Using geometric information in recognition and scene analysis

Martial Hebert

Carnegie Mellon University

Collaborators

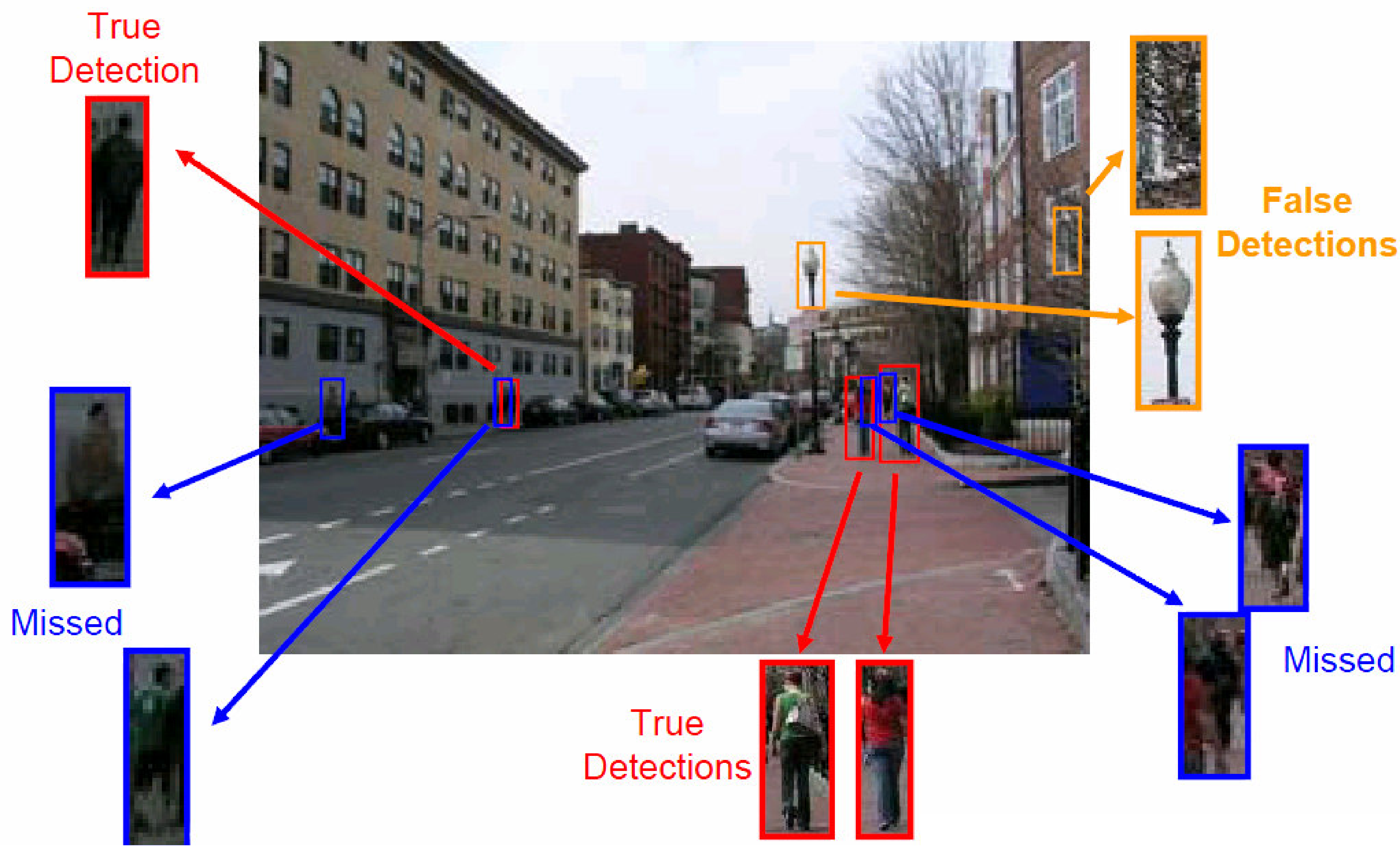
- Aloysha Efros
- Derek Hoiem

- David Lee

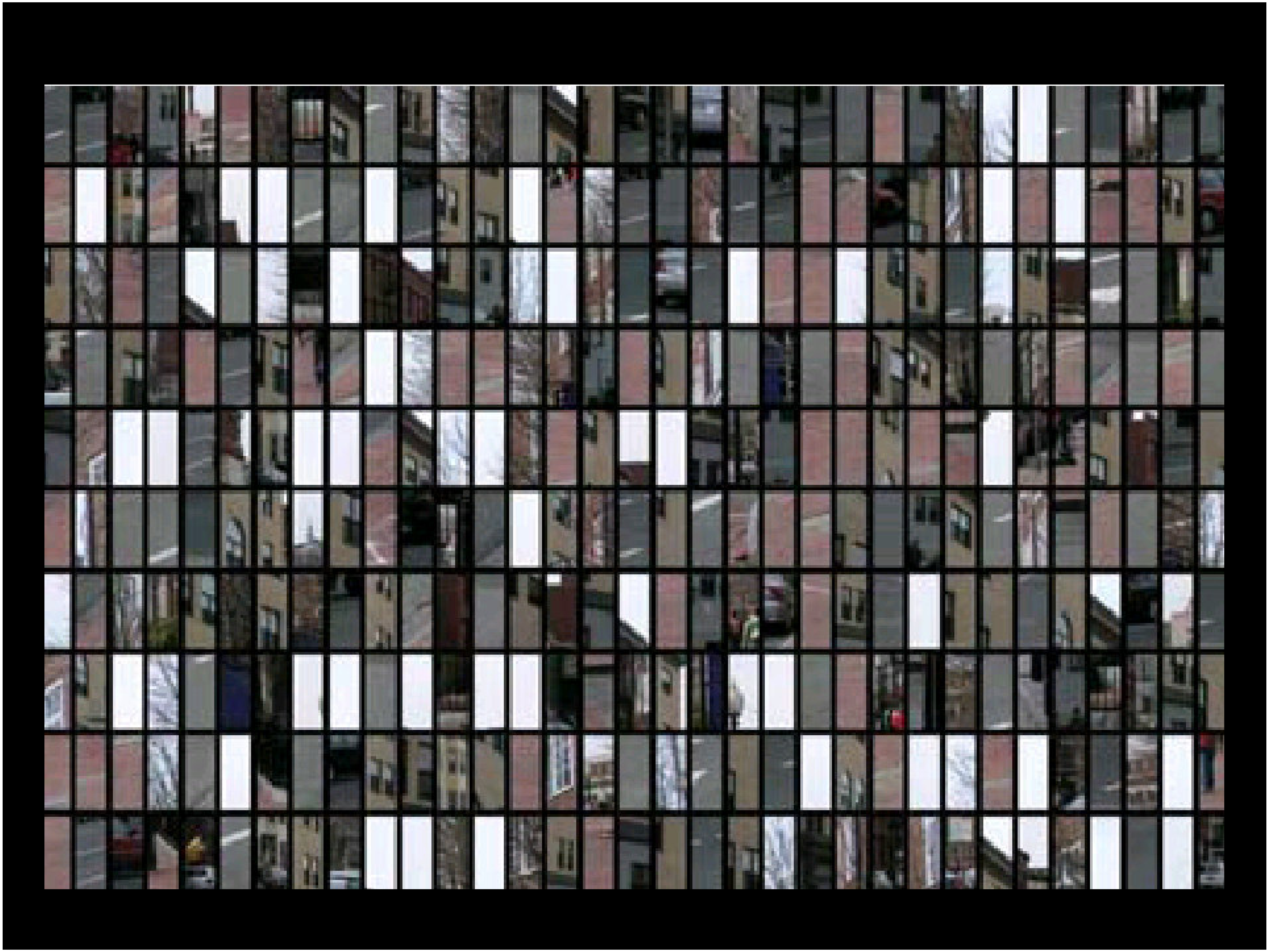
- Abhinav Gupta

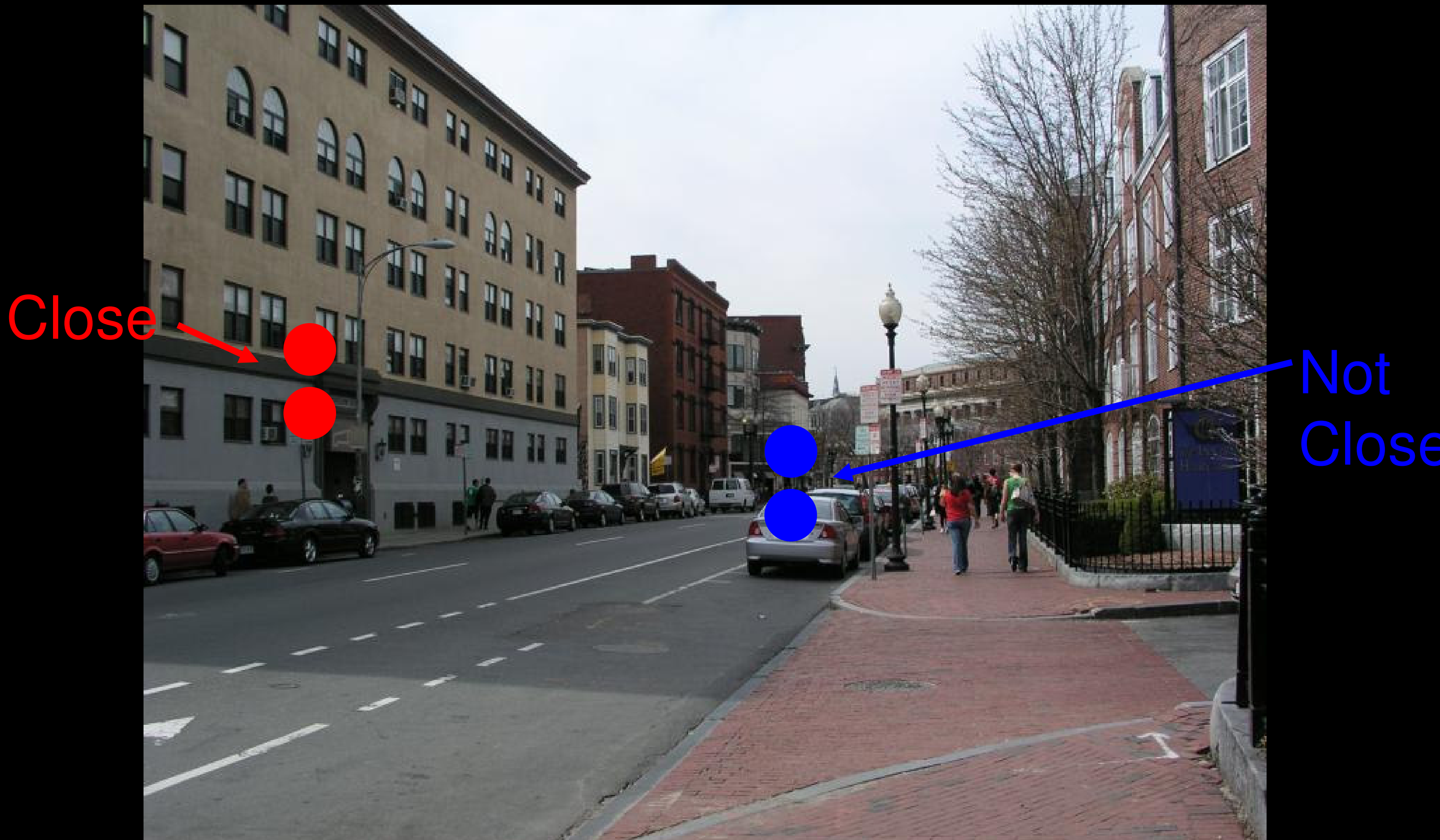
- Drew Bagnell
- Daniel Munoz
- Nicolas Vandapel

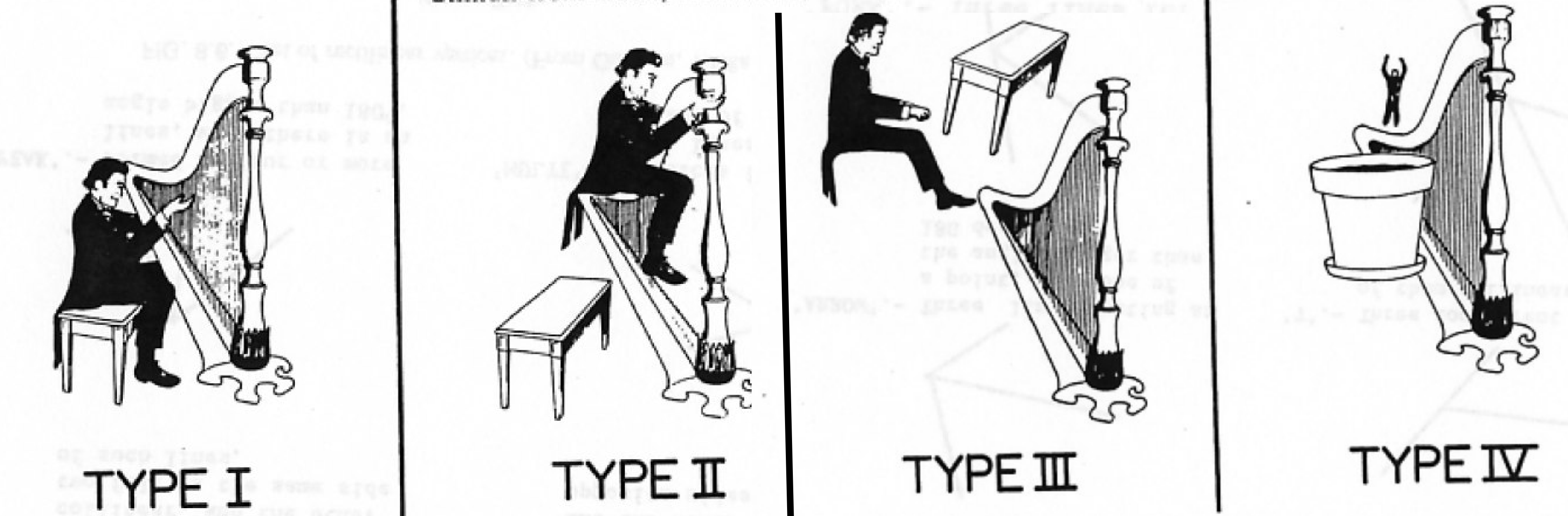
[redacted]



[redacted]



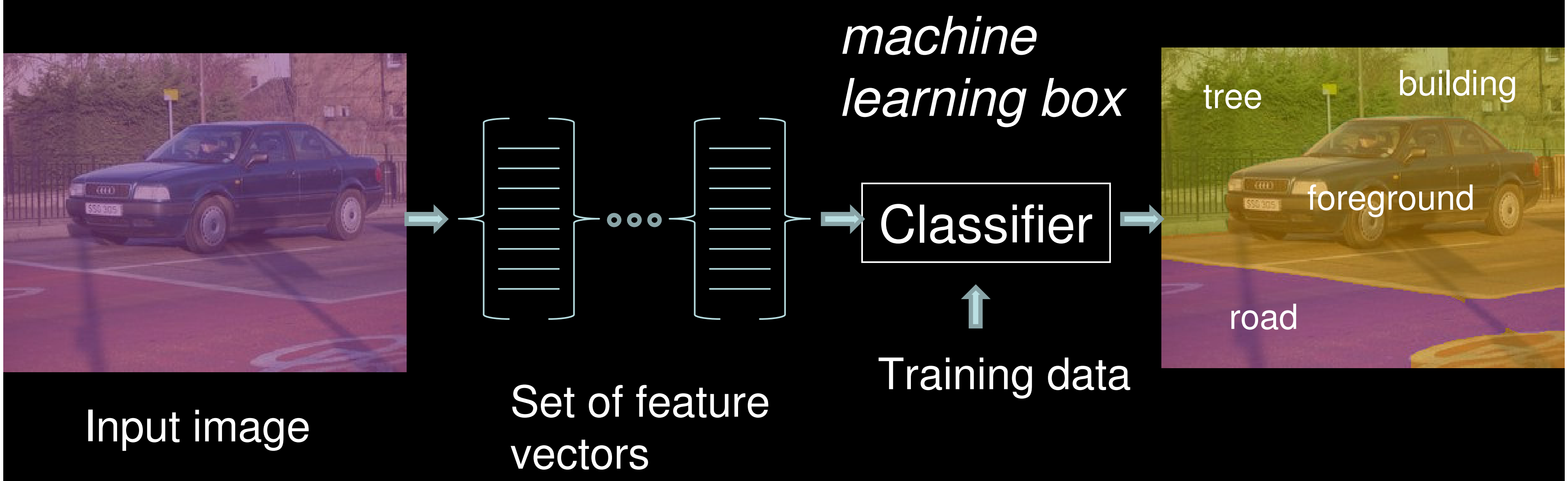

Close
Not
Close

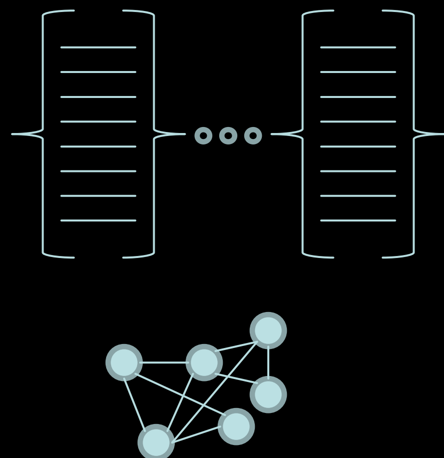


Stimuli from Hock, Romanski, Galie, and Williams (1978).

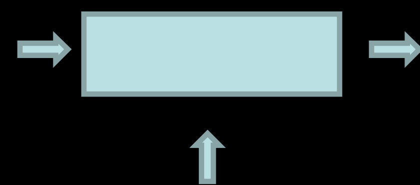
- Biederman's relations in a well-formed scene (1981):

1. *Support* (e.g., a floating fire hydrant). The object does not appear to be resting on a surface.
2. *Interposition* (e.g., the background appearing through the hydrant). The objects undergoing this violation appear to be transparent or passing through another object.
3. *Probability* (e.g., the hydrant in a kitchen). The object is unlikely to appear in the scene.
4. *Position* (e.g., the fire hydrant on top of a mailbox in a street scene). The object is likely to occur in that scene, but it is unlikely to be in that particular position.
5. *Size* (e.g., the fire hydrant appearing larger than a building). The object appears to be too large or too small relative to the other objects in the scene.


machine
learning box
Input image
Set of feature
vectors
Classifier
Training data
tree
building
foreground
road



Machine learning box
Reasoning



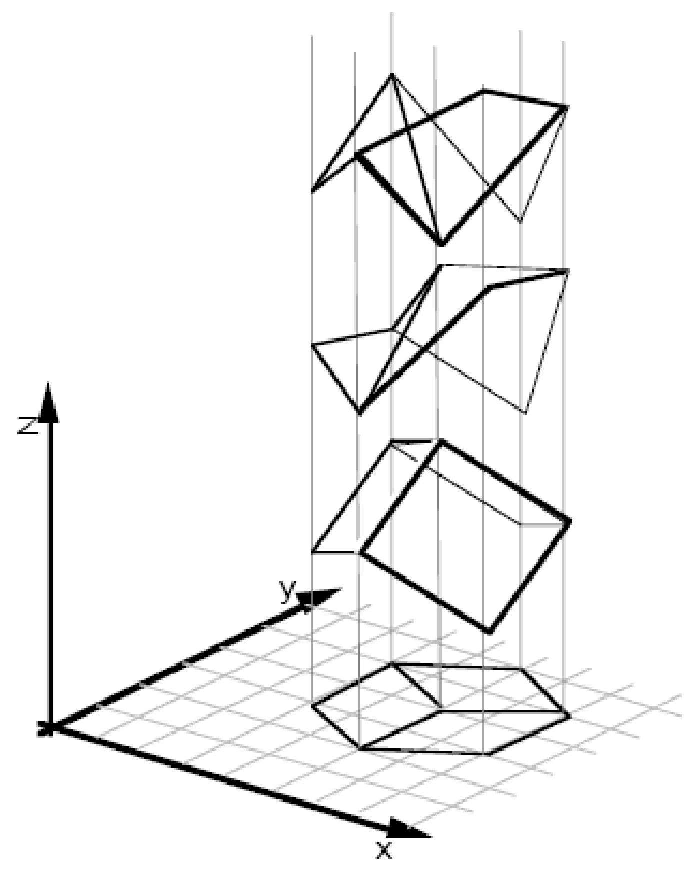
Input image

Set of feature of vectors
+ additional structure (e.g., geometry, relations)

Training data +
Geometry, relational information, physics, domain knowledge

Conclusion I

- *Qualitative* 3D information can be estimated and can be used effectively



from [Sinha and Adelson 1993]

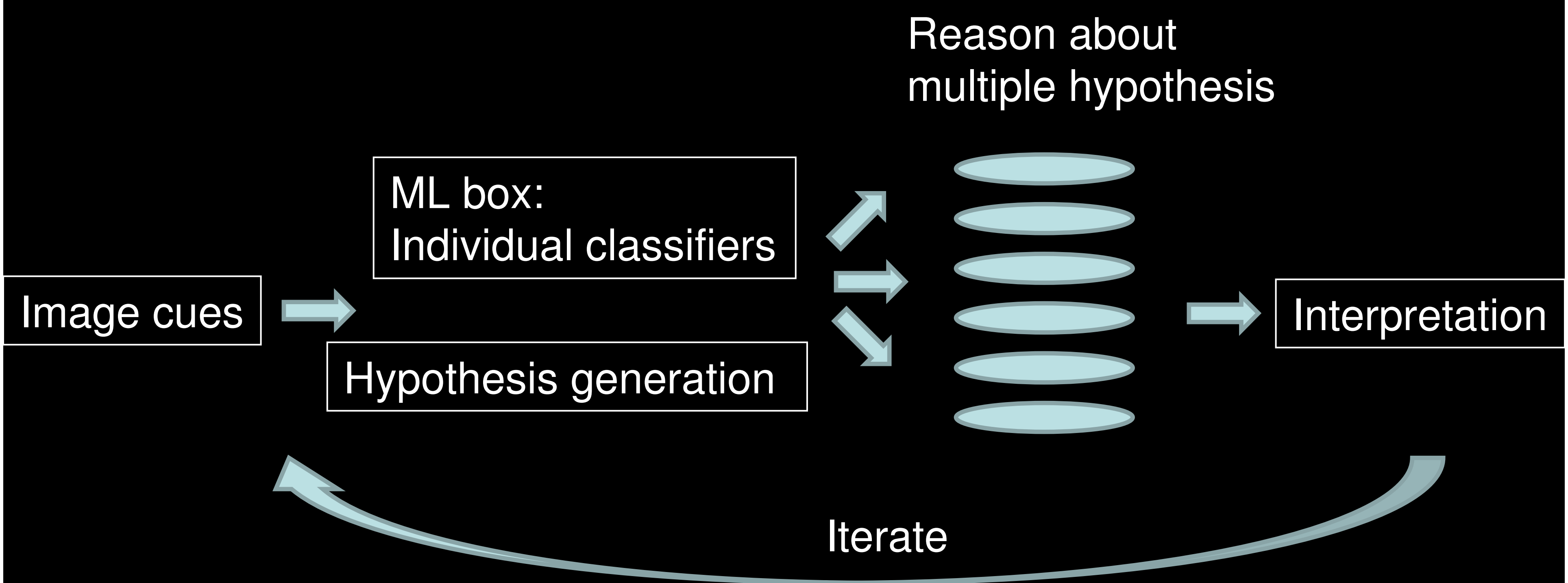
Conclusion II

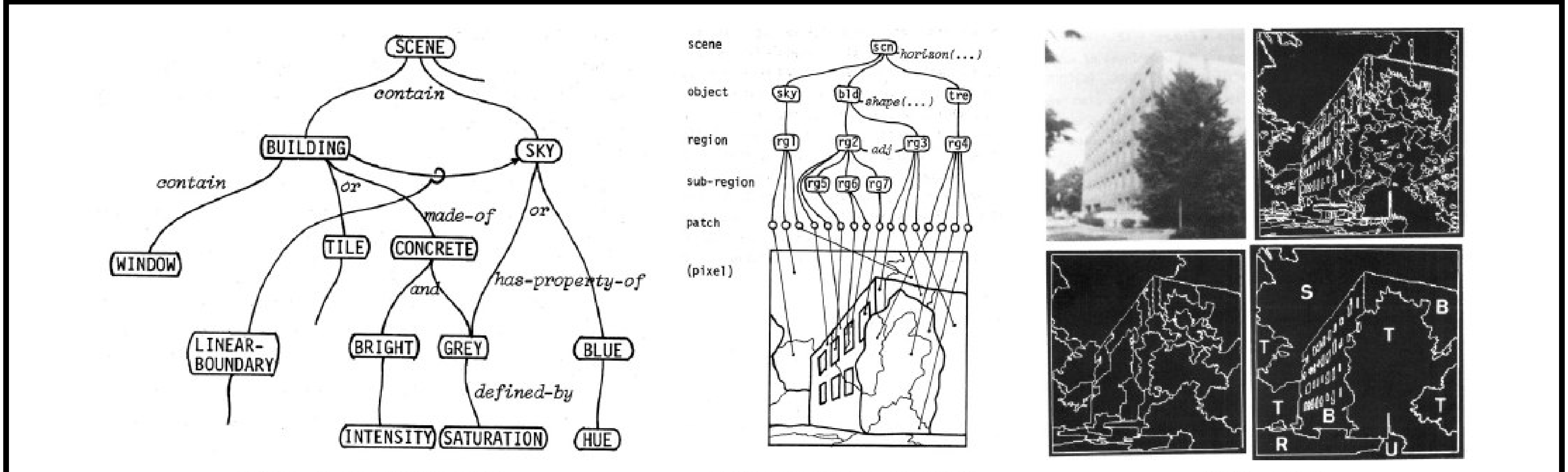
- Use reasoning/search about multiple hypotheses and interpretations in addition to “standard” learned classifiers

Image cues → ML box: Interpretation from learned models → Interpretation

Conclusion II

- Use reasoning/search about multiple hypotheses and interpretations in addition to “standard” learned classifiers





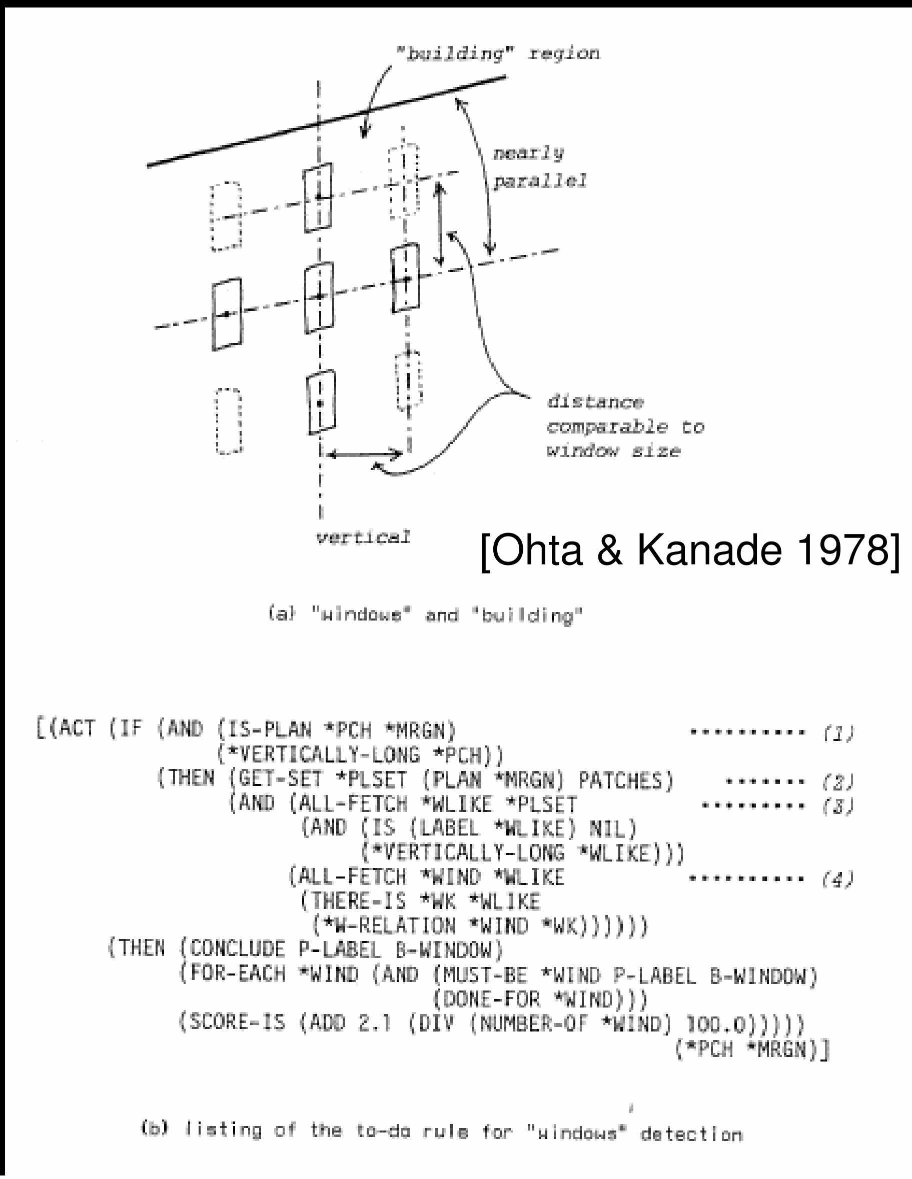
(a) Bottom-up process (b) Top-down process (c) Result

[Ohta & Kanade 1978]

- Guzman (*SEE*), 1968
- Yakimovsky & Feldman, 1973
- Hansen & Riseman (*VISIONS*), 1978
- Barrow & Tenenbaum 1978
- Brooks (*ACRONYM*), 1979
- Marr, 1982
- Ohta & Kanade, 1978

Then

Now



- Combine "modern" data-driven techniques (e.g., classifiers learned from training data) with *knowledge representations* and *reasoning tools* in integrated *control structure*

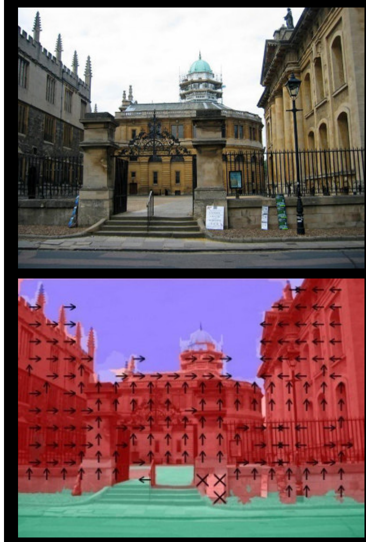
Levels of 3D-ness

Qualitative

More quantitative
more precise

Explicit

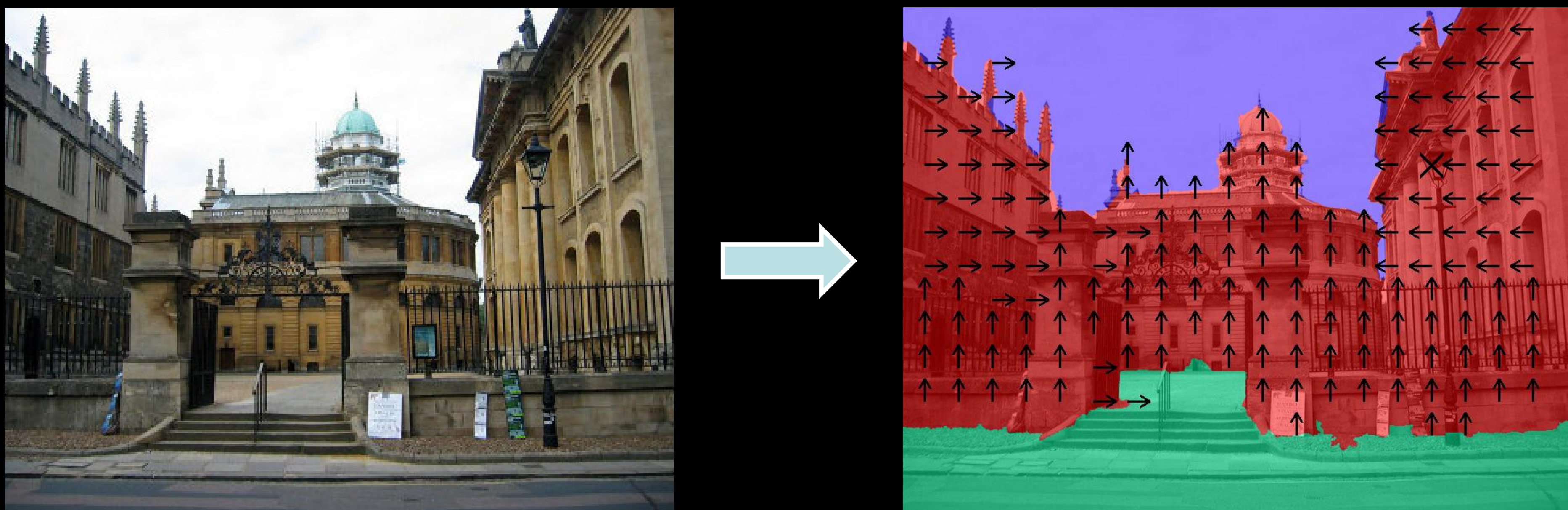
Levels of 3D-ness



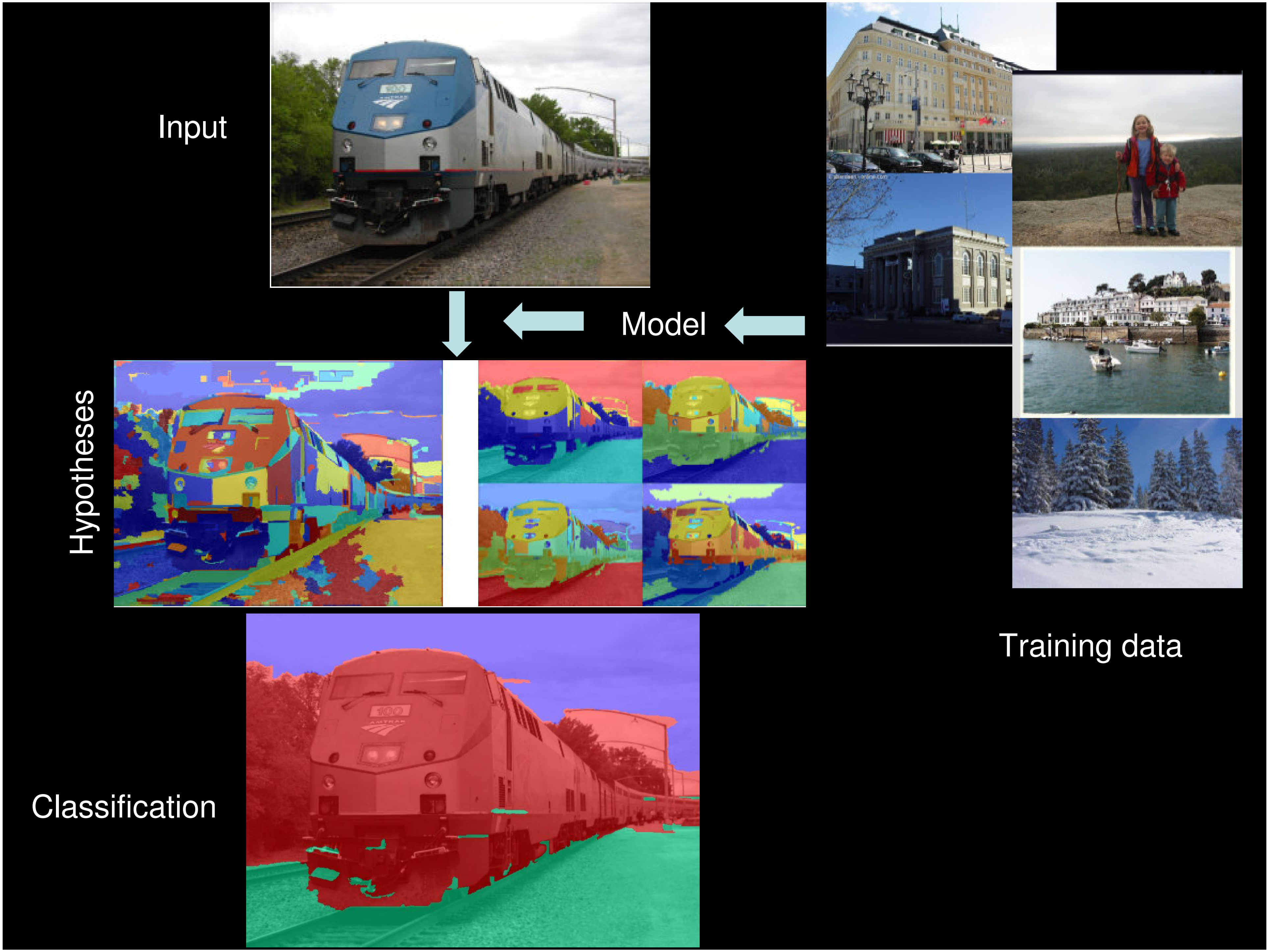
Region labels

Qualitative

First attempt: Estimate surface labels



[D. Hoiem, A. A. Efros, and M. Hebert. *Recovering surface layout from an image*. IJCV, 75(1):151–172, 2007]


Input
Model
Hypotheses
Classification
Training data

Cues used to design features

Vanishing points, lines

Color, texture, image location

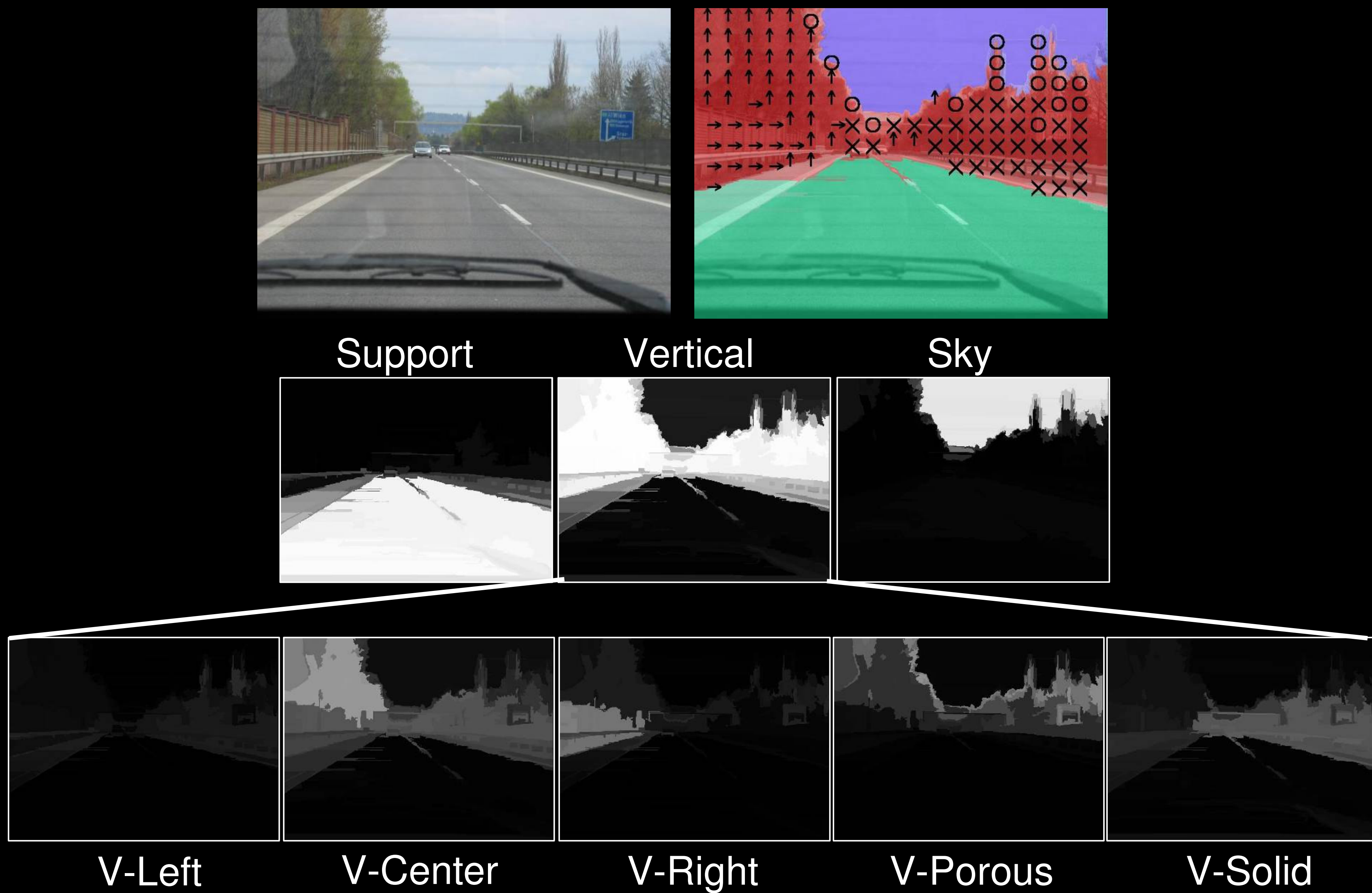
Texture gradient

Example features

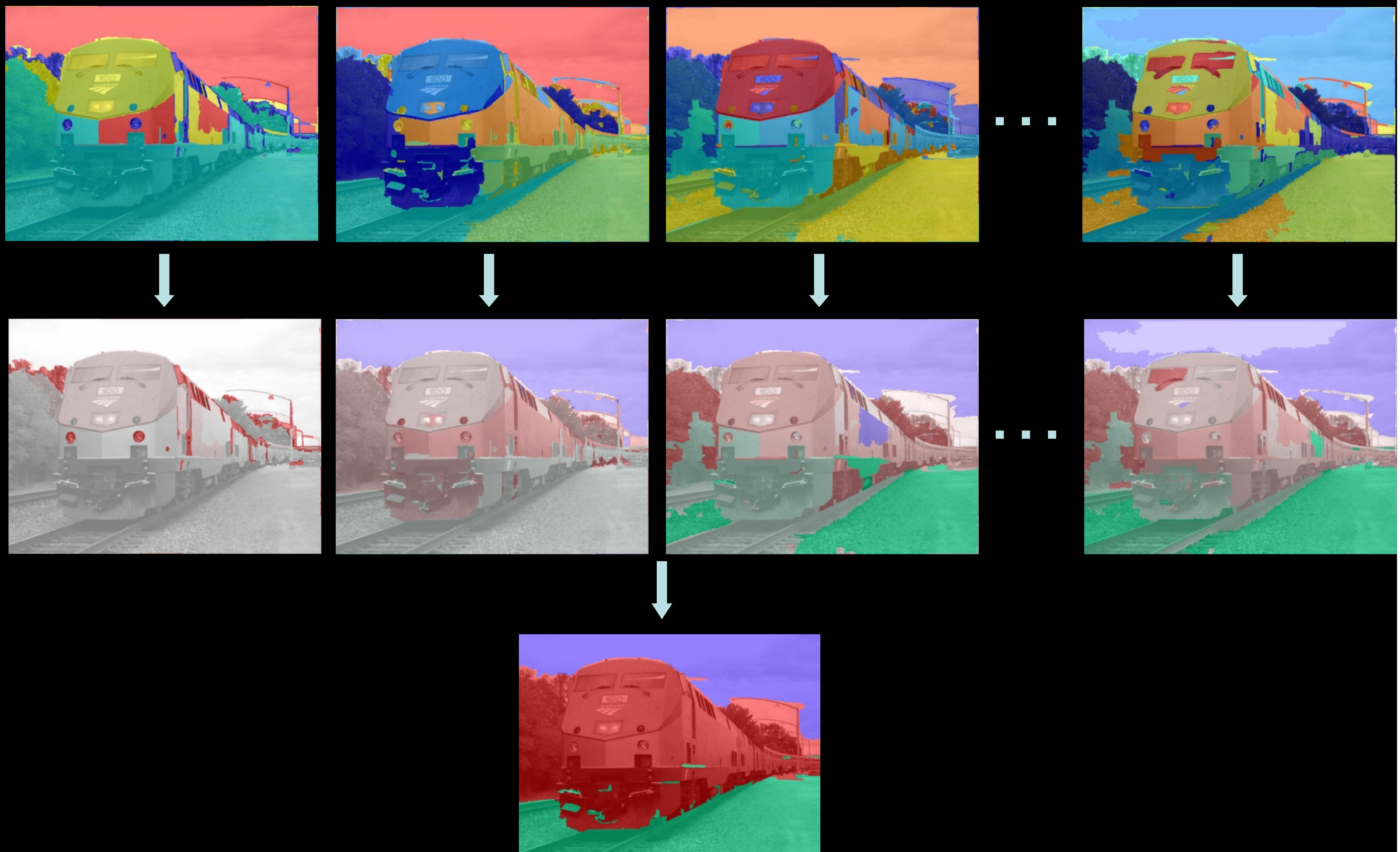
- Material
- Image Location
- Perspective
- Input to boosted decision tree classifier

SURFACE CUES
Location and Shape
L1. Location: normalized x and y, mean
L2. Location: norm. x and y, 10^{th} and 90^{th} pctl
L3. Location: norm. y wrt estimated horizon, 10^{th}, 90^{th} pctl
L4. Location: whether segment is above, below, or straddles estimated horizon
L5. Shape: number of superpixels in segment
L6. Shape: normalized area in image
Color
C1. RGB values: mean
C2. HSV values: C1 in HSV space
C3. Hue: histogram (5 bins)
C4. Saturation: histogram (3 bins)
Texture
T1. LM filters: mean abs response (15 filters)
T2. LM filters: hist. of maximum responses (15 bins)
Perspective
P1. Long Lines: (num line pixels)/sqrt(area)
P2. Long Lines: % of nearly parallel pairs of lines
P3. Line Intersections: hist. over 8 orientations, entropy
P4. Line Intersections: % right of center
P5. Line Intersections: % above center
P6. Line Intersections: % far from center at 8 orientations
P7. Line Intersections: % very far from center at 8 orientations
P8. Vanishing Points: (num line pixels with vertical VP membership)/sqrt(area)
P9. Vanishing Points: (num line pixels with horizontal VP membership)/sqrt(area)
P10. Vanishing Points: percent of total line pixels with vertical VP membership
P11. Vanishing Points: x-pos of horizontal VP - segment center (0 if none)
P12. Vanishing Points: y-pos of highest/lowest vertical VP wrt segment center
P13. Vanishing Points: segment bounds wrt horizontal VP
P14. Gradient: x, y center of gradient mag. wrt. image center

