Imagining the Unseen: Stability-based Cuboid Arrangements for Scene Understanding

Tianjia Shao*  Aron Monszpart§
Youyi Zheng†  Bongjin Koo §  Weiwei Xu‡
Kun Zhou *  Niloy J. Mitra §

* Zhejiang University
§ UCL
† Nanjing Normal University
‡ Nanyang Normal University

SIGGRAPH ASIA 2014 SHENZHEN
Background

- A fundamental problem for single view data acquisition
  - Missing observations due to scene occlusion
  - Limiting the understanding of indoor scenes

From the capture view

From another view
How do We Imagine Unseen Data?
How do We Imagine Unseen Data?
Our Motivation

Mimic the process of human’s imagination based on physical stability
Our Motivation

- Physical stability can help to reason about the scene structure
  - Relationship among parts – touching or fixed
Related Works: Proxy-based Scene Understanding

- Structured output of primitives
  - [Li et al. 2011]
  - [Lafarge et al. 2013]
- Creating abstracted geometry
  - [Arikan et al. 2013]
- Studying spatial layout
  - [Gupta et al. 2010]
  - [Lee et al. 2010]
  - [Hartley et al. 2012]
- Image manipulation
  - [Zheng et al. 2012]
Related Works: Cuboid Detection

- Statistical deformable cuboid model
  - [Fidler et al. 2012]

- Cuboid corner point model
  - [Xiao et al. 2012]

- Image space contrast-based features
  - [Hedau et al. 2012]
Related Works: Physical Validity Constraints

- Penetration free
  - [Hedau et al. 2010]
  - [Jiang and Xiao 2013]

- Improve segmentation
  - [Jia et al. 2013]

- Voxel-based scene parsing
  - [Zheng et al. 2013]
  - [Kim et al. 2013]
Our Goal

- Recover the underlying structure of indoor scenes
  - Abstracting indoor scenes as collections of cuboids
  - Hallucinate geometry in the occluded regions
  - Identifying part parameters (size and orientation)
  - Identifying part relations (touching or fixed)
Algorithm Overview
Algorithm Overview

Input scan → Initial cuboids → Optimized cuboids + recovered structure
Algorithm Overview

Input scan → Initial cuboids → Optimized cuboids + recovered structure
Inferring Geometry and Relations

Key observation: relations are coupled with geometry

G supports A & E;
E supports B & C & D

G supports E;
E supports A & C & D;
C supports B

Optimized geometry

Optimized geometry
Inferring Geometry and Relations

- Encode the discrete interaction among the cuboids as a multi-connection graph $G := (V, E)$
  - $V$: cuboids;
  - $E$: contact types (and corresponding cuboid extensions)
Inferring Geometry and Relations

Optimization formulation

- Edge selection problem
  - $x_{ij}^k = 1$: edge type $e_{ij}^k$ selected;
  - $x_{ij}^k = 0$: not selected.
- Goal: Physically stable arrangement of cuboids with minimal number of fixed contacts
  - $\min \{x_{ij}^k\} \#(e_{ij}^k = \text{fixed joint})$
- Constraints: $\sum_k x_{ij}^k = 1$
Inferring Geometry and Relations

- Pruning the solution space
  - Pre-pruning
    - Penetration check
    - Visibility check
Inferring Geometry and Relations

- Pruning the solution space
  - Pre-pruning
    - Penetration check
    - Visibility check
  - Branch-and-bound

Most stable

Fixed

Touching

unstable

stable

unstable

Prune

Expand

Prune

T. J. Shao, A. Monszpart, et al.
Inferring Geometry and Relations

- Assessing physical stability
  - Local reasoning is not enough
Measuring global stability through static equilibrium

- The net force and torque should be zero
- Decompose the contact forces to compression forces and friction forces
- $E_s(A) := \min_{\mathbf{f}} \|\mathbf{Df + w}\|^2$ s.t. $\mathbf{f}_n^i \geq 0$ and $|\mathbf{f}_u^i|, |\mathbf{f}_v^i| < \mu \mathbf{f}_n^i$
Overview

- Ground truth
- Metrics
- Robustness
- Validity
- Applications
Overview

• Ground truth
• Metrics
• Robustness
• Validity
• Applications
Evaluation – Ground Truth

- 20 scenes
- High occlusion
- 3 dimensions measured by hand
Evaluation – Ground Truth

- 20 scenes
- High occlusion
- 3 dimensions measured by hand

- 700 scenes (NYU2)
- Medium occlusion
- Cuboids, floors, walls
- Support graph
- Initialization

• Fixed
• Touching
Evaluation – Ground Truth

- 20 scenes
- High occlusion
- 3 dimensions measured by hand

- 700 scenes (NYU2)
- Medium occlusion
- Cuboids, floors, walls
- Support graph
- Initialization

Online: 720 scenes, annotator tool
Overview

- Ground truth
- Metrics
- Robustness
- Validity
- Applications
Evaluation – Metrics

- Approximation error
- $L_1$-norm
Evaluation – Metrics

• Approximation error
• $L_1$-norm

• Correctness of structure graph
• $F_1$ score
Evaluation – Approximation error

Initial
Evaluation – Approximation error

Initial

Ground truth

\( \{ \text{err}^{\text{init}}_i \} \)
Evaluation – Approximation error

- Initial
- Ground truth
- Optimized

\{err_i^{\text{init}}\} \quad \{err_i^{\text{optimized}}\}
Evaluation – Approximation error

Initial

Ground truth

Optimized

\[ \{err_i^{\text{init}}\} \rightarrow \{err_i^{\text{optimized}}\} \]

\[ err_i^{\text{init}} - err_i^{\text{optimized}} > 5\% \]
Evaluation – Approximation error

\[ \{ err_i^{\text{init}} \} \quad \{ err_i^{\text{optimized}} \} \]

\[ err_i^{\text{init}} - err_i^{\text{optimized}} > 5\% \]

35% > 3%
Evaluation – $F_1$ score

RGB(-D)

Ground Truth

Fixed
Touching
Evaluation – $F_1$ score

RGB(-D)

Ground Truth

Output

Fixed

Touching

FP
Evaluation – $F_1$ score

RGB(-D)

Ground Truth

Output

Fixed
Touching
Evaluation – $F_1$ score

<table>
<thead>
<tr>
<th>Precision</th>
<th>$\frac{TP}{TP + FP}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>$\frac{TP}{TP + FN}$</td>
</tr>
<tr>
<td>$F_1$</td>
<td>$200 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$</td>
</tr>
</tbody>
</table>

RGB(-D)

Ground Truth

Fixed

Touching

Output
Evaluation – F₁ score

<table>
<thead>
<tr>
<th></th>
<th>#Scenes</th>
<th>Initial</th>
<th>Optimized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic</td>
<td>12</td>
<td>32.2</td>
<td>87.2</td>
</tr>
<tr>
<td>NYU2</td>
<td>700</td>
<td>40.0</td>
<td>60.5</td>
</tr>
<tr>
<td>Own</td>
<td>20</td>
<td>56.5</td>
<td>95.9</td>
</tr>
<tr>
<td>Kinfu</td>
<td>2</td>
<td>32.1</td>
<td>89.7</td>
</tr>
</tbody>
</table>

RGB(-D)
Evaluation – $F_1$ score

RGB(-D)  

Ground Truth

Fixed  
Touching

Initialization
Evaluation – F₁ score

RGB(-D)  

Ground Truth

Optimized

Initialization

Fixed  

Touching
Evaluation – $F_1$ score

RGB(-D)

Ground Truth

 Optimization

Fixed

Touching

T. J. Shao, A. Monszpart, et al.
Overview

- Ground truth
- Metrics
- Robustness
- Validity
- Applications
Robustness to viewpoint

Viewpoint 1

Fixed
Touching
Robustness to viewpoint

Viewpoint 1  Viewpoint 2

Fixed  Touching
Robustness to scene complexity

Input scene
Robustness to scene complexity

Input cloud (back view)
Robustness to scene complexity

Initialization
Robustness to scene complexity

Initialization

Optimized

Structure graph

Fixed

Touching
Robustness to scene complexity

Initialization  Optimized  Structure graph

Fixed
Touching
Overview

• Ground truth
• Metrics
• Robustness
• Validity
• Applications
Evaluation – Validity

RGB-D input

Initial (side view)
Evaluation – Validity

RGB-D input

Initial (side view)

Optimized (side view)
Evaluation – Validity

Less occluded scan
Evaluation – Validity

Less occluded scan + Optimized
Evaluation – Validity

Less occluded scan + Optimized
Extra proxies removed
Front view
Overview

• Ground truth
• Metrics
• Robustness
• Validity
• Applications
Applications #1 – Model retrieval
Applications #1 – Model retrieval
Applications #1 – Model retrieval
Applications #2 – Re-arrangement
Applications #2 – Re-arrangement
Applications #3 - Completion

RGB-D Input
Applications #3 - Completion

RGB-D Input

Physically stable approximation
Applications #3 - Completion

RGB-D Input  Completed models
Applications #3 - Completion

RGB-D Input

Completed models
Limitations

heterogeneous

[http://azfoo.net/places/az/mesa/cemetery]

[GravityGlue.com]
Limitations

Fixed
Touching

[http://azfoo.net/places/az/mesa/cemetery]

[GravityGlue.com]
Code + Data:
Robustness to initialization

- No perturbation
- 5°
- 10°