Data-Driven 3D Primitives for Single Image Understanding

David Fouhey, Abhinav Gupta, Martial Hebert

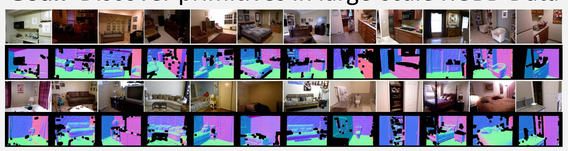
Carnegie Mellon University — The Robotics Institute

Task: 3D Understanding

What are the right primitives?

Our answer: any region that is
Visually ***Discriminative***
Geometrically ***Informative***

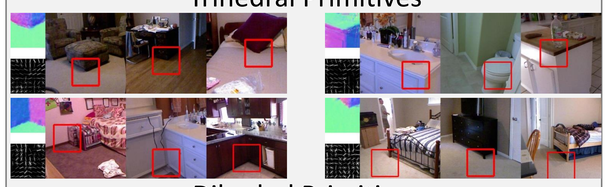
Goal: Discover primitives in large-scale RGBD Data



Results

(NYU v2)

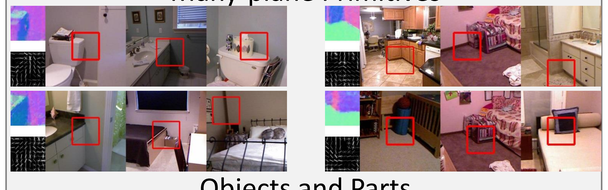
Trihedral Primitives



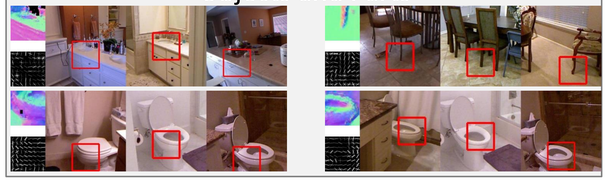
Dihedral Primitives

Planar Primitives

Many-plane Primitives



Objects and Parts



Learning

Components: Detector (w) | Canonical Form (N) | Primitive Instances (y)

Formulation

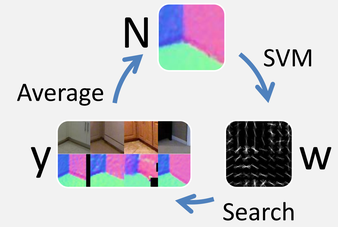
$$\min_{\mathbf{y},\mathbf{w},\mathbf{N}} \quad R(\mathbf{w}) \qquad \text{Regularization}$$

$$+\sum_{i=1}^{m}\Big[c_1 y_i \Delta(\mathbf{N},\mathbf{x}_i^G) \qquad \text{Geometric Consistency Enforces: } \textbf{\textit{Informative}}$$

$$+ c_2 L(\mathbf{w},\mathbf{N},\mathbf{x}_i^A,y_i)\Big] \qquad \text{Misclassification Loss Enforces: } \textbf{\textit{Discriminative}}$$

Hard to optimize directly – use iterative approach
Alternates ***discriminative*** (learning detector) and ***informative*** (updating canonical form)

Iterative Solution

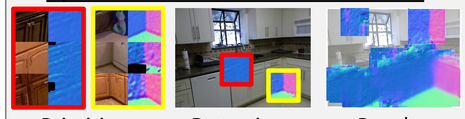


Search: Scan training set for top detections
Average: Average per-pixel surface normals
SVM: Train a linear SVM to detect instances

*Initialization (**y**):* Cluster hundreds of thousands of random patches in normal and HOG space.

