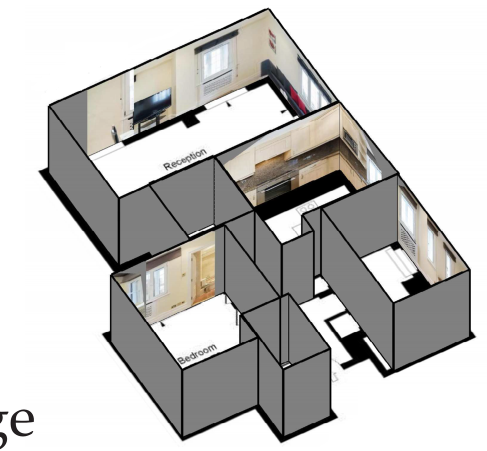
Rent3D: Floor-Plan Priors for Monocular Layout Estimation

Chenxi Liu[1,*] Alexander G. Schwing[2,*] Kaustav Kundu[2] Raquel Urtasun[2] Sanja Fidler[2]
* equal contribution [1]Tsinghua University [2]University of Toronto

Motivation

- 11.69% US citizens moved in 2012
- 10, 000 ads featuring few images per day for any big city
- Oftentimes a floor-plan illustrates the apartment layout

Enable 3D virtual tour of an apartment given
- Small set of monocular images
- Floor plan
- Rental metadata (ceiling height)



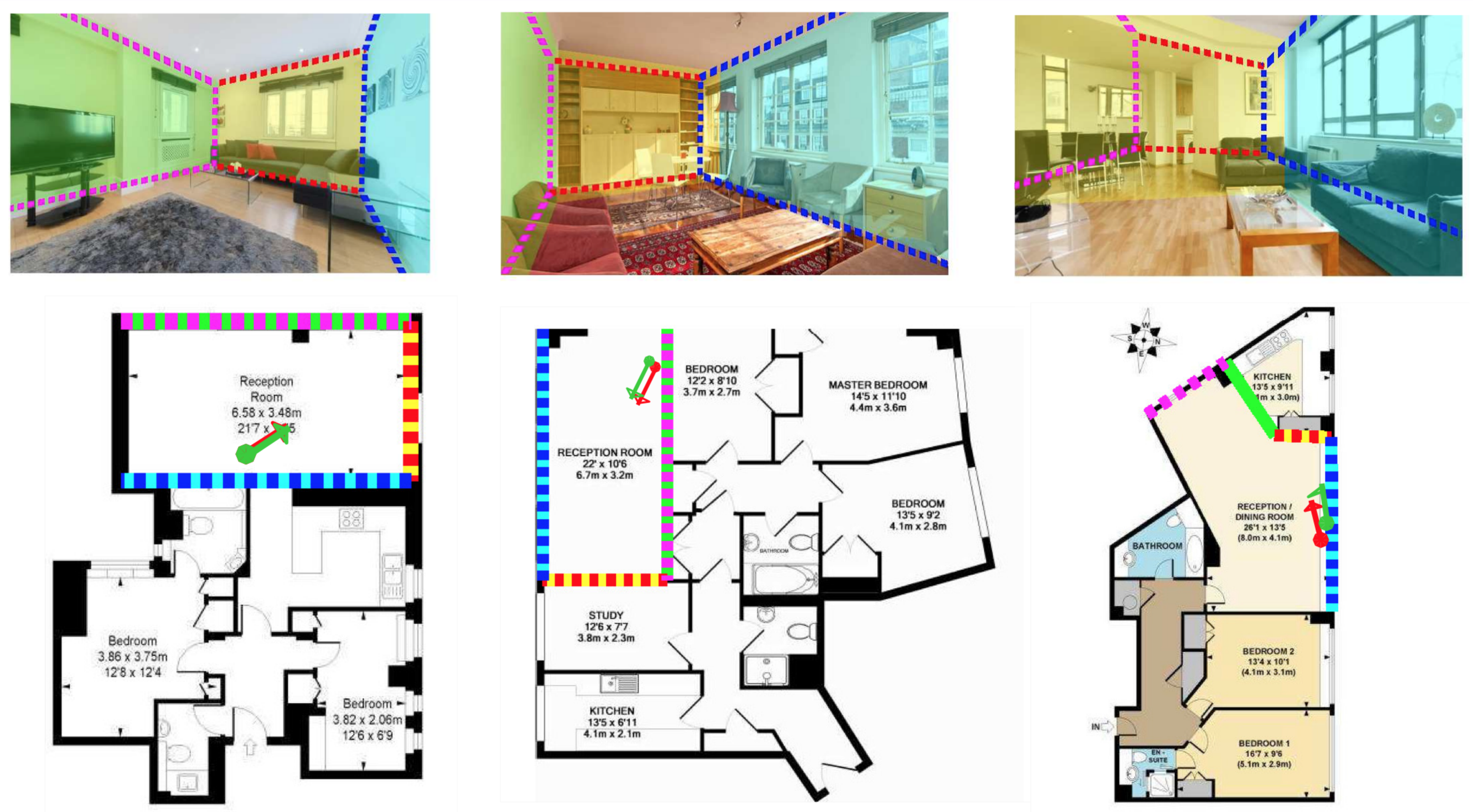
Contributions:
- Floor plans as a source of prior knowledge
- Localization of monocular image within apartment

