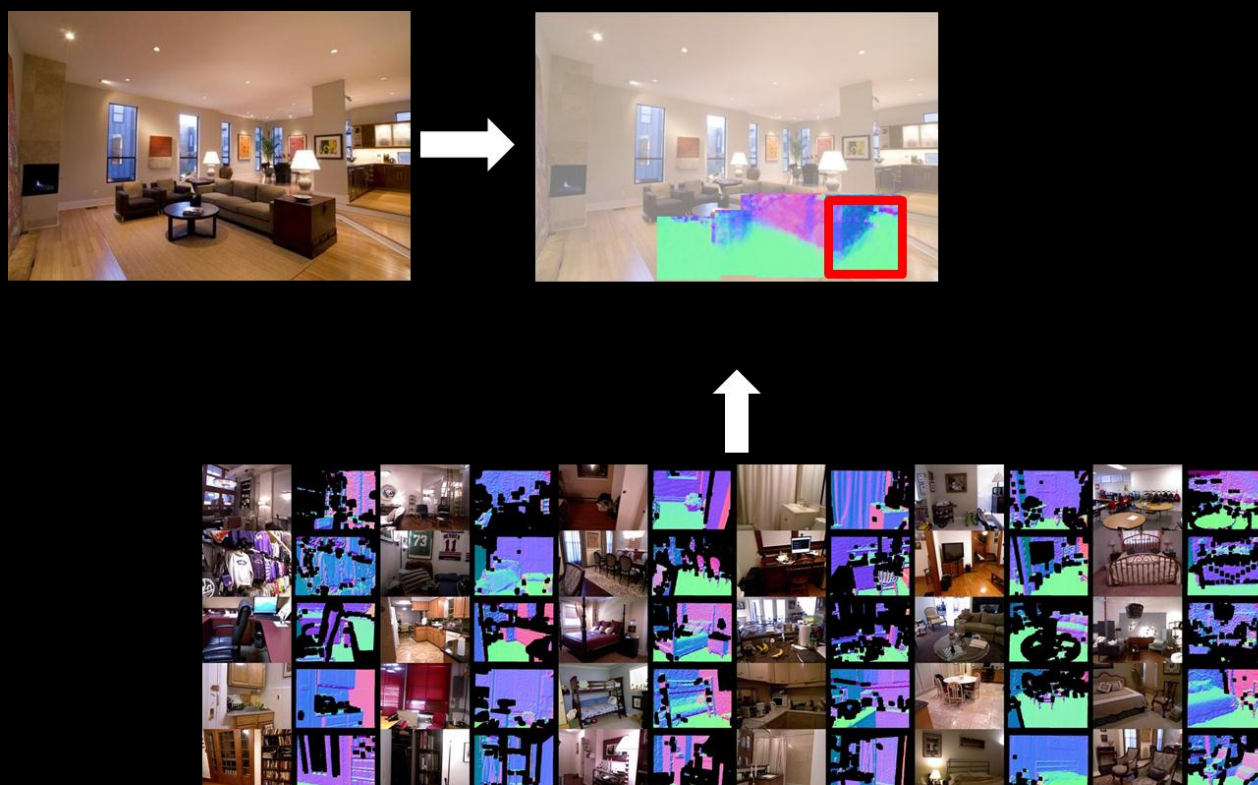
Using 3D data for image interpretation and geometric reasoning

Martial Hebert
Abhinav Gupta
David Fouhey, Adrien Matricon, Wajahat Hussain

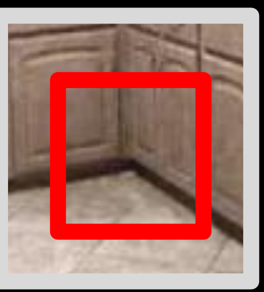
- Sparse mid-level primitives can be used to transfer geometric information?
- Can this helps in detection and matching tasks?
- Geometric reasoning can use this local evidence to produce a consistent geometric interpretation?

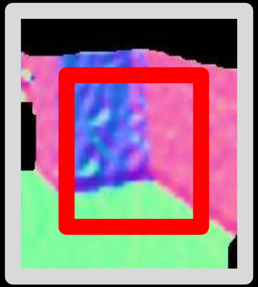


Primitives

Visually Discriminative

Geometrically Informative

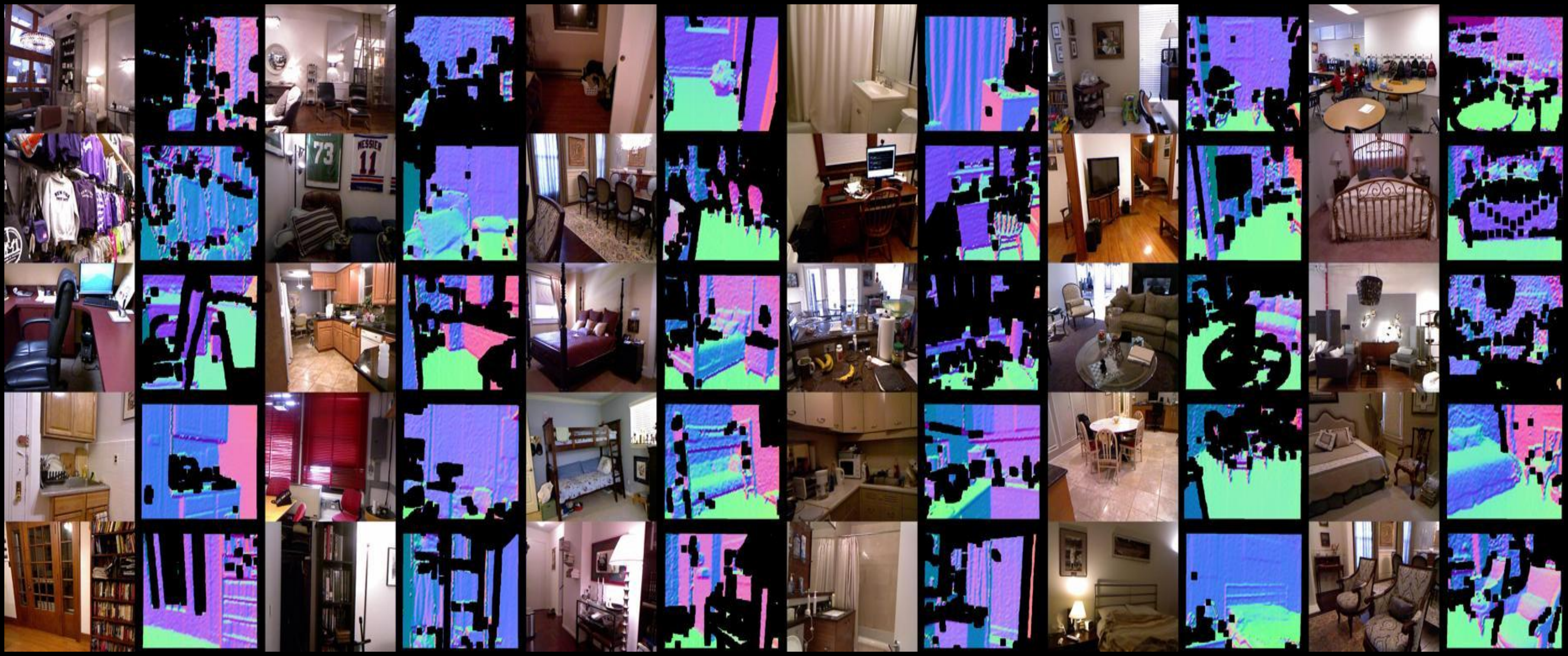




Image

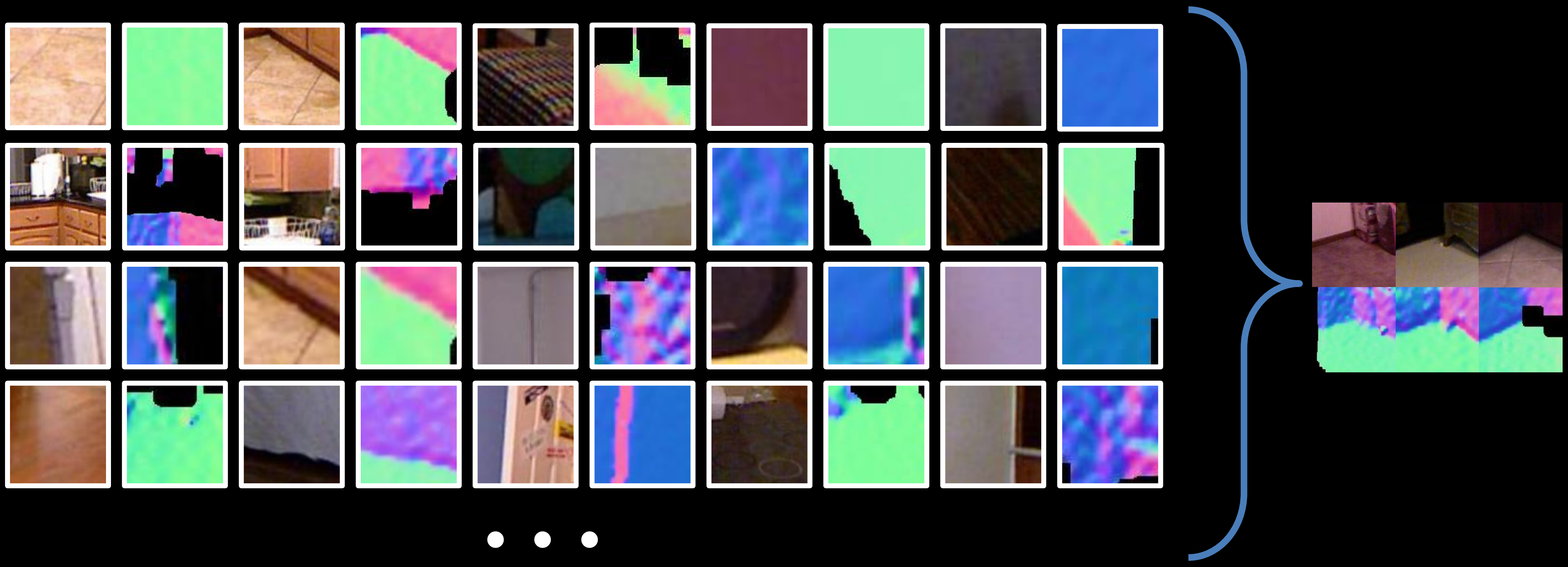
Surface Normals

Saurabh Singh et al. *Discriminative Mid-Level Patches*



NYU v2 Dataset (Silberman et al., 2012)

