Creating Consistent Scene Graphs Using a Probabilistic Grammar

Tianqiang Liu\textsuperscript{1}  Siddhartha Chaudhuri\textsuperscript{1,2}  Vladimir G. Kim\textsuperscript{3}

Qi-Xing Huang\textsuperscript{3,4}  Niloy J. Mitra\textsuperscript{5}  Thomas Funkhouser\textsuperscript{1}

Motivation

Growing number of 3D scenes online.

Google 3D warehouse
Motivation

Synthesis [Fisher et al 2012, Xu et al 2013]

Understanding [Xu et al 2014, Song et al 2014]
Goal

Input: A scene from Trimble 3D Warehouse
Goal

Output 1: Semantic segmentations
Goal

Output 2: Category labels.

- Window
- Bed frame
- Pillow
- Mattress
- Mirror
- Door
- Heater
- Nightstand
- Dresser
- Artifact
Goal

Output 2: Category labels at different levels.
Goal

Output 2: Category labels at different levels.
Challenges

Shape is not distinctive.

Night table

Console table
Challenges

Contextual information

- Study desk
- Study chair
- Meeting table
- Meeting chair
Challenges

All-pair contextual information is not distinctive.

![Diagram of a meeting and study chair setup with corresponding charts showing the number of pairs within a certain distance.](image)
Challenges
Challenges
Challenges
Challenges

All-pair contextual information is not distinctive.
Key Idea

Semantic groups

Semantic hierarchy

Library

- Study area (X2)
- Study chair (X2)
- Study desk

- Meeting area (X5)
- Meeting chair (X4)
- Meeting table
Key Idea

- **Meeting chair vs Meeting chair**
  - Graph showing the number of pairs in feet.
  - Peaks at 1 and 3 feet.

- **Study chair vs Study chair**
  - Graph showing the number of pairs in feet.
  - Peak at 1 foot.

Diagram illustrating the seating arrangement and movement between different types of chairs.
Pipeline

Probabilistic grammar
Related Work

Van Kaick et al. 2013
Related Work

Van Kaick et al. 2013

Boulch et al. 2013
Overview

Grammar Structure

Learning a Probabilistic Grammar

Scene Parsing

Results
Probabilistic Grammar

Labels

Rules

Probabilities
Labels

Examples:

bed, night table, sleeping area
Rules

Example:

sleeping area → bed, night table
Probabilities

Derivation probabilities

Cardinality probabilities

Geometry probabilities

Spatial probabilities
Derivation probability $P_{nt}$

$P = 0.4$

bed $\rightarrow$ bed frame, mattress

$P = 0.6$
Cardinality probability $P_{\text{card}}$

sleeping area $\rightarrow$ bed, night table

$P_{\text{card}}(\text{bed} \mid \text{sleeping area}) \quad P_{\text{card}}(\text{nighttable} \mid \text{sleeping area})$
Geometry probability $P_g$

$$P_g(x \mid \text{bedframe}) > P_g(y \mid \text{bedframe})$$
Spatial probability $P_s$

$$P_s(x, y \mid \text{desk, chair, studyarea}) > P_s(z, y \mid \text{desk, chair, studyarea})$$
Overview

Grammar Structure

Learning a Probabilistic Grammar

Scene Parsing

Results
Pipeline

Probabilistic grammar
Learning a Probabilistic Grammar

Identify objects

Speed X 5
Learning a Probabilistic Grammar

Label objects

Speed X 5
Learning a Probabilistic Grammar

Group objects
Learning a Probabilistic Grammar

Grammar generation

→ Labels  all unique labels
  Rules
  Probabilities
Learning a Probabilistic Grammar

Grammar generation

Labels

Rules

concatenating all children for each label

Probabilities

Training example 1:

Training example 2:
Learning a Probabilistic Grammar

Grammar generation

Labels

Rules

→ Probabilities

\[ P_{nt}, P_{\text{card}} \] : learning from occurrence statistics

\[ P_g \] : estimating Gaussian kernels

\[ P_s \] : kernel density estimation
Overview

Grammar Structure

Learning a Probabilistic Grammar

Scene Parsing

Results
Pipeline

Probabilistic grammar
Pipeline

Probabilistic grammar
Scene parsing

Objective function

\[ H^* = \operatorname{arg\,max}_H P(H \mid S, G) \]

- \( H \) is the unknown hierarchy
- \( S \) is the input scene
- \( G \) is the probabilistic grammar
Scene parsing

After applying Bayes’ rule

\[ H^* = \arg \max_H P(H \mid G)P(S \mid H, G) \]
Scene parsing

After applying Bayes’ rule

$$H^* = \arg\max_H P(H \mid G) P(S \mid H, G)$$

Prior of hierarchy

$$P(H \mid G) = \prod_{x \in H} P_{prod}(x)^{T(x)}$$
Scene parsing

After applying Bayes’ rule

\[ H^* = \arg\max_H P(H \mid G)P(S \mid H, G) \]

Prior of hierarchy

\[ P(H \mid G) = \prod_{x \in H} P_{\text{prod}}(x)^{T(x)} \]

\( P_{\text{prod}}(x) \) : probability of a single derivation

formulated using \( P_{\text{nt}}, P_{\text{card}} \)
Scene parsing

After applying Bayes’ rule

\[ H^* = \arg\max_H P(H \mid G)P(S \mid H, G) \]

Prior of hierarchy

\[ P(H \mid G) = \prod_{x \in H} P_{prod}(x)^{T(x)} \]

\( T(x) \) compensates for decreasing probability as \( H \) has more internal nodes.
Scene parsing

