Rent3D: Floor-Plan Priors for Monocular Layout Estimation

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

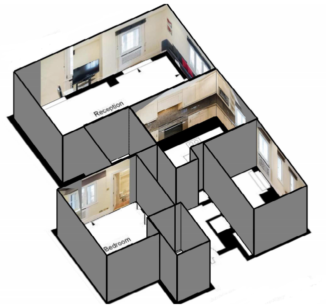

Figure 1: We reconstruct rental apartments from a set of monocular images and a floor plan.

The goal of this paper is to enable a 3D "virtual-tour" of an apartment given a small set of monocular images of different rooms, as well as a 2D floor plan. We frame the problem as the one of inference in a Markov Random Field which reasons about the layout of each room and its relative pose (3D rotation and translation) within the full apartment. This gives us information, for example, about which room the picture was taken in. What sets us apart from past work in layout estimation [1, 3, 4, 6] is the use of floor plans as a source of prior knowledge. In particular, we exploit the floor plan to impose aspect ratio constraints across the layouts of different rooms, as well as to extract semantic information, *e.g.*, the location of windows (which are labeled in floor plans) and scene type. We show that this information significantly helps in resolving the challenging layout estimation and camera localization problem. We also derive an efficient exact inference algorithm which takes only a few ms per apartment. This is due to the fact that we exploit integral geometry as well as our new bounds on the aspect ratio of rooms which allow us to carve the space, significantly reducing the number of physically possible configurations. We demonstrate the effectiveness of our approach on a new dataset which contains over 200 apartments.

New Dataset: Since our goal here is to reconstruct apartments in 3D, we collected a new dataset by crawling a rental website. The 215 apartments have in total 1312 rooms, 6628 walls, 1923 doors, and 1268 windows. The number of photos in each apartment ranges from 2 to 30, with the total number of photos in our dataset being 1259, not counting the outdoor images. We collected ground-truth for room layout in each image, scene type, room in which the photo was taken, and finally, which wall in the apartment the camera is facing. The dataset is available online: www.cs.utoronto.ca/∼fidler/projects/rent3D.html.

Layout Estimation: Our goal is to estimate an accurate layout for each room from a set of monocular images and to relate each layout to the full apartment via its 3D pose (rotation and translation). We phrase the problem as minimization of an energy E. In addition to well known layout cues such as orientation maps [3] and geometric context [1, 2] subsumed in an energy term E_{layout}, we exploit rich information contained in the apartment's floor plan as additional source of information. More specifically, we make use of the floor plan in order to retrieve information about aspect ratios of the walls for each room. The room layouts across images are linked by the fact that all rooms share the same height in 3D, which imposes aspect ratio constraints to the layout estimation problem, summarized in the energy term $E_{\text{as-ratio}}$. Further we also make use of semantics in the form of windows for which typically the floor plan contains additional ratio constraints. Windows are shown to be a very useful cue for the camera localization problem since they break the symmetry of parallel walls. We therefore add a term E_{win}. We additionally predict scene type for each image which helps us position the image into the correct room in the apartment. We use structured prediction [7, 8] to train our model and follow [5] in designing a branch-and-bound inference algorithm that runs in a real-time.

	Layout error	Evaluations	Test time [s]
[5]	13.88	16012.4	0.0150
Ours	**11.90**	**1271.5**	**0.0037**

Table 1: Layout estimation accuracy – pixel wise error

	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

Table 2: Localization error (correct assignment of front to apartment wall)

Results: In Tab. 1 we evaluate a setting in which we know which wall the camera is facing. In this case, our method uses the correct aspect ratio of the walls. We compare the effectiveness of using this cue against the baseline method [5]. Not only do we improve the prediction performance, but also are we able to reduce inference time by one order of magnitude. In Tab. 2 we evaluate the camera localization problem. We demonstrate how the localization error, *i.e.*, estimation of which wall the camera is facing, improves when incorporating one cue at a time: the aspect ratio information, window energy term E_{win} using perfect features ("windowGT") and using a trained classifier for "window". Here +Room denotes the setting in which the algorithm is given information in which room each photo was taken in. Fig. 2 shows a few qualitative examples.

Figure 2: Room layout estimation and wall alignment with our model. Top row show the estimated vertical walls (solid color), and GT layout (dashed lines). Bottom row shows the floor plans with predicted and GT alignments. Solid colors show alignment predicted by the model, where the yellow solid line denotes the front wall. GT alignment is shown with dashed lines.

[1] V. Hedau, D. Hoiem, and D. Forsyth. Recovering the Spatial Layout of Cluttered Rooms. In *Proc. ICCV*, 2009.

[2] D. Hoiem, A. A. Efros, and M. Hebert. Putting Objects in Perspective. *IJCV*, 2008.

[3] D. C. Lee, M. Hebert, and T. Kanade. Geometric Reasoning for Single Image Structure Recovery. In *Proc. CVPR*, 2009.

[4] L. Pero, J. Bowdish, D. Fried, B. Kermgard, E. Hartley, and K. Barnard. Bayesian geometric modeling of indoor scenes. In *Proc. CVPR*, 2012.

[5] A. G. Schwing and R. Urtasun. Efficient Exact Inference for 3D Indoor Scene Understanding. In *Proc. ECCV*, 2012.

[6] A. G. Schwing, S. Fidler, M. Pollefeys, and R. Urtasun. Box In the Box: Joint 3D Layout and Object Reasoning from Single Images. In *Proc. ICCV*, 2013.

[7] B. Taskar, C. Guestrin, and D. Koller. Max-Margin Markov Networks. In *Proc. NIPS*, 2003.

[8] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. *JMLR*, 2005.