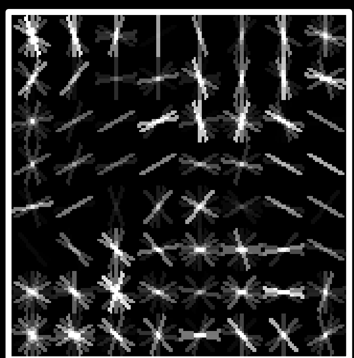
Learning primitives

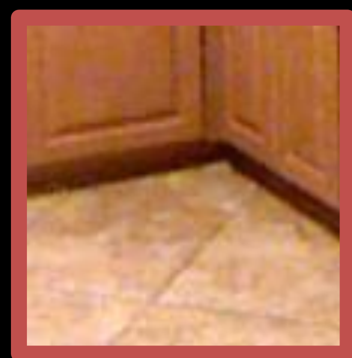


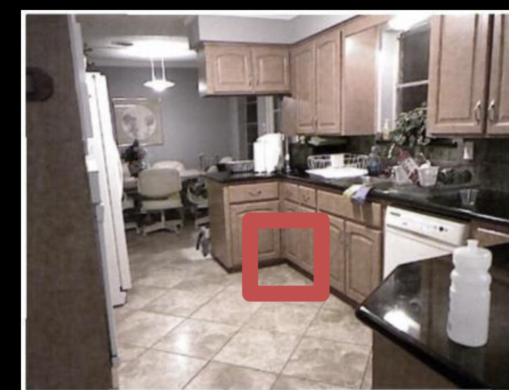
Representation

Detector

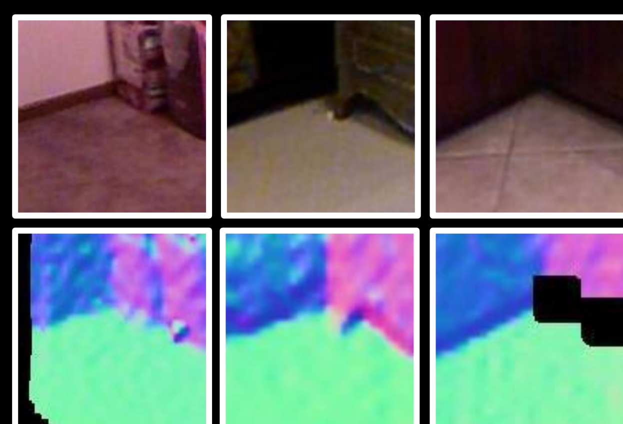
Instances

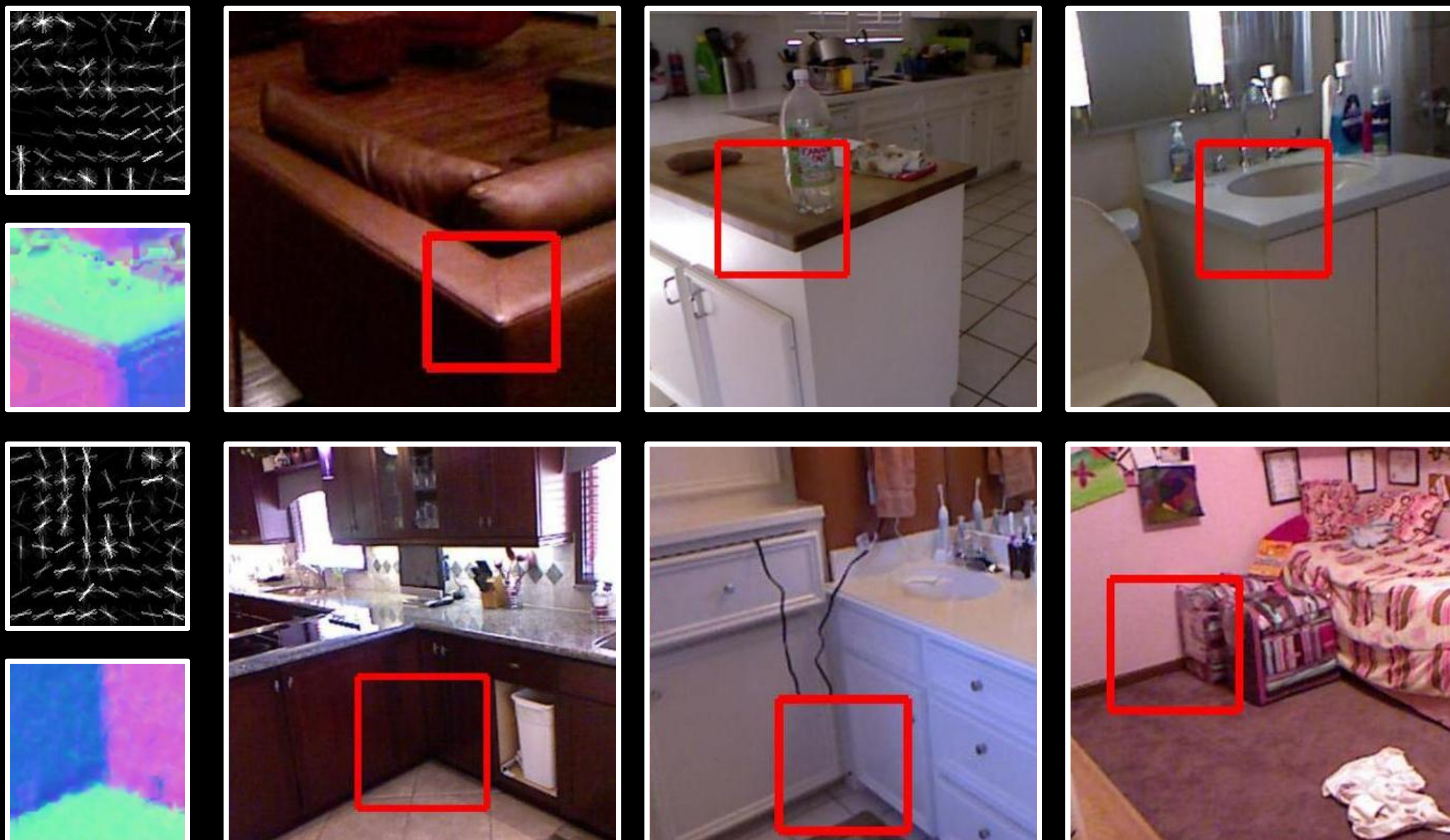


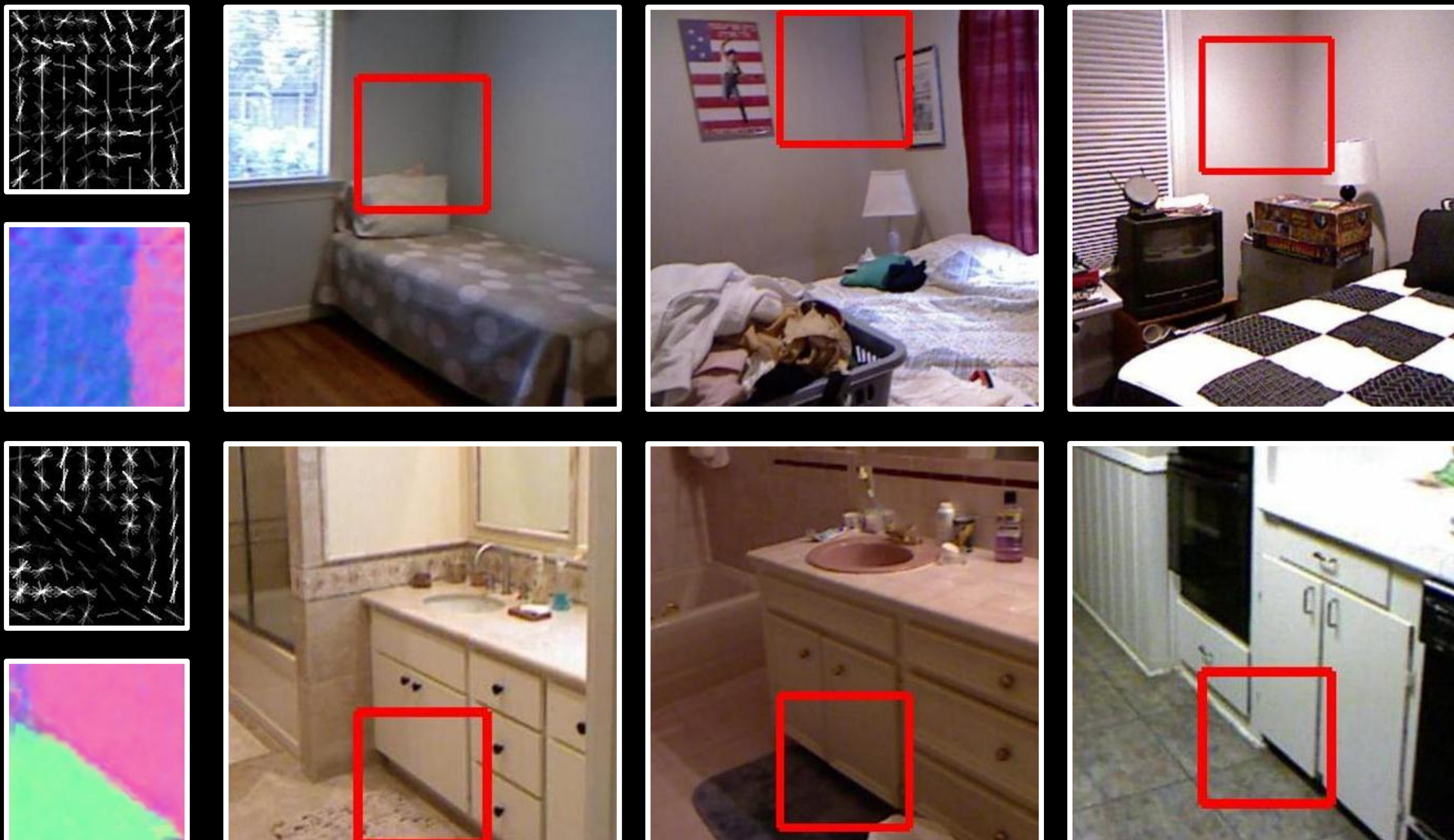




Canonical Form

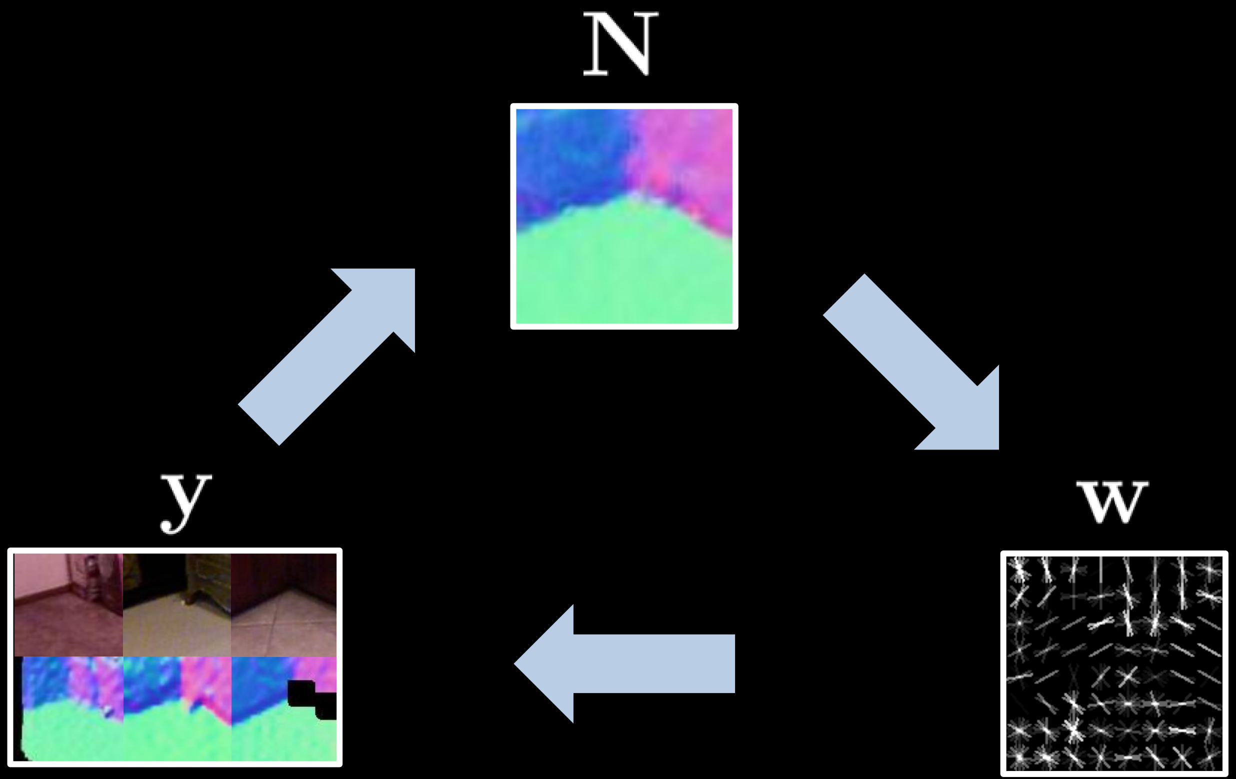






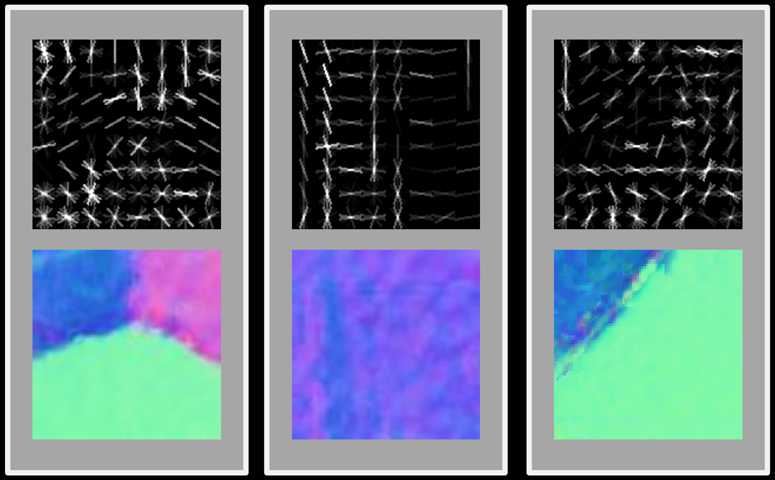
Learning Primitives

Approach: iterative procedure



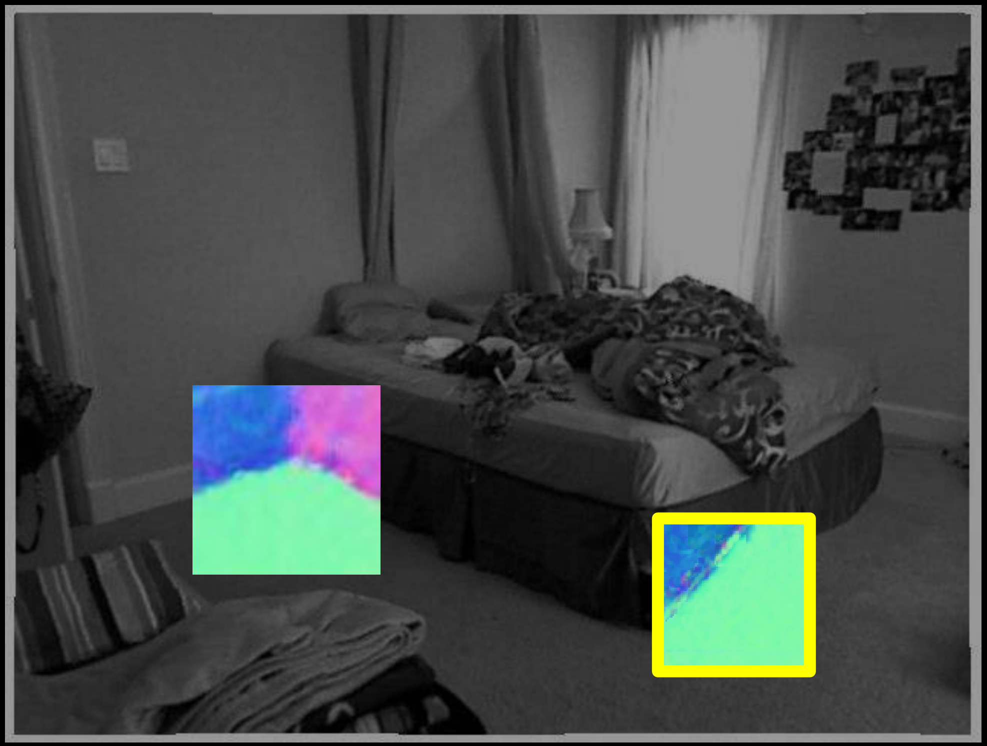
Inference

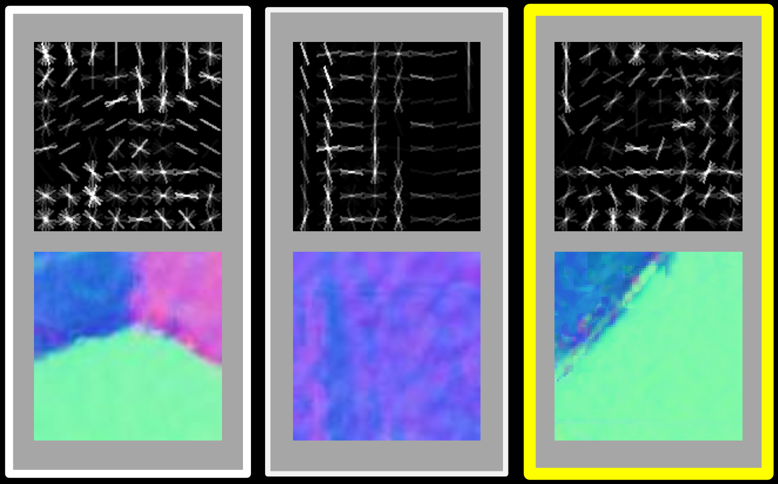
Sparse Transfer



Inference

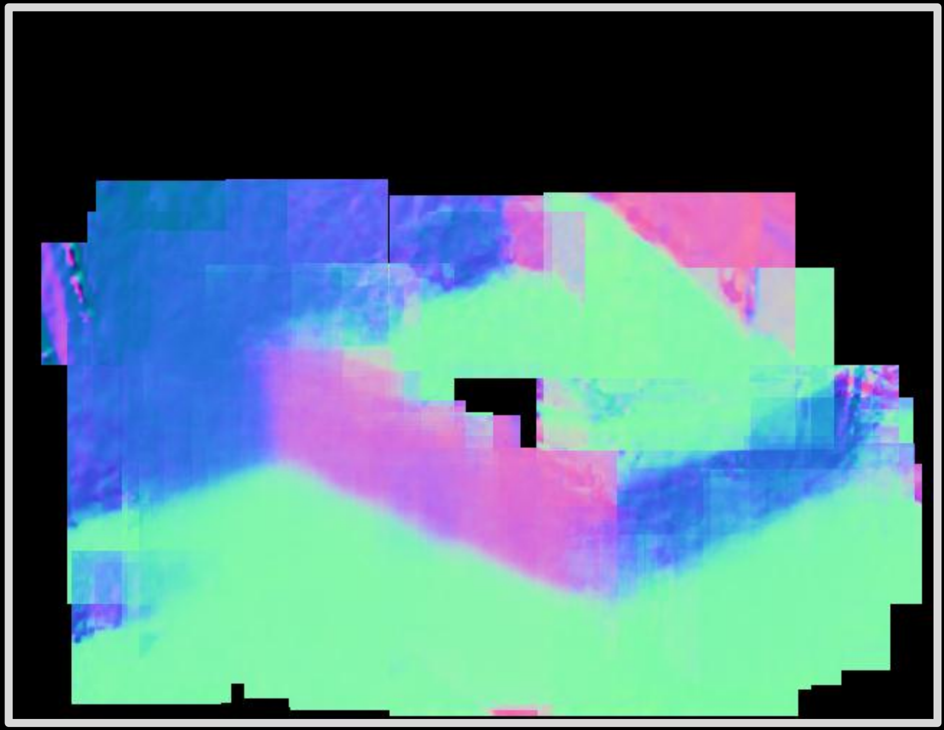
Sparse Transfer





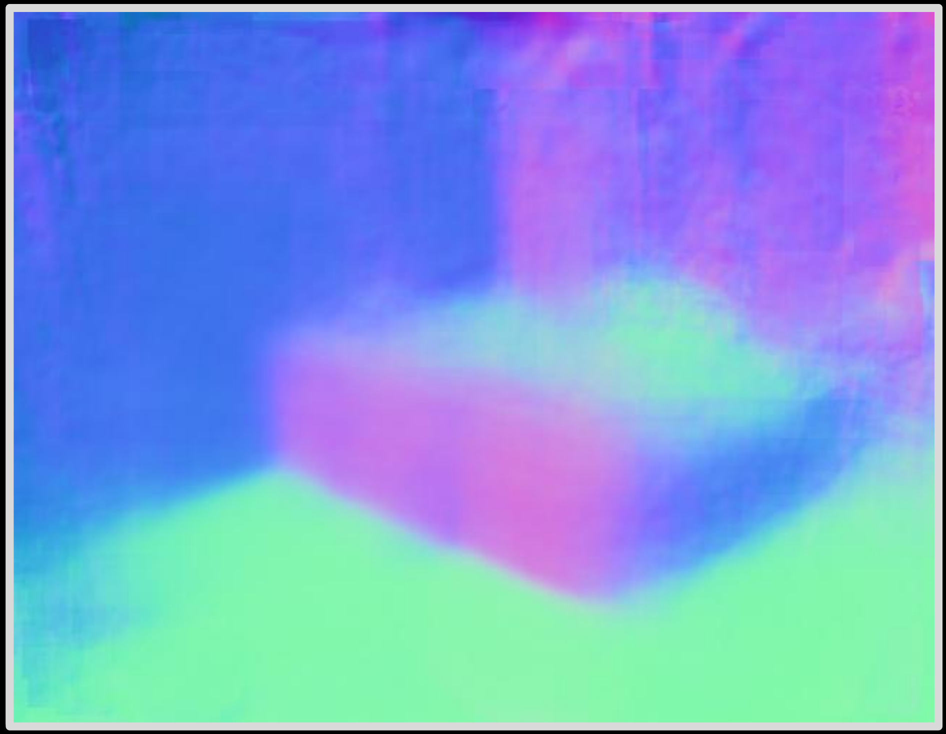
Inference

Sparse Transfer

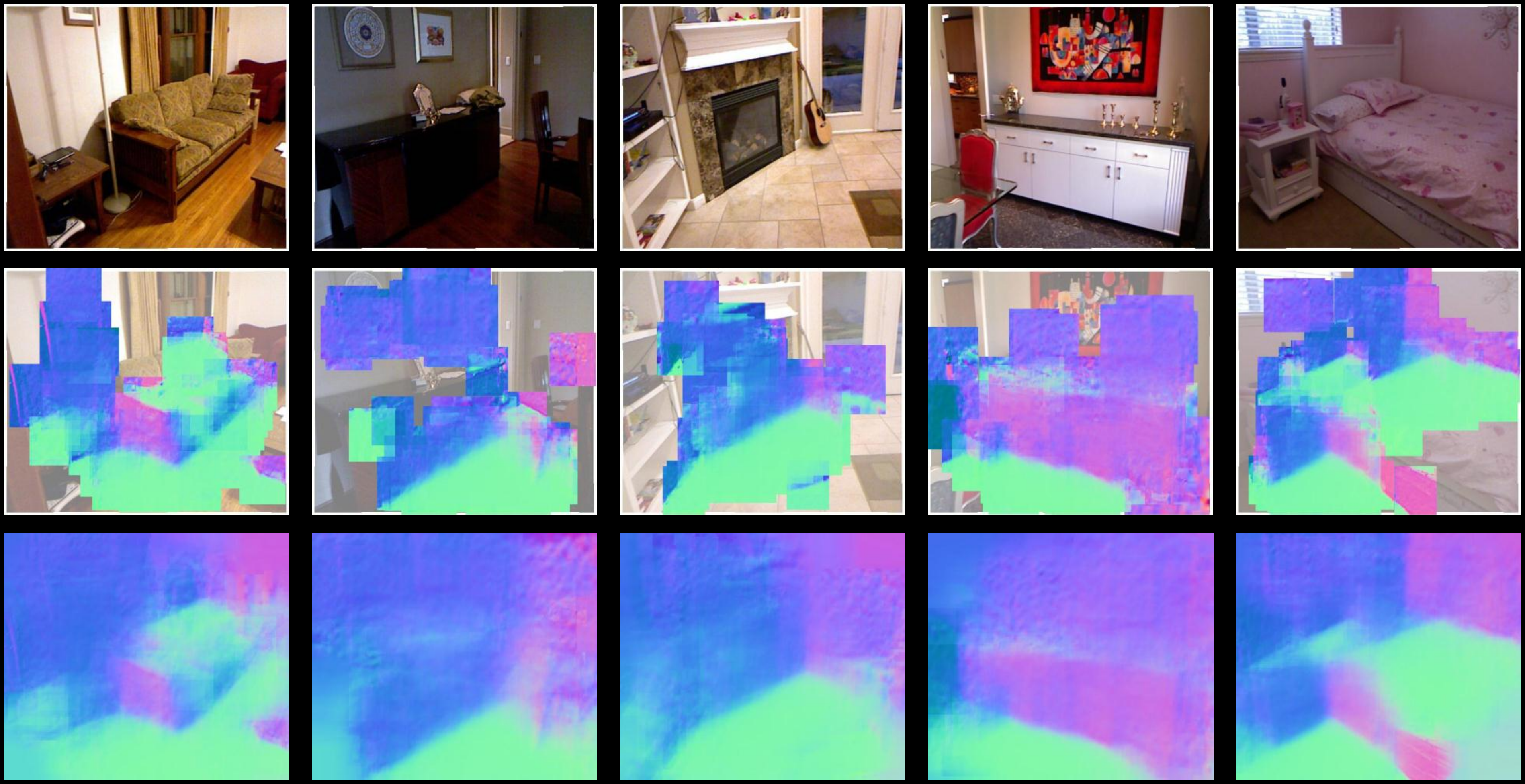


