Introduction

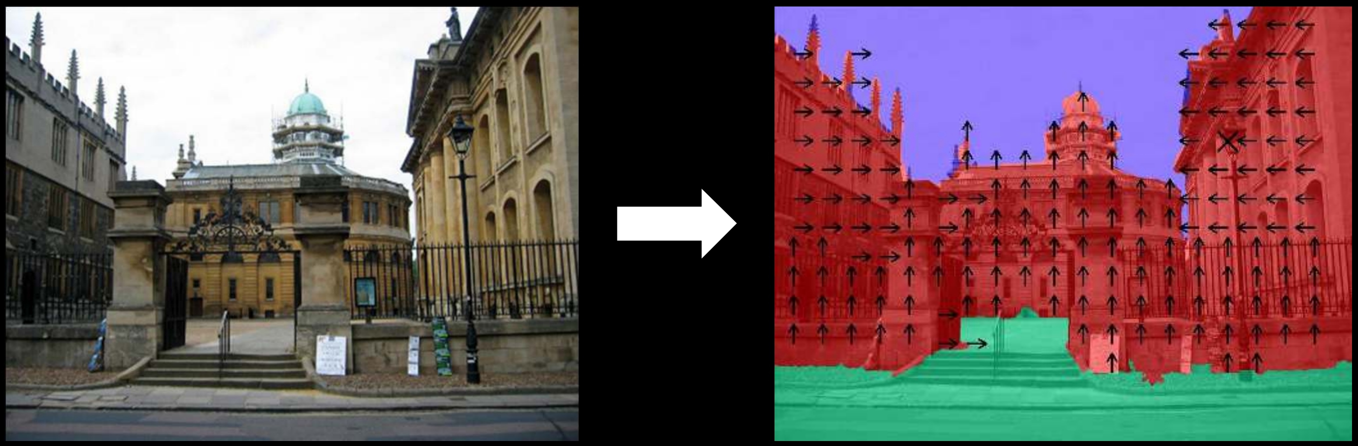
- What are we trying to achieve?
- Why are we doing this?
- What do we learn from past history?
- What will we talk about today?

What are we trying to achieve?

