

Layout Consistent Segmentation of 3-D Meshes via Conditional Random Fields and Spatial Ordering Constraints

Alexander Zouhar^{1,2}, Sajjad Baloch¹, Yanghai Tsin¹,
Tong Fang¹, and Siegfried Fuchs²

¹ Siemens Corporate Research, Inc., Princeton, USA

² Dresden University of Technology, Dresden, Germany

Abstract. We address the problem of 3-D Mesh segmentation for categories of objects with known part structure. Part labels are derived from a semantic interpretation of non-overlapping subsurfaces. Our approach models the label distribution using a Conditional Random Field (CRF) that imposes constraints on the relative spatial arrangement of neighboring labels, thereby ensuring semantic consistency. To this end, each label variable is associated with a rich shape descriptor that is intrinsic to the surface. Randomized decision trees and cross validation are employed for learning the model, which is eventually applied using graph cuts. The method is flexible enough for segmenting even geometrically less structured regions and is robust to local and global shape variations.

1 Introduction

Surface segmentation involves the partitioning of a surface mesh into non overlapping sub-meshes, each representing a part of the underlying 3-D object. Such decomposition into parts has traditionally been addressed through metrics derived from geometric properties of a surface, such as curvature, dihedral angles, and protrusion [1,2,3]. Such raw geometric properties are typically local or regional in nature, and completely ignore the compositional arrangement of constituent parts. Consequently, the accuracy of these methods gets challenged in two major ways: (1) previously unseen shape variations may, in general, result in erroneous labeling if the descriptor is not strong enough, (2) neighboring surface regions with similar geometric properties could actually belong to different parts, i.e., they are interpreted differently. These issues may be resolved through explicit constraints on the relative spatial arrangement of neighboring labels.

In this paper, we present an object model-based approach for the segmentation of categories of 3-D objects with known compositional structure. Our method employs a labeling of various parts, which densely covers a 3-D surface mesh. The parts are derived from a semantic interpretation of compact non-overlapping regions on a surface. For instance, typical parts of a car include hood, trunk, roof, and so on.

The label distribution is modeled using a CRF whose undirected graph is isomorphic to the mesh topology. Each label variable is associated with a rich local shape descriptor that is intrinsic to the surface. Prior knowledge about the object structure is incorporated via constraints on the spatial layout of neighboring labels in a mesh graph. For instance, the windshield and the roof are neighbors, but the roof must be detected “above” the windshield, not vice-versa. In short, the proposed layout constraints directly address the limitations of the aforementioned methods by penalizing the assignment of ambiguous and/or spatially incorrect labels.

The capability of the model is demonstrated on the example of human outer ear 3-D surface meshes. Mesh segmentation is highly significant in digital hearing aid design [4]. It serves as a pre-requisite for automatic surface manipulation in order to reduce the amount of human intervention in the design process [5].

1.1 Related Work and Our Contribution

An extensive review of algorithms for 3-D mesh segmentation can be found in [6]. Most of these methods achieve segmentation exclusively via raw geometric information derived from a surface. Instead, our method utilizes a rich descriptor coupled with spatial ordering constraints. There has been little work that exploits explicit constraints on the spatial arrangement of parts. Coupled with our descriptor, they help in resolving over- and under-segmentation even if the parts have boundaries in geometrically less structured regions. [7] used a CRF model for partially occluded objects in 2-D images, which may extend directly to 3-D volumetric data. However, its application to 3-D meshes is more complicated due to the absence of natural (grid) directions on a curved 2-manifold embedded in 3-D space.

2 Layout Consistent Segmentation of Surfaces

Given a surface representation of a 3-D object in the form of a polygonal mesh $X := (V, E)$, where $V = \{v_i\}$ is the set of vertices, and $E = \{e_k\}$ is the set of edges. Our goal is to infer a labeling $h : V \rightarrow L, v_i \mapsto h_i := h(v_i)$, that assigns a label h_i to each vertex $v_i \in V$.

Inference is usually based on some characteristic observations on a mesh X , in the form of local shape descriptors $\{x_i\}$. In this paper, we develop a model that structures the labels $H := \{(v_i, h_i) | v_i \in V\}$ in the form of an undirected graph G . We assume that G is isomorphic to X .

The relationship between neighboring labels is directly associated with the relationship between the corresponding local mesh characteristics. This allows the compositional constraints to be imposed through the joint probability distribution of (X, h) , which in turn may be modeled as a *Markov Random Field* (MRF) [8]. However, this requires that the local mesh characteristics are statistically independent from those of their immediate neighbors. In this paper, we allow a richer shape representation involving a complex neighborhood. Consequently, optimizing the MRF model becomes difficult, if not intractable.

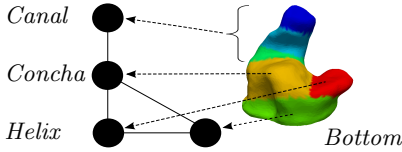


Fig. 1. Part adjacency graph and corresponding mesh representation of a human outer ear. Parts (circles) are linked (solid) according to adjacency of anatomical regions, i.e., *Canal*, *Concha*, *Helix*, and *Bottom*. The colors indicate the anatomical interpretation (dashed) of the parts. The *Canal* is composed of 3 subparts.

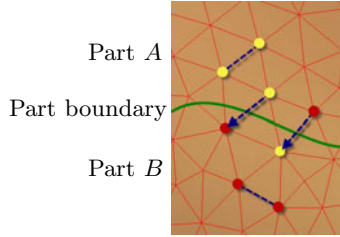


Fig. 2. Neighboring labels on G . Two nodes are connected via a link (dashed). Part membership of the nodes is indicated by the color, i.e., yellow means part *A* and red means part *B*. The labeling along the directed link from red to yellow is incorrect.

The problem can be resolved by the conditional distribution model of CRFs [9]. The joint distribution over elementary events (X, h) may be written as

$$P(X, h; \theta) = P(X; \theta)P(h|X; \theta). \quad (1)$$

The conditional distribution $P(h|X; \theta)$ is sufficient for label estimation, so that the marginal $P(X; \theta)$ can be dropped. Because the model is conditional, complex dependencies among input variables do not need to be explicitly represented which allows the use of rich local shape descriptors. Assuming pairwise statistical dependencies between neighboring labels a model for the conditional distribution $P(h|X; \theta)$ is given by

$$P(h|X; \theta) = \frac{1}{Z(\theta, X)} \prod_{(i,j)} \xi_{ij}(h_i, h_j, X; \theta) \prod_i \nu_i(h_i, X; \theta), \quad (2)$$

where ν_i is a unary potential function that models the posterior distribution of the labels at vertex v_i in a mesh X . ξ_{ij} is the pair-wise potential that constrains the spatial arrangement over the neighboring labels. Note that the partition function Z can be dropped, since we are not interested in actual probabilities. Consequently, Eqn. (1) reduces to an energy functional whose maximization leads to the optimal labeling. A “price” to pay here is that the potential functions tend to be more complex than those in a pair-wise MRF. On the other hand, the dependency on X allows us to exploit the mesh geometry for explicitly enforcing constraints on the spatial layout of neighboring labels. In the following discussion, we develop these layout constraints through a part adjacency graph.

Definition 1 Part Adjacency Graph (PAG). A PAG for a class of objects is an undirected graph with nodes representing constituent object parts. Undirected links between nodes indicate which parts are adjacent to which other parts. The underlying binary relation is symmetric and reflexive.

Fig. 1 depicts a simplified PAG of the human outer ear together with the corresponding anatomical interpretation of the parts. A natural constraint for neighboring labels in a mesh graph G can be derived from its PAG.

