Rent3D:

Floor-Plan Priors for Monocular Layout Estimation

Chenxi Liu[1,*] Alexander Schwing[2,*] Kaustav Kundu[2]
Raquel Urtasun[2] Sanja Fidler[2]

[1]Tsinghua University, [2]University of Toronto

How Many Times Have You Looked for Apartments?

How Many Times Have You Looked for Apartments?

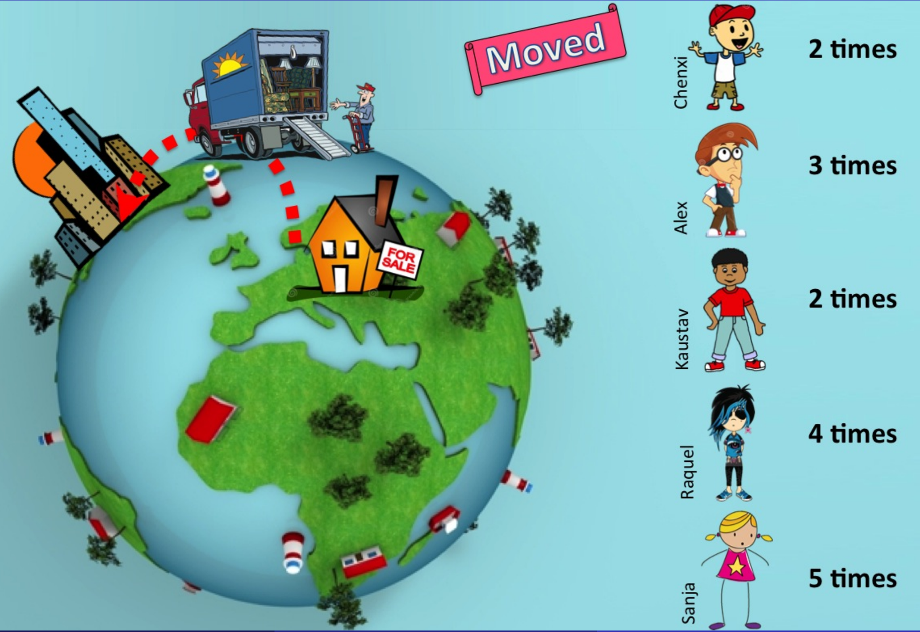
United States:

- 11.7% per year

Craigslist:

- 90,000 rental ads per day only in New York
- 10 million people visit the website per day

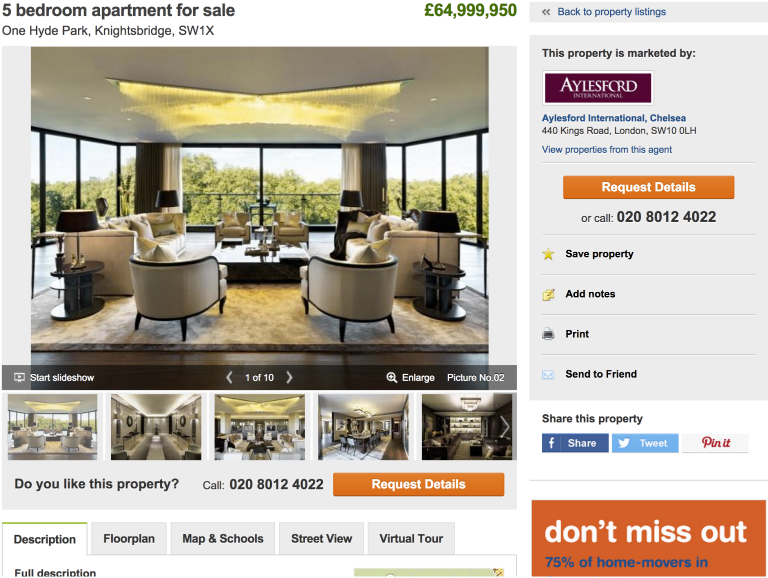
How Many Times Have You Looked for Apartments?