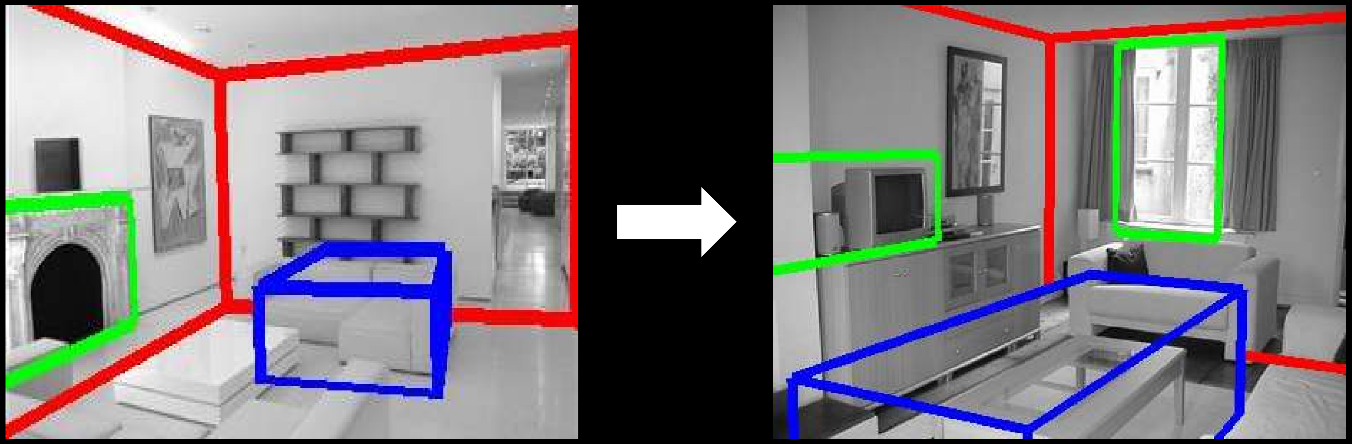
Example from
Scott Satkin

3D interpretation from image

Geometric labels

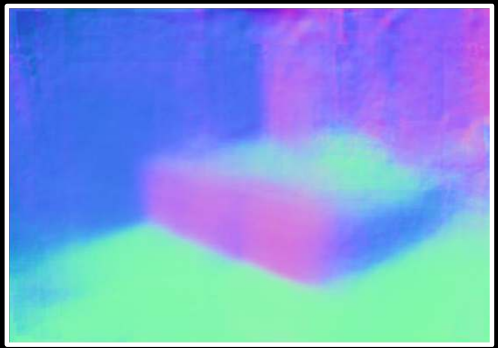


Volumetric layout

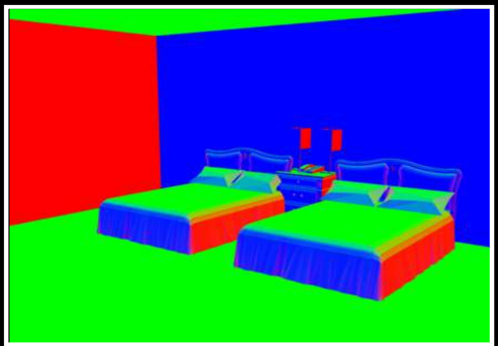


Sparse primitives

Dense reconstruction



3D scene model

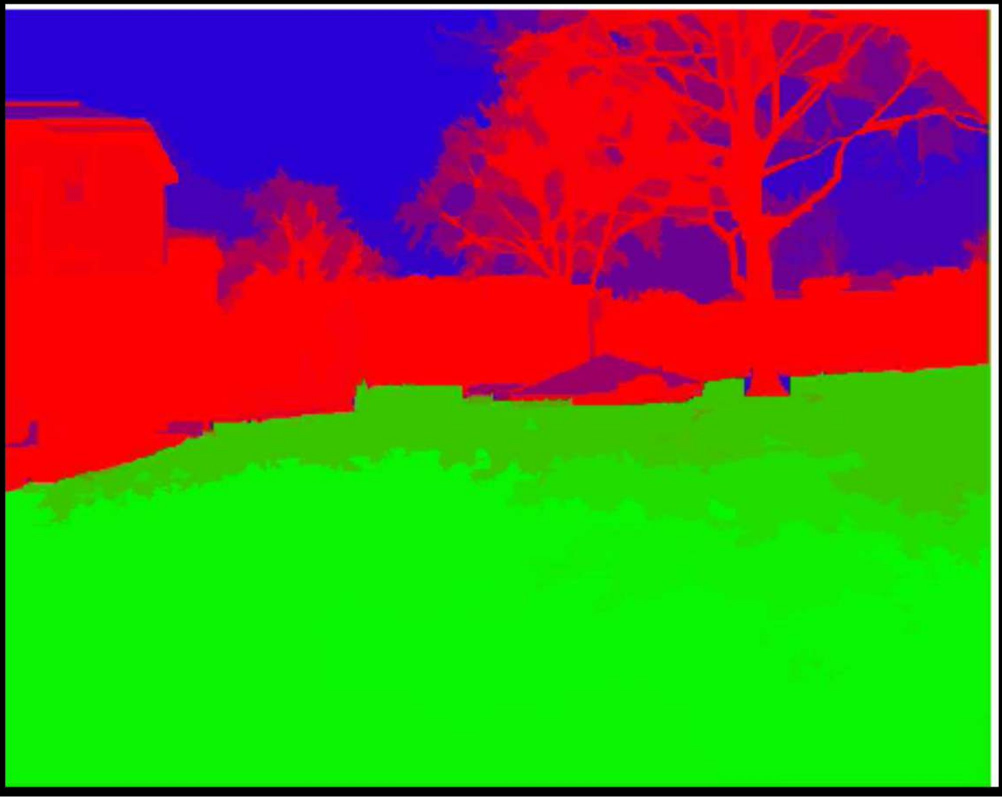


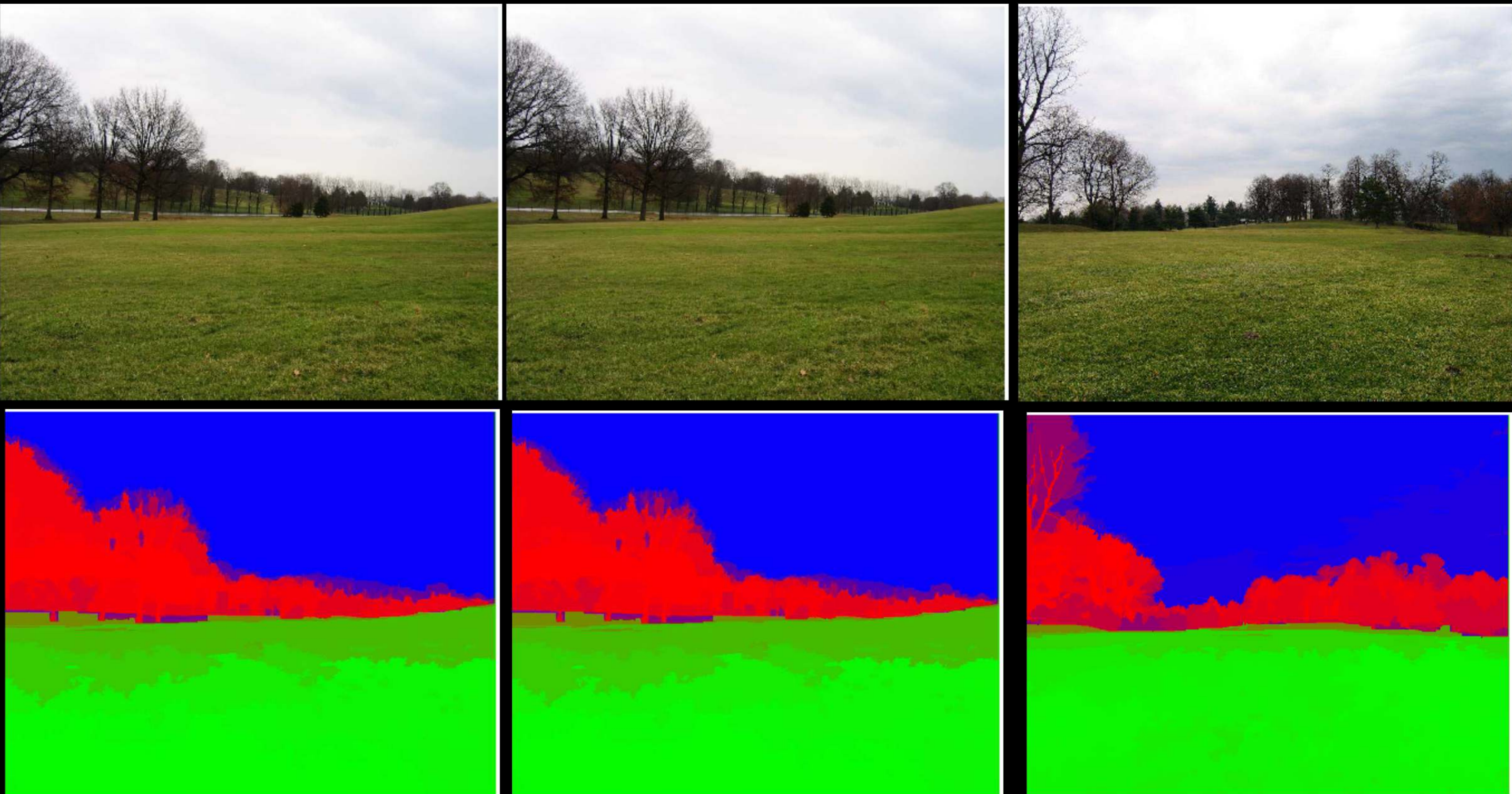
Why are we doing this?