Output

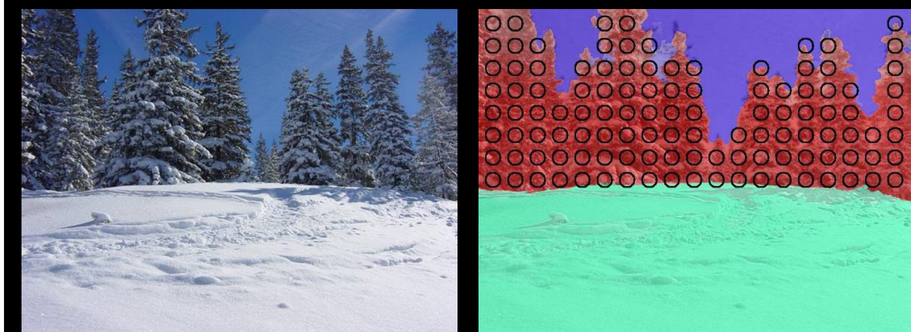


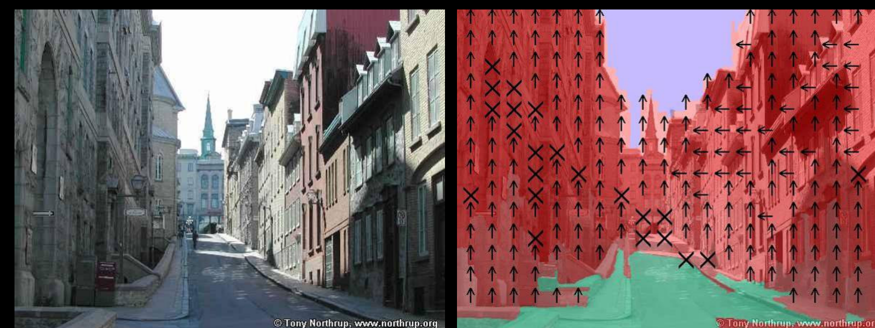
Using multiple segmentations

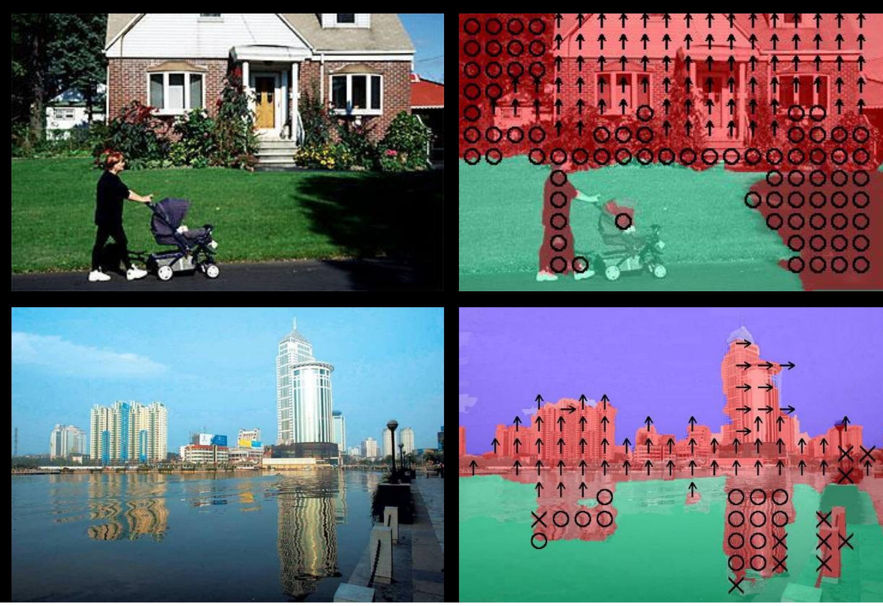


Labeled Pixels

Sample outputs



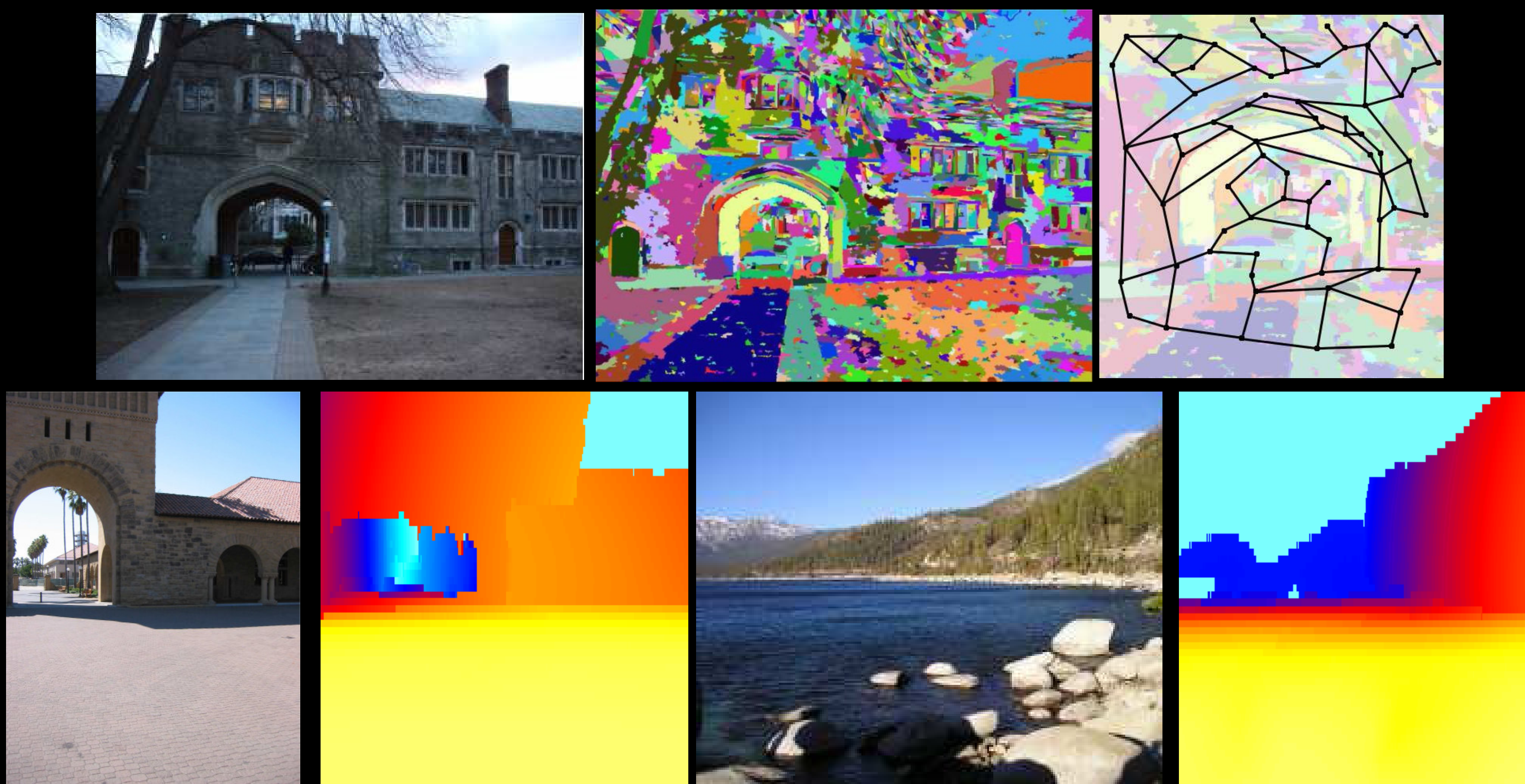




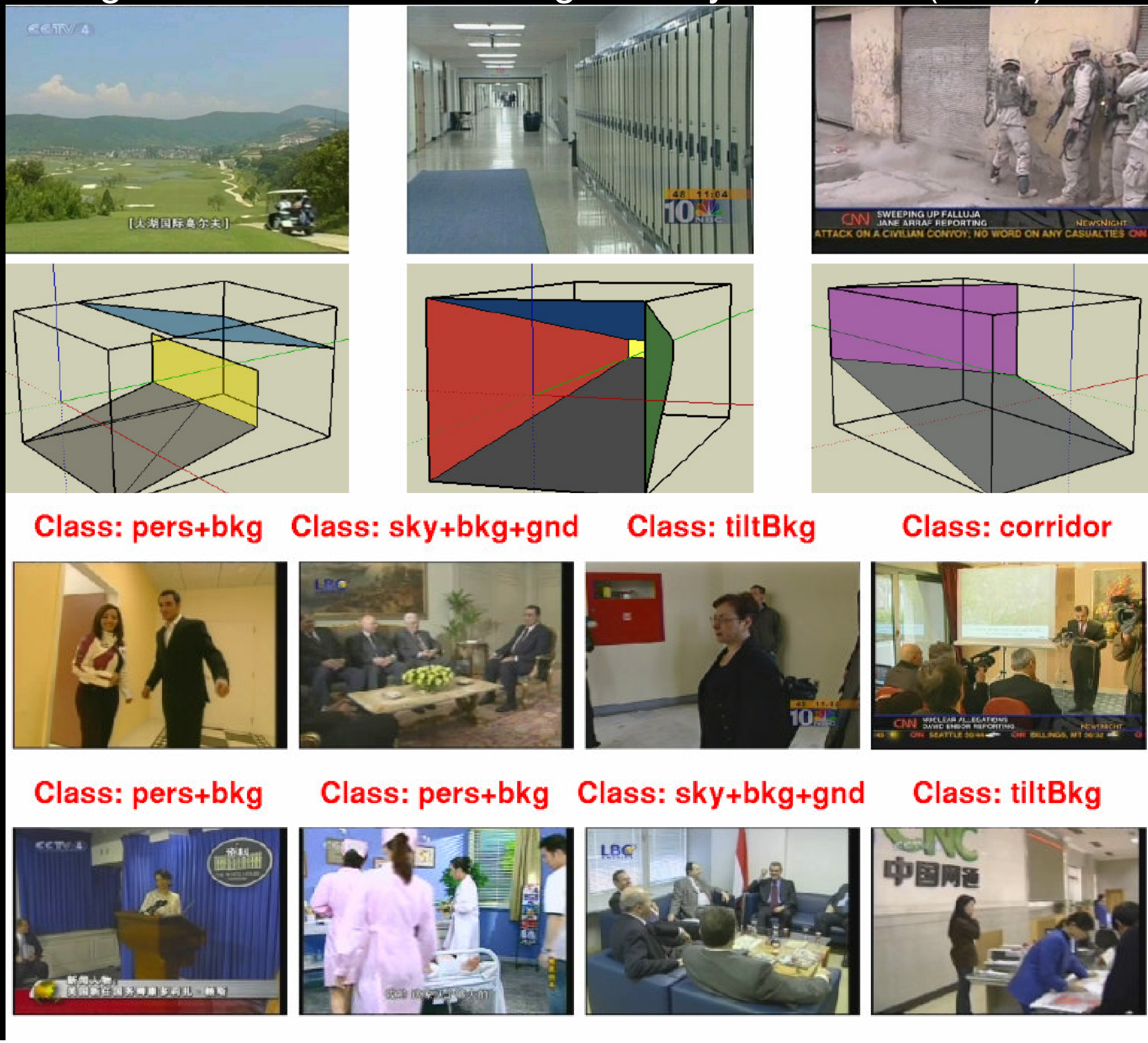
Main Class: 88.1%
Subclasses: 61.5%

[redacted]

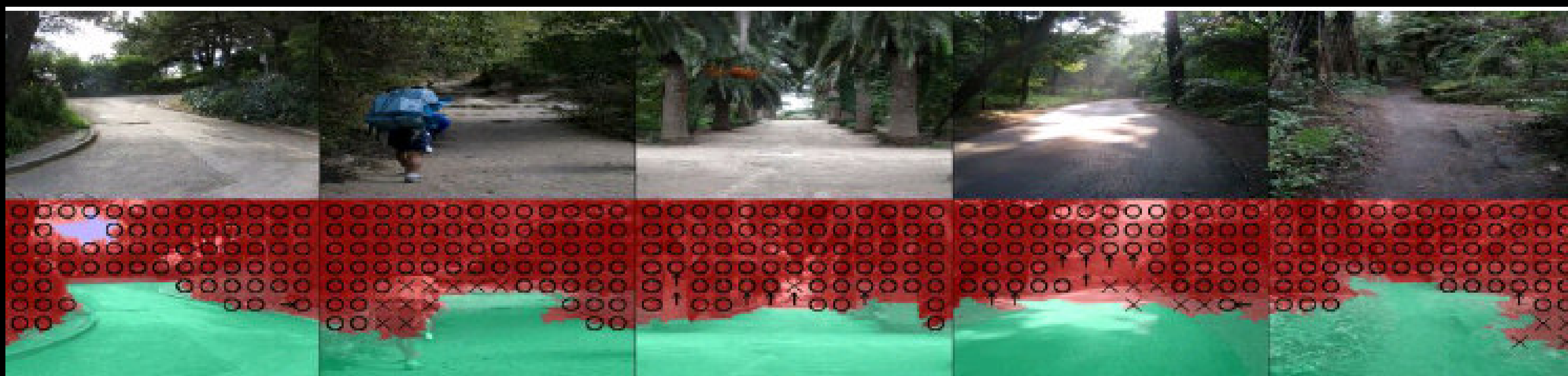
- *Learning from image features to depth + MRF:* A. Saxena, S. H. Chung, and A. Y. Ng. *3-D depth reconstruction from a single still image*. IJCV, 76, 2007.



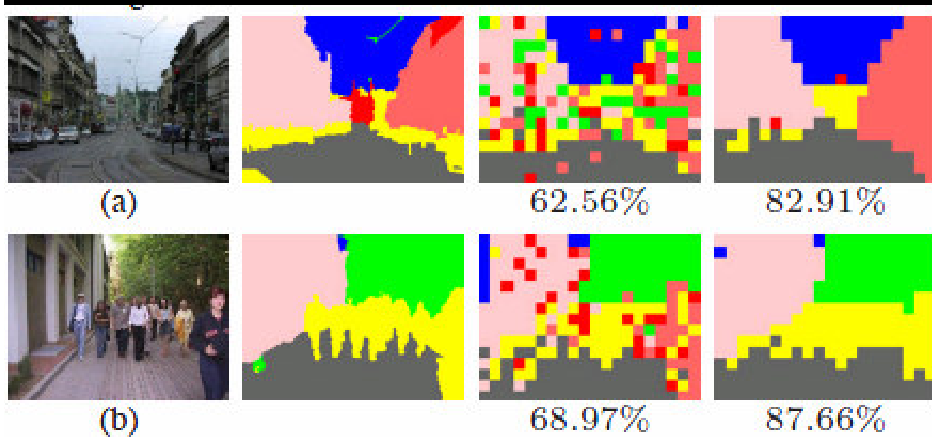
- *Stage classes:* Nedovic, V., Smeulders, A., Redert, A., Geusebroek, J.: Stages as models of scene geometry. In: PAMI (2010)

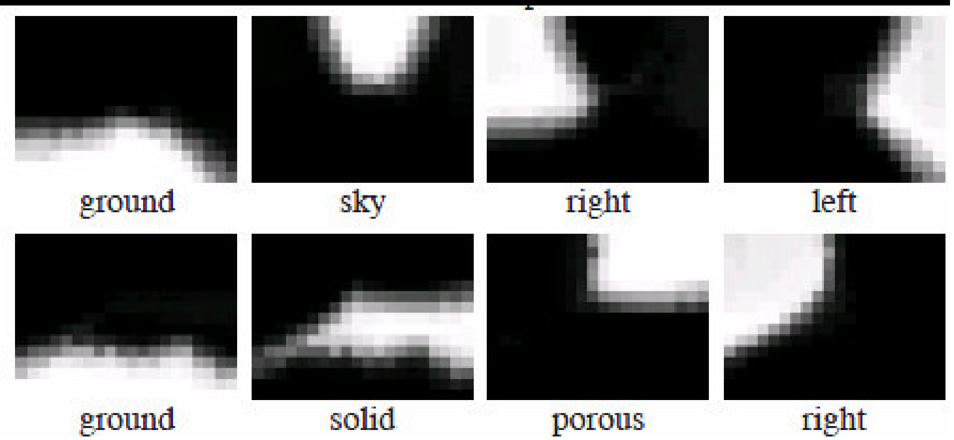


- S. Divvala, A. Efros, and M. Hebert. Can *Similar Scenes help Surface Layout Estimation?* IEEE Workshop on Internet Vision, 2008.

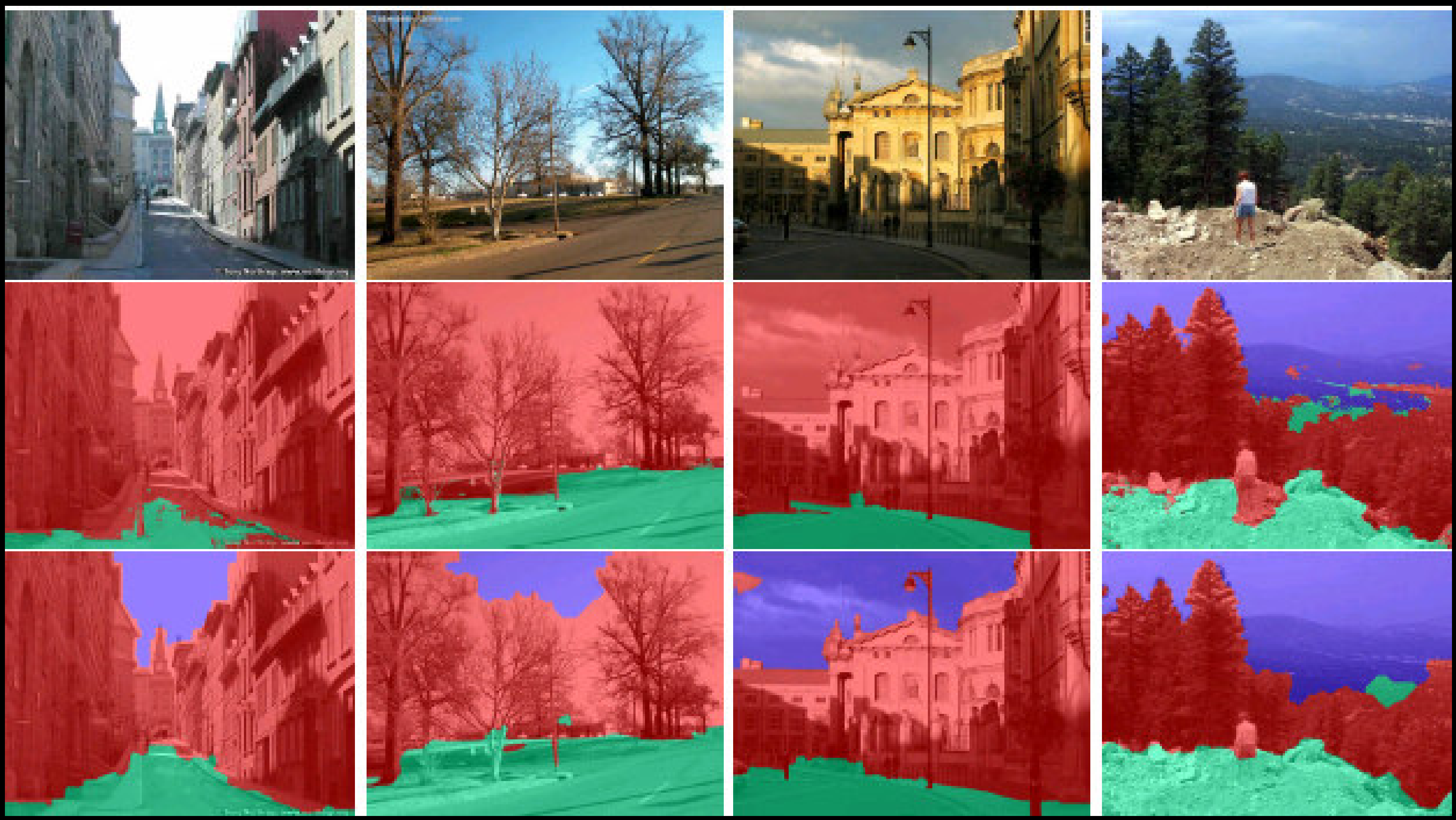


- Lazebnik, S., Raginsky, M.: *An empirical bayes approach to contextual region classification*. CVPR 2009.





- M. Szummer, P. Kohli, D. Hoiem. *Learning CRFs using Graph Cuts*. ECCV 2008.



Comments

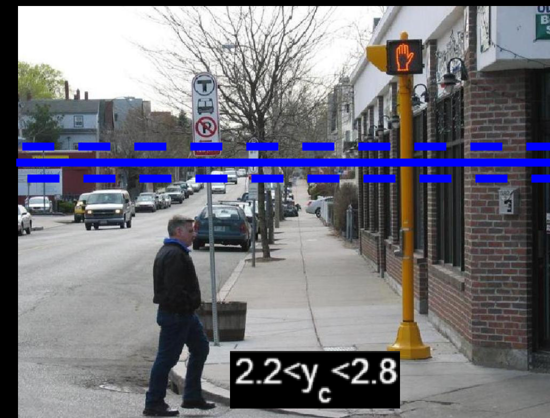
- Concept of qualitative 3D information from image cues
- Use of multiple segmentations combined with “standard” classifiers

Next….

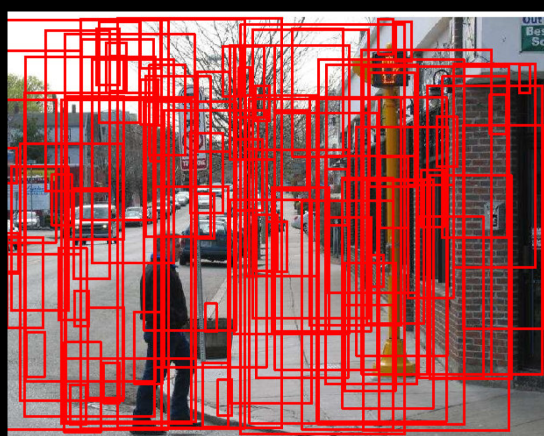
- Can coarse surface labels be used for improving object recognition and scene analysis performance through better geometric reasoning?

Image

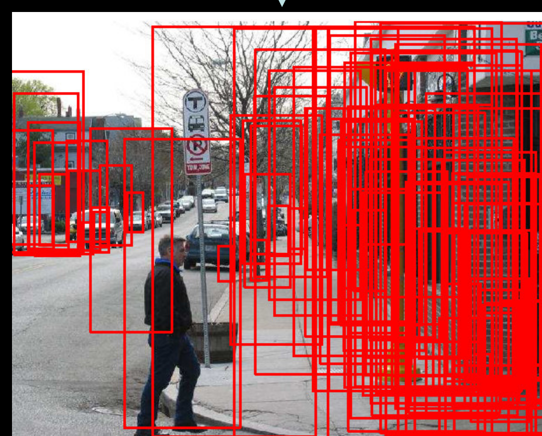
P(surfaces)



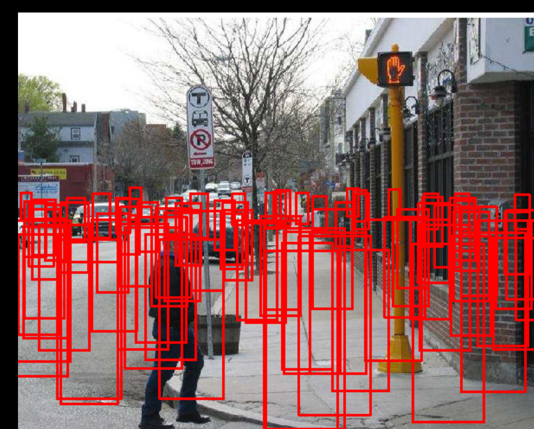
P(viewpoint)



P(object)



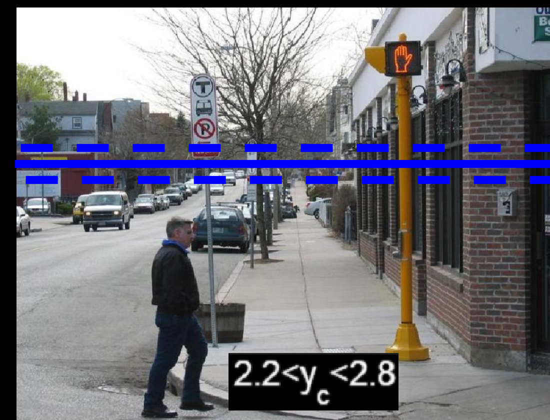
P(object | surfaces)



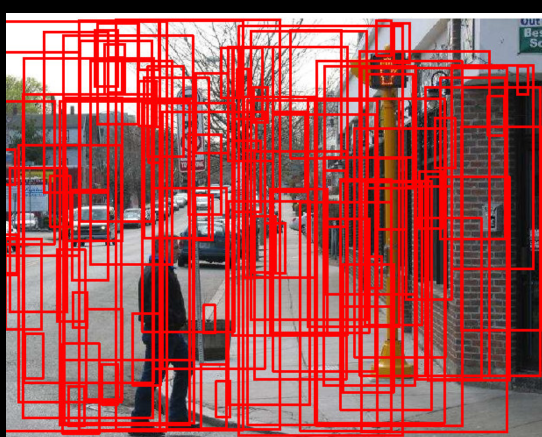
P(object | viewpoint)

Image

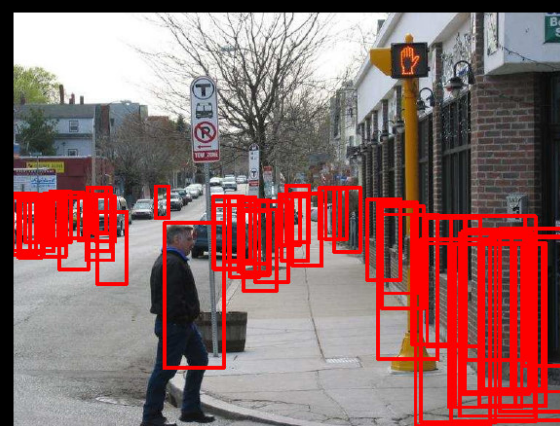
P(surfaces)



P(viewpoint)

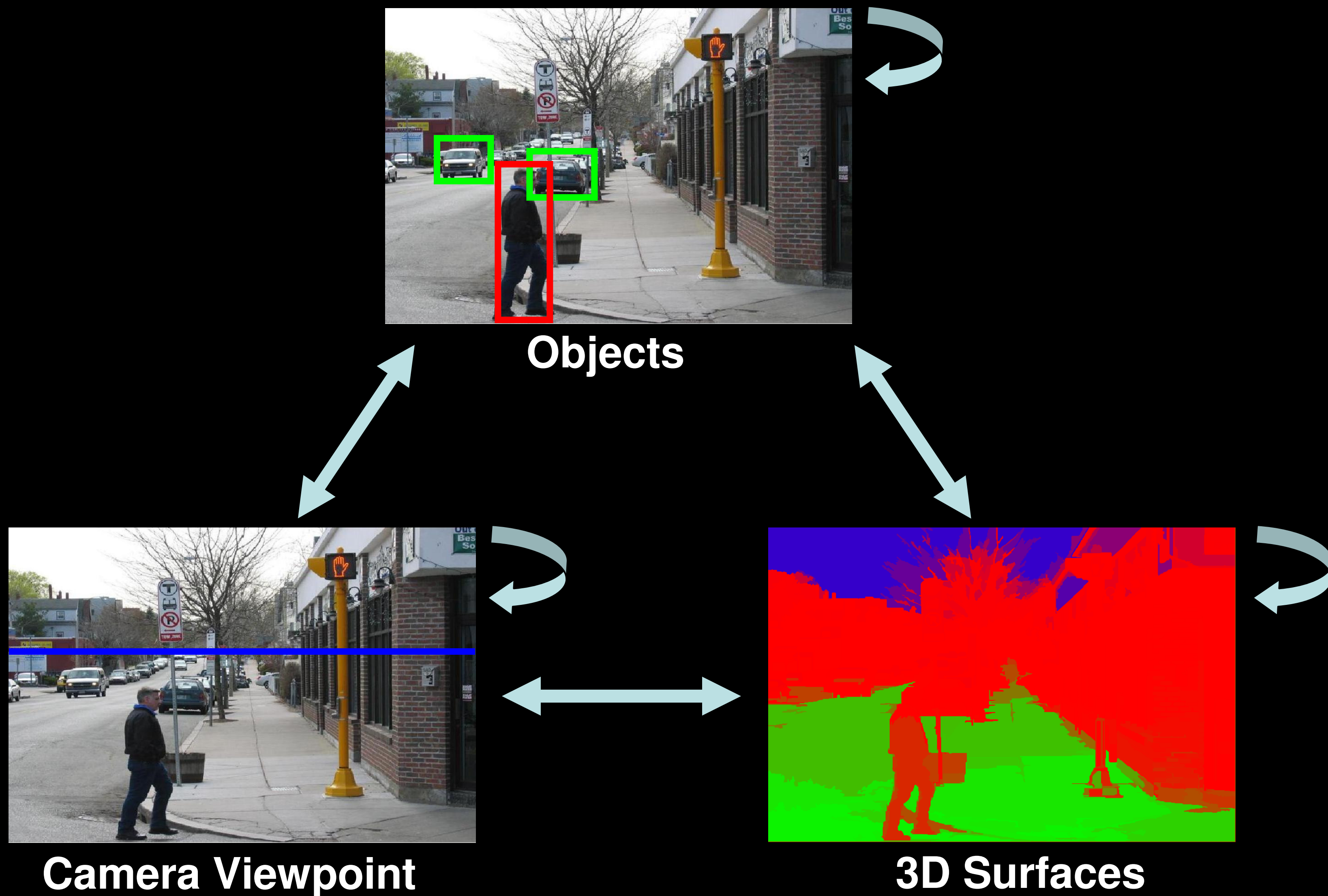


P(object)

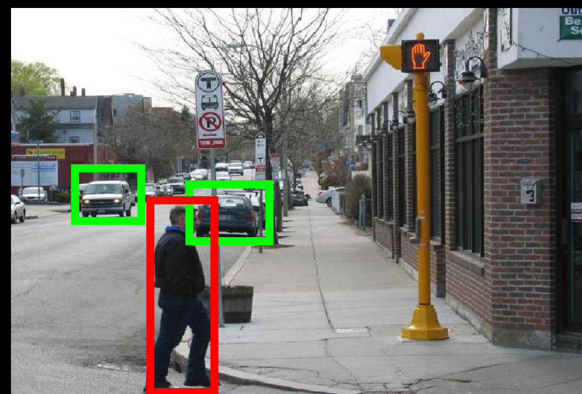


P(object | surfaces, viewpoint)

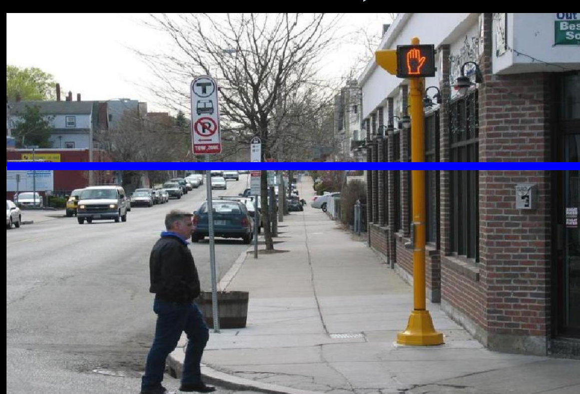
General model



Approximate model



Objects

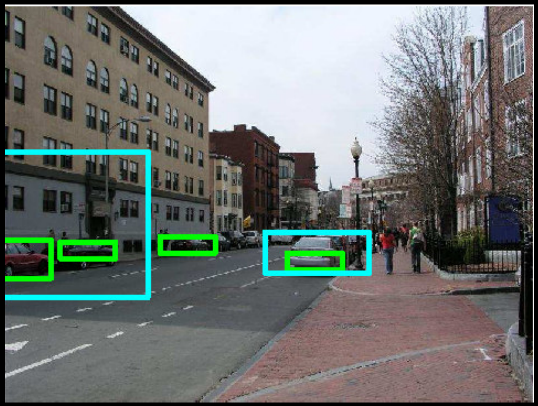


Viewpoint

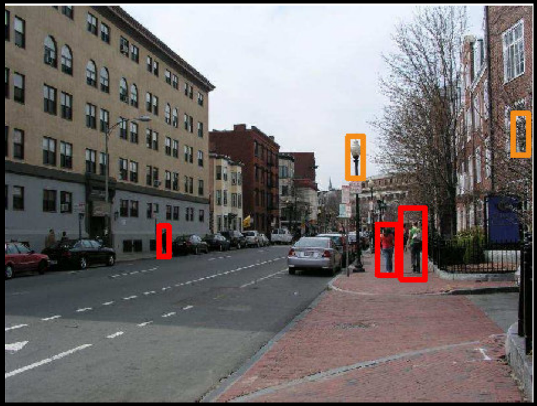
3D Surfaces

Input

Object Detection



Local Car Detector

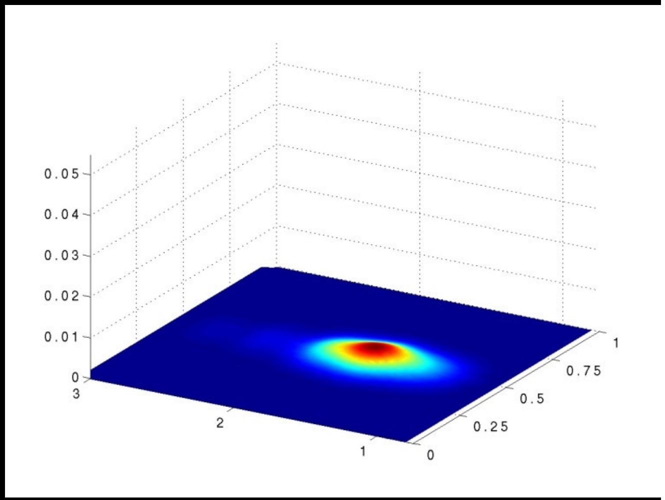


Local Ped Detector

[Dalal-Triggs 2005]

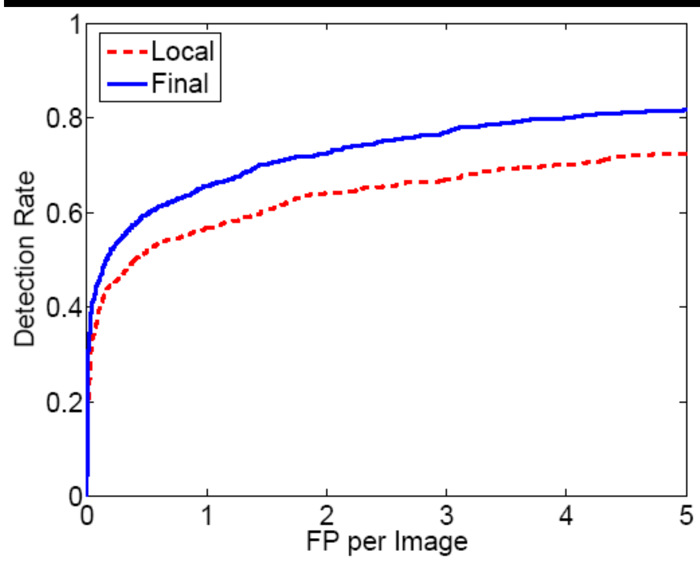
Surface Estimates

Viewpoint Prior



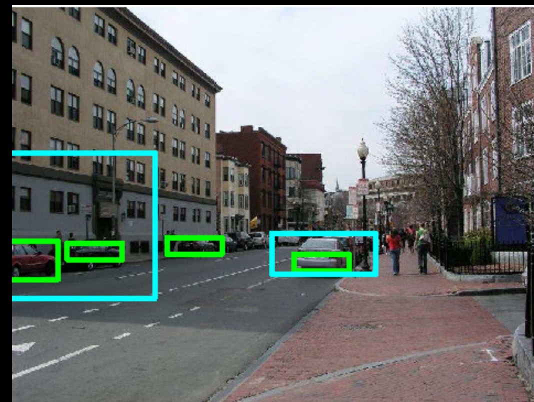
Car: TP / FP

Ped: TP / FP

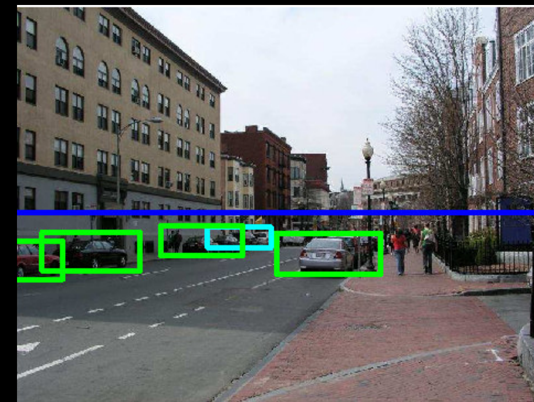


Initial (Local)

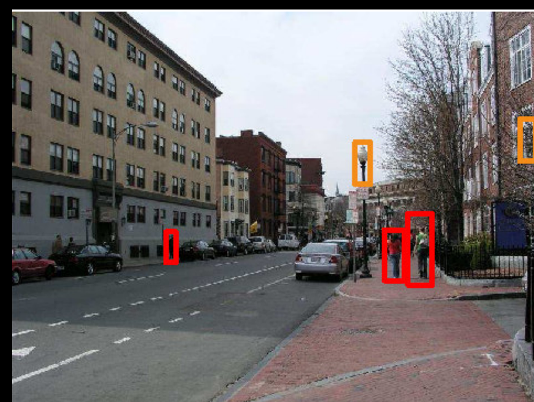
Final (Global)



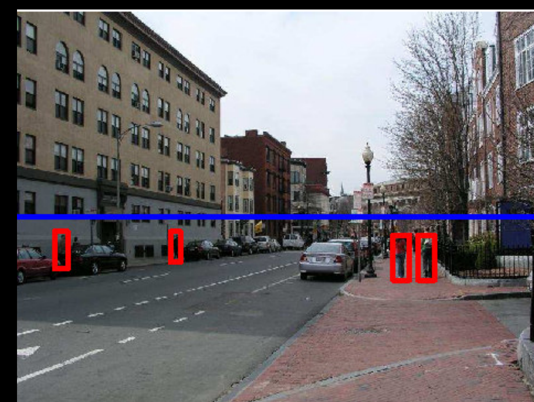
4 TP / 2 FP



4 TP / 1 FP



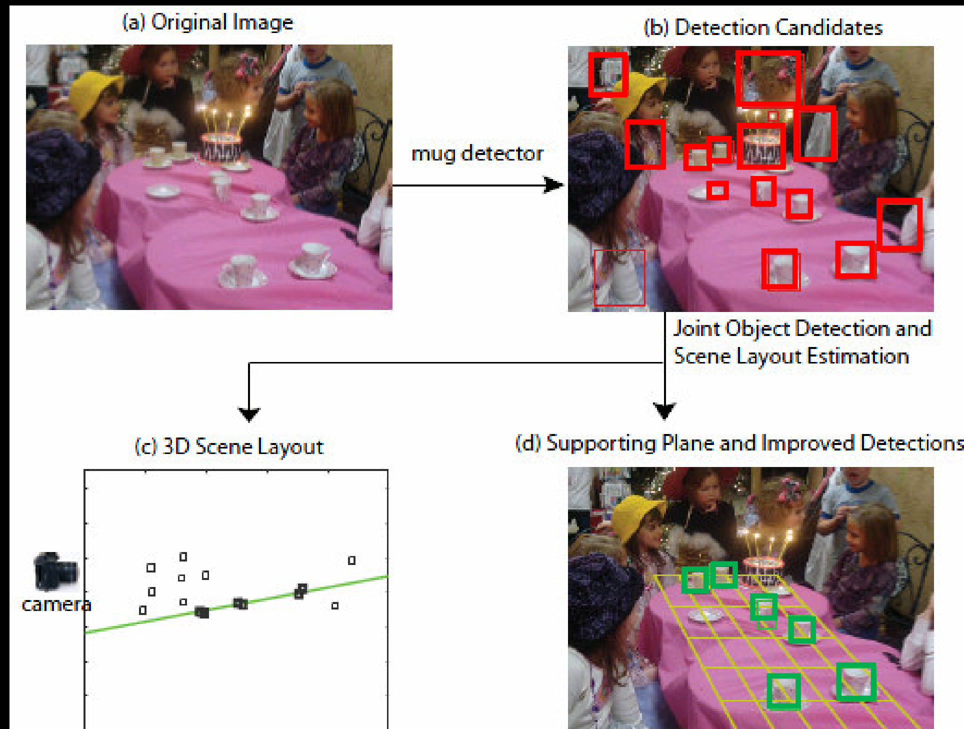
3 TP / 2 FP



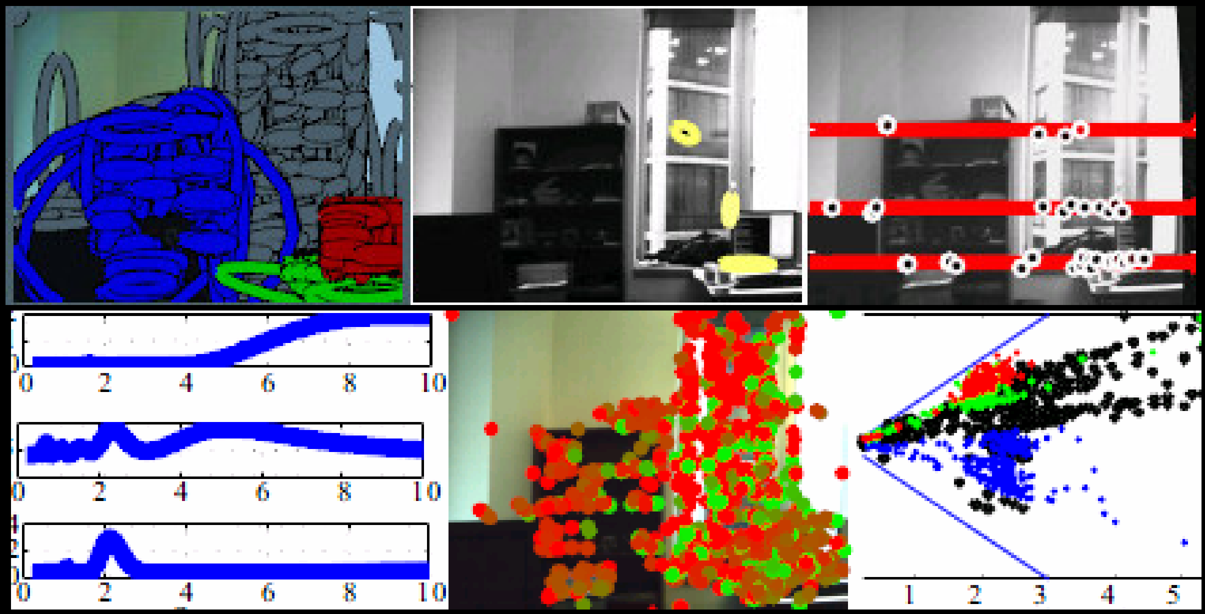
4 TP / 0 FP

[D. Hoiem, A. Efros, M. Hebert. Putting objects in perspective. IJCV 2009]

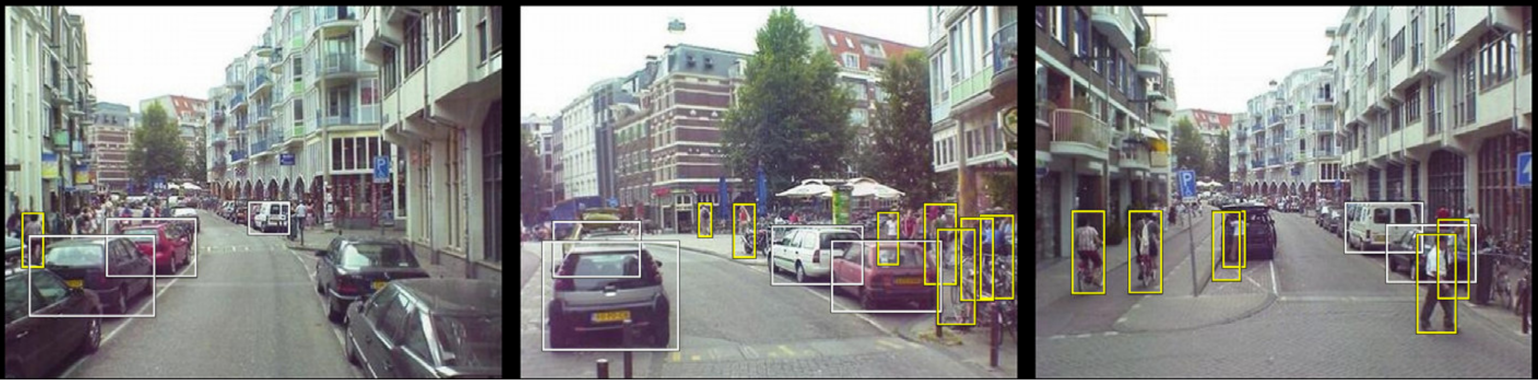
- S.Y. Bao, M. Sun, S.Savarese. *Toward Coherent Object Detection And Scene Layout Understand*ing. CVPR 2010.



- E. Sudderth, A. Torralba, W. T. Freeman, and A. S. Willsky. *Depth from Familiar Objects: A Hierarchical Model for 3D Scenes.* CVPR 2006.



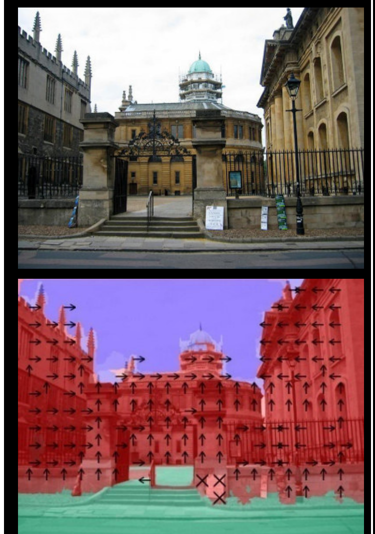
- B. Leibe, N. Cornelis, K. Cornelis, and L. Van Gool. *Dynamic 3D Scene Analysis from a Moving Vehicle*. CVPR07



- Is a more precise representation possible?