Inference

Dense Transfer



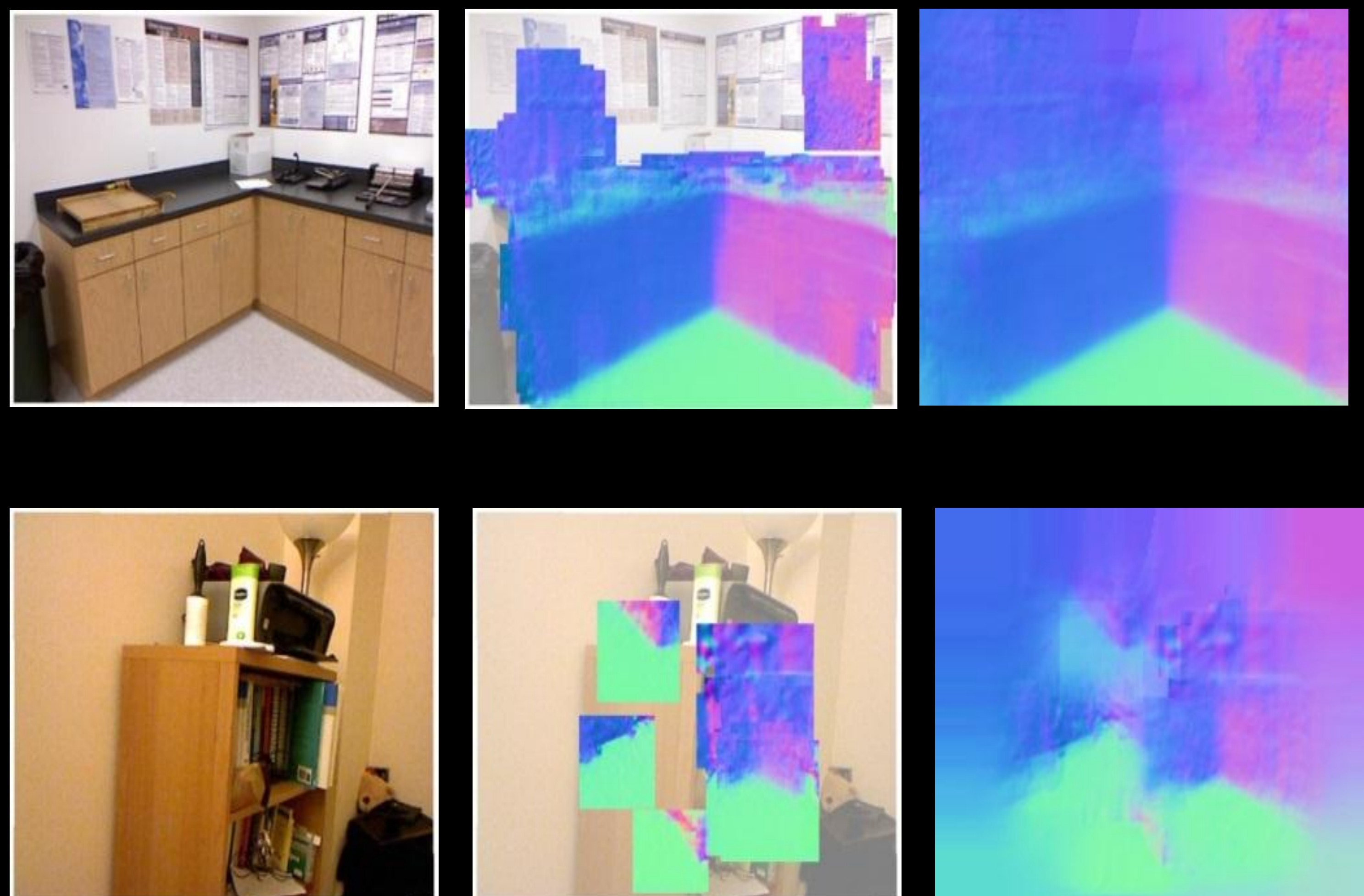
Sample Results – Qualitative



Confidence

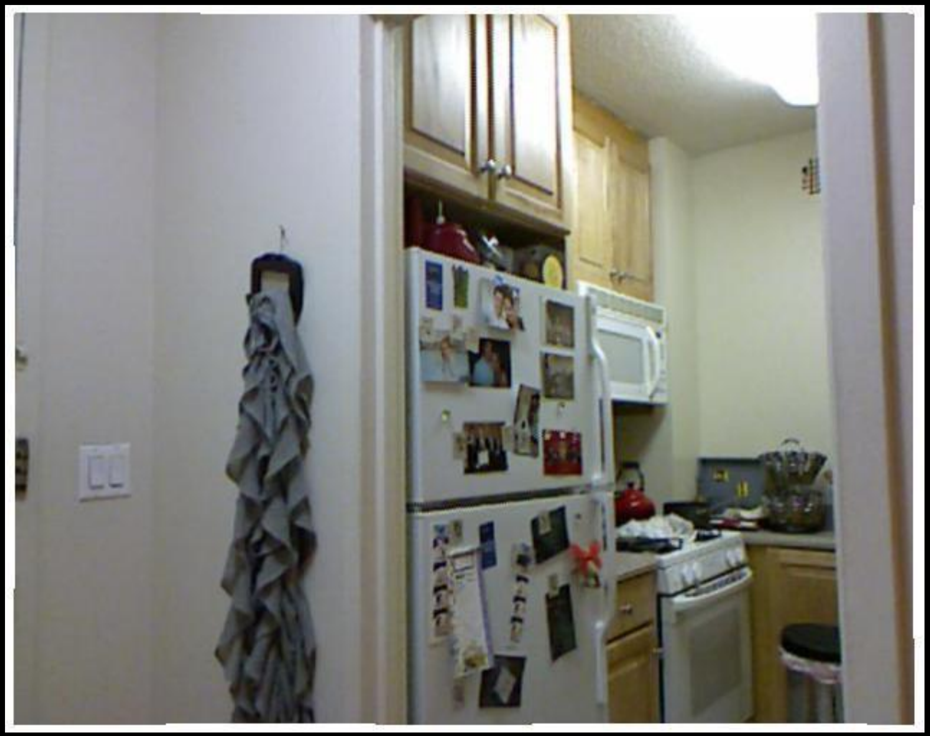
Most Confident Result

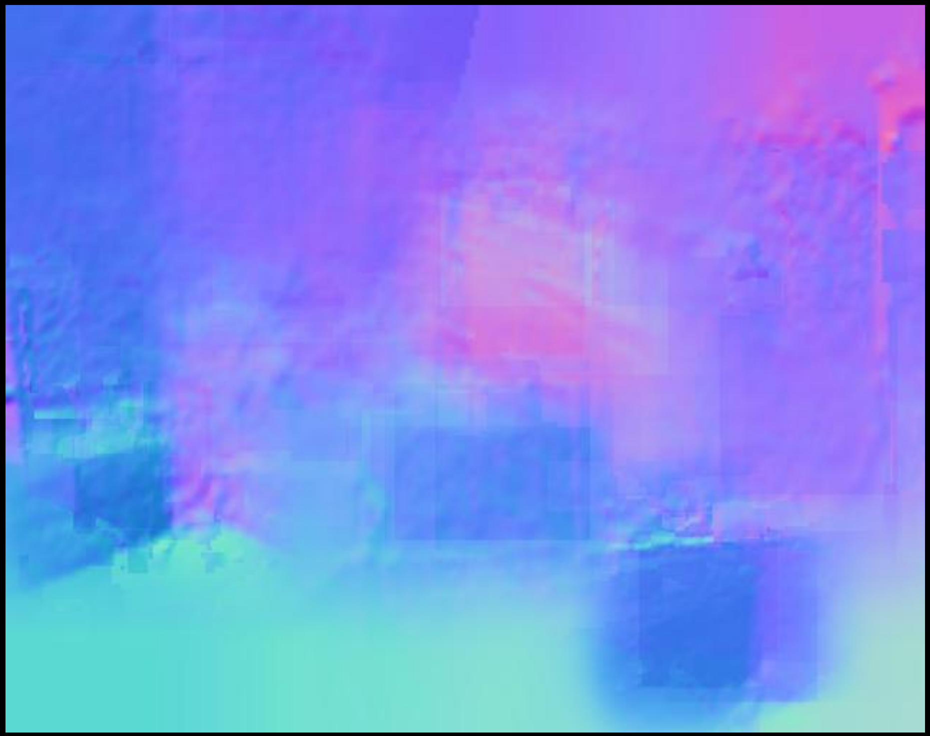
Least Confident Result



rank

Failures

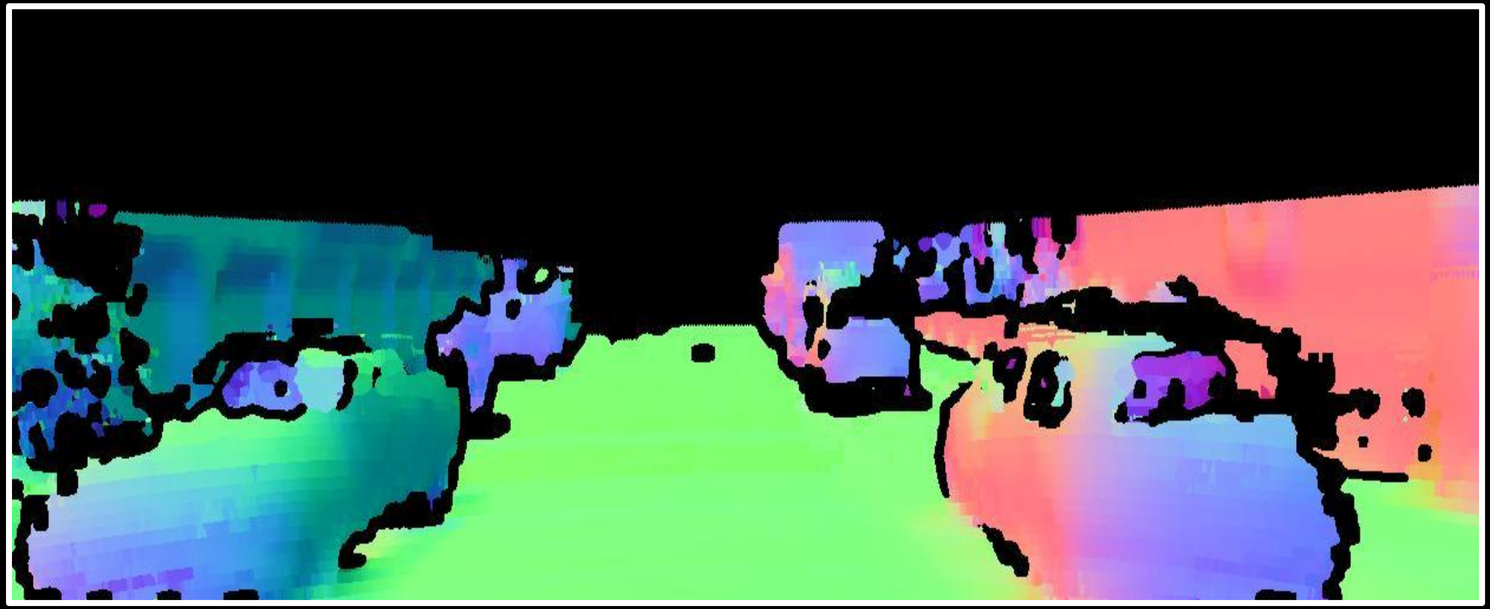




	Summary Stats (^{0}) (Lower Better)			% Good Pixels (Higher Better)		
	Mean	Median	RMSE	11.25^0	22.5^0	30^0
3D Primitives	**33.0**	**28.3**	**40.0**	**18.8**	**40.7**	**52.4**
Singh et al.	35.0	32.4	40.6	11.2	32.1	45.8
Karsch et al.	40.8	37.8	46.9	7.9	25.8	38.2
Hoiem et al.	41.2	34.8	49.3	9.0	31.7	43.9
Saxena et al.	47.1	42.3	56.3	11.2	28.0	37.4
RF + Dense SIFT	36.0	33.4	41.7	11.4	31.1	44.2

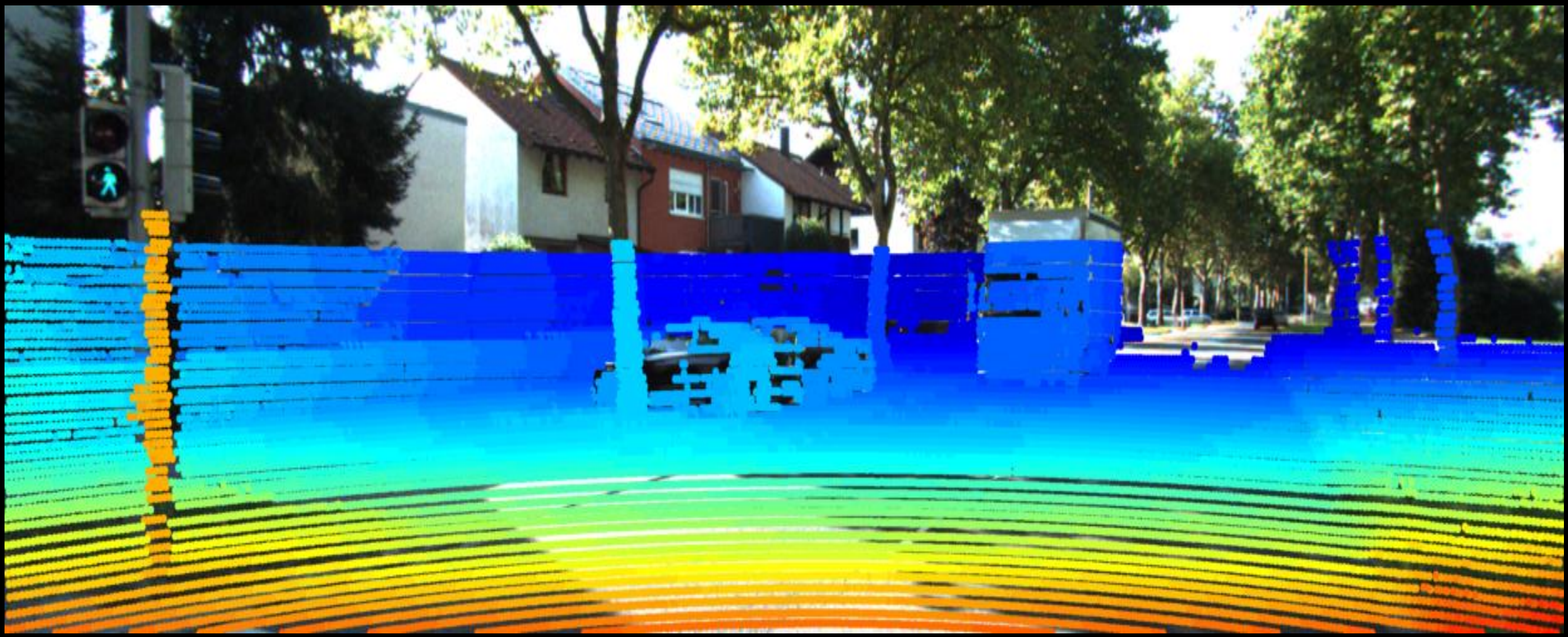
RMSE

More general environments?