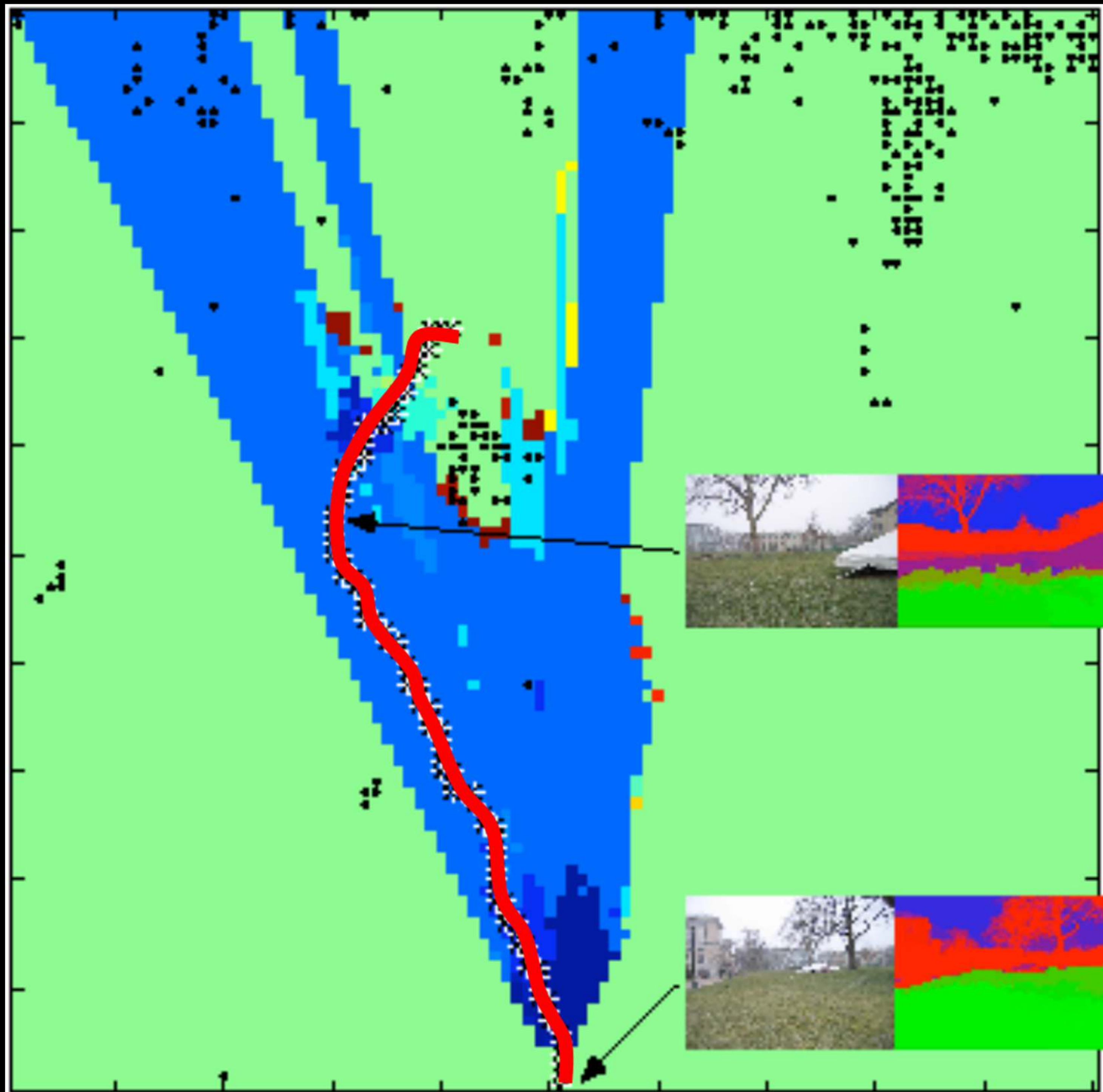
Applications

3D for motion

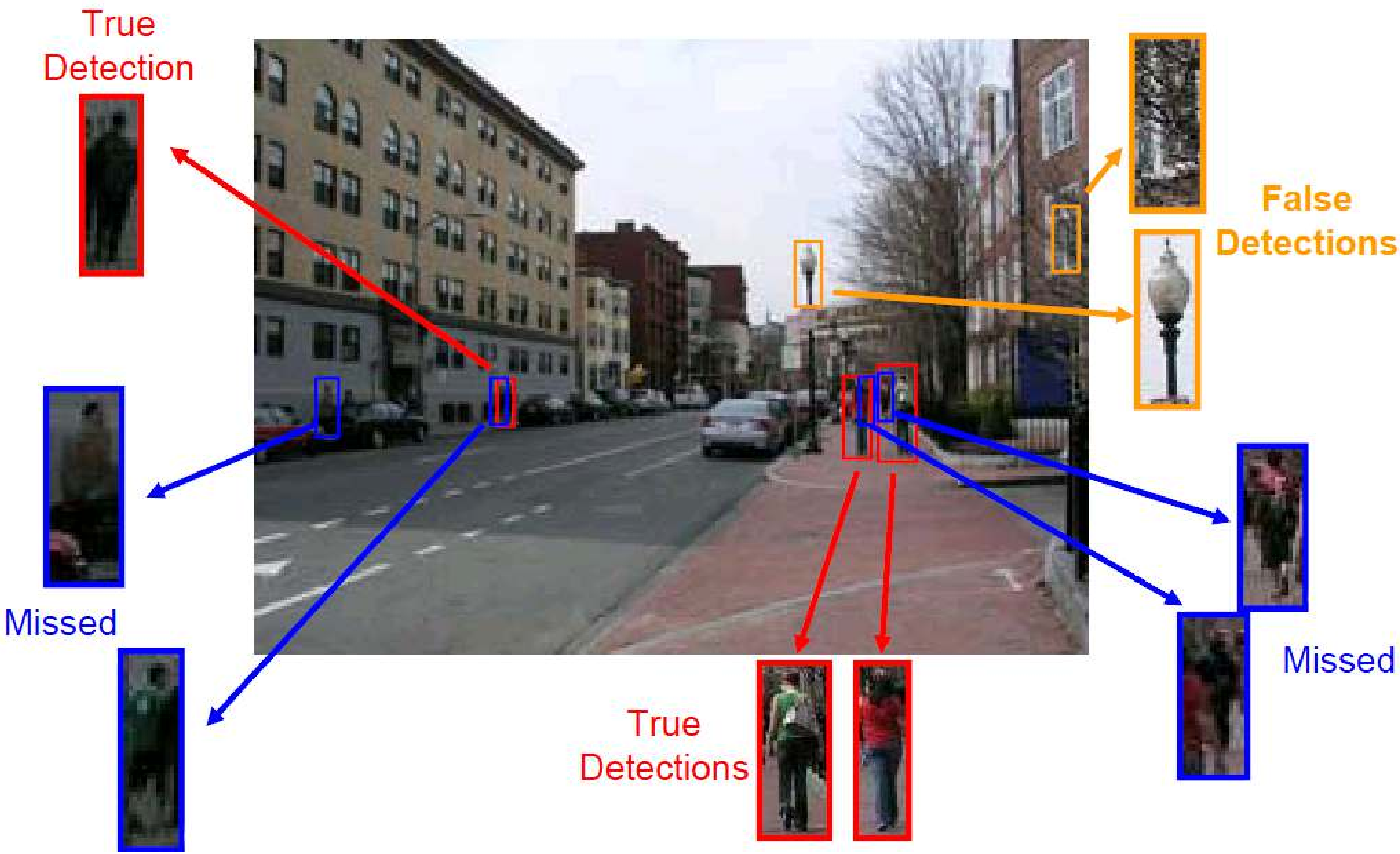
…when there is no direct way to get 3D

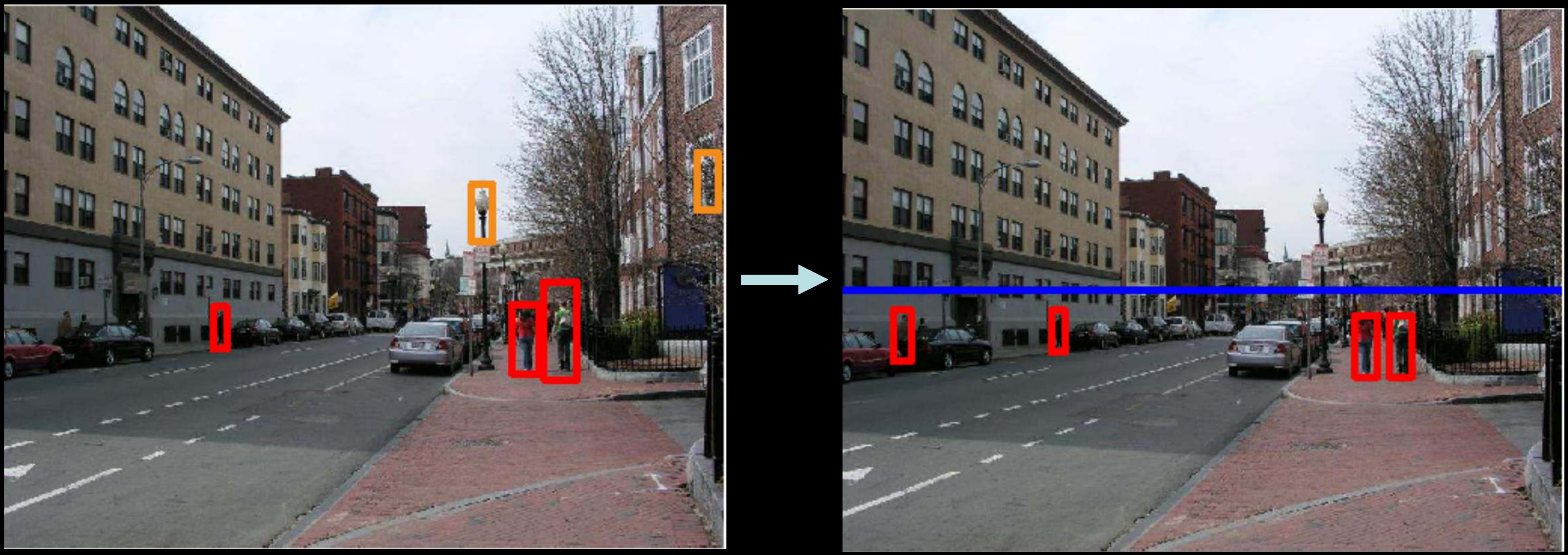




Informing detectors



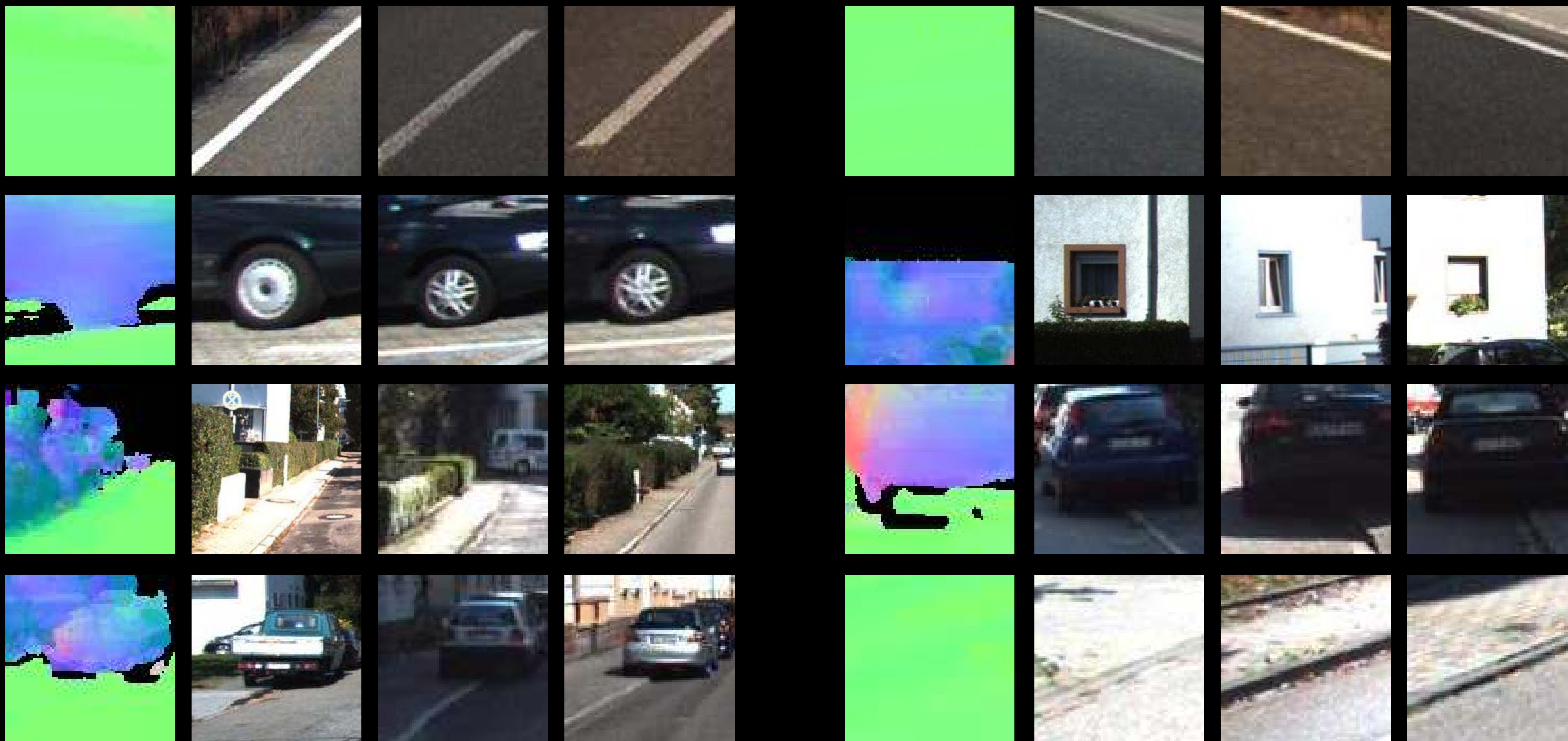
True Detection
False Detections
Missed
True Detections
Missed



