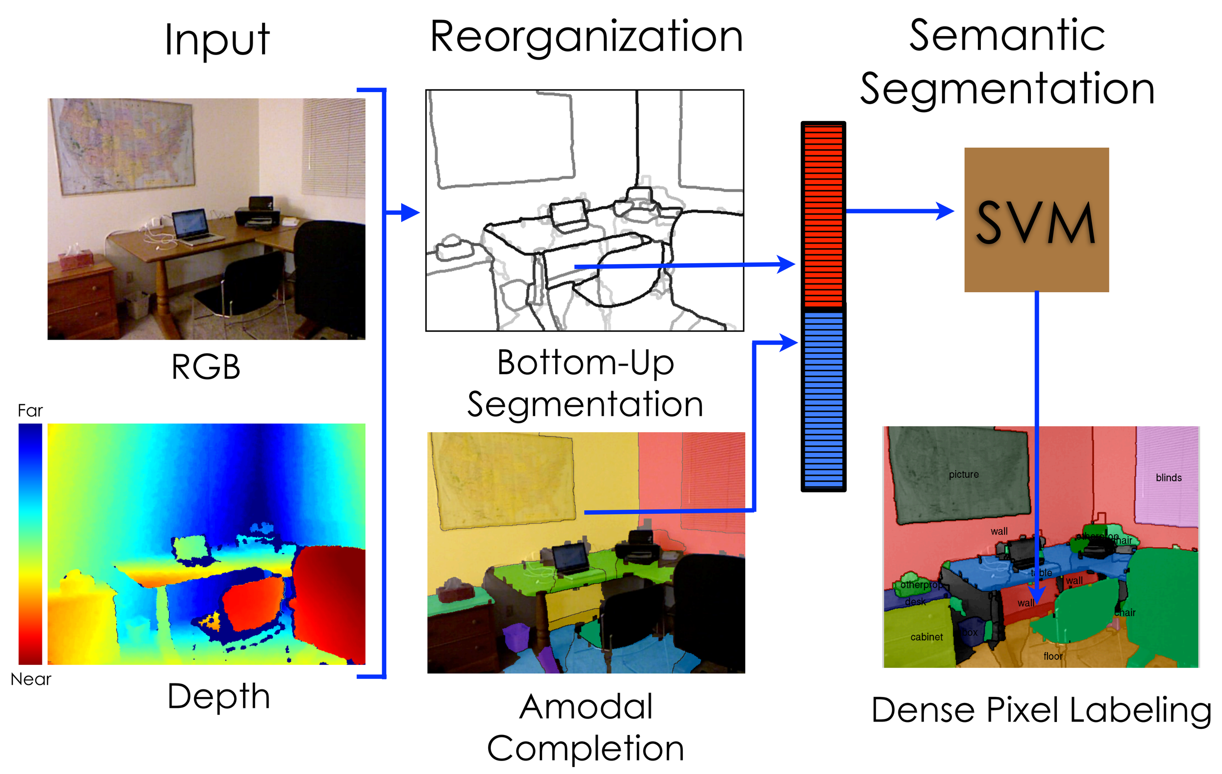
Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images

Saurabh Gupta, Pablo Arbeláez, Jitendra Malik
UC Berkeley

Overview



Related Work

Anand et al, IJRR12, Contextually Guided Semantic Labeling and Search for 3D Point Clouds

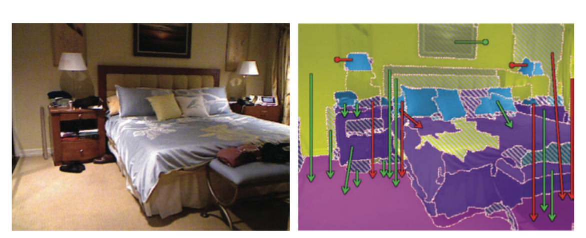
Modeled context using structural SVMs for semantic segmentation in full 3D Scenes

Silberman et al, ECCV12, Indoor segmentation and support inference from RGBD images.