- For example: We would like to include reasoning about interposition (relations between object relative to a viewpoint induced by occlusion boundaries)

Levels of 3D-ness

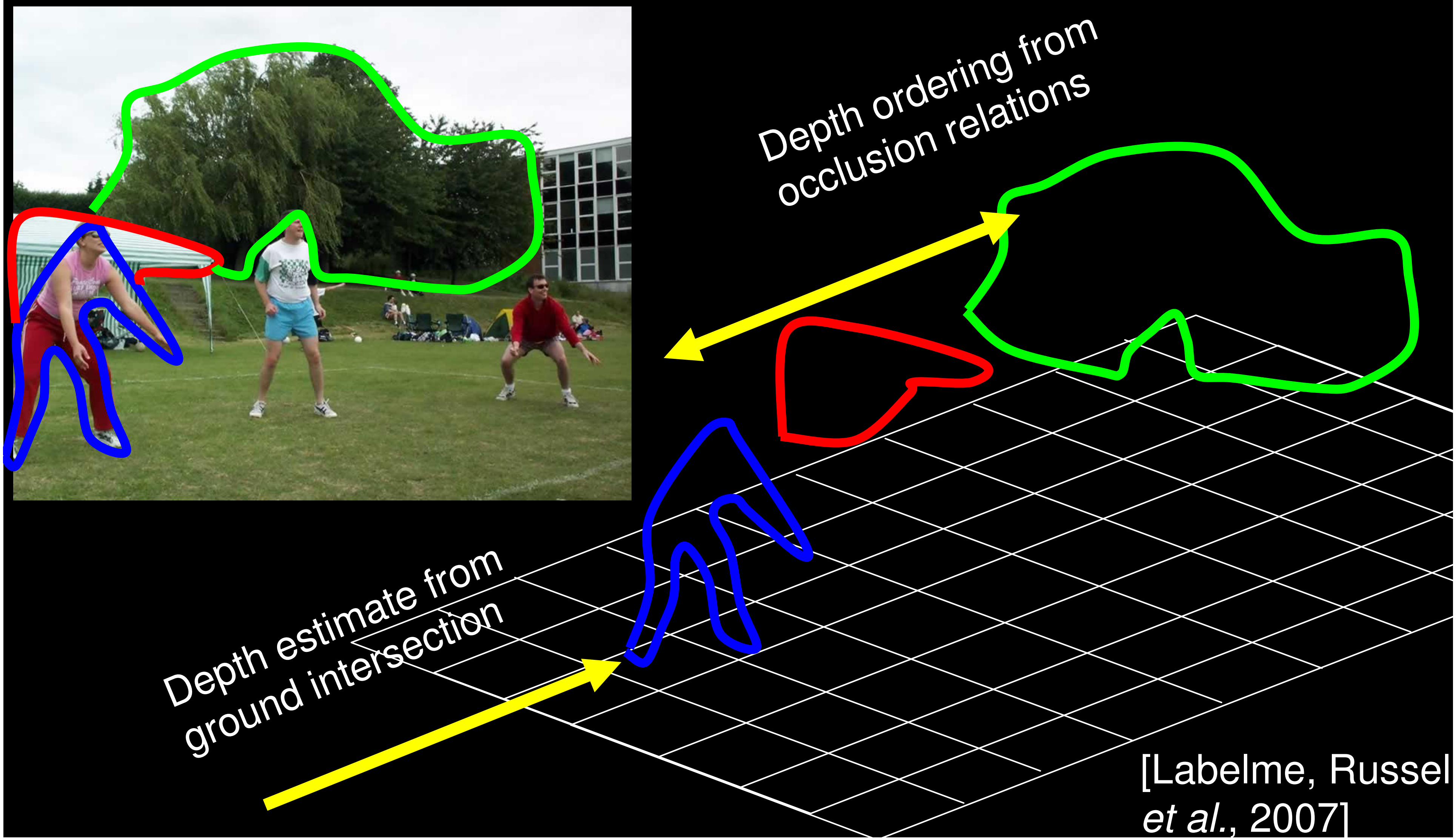


Region labels

+ Boundaries
and objects

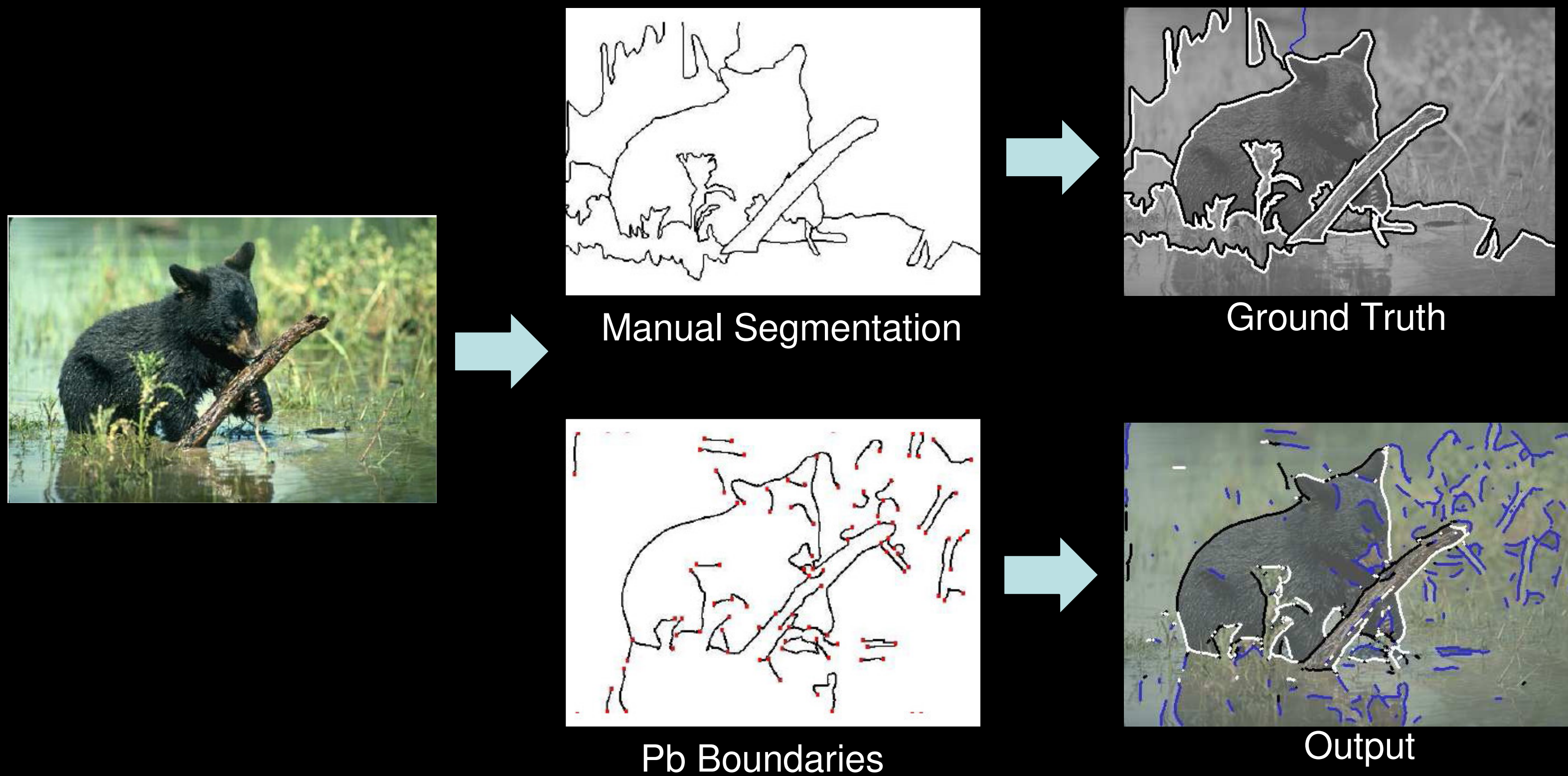
Qualitative

Using occlusion cues: Depth ordering and depth estimation



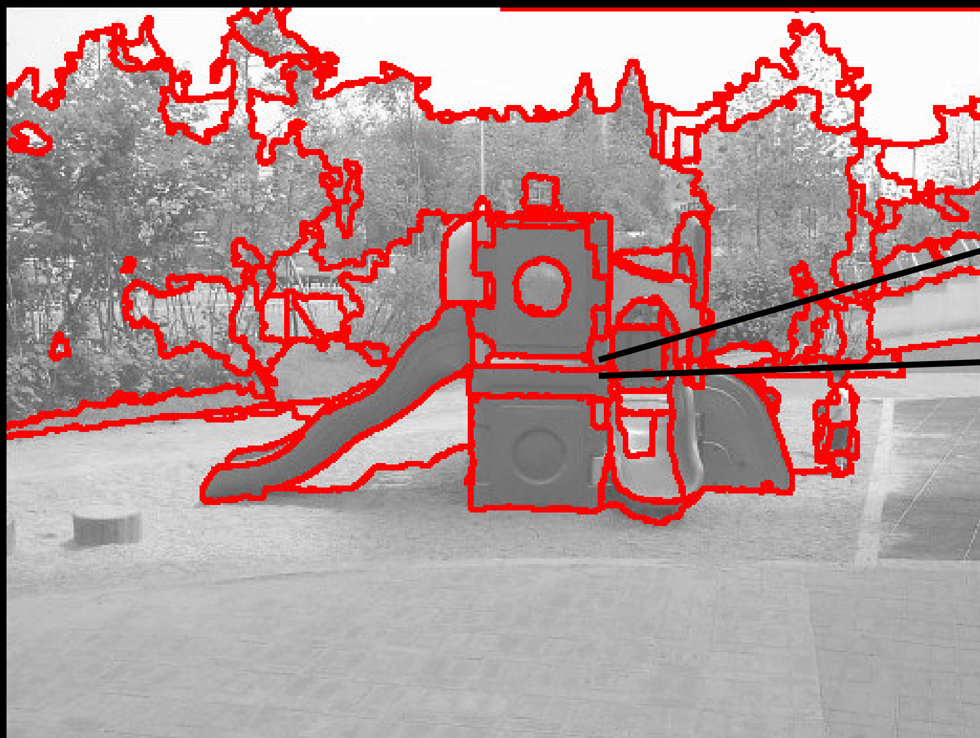
[Labelme, Russel *et al.*, 2007]

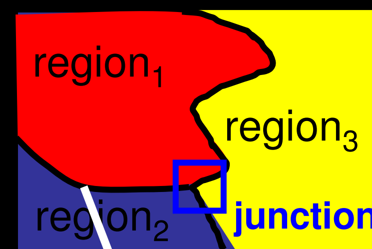
Occlusion cues from images



- M. Maire, P. Arbelaez, C. Fowlkes, and J. Malik. Using Contours to Detect and Localize Junctions in Natural Images. CVPR 2008.
- C. Fowlkes, D. Martin, and J. Malik. Local Figure/Ground Cues are Valid for Natural Images. Journal of Vision 2007.
- D. Martin, C. Fowlkes, and J. Malik. Learning to Detect Natural Image Boundaries Using Brightness and Texture. NIPS 2002.

Occlusion detection as a classification task



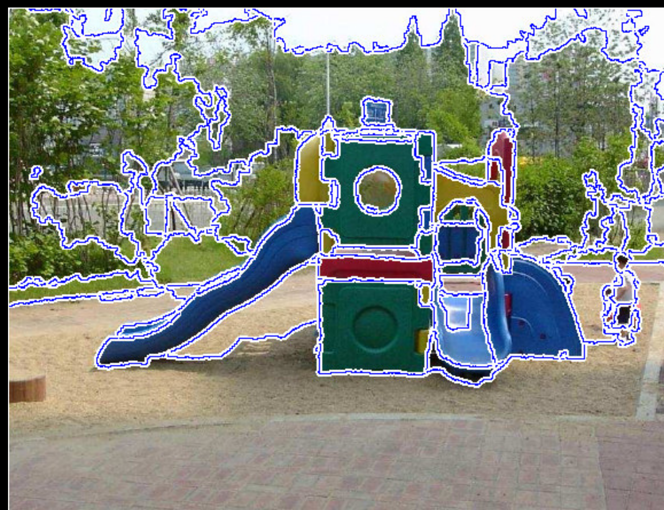


non-occlusion
region$_1$ occludes
region$_2$ occludes

[D. Hoiem, A. N. Stein, A. A. Efros, and M. Hebert. Recovering occlusion boundaries from an image. In ICCV, 2007]

Cues from images

Region Cues



Contour Cues

Surface Cues

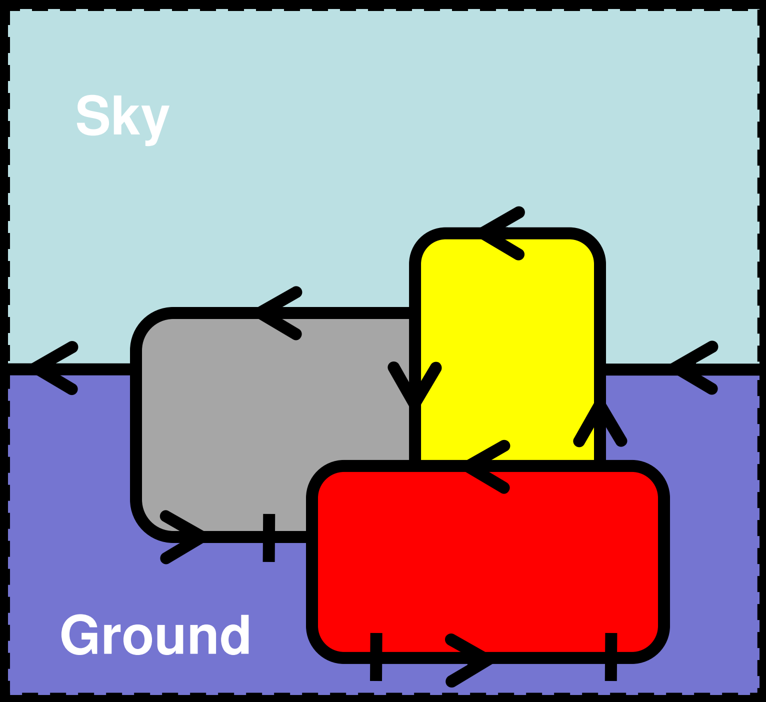
Support

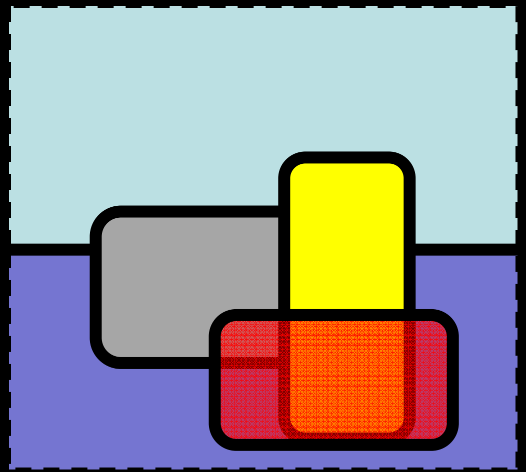
Porous

Vertical

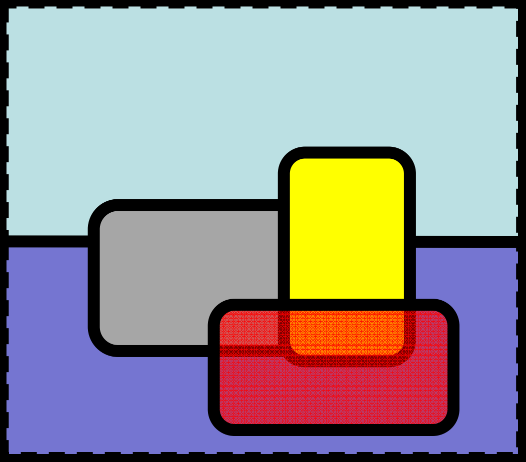
Sky

Derived cues: Depth ordering





Minimum Depth Interpretation

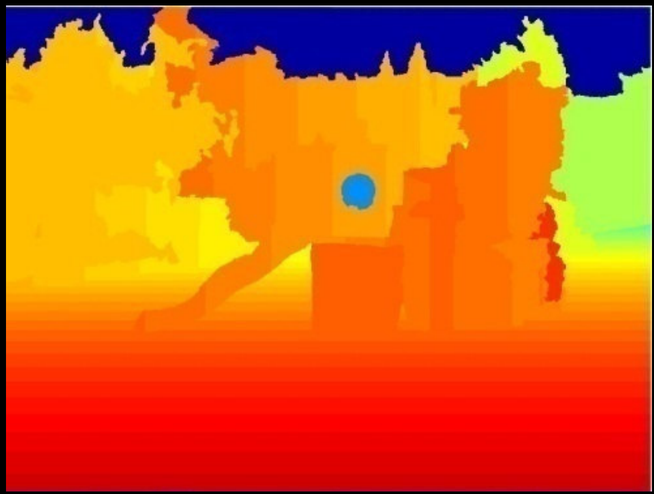


Maximum Depth Interpretation

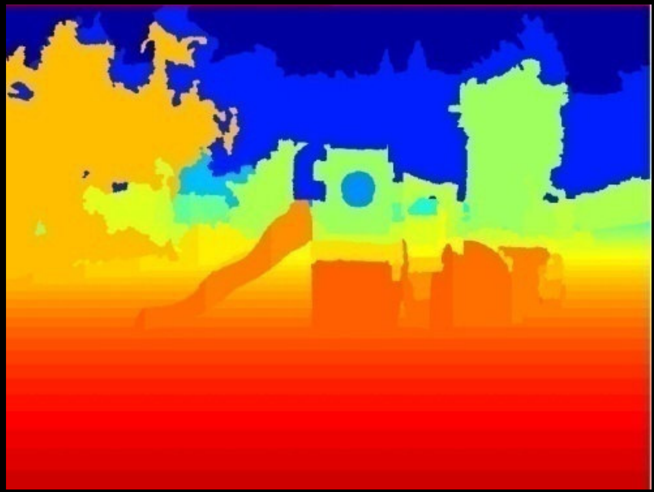
Derived cues: Depth ordering

Current Boundary Estimate

+
ground/sky labels
figure/ground labels
ground contact points



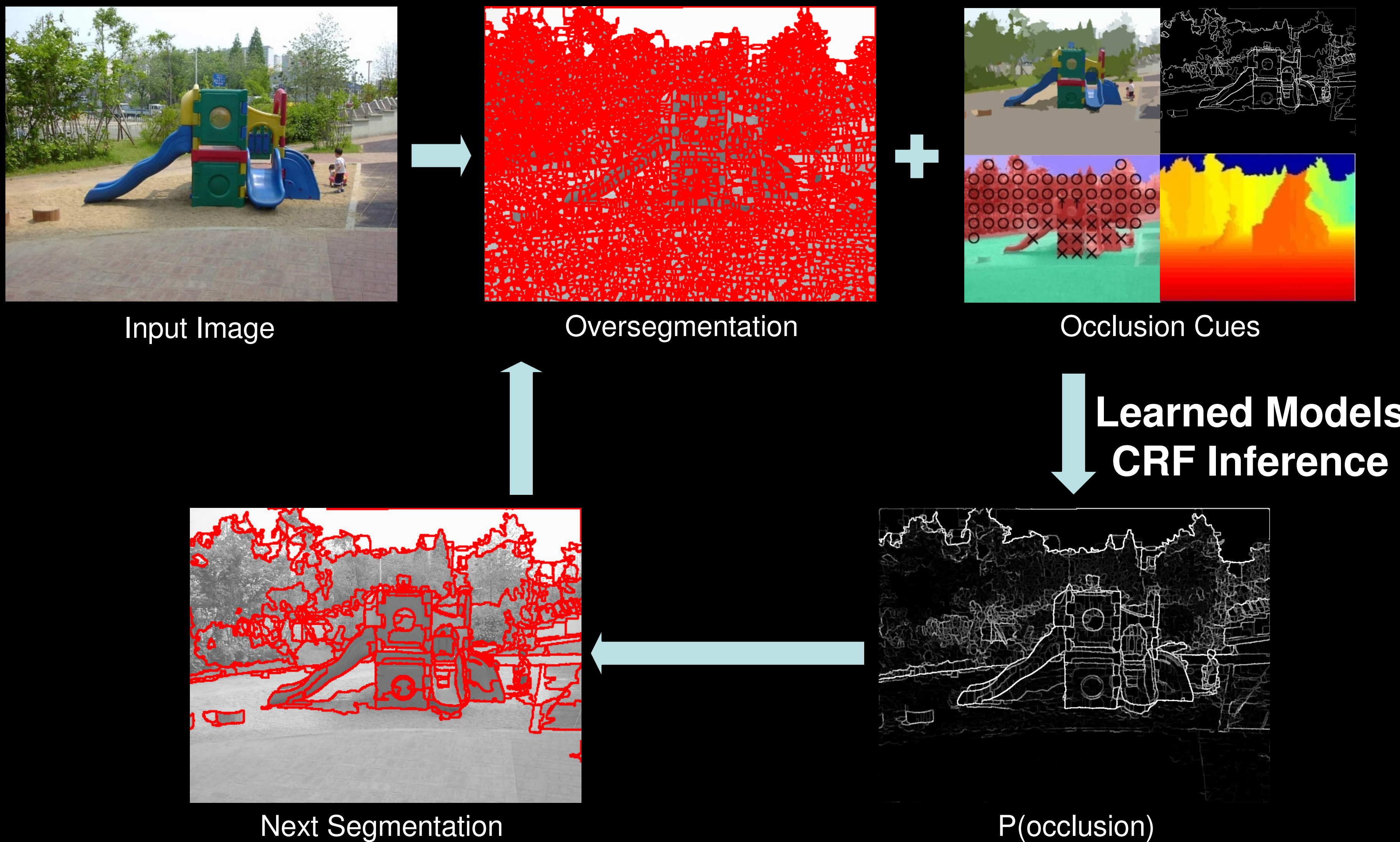
Minimum Depth Interpretation



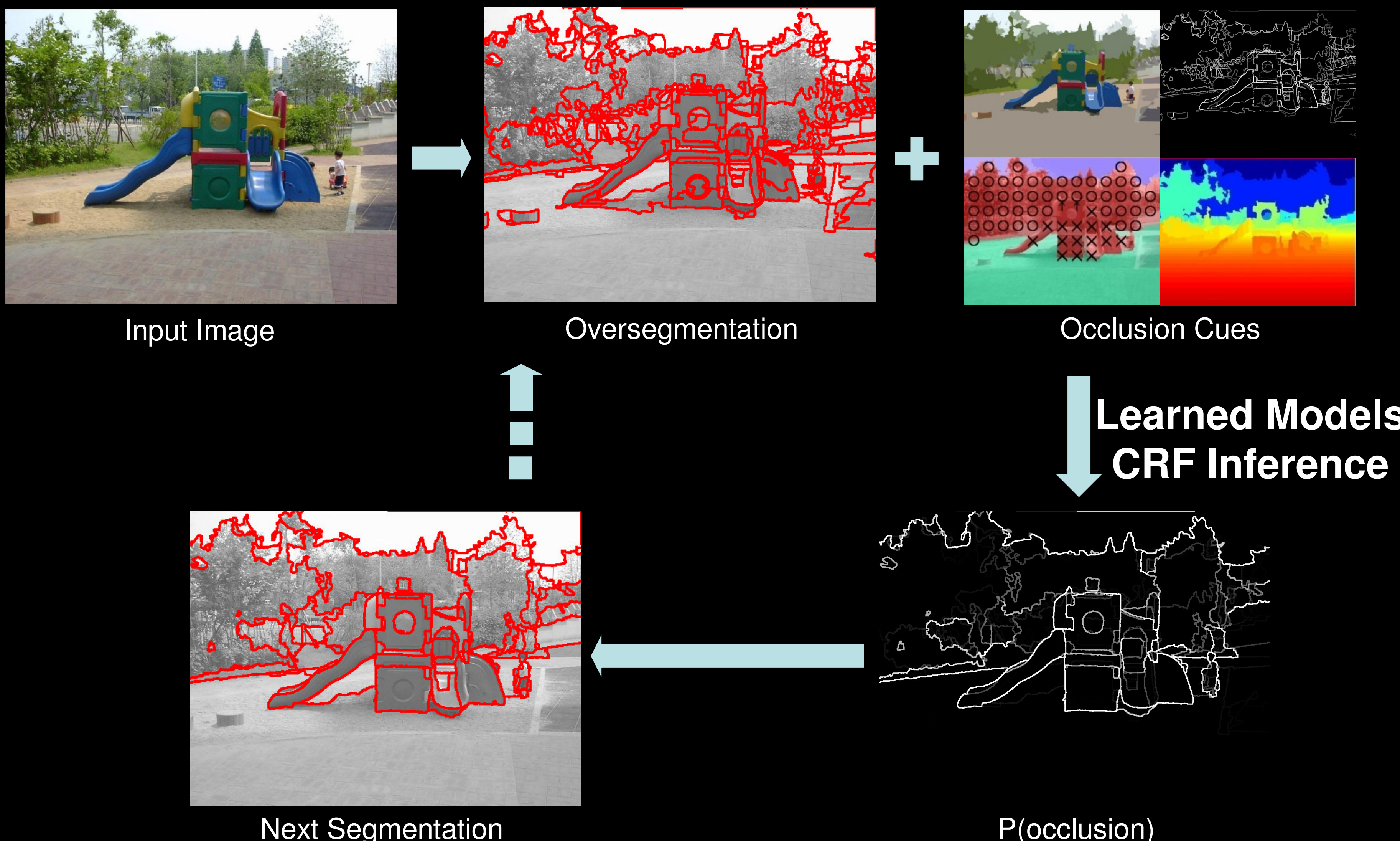
Maximum Depth Interpretation

Ground contact estimation: Lalonde et al. 2007

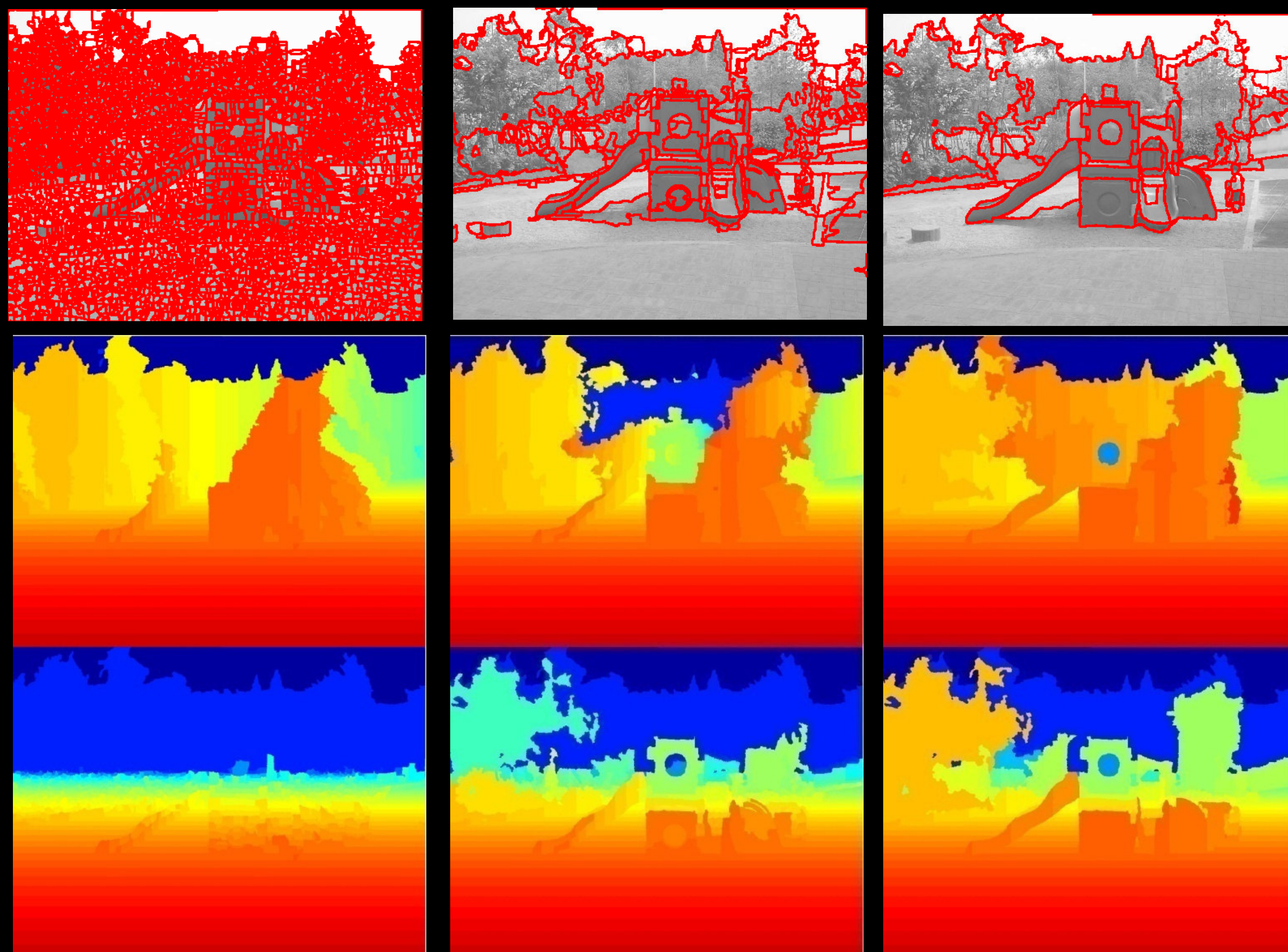
Gradual Inference of Scene Structure



Gradual Inference of Scene Structure

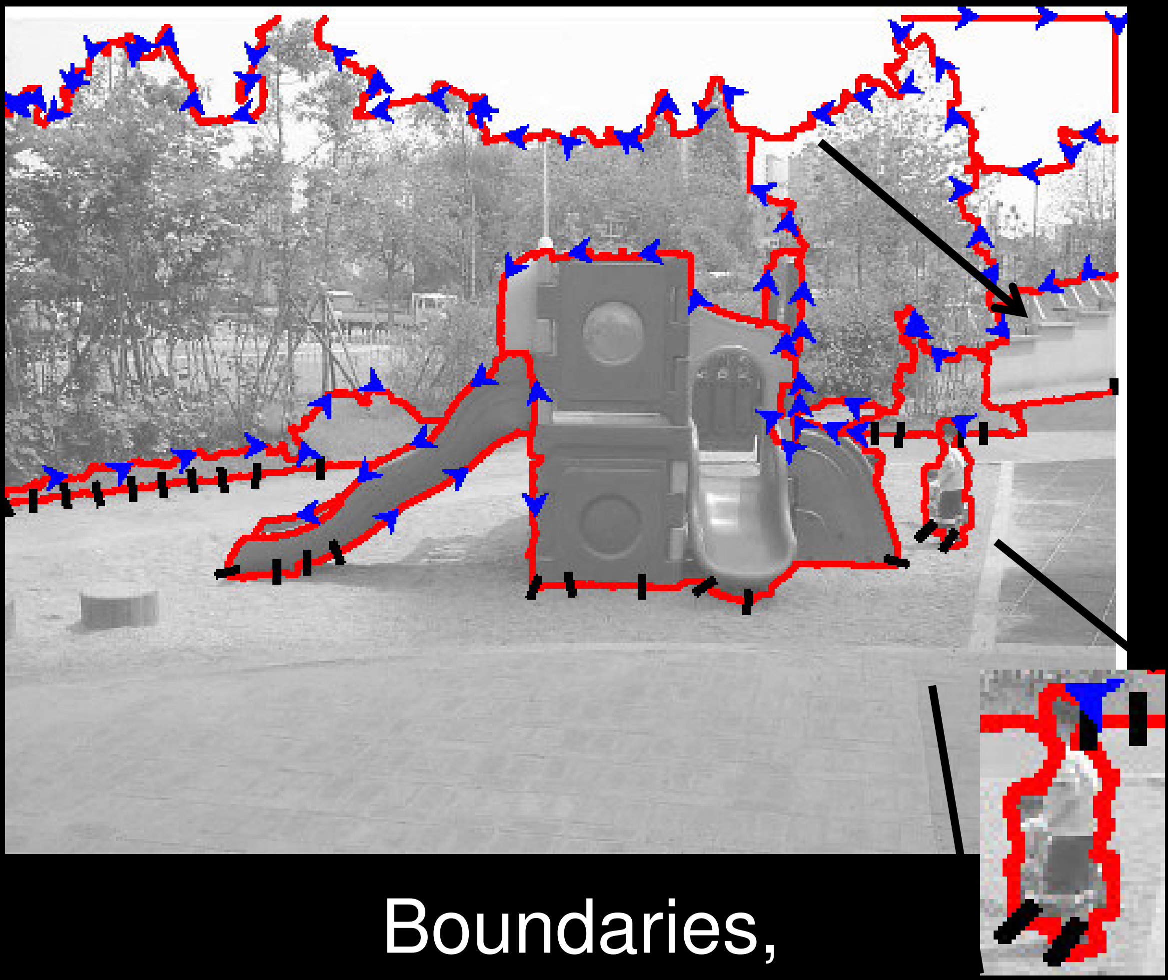


Example

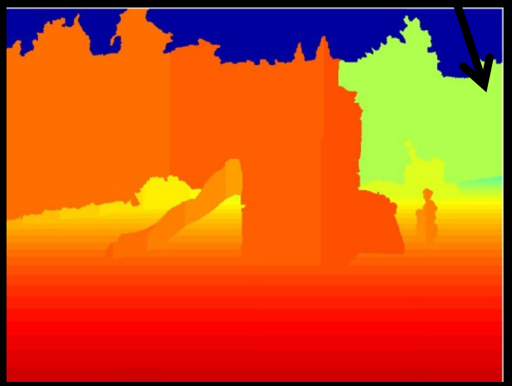


1st Iteration   2nd Iteration   3rd Iteration

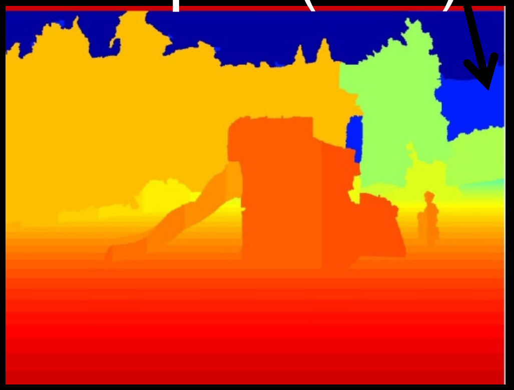
Final Estimate



Boundaries,
Foreground/Background, Contact

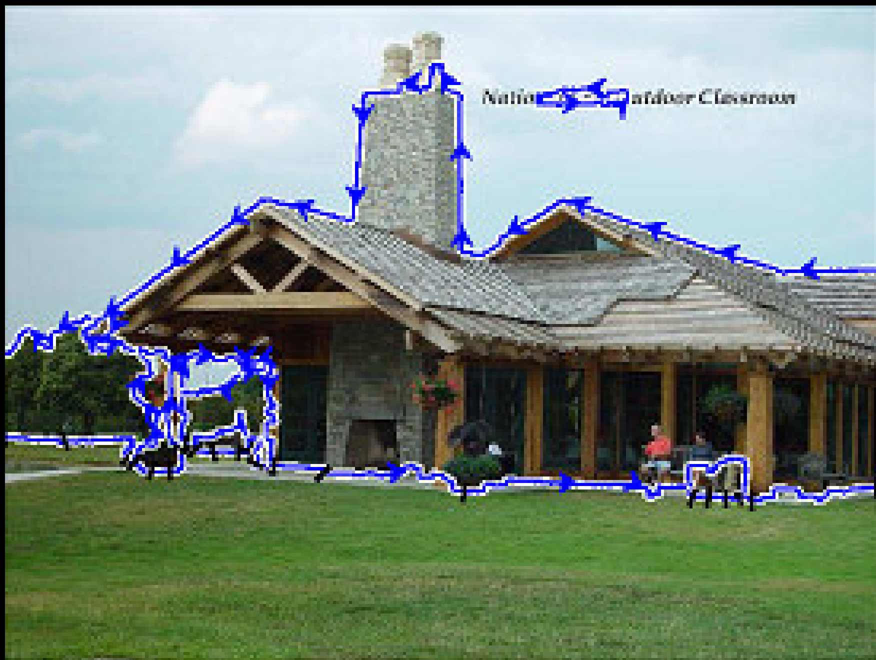


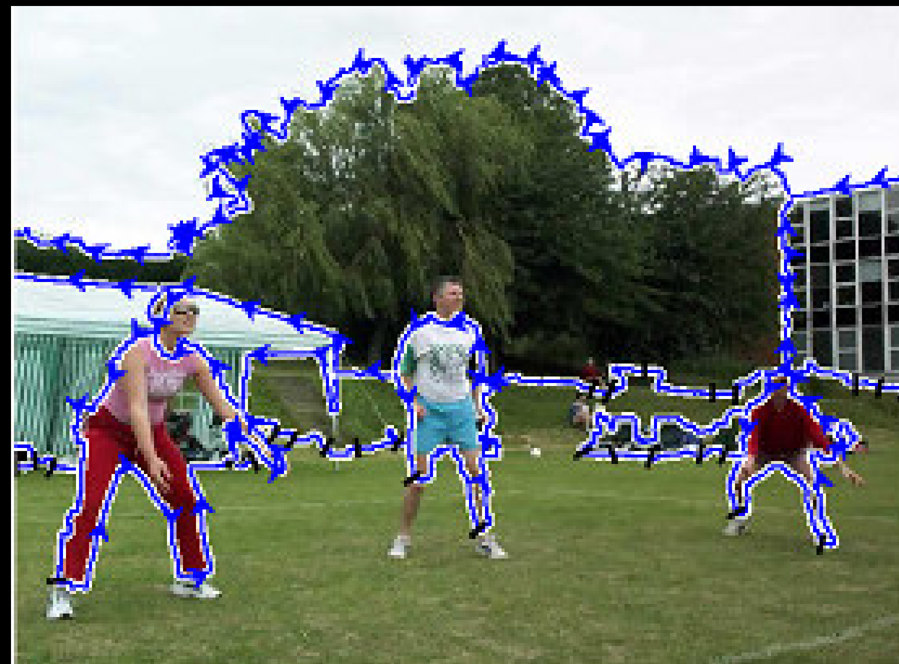
Depth (Min)

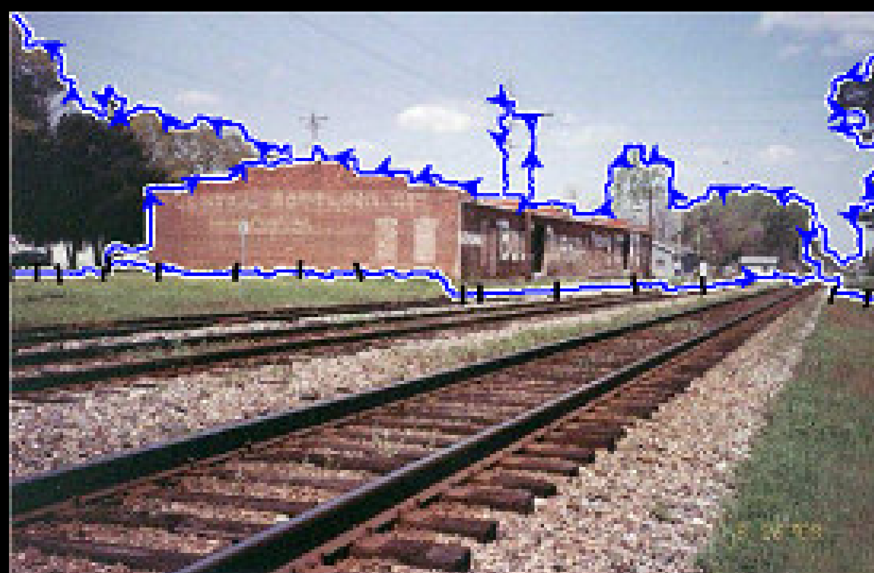


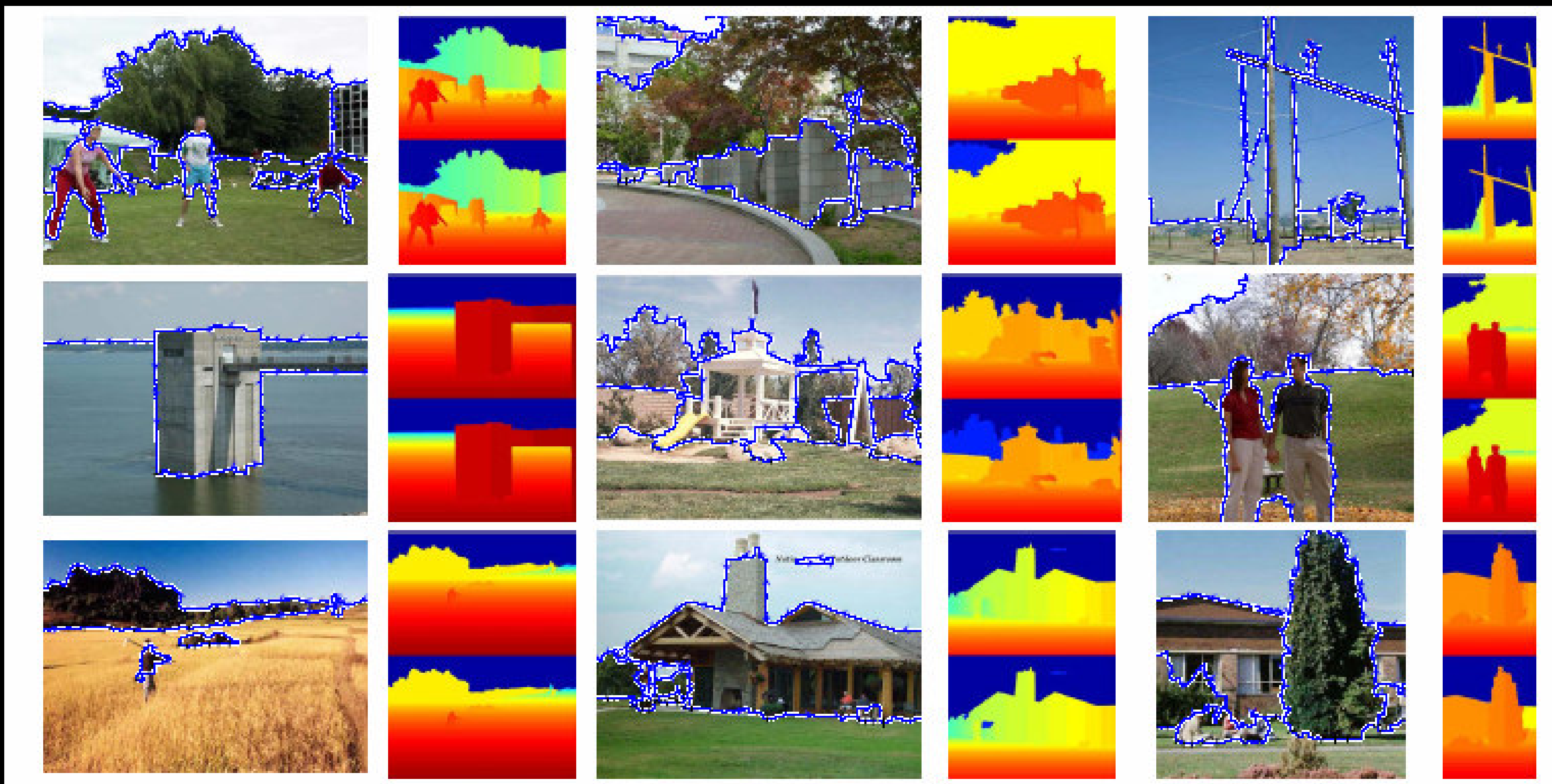
Depth (Max)

Examples









Are 3D cues useful?

Fancy CRF models help a little but not much

	Edge/Region Cues	+ 3D Cues	With CRF
Iter 1	58.7%	71.7%	Not Used
Iter 2	65.4%	75.6%	77.3%
Final	68.2%	77.1%	**79.9%**

"Reasoning" through iterative reasoning is necessary: Straight classification can't do it

3D cues necessary to boost performance

Comments

- Qualitative representation of 3D (occlusion relations and relative depth ordering rather than absolute shape)
- Multiple segmentations
- Iterative search through multiple hypothesis combined with local classifiers

- We’ve improved our understanding of the 3D structure of the scene
- Fine, but can we use this to help with scene interpretation as part of a larger reasoning system?