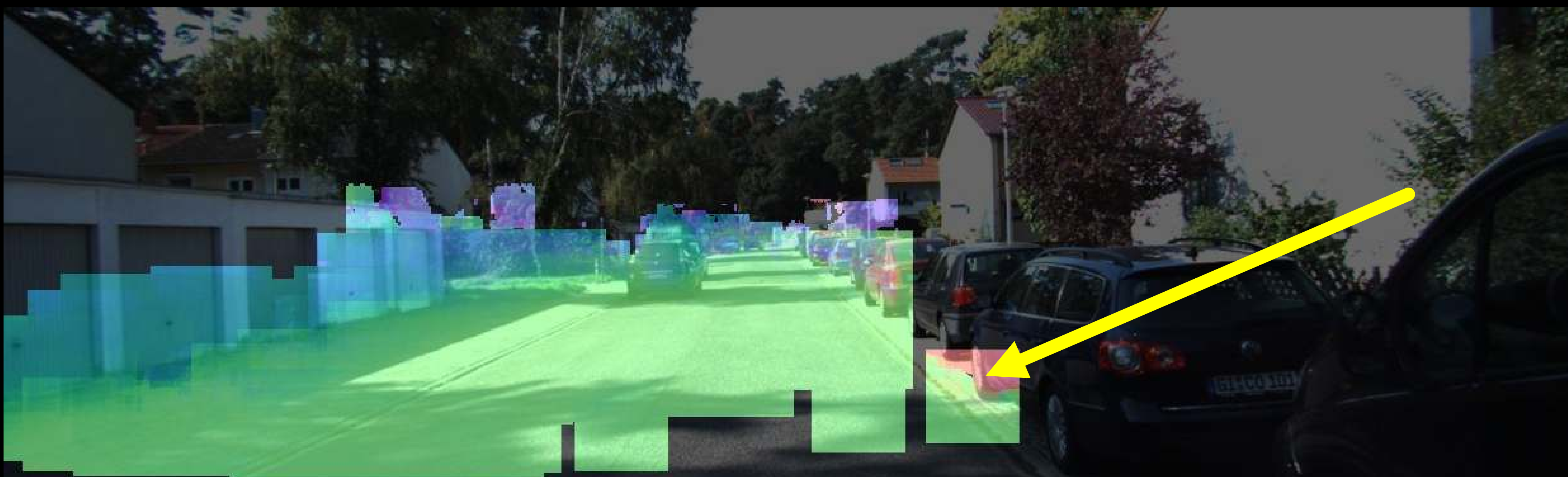
D. Hoiem, A. A. Efros, and M. Hebert. *Putting Objects in Perspective*. International Journal of Computer Vision, Vol. 80, No. 1, October, 2008.


KA HG 210
KA SW 191

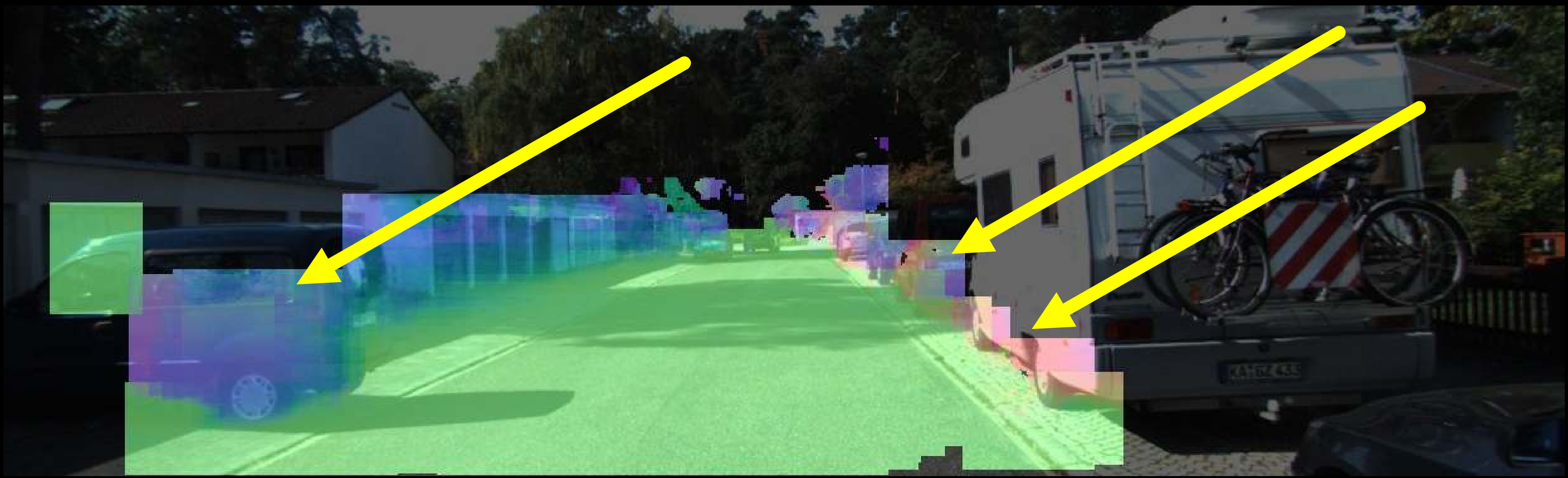
Learned Primitives (Examples)