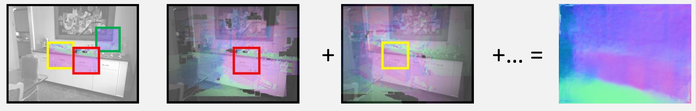
Inference

Sparse: Transfer canonical form



Primitives | Detections | Results

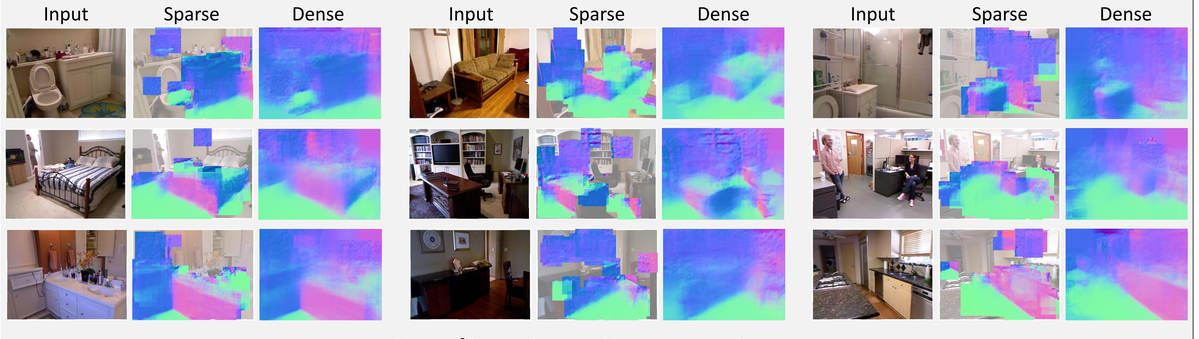
Dense: Transfer patch context



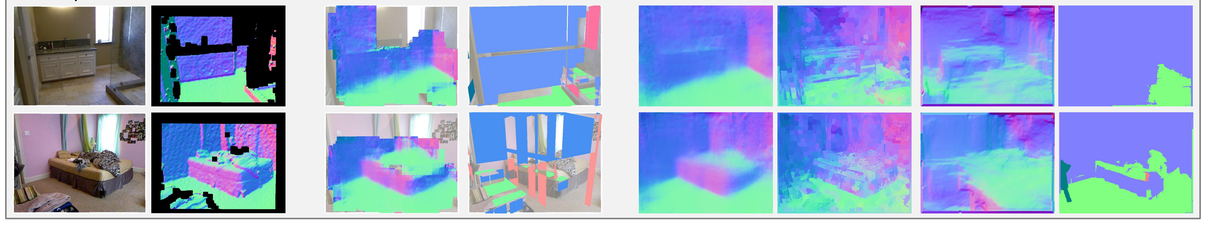
Detections | Averaged patch contexts | Results

Qualitative Results

(NYU Depth v2 Dataset)



Qualitative Comparison

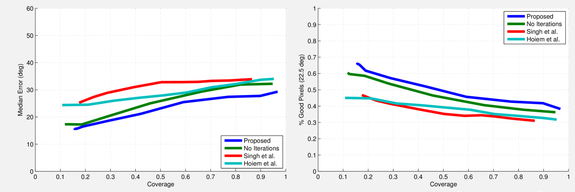


Quantitative Results

4-way split on NYU Depth v2
Per-pixel evaluation criterion: angular error

		Manhattan-World Techniques			Non-Manhattan-World Techniques				
		Lee '09	Hedau '10	3DP	Karsch '12	Saxena '08	Hoiem '07	Singh '12	3DP
Low Better	Mean (°)	44.9	41.2	**33.5**	40.8	47.1	41.2	35.0	**33.0**
	Median (°)	34.6	25.5	**18.0**	37.8	42.3	34.8	32.4	**28.3**
	RMSE (°)	54.8	55.1	**46.6**	46.9	56.3	49.3	40.6	**40.0**
High Better	Pct. <11.25°	24.8	33.2	**34.7**	7.9	11.2	9.0	11.2	**18.8**
	Pct. <22.5°	40.5	47.7	**55.0**	25.8	28.0	31.7	32.1	**40.7**
	Pct. <30°	46.7	53.0	**61.2**	38.2	37.4	43.9	45.8	**52.4**

Performance (Error) vs. Coverage (% Pixels Predicted)

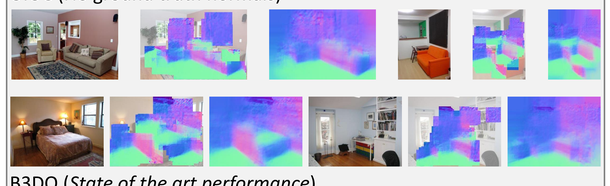


Mean Error | % Pixels < 22.5 (Precision)

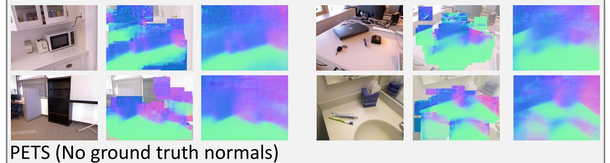
Cross-Dataset Results

Train on NYU, test on other data with identical settings.

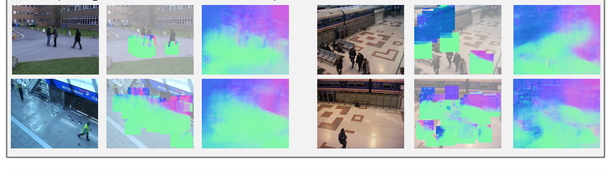
UIUC (No ground truth normals)



B3DO (*State of the art performance*)



PETS (No ground truth normals)



Code Available!
tinyurl.com/3DPrimitives
Already in use at CMU as a feature!