Definition 2 Soft Layout Consistency (SLC). *Let G be a graph representing the structure between label variables, and let h_i and h_j , ($i \neq j$) be the labels of two neighboring vertices. Let G_{adj} be the PAG for the class of objects under consideration. The labels h_i and h_j are soft layout consistent if $(h_i, h_j) \in G_{adj}$.*

Definition 2 explicitly ensures local spatial coherence, and implicitly encourages regional compatibility of the labels. However, the constraint is “soft” in the sense that it does not capture the relative spatial arrangement of neighboring parts (e.g., above, below, left, right, back, front). For instance, an inferred labeling on a human body surface may put the neck on top of the head, since this arrangement will still be consistent according to Definition 2. Figure 2 illustrates the main idea on the example of two parts. Note that while moving along the directed link, a change from label $h_i = A$ to a label $h_j = B$ is correct. However, a change from label $h_i = B$ to $h_j = A$ is incorrect.

For two neighboring labels h_i and h_j , $i \neq j$ there are three possible types of transitions:

Part interior: $h_i = h_j$,

Adjacent parts: $(h_i, h_j) \in G_{adj}$, $h_i \neq h_j$. Note the symmetry here.

Inconsistent: $(h_i, h_j) \notin G_{adj}$, $h_i \neq h_j$.

Erroneous label assignments as shown in Fig. 2 should be penalized, but this requires an understanding of the spatial order or arrangement of the neighboring labels. Similar to [7] the value of the pairwise potential varies according to transition type as follows:

$$-\log \xi_{ij}(h_i, h_j, X; \theta) = \begin{cases} 0 & \text{Part interior,} \\ \gamma_1 & \text{Adjacent parts,} \\ \gamma_2 & \text{Inconsistent.} \end{cases} \quad (3)$$

Let $g(h_i, h_j) \in \{0, 1\}$ be denoted as *directional consistency function* that characterizes the consistency of soft-layout consistent labels h_i and h_j in a mesh graph G (assuming $h_i \neq h_j$). We define the behavior of $g(\cdot)$ as follows. With reference to the spatial order defined by the arrow in Fig. 2, if $h_i = A$, $h_j = B$ then $g(A, B) = 1$, and if $h_i = B$, $h_j = A$ then $g(B, A) = 0$. We discuss the choice of the directional consistency function in Section 3.3.

Definition 3 Strict Layout Consistency (StLC). *Let h_i and h_j , ($i \neq j$) be two neighboring labels in a mesh graph G . Furthermore, let G_{adj} be the PAG for the class of objects under consideration. The labels h_i and h_j are strict layout consistent if one of the following conditions holds, (1) $h_i = h_j$, (2) $(h_i, h_j) \in G_{adj} \wedge g(h_i, h_j) = 1$.*

For two neighboring labels h_i and h_j , $i \neq j$ we define transitions to be one of the following types:

Part interior: $h_i = h_j$,

Adjacent parts: $(h_i, h_j) \in G_{adj} \wedge g(h_i, h_j) = 1$,

Inconsistent: $(h_i, h_j) \notin G_{adj} \vee g(h_i, h_j) = 0$.

The consistency function $g(\cdot)$ needs to know the spatial arrangement of the labels h_i and h_j on a mesh X . Hence, for StLC, the pairwise potentials in Eqn. (3) make use of the mesh geometry X (see Section 3.3).

3 Learning and Inference

In this section, we learn the potentials in Eqn. (1) using a supervised algorithm. This requires that all training surfaces are pre-labeled. Once the potentials are learned, the negative logarithm of Eqn. (1) may be efficiently minimized using the α -expansion algorithm [10].

3.1 Local Shape Descriptors

The unary potentials use “local” information of a mesh. As mentioned earlier a mesh X is represented by a set of local shape descriptors $\{x_i\}$. Each local shape descriptor x_i is associated with a vertex v_i . The choice of descriptor is driven by invariance requirements to geometric transformations of a surface and the need for robustness to non-ideal conditions, such as noisy 3-D scans.

We propose the *Geodesic Shape Context* (GSC) which is invariant to rigid transformations and scale. A GSC is obtained by binning the geodesic distances measured from a vertex v_i to all other vertices on a mesh [11].

We now define two types of features at each vertex: (1) the cumulative mean curvature of vertices inside each bin. This yields a function of cumulative mean curvature versus the bin index, (2) a difference between the cumulative curvatures of the neighboring bins. Both features are normalized between 0 and 1. Each vertex is hence represented by two feature vectors, which we collectively refer to as the GSC. Note that the GSC at a vertex v_i has a global support region, i.e., it covers the entire mesh relative to v_i ¹. The GSC leads to a very rich representation of the underlying geometry compared with [12] capturing variations in curvature such as the amount of bending.

