\mathcal{F}$. In this work we develop a semantic part-labeling framework of 3-D meshes of ear

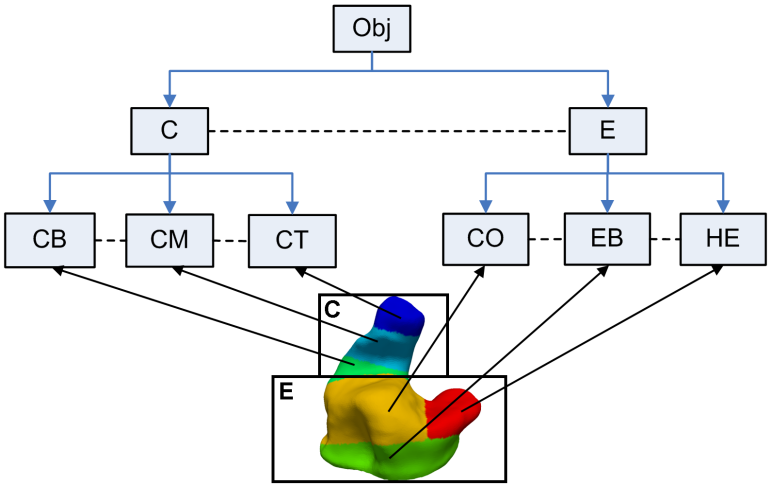


Abbildung 2: Hierarchical structure of the human outer ear. A human outer ear (Obj) is composed of a spindle-shaped *Canal* (C) that sits deep in the outer ear and a base that resides in the *External Ear* (E). The two are separated by a narrow opening called *Aperture*. Three cutting planes, passing through *Aperture*, *First Bend* and *Second Bend*, decompose the *Canal* structure into three constituent parts denoted as *Canal Tip* (CT), *Canal Middle* (CM), and *Canal Base* (CB). A horizontal cut passing through the *External Ear* separates the *Helix* (HE) and the *Conchae* (CO) from the *External Ear Base* part (EB) which forms the base of the surface. *Helix* (HE) and *Conchae* (CE) are divided by a cut passing through the *Crus*. Compositional constraints (dashed) between sibling nodes of the part hierarchy. Colors indicate the anatomical interpretation of the parts.

implants. The labeling represents the part structure of the underlying physical object where both labels and transition boundaries between the segments capture the essence of the underlying HA design process. Consequently, the labels must agree with the semantic interpretation of the object pattern. An object is recognized if its constituent parts have been recognized thereby satisfying the constraints imposed on their spatial arrangement. We hereby assume that the part structure is consistently present among all individuals of the object class under consideration. Requirements and assumptions of each component of the framework are determined by the HA manufacturing application. Key requirements of the proposed framework include to the anatomical variability of the ear, invariance to similarity transformations of an object as well as generalizability to other classes of 3-D meshes.

Overview of approach An overview of the proposed recognition framework is depicted in Fig. 3. We have divided the work according to the main components: section 3 considers the representation of 3-D objects using local shape descriptors, section 4 is concerned with the registration and shape clustering of ear implants, section 5 introduces the semantic labeling model and section 6 addresses the joint shape classification and labeling of 3-D objects.

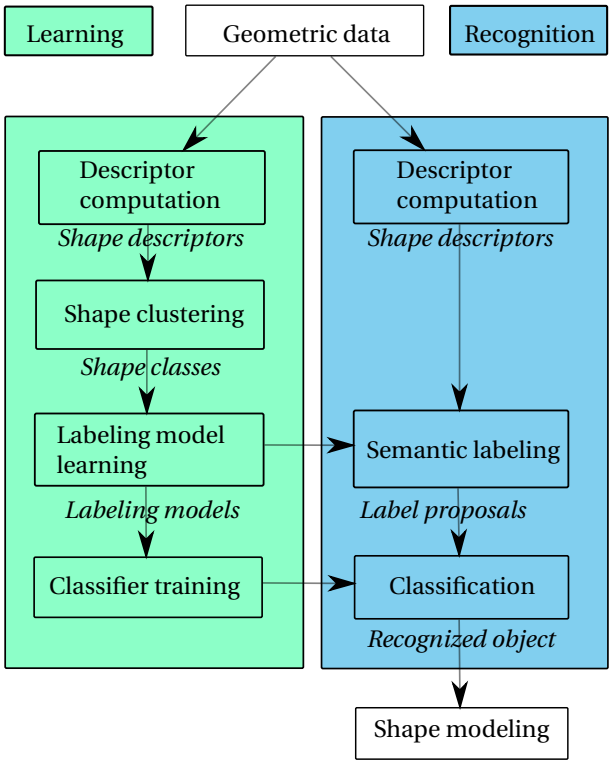


Abbildung 3: 3-D object recognition for shape modeling in HA manufacturing.

3 Representation of 3-D objects using local descriptors

Local 3-D shape descriptors capture the regional or global shape of an object relative to a point on the surface. The choice of descriptor is driven by invariance requirements and the need for robustness to non-ideal conditions, such as noisy 3-D scans. We study several descriptor schemas regarding their discriminative performance for surface registration and semantic labeling. Our performance criterion is the ability of a descriptor to discriminate between different object parts in descriptor space. Many of the considered solutions perform poorly on the ear data. Please note, that this section provides a condensed version of the corresponding chapter in the dissertation.

The investigated descriptor schemas include: curvatures at a point x on a surface (principal curvatures: κ_x^1, κ_x^2, mean curvature: H_x, Gaussian curvature: K_x) [Rus04], global geodesic function (GGF) [HK03], 3-D shape context (SC) [KPNK03], intrinsic shape context (ISC) [STdZ$^+$07], spin image (SPI) [JH99] and the *shape image (SHI)*. The key idea of the shape image descriptor is to capture the shape of a surface in terms of how similar a local descriptor is to other local descriptors in its geodesic neighborhood. This requires a suitable (dis)similarity measure in descriptor space. We define the average value of the distance between the local descriptor at a reference point and a local descriptor residing in the geodesic neighborhood of the reference point as the *shape function* at this point. Larger values of the shape function suggest that the considered region tends to be more salient compared to smaller values. Since real-world objects are typically composed of different structures at different scales it is reasonable to compute the shape function for different neighborhood sizes. There is no way, however, to know a priori what scales are appropriate for capturing the interesting structures. This is why we compute the shape function at different scale levels which we collectively refer to as *shape image*.

In our experiments we compare the discriminative performance of several local descriptors mentioned above. The evaluation was done regarding the ability of the descriptors to find anatomically meaningful correspondences between two surfaces under a bipartite matching model. Ambiguous descriptors, for example, tend to produce matchings for which the semantic labels at corresponding points disagree. In this case the point correspondences may not be anatomically meaningful.

There are three main insights gained from this analysis: (1) on average 3-D shape contexts give rise to the lowest number of mismatches on the ear data; (2) purely local descriptors (e.g. curvatures, global geodesic function) tend to be less discriminative compared to descriptor schemas that represent larger neighborhoods (e.g. spin image, shape image, 3-D shape context). The main reason for this is that organic surfaces usually undergo subtle variations in form of bends. This implies that the geometry barely

varies within smaller neighborhoods; (3) the computational complexity of the considered descriptor schemas increases with their discriminative performance. From this it follows that more complex descriptor schemas (e.g. 3-D shape contexts, shape images) are better suited for off-line recognition tasks. In the following sections we will use 3-D shape contexts as local descriptors of the ear anatomy.