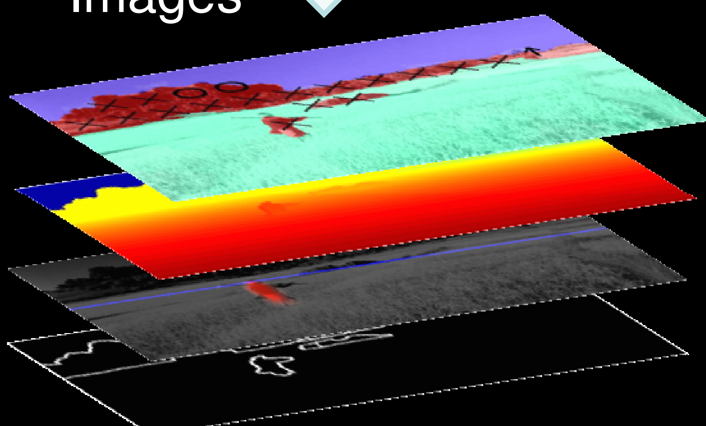
[D. Hoiem, A. A. Efros, and M. Hebert. *Closing the loop on scene interpretation*. In CVPR, 2008]

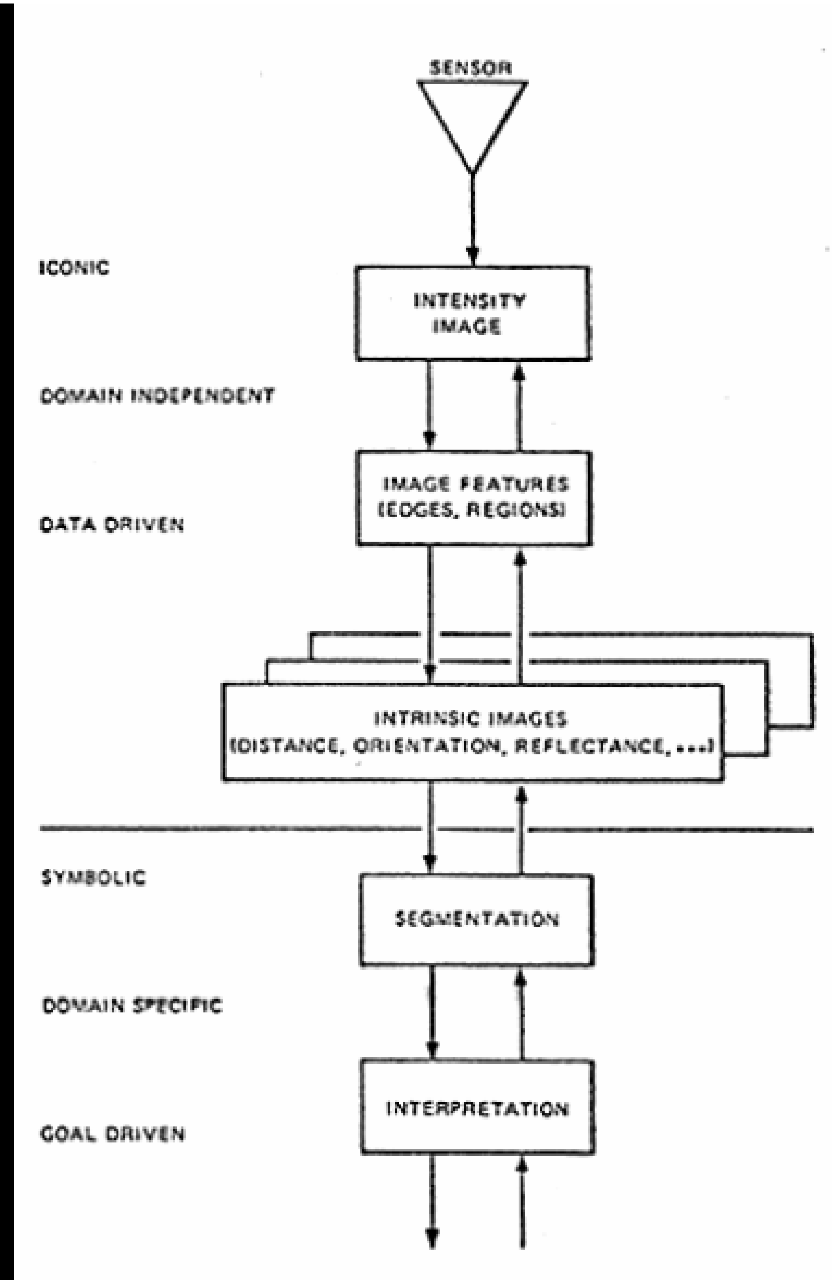
Input Image

Scene Analysis Processes

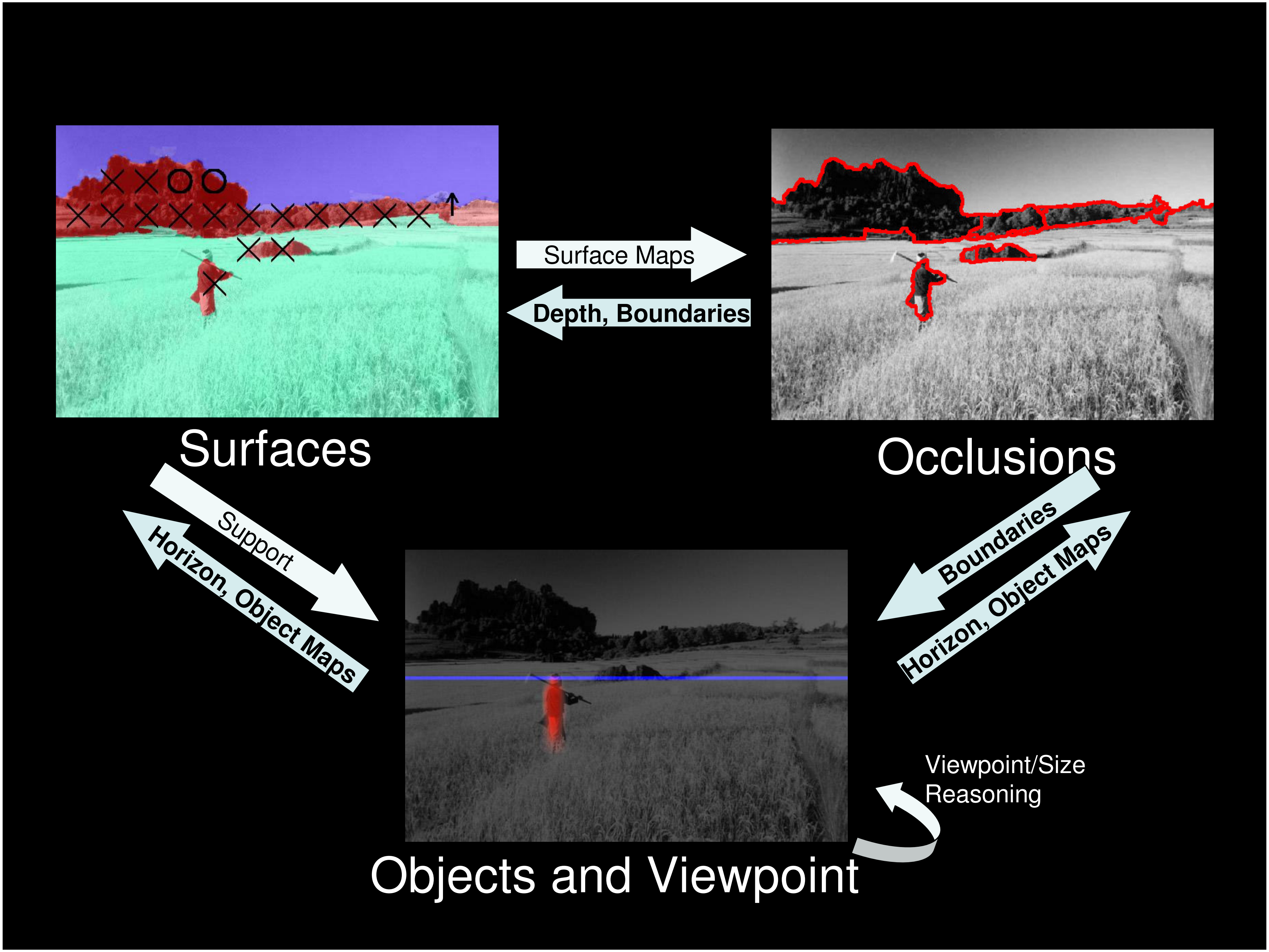
Surface Orientation
Object/Viewpoint
Occlusion/Depth

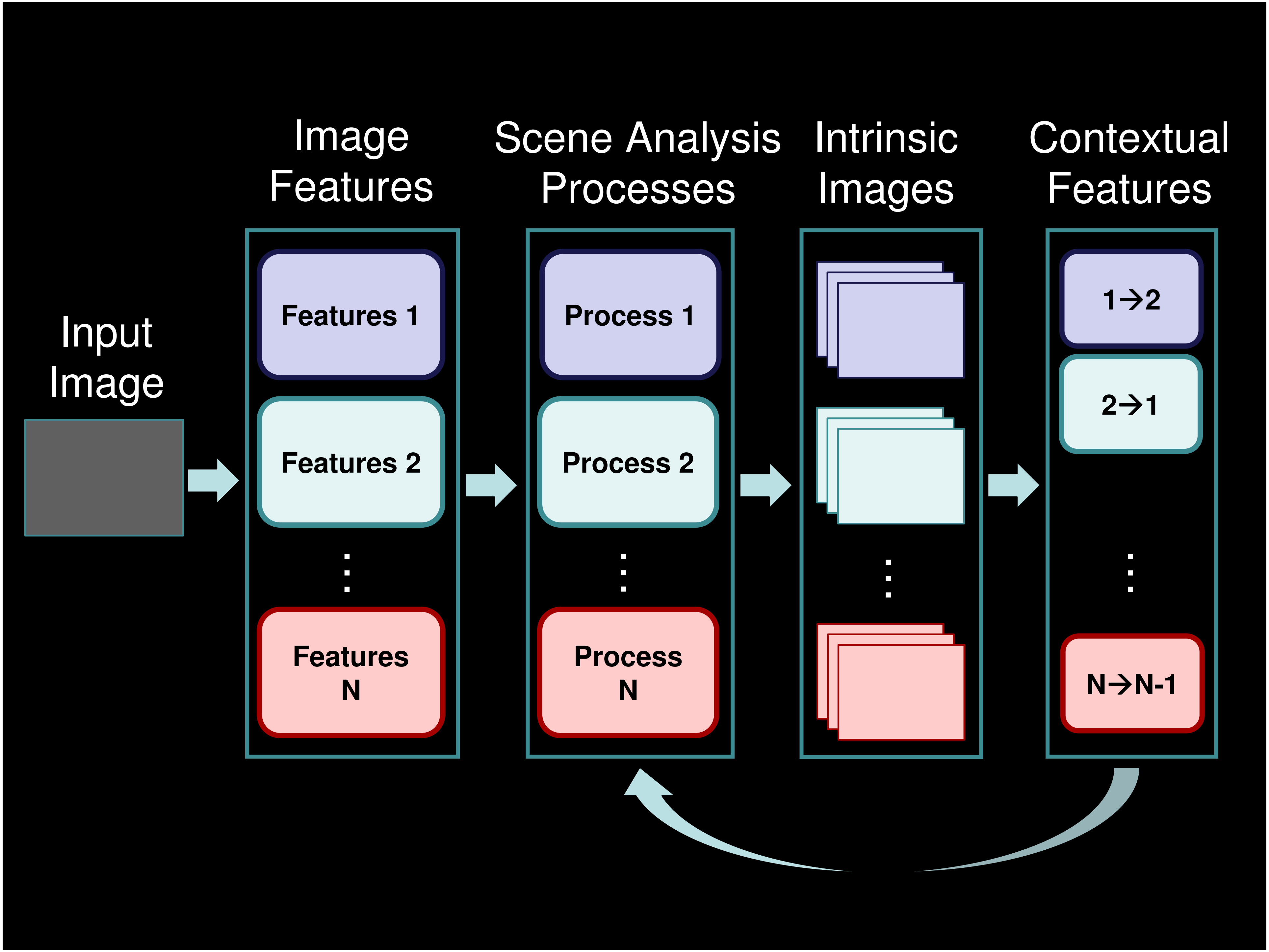
Intrinsic Images

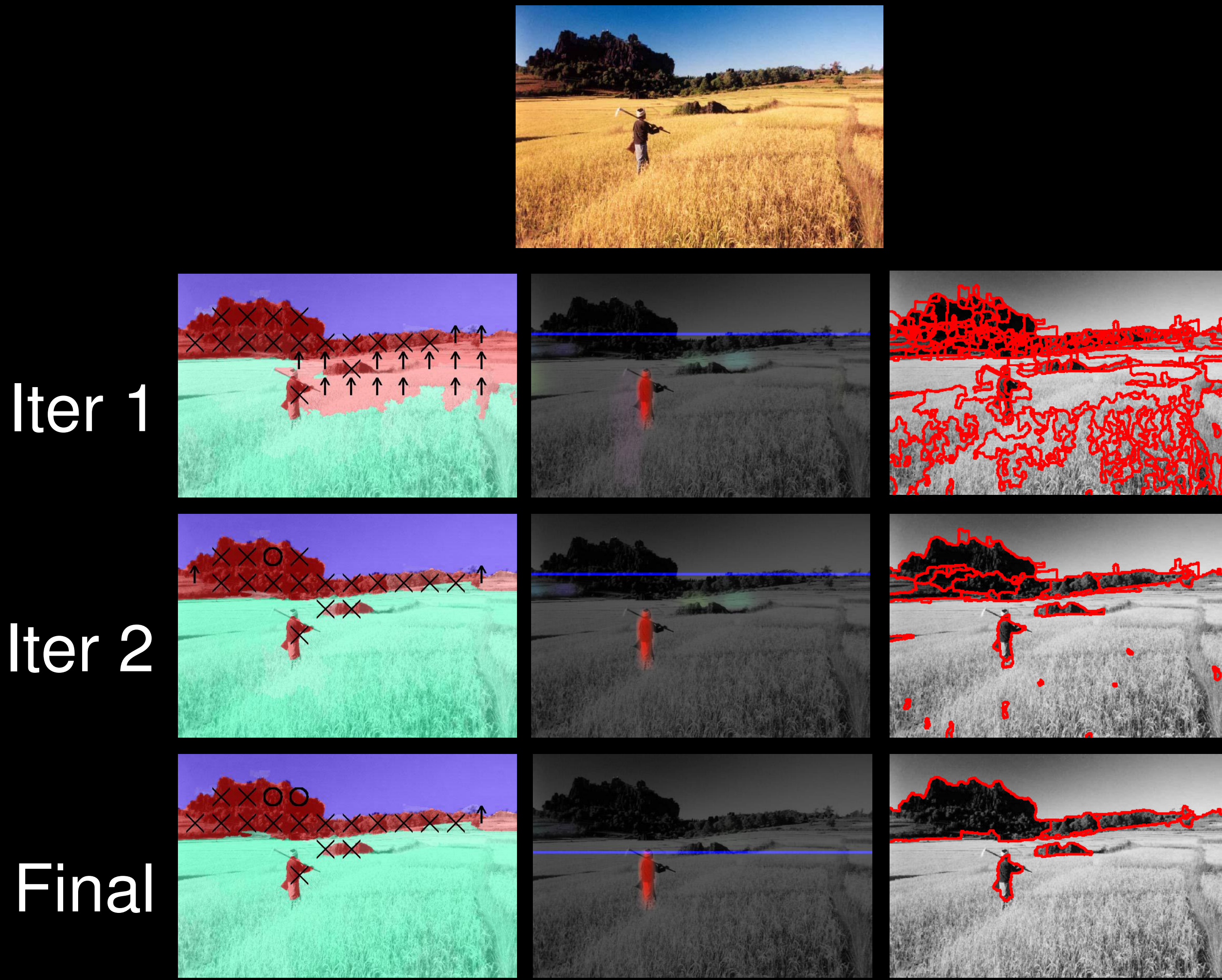




[Barrow and Tenenbaum 1978]


Surface Maps
Depth, Boundaries
Surfaces
Occlusions
Support
Horizon, Object Maps
Boundaries
Horizon, Object Maps
Viewpoint/Size
Reasoning
Objects and Viewpoint

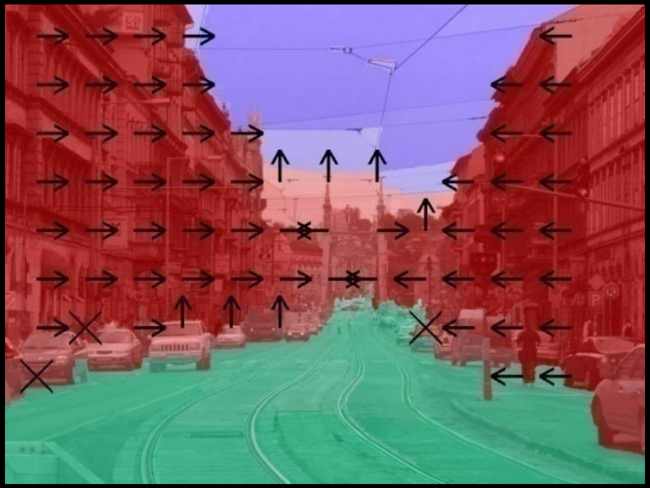

Input Image
Image Features
Features 1
Features 2
Features N
Scene Analysis Processes
Process 1
Process 2
Process N
Intrinsic Images
Contextual Features
1→2
2→1
N→N-1



- Small improvement of surface labels
- 7-10% improvement of object detections

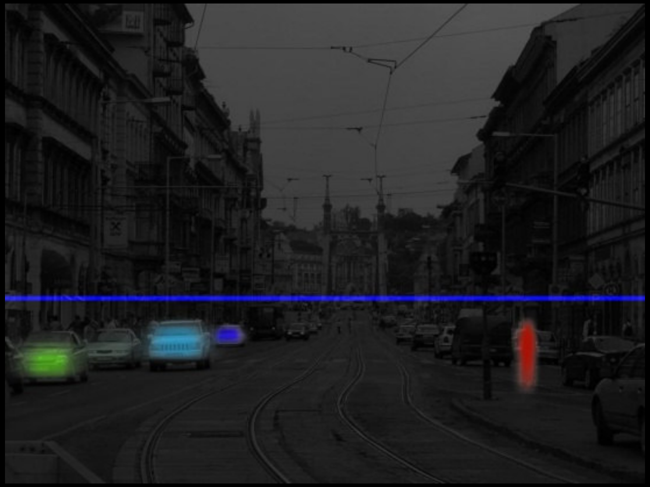
Separate cues

Input



Surfaces

Occlusion Boundaries



Objects/Horizon

Combined reasoning

Input

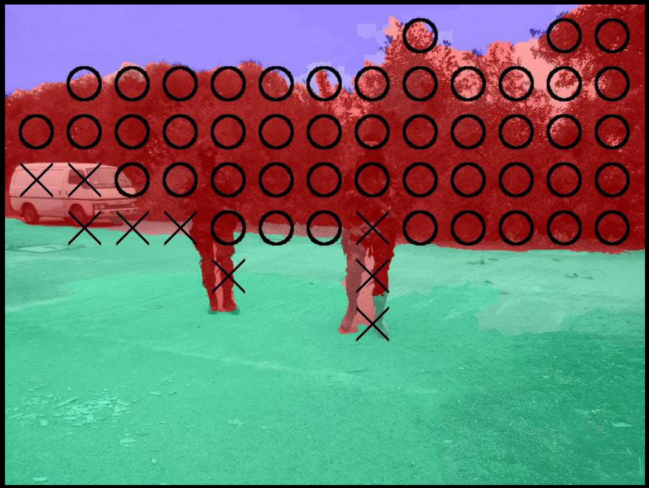
Surfaces

Occlusion Boundaries

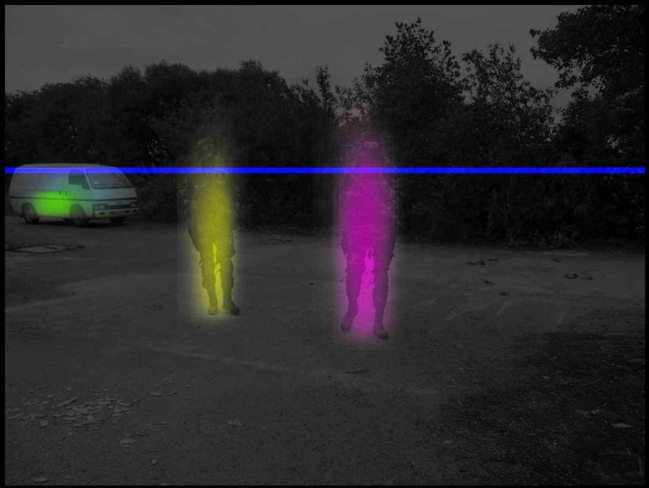
Objects and Horizon

Separate cues


Input

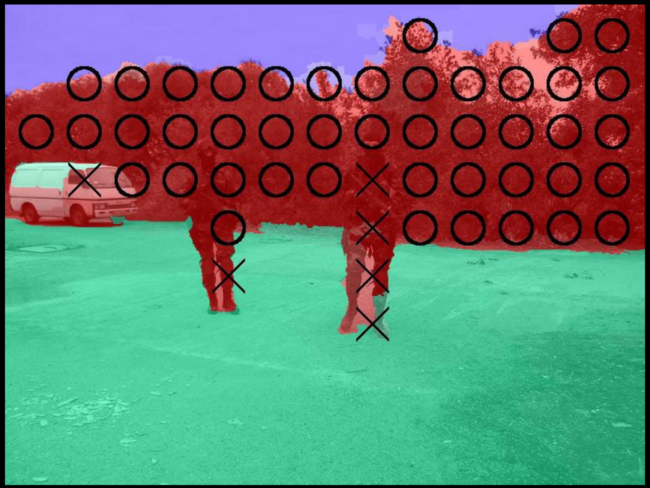

Surfaces


Occlusion Boundaries


Objects/Horizon

Combined reasoning

Input



Surfaces

Occlusion Boundaries

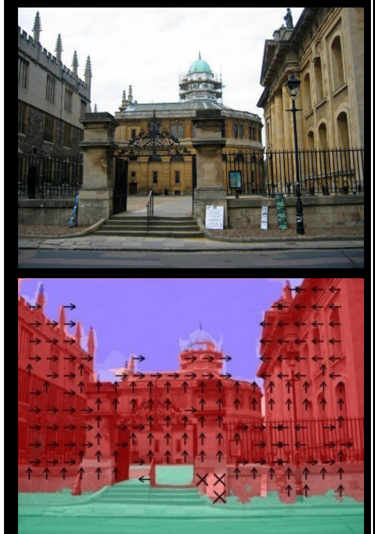
Objects and Horizon

CVPR'08

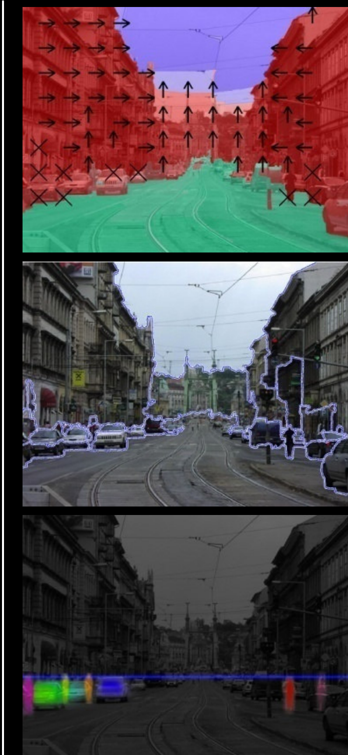
Comments

- Plus:
  - Scene geometry (surface geometry and object relations) estimated from image data
  - Scene geometry used explicit in scene understanding
- Minus:
  - Still mostly bottom-up classification approach
  - No use of domain constraints or known laws governing the physical world

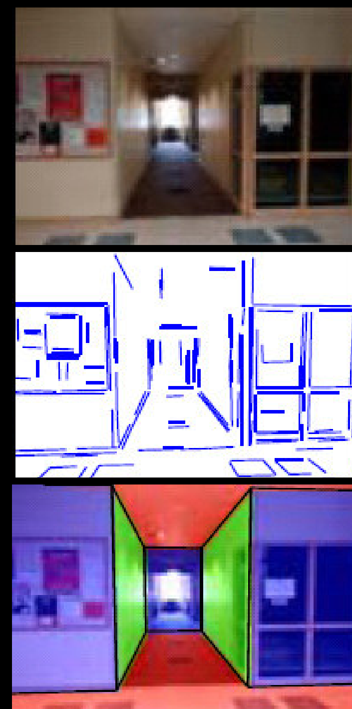
Levels of 3D-ness



Region labels



+ Boundaries and objects



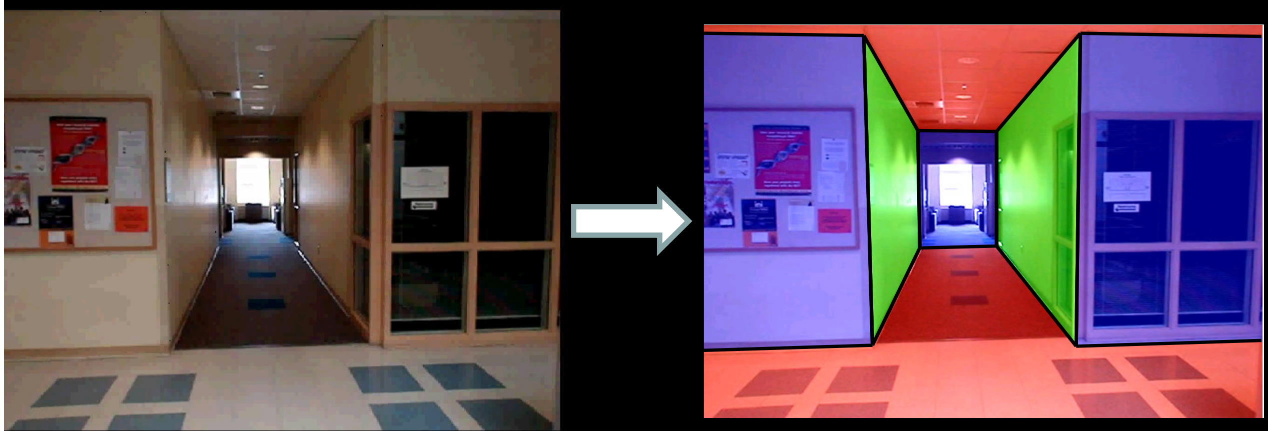
Stronger geometric constraints from domain knowledge

Qualitative

More quantitative
more precise

Example

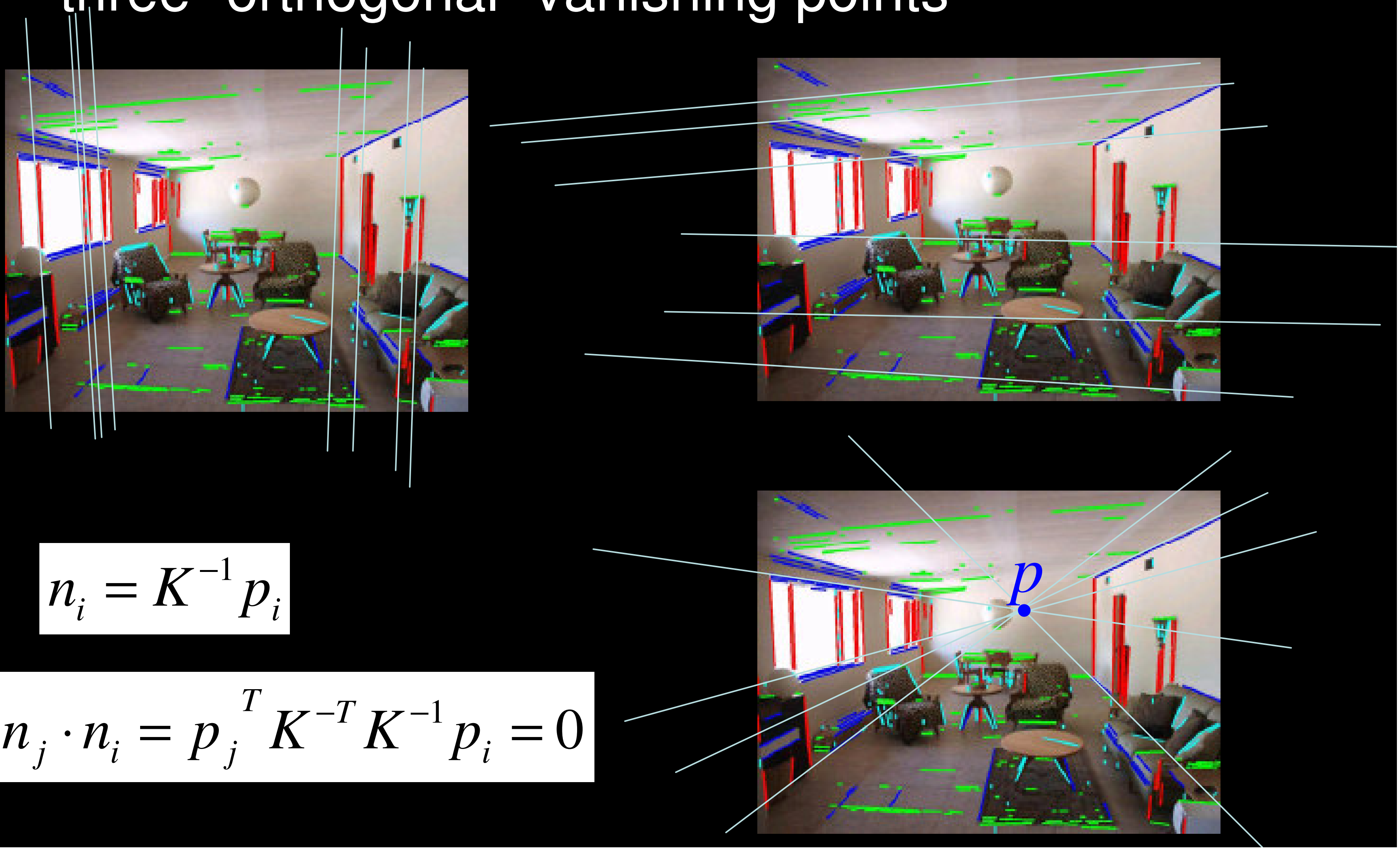
- Using constraints induced by man-made environments in interpreting images
- Examples: Manhattan world, limited vocabulary of object configurations, etc.



D. Lee, T. Kanade, M. Hebert. Geometric Reasoning for Single Image Structure Recovery. CVPR09. (+ under review, 2010)

Constraint: Manhattan world assumption

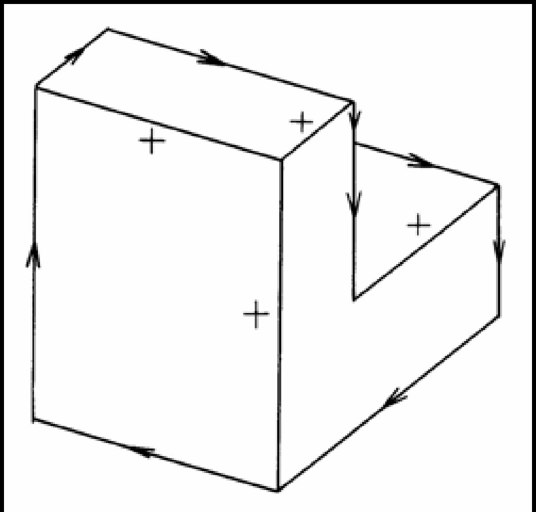
- Three dominant directions corresponding to three “orthogonal” vanishing points



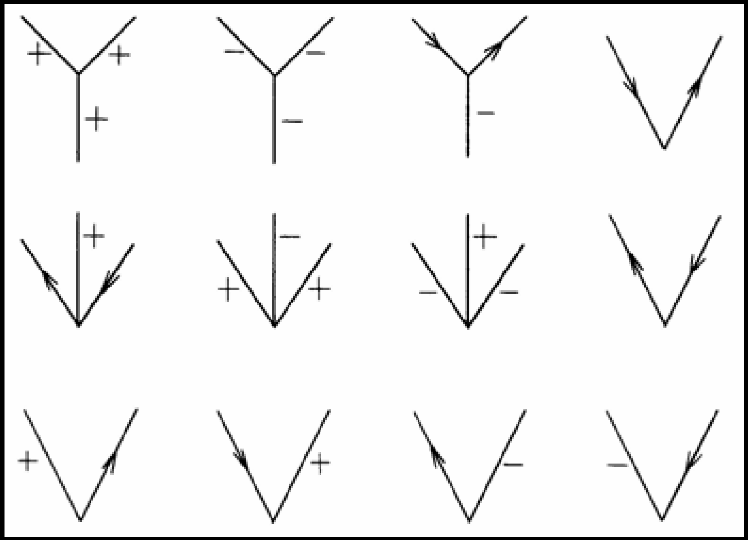
$$n_i = K^{-1} p_i$$

$$n_j \cdot n_i = p_j^{\;T} K^{-T} K^{-1} p_i = 0$$

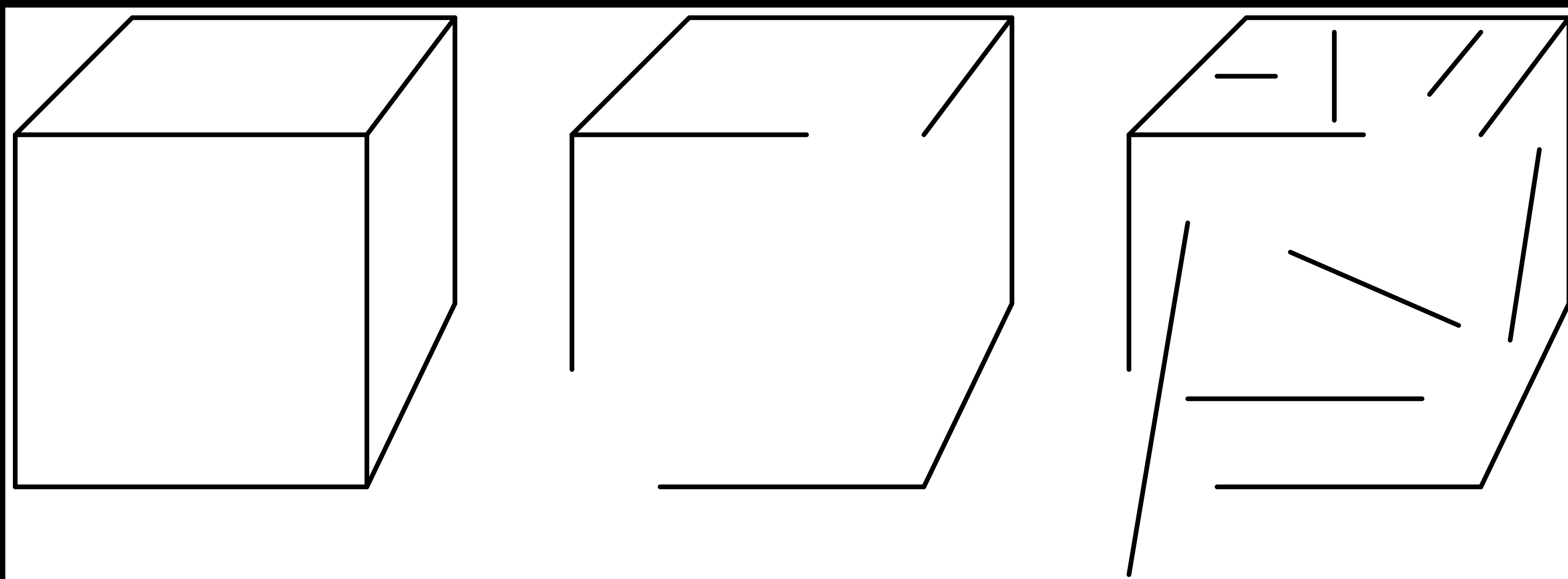
Line Drawing Interpretation (1970~)



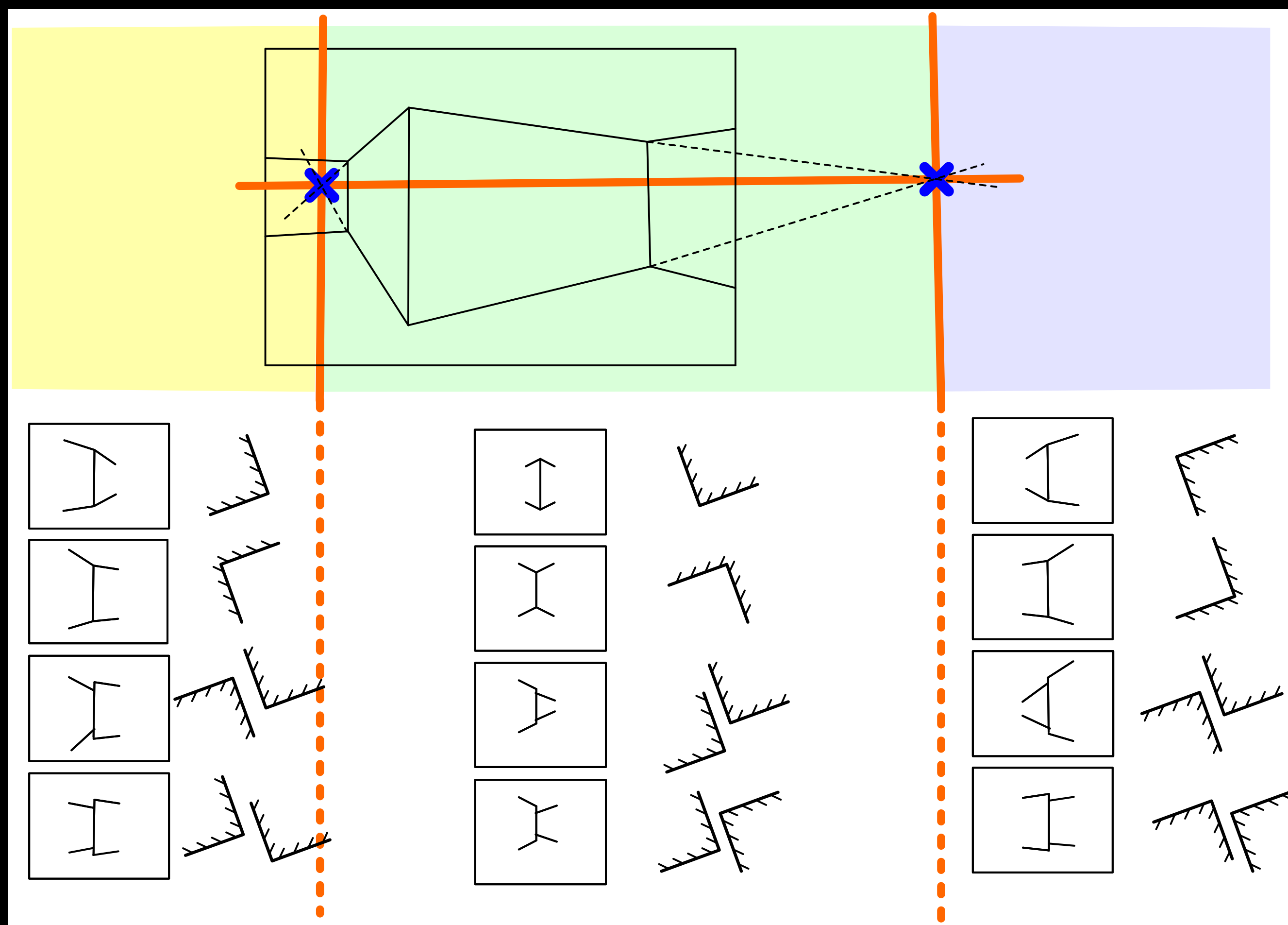
Labeled Line drawings



12 Possible Junctions



Geometric reasoning on corners

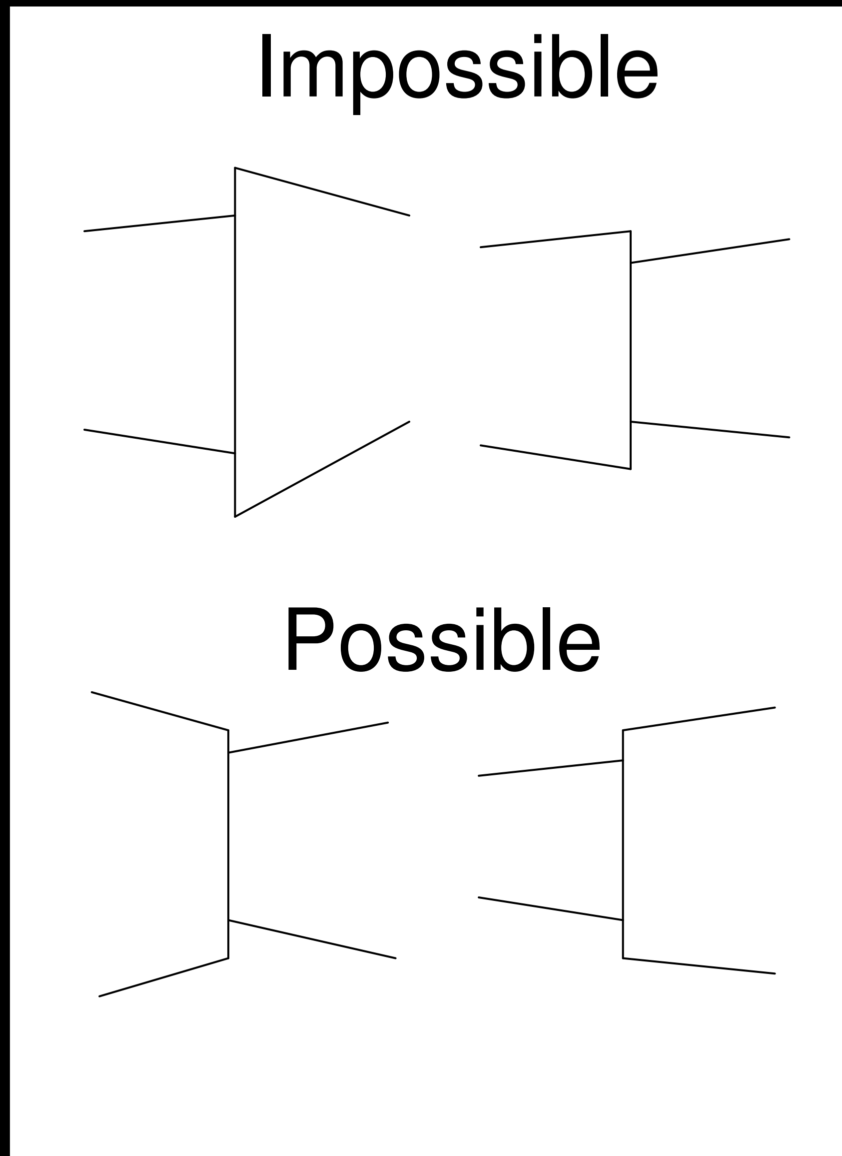


4 Possible Corners

4 Possible Corners

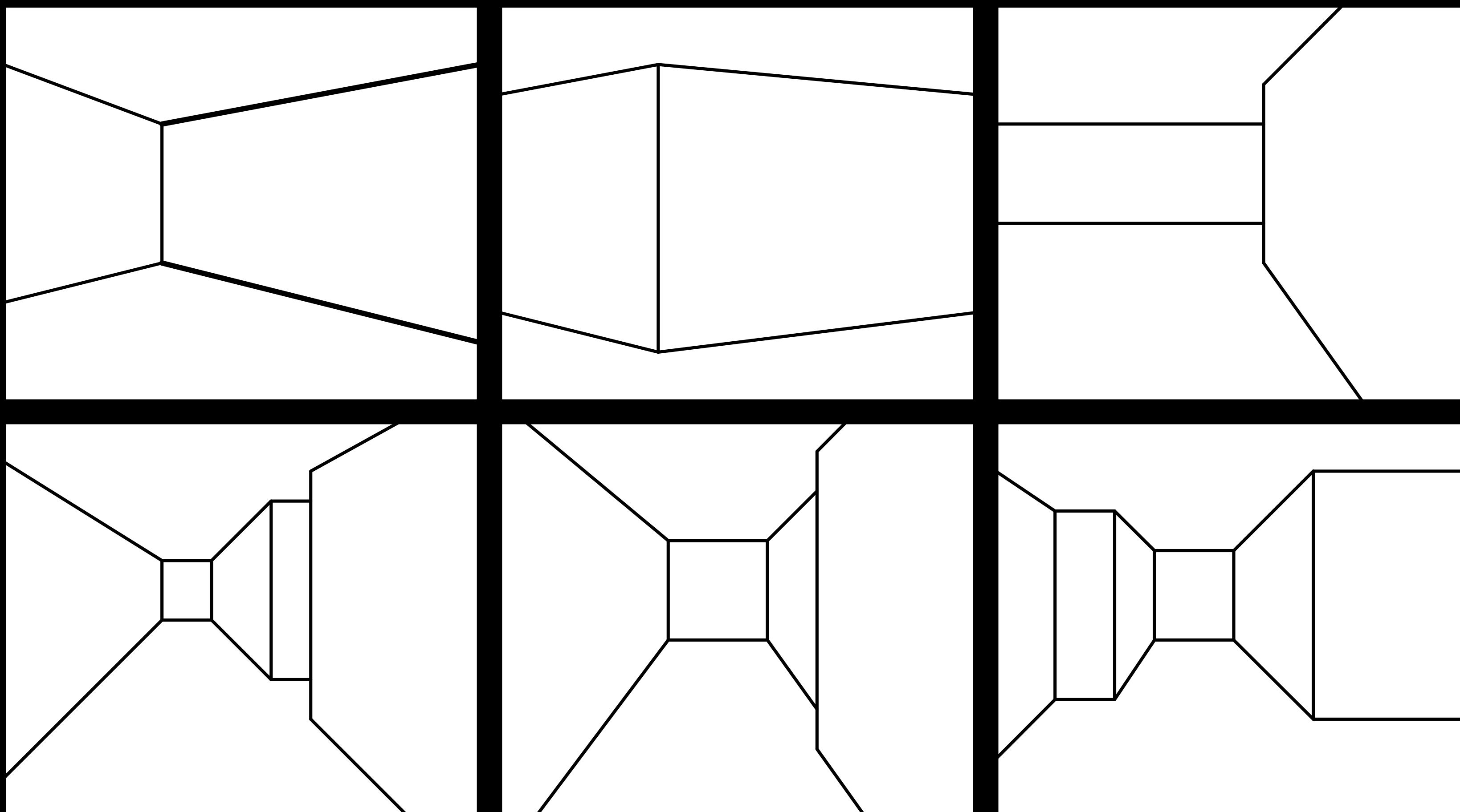
4 Possible Corners

Geometric reasoning on corners



World Model

- Manhattan World

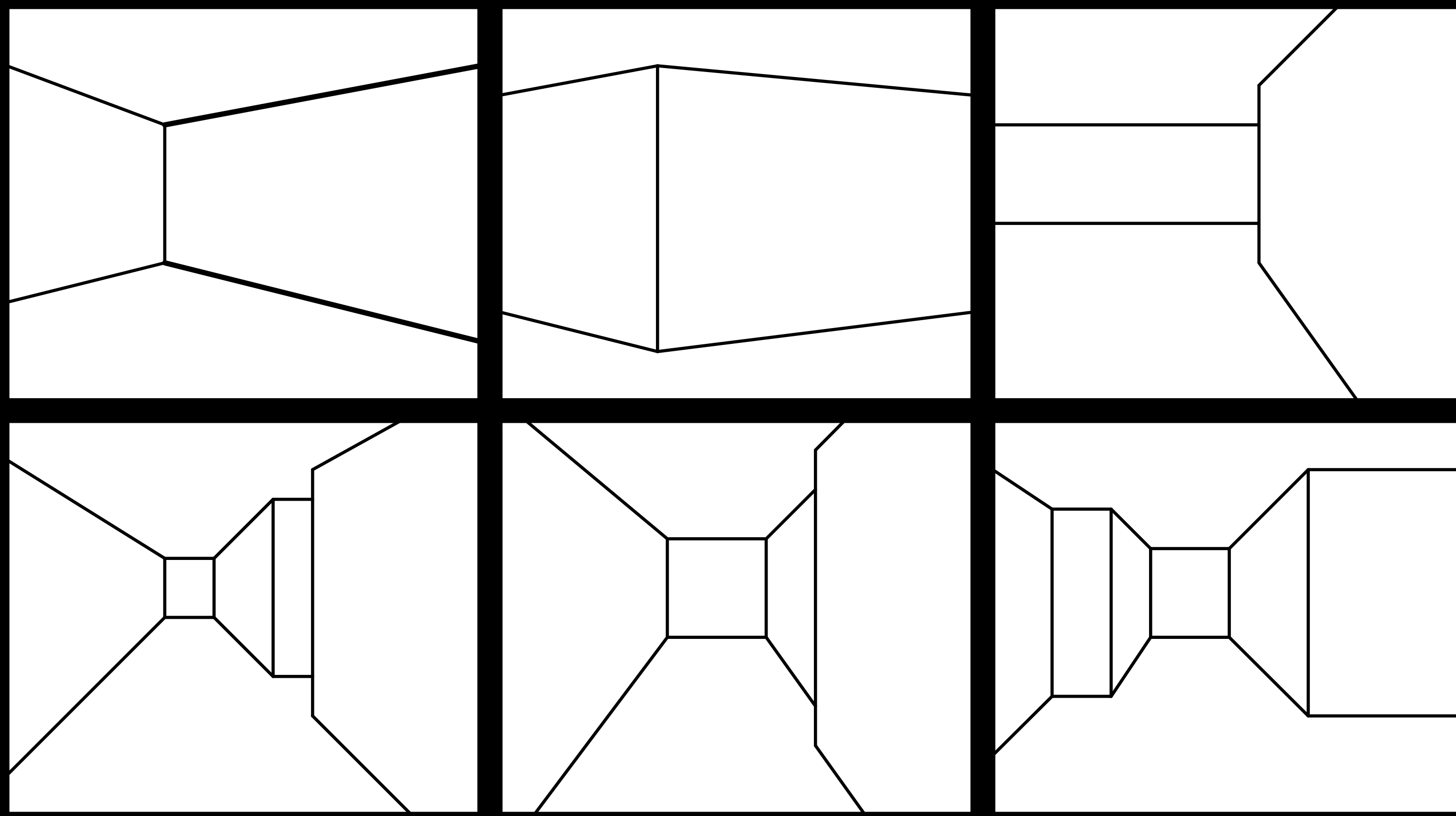


[Delage et al. CVPR'06, Kosecka et al. CVIU'05, Coughlan & Yuille Neural Computation'03]

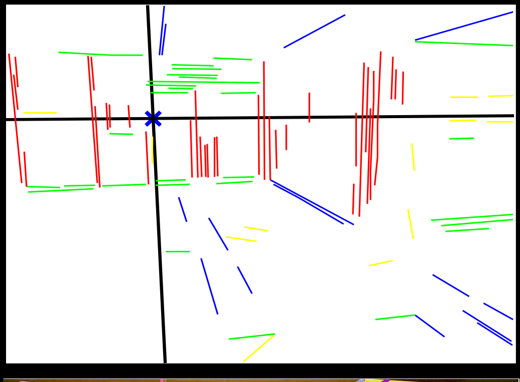
Dictionary

Concave (-) | Convex (+) | Occluding (>)

Combination

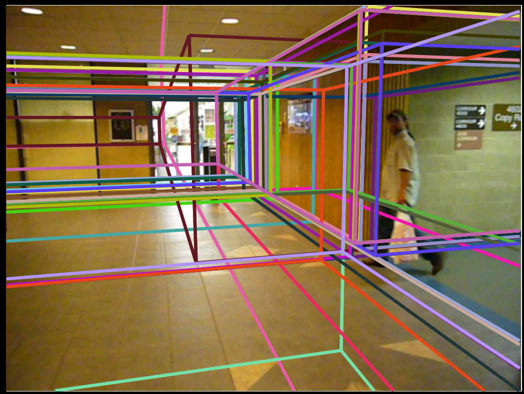


Recovering Structure

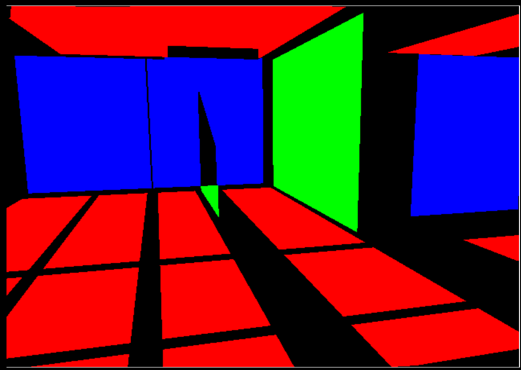


1. Detect line segments
2. Estimate vanishing points

J. Coughlan and A. Yuille. Manhattan world: Compass direction from a single image by bayesian inference. In Proceedings ICCV, 1999.
J. Kosecka andW. Zhang. Video compass. In Proceedings of European Conference on Computer Vision, pages 657 – 673, 2002.