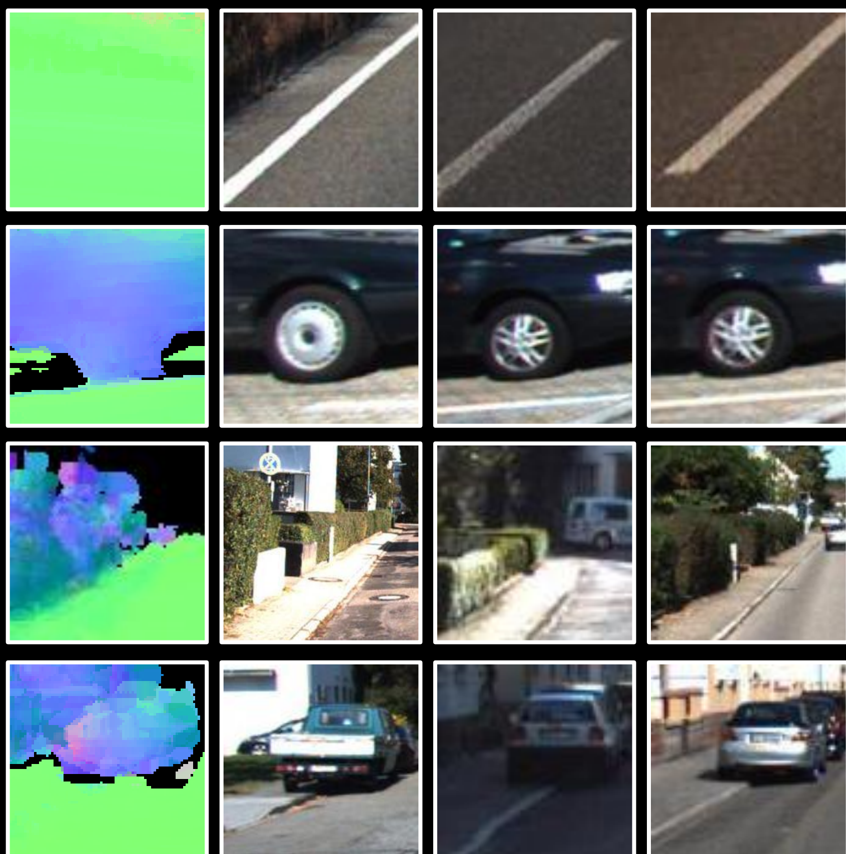


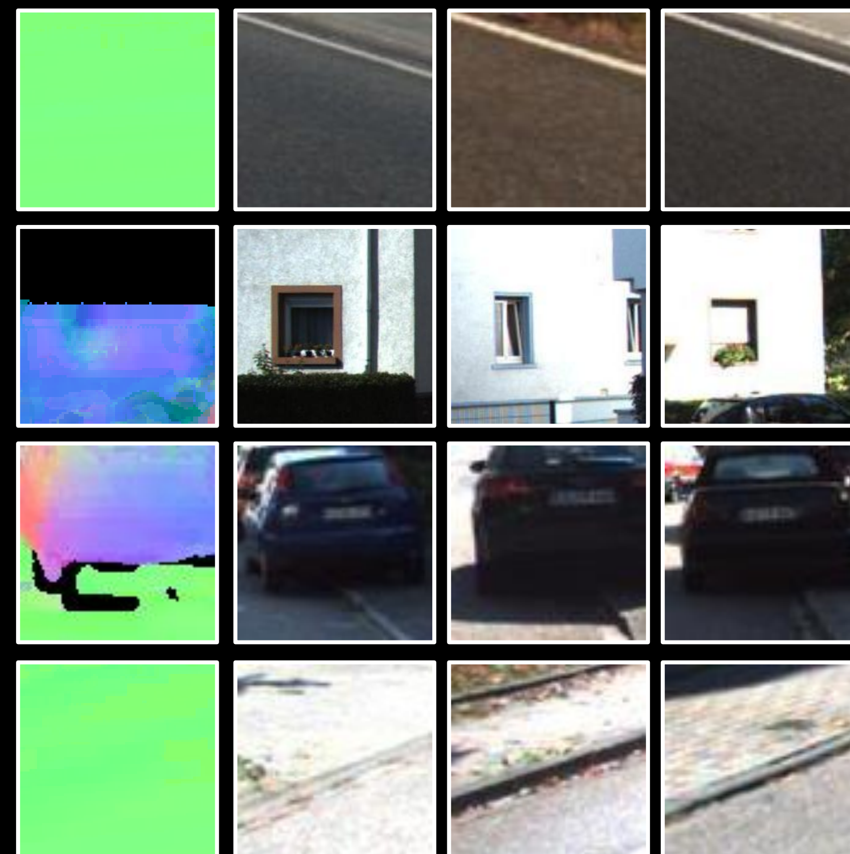
KITTI Dataset: Geiger, Lenz, Urtasun, ‘12

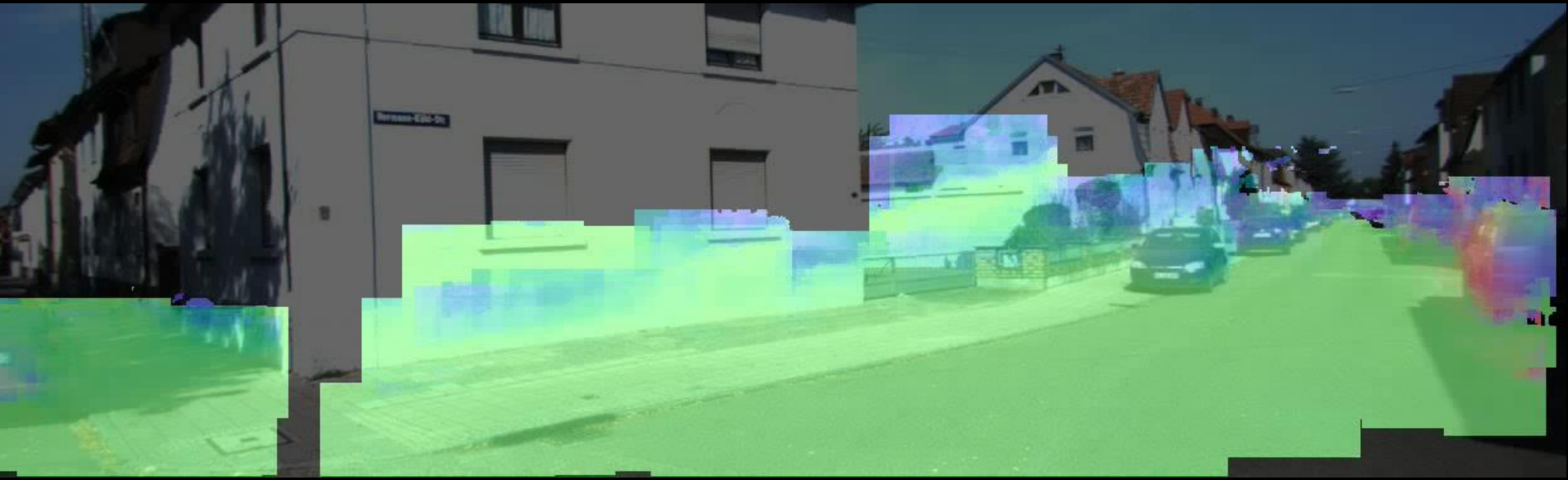
- Large regions without surface interpretation
- Fewer linear/planar structures to anchor
- Irregular distribution of 3D training data




Discovered Primitives (Examples)

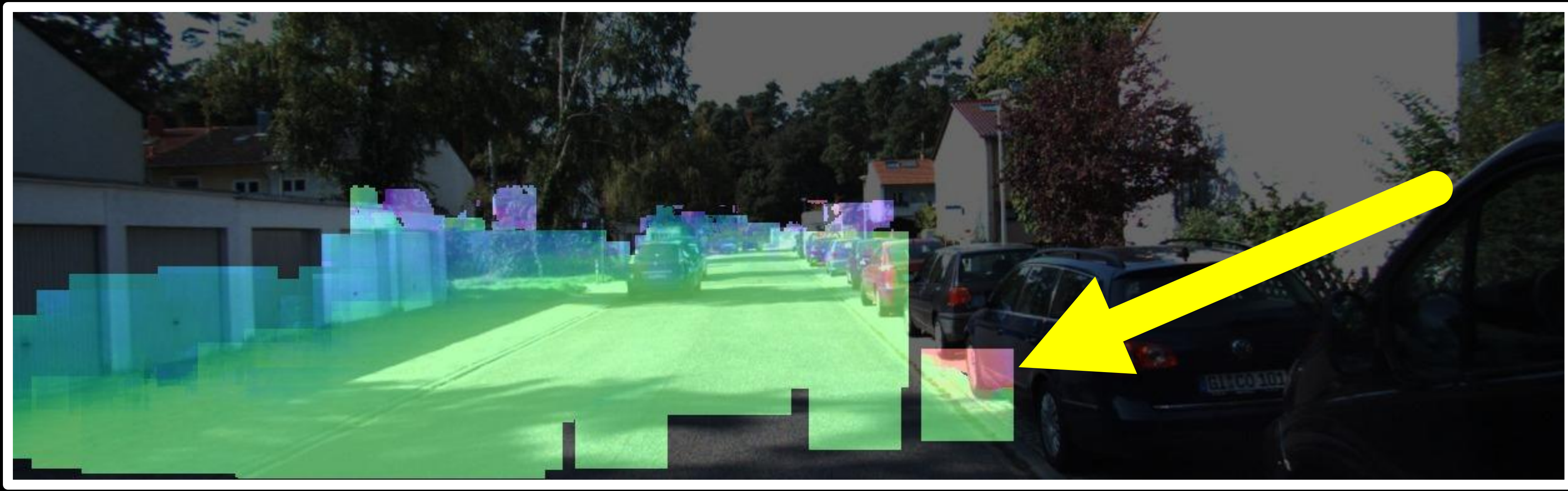




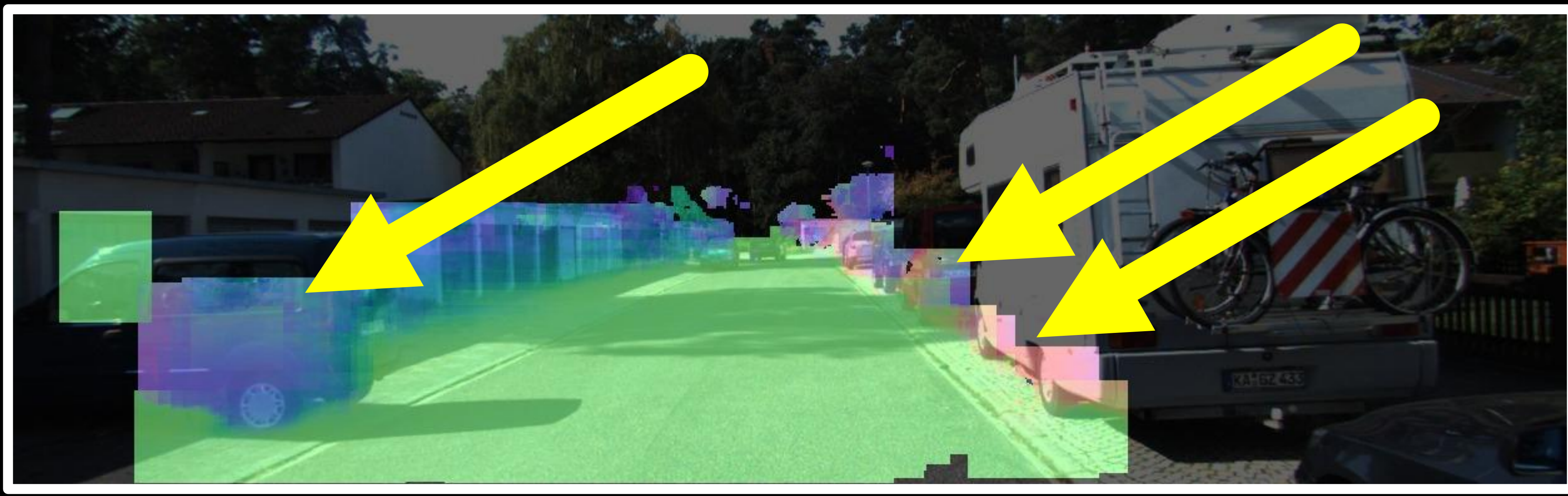


	Summary Stats (^{0}) (Lower Better)			% Good Pixels (Higher Better)		
	Mean	Median	RMSE	11.25^0	22.5^0	30^0
3D Primitives	**23.4**	**9.9**	35.4	**52.6**	**64.4**	**69.4**
RF + Dense SIFT	24.2	16.3	**32.8**	39.2	59.3	68.3