Finding an Apartment/House is a Pain...

- Particularly during a winter in Toronto

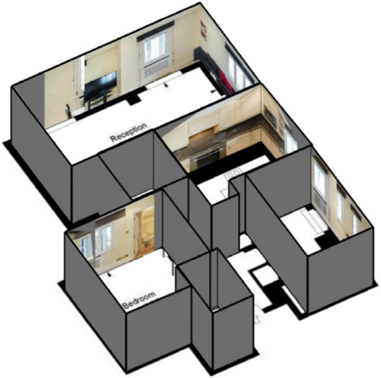
Renting Apartments



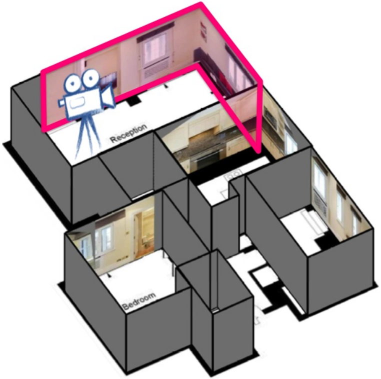
Example Rental Data

- Plus some meta information e.g. wall height

Rent3D: View Rental Ads in 3D



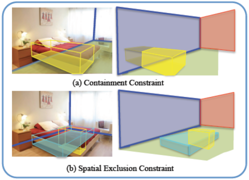
Rent3D: View Rental Ads in 3D



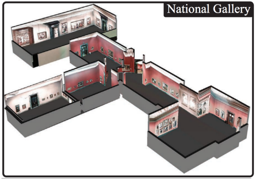
- Camera localization within apartment

Related Work

- Room layout estimation
  - ▷ Hedau et al., 2009, 2012
  - ▷ Lee et al., 2010
  - ▷ Schwing et al., 2012, 2013
  - ▷ Del Pero et al., 2011, 2012
  - ▷ Choi et al., 2013
- Virtual tours
  - ▷ Xiao & Furukawa, 2012
- 3D indoor reconstruction from large photo collections or video
  - ▷ Cabral & Furukawa, 2014
  - ▷ Brualla et al., 2014
- Indoor localization (video, depth sensors)
  - Project Tango
  - SLAM work



Lee et al., 2010

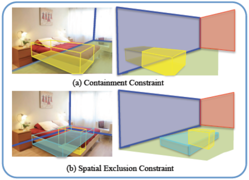


Xiao & Furukawa, 2012

Cabral & Furukawa, 2014

Related Work

- Room layout estimation
  - ▷ Hedau et al., 2009, 2012
  - ▷ Lee et al., 2010
  - ▷ Schwing et al., 2012, 2013
  - ▷ Del Pero et al., 2011, 2012
  - ▷ Choi et al., 2013



Lee et al., 2010

Our work:

3D indoor reconstruction and localization using monocular imagery

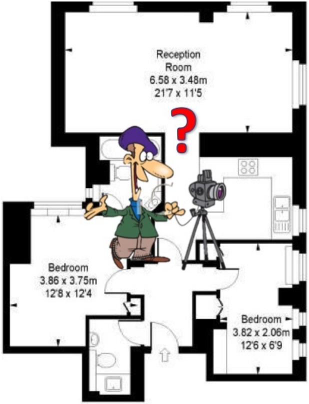
Xiao & Furukawa, 2012

  - ▷ Cabral & Furukawa, 2014
  - ▷ Brualla et al., 2014

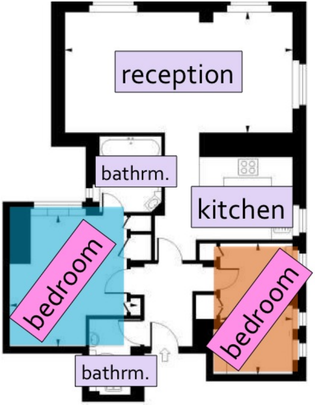
- Indoor localization (video, depth sensors)
  - Project Tango
  - SLAM work

