

Semi-supervised Mesh Segmentation and Labeling

Jiajun Lv, Xinlei Chen, Jin Huang[†], Hujun Bao

State Key Lab of CAD & CG, Zhejiang University

Abstract

Recently, approaches have been put forward that focus on the recognition of mesh semantic meanings. These methods usually need prior knowledge learned from training dataset, but when the size of the training dataset is small, or the meshes are too complex, the segmentation performance will be greatly effected. This paper introduces an approach to the semantic mesh segmentation and labeling which incorporates knowledge imparted by both segmented, labeled meshes, and unsegmented, unlabeled meshes. A Conditional Random Fields (CRF) based objective function measuring the consistency of labels and faces, labels of neighbouring faces is proposed. To implant the information from the unlabeled meshes, we add an unlabeled conditional entropy into the objective function. With the entropy, the objective function is not convex and hard to optimize, so we modify the Virtual Evidence Boosting (VEB) to solve the semi-supervised problem efficiently. Our approach yields better results than those methods which only use limited labeled meshes, especially when many unlabeled meshes exist. The approach reduces the overall system cost as well as the human labelling cost required during training. We also show that combining knowledge from labeled and unlabeled meshes outperforms using either type of meshes alone.

Categories and Subject Descriptors (according to ACM CCS): I.3.5 [Computer Graphics]: Computer Graphics—Computational Geometry and Object Modeling

1. Introduction

Consistent semantic segmentation and labelling of multiple 3D shapes is fundamental to shape understanding and processing. Many tasks in shape processing, 3D modeling, animation and texturing of 3D meshes benefit from automatic segmentation of shapes into components that appear natural [CGF09, CKGK11].

Classical approaches to mesh segmentation base their computation purely on geometric structure of individual shapes, in particular geometric boundaries, of the meshes. A variety of geometric features have been investigated, but no single feature or collection of features is known to produce high quality results for all classes of shapes [CGF09]. The underlying difficulty is that a perceptually natural segmentation of a shape is often the result of prior familiarity with other similar shapes and their function. The surface geometry of an individual shape may lack sufficient cues to identify all parts that would be perceived as meaningful to a human observer [VKTS*11].

[†] Corresponding author: hj@cad.zju.edu.cn

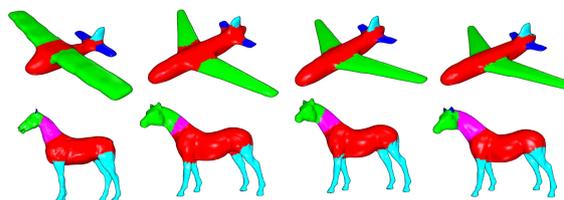


Figure 1: Illustrates with more unlabeled meshes, the segmentation results will improve. The first column is one of 3 labeled training meshes, the second is the segmentation results with no unlabeled meshes, the third is with 7 unlabeled meshes and the last is with 17 unlabeled meshes.

Recently, data-driven semantic segmentation approaches and unsupervised co-segmentation approaches are proposed which can segment and label parts in various 3D meshes and the results are comparable to human work. But the data-driven methods need prior knowledge learned from training dataset and the size of the training dataset has a great impact on the segmentation performance [KHS10]. Even though the unsupervised joint segmentation approaches do not need