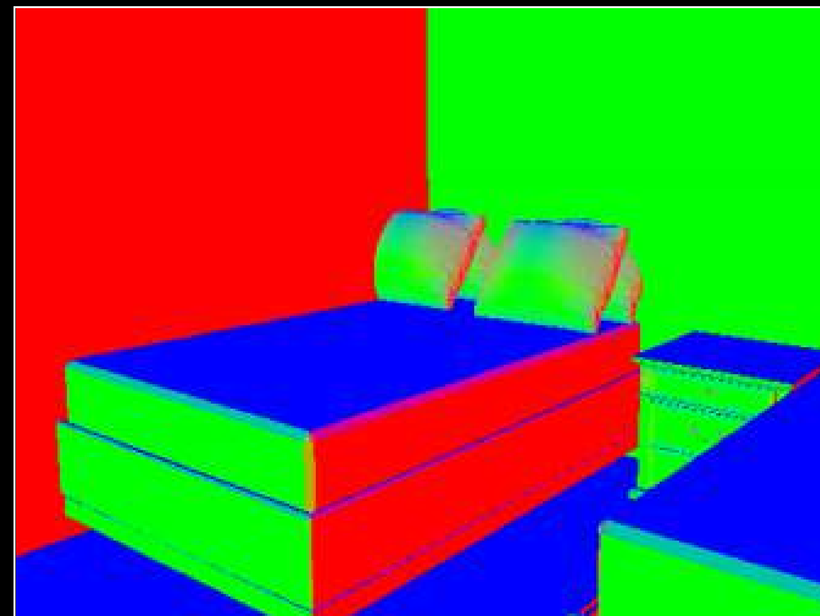
Contact points

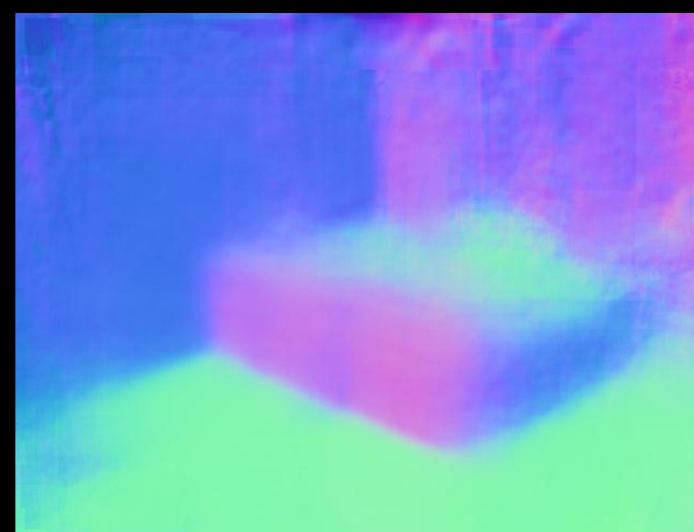


Object surfaces + Contact points



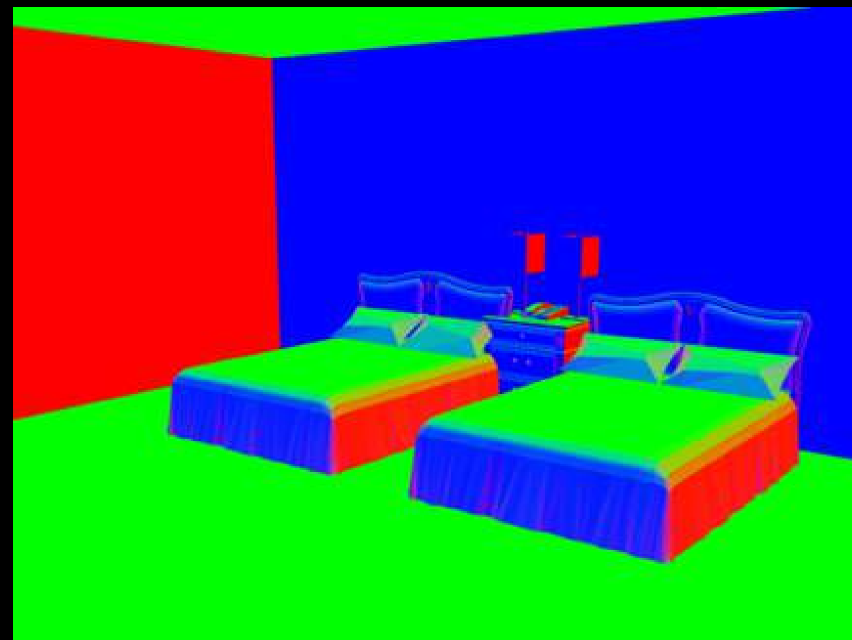
Editing images



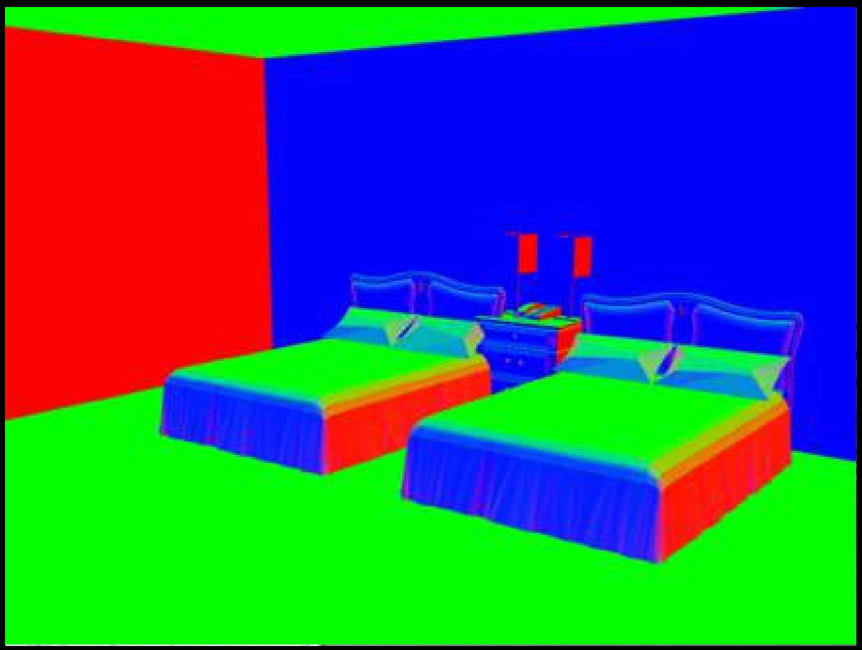
Predicting actions

Input Image



Estimated Geometry

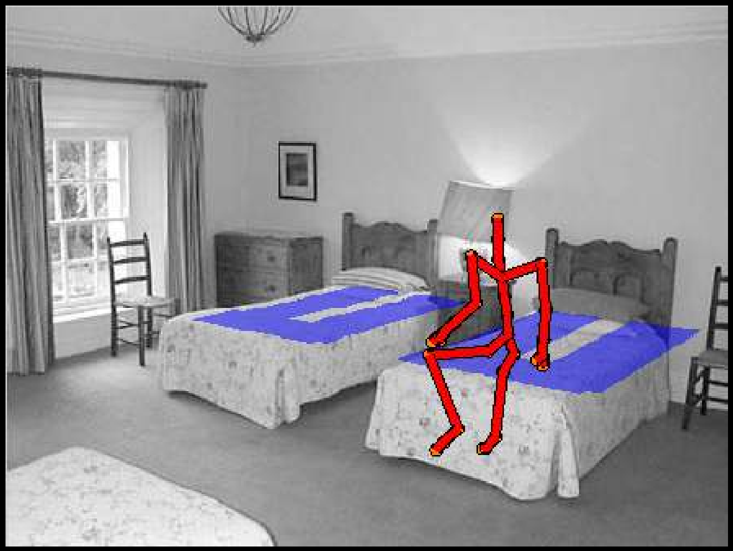
Application: Affordance Estimation



Estimated Geometry

Predicted Sitting Locations