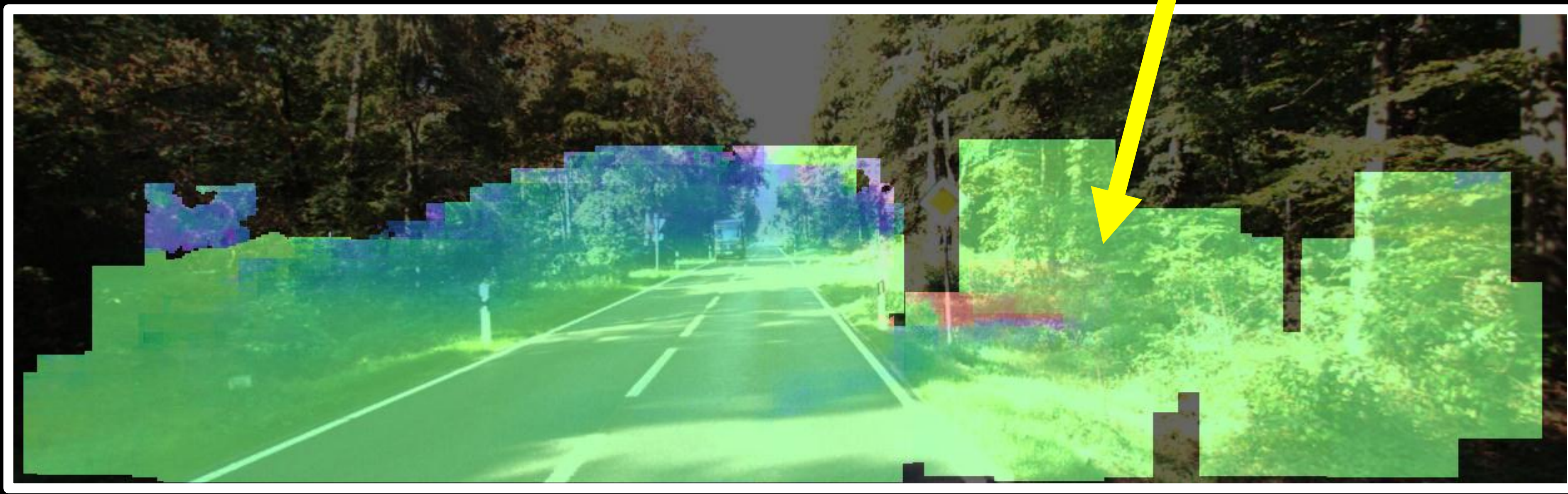
Contact points



Object surfaces + Contact points

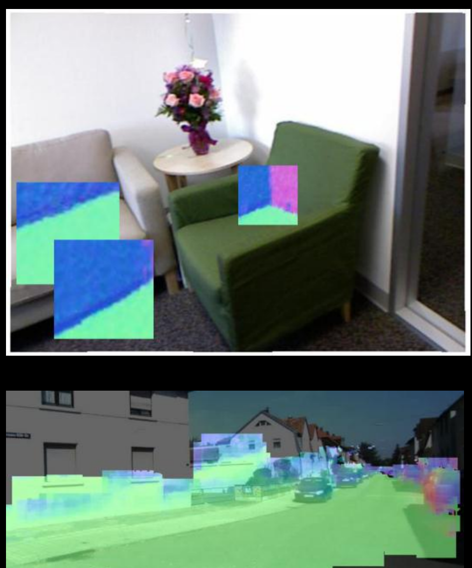


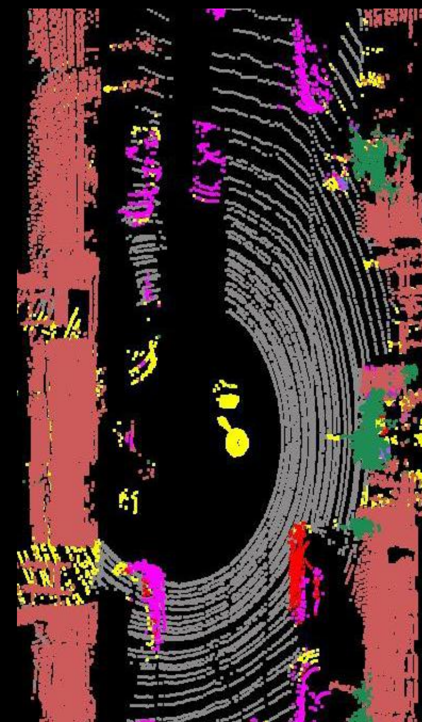
Failures



Failures

Digression

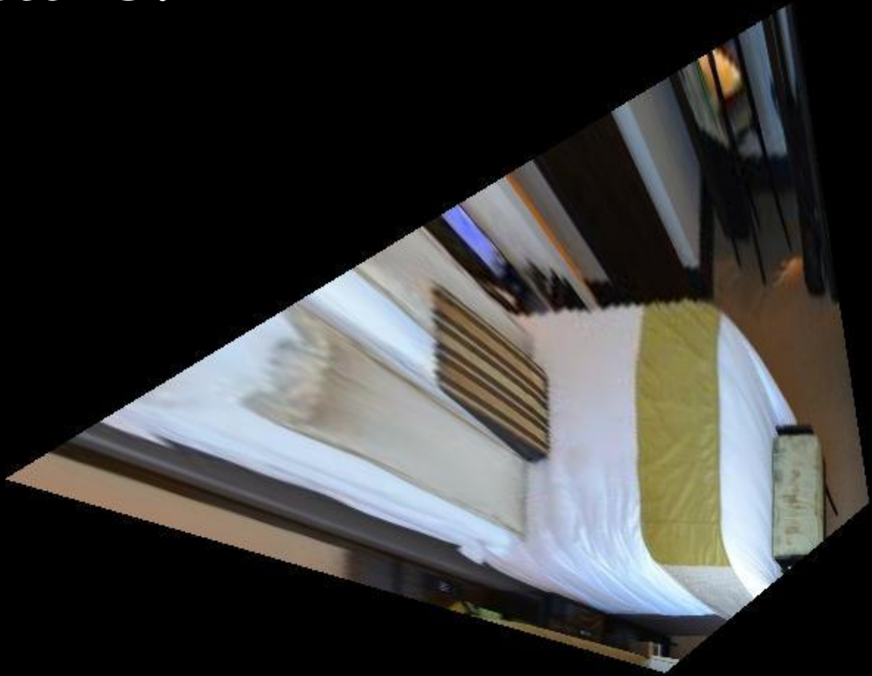




Style and structure

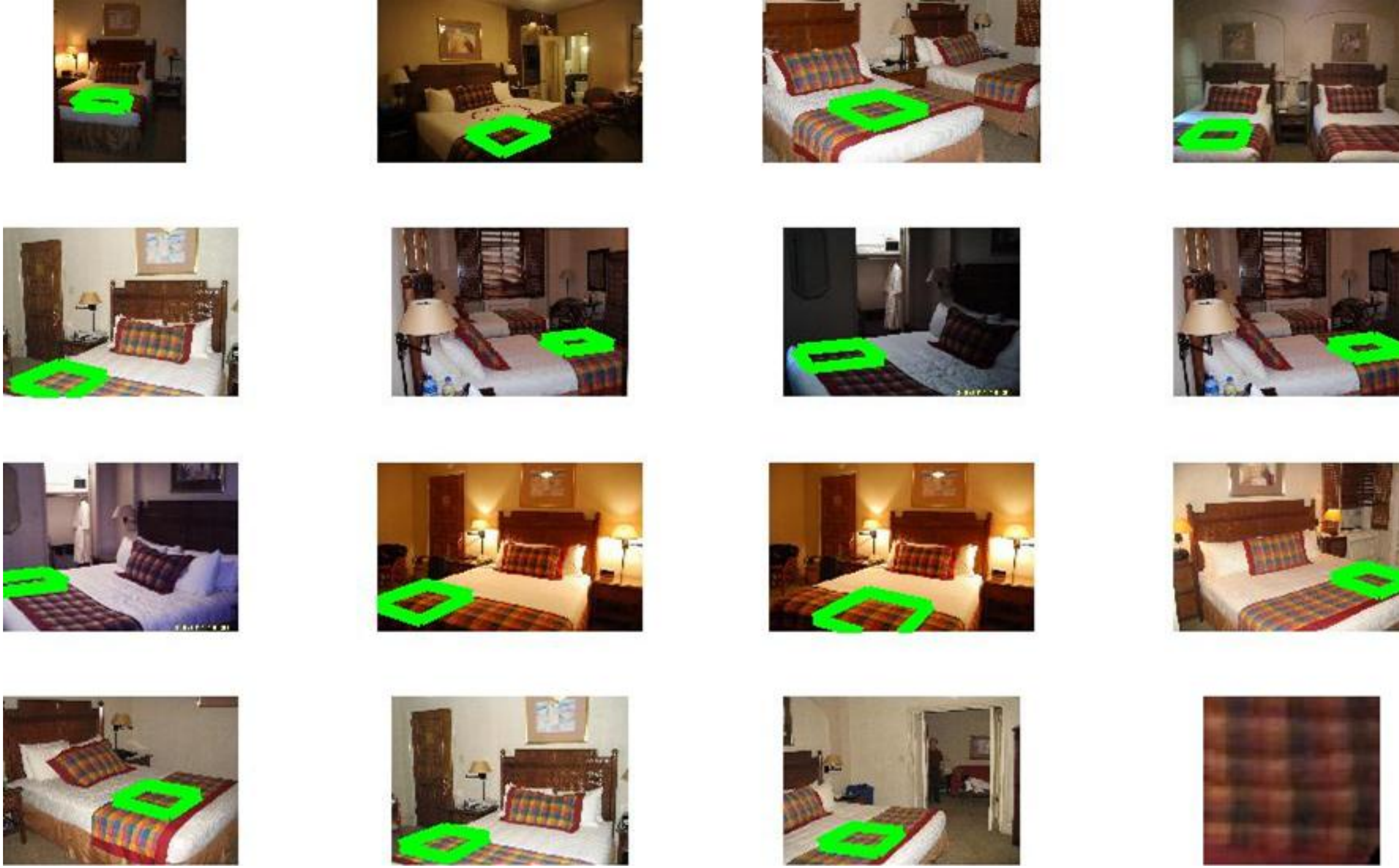


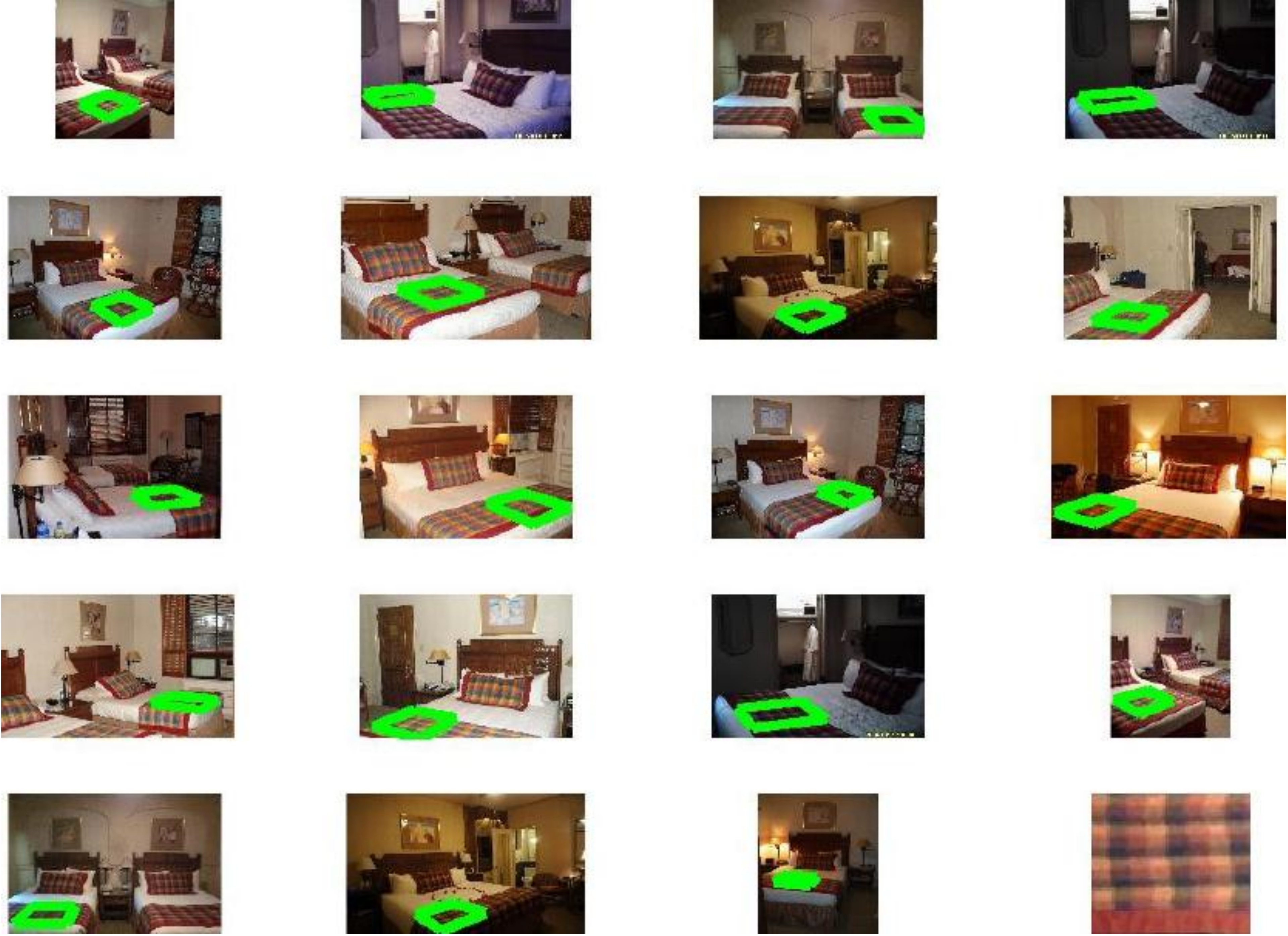

Style vs. structure?

Tenenbaum & Freeman. Separating Style and Content with Bilinear Models. Neural Computation. 2000.
Lee, Efros, Hebert. Style-aware Mid-level Representation for Discovering Visual Connections in Space and Time. 2013.

Casablanca Hotel, New York



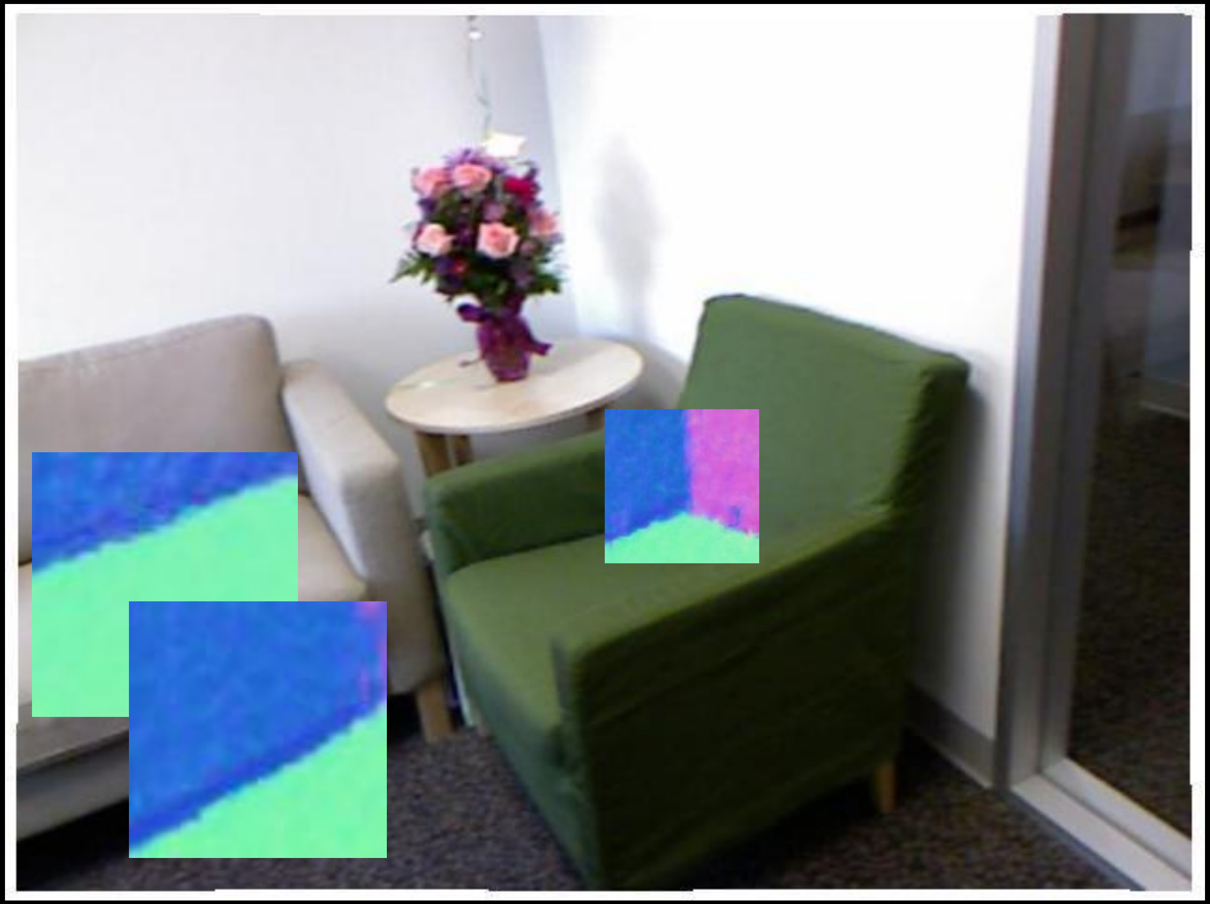


Meritan Apartments Sydney

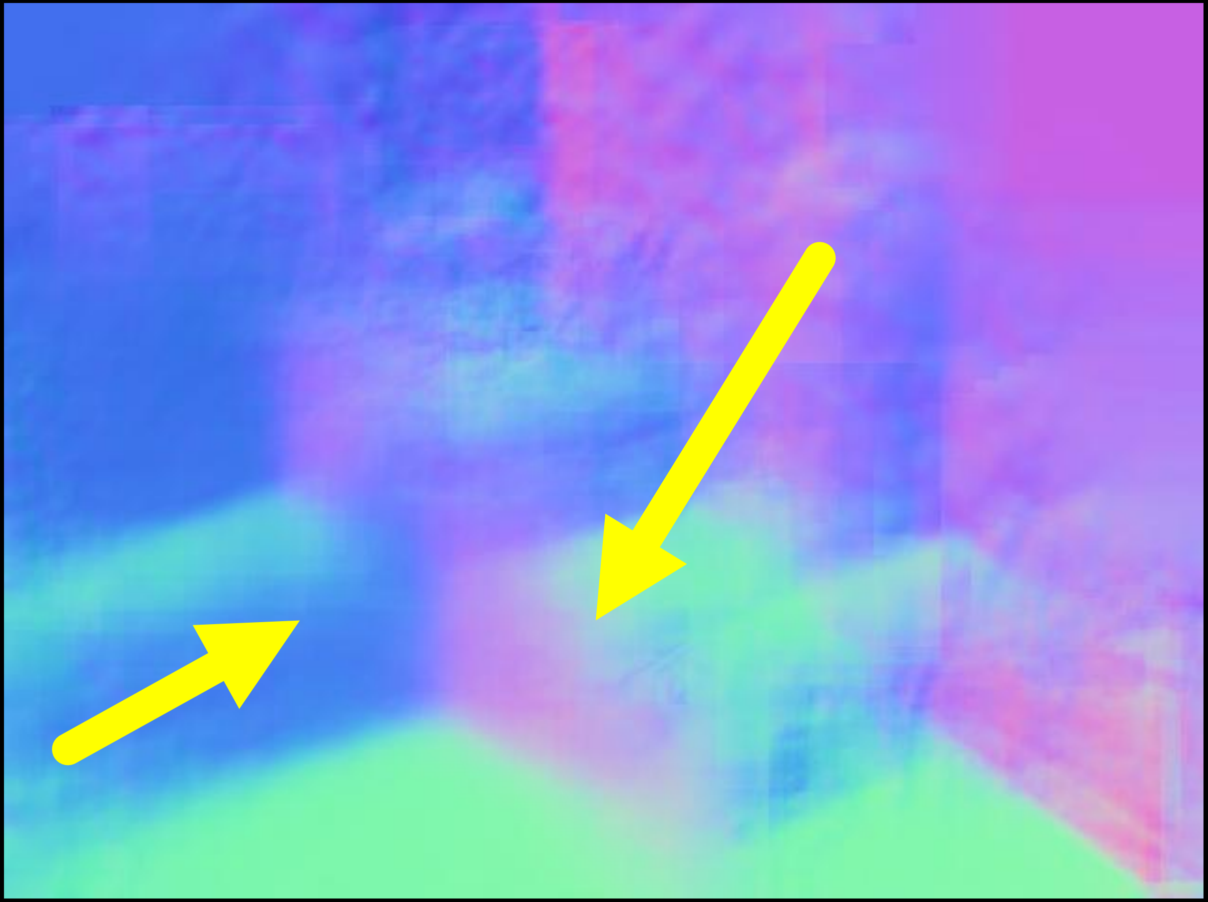
Sheraton Hotels (North America)