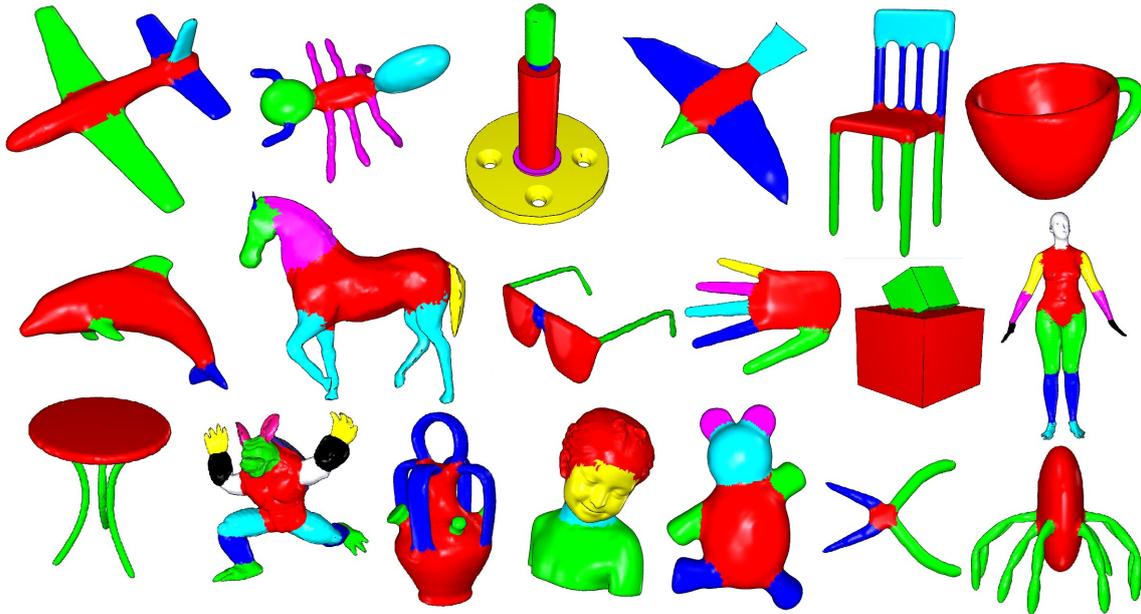


Figure 2: The segmentation and labeling results of our semi-supervised mesh segmentation algorithm on the whole Princeton Segmentation Benchmark [CGF09]. The training process is performed using other 3 labeled meshes with all the unlabeled meshes in the same set.

training data, their results are, not surprisingly, inferior to the supervised ones.

In complex segmentation tasks, obtaining a large amount of labeled data for training is not practical. Yet, large unlabeled datasets are often easy to obtain, making semi-supervised learning methods appealing. Besides, considering the great complexity and variance of meshes, even hundreds of labeled examples can only sparsely cover the parameter space, which makes semi-supervised learning even more attractive [Zhu07, YNX*12]. Additionally, the human labeled training datasets are also likely to be inconsistent or mislabeled and for the inconsistently labeled training datasets, where semi-supervised method is more robust than supervised method [GB04]. Based on these observations, we present a semi-supervised mesh segmentation approach that is not only accurate, robust but also scalable, efficient, and easy to train and deploy. In addition to be computationally efficient, our proposed method reduces the amount of labeling required during training, which makes it appealing for use in the segmentation and labeling of 3D meshes.

We evaluate the proposed approach on the complete Princeton shape benchmark [CGF09] and compare it to the supervised approach of [KHS10]. Our results demonstrate that making use of unlabeled information produces better segmentation than supervised approaches. Figure 1 illustrates the power of incorporating more unlabeled meshes. The contributions of our work are:

- Semi-supervised mesh segmentation model using Condi-

tional Random Fields (CRF) [LMP01, MC07] - a model uses unlabeled mesh information, considering mesh face features and mesh topology.

- Efficient training method for the semi-supervised mesh segmentation model using Virtual Evidence Boosting (VEB) [Obe07, LCFK07] - a method combines logit-boost [FHT00] and belief propagation [MWJ99].
- Experimental results demonstrate the advantages of our model, which outperforms other mesh segmentation models and is robust to mislabeling of training data.

2. Related Work

As an important research topic, shape segmentation has attracted much attention in years. We only review the most related work here, and a comprehensive survey can be found in [Sha08]. A large number of approaches have been developed for decomposing a single shape into parts [Sha08, GF08]. Such approaches suffer when the individual shape does not provide sufficient geometric cues to distinguish its meaningful parts, or when strong extraneous geometric features are present. In a recent evaluation, no segmentation algorithm performed well across all tested datasets [CGF09].

To overcome the limitations inherent in single-shape, limited features analysis, researchers have turned to data-driven techniques that utilize large feature information from multiple shapes in order to segment a given shape. Data-driven techniques are well-established in image processing: top-performing image segmentation algorithms all utilize man-

ually segmented training sets [SWRC09]. This supervised approach to segmentation has recently been extended to 3D shapes by [KHS10], who demonstrated significant improvement over previous segmentation algorithms. [BLVD11] further introduces this learning approach into identifying boundary edges. However, supervised approaches require a substantial number of manually segmented training shapes. When the number of training shapes decreases, their performance drops dramatically.

Without enough labeled meshes, unsupervised approaches are proposed such as [HKG11], in which linear programming is used to jointly segment the meshes. Oana Sidi has another unsupervised segmentation approaches which use descriptor-space spectral clustering [SvKK*11]. However, without prior knowledge of label information, their segmentation results are naturally inferior to the supervised one, especially when the meshes are complicated.

Our approach uses information from both labeled meshes and unlabeled ones, overcoming the difficulties of requiring a large amount of labeled meshes and the disadvantage of inability to use unlabeled meshes. The method generates better results than both supervised and unsupervised method, and is more robust to inconsistently labeled training set which may result from human mislabeling or label missing.

3. Overview

In this section, we present an overview of our semi-supervised approach, full details of the algorithmic components are covered in subsequent sections. We begin by first give a formal definition of the problem:

Definition 1 *Semi-supervised mesh segmentation problem is to learn a mesh segmentation model from a collection of labeled meshes D^l and unlabeled meshes D^u with unary features X and pairwise features ω , which can be formulated into the optimization framework:*

$$\max_{\Theta} f(D^l, D^u, X, \omega; \Theta), \quad (1)$$

where $f(\cdot)$ is some objective function, and Θ is the model parameters.

Per-mesh CRF model. First, much recent work [KHS10, VKTS*11] shows that the objective function $f(\cdot)$ of a good mesh segmentation approach should contain two parts, with one measuring the consistency between the label and its face, and another measuring the consistency between the label and its neighborhood. Therefore, Conditional Random Fields (CRF) is well fit to fulfill this job.