Cabral & Furukawa, 2014

Overview


Reception
Room
6.58 x 3.48m
21'7 x 11'5
Bedroom
3.86 x 3.75m
12'8 x 12'4
Bedroom
3.82 x 2.06m
12'6 x 6'9

Overview

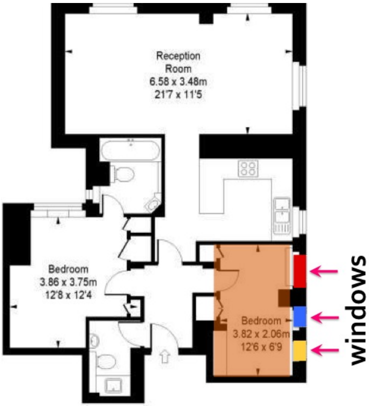


bedroom

Accurate **camera localization**:

- **Scene cues**

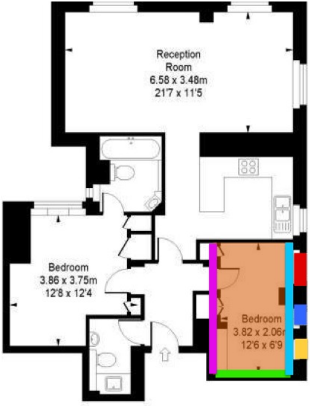
Overview

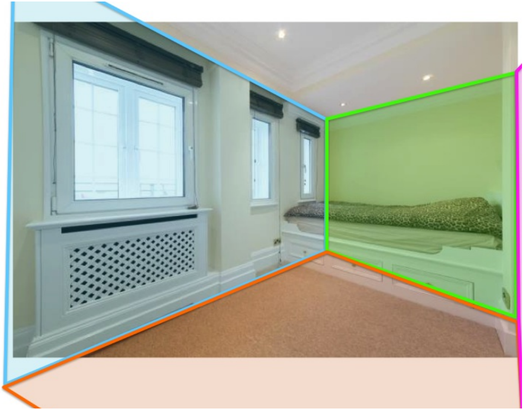


Accurate **camera localization**:

- **Scene cues**
- **Semantic cues**

Overview



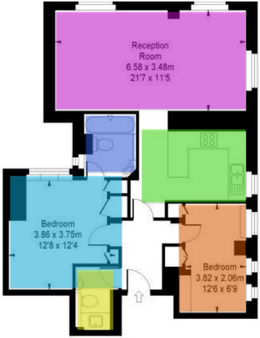


Accurate **camera localization**:

- **Scene cues**
- **Semantic cues**
- **Geometric cues** by exploiting the dimension information

Formulation

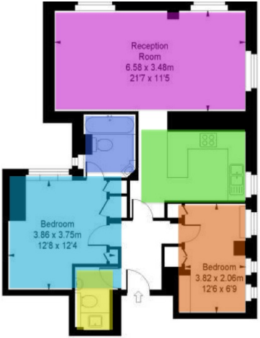
- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room