Application: Affordance Estimation



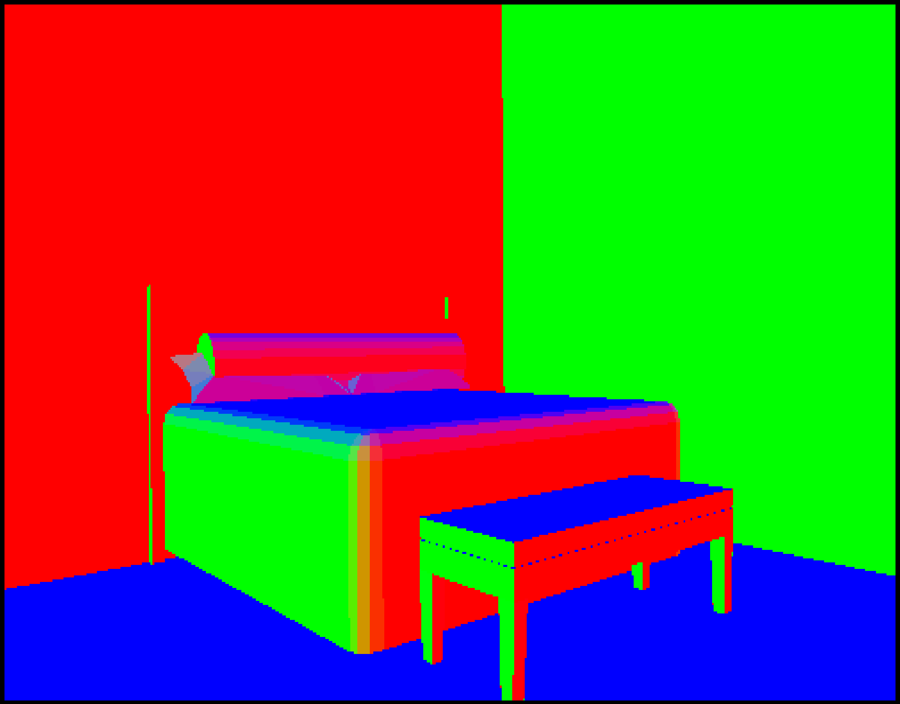
Sitting Upright

Sitting Reclined

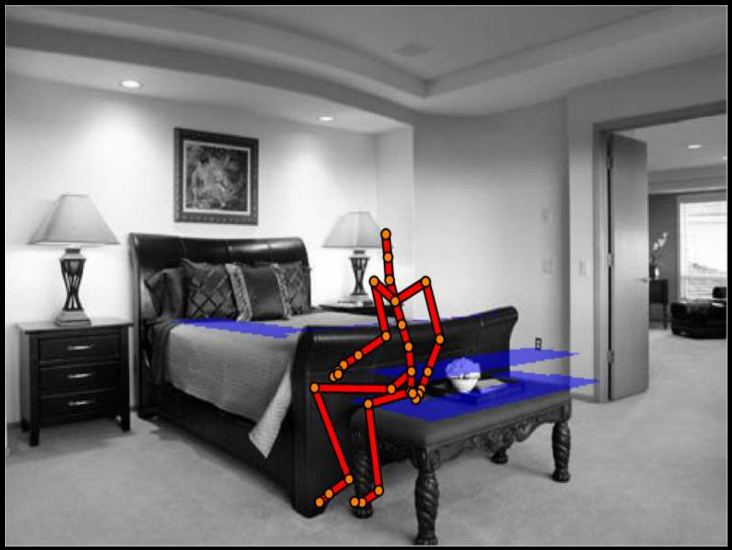
Laying Down

Reaching (4 poses)

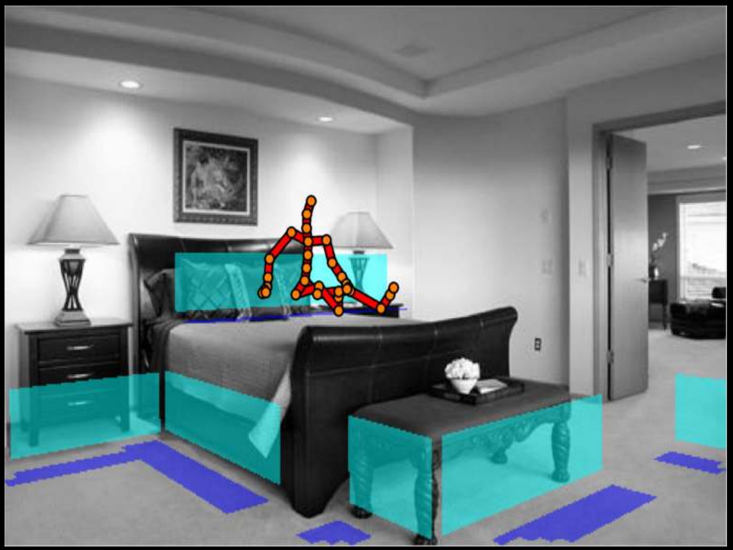
Application: Affordance Estimation



Application: Affordance Estimation



Sitting Upright



Sitting Reclined

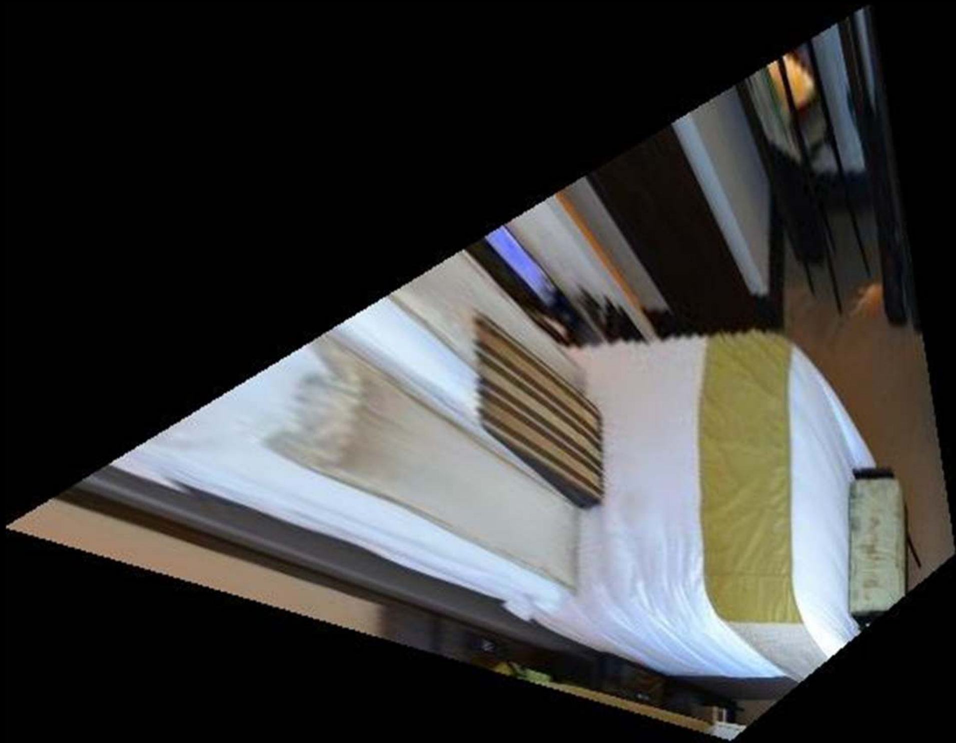
Laying Down

Reaching (4 poses)