Incorporation of unlabeled mesh information. To incorporate the unlabeled data information into the mesh segmentation problem, we add an entropy term to the original objective function [JWL*06]. As a result, the putative labeling for the unlabeled data can in turn reinforce with the

supervised labels, making the results more accurate and robust.

Learning methods. Training complex CRF models with many features is slow, and exact inference is often intractable. In mesh segmentation, we face thousands of faces, and each is represented by hundreds of features. Additionally, the added entropy term makes the objective function non-convex and thus more hard to optimize. To efficiently solve the problem, we adopted the Virtual Evidence Boosting (VEB) [LCFK07] to train, which performs feature selection and parameter estimation in a unified and efficient manner.

4. Semi-supervised CRF Model Description

This section describes the semi-supervised CRF model we used. First, we would introduce the per-mesh CRF model, then adding an entropy term to incorporate the unlabeled mesh information into the model.

4.1. Per-Mesh CRF Model

Conditional Random Fields [LMP01, MC07] is a kind of discriminative undirected probabilistic graphical model widely used for segmenting and labeling data based on conditional approach. It has some special advantages. Both the unary term $F(X, Y)$ and the pairwise term $G(Y, \omega)$ in CRF are learned from the input data rather than predefined, where X , Y and ω are the features, labelings and pairwise features of the whole mesh.

A per-mesh CRF model was proposed by [KHS10]. However, their model cannot directly lead to an optimization problem in the semi-supervised framework. To overcome this, we introduce another per-mesh CRF model whose whole objective function is learned from function templates and can be optimized in a semi-supervised manner. Our model also captures the information from both the features of a face, and from its neighborhood. Our goal is to label each mesh face $s \in S$ with a label $y_s \in J$, where J is a predefined set of possible labels, such as “body”, “head”, or “wing”. The unary features x_s for each face s , is the same as [KHS10], including curvature features, PCA features, shape diameter, distance from medial surface, average geodesic distance, shape context and so on. Pairwise features ω used in our approach are the dihedral angles of face pairs (the smaller the angle, the more likely the labels of the face pair have little relationship), and potentially more topological features can be included. We compute all mesh labels by maximizing the conditional probability function:

$$P(Y|X, \omega) = \frac{\exp(F(X, Y) + G(Y, \omega))}{Z(X, Y, \omega)}, \quad (2)$$

where $F(X, Y) = \sum_{f_m} \sum_{s=1}^n f_m(x_s, y_s)$, $G(Y, \omega) = \sum_{g_n} \sum_{s_1, s_2} g_n(y_{s_1}, y_{s_2})$, s_1 and s_2 are neighbouring faces and $Z(X, Y, \omega)$ is a normalization factor.

Unary Energy Term. The unary energy term $F(X, Y) = \sum_{f_m} \sum_{s=1}^n f_m(x_s, y_s)$ measures the consistency between the features and the labels. In each iteration, parameters of template functions f_ψ are learned by virtual evidence boosting and the estimated function is added to $F(X, Y)$.

Pairwise Energy Term. The pairwise energy term $G(Y, \omega) = \sum_{g_n} \sum_{s_1, s_2} g_n(y_{s_1}, y_{s_2})$ measures the consistencies of the labels in certain geometry circumstances, which means evaluating the cost of label combinations that the neighboring vertex takes. The main role of the pairwise term is to improve boundaries between segments and to prevent incompatible segments from being adjacent.

4.2. Semi-supervised Objective Function

To make full use of the available training data, we propose a semi-supervised mesh segmentation approach that exploits the data information in the form of entropy regularization.

Assume we have a set of labeled meshes $D^l = \{(X^{(1)}, Y^{(1)}), \dots, (X^{(N)}, Y^{(N)})\}$, and unlabeled meshes $D^u = \{(X^{(N+1)}, Y^{(N+1)}), \dots, (X^{(M)}, Y^{(M)})\}$, where $X^{(i)}$ is the face features of mesh i , $Y^{(i)}$ is the face labels of mesh i . Specifically, for a semi-supervised CRF, we maximize the objective function:

$$f(\cdot) = \sum_{i=1}^N \log P(Y^{(i)} | X^{(i)}) + \gamma \sum_{i=N+1}^M \sum_Y P(Y | X^{(i)}) \log P(Y | X^{(i)}) \quad (3)$$

The first term is the penalized log conditional likelihood of the labeled meshes of the CRF model, and the second term is the negative conditional entropy of the CRF on the unlabeled meshes, \sum_Y means the sum over all combinations of mesh face labels, which is intolerable in mesh segmentation. Here, γ is a tradeoff parameter that controls the influence of the unlabeled data, in our experiments, an $\gamma \in [0.01, 0.05]$ yields desired results.

The motivation is that minimizing conditional entropy over unlabeled data encourages the algorithm to find putative labelings for the unlabeled data that are mutually reinforcing with the supervised labels; that is, greater certainty on the putative labelings coincides with greater conditional likelihood on the supervised labels, and vice versa [JWL*06]. More motivations can be found in [GB04].

5. Learning Parameters

We slightly modified the Virtual Evidence Boosting (VEB) [LCFK07] to train CRF. It runs two procedures iteratively. The first is Belief Propagation (BP) [MWJ99], which transfers information between labels to get a unified belief for the marginal distribution of a set of variables in the graph. Then second part is the normal Logitboost [FHT00]. Since after BP every node shares the same information, the information propagated from the neighborhood is inexplicitly included.

From a learning perspective, this approach includes three steps. First, knowledge is learned from a set of both labeled and unlabeled meshes in the form of per-label classifiers and pairwise label relationship. Next, given a query mesh to be segmented and labeled, we apply the classifiers to assign a probabilistic label to each mesh face. Finally, the probabilistic labels are used synergistically with label pairwise relationship in belief propagation to yield the resulting mesh segments.

5.1. Belief Propagation

The VEB considers two types of evidences for a node y_s . The first type is the feature x_s in the training data. The second type is the message from neighboring nodes $n(y_s)$; the message is obtained by running belief propagation with the current functions $F(X, Y), G(Y, \omega)$.

The likelihood of neighbouring faces taking different labels increases when the angle of the face normals is large, so we modify BP to incorporate this information in the pairwise term in the form of dividing $G(y_{s_1}, y_{s_2})$ by $\rho(2 - \cos \theta)$, where θ is the angle of neighbour face normals and we use $\rho \in [1, 2]$. In BP, messages are distinguished before and after multiplying the compatibility potentials $e^{\frac{G(y_{s_2}, y_{s_1})}{\rho(2 - \cos \theta)}}$. We denote these messages as $\lambda_{s_2 \rightarrow s_1}(y_{s_2})$ and $\mu_{s_2 \rightarrow s_1}(y_{s_1})$, respectively. These messages are distributed normally at first and computed iteratively [Pea88].

$$\lambda_{s_2 \rightarrow s_1}(y_{s_2}) = \alpha e^{F(y_{s_2}, x_{s_2})} \prod_{s_3 \in n(y_{s_2}), s_3 \neq s_1} \mu_{s_3 \rightarrow s_2}(y_{s_2}), \quad (4)$$

and

$$\mu_{s_2 \rightarrow s_1}(y_{s_1}) = \beta \sum_{y_{s_2}} e^{\frac{G(y_{s_2}, y_{s_1})}{\rho(2 - \cos \theta)}} \lambda_{s_2 \rightarrow s_1}(y_{s_2}), \quad (5)$$

where α and β are used for normalization. As can be seen, λ -messages contain information about the distribution of the sending node y_{s_2} , and μ -messages contain information about which values the recipient node y_{s_1} should prefer. So the μ -messages correspond exactly to the messages sent in regular belief propagation.

Equivalently, from belief propagation we have

$$p(y_{s_1} | X) = \eta e^{F(y_{s_1}, x_{s_1})} \prod_{s_2 \in n(y_{s_1})} \mu_{s_2 \rightarrow s_1}(y_{s_1}), \quad (6)$$

where η is used for normalization.

5.2. LogitBoost

LogitBoost is a boosting algorithm that has many appealing properties: it performs automatic feature selection and can handle large numbers of input features for multiclass classification, it has a fast sequential learning algorithm, and it is suitable for combination with other parts in the model.

This section introduces the LogitBoost we modified to incorporate the unlabeled meshes. For the unary energy term, there are J functions F_j for each label, and the probability of vertex x_s taking label $j \in J$ is computed as $p_j(x_s) = \frac{e^{F_j(x_s)}}{\sum_{k=1}^J e^{F_k(x_s)}}$. For the pairwise term, there are also J functions G_j , and if vertex x_s takes label j , the consistency to the neighbouring label j^* is computed as $G_j(j^*) = \sum_{k=1}^J \alpha_{j,k} \theta(k = j^*)$.

Before the update, the first derivation ∇_s and second derivation W_s for each face s of each label should be computed. Then, the boosting weights, w_s , and working responses, z_s , are calculated as $w_s = -W_s$, $d_s = \nabla_s$, $z_s = \frac{\nabla_s}{w_s}$. For the labeled faces, $w_s = p_j(x_s)(1 - p_j(x_s))$, $d_s = 1_{\langle y_s \rangle^* = j} - p_j(x_s)$, where $p_j(x_s)$ is the probability of face s taking label j after belief propagation. For the case of unlabeled data, we derived the expressions for d_s and w_s as follows:

$$d_s = \alpha P_j(x_s)(\log P_j(x_s) - \sum_j P(j|x_s, y_s) \log P(j|x_s, y_s)) \quad (7)$$

$$w_s = \alpha P_j(x_s)(1 - \log P_j(x_s) + (1 - 2 \log P_j(x_s)) (\sum_j P(j|x_s, y_s) \log P(j|x_s, y_s) - \log P_j(x_s))) \quad (8)$$

The updating algorithm is described below [Obe07].