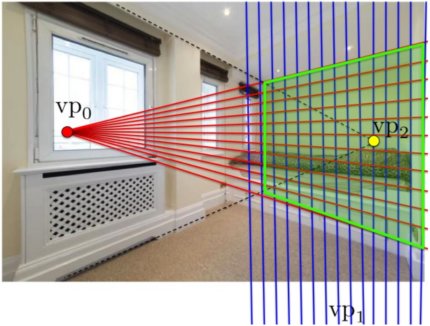


Formulation

- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room

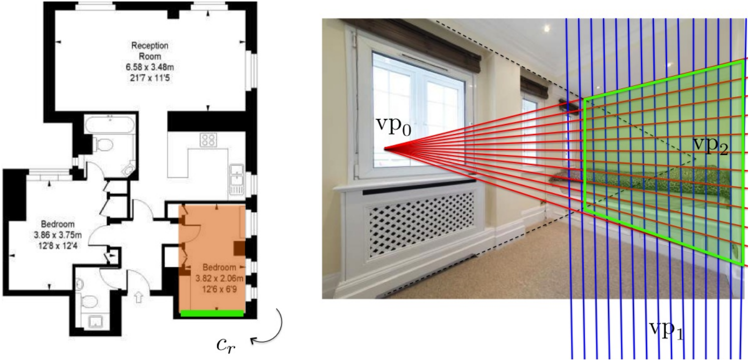
Front wall is the plane defined by vp_0 and vp_1





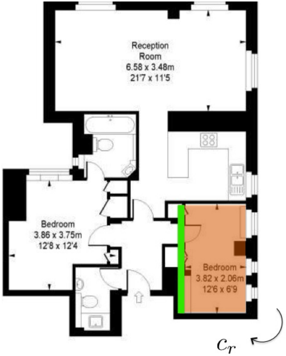
Formulation

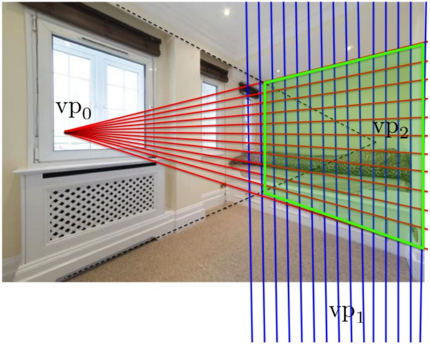
- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room
- $c_r \in \{1, \ldots, |C_r|\}$ ... a discrete variable representing within room r which wall the picture is facing ($|C_r|$ the number of walls in a room)



Formulation

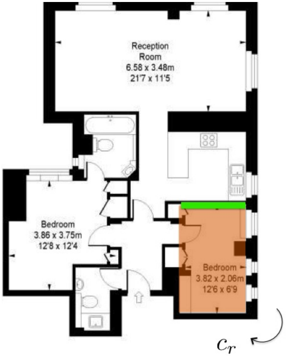
- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room
- $c_r \in \{1, \ldots, |C_r|\}$ ... a discrete variable representing within room r which wall the picture is facing ($|C_r|$ the number of walls in a room)

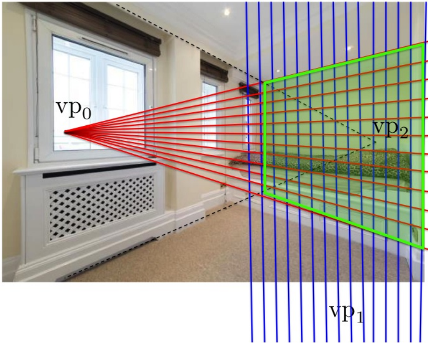




Formulation

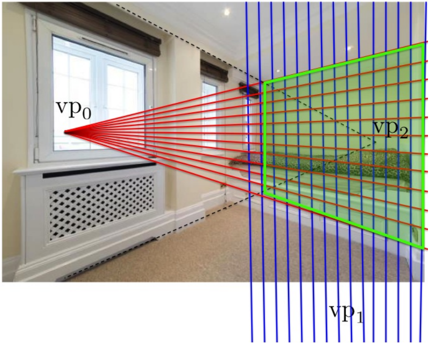
- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room
- $c_r \in \{1, \ldots, |C_r|\}$ ... a discrete variable representing within room r which wall the picture is facing ($|C_r|$ the number of walls in a room)





Formulation

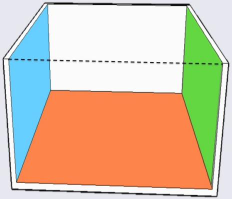
- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room
- $c_r \in \{1, \ldots, |C_r|\}$ ... a discrete variable representing within room r which wall the picture is facing ($|C_r|$ the number of walls in a room)

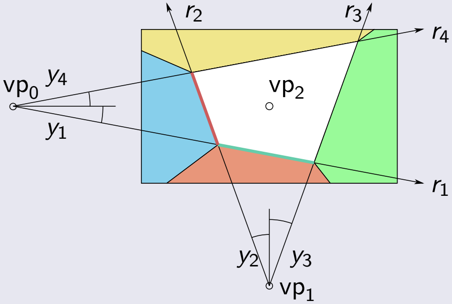


Formulation

- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room
- $c_r \in \{1, \ldots, |C_r|\}$ ... a discrete variable representing within room r which wall the picture is facing ($|C_r|$ the number of walls in a room)
- $\mathbf{y}$ ... rays representing a room layout

Typical parametrization for room layout [Hedau et al., 2009]:





- Room is a 3D cuboid
- $\mathbf{y} = (y_1, y_2, y_3, y_4)$
- 4 rays needed to define it

Formulation

- $r \in \{1, \ldots, R\}$ ... discrete random variable representing the room
- $c_r \in \{1, \ldots, |C_r|\}$ ... a discrete variable representing within room r which wall the picture is facing ($|C_r|$ the number of walls in a room)
- $\mathbf{y}$ ... rays representing a room layout
- We formulate the problem as inference in a Conditional Random Field with the following energy:

$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$

Energy Terms: Scene Type

$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$

- **Potential:** Score of a scene classifier predicting scene type (e.g., bedroom, kitchen, reception)

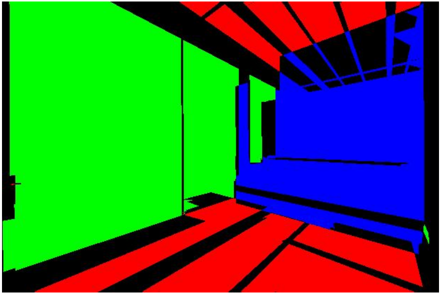
Energy Terms: Scene Type

$$E(r, c_r, \mathbf{y}) = {\color{magenta}E_{scene\_type}(r)} + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$

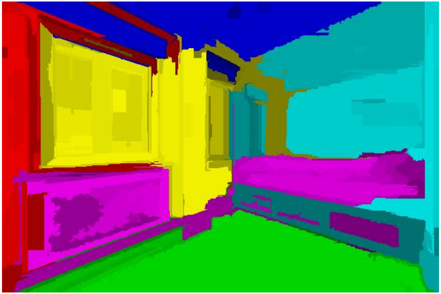
- **Potential:** Score of a scene classifier predicting scene type (e.g., bedroom, kitchen, reception)

Energy Terms: Layout

$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$



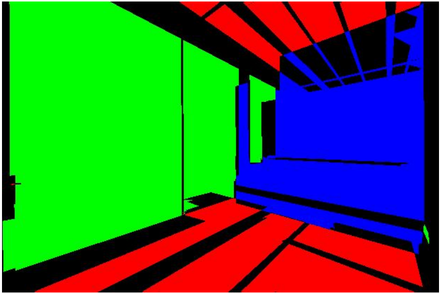
Orientation Map [Lee et al., 2009]



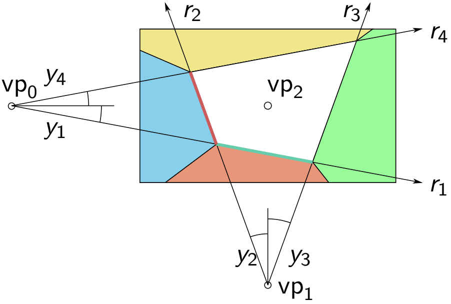
Geometric Context [Hedau et al., 2009]