Separating style and structure

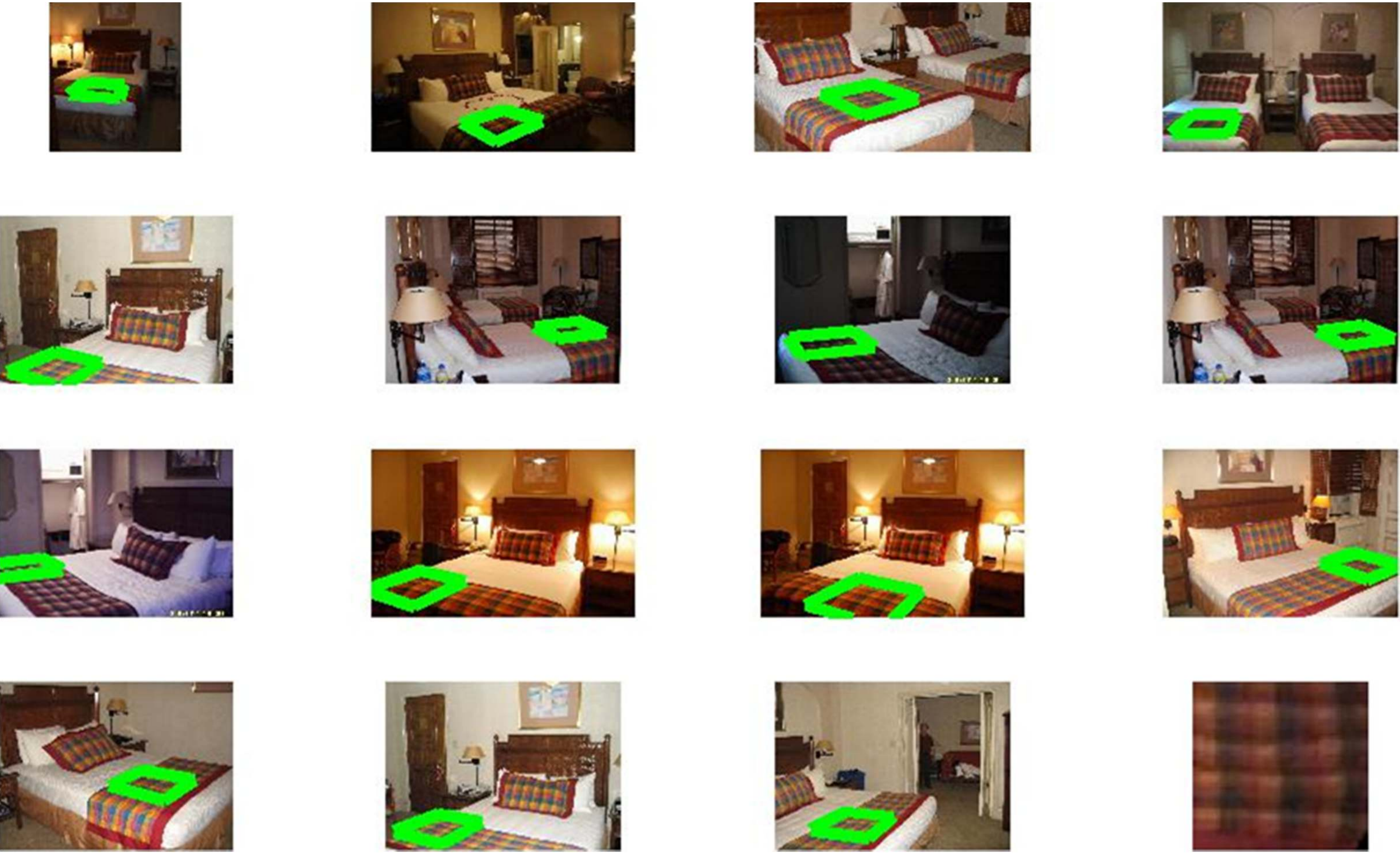


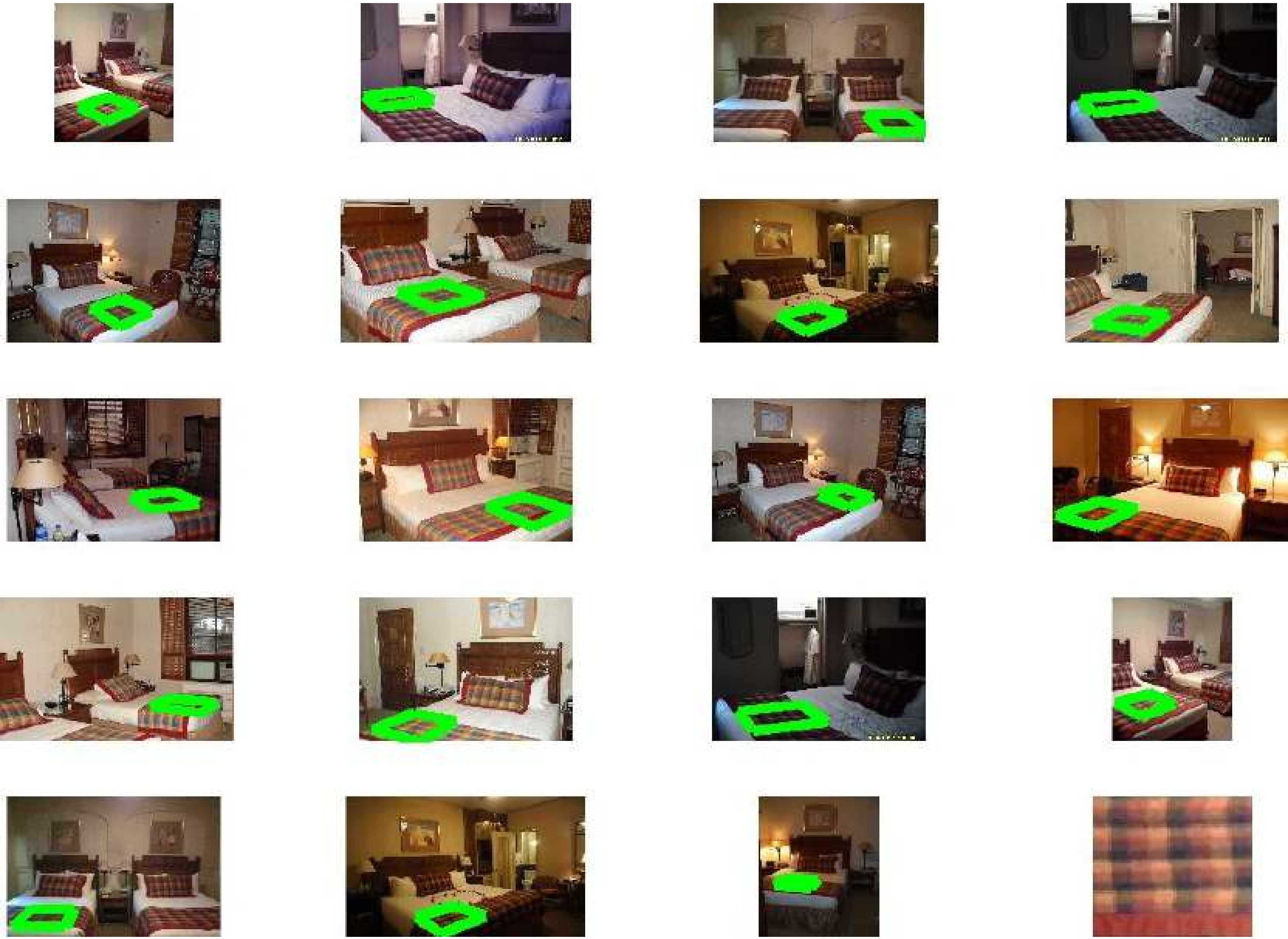




Tenenbaum & Freeman. Separating Style and Content with Bilinear Models. Neural Computation. 2000.

Casablanca Hotel, New York

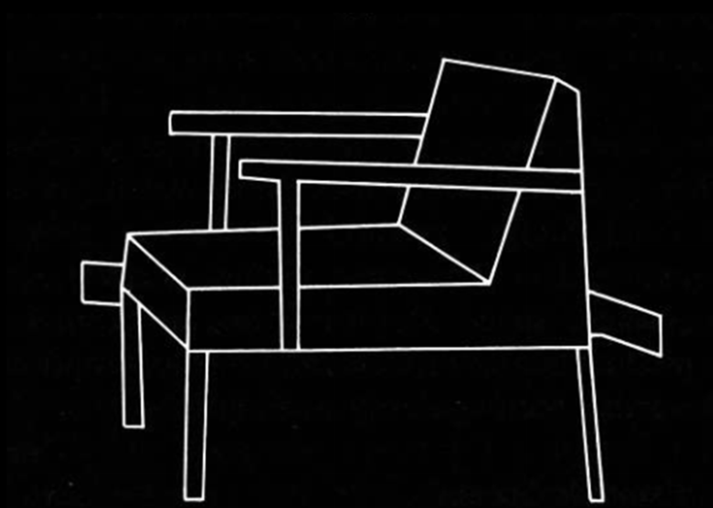


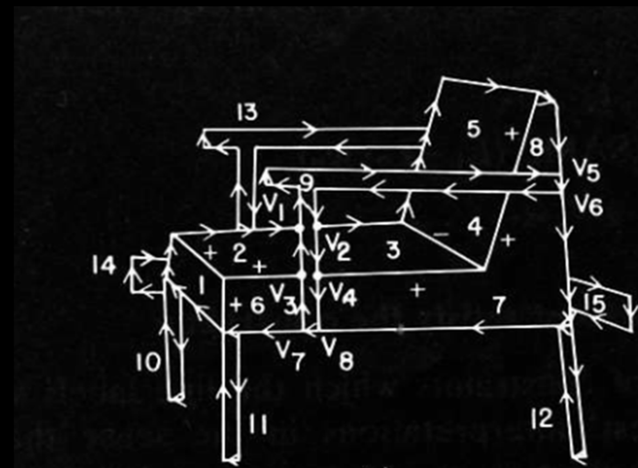


What do we learn from past history?