4 Registration and shape clustering of 3-D objects

A capable recognition approach copes with the anatomical variability in order to consistently interpret the ear anatomy across individuals. The nature of the variation of the ear, however, tends to be complex and is difficult to model. In order to capture the anatomical variability of the human outer ear we adopt the idea in [ZBA+09] and partition the surfaces into non-overlapping clusters of objects with a similar shape. The clusters implicitly capture the anatomical variability in a non-parametric way. This is much simpler than defining a parametric model for all possible shapes an object may take but it requires a larger number of training examples to capture the variability reasonably well.

Registration In this paragraph our objective is to transform the ear shapes into a common frame of reference with the ultimate goal of deriving a notion of distance between shapes as a cost function for clustering.

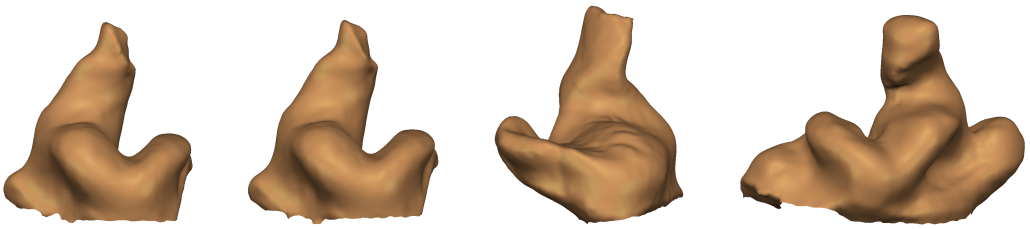
In order to ensure that all given surfaces reside in the same space we solve the correspondence problem jointly across the surfaces to bring them into alignment. In the literature this task is referred to as *groupwise registration* problem. Groupwise registration typically involves iteratively improving the registration between shapes and an evolving average shape. After the registration is complete, correspondences between any point on one shape and any point on another shape are known via the common reference model. If we assume that the surfaces in our data set are independent given a reference object representative of the ear population then the groupwise registration problem reduces to a pairwise registration between each candidate surface and the reference model. We developed and tested three pairwise registration techniques with application to ear shapes: (1) binaural registration using feature correspondences [ZUF+06], (2) registration via correlation of extended Gaussian images [MID06], (3) registration via bipartite shape matching [BMP02].

The binaural registration algorithm and the correlation based alignment yield excellent performance in intra-patient registration. The two methods were successfully integrated into an existing CAD software system. However, the two methods were found insufficient for the inter-patient registration of ear shapes due to the inconsistent occurrence of characteristic surface features across individuals. The bipartite matching method overcomes this limitation by establishing surface correspondence between points using rich local descriptors under a bipartite matching model similar to [BMP02]. The points were sampled from the surfaces at random with a roughly uniform spacing. In general such points will not and need not correspond to characteristic surface features.

Shape distance After transforming the surfaces into a common reference space the *shape distance* between any two surfaces $\mathcal{X}_1, \mathcal{X}_2$ may be estimated via the corresponding points on the reference model $\mathcal{X}_r$. Specifically, let $i, 1 \leq i \leq |\mathcal{P}|$ be a vertex on $\mathcal{X}_1$ and let $j, 1 \leq j \leq |\mathcal{Q}|$ be a vertex on $\mathcal{X}_2$. When the corresponding point of i on $\mathcal{X}_r$ is equal to the corresponding point of j on $\mathcal{X}_r$ then i and j are regarded as corresponding points. A meaningful notion of shape distance, say, $\text{dist}(\mathcal{X}_1, \mathcal{X}_2)$ between $\mathcal{X}_1$ and $\mathcal{X}_2$ derives from the matching cost $d(\xi_i, \xi_j) \geq 0$ via

$$\text{dist}(\mathcal{X}_1, \mathcal{X}_2) = \frac{1}{N} \sum_{i \in \mathcal{V}_1} \sum_{j \in \mathcal{V}_2} \begin{cases} d(\xi_i, \xi_j), \text{if} \, \xi_i, \xi_j \, \text{match on} \, \mathcal{X}_r, \\ 0, \text{otherwise}, \end{cases} \tag{4.1}$$

where in this case N is the number of corresponding points on the reference model. Fig. 4 illustrates the matching performance by showing the two closest surfaces in the data set (a),(b) and the two most distant surfaces (c),(d) in terms of Eqn. (4.1) where we have used 3-D shape contexts as descriptors and the χ^2 measure as distance function between the descriptors. The distance value between the surfaces (a),(b) is 0.025 and the distance value between the surfaces (c),(d) is 0.391. The differences between the former two are barely visible. In fact, the two surfaces seem to represent the same anatomy. Note, that the right surface in the bottom row in Fig. 4 contains a lot of excess material at its base which is not part of the anatomy.



(a) (b) (c) (d)

Abbildung 4: Shape matching and distance. The first two surfaces (a), (b) give rise to the smallest distance in Eqn. (4.1) with a value of 0.025. The second two surfaces (c), (d) yield the largest distance with a value of 0.391. A visual inspection confirms the notion of distance between the shapes.

Despite the presence of mismatches between the point sets we find that the quantity in Eqn. (4.1) supports the visual intuition of shape similarity, i.e., with increasing values the shapes look more similar and more different otherwise.

Shape clustering The nature of the underlying variation of organic shapes tends to be complex and is often unknown. Non-parametric clustering is a common way to explore the structure of a data set with the objective to arrange the objects into groups with strong internal similarities. Like in [ZBA+09] we attempt to cluster the ear data into non-overlapping subsets of objects with a similar shape thereby assuming that the number of clusters is significantly smaller than the number of data examples. The surfaces inside a cluster are expected to look more similar compared to those residing in different clusters.

Non-parametric clustering typically involves defining (1) a measure of (dis)similarity between the data examples, (2) a criterion function for clustering, (3) an algorithm which optimizes the clustering criterion. The choices are largely driven by the underlying application.

We use affinity propagation (AP) [FD07] to cluster the ear shapes while simultaneously identifying object candidates (prototypes) that best represent other cluster members. AP solves a combinatorial optimization problem in which the distance function in Eqn. (4.1) gives rise to a clustering objective referred to as *net similarity* $S(c)$. The vector $c=(c_1,...,c_N)$ denotes a configuration of N hidden labels with $c_i \in \{1,...,N\}$ indicating the prototype (the cluster) to which each of the N data examples $\mathcal{X}_i$ has the highest affinity. Please refer to the dissertation for more details about the clustering objective.