Rent3D Apartment Dataset

Crawling a website featuring London rental ads providing a floor plan and at least one photo

Statistics:
- 215 apartments with 1312 rooms, 6628 walls, 1268 windows
- 1259 photos with 2 to 30 images per apartment

Annotations:
- Global scale from physical dimensions given in the plan
- Room outline, not necessarily rectangular
- Room type (reception, bedroom, kitchen, bathroom, outdoor)
- Position of doors and windows
- Wall and Ceiling annotations for photos
- Mapping of photo walls to floor plan walls

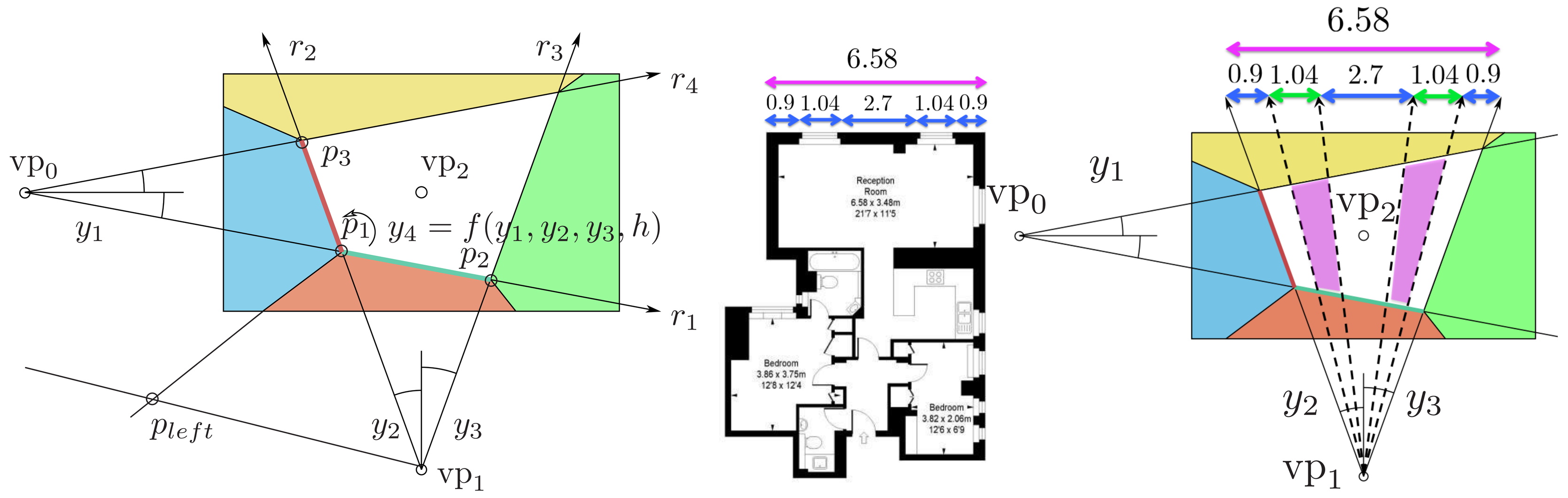


Publicly available:
http://www.cs.toronto.edu/~fidler/projects/rent3D.html

Layout Estimation with Floor Plans

Joint solution for two tasks:
- Estimation of room layout given monocular image
- Prediction of 3D pose (room & wall) relative to floor plan