3. Generate scene hypotheses

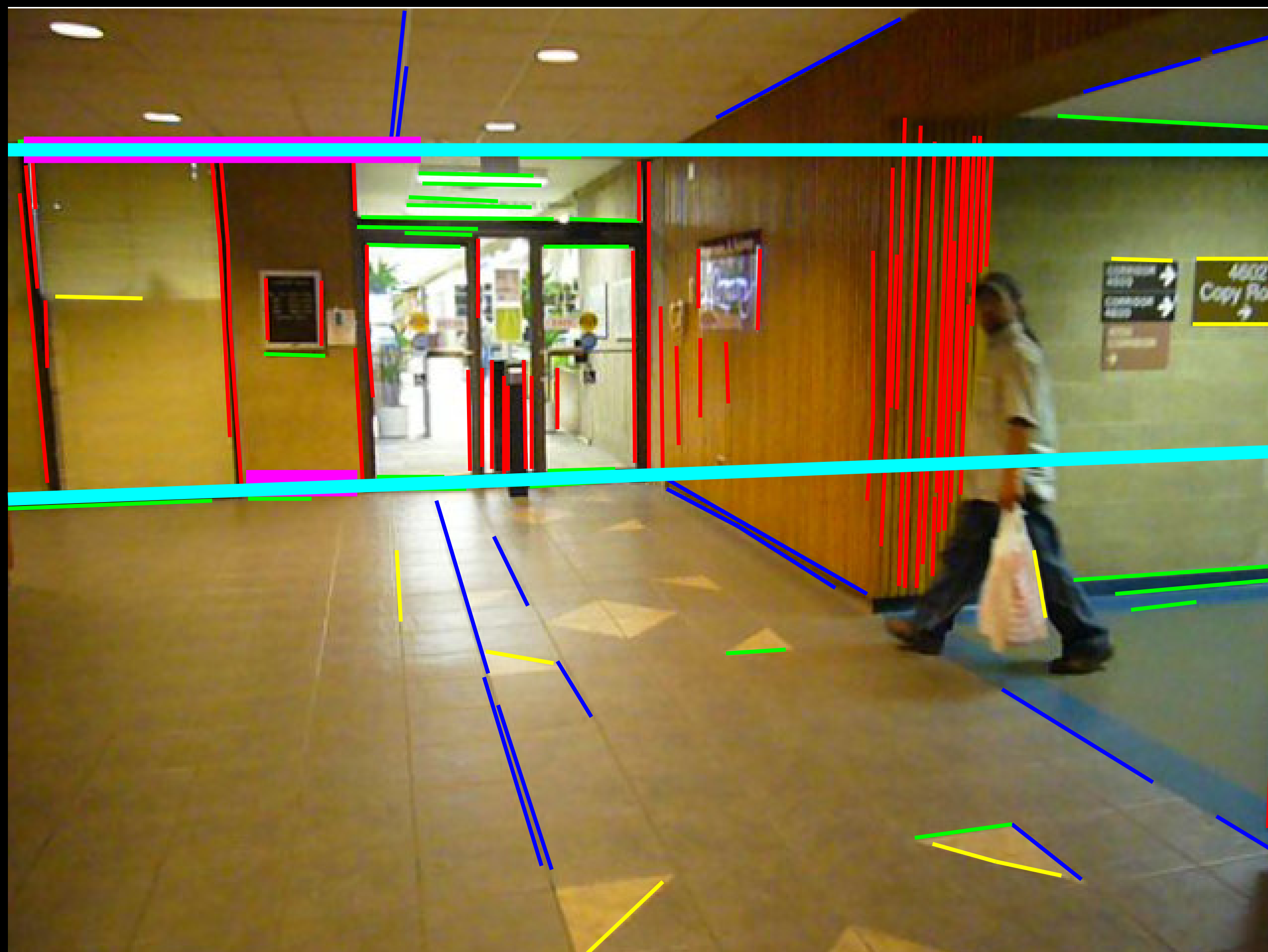


4. Evaluate scene hypotheses

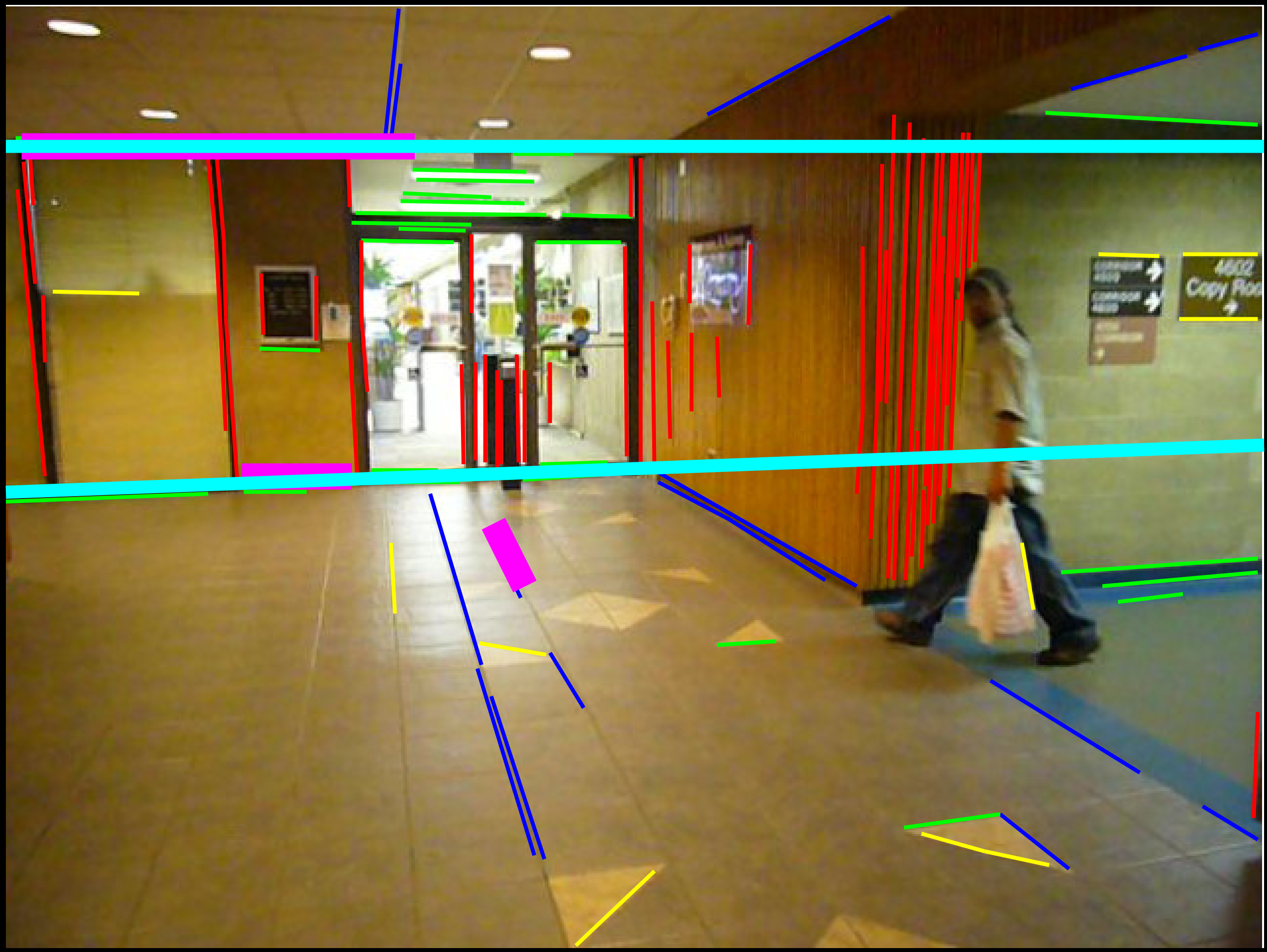
Generating Hypotheses

Generating Hypotheses

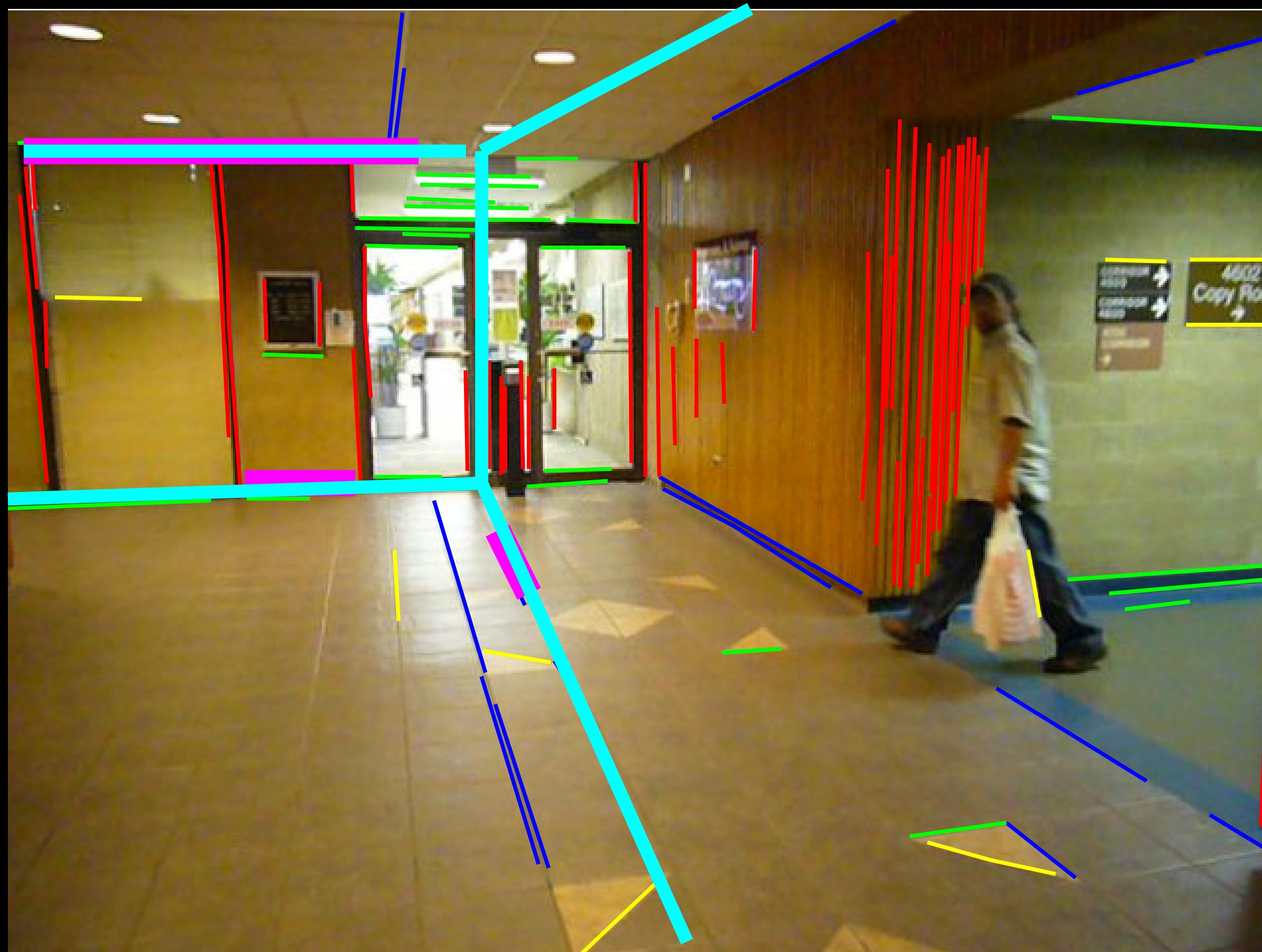
Generating Hypotheses



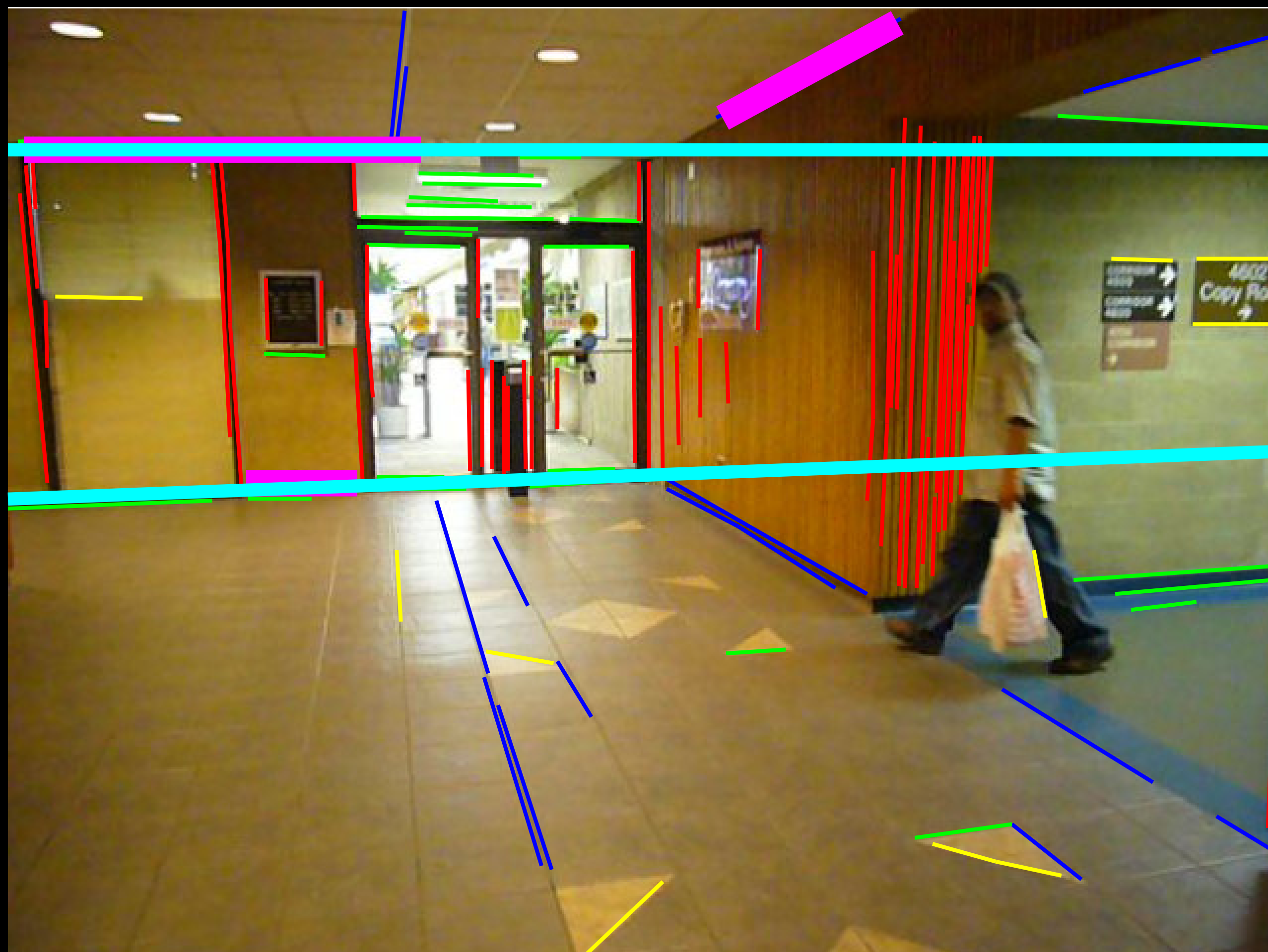
Generating Hypotheses



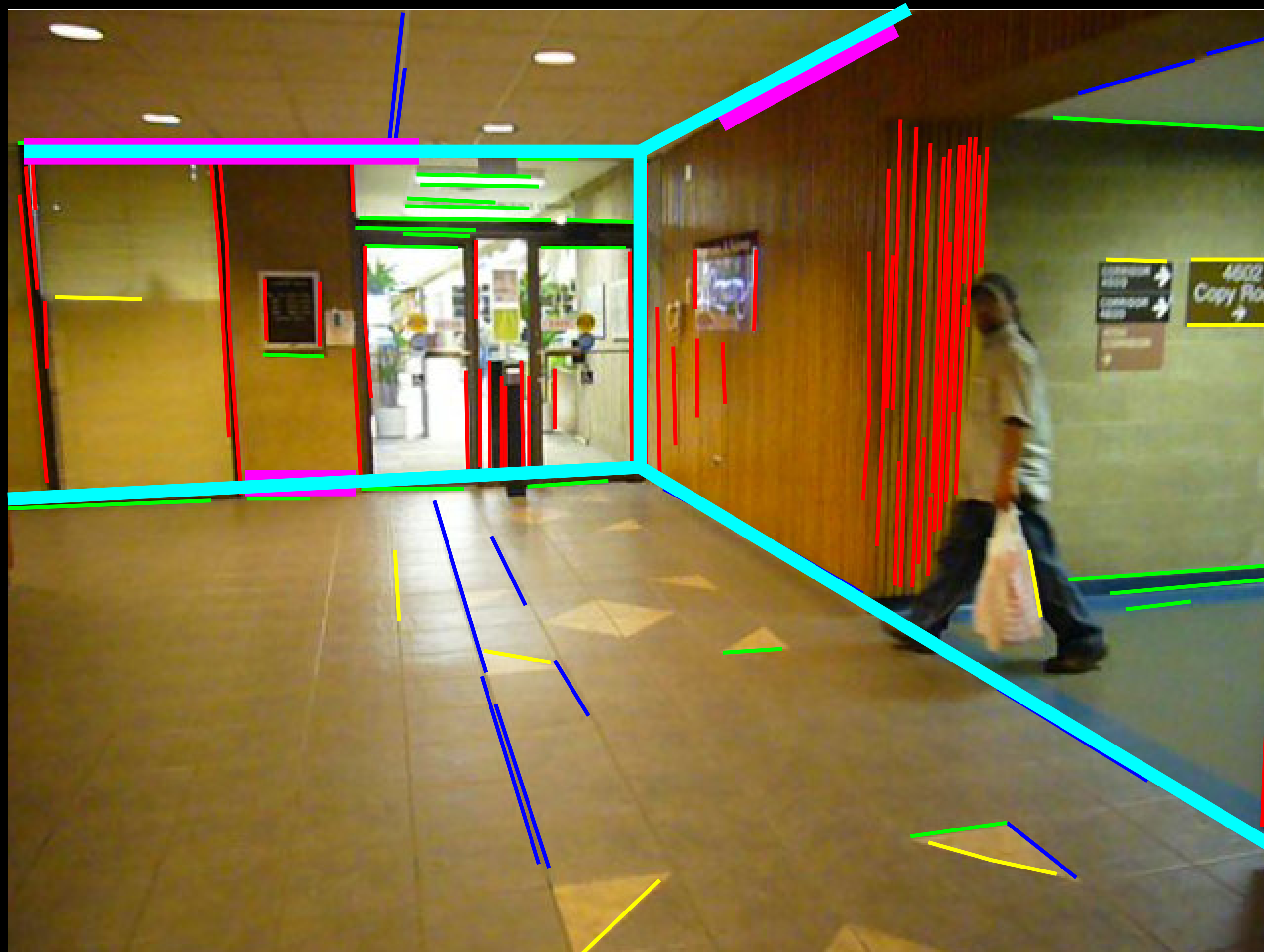
Generating Hypotheses



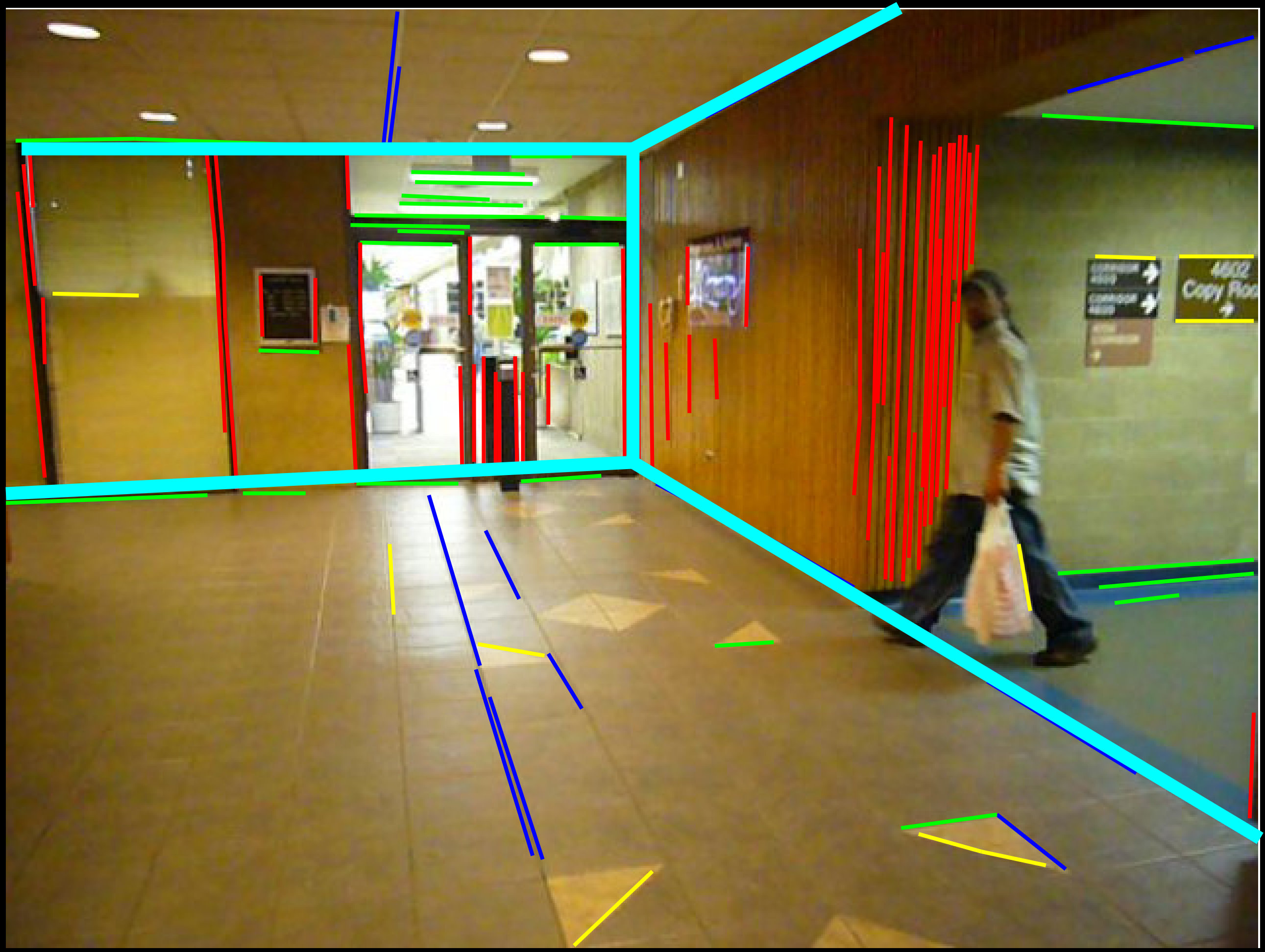
Generating Hypotheses



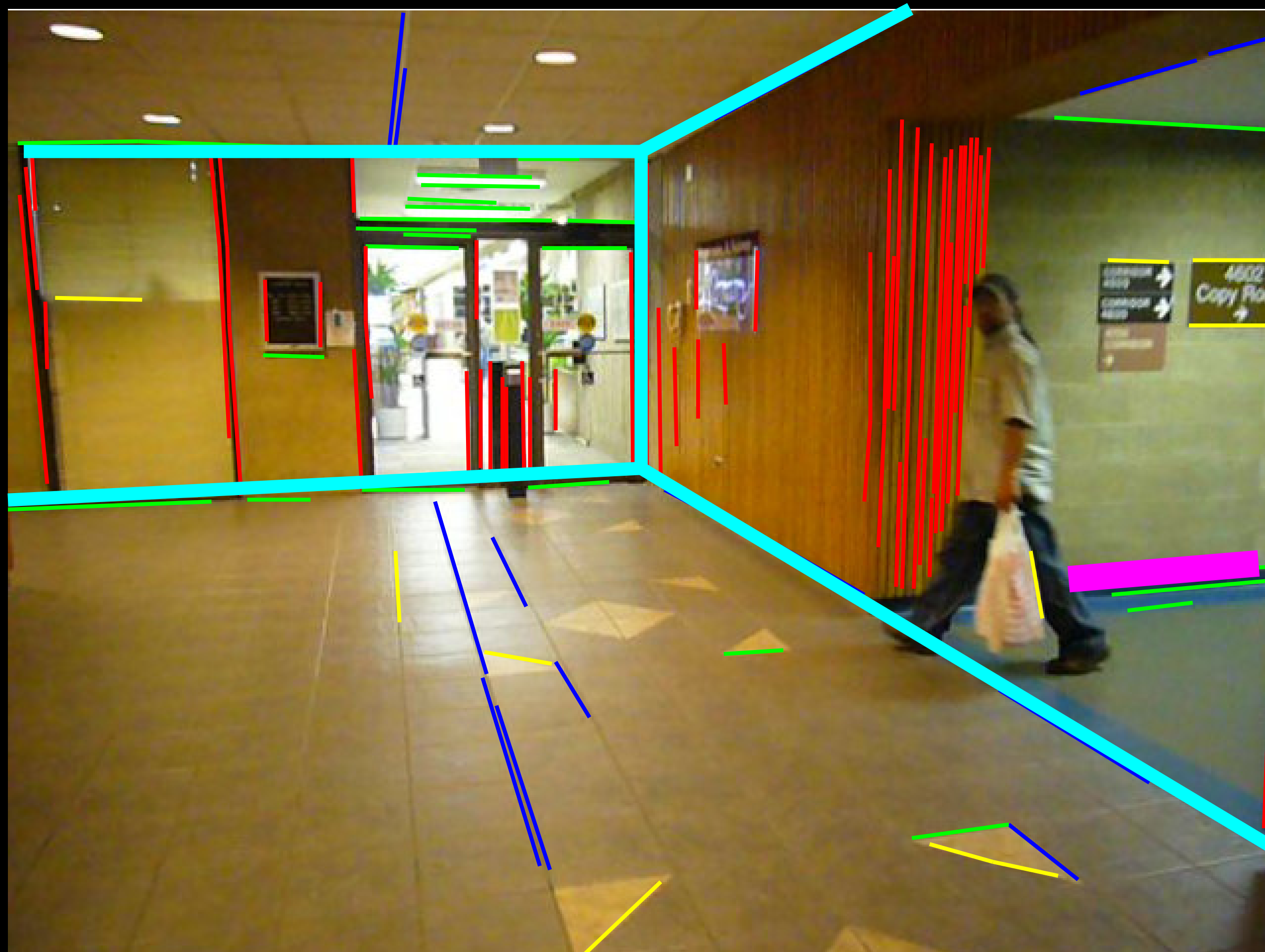
Generating Hypotheses



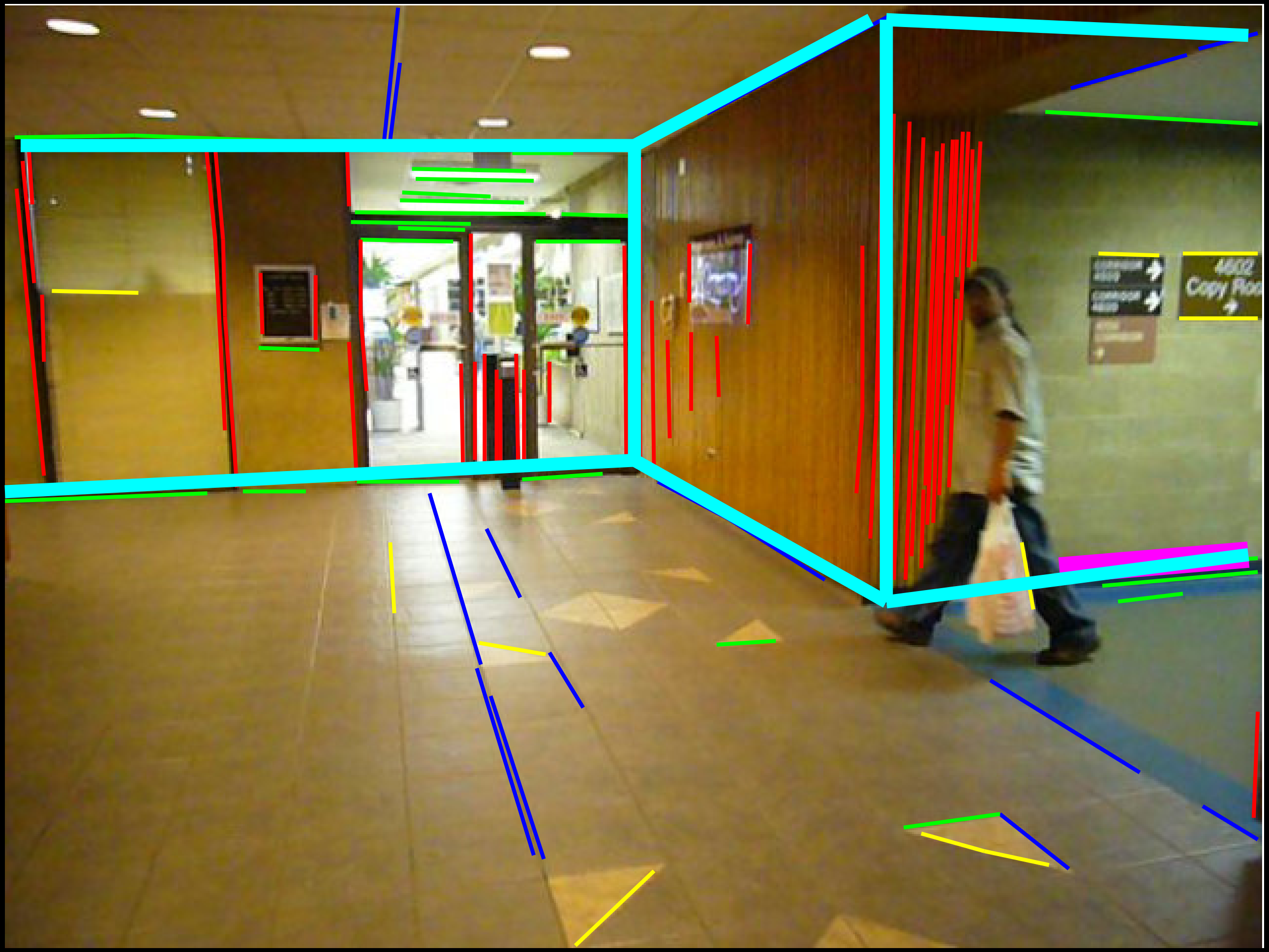
Generating Hypotheses



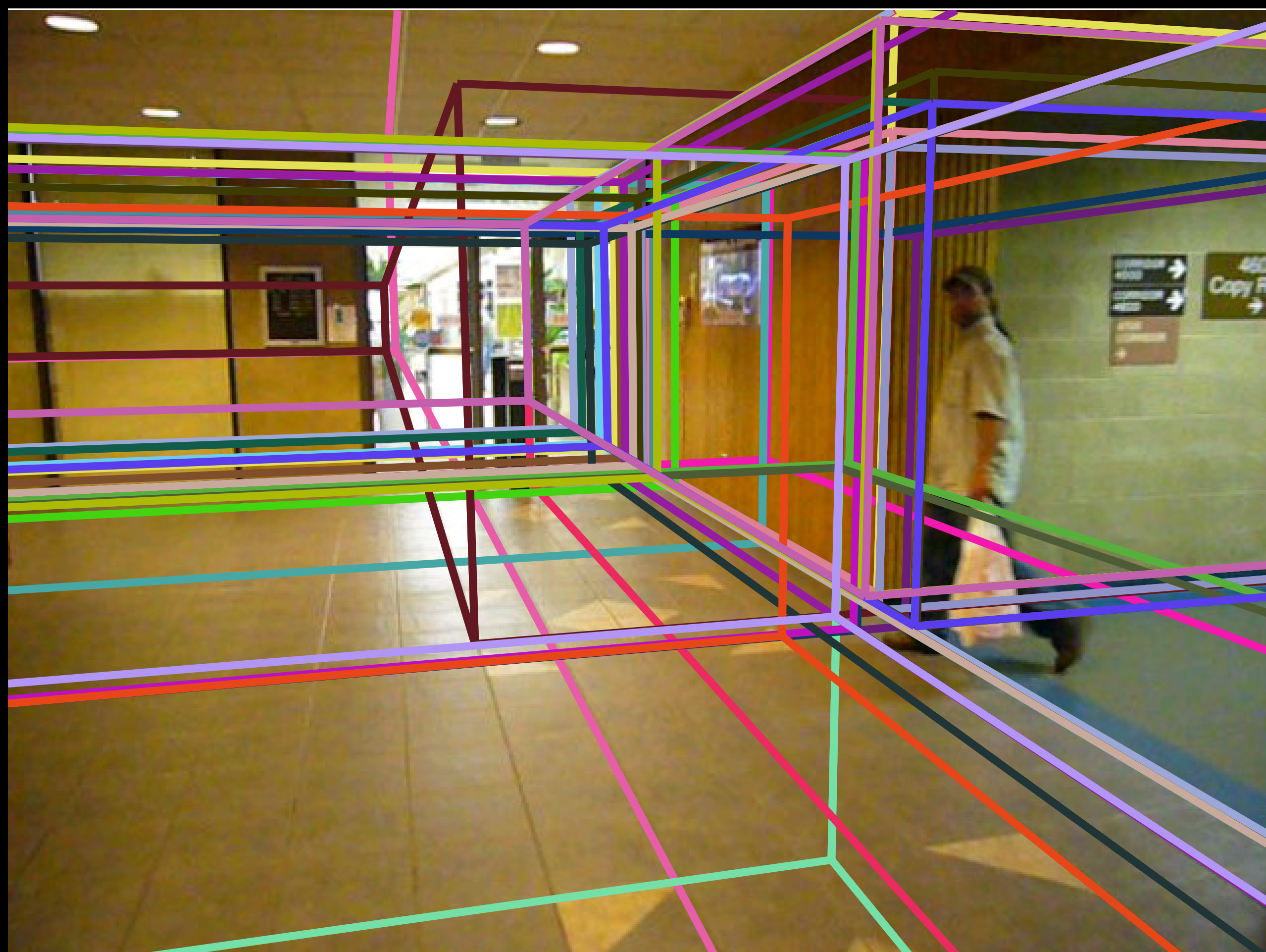
Generating Hypotheses



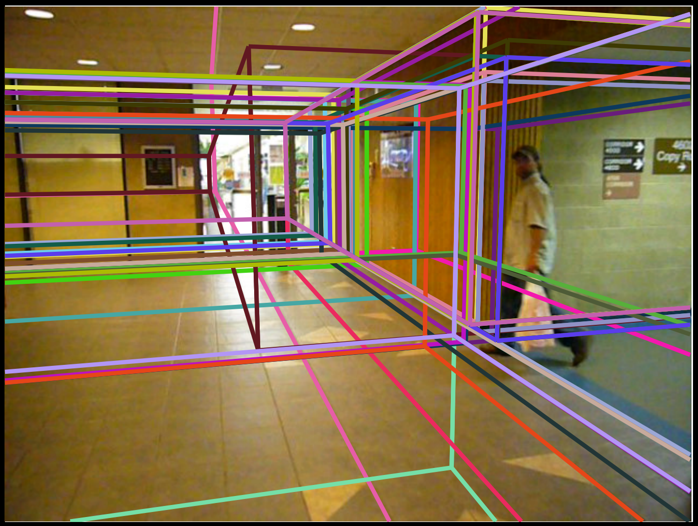
Generating Hypotheses

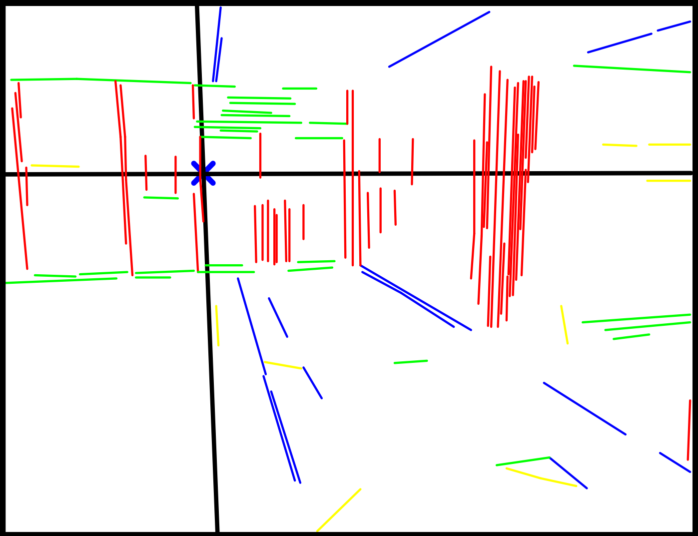


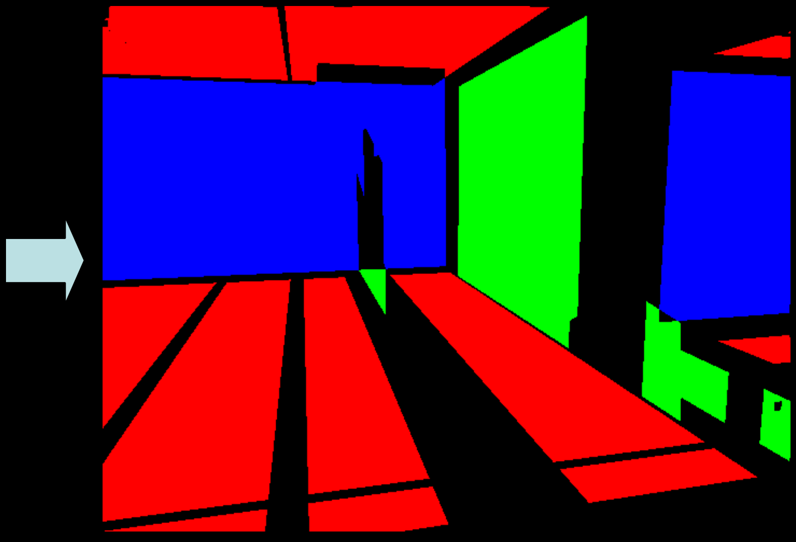
Generating Hypotheses

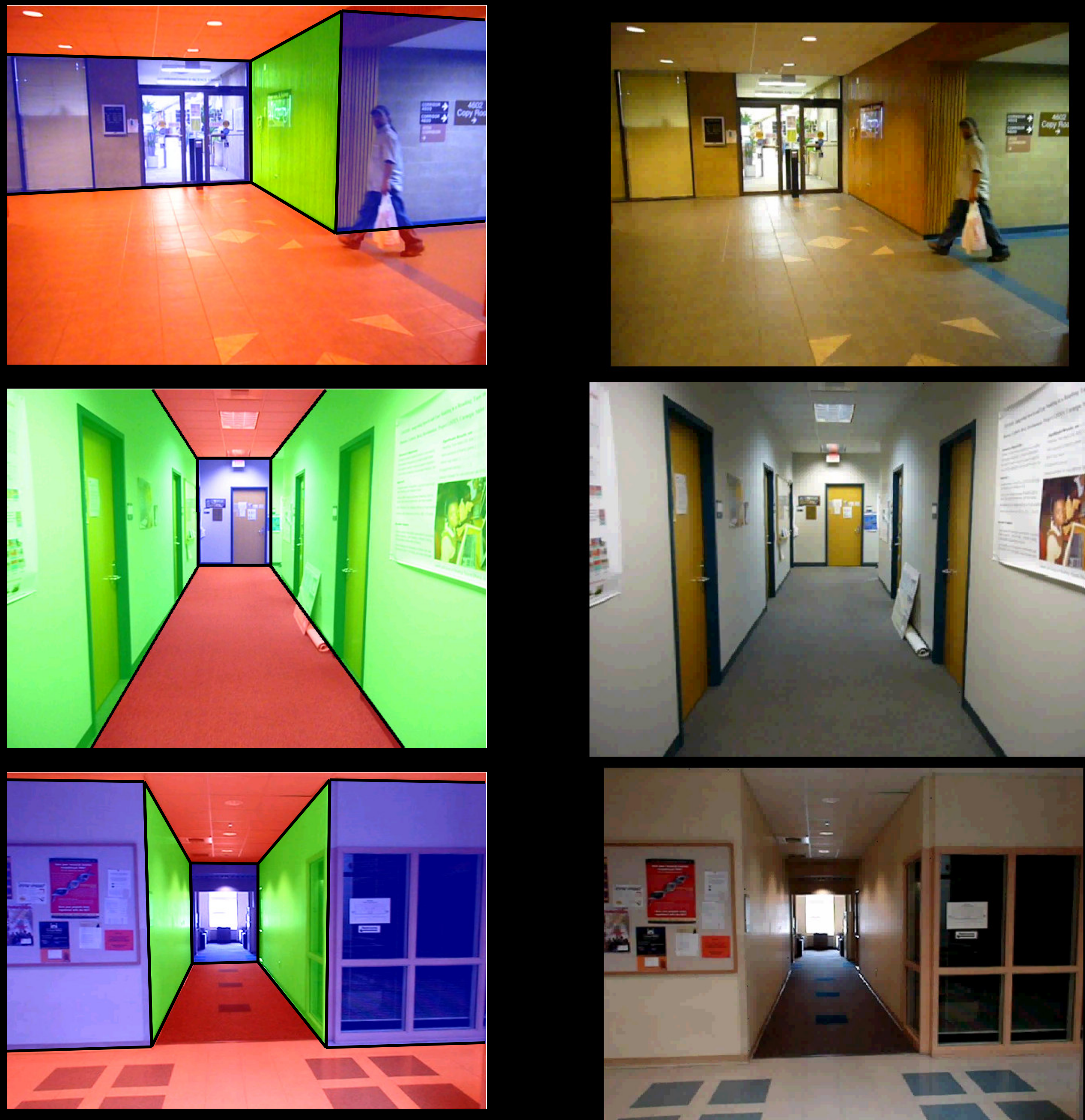


Evaluating scene hypotheses

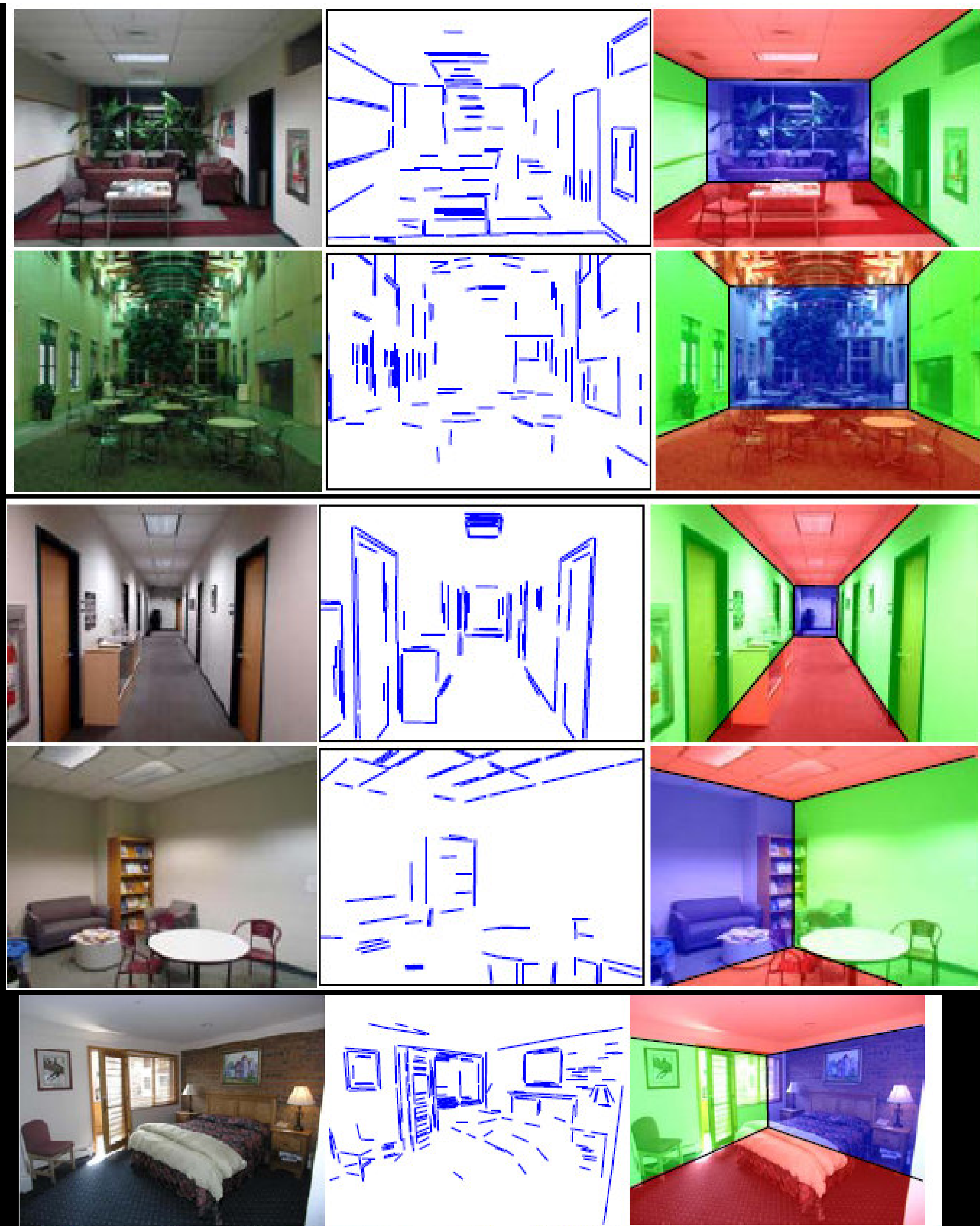




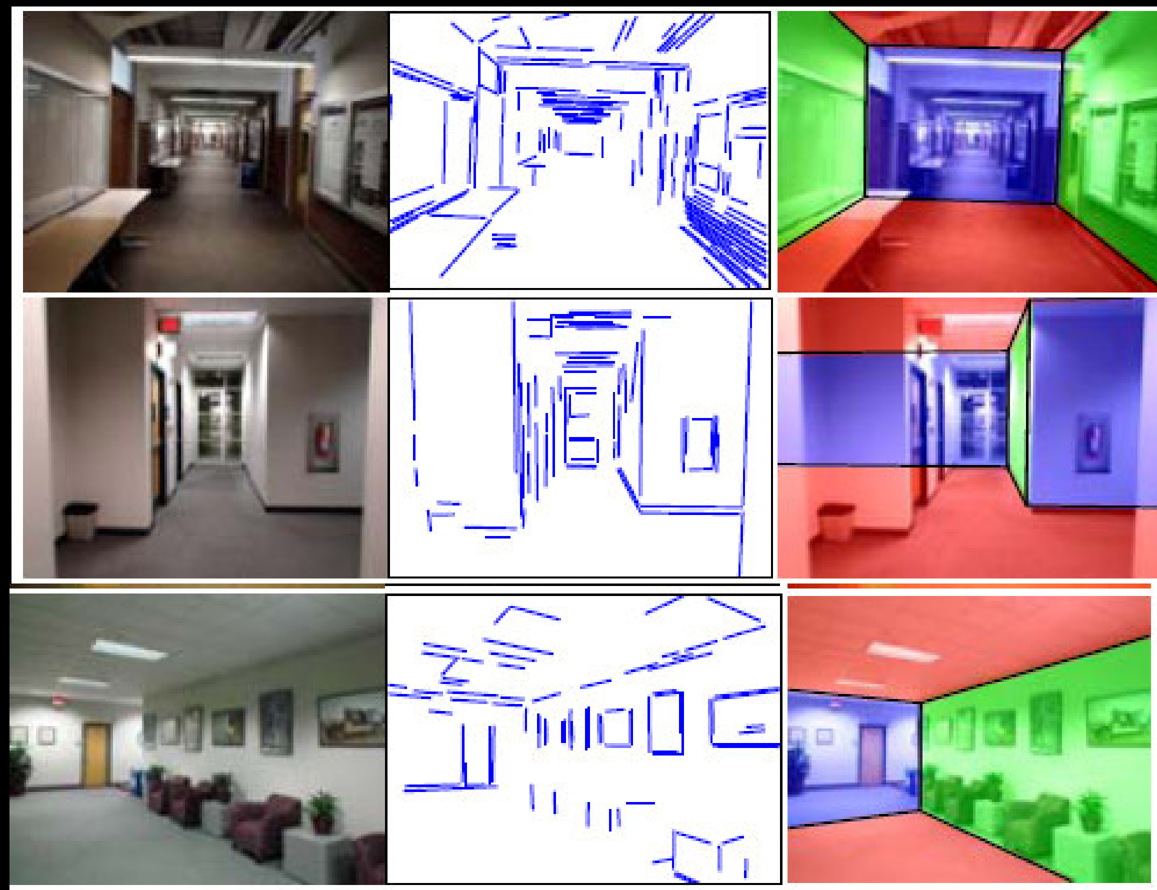




[Lee *et al.* CVPR'09]