Experiments In this paragraph we use AP to cluster the ear data into distinct groups of surfaces with a similar shape. The data set contains about 500 surfaces. Our main objective here is to examine how similar the shapes inside the clusters are and how well the clusters are separated. Input to AP is a square matrix of pairwise similarities between the surfaces for which we employed the negative pairwise distance in Eqn. (4.1) with $d(\cdot,\cdot)$ denoting the χ^2 matching cost between the corresponding 3-D shape contexts. Fig. 5 plots the value of the clustering objective $S(c^*)$ against the number of clusters

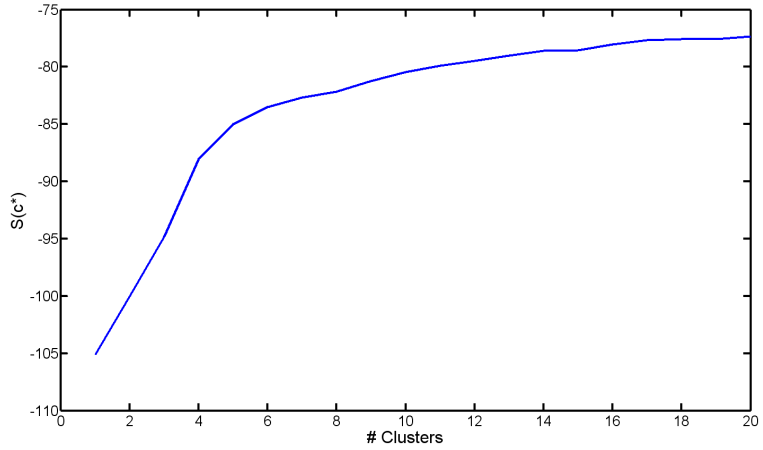


Abbildung 5: Clustering objective $S(c^*)$ (see dissertation for details) plotted against the number of clusters. Note how the value of $S(c^*)$ increases more rapidly to the point of five clusters.

where c^* is a vector in which the elements denote "optimal" assignments of the surfaces to K clusters. From the figure one can see that the value of the objective function increases more rapidly to the point of $K=5$, decreasing much more slowly thereafter. We therefore assume that the data are grouped into $K=5$ reasonably well separated clusters.

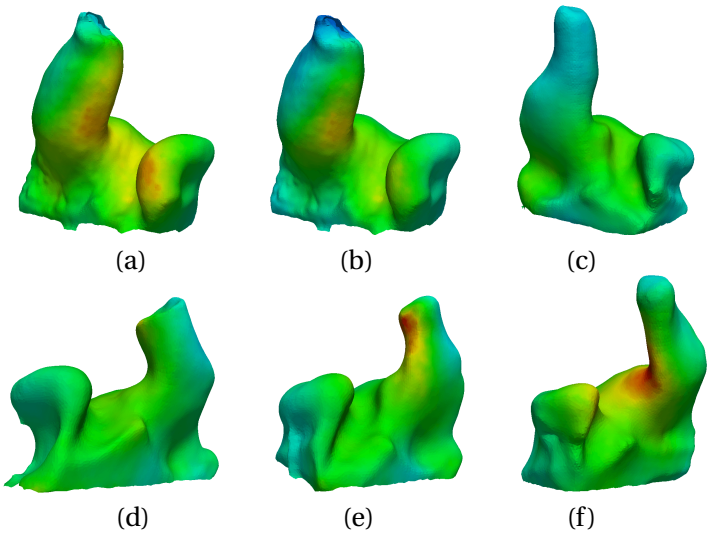


Abbildung 6: Visualization of the intra-cluster variability. Shown are the cluster prototypes for $K=1$ (a) and for $K=5$ (b)–(f). The color map encodes the intra-cluster variability ranging from low (blue) to high (red) variability. The surface regions with a larger variability tend to become more localized when the number of clusters increases from $K=1$ to $K=5$.

Fig. 6 illustrates the intra-cluster variability for $K=1$ (Fig. 6(a)) and for $K=5$ (Fig. 6(b)–(f)). Shown is the cluster prototype together with a color map which encodes the local variation of the shapes within the clusters ranging from zero (blue, small variation) to one (red, high variation). The color values capture the variance of the matching cost between the corresponding points relative to the cluster prototype.

From the color maps one can see that the regions of higher variability (yellow, red) become more localized when the number of clusters K increases from $K=1$ to $K=5$. For $K=1$ we observe the largest variation alongside the canal structure, near the tip of the helix and in the transition area between the two. The largest residual variation for $K=5$ occurs in the aperture region (Fig. 6(f)) and at the tip of the auditory canal (Fig. 6(e)). Notice, that the cluster prototype for $K=1$ shown in Fig. 6(a) is also a prototype for $K=5$ shown in Fig. 6(b). The reduced intra-cluster variability suggests that the shapes inside a cluster tend to look more similar than those residing in different clusters. In section 6 we use the clusters as a non-parametric shape model.

5 Semantic part-labeling of 3-D objects

In this section we consider the problem of sematic part-labeling of 3-D meshes of ear implants. Our contribution 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.

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. 7 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.

The most recent state of the art approaches segment and label 3-D surface meshes jointly. The

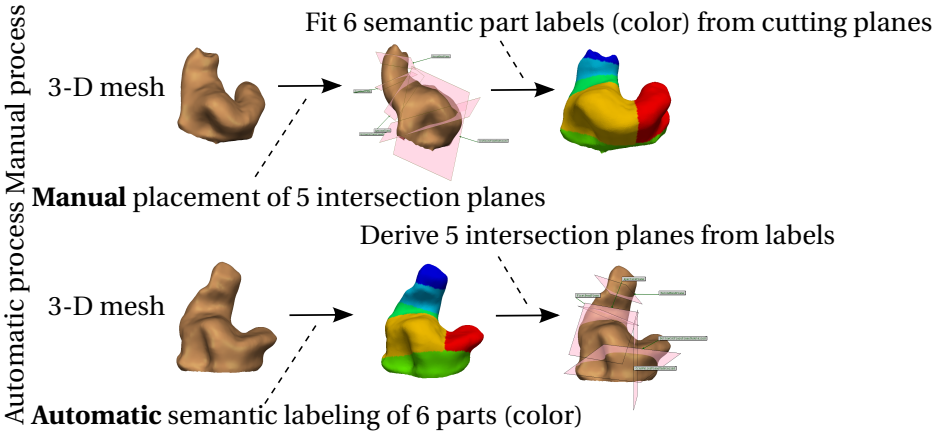


Abbildung 7: Semantic 3-D labeling of ear implants in hearing aid (HA) design. **(Top row)**: A domain expert manually places 5 cutting planes along anatomical lines based on hearing aid design rules. This manual procedure is very cumbersome. Part labels (color) corresponding to 6 anatomical regions 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).

authors in [KHS10], for example, collect statistics of neighboring surface features to learn a CRF with local pairwise interactions. The prior model in [KHS10], 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 [KHS10], tends to be constant (or zero) across the surface.

Not much work exists on the labeling of ear shapes. Similar to [KHS10] the authors in [Zea10] use data dependent pairwise terms to penalize inconsistent labels based on local feature statistics. An alternative approach is presented in [ZSF13] 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 [BMF+10] 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. 7. 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 e.g. [BZF10]). 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. 7. 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.

5.1 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 Fig. 8.