The algorithm description of unary energy term update is in Algorithm 1. Unary Energy Term Update Algorithm training data, likelihood from last belief propagation update function f_m , $gain_c$ initialize $gain_c = 0$ F_j , $j \in J$ node index $s \in S$ compute d_s , w_s , z_s compute best update function $f_{m,j}$ accumulate likelihood gain $gain_c \leftarrow gain_c + \sum_{s, d_s \neq 0} \frac{d_s^2}{w_s} - w_s(f_{m,j}(x_s) - \frac{d_s}{w_s})^2$

For the unary energy term $f_{m,j}$, we add the function from the function template

$$f_{m,j}(x^k) = \theta_1 \delta(x^k \geq h) + \theta_2 \delta(x^k < h) \quad (9)$$

to F_j iteratively, where x^k means the k -th feature.

The best parameter satisfies [LCFK07]:

$$\theta_1 = \frac{\sum_{s=1}^N w_s z_s \delta(x_s^k \geq h)}{\sum_{s=1}^N w_s \delta(x_s^k \geq h)} \quad (10)$$

$$\theta_2 = \frac{\sum_{s=1}^N w_s z_s \delta(x_s^k < h)}{\sum_{s=1}^N w_s \delta(x_s^k < h)} \quad (11)$$

The algorithm description of pairwise energy term update is in Algorithm 2. Pairwise Energy Term Update Algorithm training data, belief propagation message update function g_n , $gain_c$ initialize $gain_c = 0$ G_j , $j \in J$ node index $s \in S$ compute d_s , w_s , z_s label $j^* \in J$ compute best update function $g_{n,j}(j^*)$ accumulate likelihood gain $gain_c \leftarrow gain_c + \sum_{s, d_s \neq 0} \frac{d_s^2}{w_s} - w_s(g_{n,j}(j^*) - \frac{d_s}{w_s})^2$

For the pairwise term $g_{n,j}(j^*)$, the function templated is:

$$G_j(j^*) = \sum_{k=1}^J \alpha_{j,k} \theta(k = j^*). \quad (12)$$

The parameters of the update function can be calculated as

$$\alpha_{j,j^*} = \frac{\sum_{s=1}^n w_s z_s \lambda_{s^* \rightarrow s}(y_{s^*} = j^*)}{\sum_{s=1}^n w_s \lambda_{s^* \rightarrow s}(y_{s^*} = j^*)} \quad (13)$$

In each iteration, we pick a function from f_m and g_n with the most likelihood gain and add it to $F(X, Y)$ or $G(Y, \omega)$.

6. Experimental Results

In this section, we evaluate our semi-supervised segmentation approach and present qualitative and quantitative results. We make comparisons to both the supervised method of [KHS10] and the baseline supervised method of our semi-supervised approach. The robustness of our method to inconsistently labeled training meshes, such as human mislabeling and label missing, is also illustrated. Finally, we analyze the complexity of our algorithm.

Dataset	Sup [KHS10]	Sup	Semi(7)	Semi(17)
Airplane	91.2	91.1	94.2	95.9
Ant	97.4	93.8	97.1	98.7
Armadiillo	83.7	84.1	86.6	87.3
Bearing	61.3	76.2	79.2	82.4
Bird	76.3	81.6	82.1	77.5
Bust	52.2	55.6	58.2	63.4
Chair	97.1	97.6	97.9	98.3
Cup	96.3	91.4	93.8	95.3
Fish	94.1	92.3	94.8	95.7
FourLeg	82.0	83.8	87.2	88.1
Glasses	94.4	93.5	96.3	97.1
Hand	74.9	91.0	95.4	94.7
Human	83.2	84.1	86.7	87.3
Mech	82.4	81.2	81.9	85.6
Octopus	98.3	98.8	98.6	98.7
Plier	92.2	82.1	94.7	94.2
Table	99.0	98.3	93.2	89.1
Teddy	93.1	79.1	92.1	97.9
Vase	74.3	77.4	79.8	81.2
Average	85.4	86.0	88.9	89.9

Table 1: The first method is the supervised approach in [KHS10], The second is our supervised method, the third is our semi-supervised method with 7 unlabeled meshes and the last is our semi-supervised method with 17 unlabeled meshes(3 labeled meshes)

6.1. Segmenting Princeton Segmentation Benchmark

We applied our semi-supervised approach to the same benchmark of [KHS10], processed Princeton Segmentation Benchmark [CGF09]. In this section, all the results shown are done with 3 labeled training meshes (the first 3 meshes