Bottom-up and semantic segmentation, and inference of support relations in RGBD Scenes

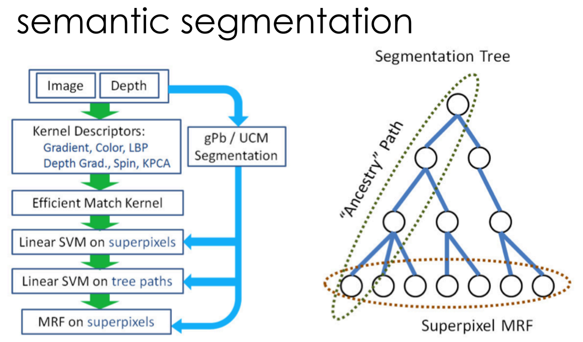
- Also introduced a RGBD dataset (NYUD2) with semantic segmentation labels

- Looked at a 4 class semantic segmentation (floor, structure, furniture, props)



Ren et al, CVPR12, RGB-(D) scene labeling: Features and algorithms.

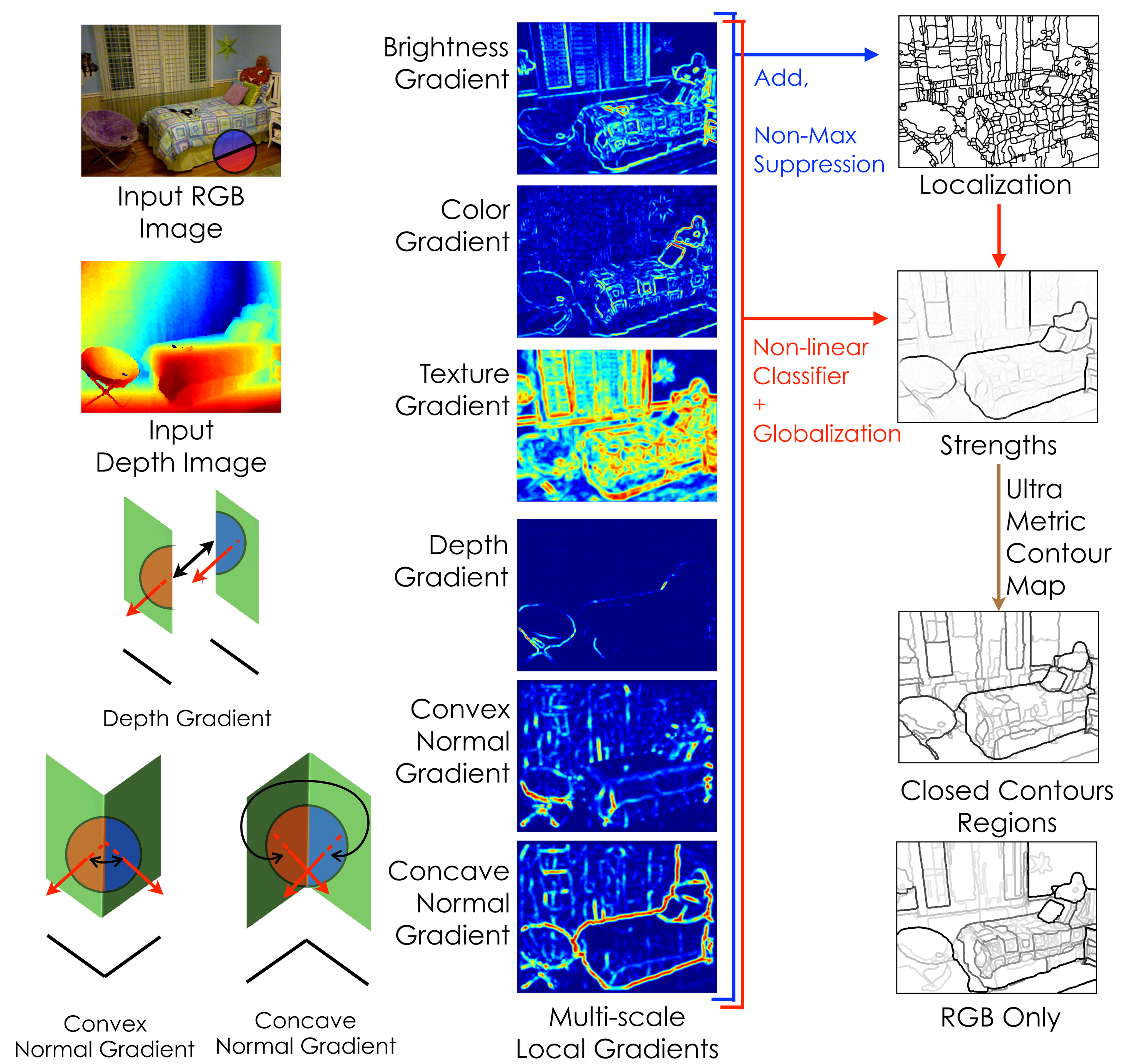
Using Kernel descriptor features and Tree path context for semantic segmentation