After applying Bayes’ rule

\[ H^* = \arg\max_H P(H \mid G)P(S \mid H,G) \]

Likelihood of scene

\[ P(S \mid H,G) = \prod_{x \in H} P_g(x)^{T(x)} P_s^*(x)^{T(x)} \]
Scene parsing

After applying Bayes’ rule

\[ H^* = \arg \max_H P(H \mid G)P(S \mid H, G) \]

Likelihood of scene

\[ P(S \mid H, G) = \prod_{x \in H} P_g(x)^{T(x)} P_s^*(x)^{T(x)} \]

\[ P_g(x) : \text{geometry probability} \]
Scene parsing

After applying Bayes’ rule

\[ H^* = \arg\max_H P(H \mid G)P(S \mid H,G) \]

Likelihood of scene

\[ P(S \mid H,G) = \prod_{x \in H} P_g(x)^{T(x)} P_s^*(x)^{T(x)} \]

\[ P_s^*(x) : \text{sum of all pairwise spatial probabilities } P_s(x) \]
Scene parsing

We work in the negative logarithm space

\[ E(H) = \log P(H \mid G)P(S \mid H,G) \]

\[ = -\sum_{x \in H} T(x) \log \left( P_{\text{prod}}(x)P_g(x)P^*_s(x) \right) \]
Scene parsing

Rewrite the objective function recursively

\[ E(H) = \overline{E}(R) \]
\[ \overline{E}(x) = E(x) + \sum_{y \in x.\text{children}} \overline{E}(y) \]

where \( R \) is the root of \( H \), \( \overline{E} \) is the energy of a sub-tree.
Scene parsing

The search space is prohibitively large ...

• Problem 1: #possible groups is exponential.

• Problem 2: #label assignments is exponential.
Scene parsing

Problem 1: #possible groups is exponential.
Problem 1: #possible groups is exponential.

Solution: proposing candidate groups.
Scene parsing

Problem 2: #label assignments is exponential.
Scene parsing

Problem 2: \#label assignments is exponential.

Solution: bounding \#RHS by grammar binarization

where $x'$ is partial label of $x$, $k \in \{a_1, a_2, \ldots, a_n\}$
Scene parsing

Problem 2: \#label assignments is exponential.

Solution: bounding \#RHS by grammar binarization

where \( x' \) is partial label of \( x \), \( k \in \{a_1, a_2, ..., a_n\} \)

Now \#rules and \#assignments are both polynomial.

The problem can be solved by dynamic programming.
Scene parsing

Problem 2: \#label assignments is exponential.

Solution: bounding \#RHS by grammar binarization

Convert the result to a parse tree of the original grammar.
Overview

Grammar Structure

Learning a Probabilistic Grammar

Scene Parsing

Results
Benefit of hierarchy

Meeting table

Study chair

Shape only
Benefit of hierarchy

Shape only

Flat grammar
Benefit of hierarchy

Shape only

Flat grammar

Ours
Benefit of hierarchy

Shape only

Flat grammar

Ours

Meeting table

Study chair

Meeting area

Study area
Benefit of hierarchy

Shape only
Benefit of hierarchy

Shape only

Flat grammar
Benefit of hierarchy

Shape only

Flat grammar

Ours
Benefit of hierarchy

Shape only

Flat grammar

Ours
Datasets

77 bedrooms

30 classrooms

8 libraries

17 small bedrooms

8 small libraries
Benefit of hierarchy

Object classification

F1 Score

- Shape-only
- Flat grammar
- Ours

Small-library | Small-bedroom | Library | Bedroom | Classroom
Generalization of our method

Parsing Sketch2Scene data set
Take-away message

- Modeling hierarchy improves scene understanding.
Limitation and Future Work

• Modeling correlation in probabilistic grammar.
Limitation and Future Work

• Modeling correlation in probabilistic grammar.
• Grammar learning from noisy data.
Limitation and Future Work

- Modeling correlation in probabilistic grammar.
- Grammar learning from noisy data.
- Applications in other fields.

Modeling from RGB-D data [Chen et al. 2014]
Acknowledgement

Data
• Kun Xu

Discussion
• Christiane Fellbaum, Stephen DiVerdi

Funding
• NSF, ERC Starting Grant, Intel, Google, Adobe
Code and Data

Back up slides
Learning a Probabilistic Grammar

Creating a training set manually
1. Identify leaf-level objects.
2. Provide a label for each object in a scene.
3. Group objects and provide group labels.
4. Maintain consistency.
Sensitivity analysis

Impact of training set size

Impact of energy terms
Benefit of hierarchy

Object classification

<table>
<thead>
<tr>
<th></th>
<th>Shape-only</th>
<th>Flat grammar</th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small-library</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Small-bedroom</td>
<td>0.8</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>Library</td>
<td>0.6</td>
<td>0.7</td>
<td>0.9</td>
</tr>
<tr>
<td>Bedroom</td>
<td>0.6</td>
<td>0.7</td>
<td>0.8</td>
</tr>
<tr>
<td>Classroom</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Object grouping

<table>
<thead>
<tr>
<th></th>
<th>Ours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small-library</td>
<td>1.0</td>
</tr>
<tr>
<td>Small-bedroom</td>
<td>1.0</td>
</tr>
<tr>
<td>Library</td>
<td>0.8</td>
</tr>
<tr>
<td>Bedroom</td>
<td>0.8</td>
</tr>
<tr>
<td>Classroom</td>
<td>0.8</td>
</tr>
</tbody>
</table>