3.2 Unary Potentials

Local shape descriptors, including GSCs, normally reside in a high dimensional space which makes learning of a model for the unary potentials challenging due to the amount of required training data. We use randomized binary decision trees for the unary potentials. The procedure for learning a binary decision tree and inference of $\nu_i(h_i, X; \theta)$ are similar to [7]. A set of trained decision trees (a forest) returns a distribution over the labels at a vertex $v_i \in V$.

¹ We call it local shape descriptor only to reflect that it is computed at a location v_i and is not a global representation of the entire shape.

3.3 Pairwise Potentials

The parameters of the pairwise potentials, γ_1, γ_2 are learned using cross validation by a search over a reasonable range of positive values of the parameters.

Spatial ordering of neighboring parts. We characterize the spatial order of neighboring labels h_i, h_j , ($h_i \neq h_j$) in terms of the relative spatial arrangement of neighboring parts on a mesh. First, we define the *geodesic center* of a part on a mesh as the average geodesic distance of all vertices in this part from a pole. To this end, we assume that a consistent pole can be detected on each mesh. The average of this measure across all training surfaces is denoted as *expected geodesic center* of the part.

The notion of ordering is established by noting that a configuration of neighboring labels is more likely when each label is assigned closer to its expected geodesic center. Formally, let q_i, q_j denote the geodesic distances of v_i and v_j from the pole, and let q_A and q_B denote expected geodesic centers of parts A and B . q_A and q_B are pre-computed across training surfaces as mentioned in the previous paragraph. We define the directional consistency function $g(h_i, h_j)$ for neighboring vertices v_i , and v_j as:

$$g(h_i, h_j) = \begin{cases} 1 & |q_i - q_A| + |q_j - q_B| < |q_i - q_B| + |q_j - q_A|, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Note, that the mesh X is required as input of the pairwise potentials in order to evaluate the directional consistency function among neighboring labels.

3.4 Inference

The α -expansion algorithm was used for approximate MAP inference of the labels due to its convergence properties (see [10,13,14] for details).

4 Evaluation

We validated the proposed method on a data set of 216 outer ear impressions, which in turn were laser scanned to reconstruct 3-D triangular mesh representations. The resulting meshes had open boundaries and were composed of roughly 5000 vertices. An expert was asked to manually label the meshes along anatomical lines using a CAD software system. In this way 6 compact regions are obtained as illustrated in Fig. 1. Such anatomical regions play a significant role in the design of personalized hearing aid devices [4]. Note that various boundaries divide geometrically less structured regions into parts, which makes segmentation challenging.

The data set was randomly divided into a training set of 180 meshes and a test set of 36 meshes. First, the unary potentials were learned on the training set. A resolution of 20 bins turned out to be reasonable for the GSCs. Next, the pairwise potential parameters were found via cross-validation against the

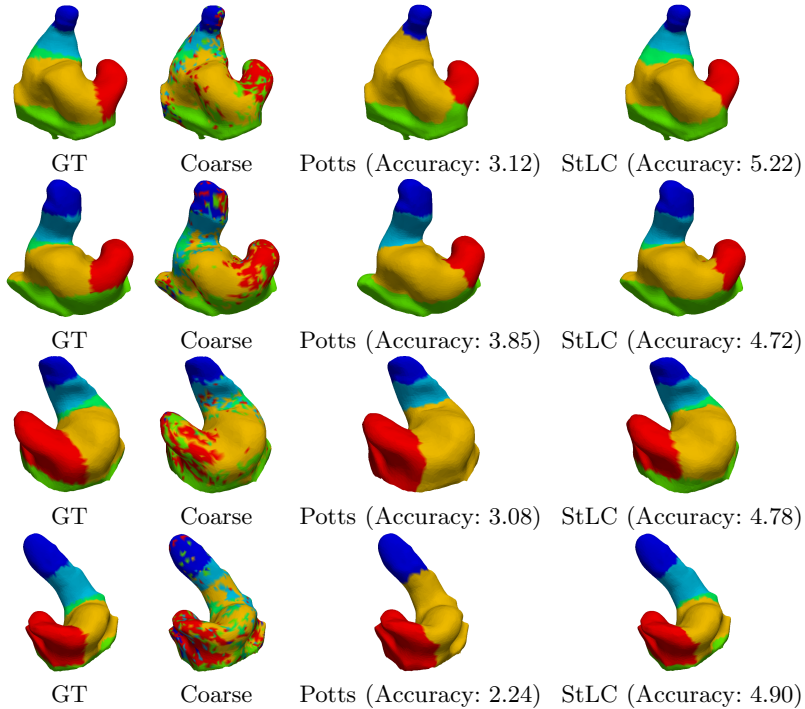


Fig. 3. Example segmentations. “Coarse” means: $\gamma_1 = \gamma_2 = 0$. For comparison the segmentation accuracy is depicted for the Potts model and the StLC model.

training set. The final parameters for StLC were: $\gamma_1 = 1$, $\gamma_2 = 4$. We provide comparison with the Potts model [15] with the cost of dissimilar neighboring labels equal to one and zero otherwise.

Some examples are depicted in Fig. 3. The first column shows the ground-truth (GT). The second column presents the coarse segmentation result, achieved as the MAP estimate from the decision trees. This corresponds to $\gamma_1 = \gamma_2 = 0$, which amounts to no smoothness or layout constraints. The third column represents the Potts model with $\lambda = 1$. Notice how various regions have become smoother. The fourth column represents the StLC model. The results indicate that StLC yields the best agreement with the GT.

Quantitative comparison was carried out as follows. For a label, first the intersection and the union of the estimated region and the corresponding GT is computed. The ratio of area of the former to that of the latter yields a measure of segmentation accuracy per part. The *segmentation accuracy* of a test candidate is defined as the sum of the part scores. The average of this measure across all test individuals using the coarse model was 2.31, for the Potts model 2.97, and for the StLC model 4.52 with 6 being the best score.

5 Conclusions

We have presented an object model-based 3-D mesh segmentation algorithm for 3-D objects with known part structure, and evaluated its performance on a data set of outer ear 3-D meshes with promising results. The method is particularly attractive for the segmentation of organs where the layout of different anatomical regions is already known, such as functional segmentation of the brain.

References

1. Golovinskiy, A., Funkhouser, T.: Randomized cuts for 3D mesh analysis. *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia)* 27 (2008)
2. Shapira, L., et al.: Consistent mesh partitioning and skeletonisation using the shape diameter function. *Visual Computing* 24 (2008)
3. Liu, R., et al.: A part-aware surface metric for shape analysis. *Eurographics* 28 (2009)
4. Slabaugh, G., et al.: 3D shape modeling for hearing aid design. *IEEE Signal Processing Magazine* (2008)
5. Zouhar, A., et al.: Freeform shape clustering for customized design automation. In: *ICCV Workshop on 3-D Digital Imaging and Modeling* (2009)
6. Agathos, A., et al.: 3D mesh segmentation methodologies for CAD applications. *Computer-Aided Design and Applications* 4 (2007)
7. Winn, J., Shotton, J.: The layout consistent random field for recognizing and segmenting partially occluded objects. In: *CVPR* (2006)
8. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *PAMI* 6 (1984)
9. Lafferty, J., et al.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *ICML* (2001)
10. Boykov, Y., et al.: Efficient approximate energy minimization via graph cuts. *PAMI* (2001)
11. Surazhsky, V., et al.: Fast exact and approximate geodesics on meshes. *SIGGRAPH* (2005)
12. Shi, Y., et al.: Direct mapping of hippocampal surfaces with intrinsic shape context. *Neuroimage* (2007)
13. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? *PAMI* (2002)
14. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *PAMI* (2004)
15. Potts, R.: Some generalized order-disorder transformation. In: *Proceedings of the Cambridge Philosophical Society* (1952)