Energy Terms: Layout

$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$

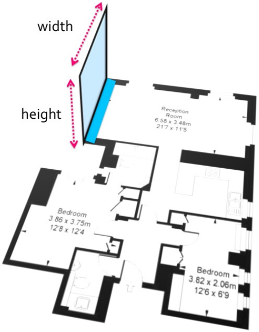


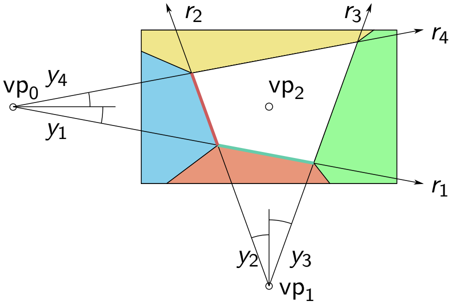
Orientation Map [Lee et al., 2009]



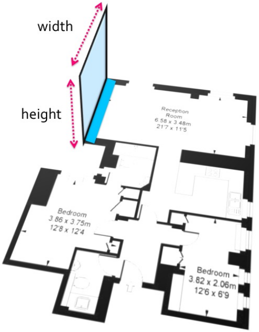
- **Potential**: Counts of blue, red, etc, pixels inside and outside of each wall
- Fast computation using *integral geometry* [Schwing et al., 2012]

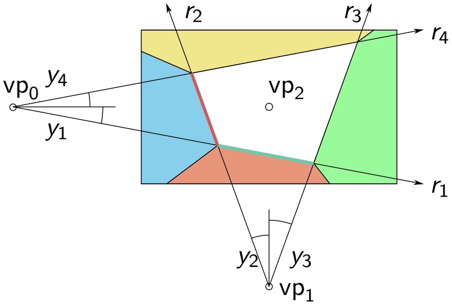
$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(\boxed{r, c_r}, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$





$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(\boxed{r, c_r}, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$

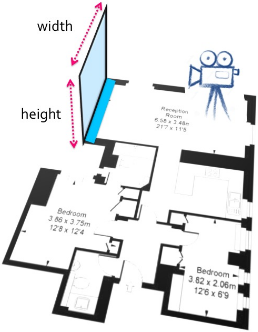


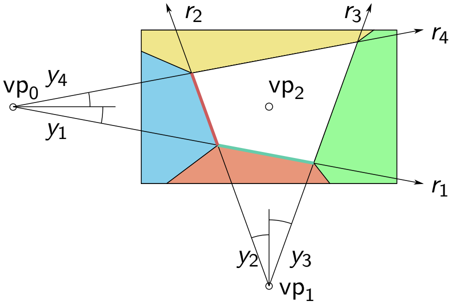


- $\mathbf{y} = (y_1, y_2, y_3, \cancel{y_4})$, $\quad y_4 = f(r, c_r, y_1, y_2, y_3)$

Energy Terms: Layout

$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(\boxed{r, c_r}, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$





- $\mathbf{y} = (y_1, y_2, y_3, \cancel{y_4})$, $\quad y_4 = f(r, c_r, y_1, y_2, y_3)$
- Additional constraint on $\mathbf{y}$: Camera is **inside** the room

Energy Terms: Windows

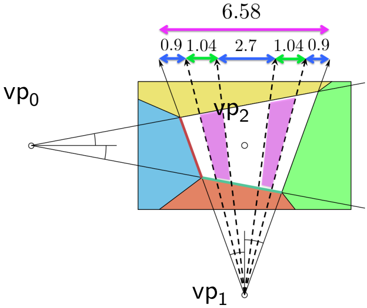
$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})$$

- Window-background segmentation

Energy Terms: Windows

$$E(r, c_r, \mathbf{y}) = E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(\boxed{r, c_r}, \mathbf{y})$$

- Window-background segmentation
- **Potential**: count window pixels inside and outside the window area



Learning and Inference

- We are minimizing the energy:

$$(r^*, c_r^*, \mathbf{y}^*) = \underset{r, c_r, \mathbf{y}}{\operatorname{argmin}} \ \left(E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y})\right)$$

Learning and Inference

- We are minimizing the energy:

$$(r^*, c_r^*, \mathbf{y}^*) = \underset{r, c_r, \mathbf{y}}{\operatorname{argmin}} \left(E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y}) \right)$$