(a) Flowchart (b) Tree path + MRF

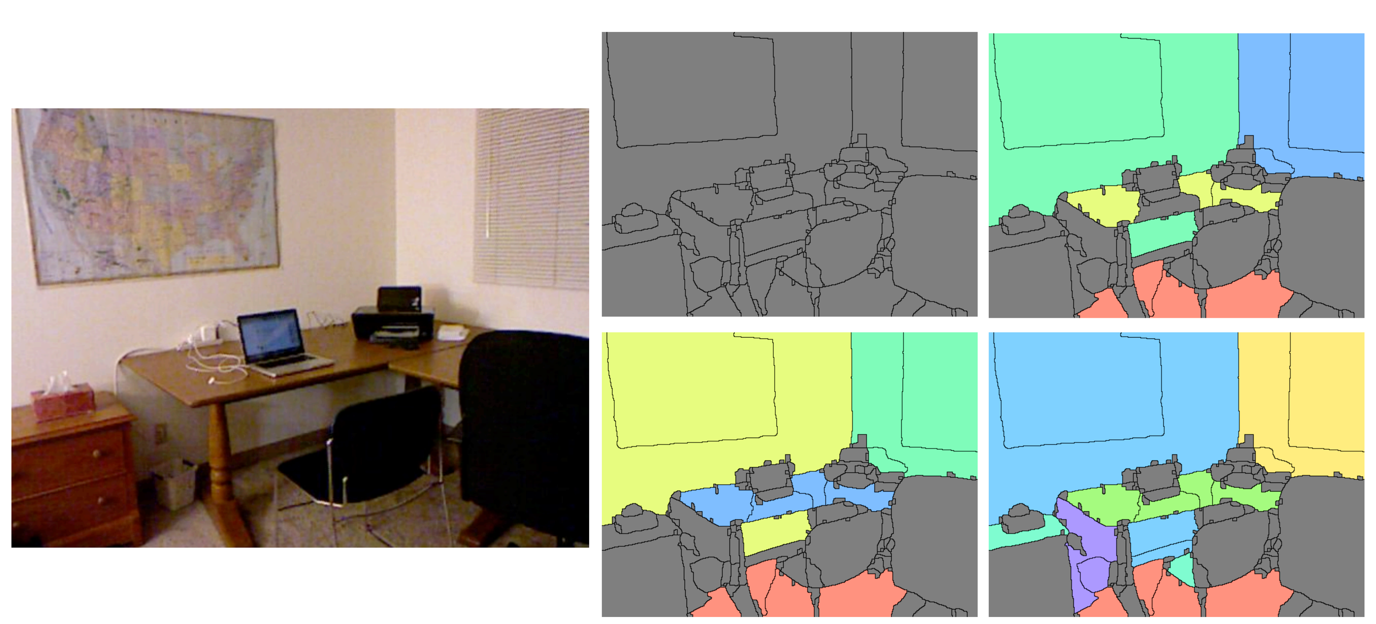
Bottom Up Segmentation

The Berkeley gPb detector (Arbeláez et al. 2011), uses brightness, color and texture gradients to find edges and regions. Here we augment it with depth data. We do this by local planar fits in oriented half-disks, and measuring depth and orientation differences.



Amodal Completion

Completing contiguous surfaces behind occluders



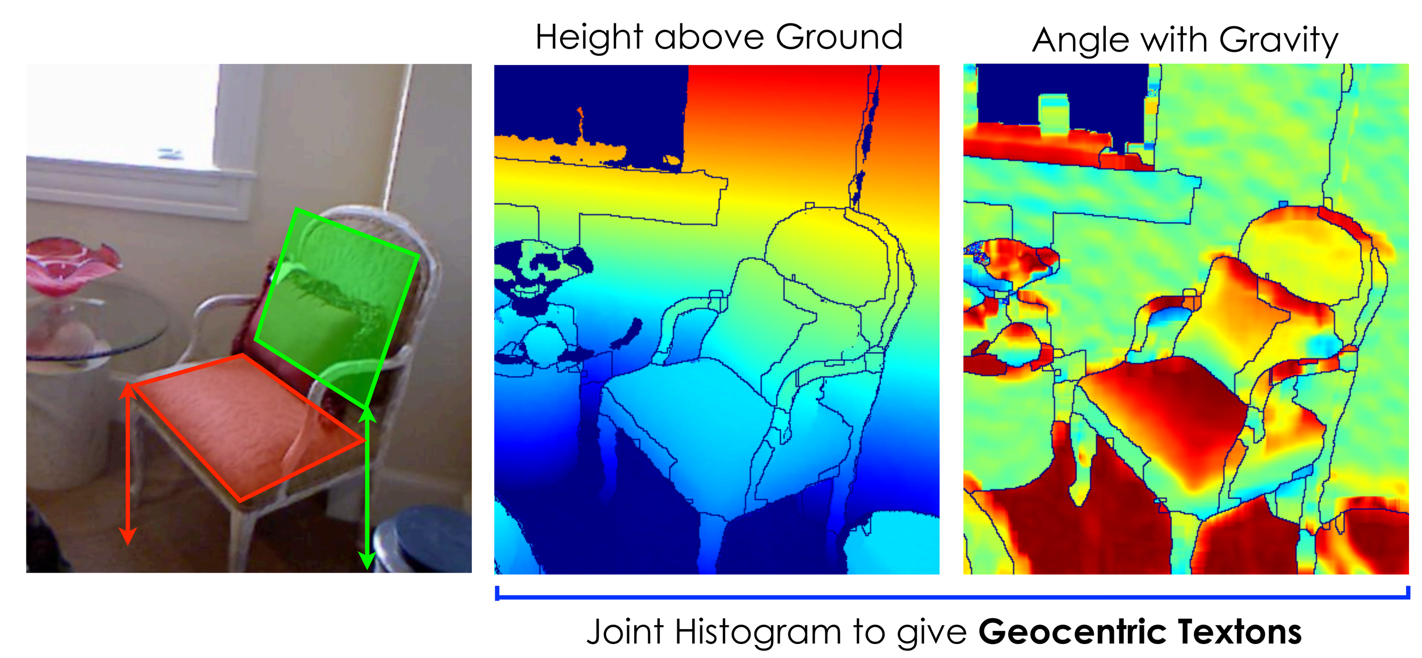
Semantic Segmentation

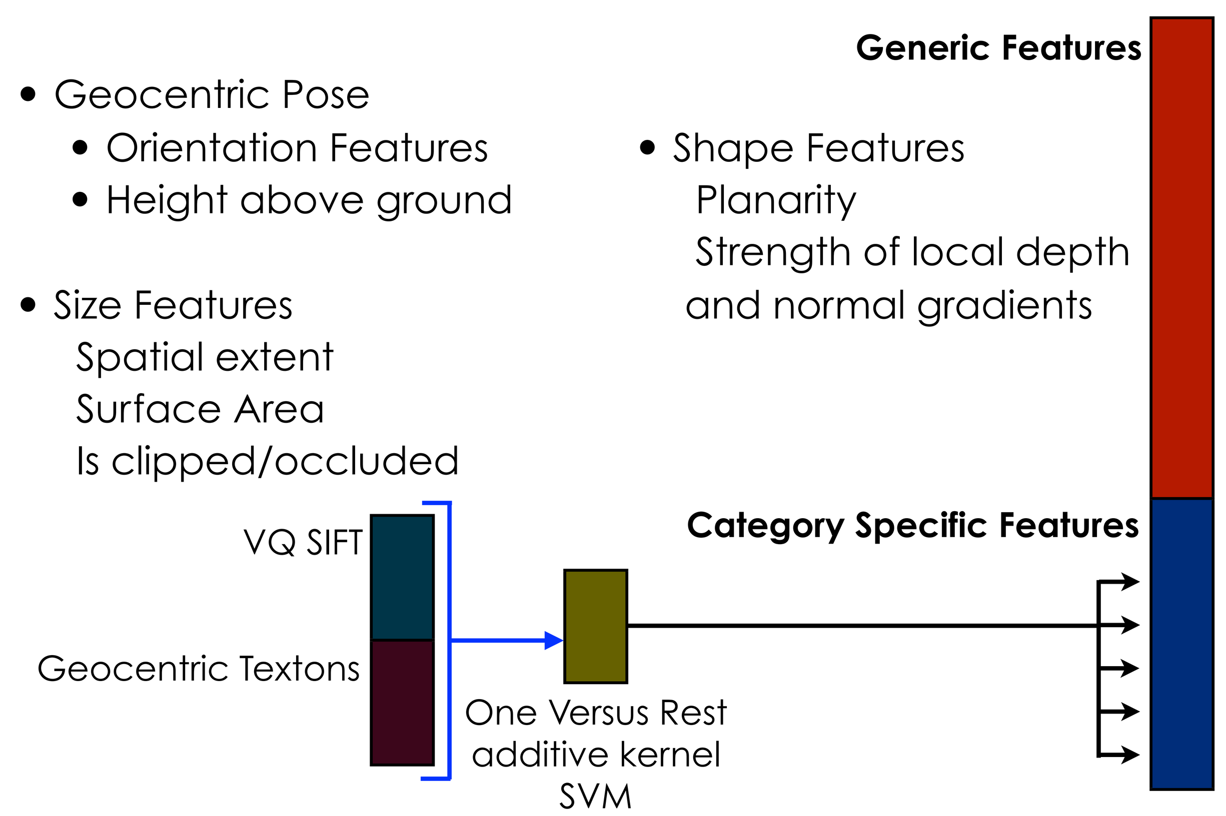
Frame it as a superpixel classification task.

Use features from both the superpixel and its amodal completion.

Features

- Appearance - SIFT on La*b*, orientation energy
- Geometry - Absolute size, height above ground and angle with gravity ...





Gravity Estimation

Estimate the direction of gravity from the depth image

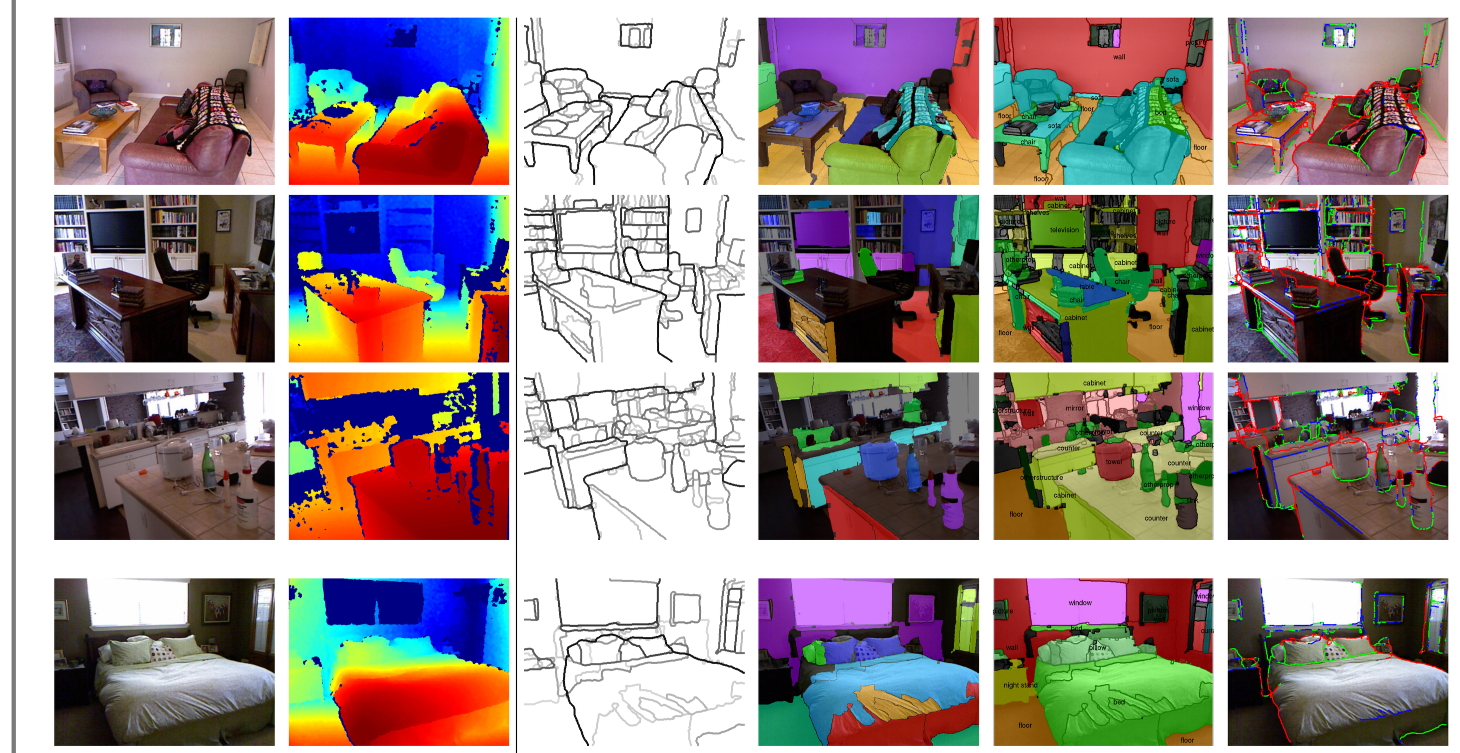
Find the direction as perpendicular to or as parallel to local normals at as many points as possible.

$$\min_{\mathbf{g}:||\mathbf{g}||_2=1} \underbrace{\sum_{\mathbf{n}:\theta(\mathbf{n},\mathbf{g})\geq d} cos^2\left(\theta(\mathbf{n},\mathbf{g})\right)}_{\text{as perpendicular to vertical surfaces}} + \underbrace{\sum_{\mathbf{n}:\theta(\mathbf{n},\mathbf{g})\leq d} sin^2\left(\theta(\mathbf{n},\mathbf{g})\right)}_{\text{as parallel to horizontal surfaces}}$$

$$\text{where, } \theta(\mathbf{n},\mathbf{g}) = \text{Angle between } \mathbf{g}, \mathbf{n}.$$

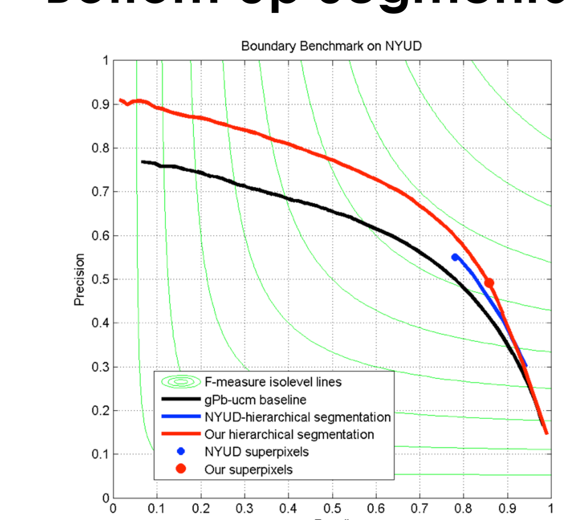
Simplifies to an eigen value problem of a 3x3 matrix

Results



(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Bottom Up Segmentation



Precision Recall Curve on Contours

Contours

		ODS	OIS	AP
Arbeláez etal	RGB	0.62	0.65	0.55
Silberman etal	RGBD	0.65	0.65	-
Our	RGBD	**0.69**	**0.71**	0.70

Regions

		ODS	OIS	bestC
Arbeláez etal	RGB	0.55	0.60	0.69
Silberman etal	RGBD	0.61	0.61	0.63
Our	RGBD	**0.62**	**0.67**	**0.75**

Semantic Segmentation

Aggregate Performance (Freq Wt I/U)

	Silberman etal	Ren etal	Our (RF)	Our (SVM)	Our(RF + Scene)	Our(SVM + Scene)
4 class task	56.31	59.19	64.36	64.81	64.97	64.9
40 class task	38.23	37.64	40.88	43.98	43.01	45.29

Performance (I/U)

	Silberman et al.	Ren et al.	Our
wall	61	60	**68**
floor	78	75	**81**
cabinet	33	37	**48**
bed	40	42	**55**
chair	32	33	**40**
sofa	25	28	**44**
table	21	17	**30**
door	5.9	**13**	8.3
window	30	28	**33**
bookshelf	**23**	17	20

	Silberman et al.	Ren et al.	Our
picture	36	32	**40**
counter	33	39	**47**
blinds	40	27	**44**
desk	4.6	**10**	**10**
shelves	3.3	**6.1**	5.1
curtain	27	28	**34**
dresser	13	7	**22**
pillow	19	20	**28**
mirror	4.4	18	**19**
floor mat	7.2	20	**22**

Scene Classification

Using a SPM on predicted labels, we can correctly classify 58% of scenes into categories like bedroom, living room, kitchen, office, ...

Acknowledgements

This work was sponsored by ONR SMARTS MURI N00014-09-1-1051, ONR MURI N00014-10-10933, and a Berkeley Fellowship.