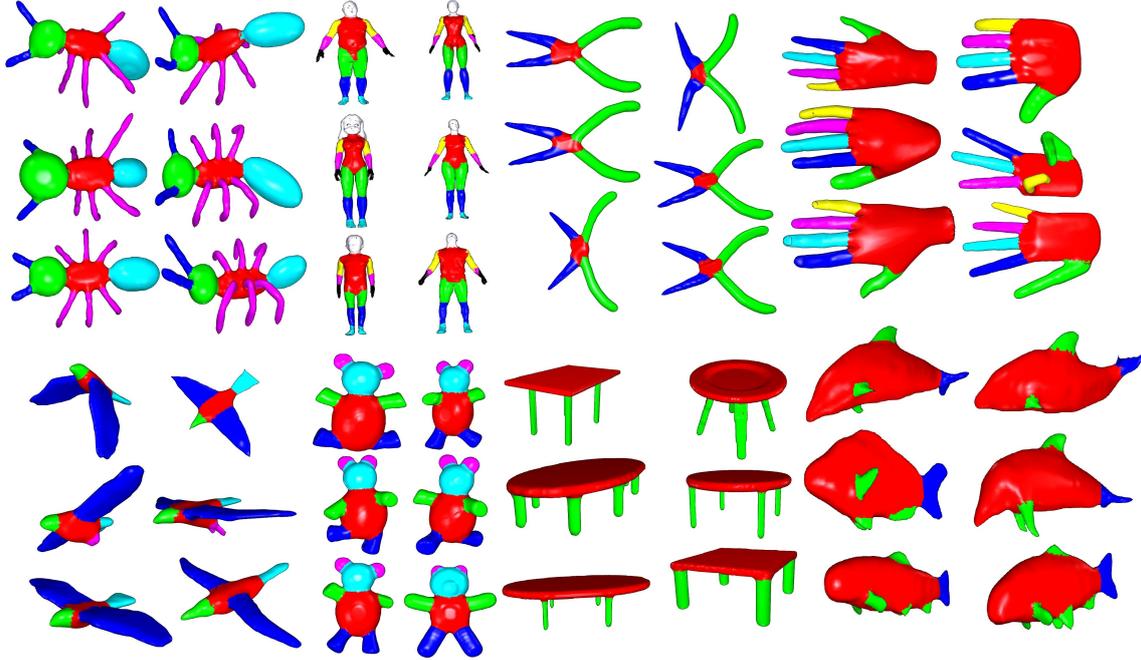


Figure 3: Experimental results of the semi-supervised mesh segmentation method. For each kind of dataset, the left column three are the labeled training dataset, and the right column three are the segmented meshes.

or the 3 meshes with the highest segment accuracy under the premise that all labels are included). Figure 2 contains the results of all datasets in the benchmark and Figure 3 contains the training sets and the corresponding segment results.

To quantitatively analysis the results, we use accuracy to measure the resulting performance, as [KHS10] and [HKG11] showed that different measuring criteria yields consistent results. The accuracy is calculated as the percentage of the mesh surface area that is correctly labeled: $Acc = \frac{(\sum_i a_i \frac{I(y_i, y_i^*) + 1}{2})}{\sum_i a_i}$, where a_i is the area of face i , y_i is the ground-truth label for face i , $y_i^* = \arg \max P(y|x_i)$, $P(y|x_i)$ is the last output of the belief propagation for face i . $I(y, y^*) = 1$ when $y = y^*$, $I(y, y^*) = -1$ when $y \neq y^*$.

Table 1 shows the results of the whole benchmark. To determine the effect of the number of unlabeled training meshes in our semi-supervised approach, we respectively use 7 unlabeled meshes and 17 unlabeled meshes, and the right three columns of Table 1 uses same labeled training sets and features. The two supervised approaches have similar performance. For the semi-supervised approach with 17 unlabeled meshes and 3 labeled meshes, the accuracy is 89.9%, which is approximately 4.5% better than the supervised approaches. Table 2 demonstrates that with more labeled training meshes, the improvement over supervised learning scheme will decline, especially when the original accuracy is already high. However, it is still better compared to the supervised methods.

The experimental results demonstrate that with the information from unlabeled meshes, the segmentation and labeling accuracy tends to improve. Also, Table 1 shows that the improvement slows down as the unlabeled training set becomes larger.

Dataset	Sup(3)	Semi(3 ¹⁷)	Sup(6)	Semi(6 ¹⁴)
Airplane	91.1	95.9	94.6	96.4
Ant	93.8	98.7	97.9	98.8
Bearing	76.2	82.4	78.3	87.3
Fish	92.3	95.7	93.6	95.1
Hand	91	94.7	93.1	94.9
Vase	77.4	81.2	78.4	82.3

Dataset	Sup(9)	Semi(9 ¹¹)	Sup(12)	Semi(12 ⁸)
Airplane	95.3	96.3	95.3	96.1
Ant	98.3	98.4	98	98.8
Bearing	84.7	88.1	84.9	88.3
Fish	94.1	95.6	94.7	95.5
Hand	94	95.1	93.9	94.8
Vase	79.7	84.6	82.2	86.1

Table 2: The comparative results of our supervised method and semi-supervised method with different number of labeled meshes and all left unlabeled meshes.