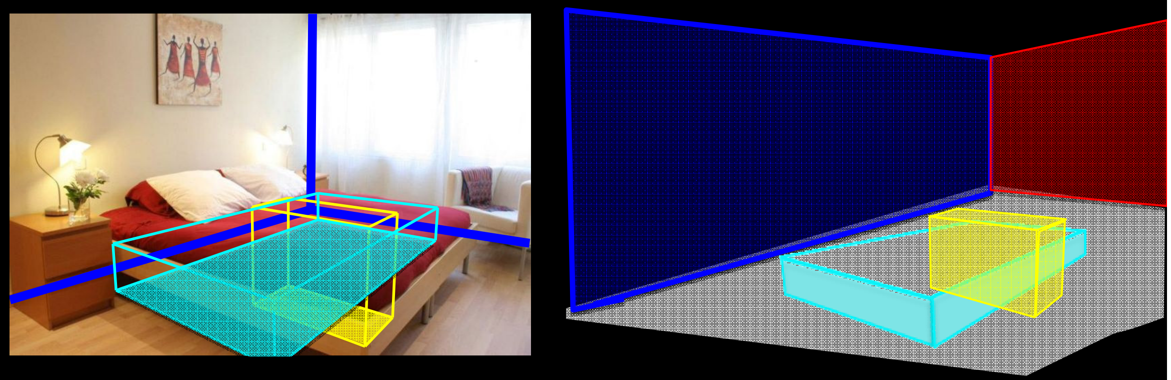


Evaluated on data from UIUC (Hoiem) and BC (Yu)

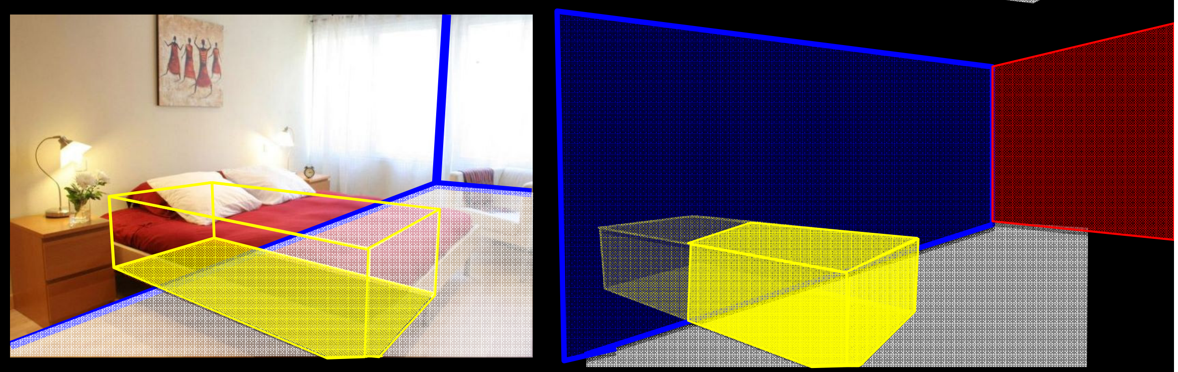


Constraints: Solid objects must satisfy physical constraints

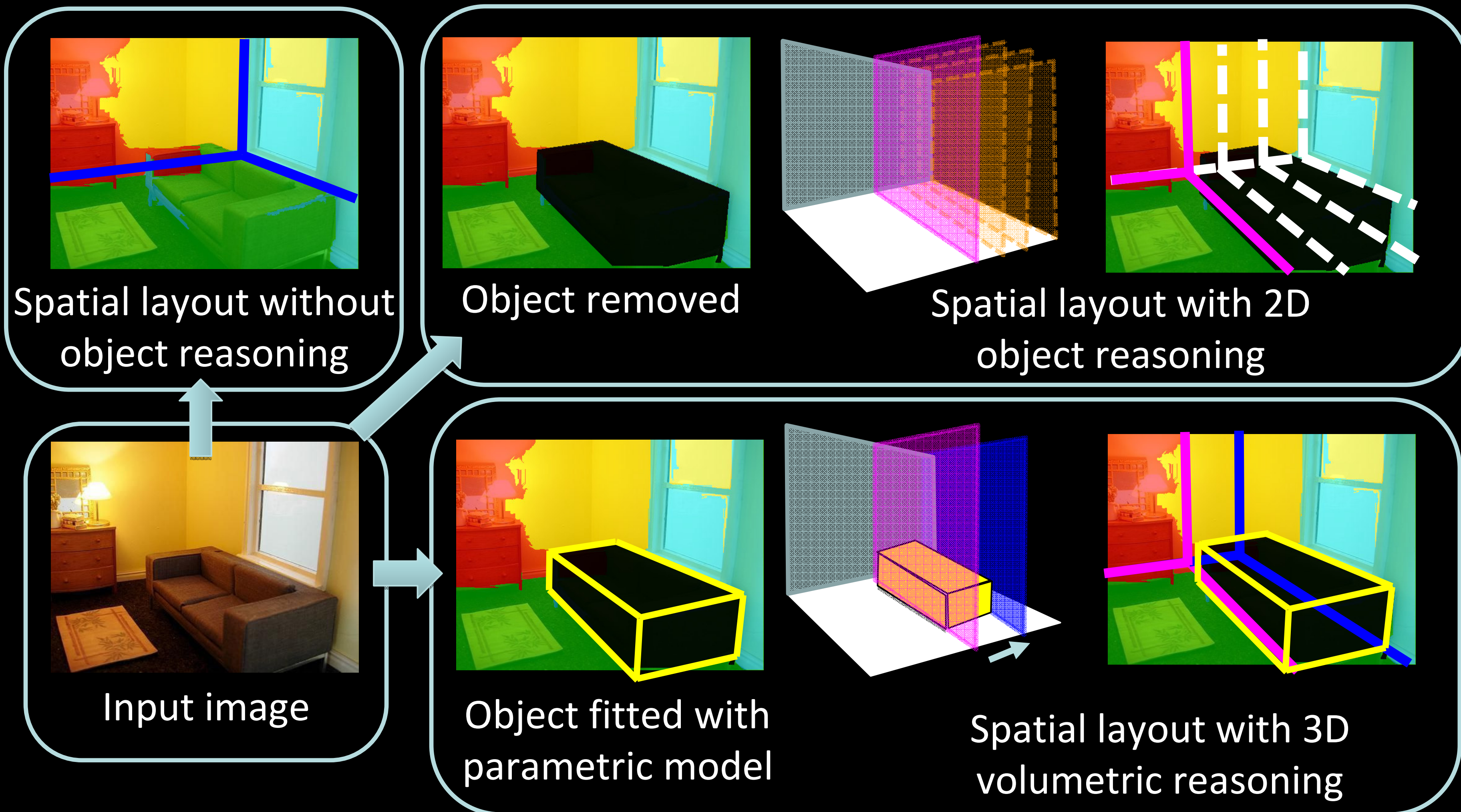
- Finite volume
- Spatial exclusion

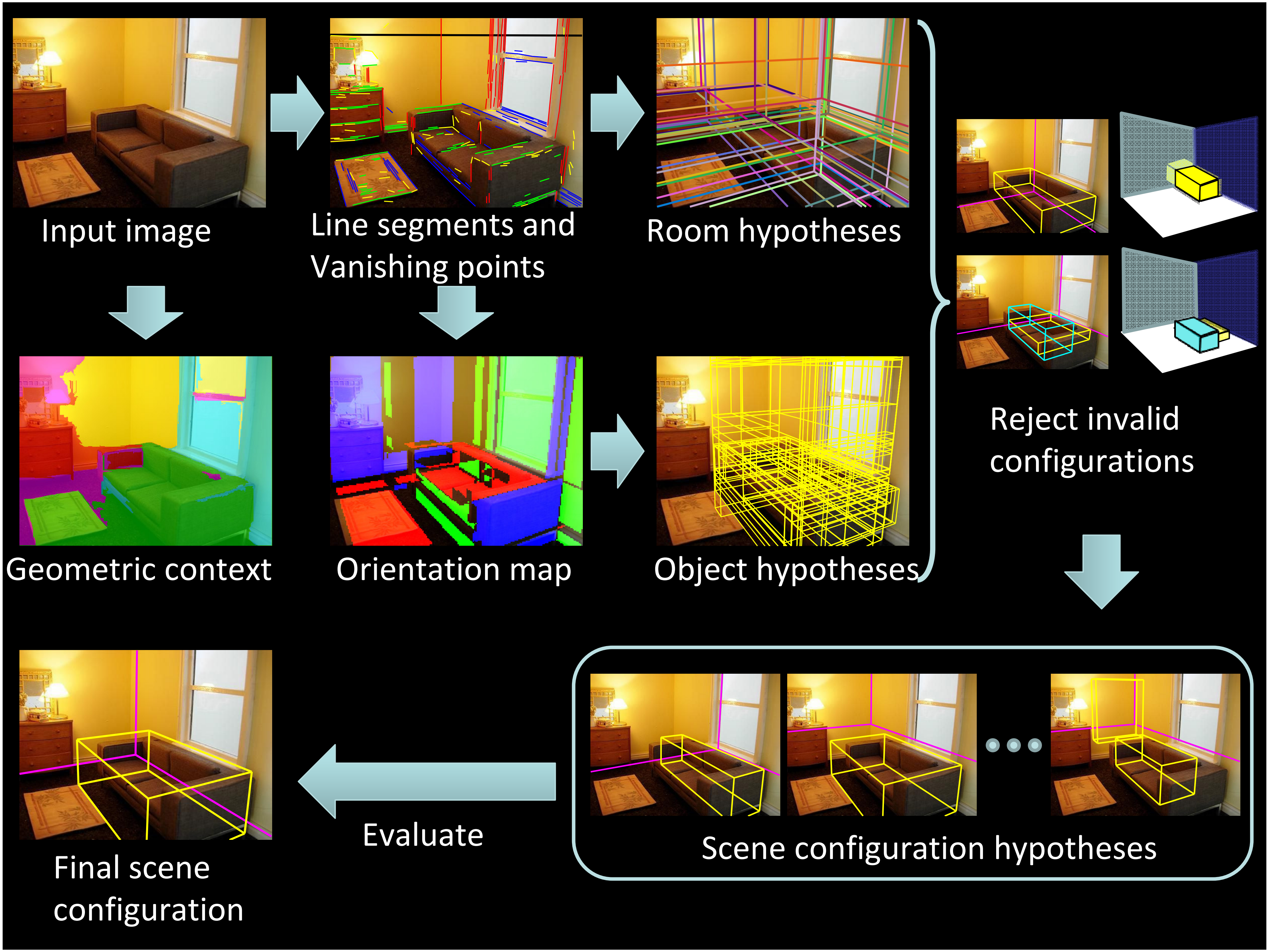


- Containment

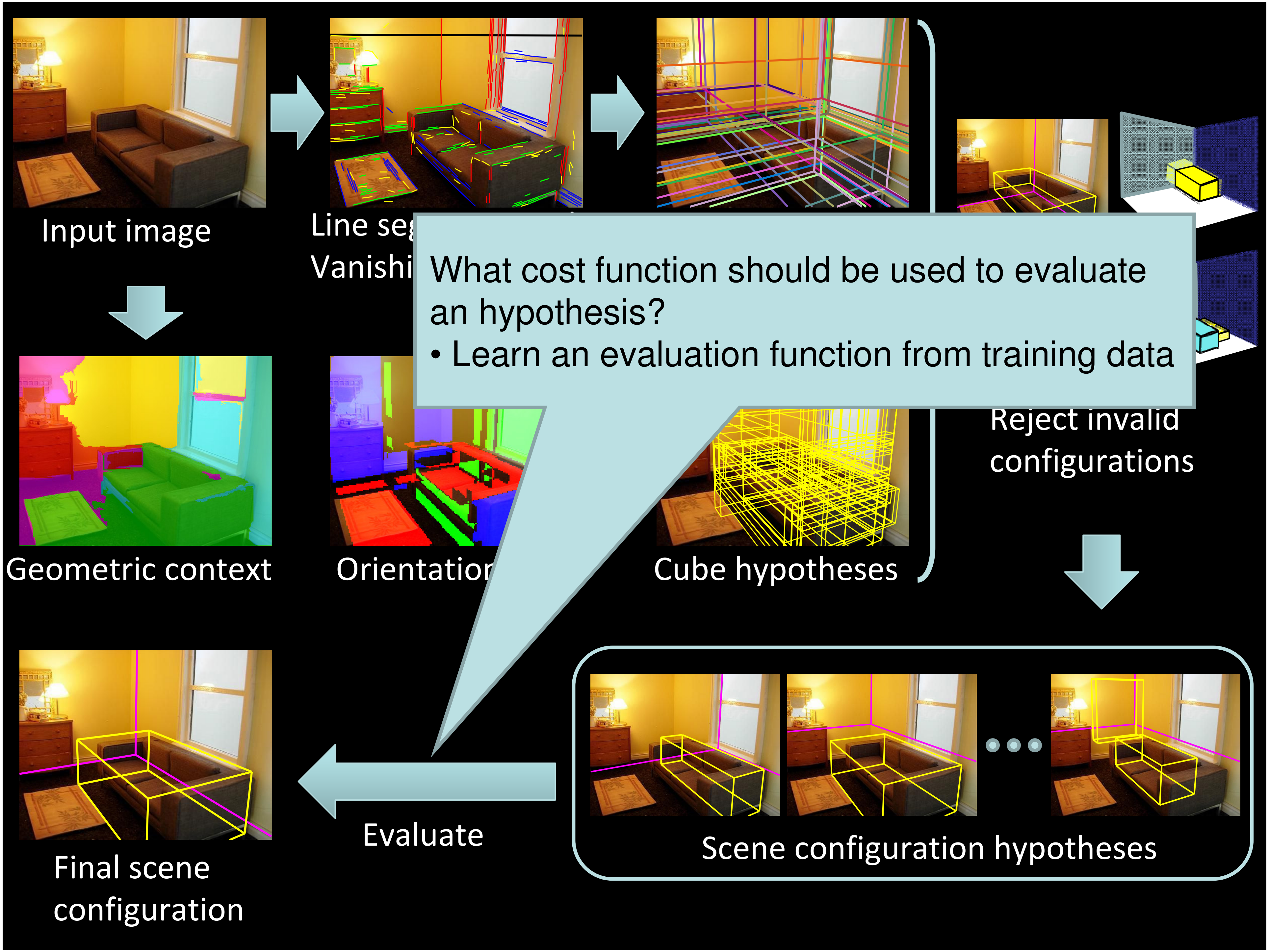


Why more constraints? Volume vs. surface reasoning



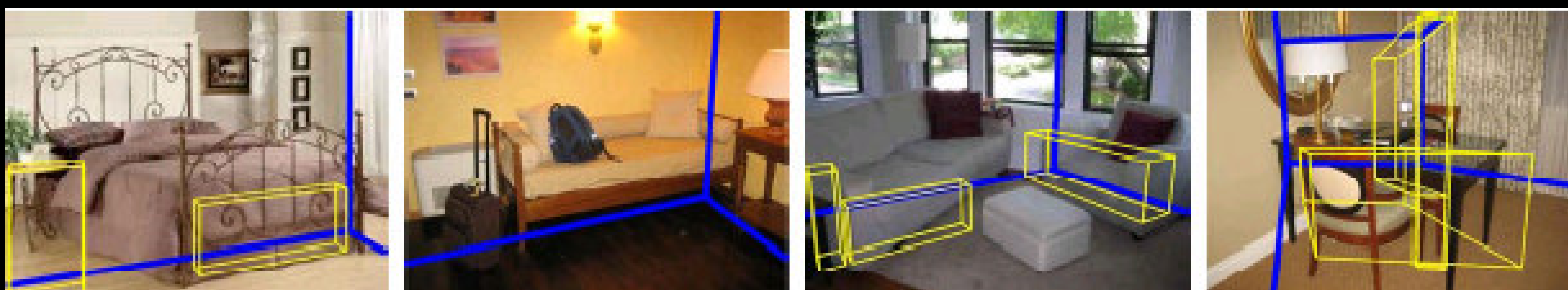


Input image
Line segments and Vanishing points
Room hypotheses
Geometric context
Orientation map
Object hypotheses
Reject invalid configurations
Final scene configuration
Evaluate
Scene configuration hypotheses

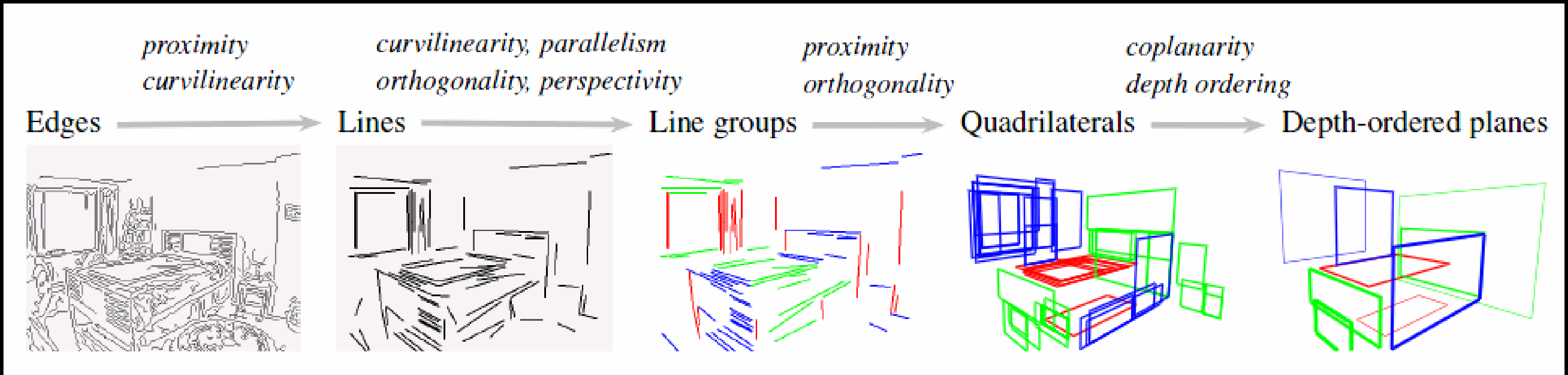


Input image
Line seg
Vanishi
What cost function should be used to evaluate an hypothesis?
• Learn an evaluation function from training data
Reject invalid configurations
Geometric context
Orientation
Cube hypotheses
Final scene configuration
Evaluate
Scene configuration hypotheses

Relative improvement on surface labeling >10% on UIUC data set

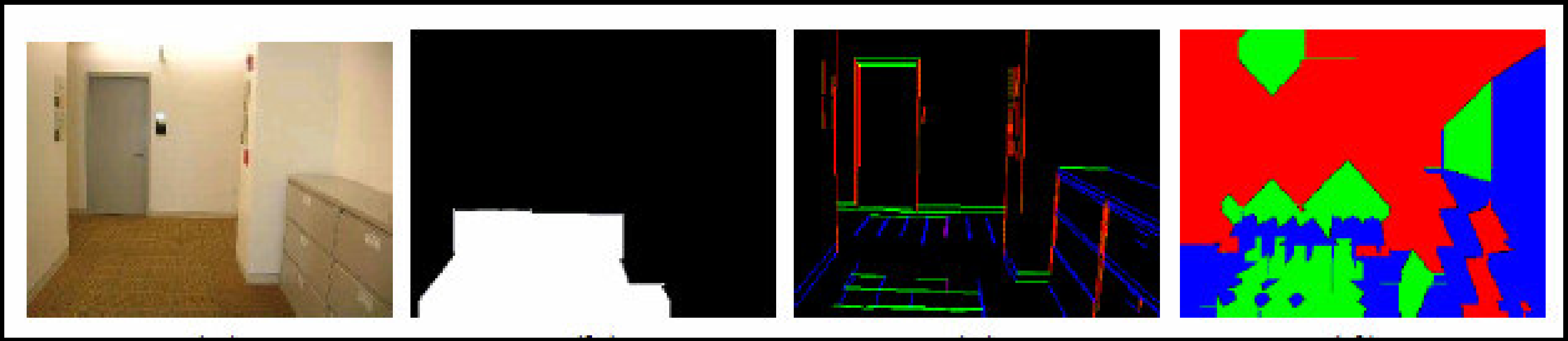


- *Iterative grouping:* X. Yu, Hao Zhang, and Jitendra Malik. Inferring Spatial Layout from A Single Image via Depth-Ordered Grouping. Workshop on Perceptual Organization in Computer Vision, 2008.

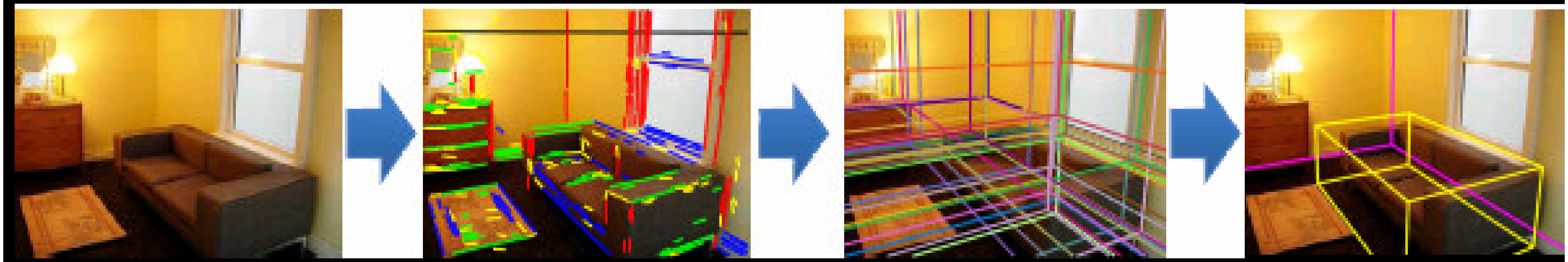


- *Fusion line features + surface labels:* V. Hedau , D. Hoiem, D.Forsyth. Recovering the Spatial Layout of Cluttered Rooms. ICCV09.

- *Line and color features + MRF:* E. Delage, H. Lee, and A. Y. Ng. Automatic Single-Image 3d Reconstructions of Indoor Manhattan World Scenes. ISRR05.



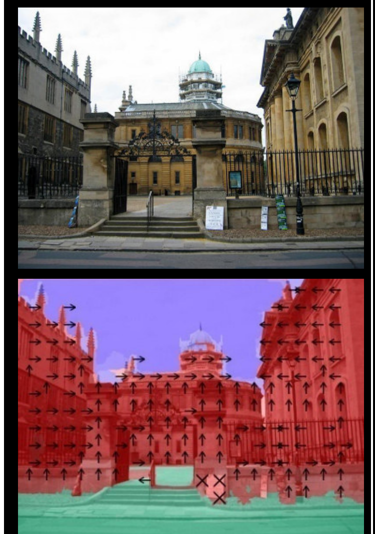
- *Hypothesis generation and verification:* D. Lee, T. Kanade, M. Hebert. Geometric Reasoning for Single Image Structure Recovery. CVPR09. (+ NIPS 2010)



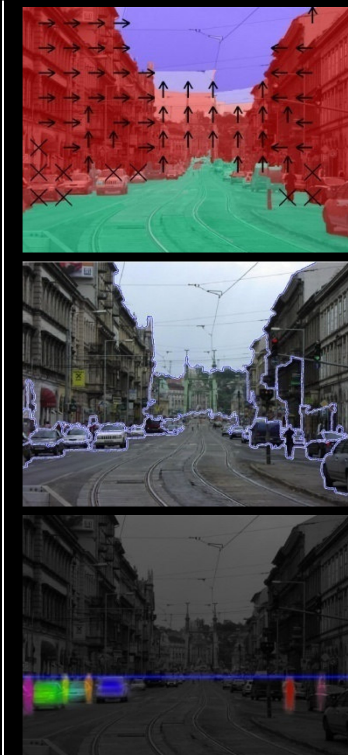
Comments

- Plus:
  - Added explicit reasoning about domain constraints
  - Combine reasoning through multiple hypotheses with learning task
- Minus:
  - Relies mostly on top-down constraint satisfaction with limited use of bottom-up learned models
  - Incorporates specific domain knowledge about geometry, little knowledge about other constraints of the real world

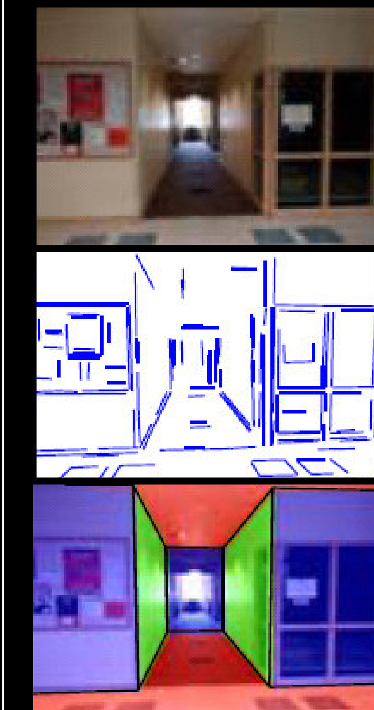
Levels of 3D-ness



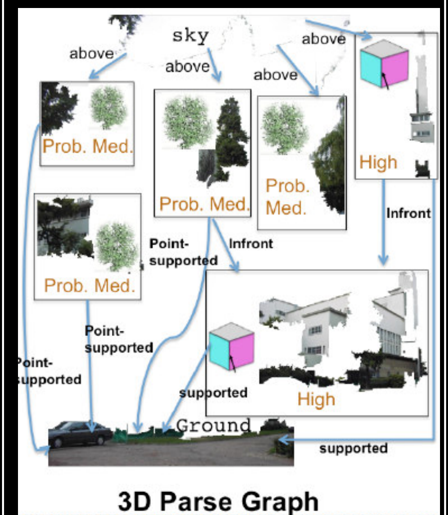
Region labels



+ Boundaries and objects



Stronger geometric constraints from domain knowledge



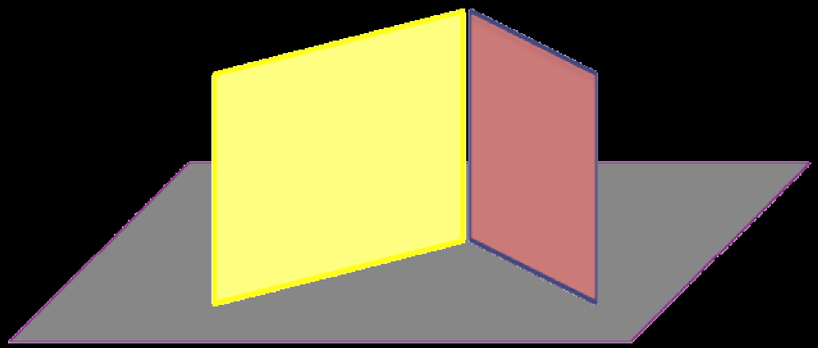
+ constraints from statics of solids

Qualitative

More quantitative
more precise

Explicit

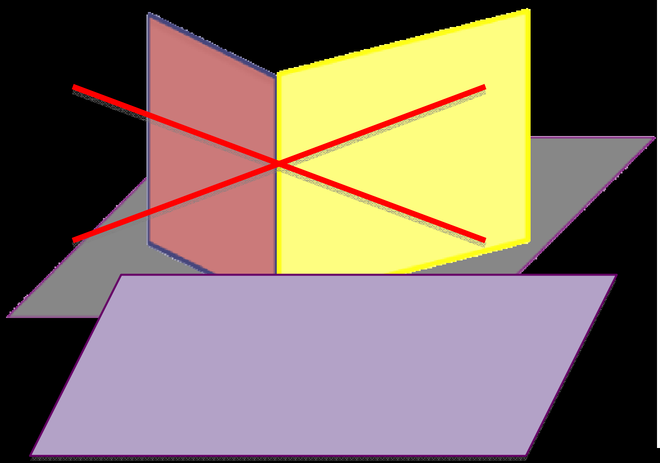
Moving along: From surfaces to objects

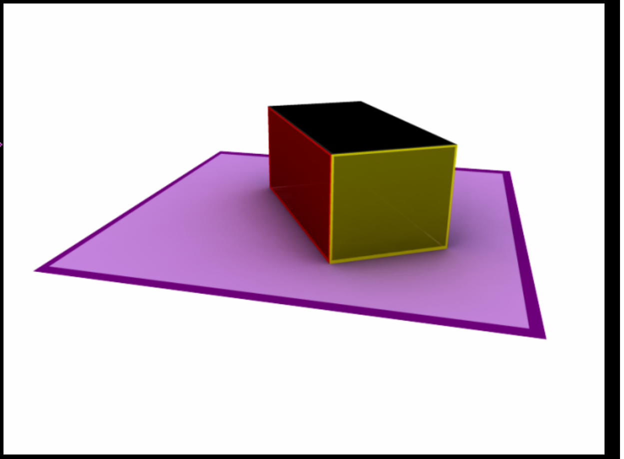


Qualitative 3D surface model

Moving along: From surfaces to objects

- But the world is not a set of surfaces
- It is a set of solid objects
- First approximation: solid objects = blocks
- We can define a richer set of constraints once we recognize that the world is populated by solid objects





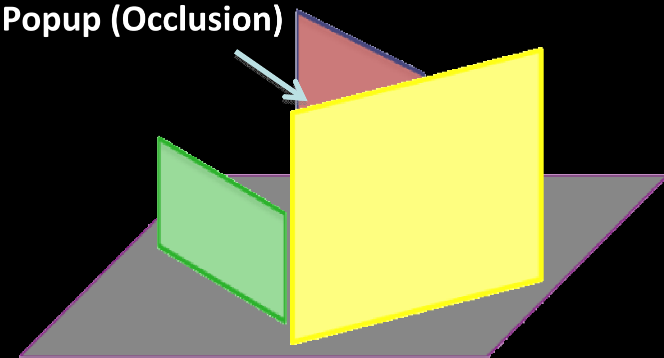
[A. Gupta, A. Efros, and M. Hebert. *Blocks World Revisited: Image Understanding Using Qualitative Geometry and Mechanics.* ECCV 2010]