Using geometric and physical constraints

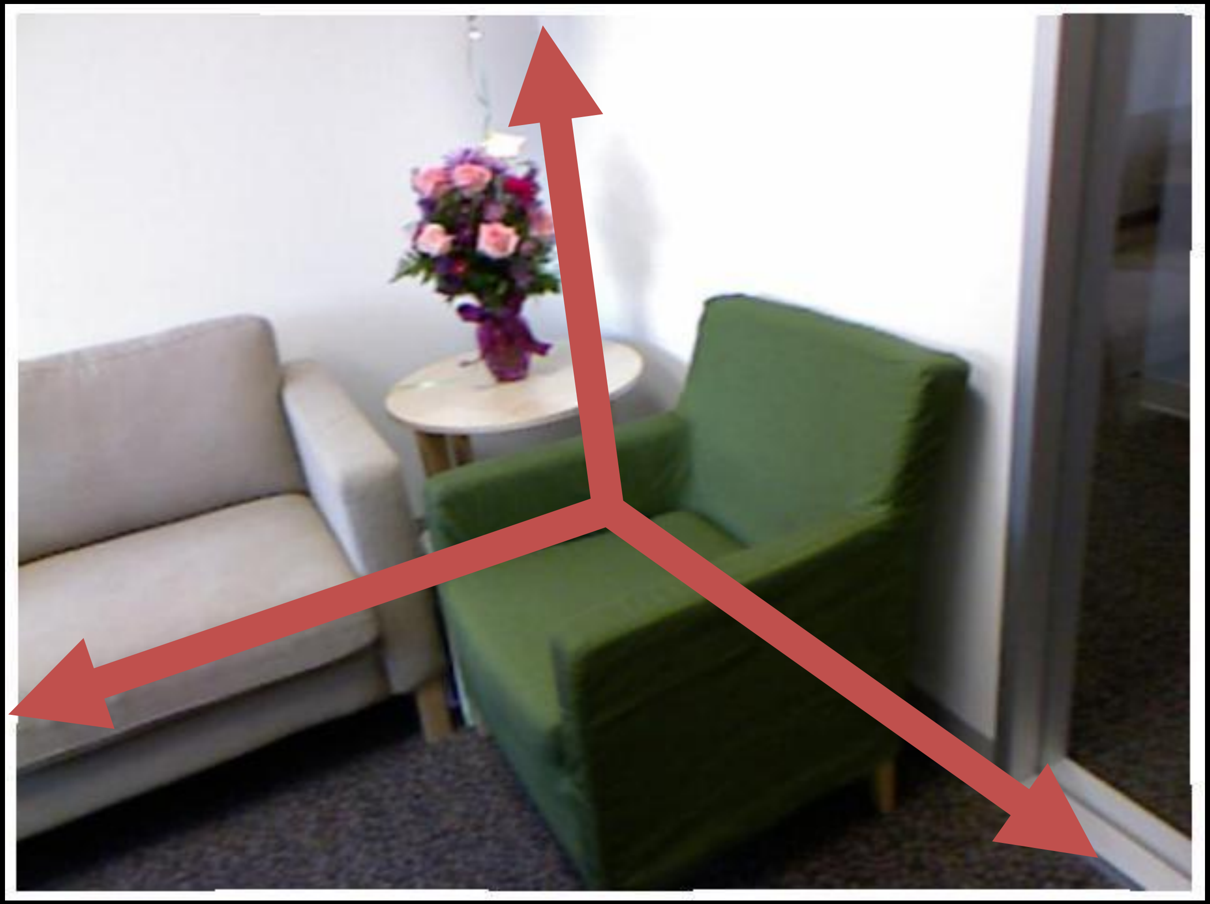
The Story So Far



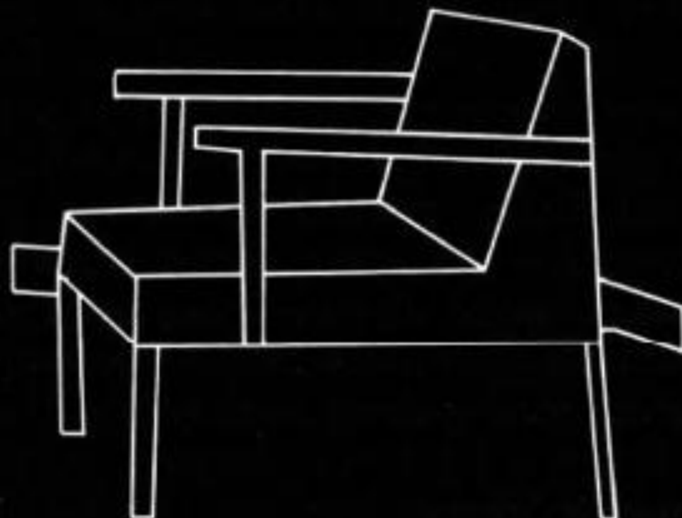
The Story So Far

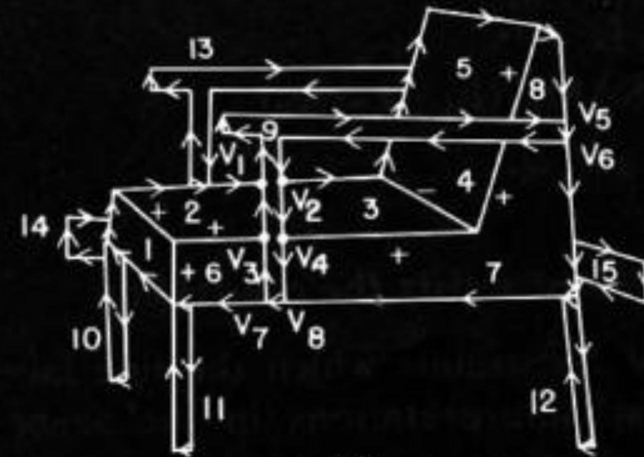


Adding Physical/Geometric Constraints

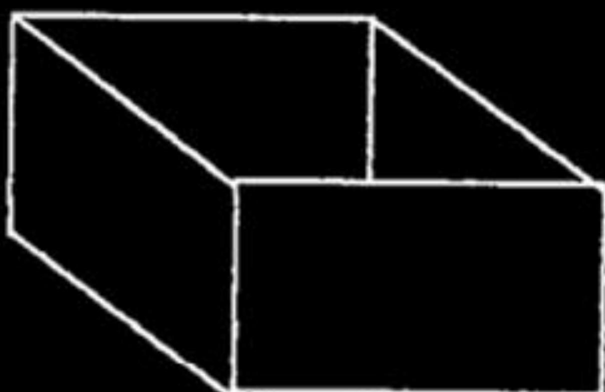


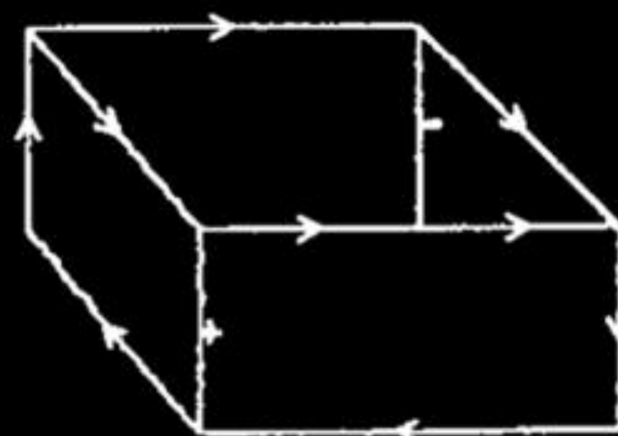
Adding Physical/Geometric Constraints





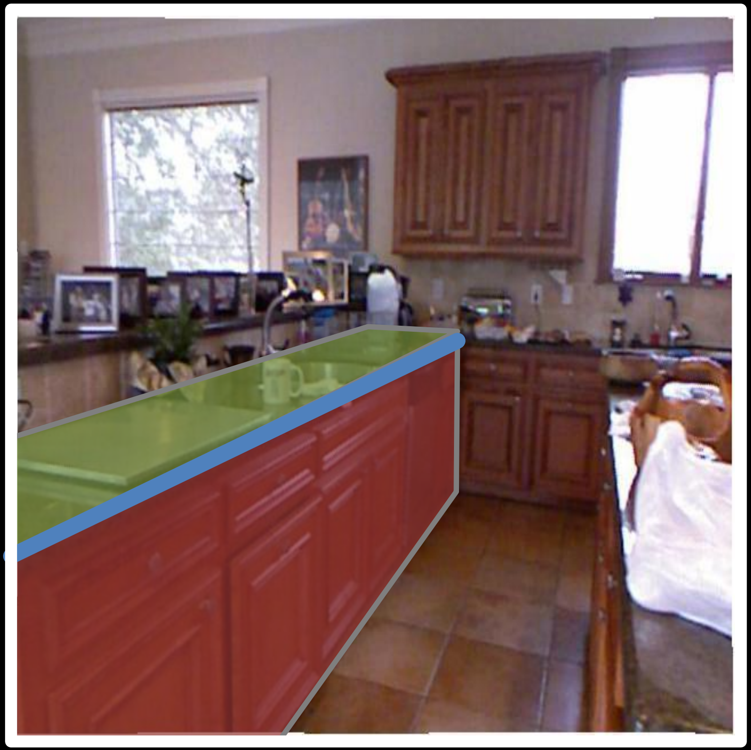
(a)





Huffman 71, Clowes 71, Kanade 80, 81 Sugihara 86, Malik 87, etc.

Edges between surfaces

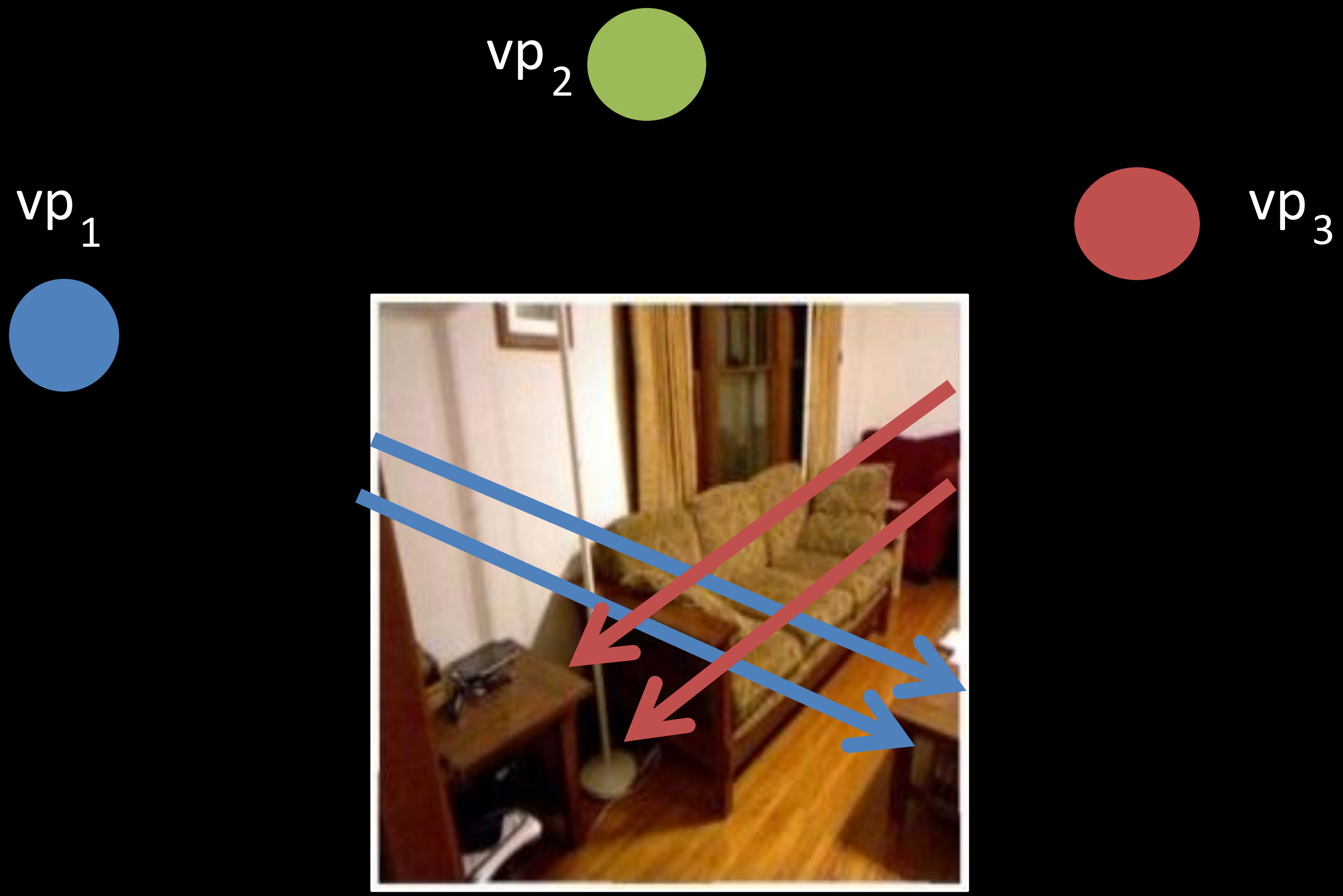


Concave
(-)

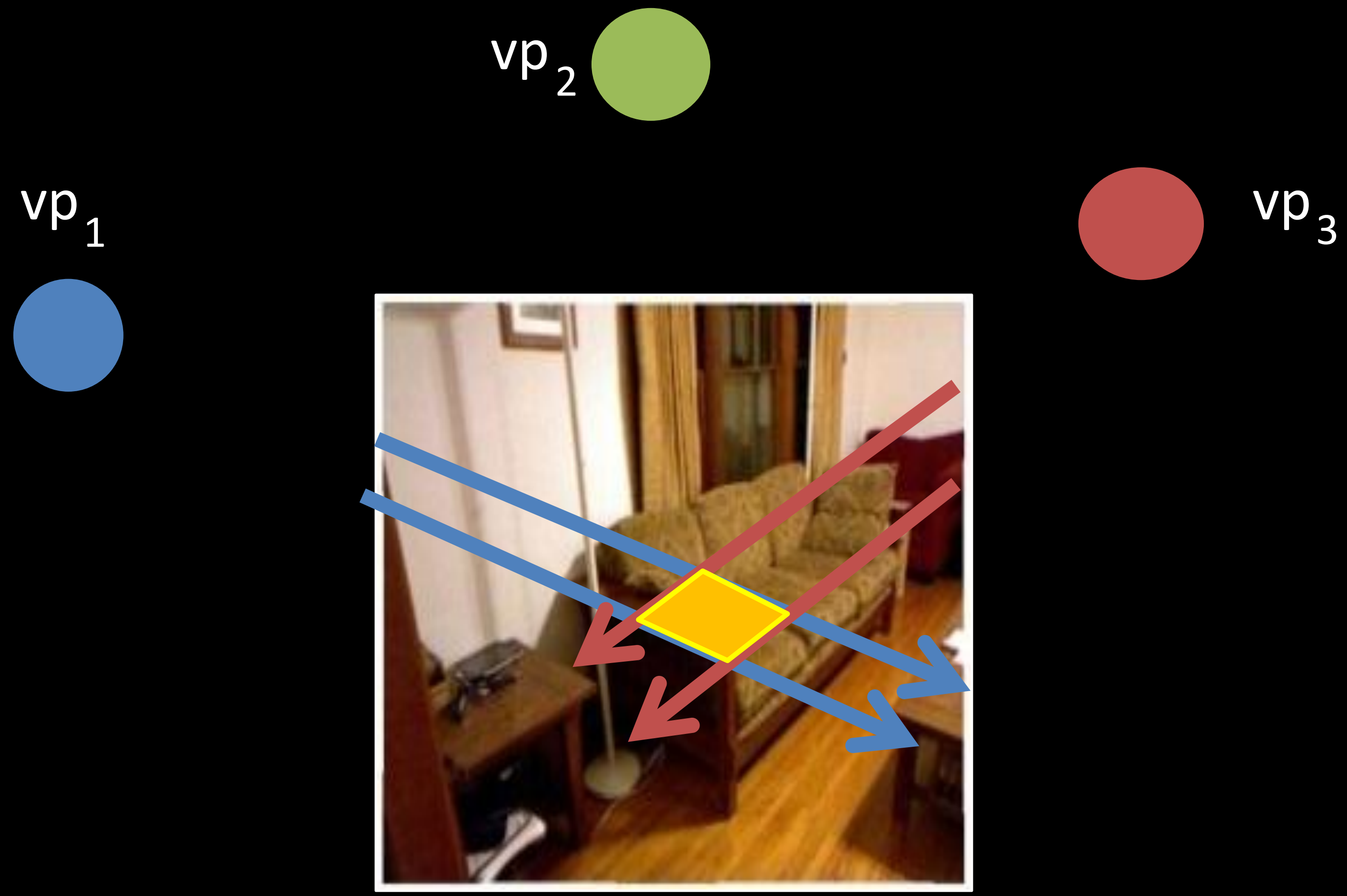
Convex
(+)