6.2. Robustness to Inconsistently Labeled Data

The supervised mesh segmentation method requires consistently labeled meshes, otherwise the results may be undesirable. However, due to the likelihood that different people



Figure 4: Comparative results on the human mislabeled training set. The first three are the labeled training meshes, but the first one is mislabeled. The next three are the results of the baseline supervised method, they are greatly affected by the mislabeled meshes, the last three are the results of our semi-supervised method, which are less influenced.

may have different segmentation boundaries to the same object, noise of the labeled set is inevitable. On the other hand, since the semi-supervised learning scheme gains information from large sets of unlabeled features to de-noise, it is more robust to noisy supervision [Zhu07].

Intuitively, the entropy term acts as a regularizer, which exploits information from the unlabeled data, and can thus avoid the model from over-fitting to the training data, with possibly mislabeled outliers. Because the regularizer essentially serves as a prior of the model, the parameters learned from our framework are more robust to the purely supervised version, which solely learns model information from the data, and is thus totally data-dependent [Bis06]. Additionally, the entropy term actually measures the class overlap [GB04], which stresses the usefulness of incorporating the unlabeled data where the labeling is indeed ambiguous.

Our result Figure 4 shows that with the inconsistent labeled data, the accuracy of supervised learning method drops quickly while the semi-supervised learning method is relatively robust. We also tested on some dataset to randomly choose 5% faces of the labeled training dataset and changed its label to another random label, simulating the process of label missing and label error. Our experimental results show that the semi-supervised method is relatively robust to such interference, the results are shown in Table 3.

Dataset	Sup	Semi(17)	Dataset	Sup	Semi(17)
Airplane	87.6	94.5	FourLeg	78.4	81.4
Ant	93.4	97.8	Hand	81.2	89.5
Bearing	54.1	75.3	Human	55.0	76.7
Chair	83.2	90.4	Table	76.7	92.6
Fish	85.7	90.2	Teddy	87.2	91.1

Table 3: The comparative results of our supervised method and semi-supervised method, with 5% random noise added, which can test the robustness of the algorithms.

6.3. Complexity Analysis

In our approach, for the training process, the cost of each belief propagation is $\Theta((D_l + D_u)NJ)$, and the cost of each boosting iteration is $\Theta((D_l + D_u)NJ\Omega)$. And for the labeling process, the application of boosting classifier costs $\Theta(NJ\Omega_{selected})$ and the belief propagation costs $\Theta((D_l + D_u)NJ)$. D_l is the number of labeled meshes, D_u is the number of unlabeled meshes, N is the size of each mesh, J is number of labels, and Ω is the dimension of features. The number of training iterations needed in [KHS10] is almost three times of our method, and the cost of each iteration is relatively the same. The space cost of this method is $\Theta(NJ^2\Omega_{selected})$, because of the feature selection step, we have $\Omega_{selected} \ll \Omega$. Therefore, the feature selection becomes very useful when faced with high dimensional data, particularly if they include a large number of redundant features.

The computational cost of this algorithm is almost the same to the baseline supervised version. Even though the algorithm converges faster, the unsupervised method always need more training data to be processed. The experimental results are performed on I7 2600 3.4GHz processor with 12G RAM. For each training process, we need to process 20 meshes, each has about 20-30K faces. This process takes about 7 hours, but when the model is learned, applying it to new meshes cost only a few minutes per mesh.

7. Discussion and Future Works

The method proposed in this paper is the *first* approach in mesh segmentation that uses both labeled and unlabeled meshes. Although it achieves some success, the result is still data-dependent. The parameters we used are manually adjusted by checking the results, actually more accurate method can be used based on the approach of [SWRC07], as in [KHS10]. Similar to [KHS10], mesh faces with weak features, such as the human face, can not be accurately segmented and the meshes with sharp features such as mechanical parts also cannot work well here. Besides, there are following two major limitations:

Jagged and Disconnect Patches Because only dihedral angles are used as geometry features of pairwise term, which abandons much very important information, making the boundaries in some meshes jagged. Additionally, for some meshes such as horses, there are small disconnected components, which degrade the segmentation quality. To alleviate these problems, more emphasis should be put on pairwise features in the belief propagation process. Integrating graph cut as a post-processing to the overall pipeline may also help.

Human Mislabeled. The human mislabeling problem arises from the fact that the true boundaries of the meshes are hard to mark for human beings. Actually, it is one big source of noises and many typical de-noising methods can help. The entropy regularizer we added actually plays such a

de-noising role and shows desired results on several datasets. However, our method cannot totally solve the issue, and more further works are required on it.

Besides solving above problems, our approach can be expanded in the following two aspects, and providing a performance statistics of many prominent learning techniques on different datasets is valuable.

Hierarchical Models. Most methods for mesh segmentation are based on a single choice of quantisation of a mesh space - faces, segments or group of segments, which has its pros and cons. Our semi-supervised CRF model can be extended to a hierarchical model where graphs combining different quantisation levels can be built. The hierarchical model is likely to integrate the benefits from features of different quantisations and yields better results.