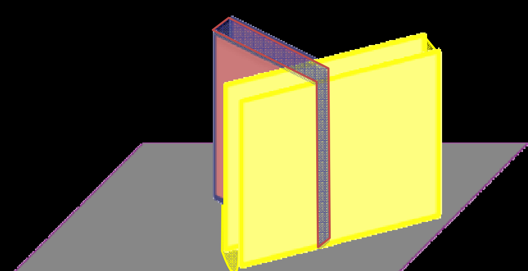
Before

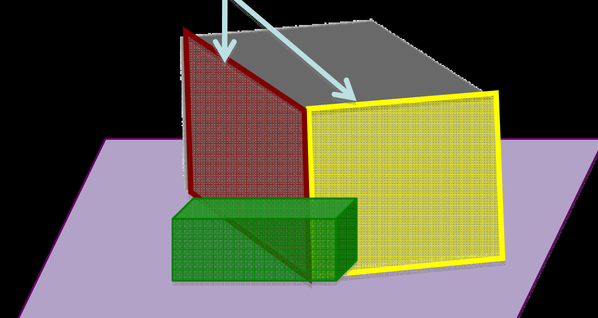
Now

Physical constraints

***1. Volumetric constraint:* Surfaces must form (partially visible) blocks**

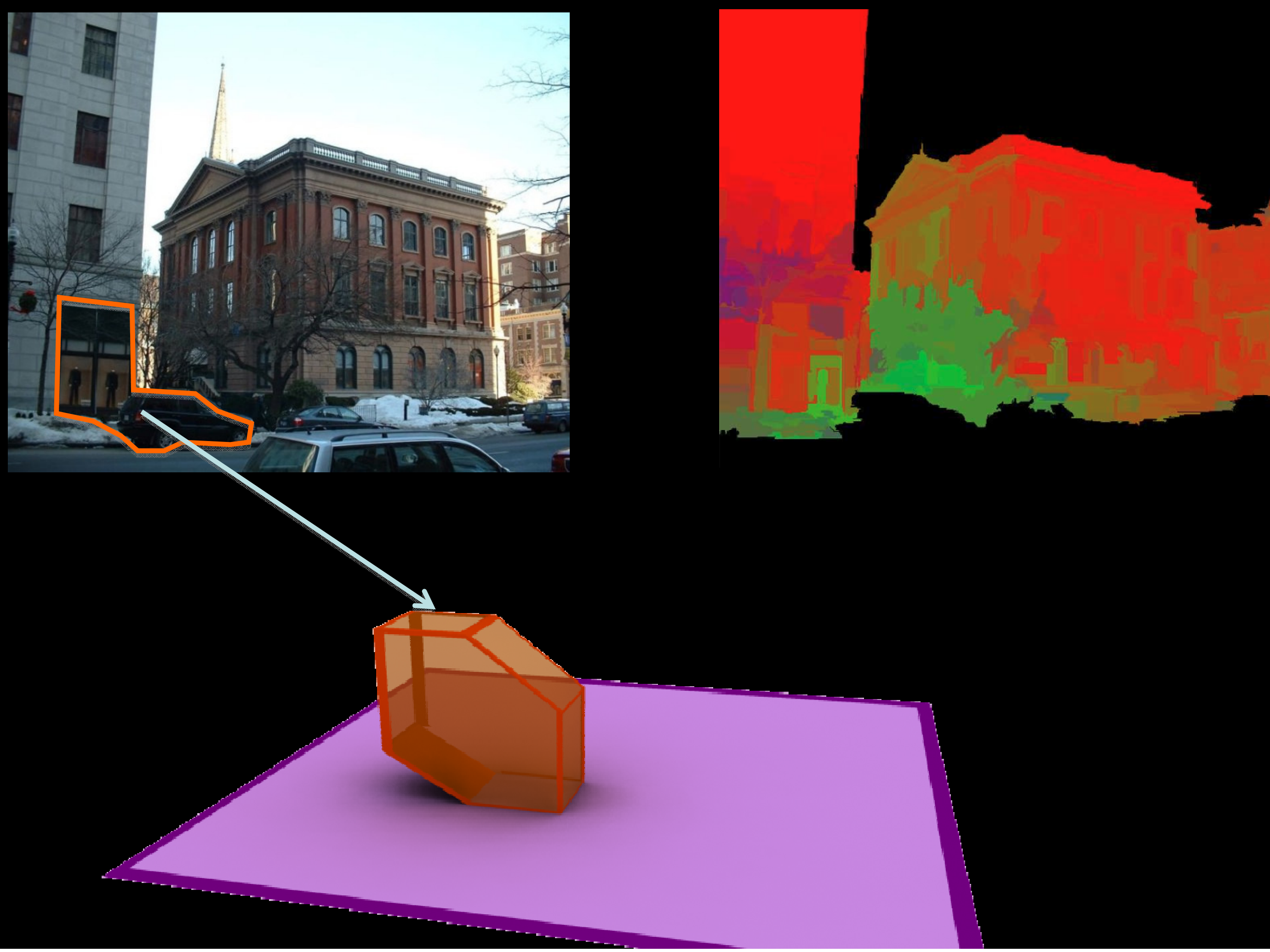






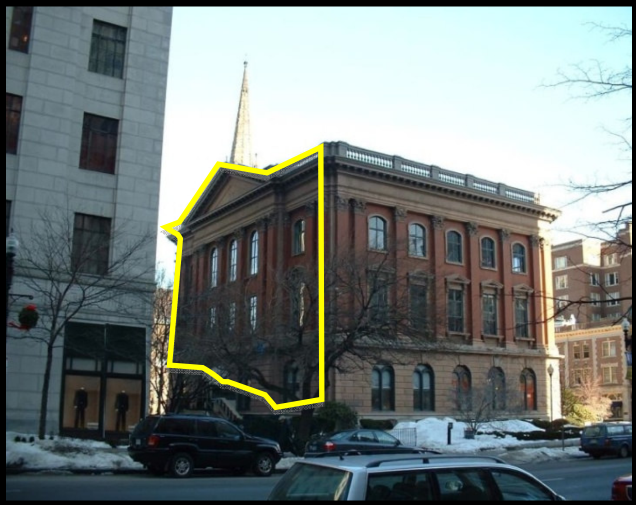
Physical constraints

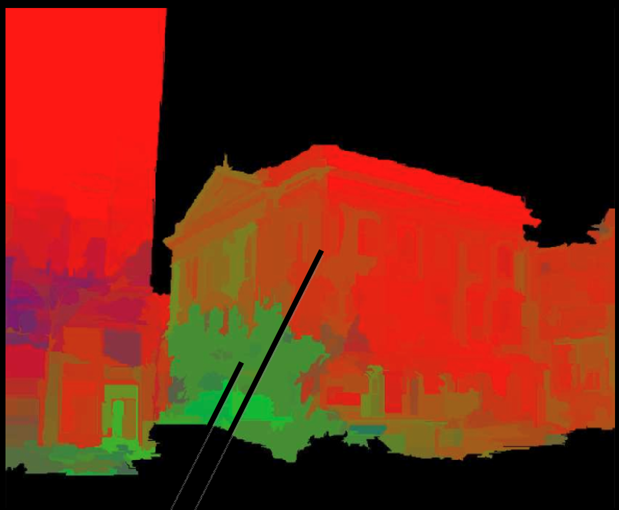
2. Static equilibrium

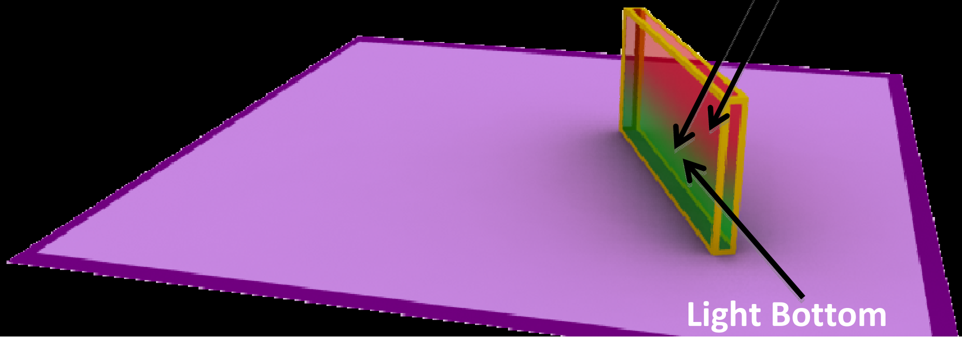


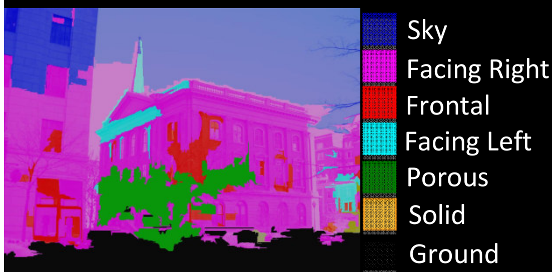
Physical constraints

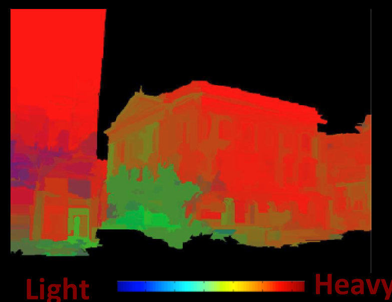
3. Internal stability











Geometry

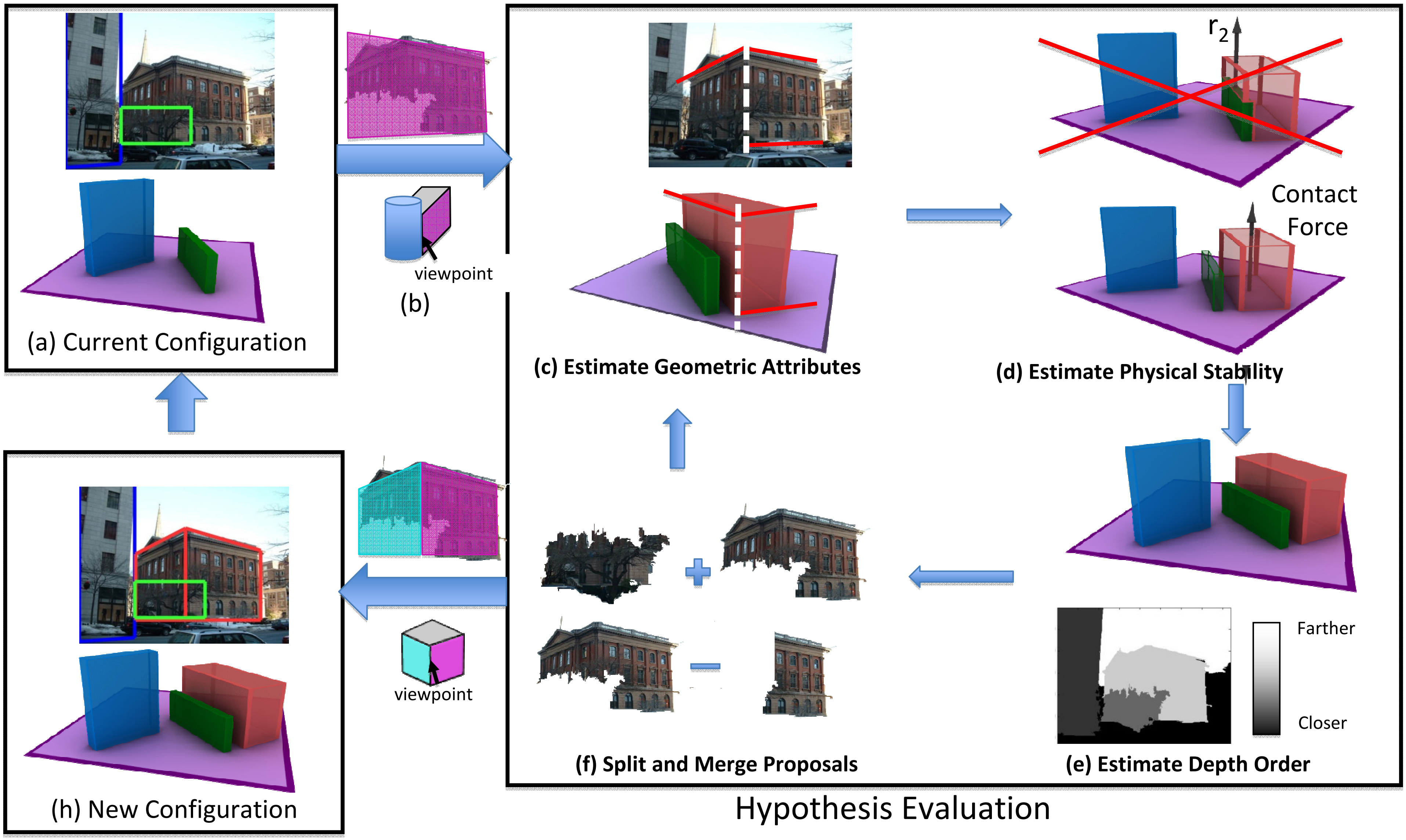
Density

Bag of Segments

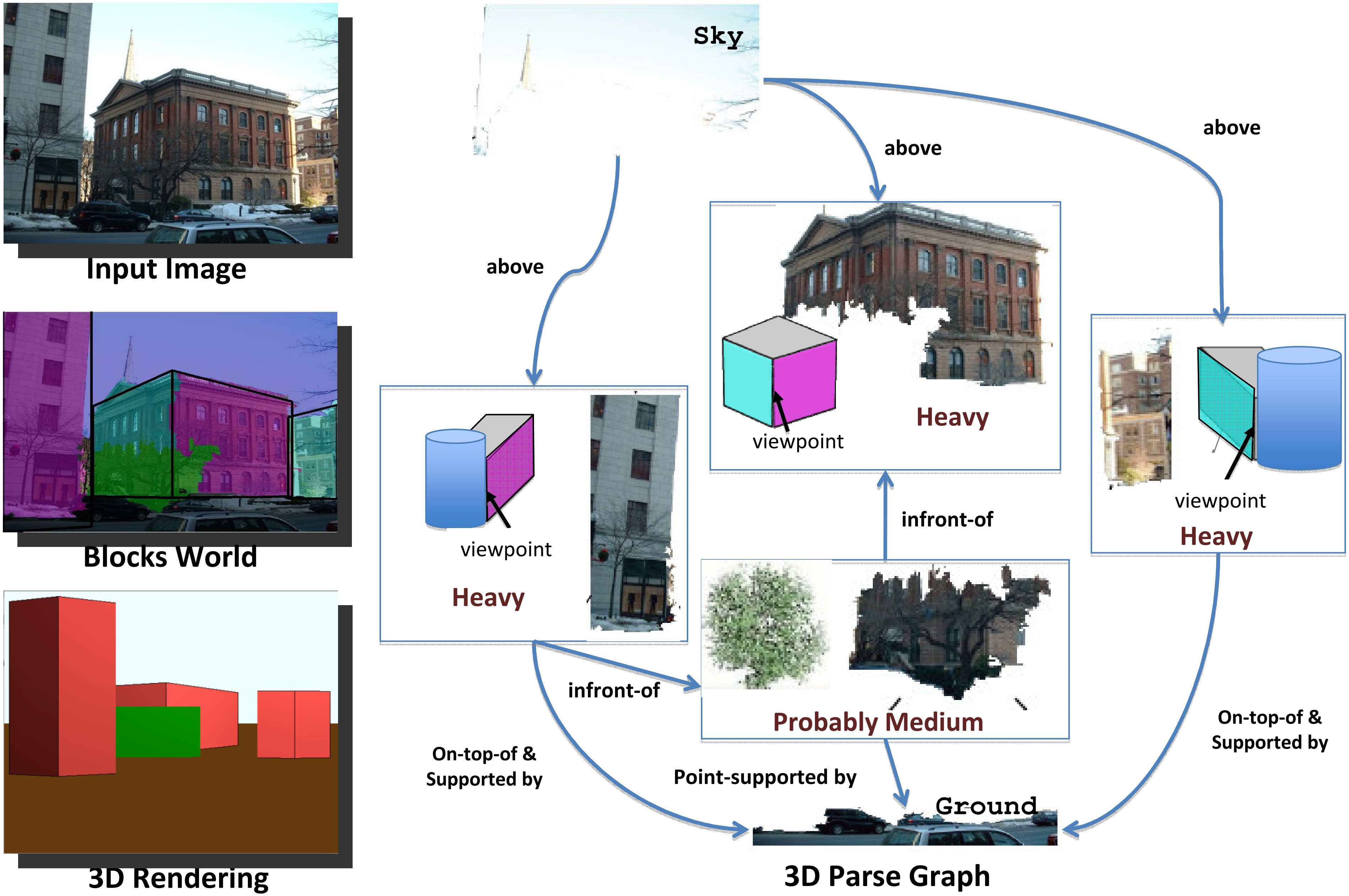
Initialize: Estimate cues from image

Iterate: Place blocks in the scene one by one

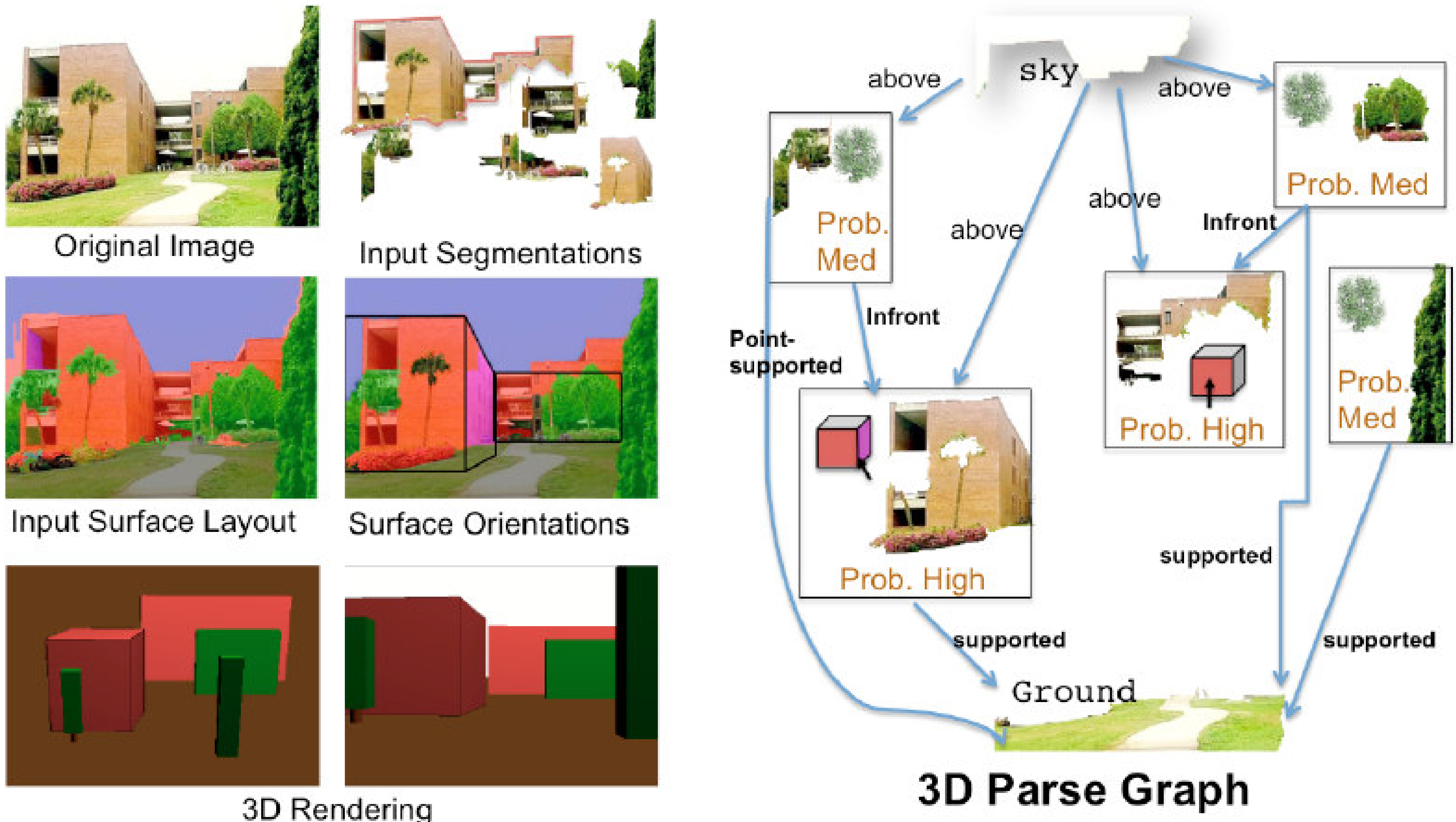
Building Blocks World



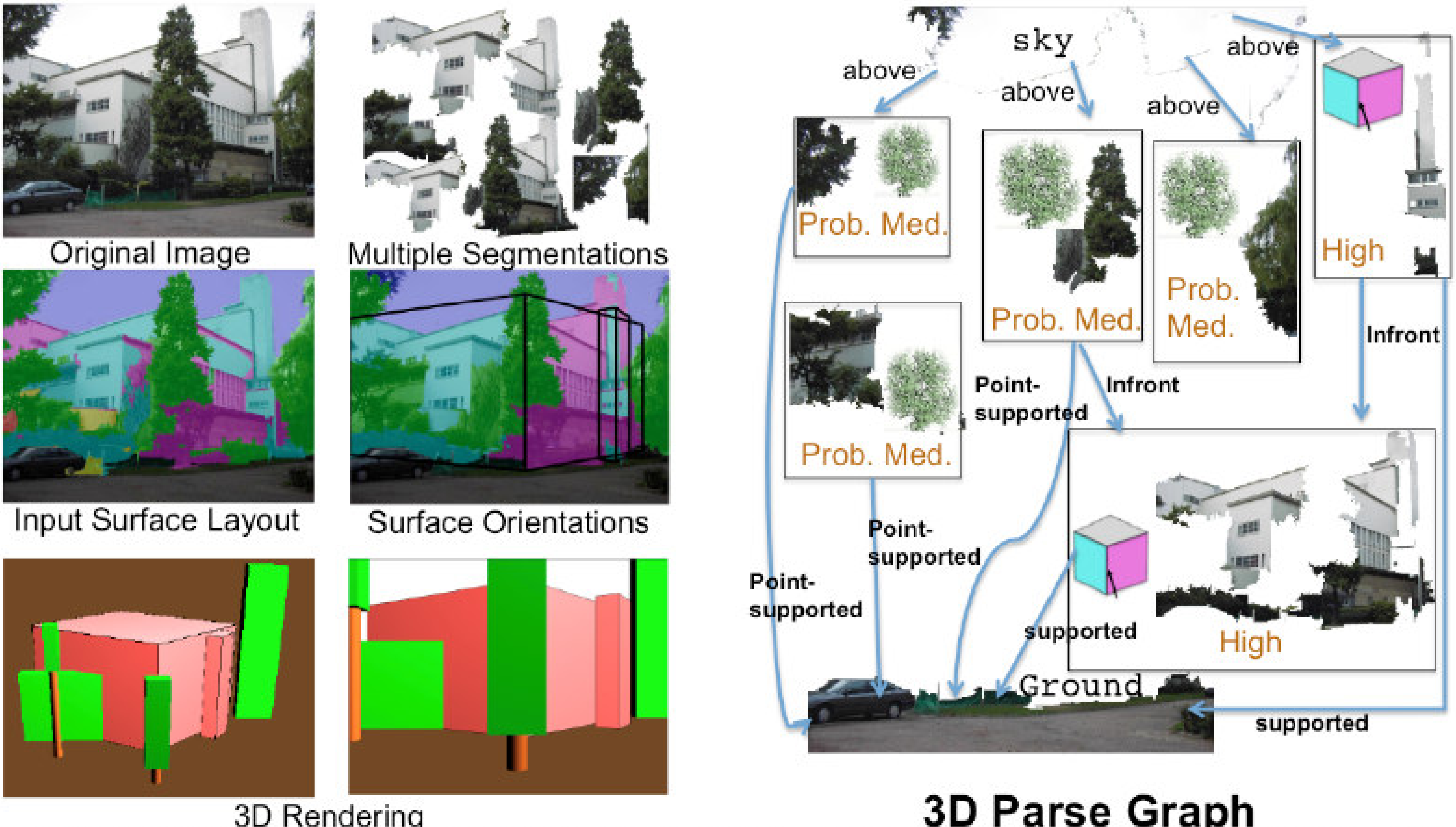
http://www.cs.cmu.edu/~abhinavg/blocksworld/



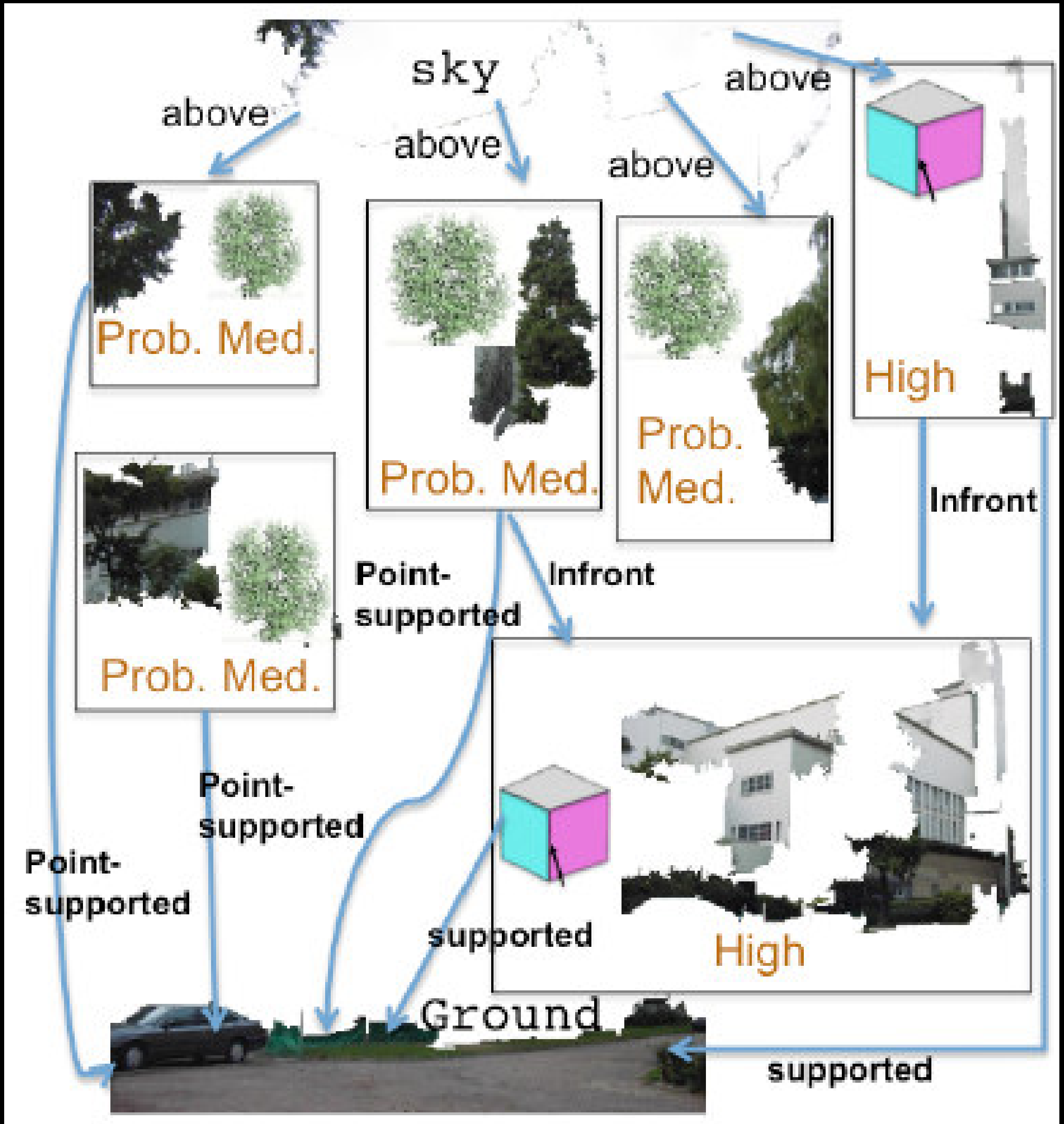
http://www.cs.cmu.edu/~abhinavg/blocksworld/



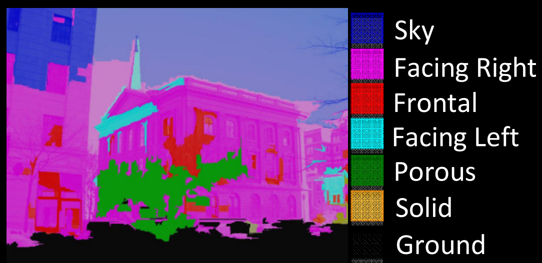
http://www.cs.cmu.edu/~abhinavg/blocksworld/



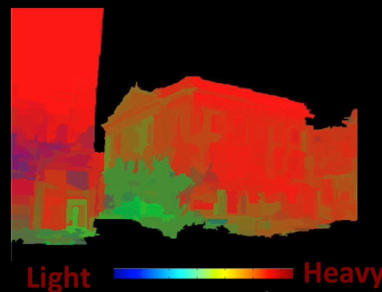
http://www.cs.cmu.edu/~abhinavg/blocksworld/



3D Parse Graph



Geometry



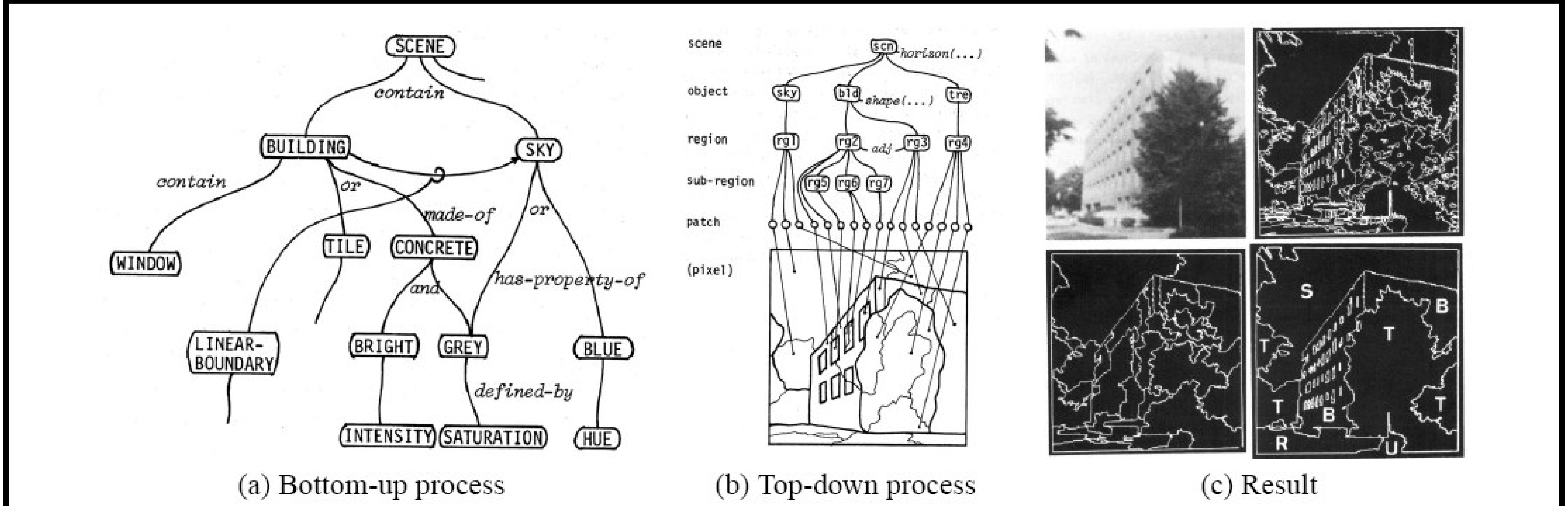
Density

Bag of Segments

$$\mathcal{C}(\mathcal{B}_i) = \mathcal{F}_{geometry}(\mathcal{G}_i) + \sum_{S \in ground, sky} \mathcal{F}_{contacts}(\mathcal{G}_i, S) + \mathcal{F}_{intra}(\mathcal{S}_i, \mathcal{G}_i, d)$$
$$+ \sum_{j \in blocks} \mathcal{F}_{stability}(\mathcal{G}_i, S_{ij}, \mathcal{B}_j) + \mathcal{F}_{depth}(\mathcal{G}_i, S_{ij}, \mathcal{D}),$$

- Approach combines:
  - Multiple segmentations
  - Set of bottom-up learned classifiers, each with a well-defined, “simple” task
  - *Control structure* to enable search through combinatorial set of hypothesis

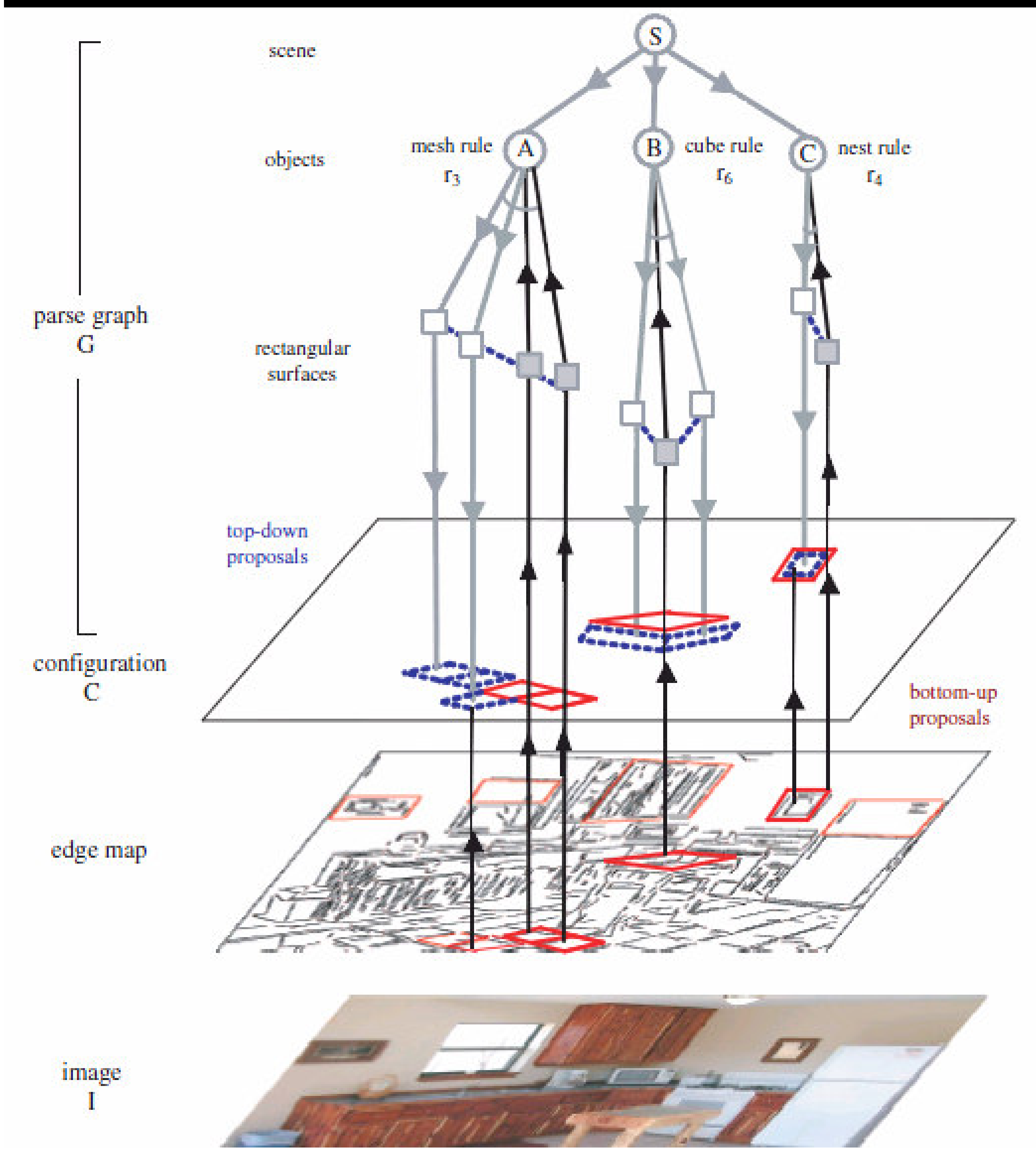
[See also: Munoz ECCV2010]

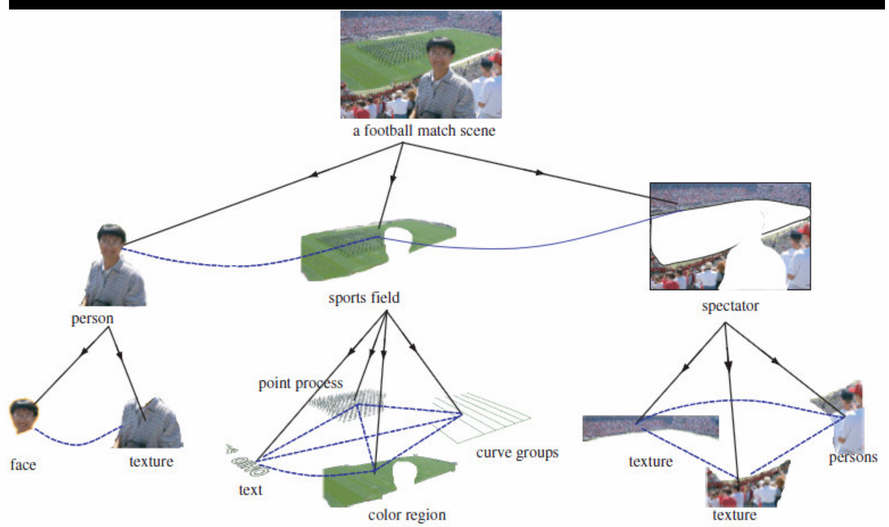


[Ohta & Kanade 1978]

- Guzman (*SEE*), 1968
- Yakimovsky & Feldman, 1973
- Hansen & Riseman (*VISIONS*), 1978
- Barrow & Tenenbaum 1978
- Brooks (*ACRONYM*), 1979
- Marr, 1982
- Ohta & Kanade, 1978

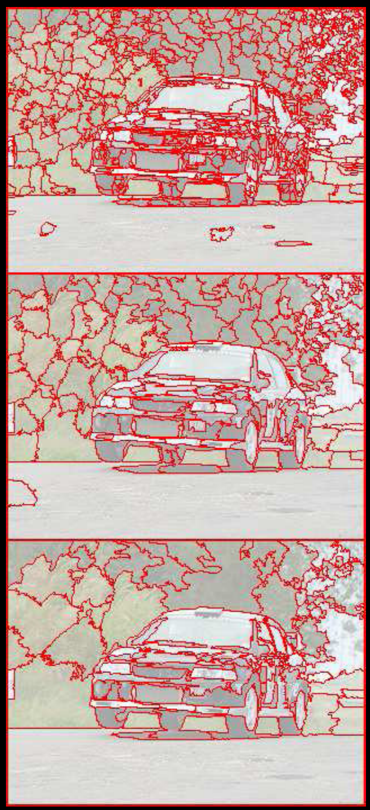
- *Stochastic grammars:* Zhu, S., Mumford, D.: A stochastic grammar of images. In: Found. and Trends. In Graph. and Vision (2006)

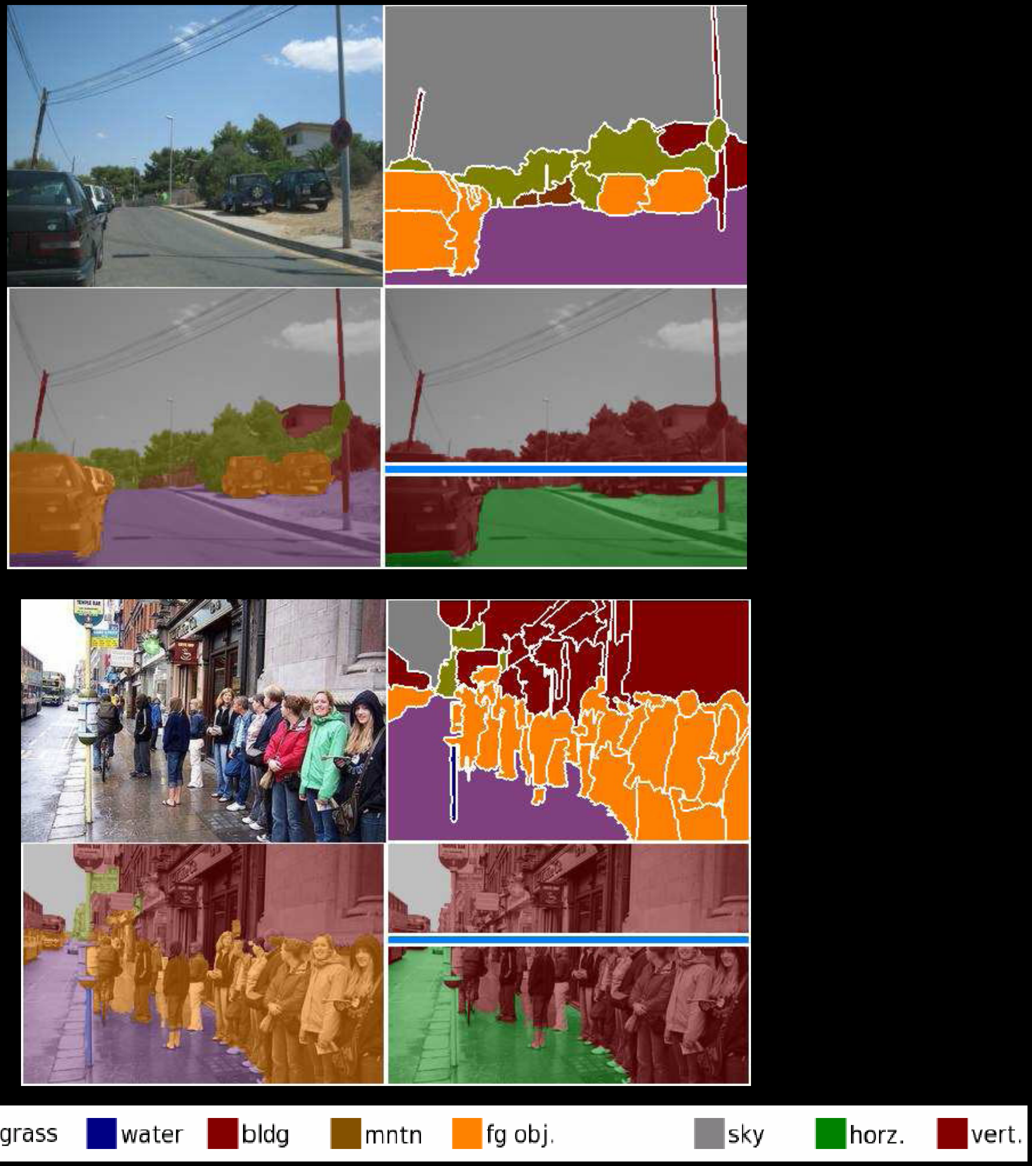




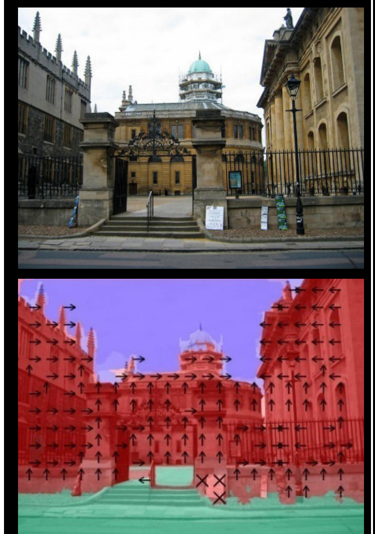
- *2D labeling:*

Gould, S., Fulton, R., Koller, D.: Decomposing a scene into geometric and semantically consistent regions. In: ICCV (2009)

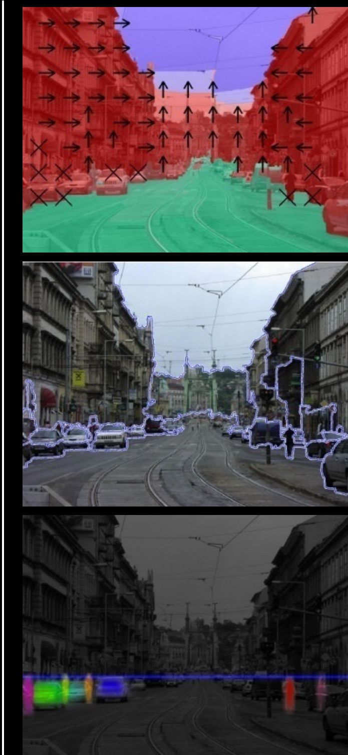




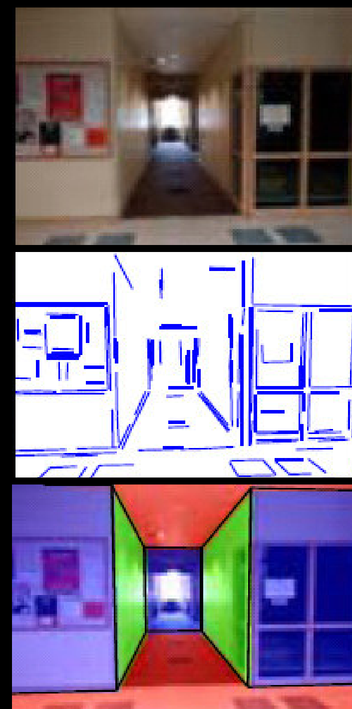
Levels of 3D-ness



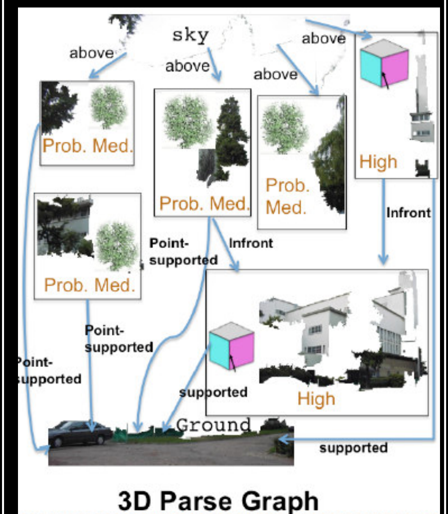
Region labels



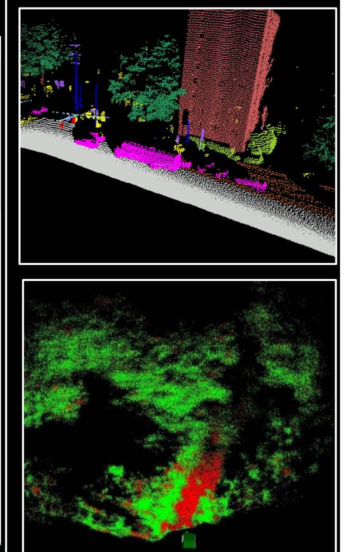
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



+ constraints from statics of solids



3D point clouds

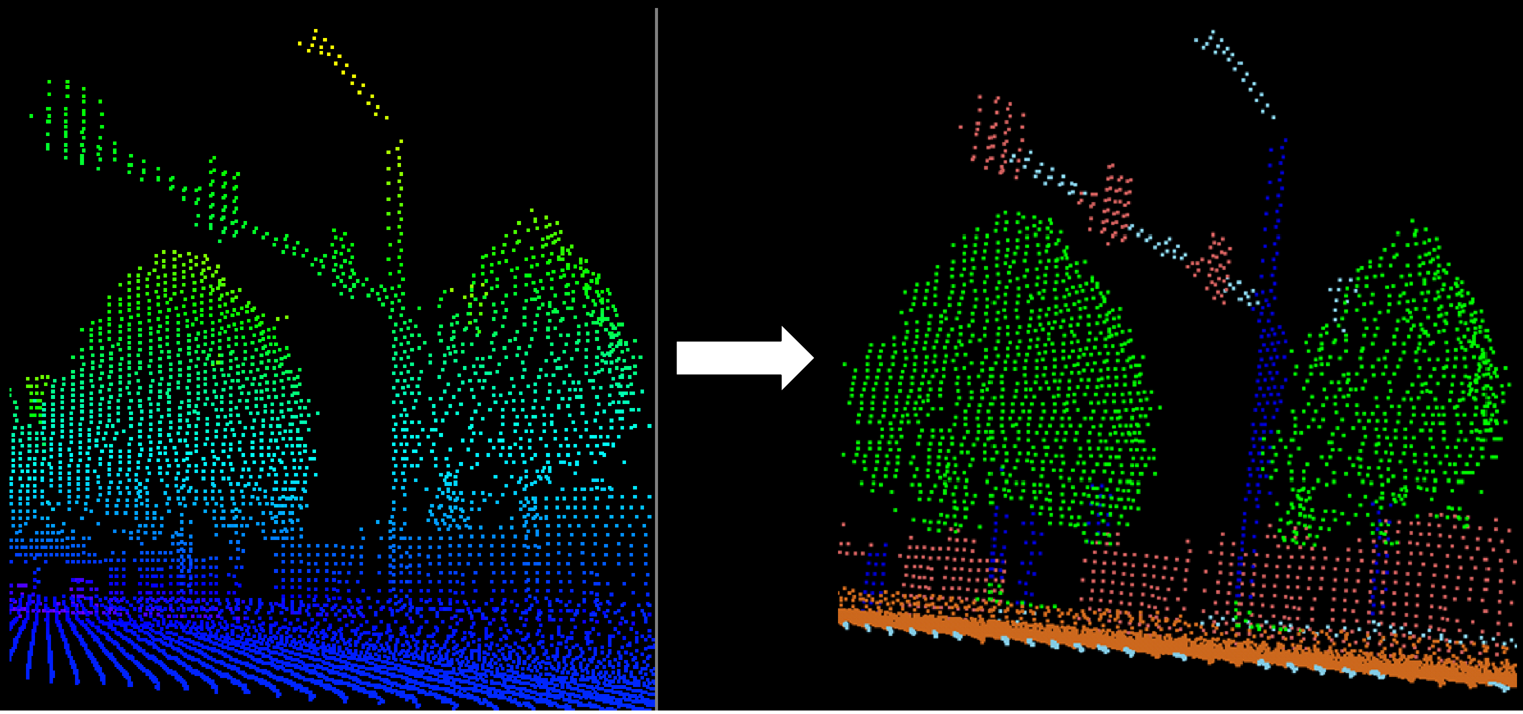
Qualitative

More quantitative
more precise

Explicit

An exercise in using explicit 3D data

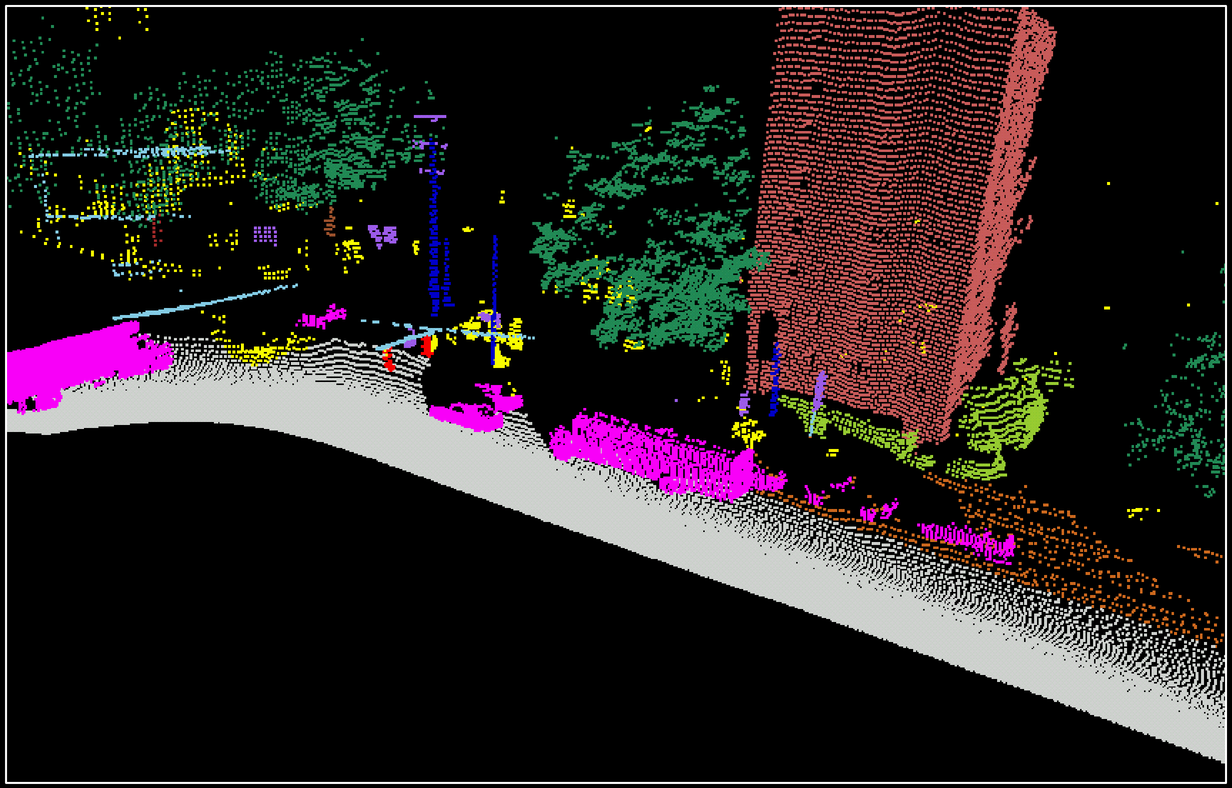
- What if we have explicit 3D data (from stereo, SFM, or other sensors)?
- How far can we go in scene interpretation from 3D data only?
- Can we adapt reasoning or classification tools from the image domain?



Motivation

- Specialized sensors available in robotics applications
- Cheap depth cameras
- Large-scale Structure From Motion (SFM) systems

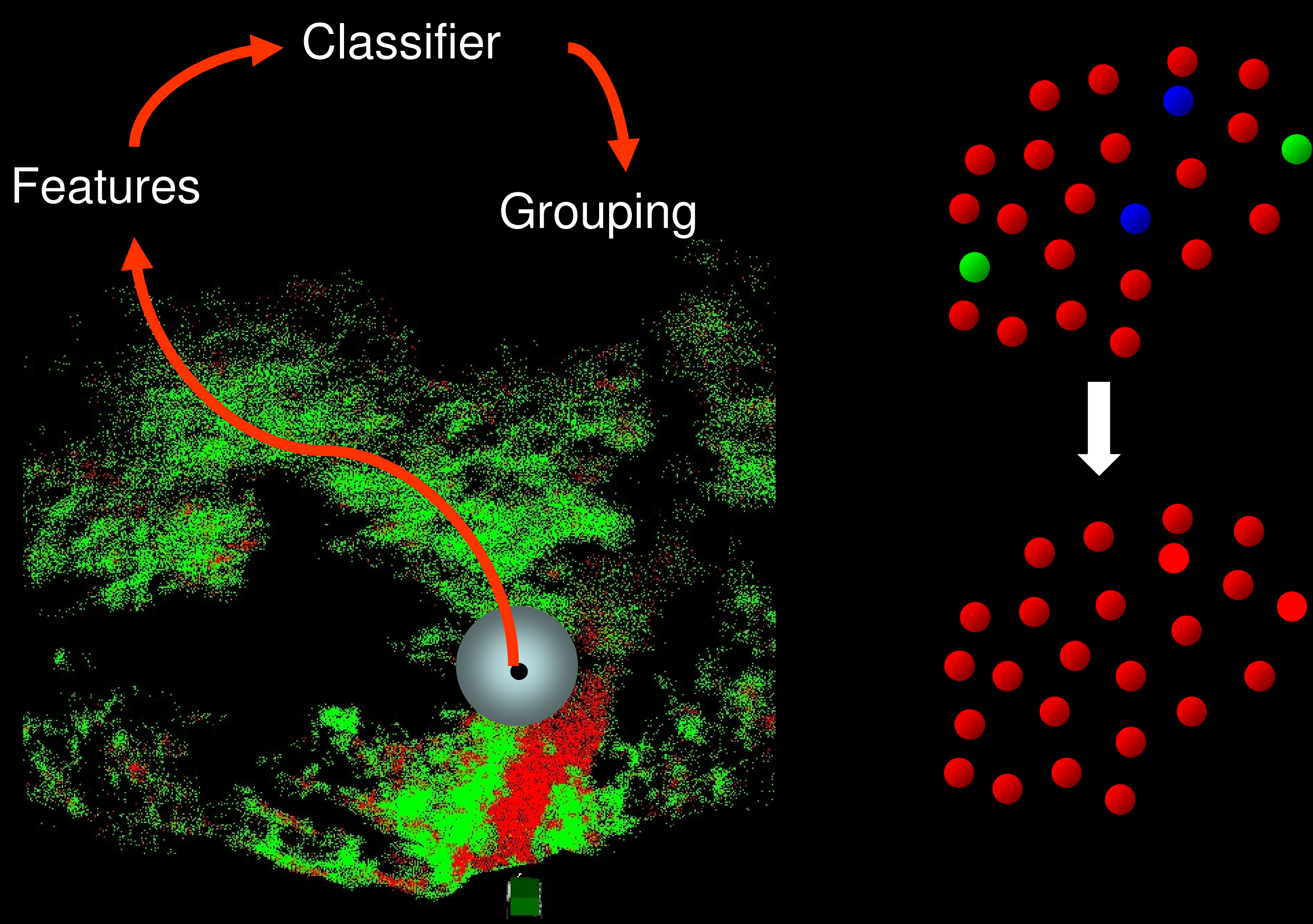
Snavely *et al.* ICCV'09



Ground truth labels

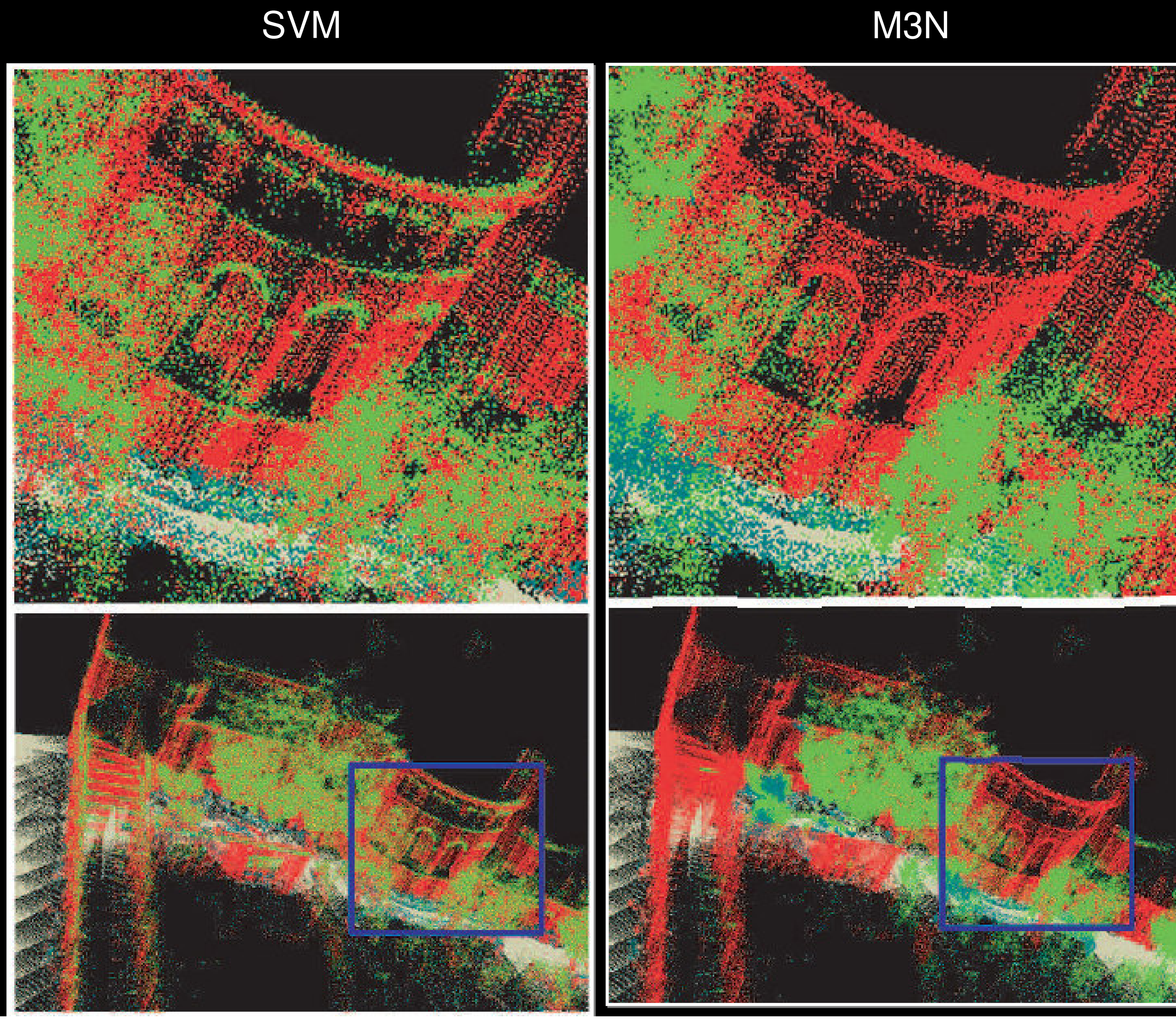
Wire
Load bearing
Pole/trunk
Vehicle
Shrub
Foliage
Facade
Paved road
Traffic lights
Cross-arm

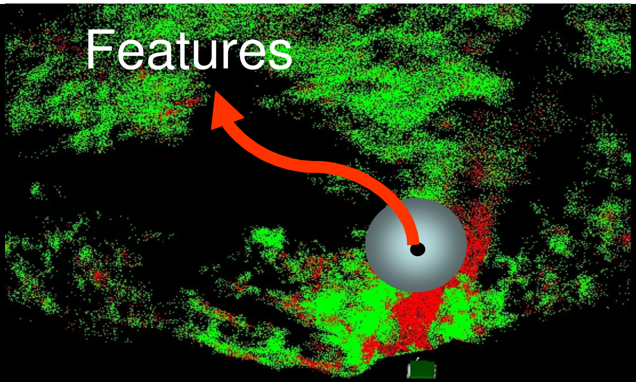
General approach



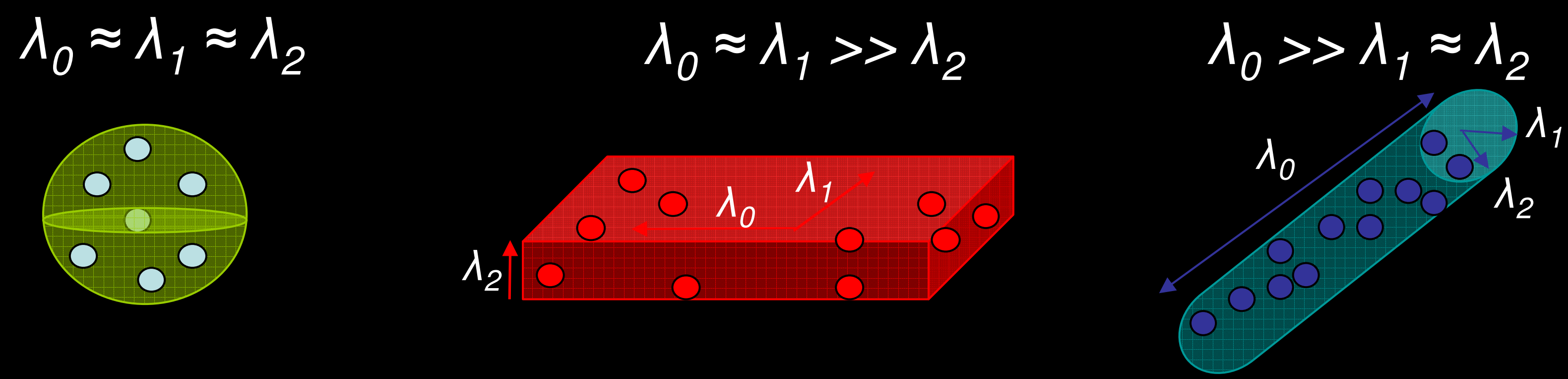
- What features and classifiers?
- What neighborhood and scale?
- What model for grouping and consistency?