Energy formulation:
- Variable representing room: $r \in \{1, \ldots, R\}$
- Variable denoting wall within room: $c_r \in \{1, \ldots, C_r\}$
- Camera location/orientation (Manhattan world): $y_{1,2,3}$



Aspect ratio permits parametrization with three variables

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

- Scoring fit of walls to orientation maps, geometric context and room interior: E_{lay}
- Window fit of layout $\mathbf{y}$ given r, c_r and pixel classifier: E_{win}
- Scene classifier score: E_{scene}

Algorithm: combines exhaustive search and branch&bound

for all r, c_r **do**
  put pair $(\bar{f}(\mathcal{Y}), \mathcal{Y})$ into queue and set $\hat{\mathcal{Y}} = \mathcal{Y}$
  repeat
    split $\hat{\mathcal{Y}} = \hat{\mathcal{Y}}_1 \times \hat{\mathcal{Y}}_2$ with $\hat{\mathcal{Y}}_1 \cap \hat{\mathcal{Y}}_2 = \emptyset$
    put pair $(\bar{f}(\hat{\mathcal{Y}}_1), \hat{\mathcal{Y}}_1)$ into queue
    put pair $(\bar{f}(\hat{\mathcal{Y}}_2), \hat{\mathcal{Y}}_2)$ into queue
    retrieve $\hat{\mathcal{Y}}$ having highest score
  until $|\hat{\mathcal{Y}}| = 1$
end for

Image and Floor Plan Geometry

Compute p_3 on ray (vp_1, p_1) to ensure aspect ratio constraint

$$\frac{||\hat{p}_2 - \hat{p}_1||}{||\hat{p}_3 - \hat{p}_1||} = a$$

- Obtain λ from $p_2 = p_1 + \lambda(p_1 - \mathrm{vp}_0)$
- Obtain camera intrinsics K and rotation R from vp
- Parametrize $p_3 = p_1 + \mu(p_1 - \mathrm{vp}_1)$
- From projective geometry:

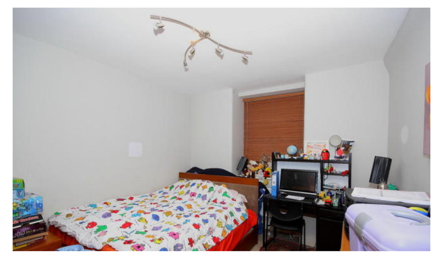
$$\mu = \frac{1}{\tilde{a} \cdot |1 + \frac{1}{\lambda}| - 1}, \quad \mu > 0, \tilde{a} = a\frac{||H\mathrm{vp}_0||}{||H\mathrm{vp}_1||}$$

with homography $H = KRK^{-1}$

Results

Scene classification:

	Recep.	Bedr.	Kitch.	Bathr	Outdr.
Recep.	0.82	0.14	0.03		
Bedr.	0.11	0.87		0.03	
Kitch.	0.06		0.89	0.05	
Bathr				1.00	
Outdr.		0.01			0.99



GT: Bedroom
Pred: Reception

GT: Bedroom
Pred: Reception

Pixelwise classification accuracy (window): Test set results
- 51.8% window prediction accuracy
- 95.6% prediction accuracy for background

Pixelwise classification error (layout prediction):

	Train [%]	Test [%]	Test Eval	Test Time [s]
(no floor plan), Transfer	16.70	17.00	22266.2	0.0208
(no floor plan), Train	13.57	13.88	16012.4	0.0150
Ours (aspect only)	11.80	11.81	1269.5	0.0019
Ours (windowGT)	11.78	11.79	1250.0	0.0026
Ours (window)	11.73	11.90	1258.7	0.0029

Localization accuracy:

	Accuracy, **Test**		
	Window+Aspect	+Scene	+Room
Random	0.0328	0.1138	0.1954
Ours (aspect only)	0.0686	0.1945	0.2654
Ours (windowGT)	0.2128	0.4737	0.5995
Ours (window)	0.1670	0.3982	0.5080

	Top5 Accuracy, **Test**		
	Window+Aspect	+Scene	+Room
Ours (aspect only)	0.2769	0.6682	0.8970
Ours (windowGT)	0.5858	0.8009	0.9519
Ours (window)	0.5492	0.7895	0.9474

Visual results: