Estimating the 3D Layout of Indoor Scenes and its Clutter from Depth Sensors

ETH zürich
USC Viterbi School of Engineering

Jian Zhang[1] Kan Chen[2] Alexander G. Schwing[1] Raquel Urtasun[3]

[1]ETH Zurich [2]USC [3]TTI Chicago

TOYOTA TECHNOLOGICAL INSTITUTE CHICAGO

Overview

Goal

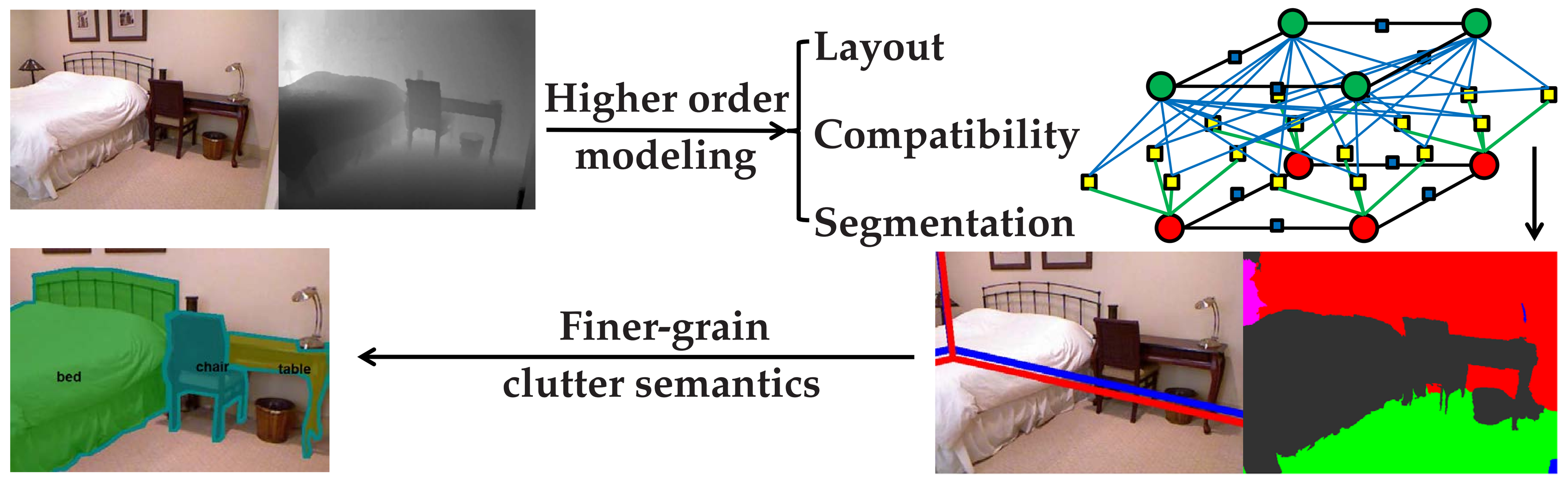
Boost layout estimation and clutter labeling with RGBD data

Motivation

- Indoor layout estimation and clutter labeling are correlated
- Kinect provides more reliable 2.5D cues for scene understanding

Approach

Jointly estimate layout and segment clutter and bounding faces



Model

Overall model

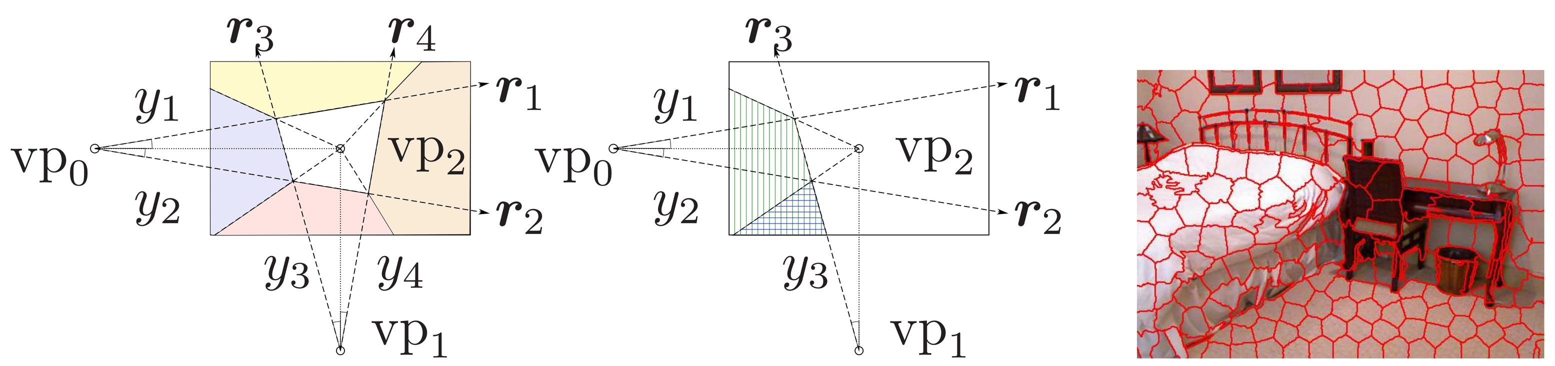
- Let $\boldsymbol{x}$ be clutter labeling and $\boldsymbol{y}$ the layout
- Joint inference on layout, clutter labeling and compatibility with

$$\min_{\boldsymbol{x},\boldsymbol{y}} E_{layout}(\boldsymbol{y}) + E_{labeling}(\boldsymbol{x}) + E_{comp}(\boldsymbol{x},\boldsymbol{y})$$

Layout modeling

Parameterization:

- y_i are discretized states for the horizontal and vertical rays $\boldsymbol{r}_i$
- Face $\alpha \in \mathcal{F} = \{$*left-wall, right-wall, front-wall, ceiling, floor*$\}$
- Front-wall is defined by 4 rays and other faces by 3 rays



Energy of layout:

- Utilize Geometric Context (GC) and Orientation Map (OM) as cues
- $\phi_{lay,\alpha}(\boldsymbol{y})$ is the sum of image cues in face α. The layout energy is

$$E_{layout}(\boldsymbol{y}) = \sum_{\alpha\in\mathcal{F}} w_{lay,\alpha}^{\top}\phi_{lay,\alpha}(\boldsymbol{y})$$

- High order energy decomposed to pairwise [Schwing et al., 2012] (See middle figure)

$$\phi_{lay,\,\text{left-wall}}(y_1,y_2,y_3) = \phi_{lay,green}(y_1,y_3) - \phi_{lay,blue}(y_2,y_3)$$

Clutter labeling

- We extend LAB SLIC superpixel with depth gradient information
- Variable $x_i \in \mathcal{L} = \mathcal{F}\cup\{$*clutter*$\}$
- GC, OM, superpixel centroid and normal as unary cues
- Binary physical relation (e.g. floor is below clutter) as pairwise cues

Model

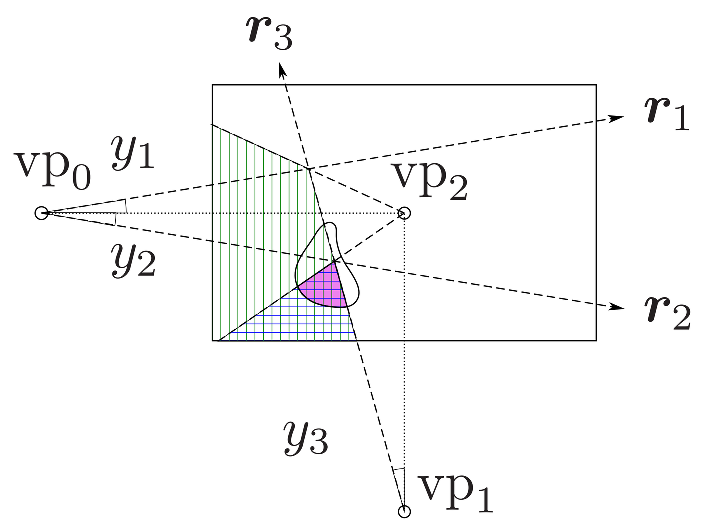
Energy of labeling:

$$-E_{labeling}(\boldsymbol{x}) = \sum_{\beta\in\mathcal{L}}\sum_{i=1}^{n} w_{lab,\beta}^{\top}\phi_{lab,\beta,i}(x_i) + \sum_{\beta,\gamma\in\mathcal{L}}\sum_{(i,j)\in E} w_{lab,\beta,\gamma}^{\top}\phi_{lab,\beta,\gamma,i,j}(x_i,x_j)$$

Compatibility modeling

Idea:

- Pixel-wise consistent estimation of bounding faces from layout estimation and clutter labeling



Energy of compatibility:

- $\boldsymbol{p}$ is an arbitrary pixel in a superpixel labeled as class $x(\boldsymbol{p}) \in \mathcal{L}$
- Accumulate overlapping of face $\alpha\in\mathcal{F}$ from layout and class $\beta\in\mathcal{L}$ from clutter labeling. $\gamma(\boldsymbol{y},\alpha)$ is the face α in layout $\boldsymbol{y}$.

$$\phi_{comp,\alpha,\beta}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\boldsymbol{p}\in\gamma(\boldsymbol{y},\alpha)}\delta(x(\boldsymbol{p})=\beta)$$