Example [anguelov-cvpr-05, triebel-icra-06, triebel-ijcai-07]





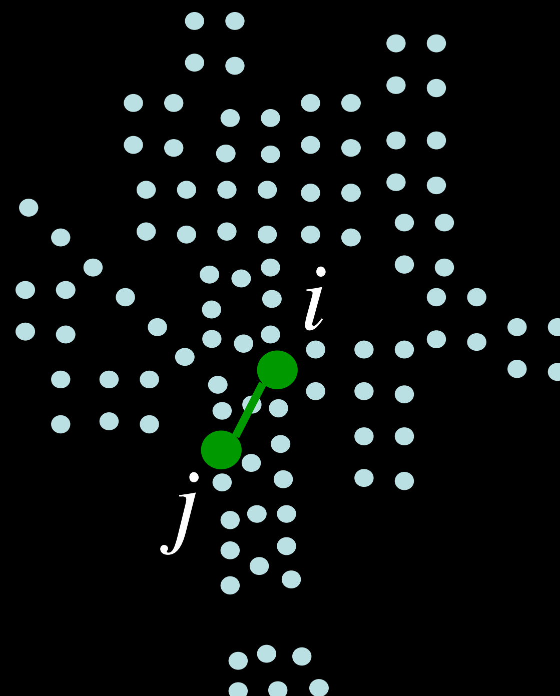
- Spectral features (inspired from tensor voting work)



$\sigma_{point} = \lambda_0$ $\quad$ $\sigma_{surface} = \lambda_1 - \lambda_2$ $\quad$ $\sigma_{linear} = \lambda_0 - \lambda_1$

- Directional features from tangent/normal
  - How to select scale: Unnikrishnan et al., "*Scale Selection for Geometric Fitting in Noisy Point Clouds*", IJCGA 2010.
  - How to deal with unstructured point clouds: Lalonde et al., "*Data Structures for Efficient Dynamic Processing in 3-D*", IJRR 2007.

Model



$$P(y \mid x) = \frac{1}{Z} \prod_i \varphi(y_i, x_i) \prod_{ij} \varphi(y_i, y_j, x_{ij})$$

Labels

Data

Features of node i

Features of relationship between i and j

Model

$$P(y \mid x) = \frac{1}{Z} \prod_i \varphi(y_i, x_i) \prod_{ij} \varphi(y_i, y_j, x_{ij})$$

Labels

Data

Features of node i

Features of relationship between i and j

AMN: Potentials favor all variables in the clique to take the same assignment of labels:

Learn w by maximizing $P(y|x)$ over training data

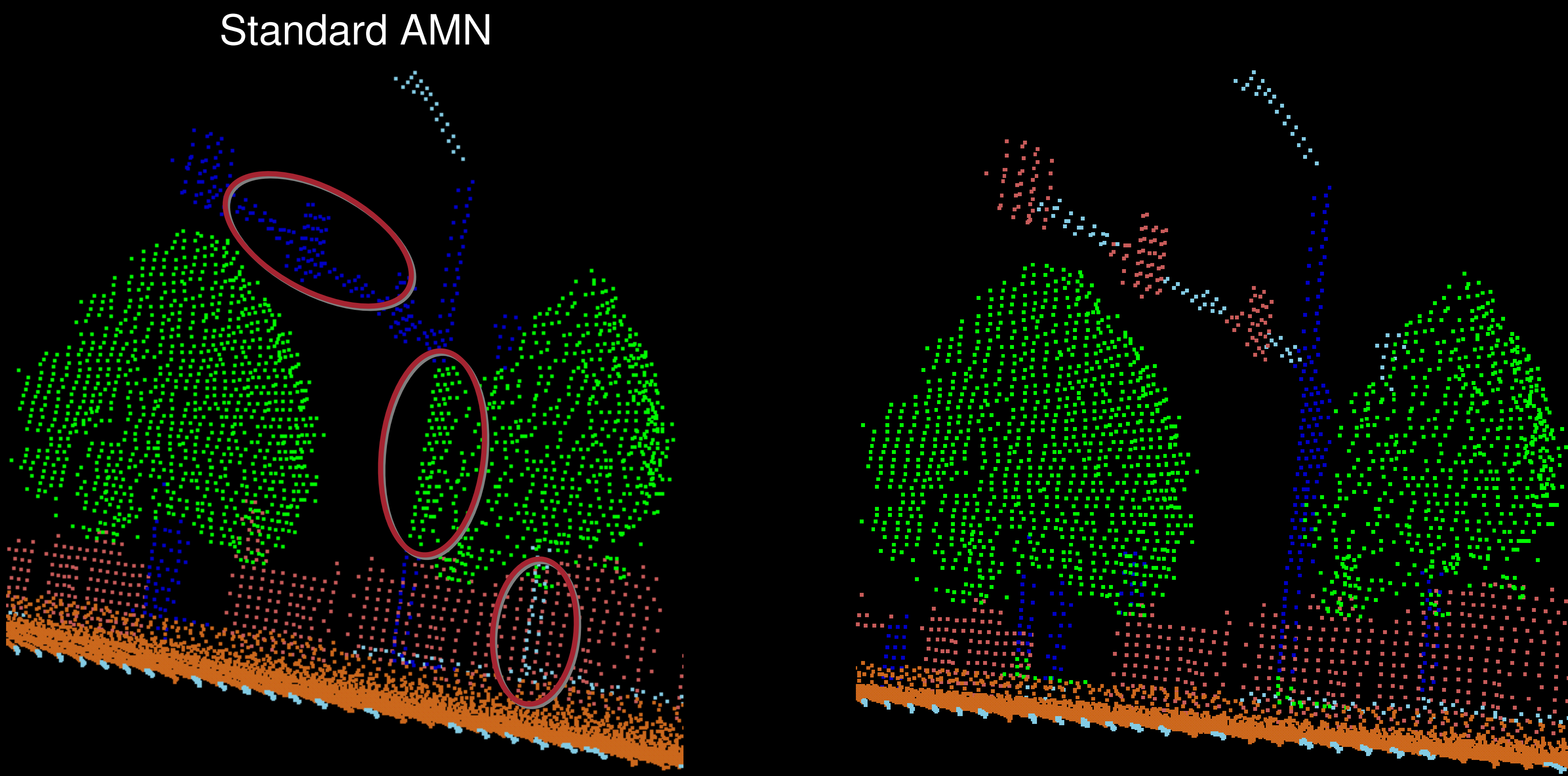
(Taskar'04)

$$\log \varphi_{ij}(k, l) = 0 \; k \neq l$$

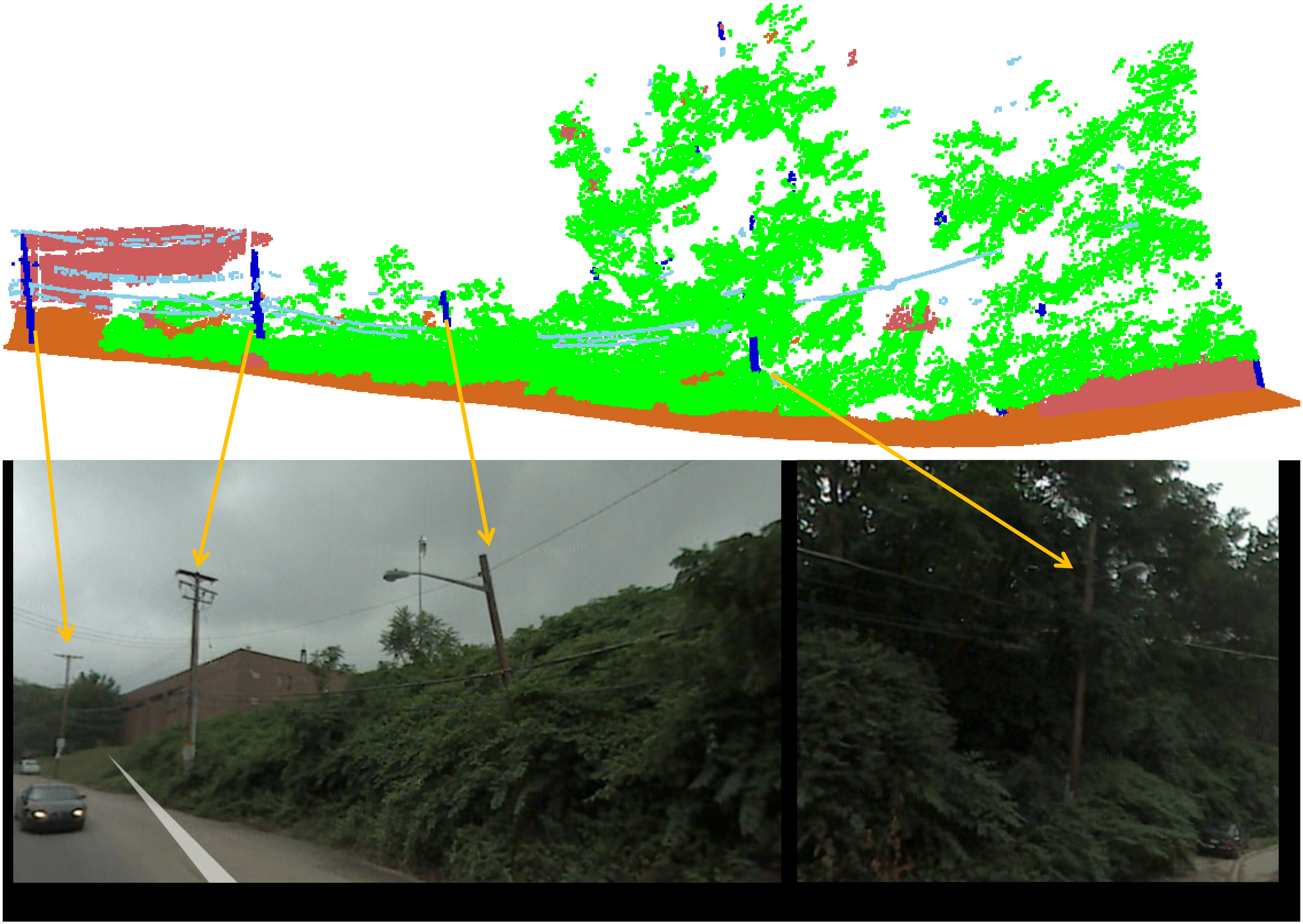
$$\log \varphi_{ij}(k, k) = w_e^{k,\theta} . x_{ij}$$

Directional model: Even more parameters to learn

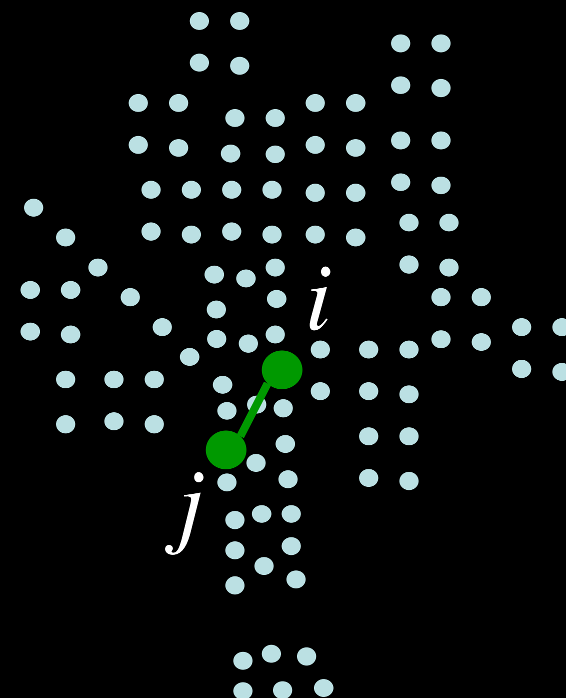
$$\log \varphi_i(k) = w_n^k . x_i$$



D. Munoz, N. Vandapel, and M. Hebert. Directional AMN for 3D point cloud classification. In 3DPVT, 2008.



Model



$$P(y \mid x) = \frac{1}{Z} \prod_{i} \varphi(y_i, x_i) \prod_{ij} \varphi(y_i, y_j, x_{ij})$$

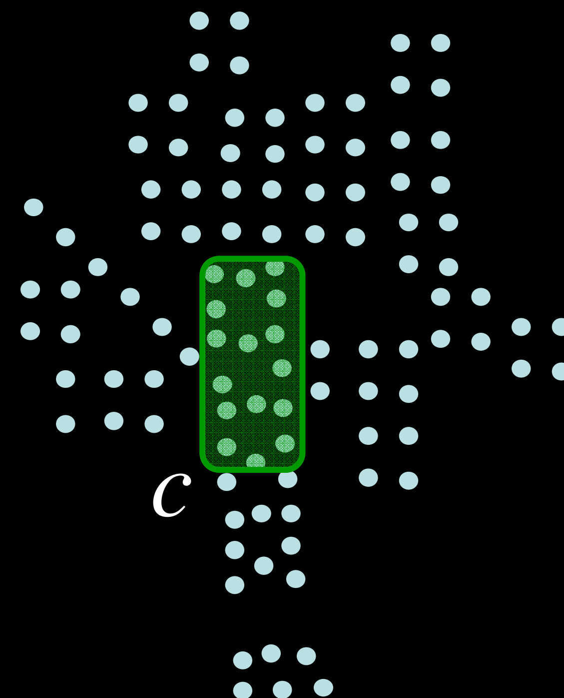
Labels

Data

Features of node i

Features of relationship between i and j

Larger support regions

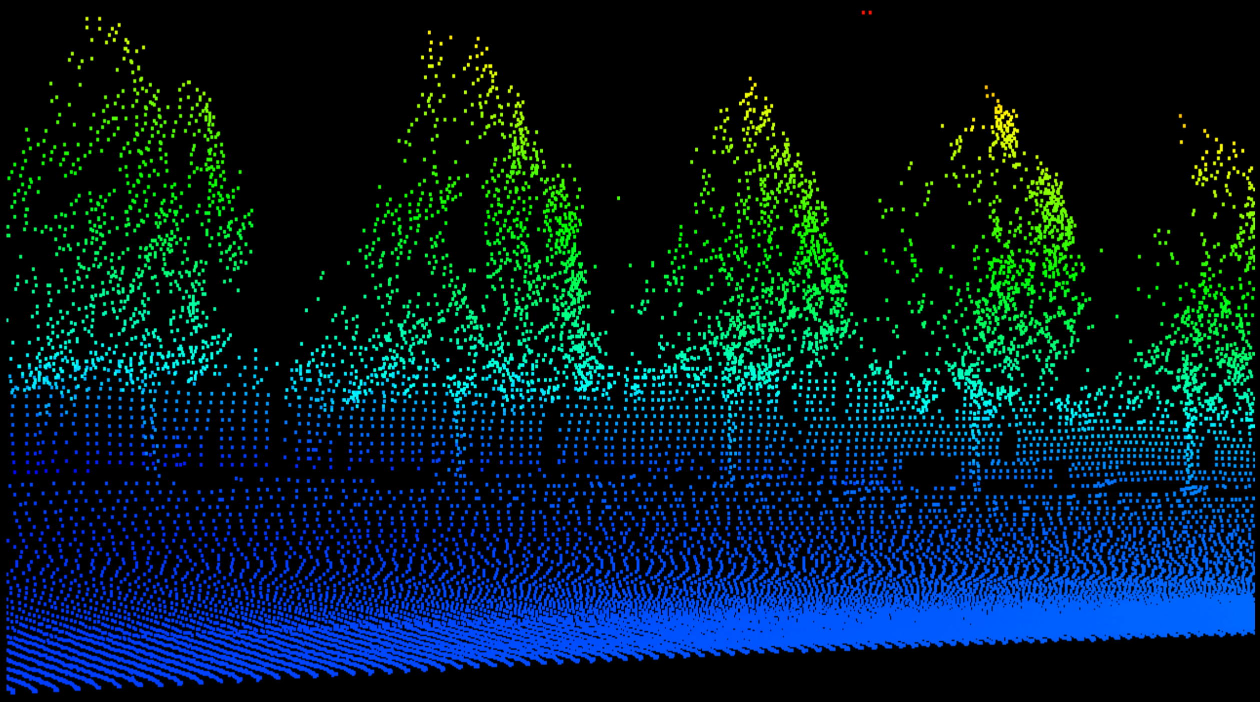


$$P(y \mid x) = \frac{1}{Z}\prod_i \varphi(y_i, x_i)\prod_{ij}\varphi(y_i, y_j, x_{ij})\prod_c \varphi(y_c, x)$$

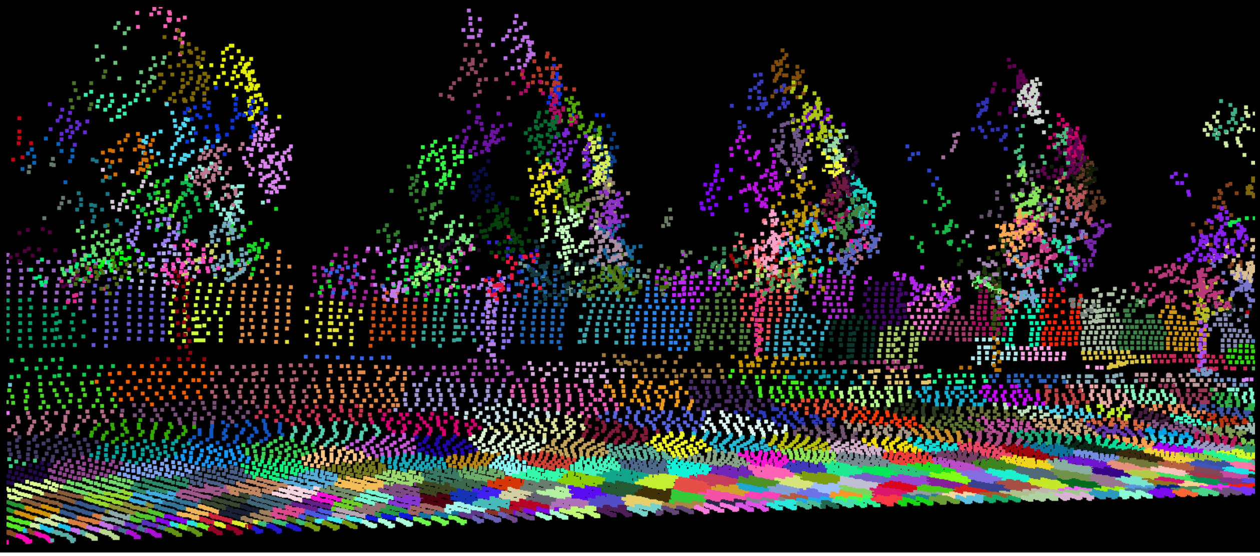
Inference possible with appropriate φ (P^n Potts model) [Kohli *et al.* 2007, 2008]

Larger support regions

Elevation coded point cloud



Regions from over-segmentation

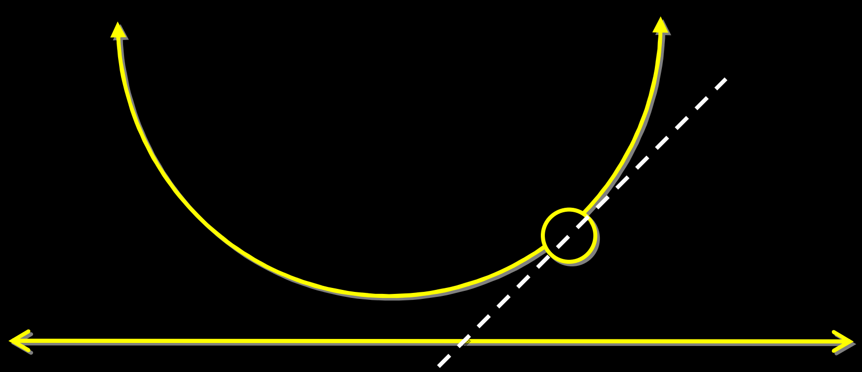


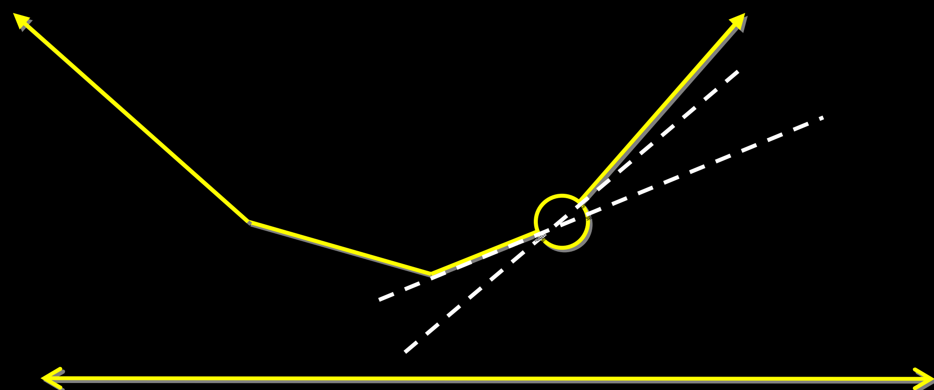
$$\min_{w}$$ **<u>Best</u> score over <u>all</u> labelings (+*margin*)** − **Score with <u>ground truth</u> labeling**

- *Convex program* [Taskar *et al*. ICML'04]

- *Subgradient* [N. Ratliff, J. Bagnell, and M. Zinkevich. Online subgradient methods for structured prediction. In AISTATS, 2007]

$$w_{t+1} \leftarrow w_t + \alpha g_w$$





$$\min_{w} \quad \boxed{\text{\underline{Best} score over \underline{all} labelings (+\textit{margin})}} \; - \; \boxed{\text{Score with \underline{ground truth} labeling}}$$

- *Convex program* [Taskar *et al*. ICML'04]
- *Subgradient* [Ratliff *et al*. AIStats'07]

$$w_{t+1} \leftarrow w_t + \alpha g_w$$

- *Functional subgradient* [N. Ratliff, D. Bradley, J. Bagnell, and J. Chestnutt. *Boosting structured prediction for imitation learning*. NIPS, 2007; D. Munoz, J. Bagnell, N. Vandapel, *Contextual Classification with Functional Max-Margin Markov Networks*. CVPR'09]

$$\varphi_{t+1} \leftarrow \varphi_t + \alpha_t h_t$$

h_t trained to:

increase the score of correctly classified nodes

decrease the score of incorrectly classified nodes

Efficient + enables more general potential

$\min_w$ **<u>Best</u> score over <u>all</u> labelings (+*margin*)** $-$ **Score with <u>ground truth</u> labeling**

- *Convex program* [Taskar *et al*. ICML'04]
- *Subgradient* [Ratliff *et al*. AIStats'07]

$$w_{t+1} \leftarrow w_t + \alpha g_w$$

- *Functional subgradient* [Ratliff *et al*. NIPS'07, Munoz *et al.* CVPR'09]

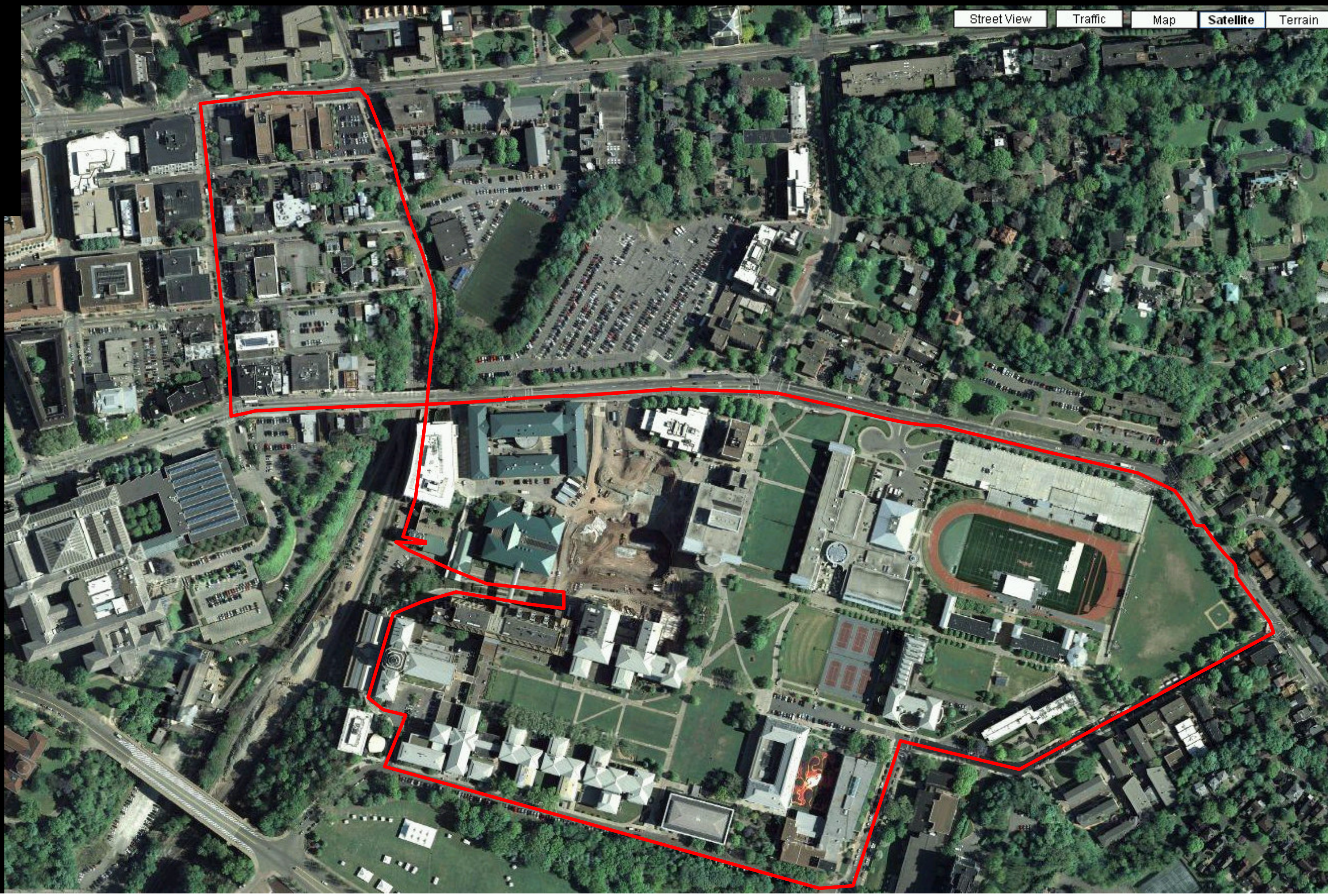
h_t trained to: $\varphi_{t+1} \leftarrow \varphi_t + \alpha_t h_t$

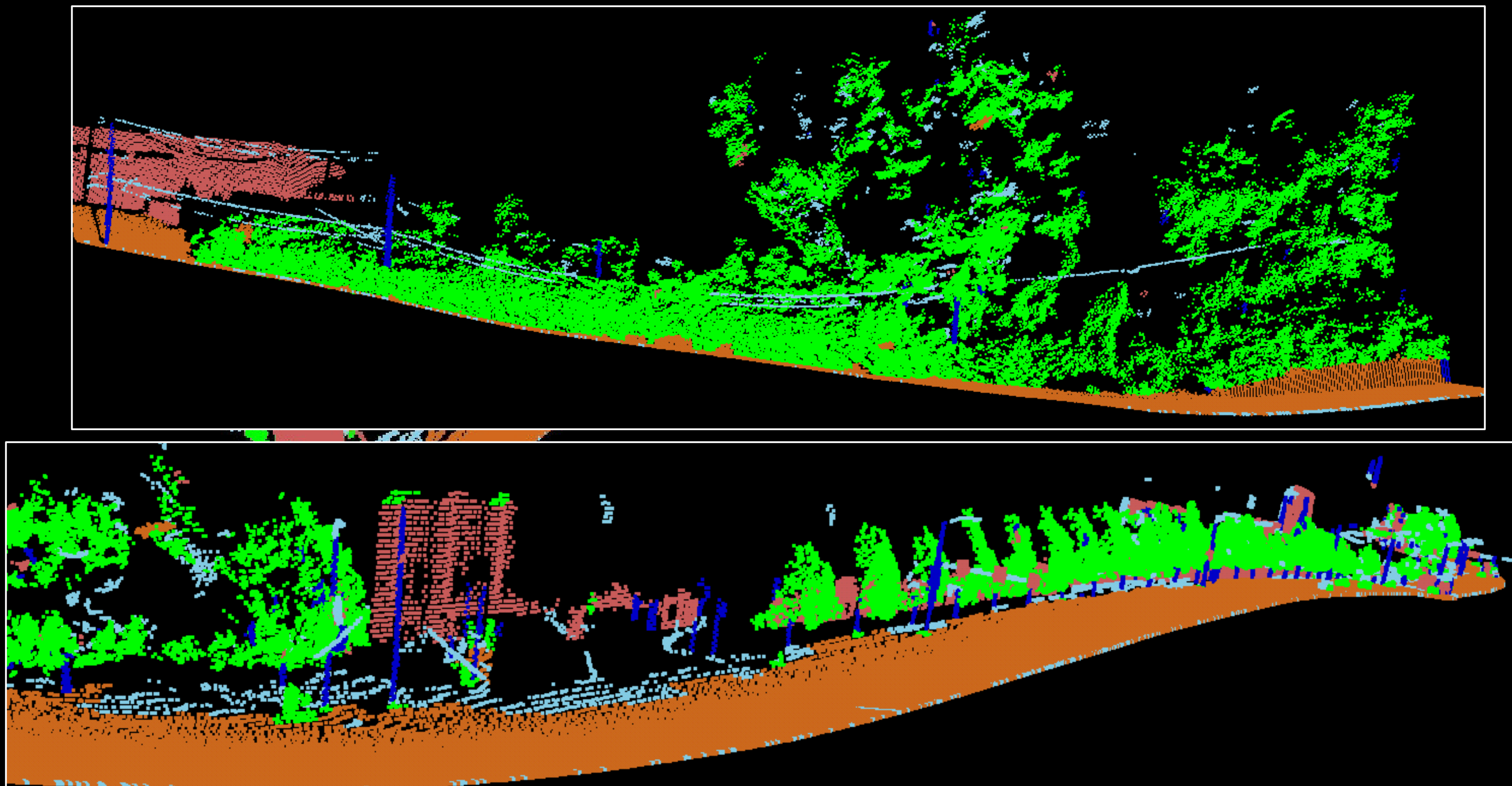
increase the score of correctly classified nodes

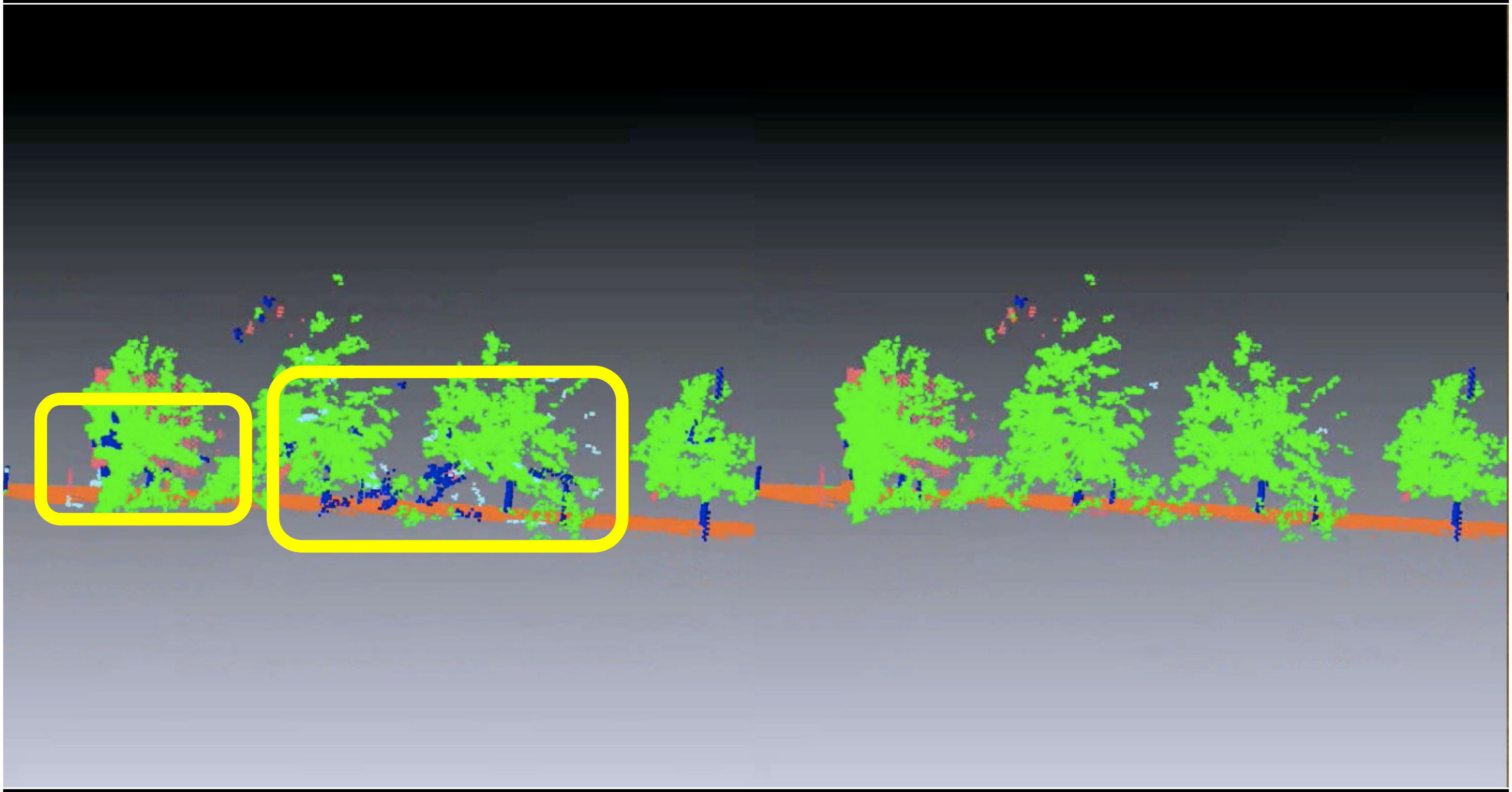
decrease the score of incorrectly classified nodes

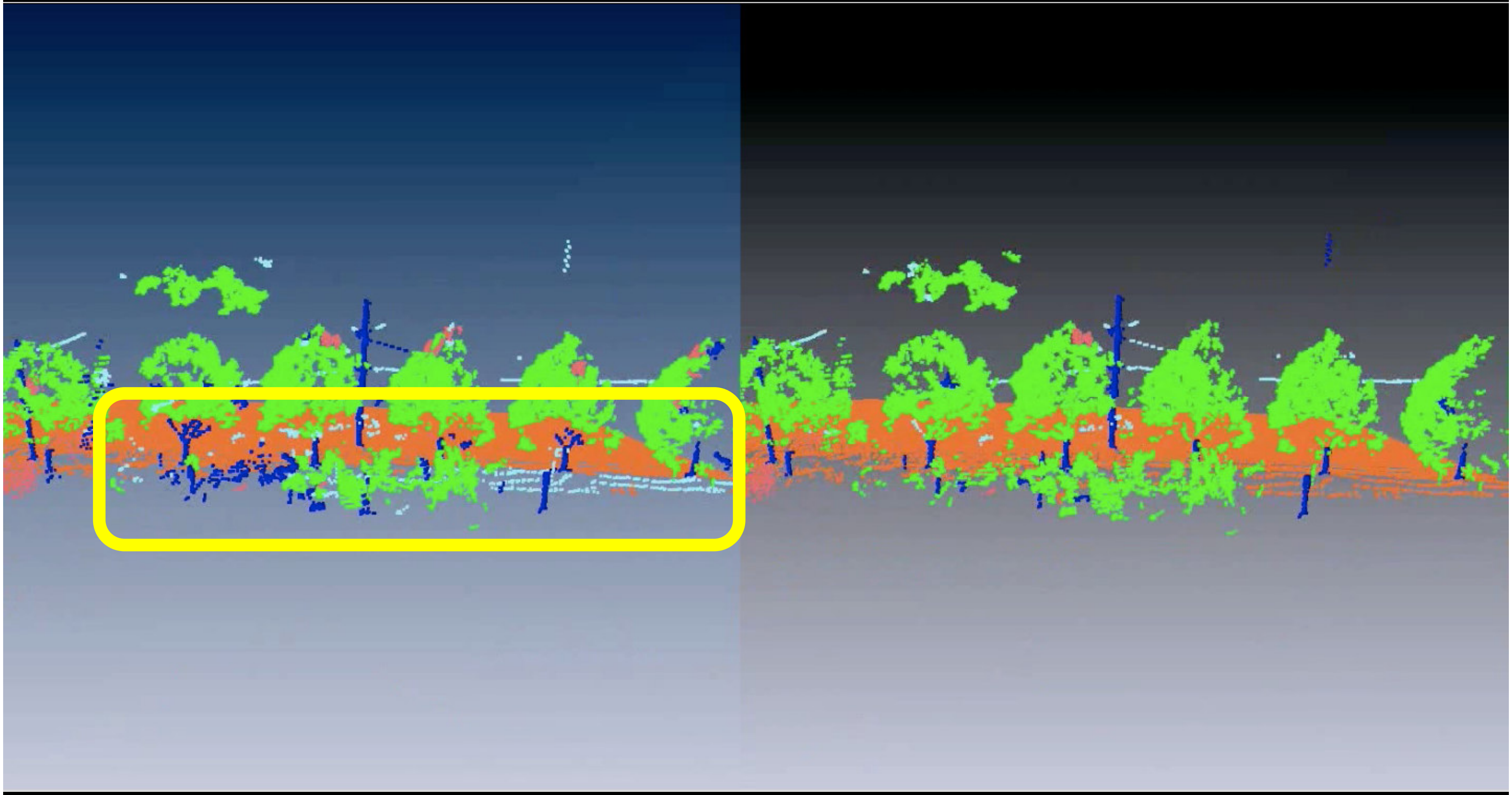
Efficient + enables more general potential

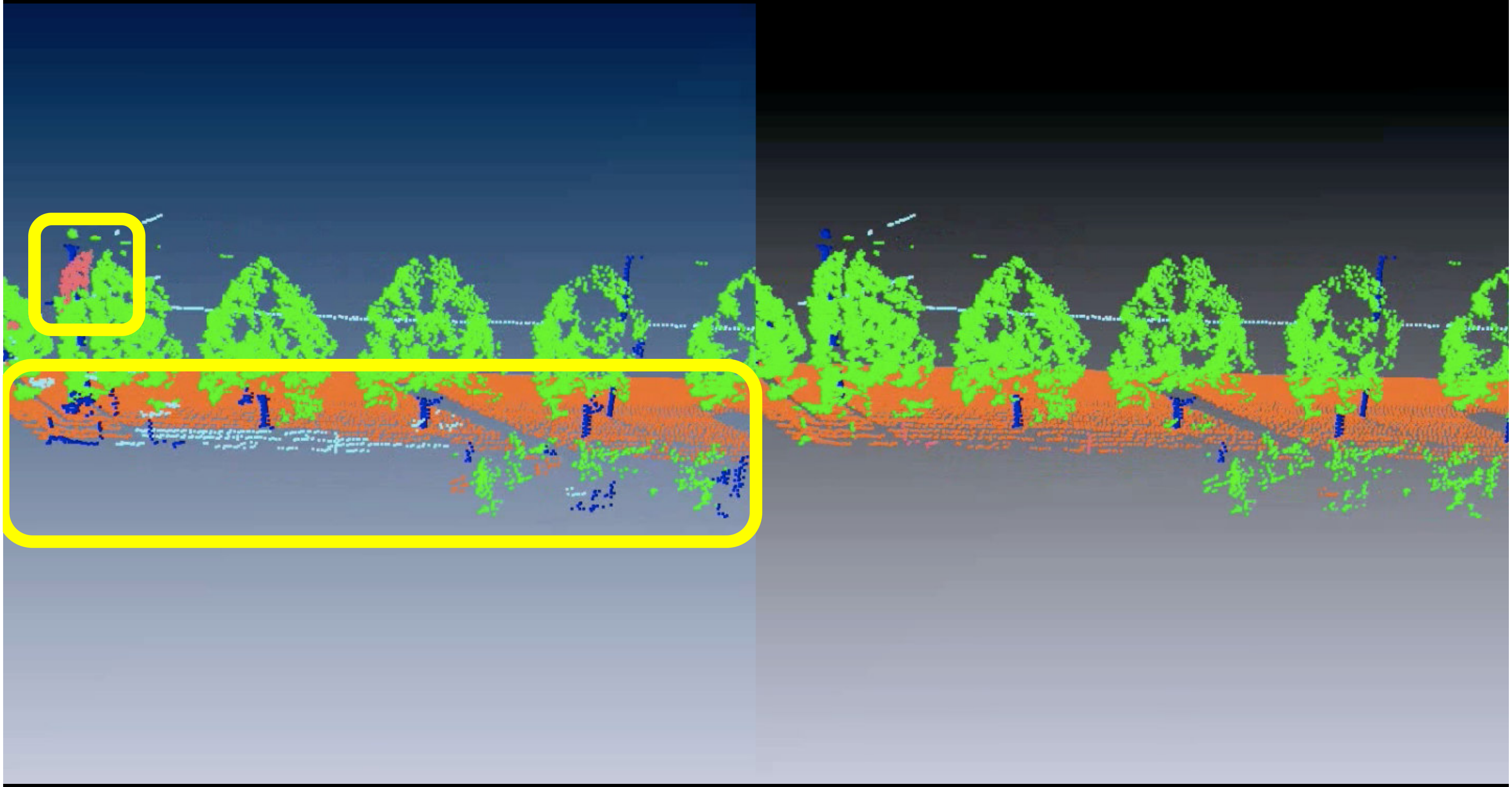
Gradient Tree Boosting for CRFs [Dietterich *et al.* 2004]; Boosted Random Fields [Torralba*et al.* 2004]; Virtual Evidence Boosting for CRFs [Liao *et al.* 2007]


Street View
Traffic
Map
Satellite
Terrain









Key issues

- Unstructured geometric data
- Incremental processing
- Efficient, online computation
- Alternate learning/inference models
- Un/Semi-supervised learning
- Online learning and adaptation
- Data fusion

- Aloysha Efros
- Derek Hoiem

- David Lee

- Abhinav Gupta

- Drew Bagnell
- Daniel Munoz
- Nicolas Vandapel