Parameterization

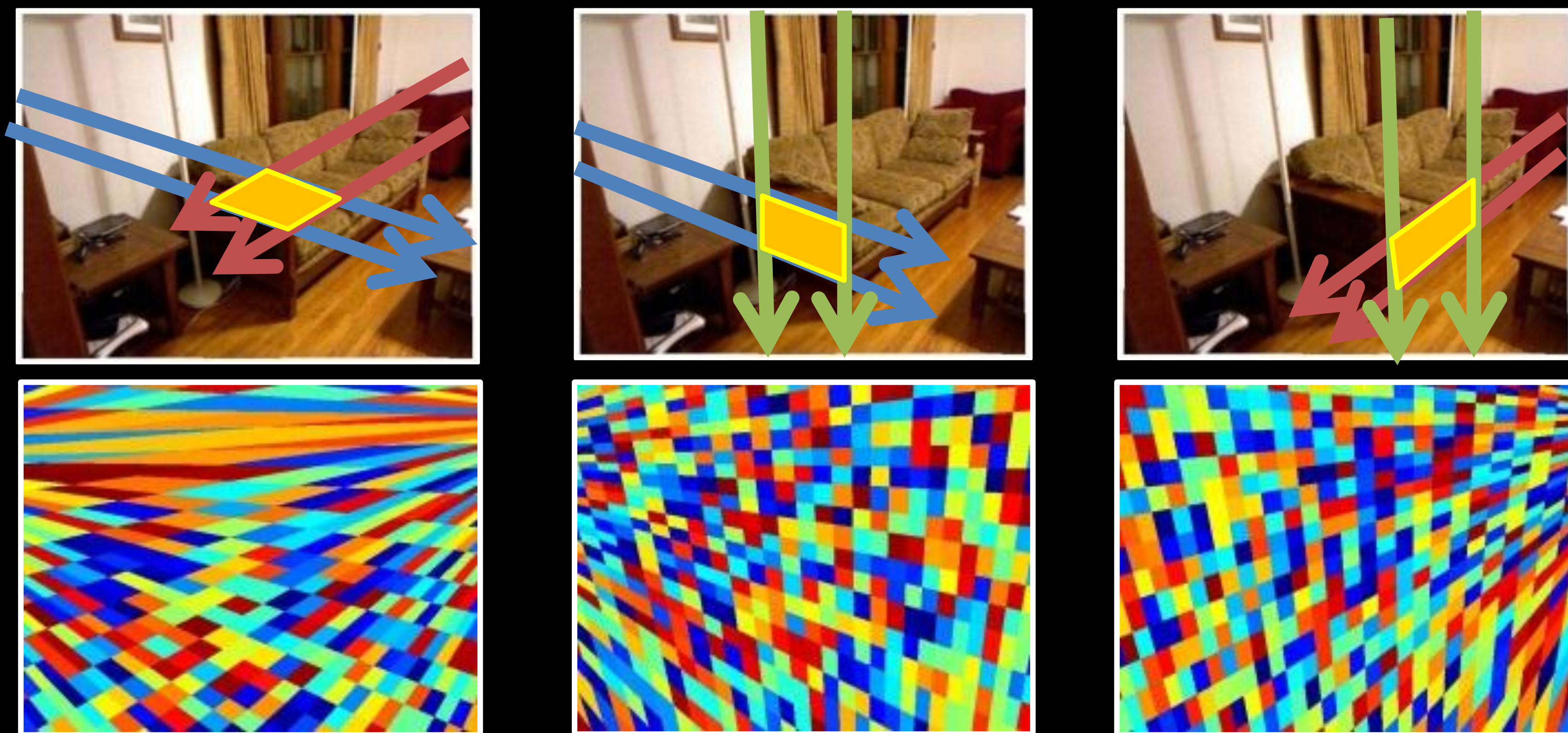
Parameterization



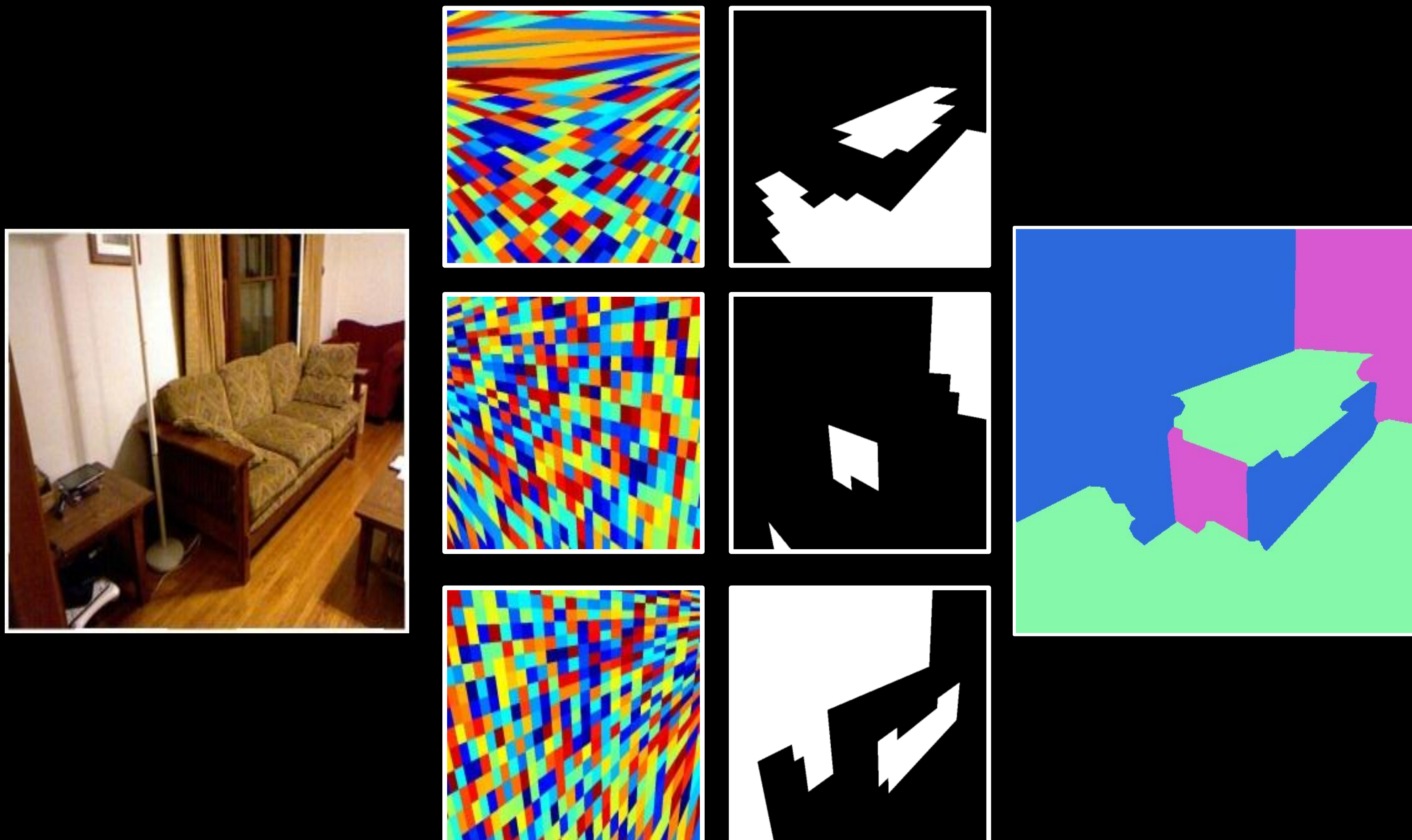
Parameterization



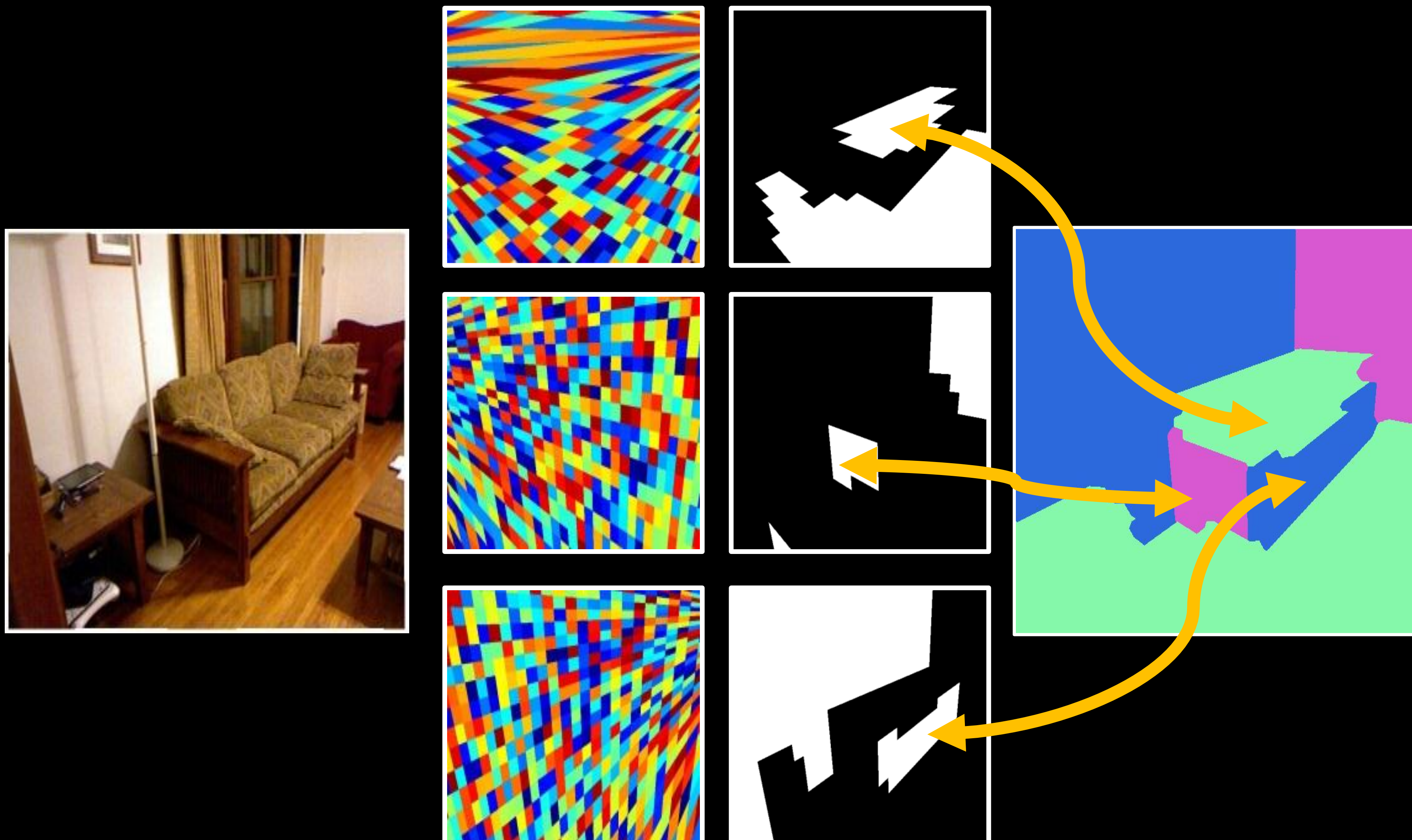
Parameterization



Parameterization

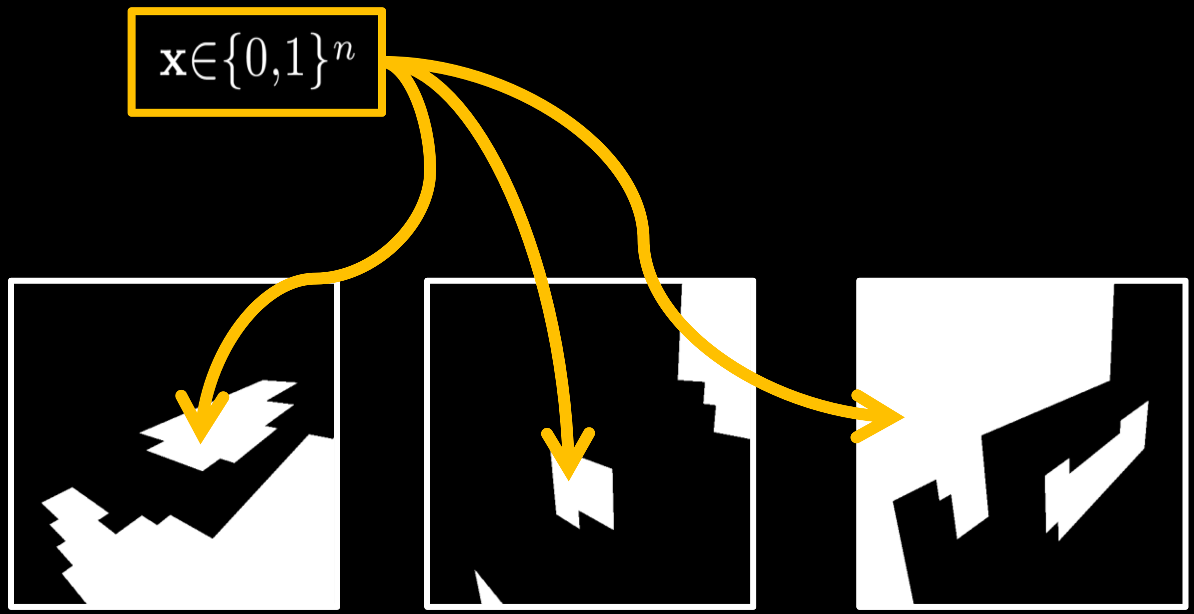


Parameterization



Labeling

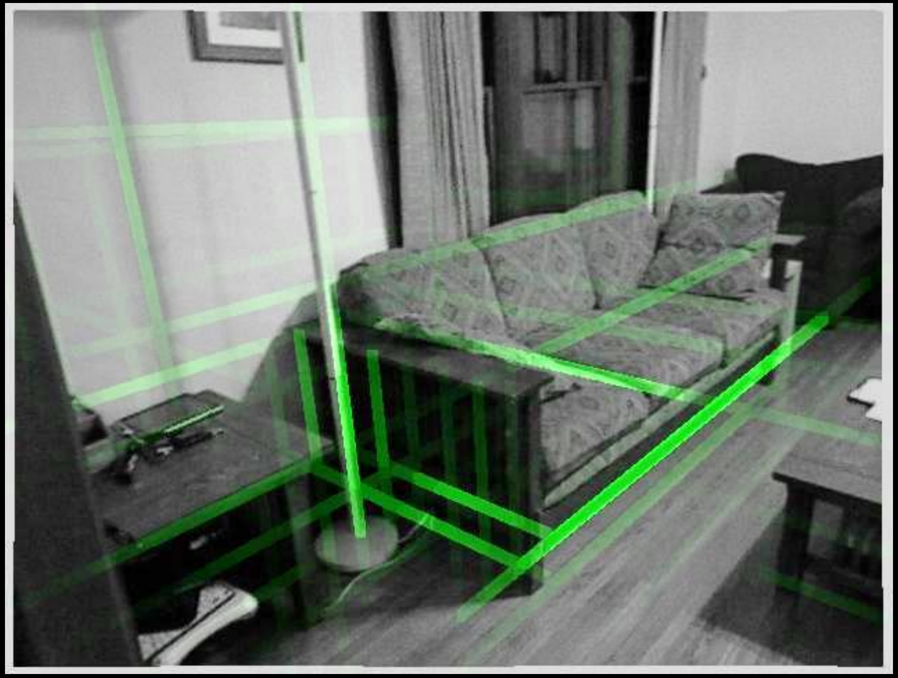
x_i: is cell i on?



Unary terms

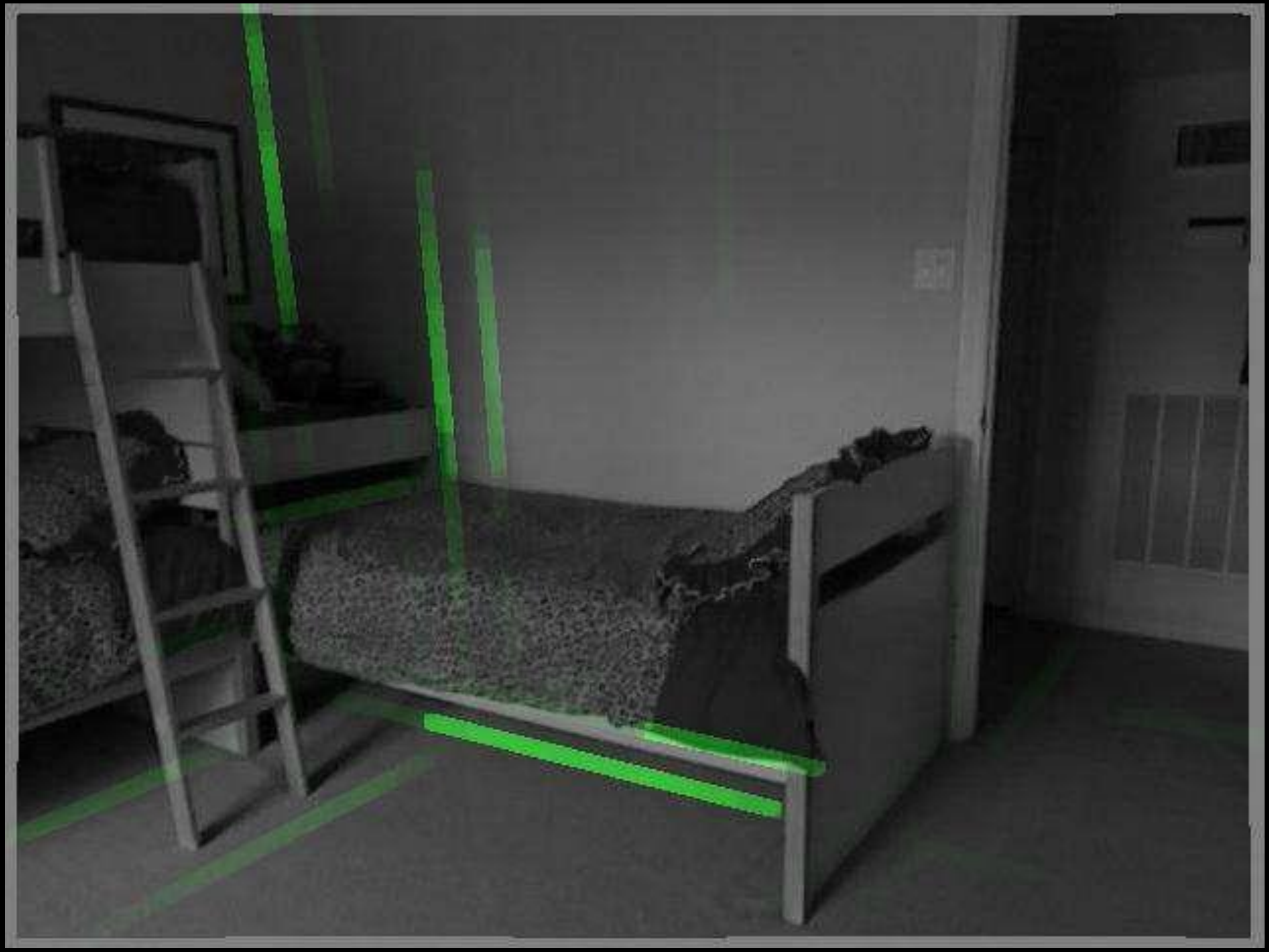
Should cell i be on?

Binary Potentials

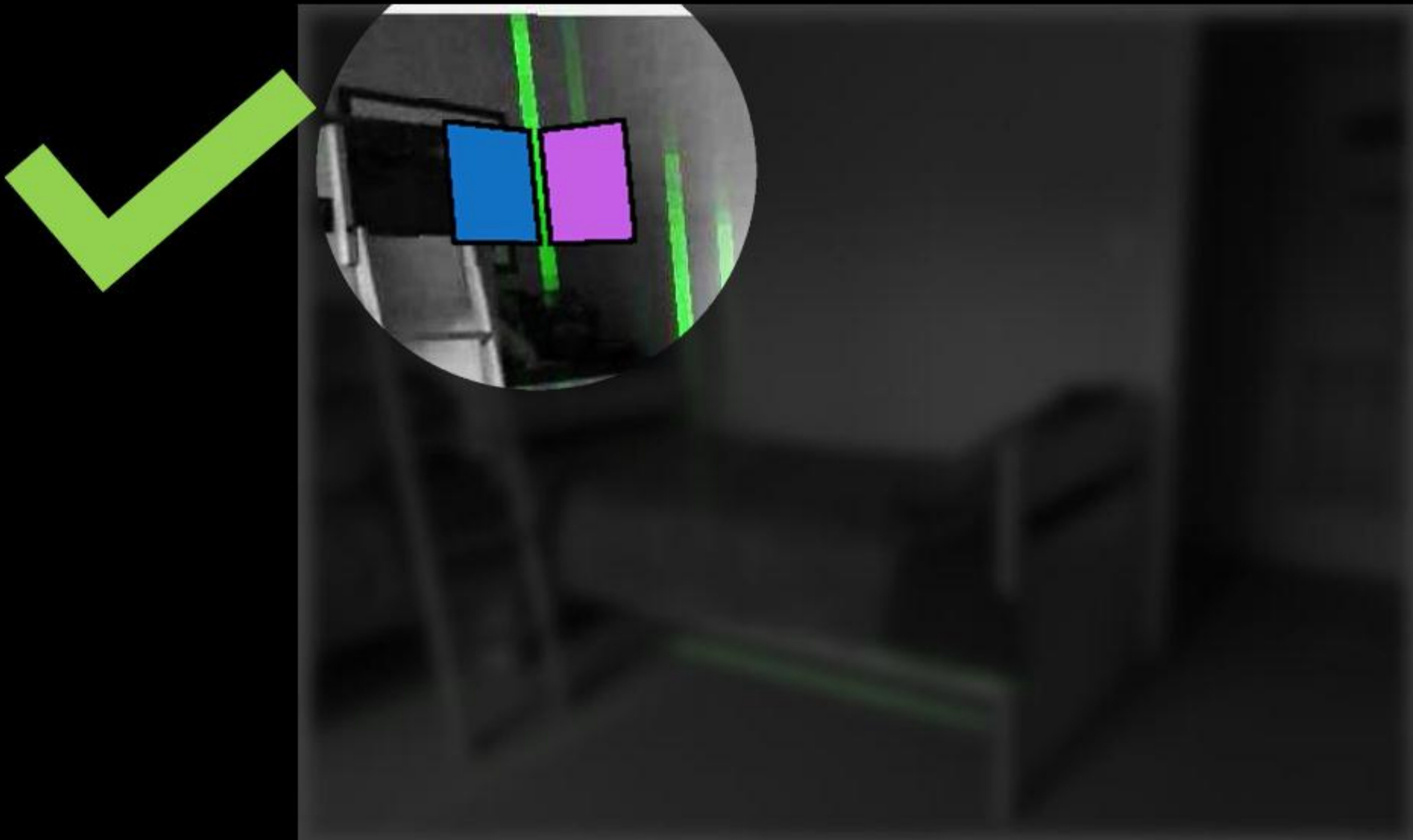


807s

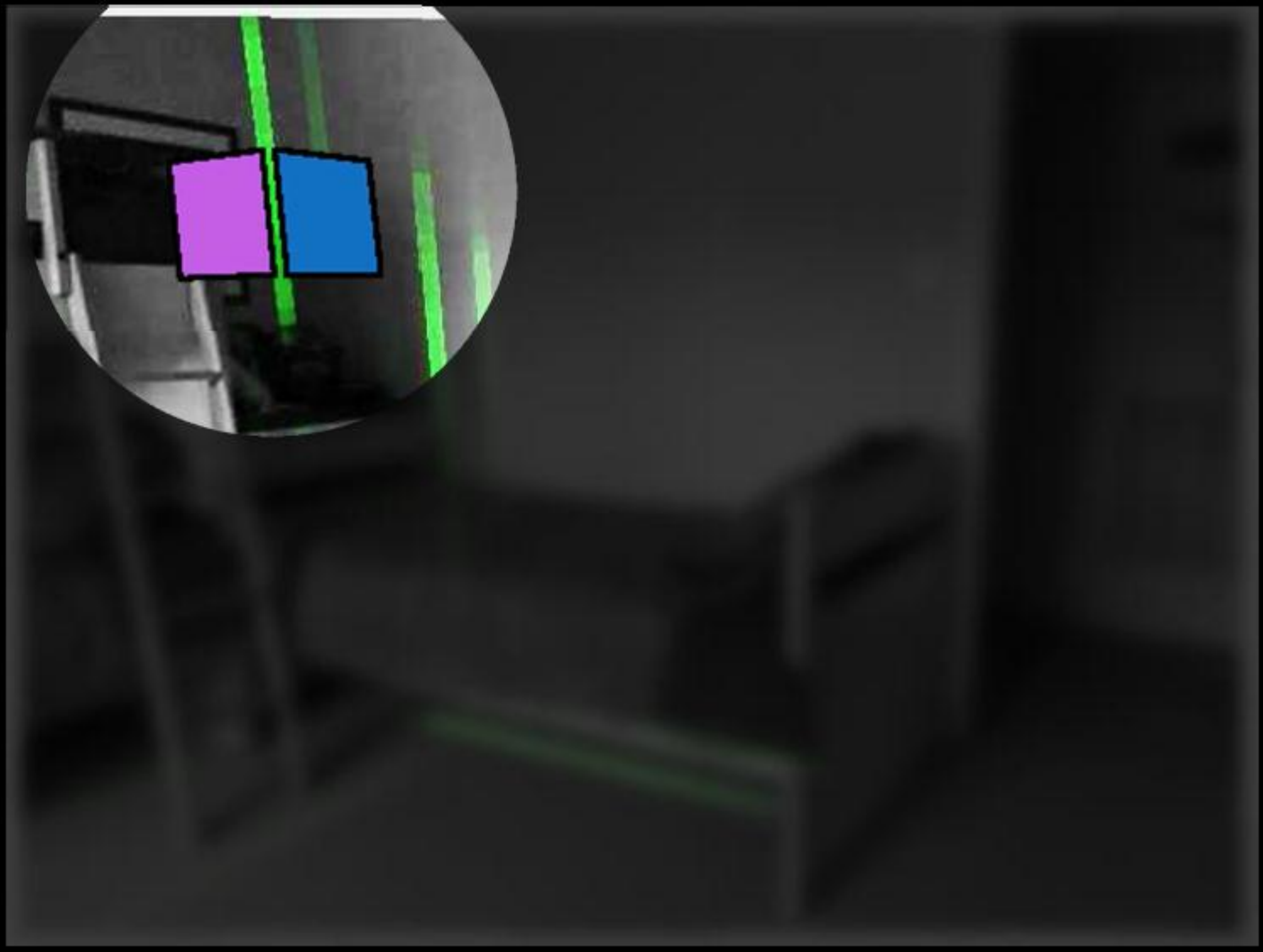
Binary terms



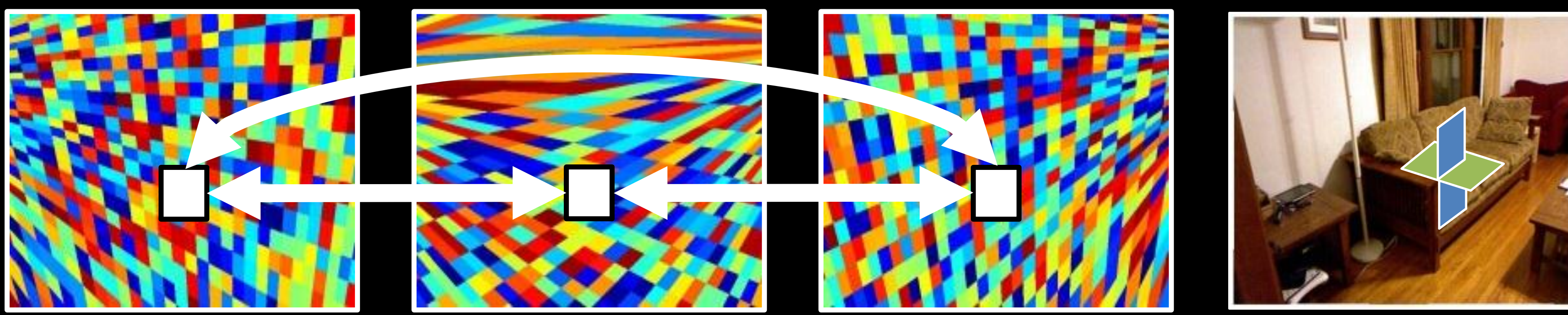
Binary terms



Binary terms



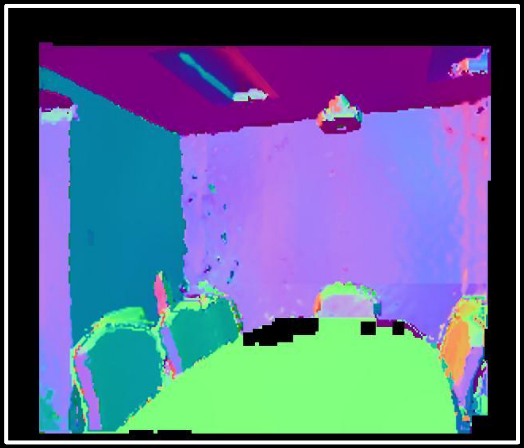
Constraints



Gurobi BB

Qualitative Results

Input

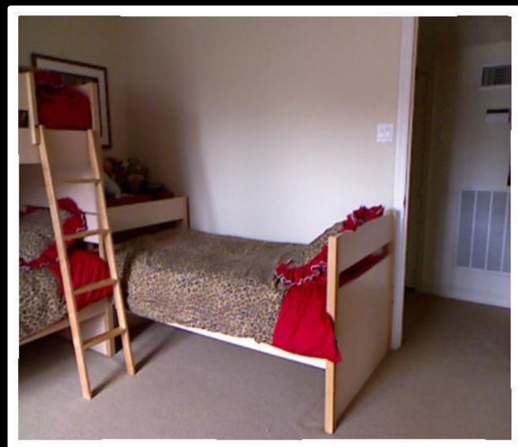


Ground Truth

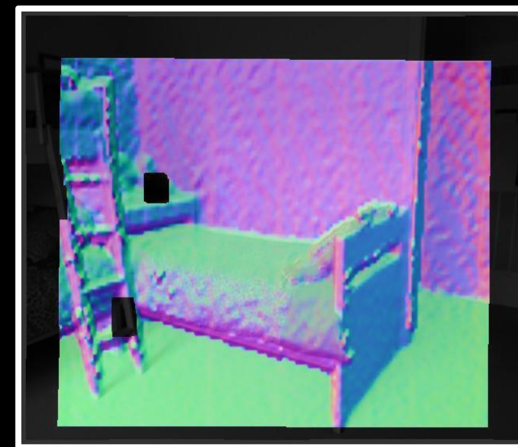
3D Primitives

Projected 3D Primitives

Proposed

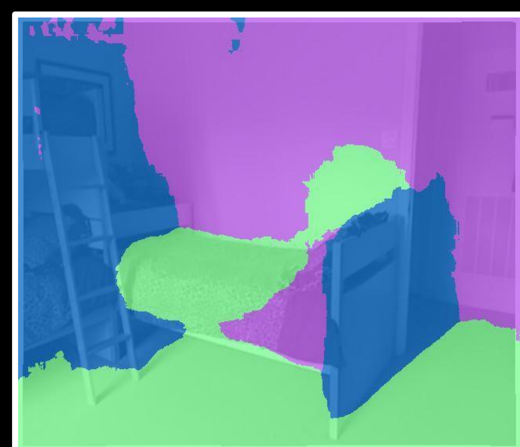


Input

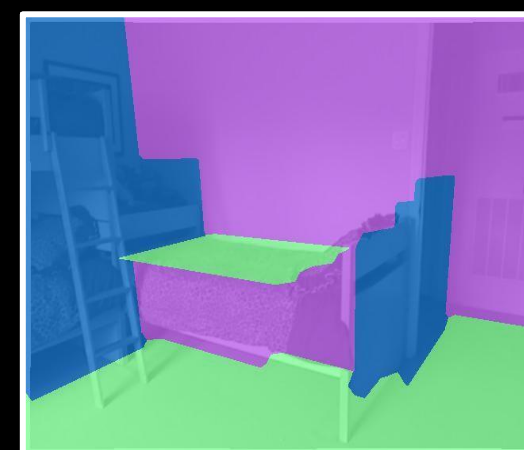


Ground Truth

3D Primitives



Projected 3D Primitives



Proposed

Random Qualitative Results

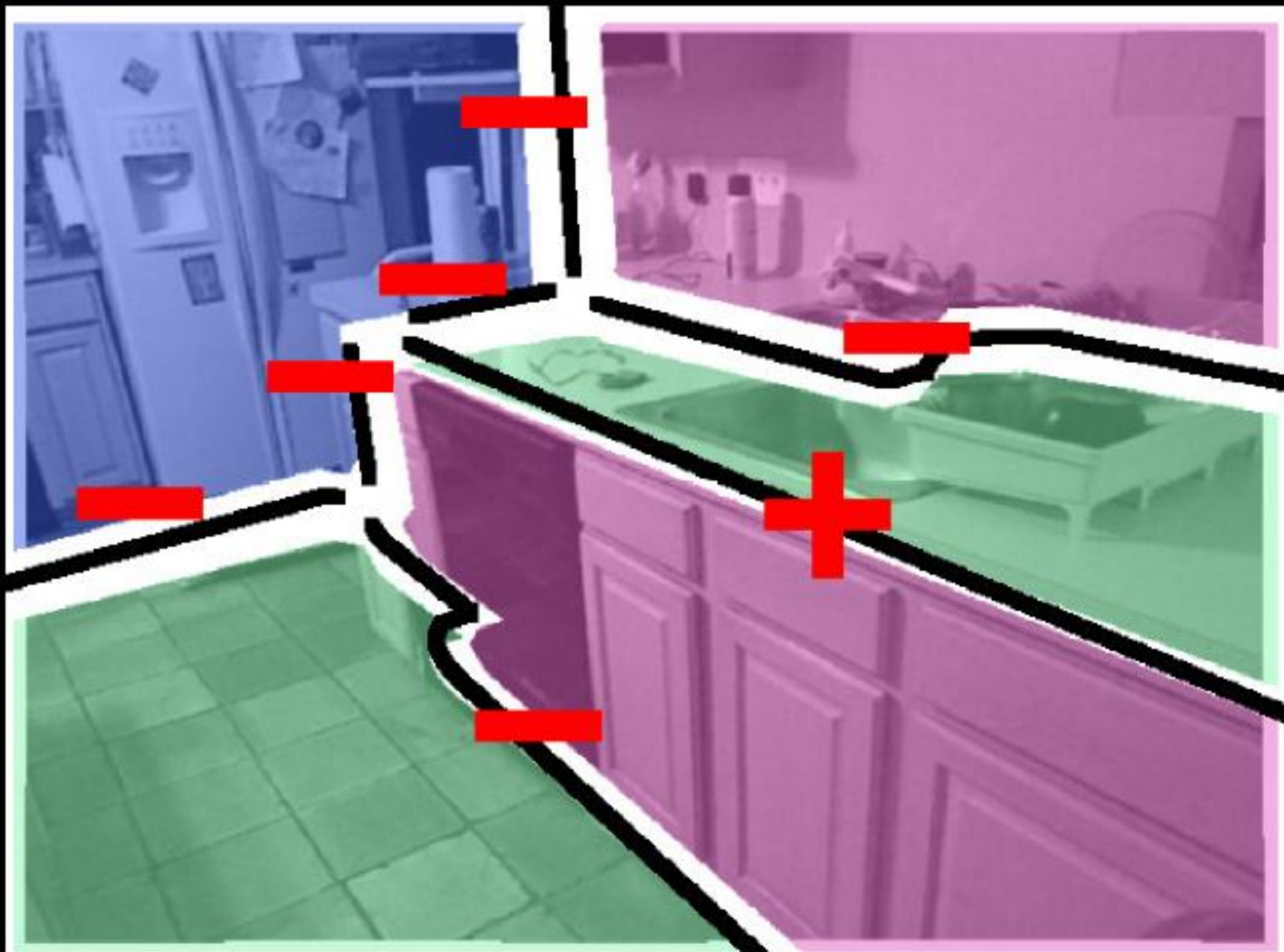
3D Primitives

Proposed

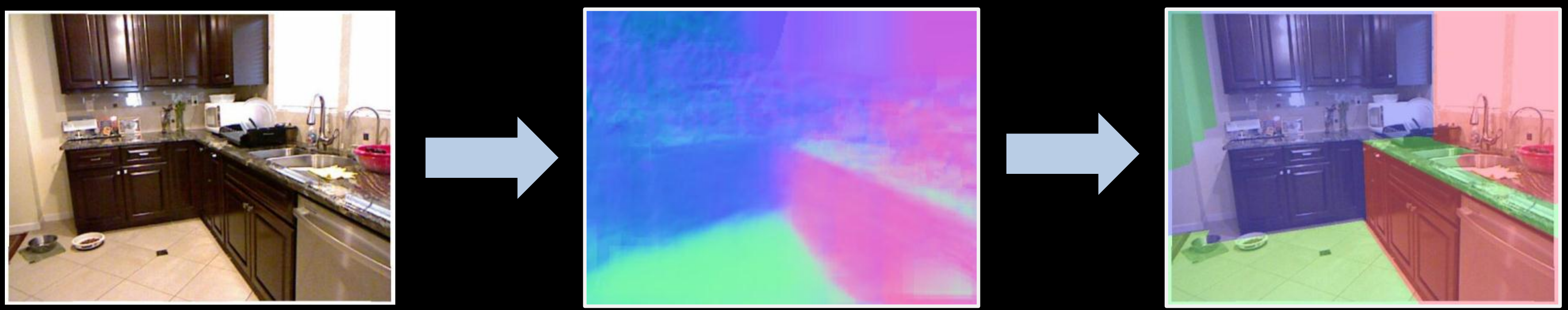
Quantitative Results

	Summary Stats (^{0}) (Lower Better)			% Good Pixels (Higher Better)		
	Mean	Median	RMSE	11.25^0	22.5^0	30^0
Proposed	**37.5**	**17.2**	**53.2**	**41.9**	**53.9**	**58.0**
3D Primitives	38.5	19.0	54.2	41.7	52.4	56.3
Hedau et al.	43.2	24.8	59.4	39.1	48.8	52.3
Lee et al.	47.6	43.4	60.6	28.1	39.7	43.9
Karsch et al.	46.6	43.0	53.6	5.4	19.9	31.5
Hoiem et al.	45.6	38.2	55.1	8.6	30.5	41.0

rank



+ Convex − Concave



Now:

Better reasoning

Semantic information

Less structured environments

Coarse-to-fine depth

Martial Hebert
Abhinav Gupta
David Fouhey, Adrien Matricon, Wajahat Hussain

ONR MURI
NDSEG
Bosch R&D