- Inference:
  - Exhaustive enumeration of r and c_r
  - Exact branch and bound inference for $\mathbf{y}$ [Schwing & Urtasun, 2012]

Learning and Inference

- We are minimizing the energy:

$$(r^*, c_r^*, \mathbf{y}^*) = \underset{r, c_r, \mathbf{y}}{\operatorname{argmin}} \left(E_{scene\_type}(r) + E_{layout}(r, c_r, \mathbf{y}) + E_{win}(r, c_r, \mathbf{y}) \right)$$

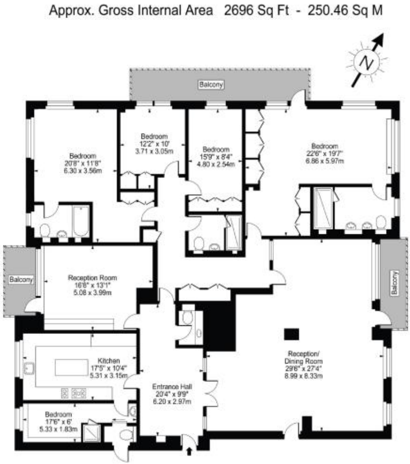
- Inference:
  - Exhaustive enumeration of r and c_r
  - Exact branch and bound inference for $\mathbf{y}$ [Schwing & Urtasun, 2012]
- We use S-SVM for training

Dataset

- We crawled a London apartment rental site

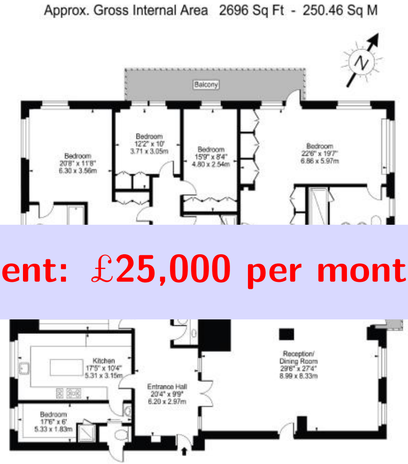
# apartments	215
# of images	1570
# of indoor images	1259
# images without GT alignment	82
avg. # rooms per apt	6
avg. # walls per apt	31
avg. # windows per apt	6
avg. # doors per apt	9

Apartments in Central London Are Not Small



Biggest apartment in dataset: 16 rooms, 5 bedrooms, 88 walls

Apartments in Central London Are Not Small



Rent: £25,000 per month

Biggest apartment in dataset: 16 rooms, 5 bedrooms, 88 walls.

Results: Layout Estimation

- We assume we know which wall the camera is facing
- **Metrics**: Pixel accuracy for predicting 5 walls

	Layout error	Evaluations	Test time [s]
Schwing'12	13.88	16012.4	0.0208
Ours	**11.81**	**1269.5**	**0.0019**

Results: Layout Estimation

- We assume we know which wall the camera is facing
- **Metrics**: Pixel accuracy for predicting 5 walls

	Layout error	Evaluations	Test time [s]
Schwing'12	13.88	16012.4	0.0208
Ours	**11.81**	**1269.5**	**0.0019**

- 2% reduction in layout error

Results: Layout Estimation

- We assume we know which wall the camera is facing
- **Metrics**: Pixel accuracy for predicting 5 walls

	Layout error	Evaluations	Test time [s]
Schwing'12	13.88	16012.4	0.0208
Ours	**11.81**	**1269.5**	**0.0019**

- 2% reduction in layout error
- 10 times less branching operations

Results: Layout Estimation

- We assume we know which wall the camera is facing
- **Metrics**: Pixel accuracy for predicting 5 walls

	Layout error	Evaluations	Test time [s]
Schwing'12	13.88	16012.4	0.0208
Ours	**11.81**	**1269.5**	**0.0019**