Partially Labeled Semi-supervised Mesh Segmentation. The nature of semi-supervised methods determines that our method can work well even if the labeled meshes are just partially labeled, which may result from missing data or human error. So this semi-supervised mesh segmentation method can be applied to this kind of meshes and likely to produce much better results than the supervised one.

Acknowledgements

This work was partially supported by China 973 Program (No.2009CB320801), NSFC (No. 61170139 and 60933007), the National High Technology Research and Development (863) Program of China (No.2012AA011503). We would also like to thank the anonymous reviewers for their valuable comments and suggestions.

References

- [Bis06] BISHOP C. M.: *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. 7
- [BLVD11] BENHABILES H., LAVOUËL G., VANDEBORRE J. P., DAUDI M.: Learning boundary edges for 3D-mesh segmentation. *Computer Graphics Forum* 30, 8 (Dec. 2011), 2170–2182. 3
- [CGF09] CHEN X., GOLOVINSKIY A., FUNKHOUSER T.: A benchmark for 3D mesh segmentation. *ACM Transactions on Graphics* 28, 3 (Aug. 2009). 1, 2, 5
- [CKGK11] CHAUDHURI S., KALOGERAKIS E., GUIBAS L., KOLTUN V.: Probabilistic reasoning for assembly-based 3D modeling. *ACM Transactions on Graphics* 30, 4 (2011). 1
- [FHT00] FRIEDMAN J., HASTIE T., TIBSHIRANI R.: Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics* 28, 2 (2000), 337–407. 2, 4
- [GB04] GRANDVALET Y., BENGIO Y.: Semi-supervised learning by entropy minimization. In *NIPS'04* (2004), pp. –1–1. 2, 4, 7
- [GF08] GOLOVINSKIY A., FUNKHOUSER T.: Randomized cuts for 3D mesh analysis. *ACM Transactions on Graphics* 27, 5 (Dec. 2008). 2
- [HKG11] HUANG Q.-X., KOLTUN V., GUIBAS L. J.: Joint shape segmentation with linear programming. *ACM Trans. Graph.* 30, 6 (2011), 125. 3, 6
- [JWL*06] JIAO F., WANG S., LEE C.-H., GREINER R., SCHURMANS D.: Semi-supervised conditional random fields for improved sequence segmentation and labeling. In *ICCL, ACL* (Stroudsburg, PA, USA, 2006), ACL-44, Association for Computational Linguistics, pp. 209–216. 3, 4
- [KHS10] KALOGERAKIS E., HERTZMANN A., SINGH K.: Learning 3d mesh segmentation and labeling. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 102. 1, 2, 3, 5, 6, 7
- [LCFK07] LIAO L., CHOUDHURY T., FOX D., KAUTZ H.: Training conditional random fields using virtual evidence boosting. In *IJCAI* (2007), vol. 35. 2, 3, 4, 5
- [LMP01] LAFFERTY J., MCCALLUM A., PEREIRA F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. of ICML* (2001). 2, 3
- [MC07] MAHDAVIANI M., CHOUDHURY T.: Fast and scalable training of semi-supervised crfs with application to activity recognition. *NIPS 20* (2007), 977–984. 2, 3
- [MWJ99] MURPHY K., WEISS Y., JORDAN M.: Loopy belief propagation for approximate inference: An empirical study. In *UAI* (1999), Morgan Kaufmann Publishers Inc., pp. 467–475. 2, 4
- [Obe07] OBERLIES T.: Understanding virtual evidence boosting : A new derivation and application to recognizing spatial context with wearable cameras. *Graphical Models* (2007). 2, 5
- [Pea88] PEARL J.: *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. 4
- [Sha08] SHAMIR A.: A survey on mesh segmentation techniques. *Computer Graphics Forum* 27, 6 (2008), 1539–1556. 2
- [SvKK*11] SIDI O., VAN KAICK O., KLEIMAN Y., ZHANG H., COHEN-OR D.: Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering. *ACM Transactions on Graphics (Special Issue of SIGGRAPH Asia)* 30, 6 (2011), 126:1–126:9. 3
- [SWRC07] SHOTTON J., WINN J., ROTHER C., CRIMINISI A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context, 2007. 7
- [SWRC09] SHOTTON J., WINN J., ROTHER C., CRIMINISI A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comput. Vision* 81, 1 (Jan. 2009), 2–23. 3
- [VKTS*11] VAN KAICK O., TAGLIASACCHI A., SIDI O., ZHANG H., COHEN-OR D., WOLF L., HAMARNEH G.: Prior knowledge for part correspondence. *Computer Graphics Forum* 30, 2 (2011), 553–562. 1, 3
- [YNX*12] YANG Y., NIE F., XU D., LUO J., ZHUANG Y., PAN Y.: A multimedia retrieval framework based on semi-supervised ranking and relevance feedback. *IEEE Trans. Pattern Anal. Mach. Intell.* 34, 4 (Apr. 2012), 723–742. 2
- [Zhu07] ZHU X.: Semi-supervised learning literature survey. *SciencesNew York* 10, 1530 (2007), 1–59. 2, 7