- Sum over all the (α,β) pair, we get third order terms

$$E_{comp}(\boldsymbol{x},\boldsymbol{y}) = \sum_{\alpha\in\mathcal{F}}\sum_{\beta\in\mathcal{L}} w_{comp,\alpha,\beta}\phi_{comp,\alpha,\beta}(\boldsymbol{x},\boldsymbol{y})$$

Alternating Inference

Algorithm 1 Alternating Inference

1: $\boldsymbol{y}^{(0)} = \min_{\boldsymbol{y}} E_{layout}(\boldsymbol{y})$
2: **for all** $i = 1 : M$ **do**
3: $\boldsymbol{x}^{(i)} = \min_{\boldsymbol{x}} E_{labeling}(\boldsymbol{x}) + E_{comp}(\boldsymbol{x},\boldsymbol{y}^{(i-1)})$
4: $\boldsymbol{y}^{(i)} = \min_{\boldsymbol{y}} E_{layout}(\boldsymbol{y}) + E_{comp}(\boldsymbol{x}^{(i)},\boldsymbol{y})$
5: **end for**
6: Return $\boldsymbol{x}^{(M)}, \boldsymbol{y}^{(M)}$

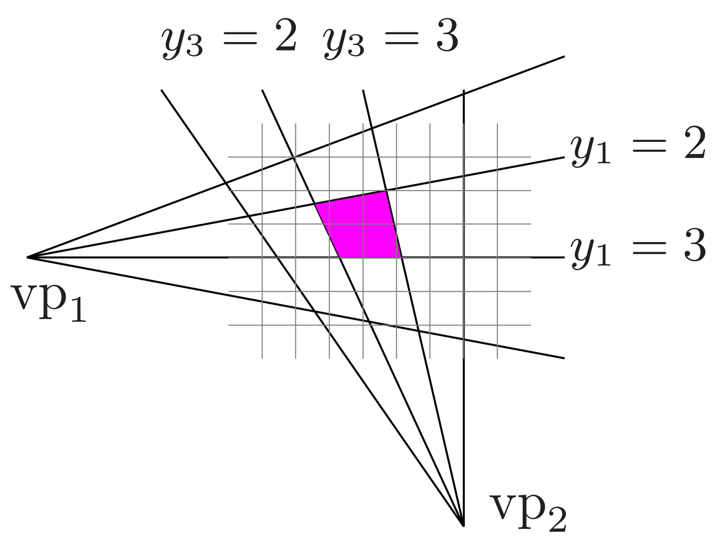
- Reduced to pairwise CRF in each iteration
- Appealing results achievable with only 2 rounds of alternating
- Alternating approximation is faster than convergency of direct BP

Fast Integral Geometry

Complexity of naive Integral Geometry computation is $O(N^2R^2)$. N^2 is the size of image and R is the cardinality of each ray variable

$O(N^2)$ solution

- Neighboring pixels lie in the same or neighboring bins of the ray grid
- Compute Integral Geometry by
  1. Assign pixles to $R\times R$ bin array
  2. Accumulate for bins with assigned pixels
  3. Do Integral Image on $R\times R$ matrix



Results

Experiment setup

- 202 training and 101 testing samples from NYU depth V2 dataset
- Inference with Convex Belief Propogation [Schwing et.al, 2011]
- Training with structured SVM [Tsochantaridis et al., 2004]

Comparison to the state-of-the-art

	layout error	labeling error
GC [Hoiem et al, 2007]	–	26.38%
[Schwing et al., 2012]	13.66%	–
Ours rgb	13.94%	23.68%
Ours rgb depth	**8.04%**	**13.65%**

Clutter labeling with finer-grain segmentation

- Six classes IOU of clutter labeling:

IOU	ceiling	floor	left	front	right	clutter	average
rgb	53.57	51.48	61.71	69.21	63.37	58.26	59.60
rgb depth	**80.36**	**85.84**	**68.93**	**77.27**	**75.15**	**72.47**	**76.67**

- We employ [Ren et al., 2012] to train classifiers for splitting clutter into 6 furniture classes. Clutter semantics classifier is trained with RGBD kernel descriptors: gradient, color, local binary pattern, depth gradient, surface normal and self-similarity
- The IOU of finer-grain segmentation:

C in structSVM	100	300	500	1000	3000	5000	10000
toilet	35.67	35.72	37.19	36.03	39.54	38.30	32.02
bed	43.69	43.59	43.42	43.46	43.00	43.47	42.49
table	16.90	20.90	21.29	21.21	20.60	20.98	21.02
cabinet	20.68	20.61	21.49	21.26	21.43	19.23	19.60
sofa	32.50	32.28	32.90	32.88	32.86	32.64	32.11
chair	35.89	35.90	36.18	36.13	35.96	35.00	35.05

Speed improvement of Integral Geometry

The speed of 640×480 Integral Geometry is reduced from $7s$ to $\mathbf{0.32s}$.

Qualitative results