- 2% reduction in layout error
- 10 times less branching operations
- 10x speedup

Results: Camera Localization

- **Metrics**: % of correct assignments of front wall to the apartment wall

	Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (no windows)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

Results: Camera Localization

- **Metrics**: % of correct assignments of front wall to the apartment wall

	Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (no windows)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

Aspect: Only aspect ratio information (and not scene) used

Results: Camera Localization

- **Metrics**: % of correct assignments of front wall to the apartment wall

	Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (no windows)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

+*Scene*: Aspect information and scene classifier are used

Results: Camera Localization

- **Metrics**: % of correct assignments of front wall to the apartment wall

	Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (no windows)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

+*Room*: We know which room the picture was taken in

Results: Camera Localization

- **Metrics**: % of correct assignments of front wall to the apartment wall

	Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (no windows)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

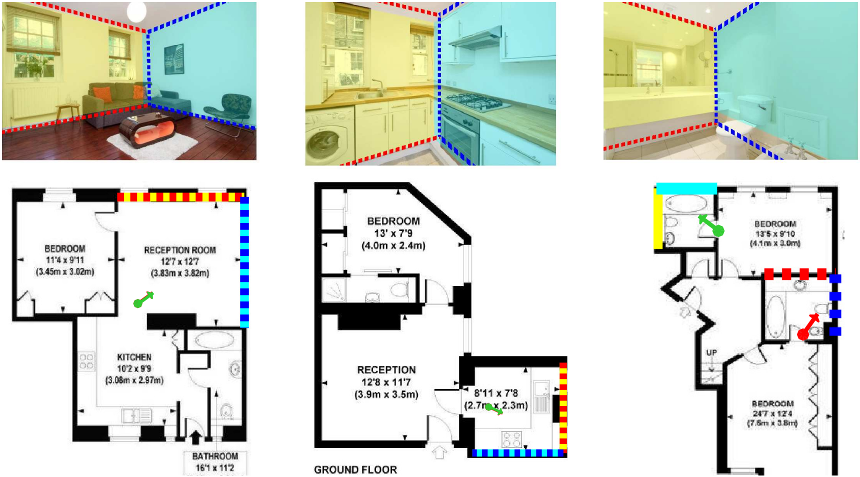
Results: Camera Localization

- **Metrics**: % of correct assignments of front wall to the apartment wall

	Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (no windows)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

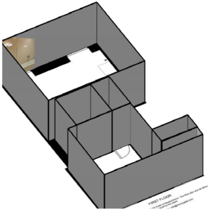
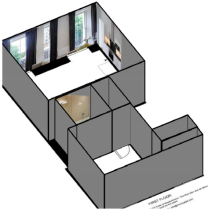
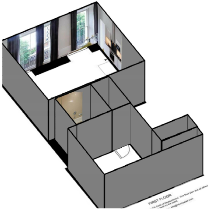
Results: Joint Layout and Localization

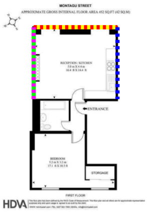
Red arrow: Groundtruth camera Green arrow: Predicted camera

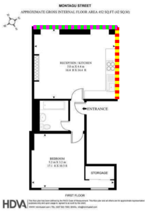


Red arrow: Groundtruth camera Green arrow: Predicted camera

Results: Reconstruction

Window+Aspect	+Scene	+Room	Ground-truth
			
1 images out of 4 2 walls out of 8	4 images out of 4 8 walls out of 8	4 images out of 4 8 walls out of 8	- -





Summary

- Problem of apartment 3D reconstruction from monocular imagery
- Model that jointly solves for localization and room layout estimation by exploiting floor-plans
- Real-time inference
- Results:
  - We improve layout prediction over past work
  - Achieve good localization performance
- Dataset with 215 apartments and all annotations available:

  `http://www.cs.toronto.edu/~fidler/projects/rent3D.html`

Alex on the Market Next Year

Thank You

Welcome to our poster at #9!