SAMPLING BEDROOMS

Del Pero L., Guan J., Schlecht J., Barnard K. - University of Arizona
{delpero, jguan1,kobus}@email.arizona.edu, {schlecht}@uni-heidelberg.de

ICVSS 2010
Registration and Video Analysis

1. Abstract

We propose a top down approach to understanding indoor scenes such as bedrooms. Here we develop a generative statistical model for rooms imaged with perspective camera, where the room boundary and objects within it are approximated with simple non overlapping blocks. We determine fits to this model by combining several data-driven sampling techniques. We argue our representation has advantages over previous ones since the 3D geometry is extracted directly, and this often helps inference.

2. A generative model for indoor scene images

- We model rooms as blocks aligned in three orthogonal directions (Manhattan world)
- Blocks are right angled parallelepipeds defined in terms of width, height, length and 3D position

$$b = (w, h, l, x, y, z)$$

- Blocks approximate the 3D geometry of the room (room boundary) and the objects within it. We constrain blocks to lie on the floor (pieces of furniture) or to be attached to a wall (windows, picture frames)
- We model an image i as a collection of blocks imaged with a perspective camera

$$\theta_i = (c_i, r_{bi}, n_i, b_{i1}, .., b_{in_i}, \phi_i)$$

θ_i	model parameters for image i	c_i	camera parameters
r_{bi}	block modeling the room boundary	n_i	number of blocks in the scene, not known a priori
b_{ij}	jth block in image i	ϕ_i	Scene orientation with respect to world reference frame

3. Camera model

- Scene size and distance to the camera can be determined only up to a scale factor (camera position is kept fixed)
- Yaw is fully determined by the scene orientation ϕ, we assume no roll
- We constrain the camera to be inside the room

$$c = (f, \psi, s)$$

f focal length, $f \in [0, +\inf]$
ψ pitch angle, $\psi \in [0, \frac{\pi}{2}]$
s relative scale of the world

4. Image model

- We assume image edge points E_i to be generated by the projected contours of the objects in the scene [1]. We define the likelihood

$$p(i|\theta_i) = p(E_i|\theta_i) \approx e_{bg}^{N_{bg}} e_{miss}^{N_{miss}} \prod_{k=1}^{K_i} e(x_{ik})^{\epsilon_k}$$

K_i	number of pixels in image i
$e(x_k)$	probability of detecting an edge point at pixel x_k
ϵ_k	$\epsilon_k = 1$ iff the kth pixel is detected as an edge point
e_{bg}	p() of a pixel not being an edge point (background point)
e_{miss}	p() of a projected point not matched to any edge point
N_{bg}	number of background points in the image
N_{miss}	number of projected model point not matched to an edge

- Edge points are treated as independent
- We find the most likely correspondece between detected edge points and projected model points [1].
- The probability of a good match between a detected edge point x_k and a projected model point m_j is defined as

$$e(x_k) = \mathcal{N}(d_{kj}, 0, \sigma_d)\mathcal{N}(\phi_{kj}, 0, \sigma_\phi)$$

d_{kj}	distance between x_k and m_j along the gradient of m_j
ϕ_{kj}	difference in orientaion between detected and projected edge

5. Inference

- To find the set of parameters that best fit the observed image, we sample from the posterior distribution

$$p(\theta_i|i) = p(i|\theta_i)p(\theta_i)$$

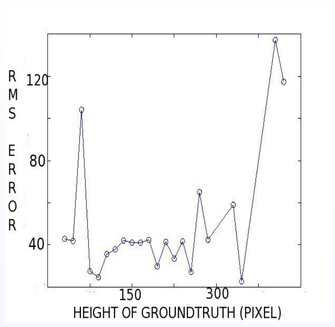
- We combine reversible jump Metropolis Hastings samples [8] for discrete changes in the model (number of blocks), and stochastic dynamics [7] to estimate continues parameter values in a particular parameter space.
- The room and the camera are fit simultaneously
- Block proposals are data-driven. We use detected image corners and vanishing points estimated from detected line segments [2,5]
- We sample over phase space [7] using energy function $E(\theta_i) = -log(p(\theta_i, i)) = -log(p(\theta_i|i) - log(p(i))$

6. Results

parameter	average error
f length	20%
pitch	0.12 (radian)

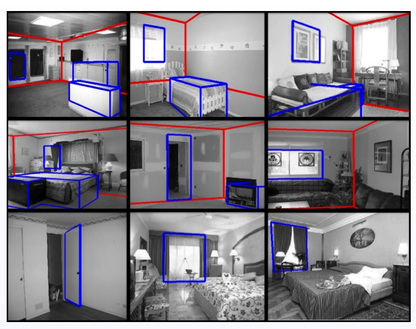
dataset	% correct orientation
no clutter (30)	78%
clutter (100)	68 %

Top row: Average error on estimated camera parameters, calculated on 30 images from the UCB room dataset. Bottom row: % of pixels labeled with the correct Manhattan world orientation [2,4,6], calculated on 30 images without hevey clutter, and on 100 cluttered images

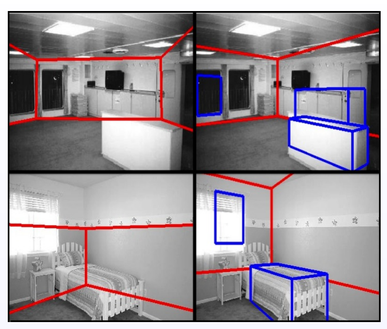


RMS difference between estimated floor boundary and groundtruth in pixel space [2]

Estimated room models backprojected onto the original image under the estimated camera parameters. Here we fit the room boundary alone, without objects



Top two rows: Examples of scenes where we found good blocks for objects (in blue). In most cases, doing so helped fit the room boundary. Bottom row: some additional detected frames



Two examples where adding blocks (right) improved the fit of room the boundary alone (left). Blocks help explain occlusions of room edges

Our results are comparable with previous work ([2],[4],[6]). Our key contribution is the top-down Bayesian approach, which is likely to prove more powerful as it integrates more information about color, texture and lighitng, and more sophisticated object models. These directions are the topic of ongoing research.

References

1. J., Schlecht, K., Barnard, Learning models of object structure , in *NIPS*, 2009
2. D.C. Lee, M., Herbert, T.,Kanade, Geometric reasoning for single image structure recovery , in *CVPR*, 2009
3. M., Hebert, A., Efros, D., Hoiem: Putting objects in perspective, in *IJCV*, 2008
4. E., Delage, H.L, Lee, A.Y, Ng, A Dynamic Bayesian Network Model for Autonomous 3D Reconstruction from a Single Indoor Image, in *CVPR*, 2006
5. C., Rother, A new approach to vanishing point detection in architectural environments, in *IVC*, Vol. 20, 2002
6. V., Hedau, D., Hoiem, D., Forsyth, Recovering the Spatial Layout of Cluttered Rooms, in *IEEE ICCV*, 2009
7. R.M., Neal, Probabilistic Inference Using Markov Chain Monte Carlo Methods, Technical report, 1993
8. P., Green, Reversible jump markov chain monte carlo computation and bayesian model determination in *Biometrika* 82, 1995