The transition boundaries are induced by 5 cutting planes passing through the mesh. Let $b=(b_1,\ldots,b_B)$ 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

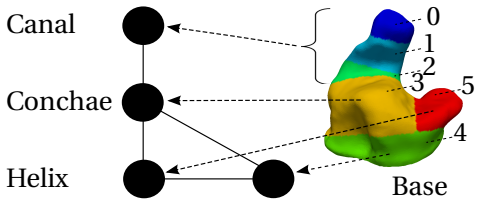


Abbildung 8: Part adjacency graph of the human outer ear (courtesy [Zea10]). Adjacent parts (black circles) in the graph are linked (solid line). The colors indicate the anatomical interpretation (dashed line) of the parts. The ear canal is composed of 3 subparts. Numbers represent anatomical part labels in the label set $\mathcal{L}$.

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:

$$U(h,b,\mathcal{X},\theta)=\sum_{i\in\mathcal{V}}U_i(h_i,\mathcal{X},\theta_1)+\sum_{\{i,j\}\in\mathcal{E}}U_{ij}(h_i,h_j,\theta_2)+\sum_{i\in\mathcal{V}}W_i(h_i,b,\mathcal{X},\theta_3), \tag{5.2}$$

where

$$U_i(h_i,\mathcal{X},\theta_1)=-\theta_1\log(p(h_i|\mathcal{X})), \tag{5.3}$$

$$U_{ij}(h_i,h_j,\theta_2)=\theta_2\delta(h_i,h_j), \tag{5.4}$$

$$W_i(h_i,b,\mathcal{X},\theta_3)=\theta_3\sum_{\ell=1}^{B}\sum_{i\in\mathcal{V}}\max\{0;1-y_\ell(h_i)<b_\ell,z_i>\}. \tag{5.5}$$

Note that the weighting parameters $\theta_1,\theta_2,\theta_3\geq 0$ regularize the influence of the individual terms. The variable z_i in Eqn. (5.5) 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. (5.5) resembles the well known *hinge loss* 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. (5.4) returns the smallest number of links connecting two nodes in the underlying part adjacency graph

$$\mathcal{G}_A=(\mathcal{L},\{\{0,1\},\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{3,5\}\}) \tag{5.6}$$

shown in Fig. 8, 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. (5.4) which also forms a metric over $\mathcal{L}$. Since the part adjacency graph $\mathcal{G}_A$ contains a loop the energy (5.2) is not submodular. For the unary terms in Eqn. (5.3) we use a randomized decision forest and 3-D shape contexts as local descriptors of the vertices $\mathcal{V}$. 3-D shape contexts [KPNK03] are rich, highly discriminative local representations of global shape which we found to work well for our data.

Optimization We use the energy minimization framework to jointly derive an estimate of the labeling h and of the transition boundaries b. Minimizing Eqn. (5.2) with respect to h and b leads to an optimization problem of the form

$$\min_{h,b}U(h,b,\mathcal{X},\theta). \tag{5.7}$$

If we fix h, the problem (5.7) reduces to

$$\min_{b=(b_1,\ldots,b_B)}\sum_{\ell=1}^{B}\sum_{i\in\mathcal{V}}\max\{0;1-y_\ell(h_i)<b_\ell,z_i>\}, \tag{5.8}$$

where we can drop the weighting parameter θ_3 in Eqn. (5.2) since it does not depend on b. Note, that we have 20 parameters, i.e., $b\in\mathbb{R}^{20}$. Since Eqn. (5.8) is convex and is subdifferentiable a subgradient method is well suited to solve the task. Given an estimate of b optimizing the energy (5.2) with respect

to h may be carried out using the expansion move algorithm [BVZ01] since the pairwise terms in Eqn. (5.4) form a metric over the part label set $\mathcal{L}$.

We get a very simple, iterative optimization schema. Given an estimate of h the problem (5.8) gives rise to a global solution since the transition prior in Eqn. (5.5) 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. (5.2) decreases monotonically. Moreover, the iterative optimization of Eqn. (5.7) 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.

Given a label estimate h the optimization problem (5.8) gives rise to a global solution since the transition prior in Eqn. (5.5) is convex. In turn, given an estimate of 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. (5.2) decreases monotonically. About 10 iterations are sufficient on our data.

Learning Given a labeled training set $\mathcal{T} = \{(\mathcal{X}, h, b)\}$ we use a supervised algorithm to learn the model parameters in Eqn. (5.2). The unary terms $U_i(h_i, \mathcal{X}, \theta_1)$ include two types of parameters: (1) the decision forest structure assuming that the forest consists of binary trees and (2) the importance weight θ_1. The parameters θ_2, θ_3 regularize the influence of the energy terms $U_{ij}(h_i, h_j, \theta_2)$ and $W_i(h_i, b, \mathcal{X}, \theta_3)$, respectively. Ideally, we would like to learn all model parameters jointly using a single objective function. However, whereas the weights $\theta_1, \theta_2, \theta_3$ 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 [WS06]. To this end we allow θ to vary over a discrete possibly very large set. In order to keep the training process as simple as possible we follow the suggestion in [CS13] and proceed by growing full trees where each leaf contains only one training example.

5.2 Experiments

For a fair comparison of our method with prior work [BMF+10],[Zea10],[ZSF13] we derive labels from the detected plane features in [BMF+10] as shown in the first row in Fig. 7 whereas for the methods in [Zea10], [ZSF13] 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 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 $\mathcal{A}_m = \{i \in \mathcal{V} | h_i = m\}$:

$$\mathrm{LA}(\mathcal{X}, m) = \frac{2|\hat{\mathcal{A}}_m \cap \mathcal{A}_m|}{|\hat{\mathcal{A}}_m \cup \mathcal{A}_m|}, \tag{5.9}$$

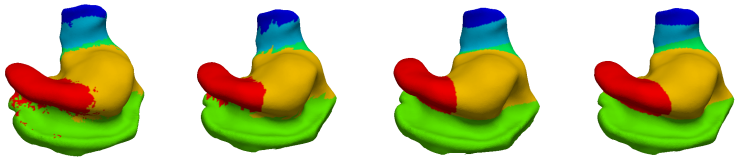
where $\hat{h}_i$ and h_i denote the estimated label and the ground-truth label of the ith vertex of $\mathcal{X}$, respectively. For a surface $\mathcal{X}$ the label accuracy $\mathrm{LA}(\mathcal{X})$ amounts to

$$\mathrm{LA}(\mathcal{X}) = \frac{1}{|\mathcal{V}|} \sum_{m=0}^{|\mathcal{L}|-1} \mathrm{LA}(\mathcal{X}, m). \tag{5.10}$$

The value $\mathrm{LA}(\mathcal{X}) \in [0,6]$ 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.

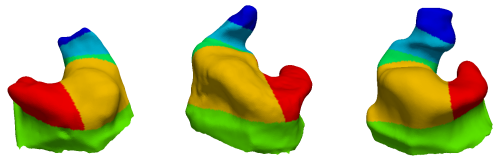
We have a novel data set of 427 human outer ear impressions at our disposal which in turn were laser scanned to reconstruct 3-D triangular surface meshes. A typical 3-D mesh of the ear is composed of roughly 20000 vertices with an average resolution of 0.22 mm. Topologically, a reconstructed outer ear surface constitutes a compact, orientable 2-manifold with boundary. We would like to stress that the manual placement of the cutting planes is very cumbersome and requires excellent understanding of the ear anatomy and of the underlying design process. This is why it is extremely difficult to get a data set of labeled meshes of the ear. Due to the practical significance of this topic we will make our data set publicly available.

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 $\theta_1, \theta_2, \theta_3$ were optimized against one half via cross-validation. The randomized decision forrest was then retrained

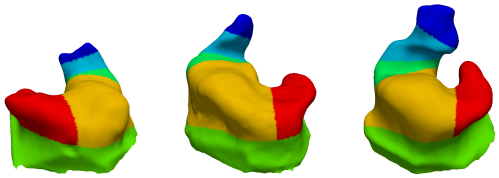


(a) LA($\mathcal{X}$):4.68, (b) LA($\mathcal{X}$):4.71, (c) LA($\mathcal{X}$):4.96, (d) Ground truth

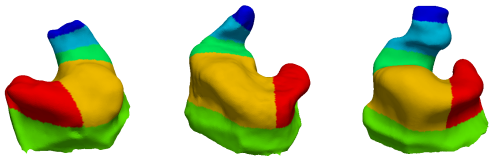
Abbildung 9: Label example using (5.2): (a) without regularization, i.e., $\theta_2,\theta_3=0$, (b) without global parametric transition prior, i.e., $\theta_3=0$. Note, in (a–b) how the segment boundaries differ from the ground truth. In (c) we show the result after least squares fitting of planes 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).



(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

Abbildung 10: Most accurate label examples on our test data using (a): [BMF$^+$10], (b): [Zea10], (c): [ZSF13]. The 2nd row depicts the result using our model (5.2) with global parametric transition prior (d)-(f). In terms of label accuracy LA($\mathcal{X}$) $\in[0,6]$ (6 is best) our model performs slightly better except for (f) when compared with (g)-(i).

using the entire training set. For the parameters we obtain $\theta_1=1,\theta_2=10,\theta_3=20$. Inference was carried out using the algorithm described in paragraph 5.1.

Fig. 9(a) shows a test surface labeled by our model (5.2) without regularization ($\theta_2,\theta_3=0$) and in Fig. 9(b) without global parametric transition prior ($\theta_2=10,\theta_3=0$). A visual comparison of the two results with the ground truth in Fig. 9(d) reveals several inaccuracies despite the overall consistent layout. Note, how the transition boundaries deviate from the ground-truth in Fig. 9(d). The label accuracy according to Eqn. (5.10) is depicted below the surfaces. In Fig. 9(c) we show the result after estimating planes from the label output in Fig. 9(a) via least squares fitting which slightly improves the label accuracy.

Next, in Fig. 10 (first row) we show the best test examples obtained by our competitors in [BMF$^+$10], [Zea10], [ZSF13] together with the result using our model (5.2) with global parametric transition prior, i.e., $\theta_3>0$ (Fig. 10 (second row)). In terms of label accuracy LA($\mathcal{X}$) our algorithm performs slightly better when compared with the ground-truth in Fig. 10 (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 (5.2) with global parametric transition prior performs best. While the methods [Zea10], [ZSF13] perform equally well they were outperformed by [BMF$^+$10]. Also note, that our model (5.2) without global parametric transition prior ($\theta_3=0$) achieves a higher label accuracy than [BMF$^+$10] after least squares fitting of the planes as illustrated in Fig. 9.

6 Joint shape classification and labeling of 3-D objects

Organic shapes such as teeth and the human outer ear are typical examples of object classes with a large variability in shape. It is extremely challenging to automatically interpret (to label) the anatomy

Tabelle 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 (5.2) w/o global parametric transition prior ($\theta_3=0$), our model (5.2), the canonical ear signature (CES) [BMF$^+$10], the layout CRF with spatial ordering constraints in [Zea10], the joint shape classification and labeling (JSCL) model in [ZSF13].

	Our (Eqn. 5.2,$\theta_3=0$)	Our (Eqn. 5.2)	CES [BMF$^+$10]	CRF [Zea10]	JSCL [ZSF13]
$\overline{\mathrm{LA}(\mathcal{X})}$	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

of such objects consistently. Image analysis methods often combine category specific segmentation models with shape prior models to cope with the visual variability and other non-ideal conditions. These combined models tend to be complex resulting in high computational costs of learning and recognition. Furthermore, the use of shape prior models is rather limited to object classes for which the nature of the underlying shape variation is known and relatively simple to model. This, however, is not the case for organic shapes in general and for the human outer ear in particular.

To overcome this difficulty the idea is to train an ensemble of CRFs in which a single CRF models the label distribution of a distinct group of objects with a similar shape. The partitioning of a data set into smaller subsets is performed prior to learning the CRFs, for example, via manual selection by domain experts or via clustering (see section 4). Instead of averaging the label estimations of the models an alternative form of model combination is to select one of the models to label a mesh.

The combination of shape specific labeling models has several advantages. Firstly, each model may be kept simple. No additional shape prior is needed to ensure consistency of the labels across all objects. Secondly, no model assumption about the nature of the shape variation is needed. The shape information is captured in terms of shape classes for which the CRFs are learned. Thirdly, the maximum a posteriori (MAP) energies of the inferred labels may be used for classification. This is a key aspect of our work, since the energy value associated with the optimal labeling can normally not be regarded as a readily useful quantity.

The shape specific labeling problem may also be formulated within the structured support vector machine (SSVM) framework. However, from a practical point of view this may be inconvenient especially when larger data sets are involved. For example, relearning of a SSVM classifier involves all training instances each time a novel observation is added to the training data. Our approach requires a single shape specific labeling model to be relearned together with a few additional classifier parameters. This only involves data members of the shape class to which a novel observation is assigned.

6.1 The model

We consider the following model. Let K denote the number of distinct groups of surfaces with a similar shape and let us denote one such group as the shape class $k \in \{1,\ldots,K\}$. As indicated above the partitioning is done via manual selection or via unsupervised clustering of a representative set of surfaces. The result is a labeled and classified training set of surfaces (see Fig. 6). To simplify the notion we define $y=(h,b)$ where h is the labeling of a surface $\mathcal{X}$ and b represents the transition boundaries between the adjacent segments like in the previous chapter.

The joint probability distribution over elementary events $(\mathcal{X},y,k)$ may be written as

$$p(\mathcal{X},y,k) \propto p(k|X)p(y|X,k). \tag{6.11}$$

The labeling model of the kth shape class gives rise to the conditional probability distribution

$$p(y|\mathcal{X},k) = \frac{\exp(-U(\mathcal{X},y,k))}{Z(\mathcal{X},k)}, \tag{6.12}$$

where $U(\mathcal{X},y,k)$ denotes the energy of y under the kth model and $Z(\mathcal{X},k)$ denotes the observation specific partition function of the kth labeling model. The energy function $U(\mathcal{X},y,k)$ is defined by (5.2). The distribution $p(k|\mathcal{X})$ on the right hand side of Eqn. (6.11) measures the confidence for a surface $\mathcal{X}$ to be a member of class k based on its shape.

We define the joint classification and labeling problem as the task of maximizing Eqn. (6.11) with respect to k and y, i.e.,

$$f(\mathcal{X}) = \arg\max_k \max_y p(k|\mathcal{X})p(y|\mathcal{X},k), \tag{6.13}$$

or equivalently

$$f(\mathcal{X}) = \arg\max_k \max_y [\log p(k|\mathcal{X}) + \log p(y|\mathcal{X},k)] \tag{6.14}$$

$$= \arg\max_k \max_y [\beta(\mathcal{X},k) - U(\mathcal{X},y,k)], \tag{6.15}$$

where

$$\beta(\mathcal{X},k) = \log p(k|\mathcal{X}) - \log Z(\mathcal{X},k). \tag{6.16}$$

Notice, that the two terms in Eqn. (6.16) have similar qualitative properties. When the K shape classes form compact clusters the posterior probability $p(k|\mathcal{X})$ will be peaked, that is, if $\mathcal{X}$ belongs to class k the first term in Eqn. (6.16) assumes a large value and a small value otherwise. Likewise, for a given $\mathcal{X}$ the partition function $Z(\mathcal{X},k)$ assumes a large value if $\mathcal{X}$ belongs to class k and a small value otherwise because, in the former case, there should exist labelings with both high and low energies. It is therefore reasonable to assume that Eqn. (6.16) may be approximated by a sum of two univariate functions, say,

$$\beta(\mathcal{X},k) \approx \beta(\mathcal{X}) + \epsilon_k, \tag{6.17}$$

where $\beta(\mathcal{X})$ depends on $\mathcal{X}$ and ϵ_k depends on k. In general, this may not be true but the assumption in Eqn. (6.17) is weaker than, e.g., to assume the decomposability of $Z(\mathcal{X},k)$. We provide empirical evidence in section 6.3. Eqn. (6.15) then simplifies to

$$f(\mathcal{X}) = \arg\max_k \left[\epsilon_k - \min_y U(\mathcal{X},y,k)\right]. \tag{6.18}$$

To further simplify the notation we define

$$q_k(\mathcal{X}) = -\min_y U(\mathcal{X},y,k), \tag{6.19}$$

and obtain the expression

$$f(\mathcal{X}) = \arg\max_k [q_k(\mathcal{X}) + \epsilon_k]. \tag{6.20}$$

The free parameters of the resulting classifier $f(\mathcal{X})$ comprise the parameters of the K energy functions as in Eqn. (5.2) and the class specific constants $\epsilon = (\epsilon_1,...,\epsilon_K)$.

6.2 Learning and inference

The classifier in Eqn. (6.20) takes a linear form with K discriminant functions $q_k(X) + \epsilon_k$. We now describe the associated learning and recognition tasks.

Learning Given a classified and labeled training set $\mathcal{T}$ we learn the unary and pairwise parameters for each of the K labeling models in Eqn. (5.2) using a supervised algorithm, such as the one described in section 5.1. Approximate MAP inference is then carried out using algorithm described in paragraph 5.1 after which the quantities $q_k(X)$ in Eqn. (6.20) are computed. We now describe the procedure for learning the class specific constants $\epsilon = (\epsilon_1,...,\epsilon_K)$.

For a surface $\mathcal{X} \in \mathcal{T}$ with known class association k the classifier $f(\mathcal{X})$ correctly decides for k if

$$q_k(\mathcal{X}) + \epsilon_k > q_{k'}(\mathcal{X}) + \epsilon_{k'}, \forall k' \neq k. \tag{6.21}$$

Thus, for each $\mathcal{X} \in \mathcal{T}$ there are $K-1$ constraints of the form (6.21) yielding a total of $(K-1)|\mathcal{T}|$ constraints for the training set $\mathcal{T}$. We follow the support vector machine (SVM) approach and minimize the upper bound of the empirical risk with respect to ϵ, i.e.,

$$L(\epsilon) = \sum_{\mathcal{X}} \sum_{k' \neq k} \max\{0, 1 - q_k(\mathcal{X}) - \epsilon_k + q_{k'}(\mathcal{X}) + \epsilon_{k'}\}, \tag{6.22}$$

where $k \leq K$ denotes the true class of a surface $\mathcal{X}$. In the literature Eqn. (6.22) is referred to as the hinge loss function. Since $L(\epsilon)$ is convex a minimizer $\epsilon^* = \arg\min_\epsilon L(\epsilon)$ can be obtained globally. Furthermore, $L(\epsilon)$ is subdifferentiable and can be minimized iteratively using a subgradient method.

A typical subgradient method iterates

$$\epsilon^{(n+1)} = \epsilon^{(n)} - \vartheta_n g^{(n)} \tag{6.23}$$

where $g^{(n)}$ denotes the subgradient of $L(\epsilon^{(n)})$ at $\epsilon^{(n)}$, ϑ_n denotes the *step size* and $n \geq 0$ is the iteration index. The subdifferential of Eqn. (6.22) is given by

$$\partial L(\epsilon) = \sum_i \partial L_i(\epsilon), \tag{6.24}$$

where the sum is over all inequalities in Eqn. (6.21) and all $\mathcal{X}$.

If for the current $\epsilon^{(n)}$ and for the tth example $\mathcal{X}$, $(t \leq |\mathcal{T}|)$ we have $q_k(\mathcal{X}) + \epsilon_k^{(n)} - q_{k'}(\mathcal{X}) - \epsilon_{k'}^{(n)} \leq 1$ then $g_k^{(n)} = -1$ and $g_{k'}^{(n)} = 1$ with $g_k^{(n)}$ and $g_{k'}^{(n)}$ denoting the kth and k'th component of the subgradient of $L_i(\epsilon^{(n)})$. Otherwise the subgradient of $L_i(\epsilon^{(n)})$ is equal to zero.

The quantity ϑ_n is a function of the step size along the direction of the negative subgradient for which many different types of rules exist, e.g.,

$$\vartheta_n \geq 0, \sum_{n=1}^{\infty} \vartheta_n^2 < \infty, \sum_{n=1}^{\infty} \vartheta_n = \infty. \tag{6.25}$$

One typical example is $\vartheta_n = a/(c+n)$, where $a \geq 0$ and $c > 0$. A common choice is $a = c = 1$.

Inference Given a test surface $\mathcal{X}$ recognition is carried out by first running algorithm described in paragraph 5.1 for each of the K labeling models. The energy values returned by the algorithm are then used to compute the quantities $q_k(\mathcal{X})$ after which Eqn. (6.20) is solved. In the next section we provide experimental results on the ear data set.

6.3 Experiments

We have experimented with the joint classification and labeling model using the same data set as in section 5.2. Our main objective here is to see how accurate the model is compared to the label performance of a single model in Eqn. (5.2).

First, the entire data set was partitioned into $K = 5$ clusters where we make use of the result in section 4. The resulting shape classes are depicted in Fig. 6. The next step is to learn the K labeling models as described in section 6.2 using the cluster members except for the test data. Prior to learning the class specific constants ϵ the quantities $q_k(X)$ are computed via Eqn. (6.19). The class specific constants ϵ are then learned by solving $\epsilon^* = \operatorname{argmin}_\epsilon L(\epsilon)$ using a subgradient method. The latter converges quickly after a few iterations.

For a test surface the solver returns the estimated class and the labels. If ϵ is set to zero, i.e., when ϵ_k is removed from Eqn. (6.20) then about 71% of the training data and 32% of the test data were assigned to the correct cluster, i.e., the learned shape class while at the same time the labeling model of this class generated the best quality labeling. On the other hand, for the learned vector $\epsilon^* \neq 0$ a correct class assignment was achieved for 93% of the training data and for 81% of the test data while at the same time the labeling model of the assigned classes performed best.

Fig. 11 illustrates the result of four test candidates. The first column shows the ground truth labels. The second column depicts the label result when a single model defined by (5.2) is used. Note, how the labels deviate from the ground truth. The third column shows the result obtained with the joint classification and labeling model for which we observe a better agreement with the ground truth. The label quality (5.10) is shown below the surfaces where a value of 6 is best.

Fig. 12 depicts the label result of four test examples when the estimated shape classes differ from the learned classes (middle) and when the surfaces are assigned to the learned classes (right). The result shows that the labeling model of the learned class achieves the best agreement with the ground truth. The outcome of the experiment provides empirical evidence for the assumption in Eqn. (6.17).

In table 2 we show the label accuracy of the combined model described in this section together with the label accuracy of the combined model in [ZSF13] where a Potts prior was used for the pairwise terms in Eqn. (5.2) and θ_3 was set to zero. To ensure a fair comparison planes were fit to the labels via least squares estimation prior to computing the accuracy over the test data. Moreover, we add the label accuracy of a single model defined by (5.2) to the table as a reference. From the test statistics one

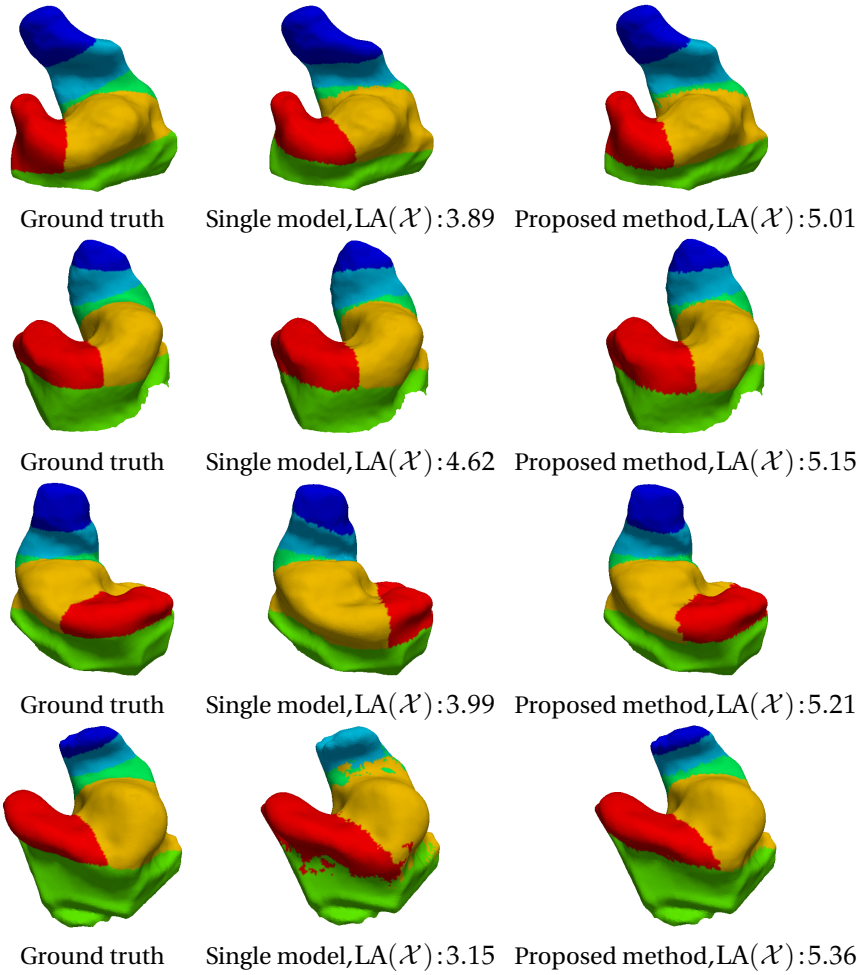


Abbildung 11: Example segmentations using a single labeling model (middle) and the joint shape classification and labeling model (right). The boundary estimation between adjacent segments fails (4th row, middle) when at least one part is missing due to over- and under-segmentation. The label accuracy $\mathrm{LA}(\mathcal{X}) \in [0,6]$ is indicated below the surfaces where a value of 6 is best.

can see that the combined model developed in this work performs better than the combined model in [ZSF13]. Also note, that the model defined by (5.2) outperforms the combined model in [ZSF13]. By comparison, on the original data set, little gain in performance is achieved by the combined model in contrast to using a single model in isolation.

7 Summary and contribution

In this work we give at least five contributions:

- (1) Feature-based surface registration algorithm with application to *binaural processing* in HA manufacturing [ZUF$^+$06]. (2) Global surface registration algorithm with application to automatic processing of HA remakes [ZUSF06]. Both techniques were successfully integrated into an existing HA manufacturing software system.

- The above registration methods assume that the 3-D models were acquired from a single individual. The methods, however, do not generalize well when the registration is carried out across individuals. This issue inspired the development of an inter-patient registration algorithm that works well when the surfaces originate from different patients. The method adapts and combines existing shape matching approaches.

- Non-parametric clustering scheme with application to outer ear surfaces. A notion of (dis)similarity was derived from the intra-patient registration algorithm and was combined with an existing non-parametric clustering schema. The original idea of the approach was published in [ZBA$^+$09].

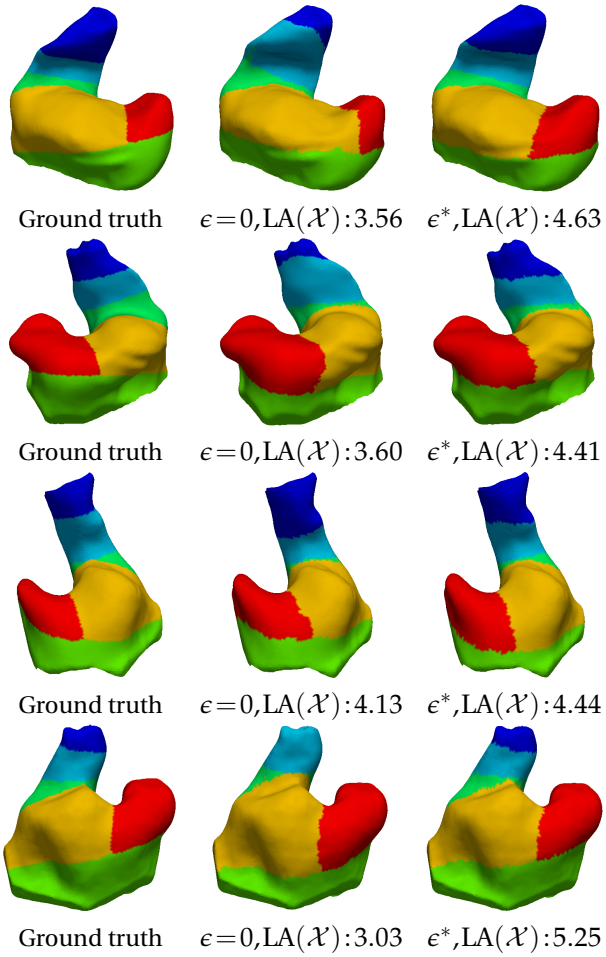


Abbildung 12: Part-labeling using the joint shape classification and labeling model with $\epsilon=0$ (middle) and with the learned $\epsilon^* \neq 0$ (right). The model associated with the learned shape class of a test candidate performs best (right). The boundary estimation between adjacent segments fails when at least one part is missing due to over- and under-segmentation (4th row, middle). The label accuracy LA($\mathcal{X}$) $\in [0,6]$ is indicated below the surfaces where a value of 6 is best.

- Semantic part-labeling model for ear implants using a CRF [Zea10]. We have explored several prior terms to accommodate high level constraints on the part layout imposed by the hearing aid design process. We introduced the concept of a global parametric transition prior which resulted in a boost in performance of the labeling model. To the best of 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 appear at MICCAI 2015*). We collected a large data set of 3-D meshes, with associated ground truth labels, which we will make publicly available.
- Joint shape classification and labeling model [ZSF13]. To better cope with the variability of the ear the method combines multiple CRFs each of which captures the label distribution of a collection of surfaces with a similar shape. For a surface the classifier returns the best model as function of label accuracy.

Literatur

[BMF+10] BALOCH, S., R. MELKISETOGLU, S. FLOERY, S. AZERNIKOV, G. SLABAUGH, A. ZOUHAR und T. FANG: *Automatic Detection of Anatomical Features on 3D Ear Impressions for Canonical Representation.* In Proceedings of MICCAI, 2010.

[BMP02] BELONGIE, S., J. MALIK und J. PUZICHA: *Shape Matching and Object Recognition Using Shape Contexts.* IEEE PAMI, 24(24), 2002.

[BVZ01] BOYKOV, Y., O. VEKSLER und R. ZABIH: *Fast approximate energy minimization via graph cuts.* IEEE Trans. Pattern Analysis, and Machine Intelligence, 2001.

Tabelle 2: Various test statistics for model comparison: $\overline{\mathrm{LA}(\mathcal{X})}$ (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: the combined model in [ZSF13] (Model 1), our combined model (Model 2), our model in Eqn. (5.2) (Model 3).

	Model 1	Model 2	Model 3
$\overline{\mathrm{LA}(\mathcal{X})}$	3.35	4.21	4.10
$\widetilde{\mathrm{LA}(\mathcal{X})}$	3.31	4.20	4.10
$\sigma_{\mathrm{LA}(\mathcal{X})}$	0.60	0.46	0.48

[BZF10] BALOCH, S., A. ZOUHAR und T. FANG: *Deformable registration of organic shapes via surface intrinsic integrals: application to outer ear surfaces.* MICCAI MCV Workshop, 2010.

[CS13] CRIMINISI, A. und J. SHOTTON: *Decision Forests for Computer Vision and Medical Image Analysis.* Springer Publishing Company, Incorporated, 2013.

[FD07] FREY, B. J. und D. DUECK: *Clustering by passing messages between data points.* Science, 315:972–976, 2007.

[HK03] HAMZA, A. BEN und HAMID KRIM: *Geodesic Object Representation and Recognition.* In: *IN PROCEEDINGS OF DGCI, VOLUME LNCS*, Seiten 378–387, 2003.

[JH99] JOHNSON, A. E. und M. HEBERT: *Using spin images for effcient object recognition in cluttered 3d scenes.* IEEE PAMI, 21(5):433–449, 1999.

[KHS10] KALOGERAKIS, E., A. HERTZMANN und K. SINGH: *Learning 3D Mesh Segmentation and Labeling.* SIGGRAPH 2010, 2010.

[KPNK03] KOERTGEN, M., G.-J. PARK, M. NOVOTNI und R. KLEIN: *3D Shape Matching with 3D Shape Contexts.* Proceedings of The 7th Central European Seminar on Computer Graphics, 2003.

[MID06] MAKADIA, A., A. PATTERSON IV und K. DANIILIDIS: *Fully Automatic Registration of 3D Point Clouds.* Proceedings CVPR'06, 2006.

[Rus04] RUSINKIEWICZ, S.: *Estimating Curvatures and Their Derivatives on Triangle Meshes.* In: *Symposium on 3D Data Processing, Visualization, and Transmission*, 2004.

[SBB+09] SICKEL, K., S. BALOCH, V. BUBNIK, R. MELKISETOGLU, S. AZERNIKOV, T. FANG und J. HORNEGGER: *Semi-Automatic Manufacturing of Customized Hearing Aids Using a Feature Driven Rule-based Framework.* In: MAGNOR, MARCUS A., BODO ROSENHAHN und HOLGER THEISEL (Herausgeber): *VMV*, Seiten 305–312. DNB, 2009.

[STdZ+07] SHI, Y., P. M. THOMPSON, G. I. DE ZUBICARAY, S. E. ROSE, Z. TU, I. DINOV und A. W. TOGA: *Direct mapping of hippocampal surfaces with intrinsic shape context.* Neuroimage, 2007.

[WS06] WINN, J. und J. SHOTTON: *The layout consistent random field for recognizing and segmenting partially occluded objects.* CVPR, 2006.

[ZBA+09] ZOUHAR, A., S. BALOCH, S. AZERNIKOV, C. BAHLMANN, G. UNAL, T. FANG und S. FUCHS: *Freeform Shape Clustering for Customized Design Automation.* ICCV 2009 Workshop on 3-D Digital Imaging and Modeling, 2009.

[Zea10] ZOUHAR, A. und ET AL.: *Layout Consistent Segmentation of 3-D meshes via Conditional Random Fields and Spatial Ordering Constraints.* MICCAI, 2010.

[ZSF13] ZOUHAR, A., D. SCHLESINGER und S. FUCHS: *Joint Shape Classification and Labeling of 3-D Objects Using the Energy Minimization Framework.* In: WEICKERT, JOACHIM, MATTHIAS HEIN und BERNT SCHIELE (Herausgeber): *GCPR*, Band 8142 der Reihe *Lecture Notes in Computer Science*, Seiten 71–80. Springer, 2013.

[ZUF+06] ZOUHAR, A., G. UNAL, T. FANG, G. SLABAUGH, H. XIE und F. MCBAGONLURI: *Anatomically-Aware, Automatic, and Fast Registration of 3D Ear Impression Models.* 3DPVT, 2006.

[ZUSF06] ZOUHAR, A., G. UNAL, G. SLABAUGH und T. FANG: *Generalized rigid alignment of 3D ear impression models.* Technical Report 2006P24212US01, Siemens Corporate Research, 2006.