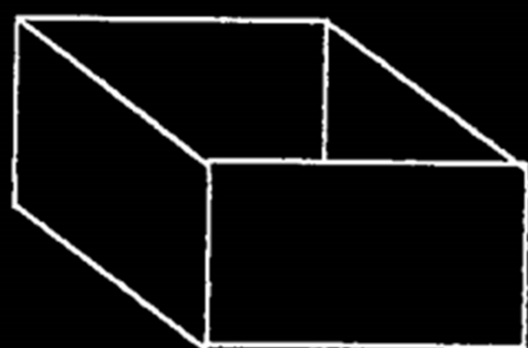
Historical perspective

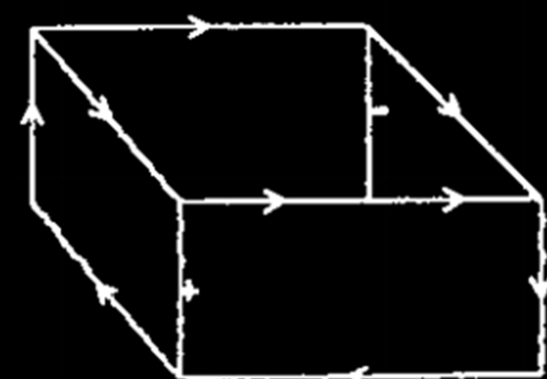
First era: Geometric/symbolic reasoning





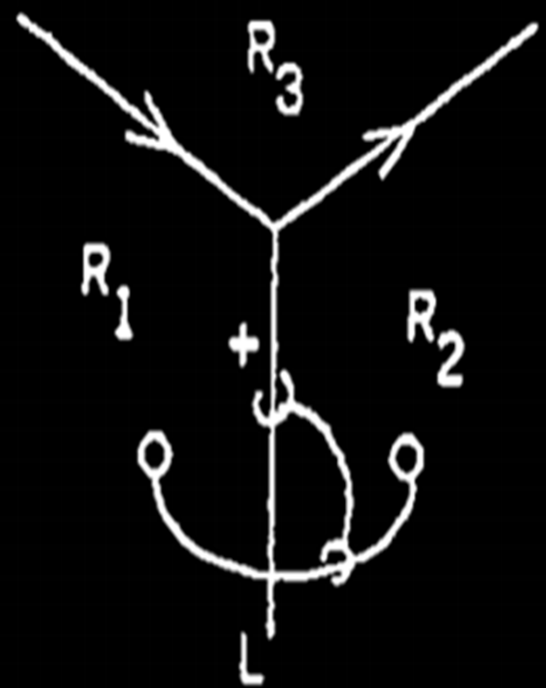
(a)



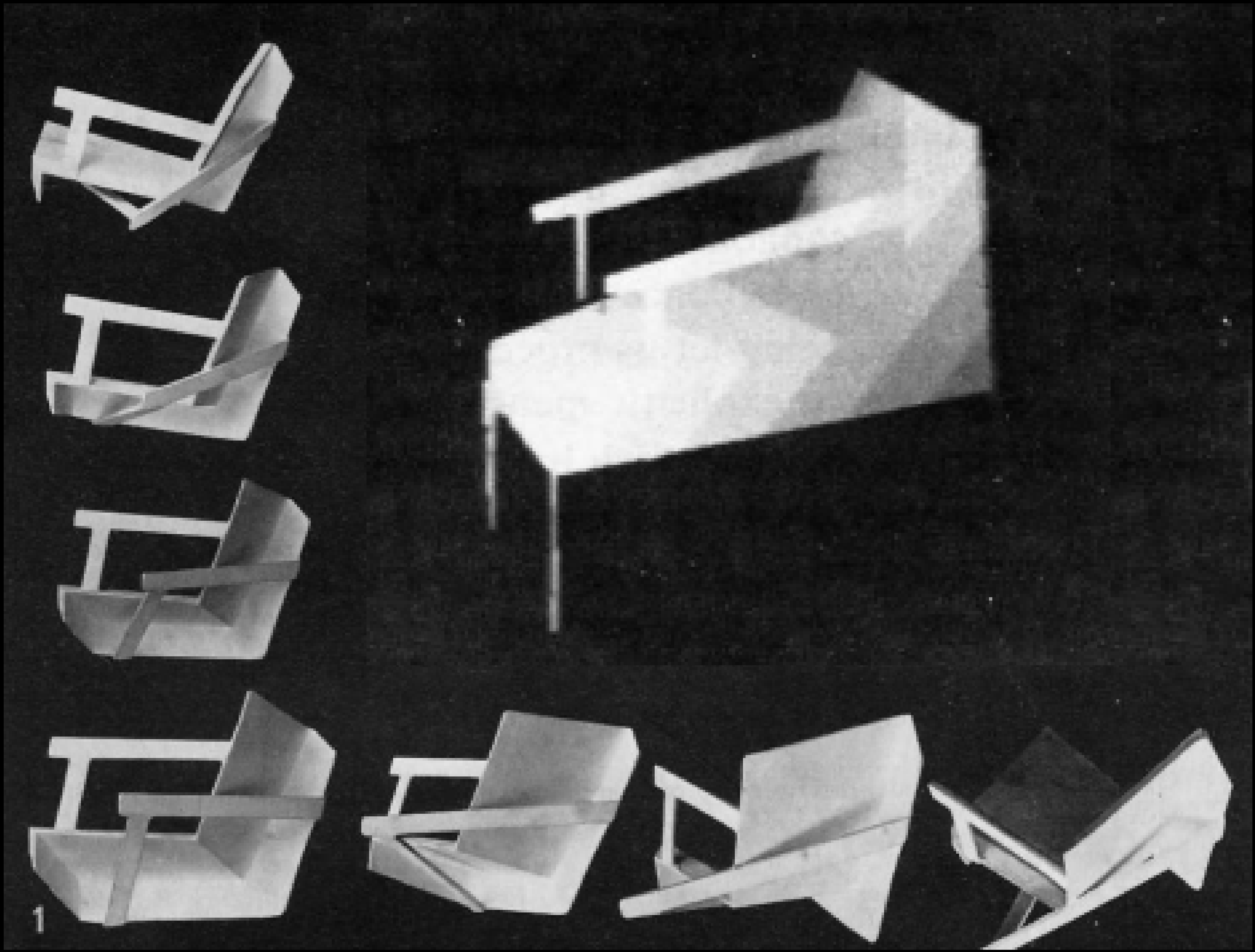


Huffman 71, Clowes 71, Kanade 80, 81 Sugihara 86, Malik 87, etc.

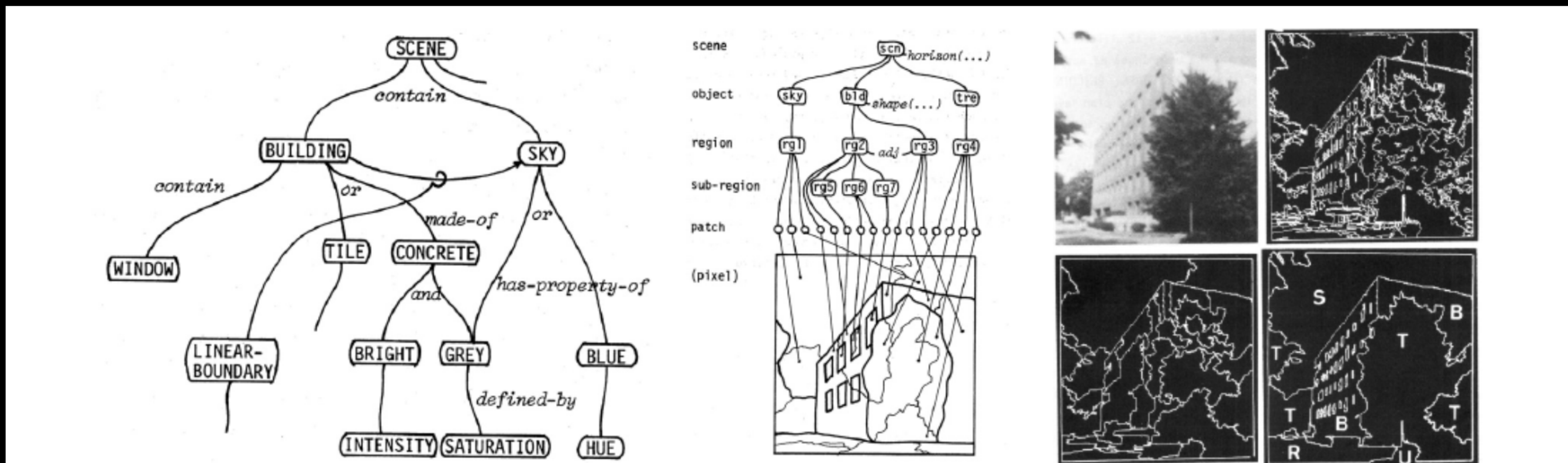
Kanade's Origami World, 1978



Kanade's chair… (Artificial Intelligence, 1981)

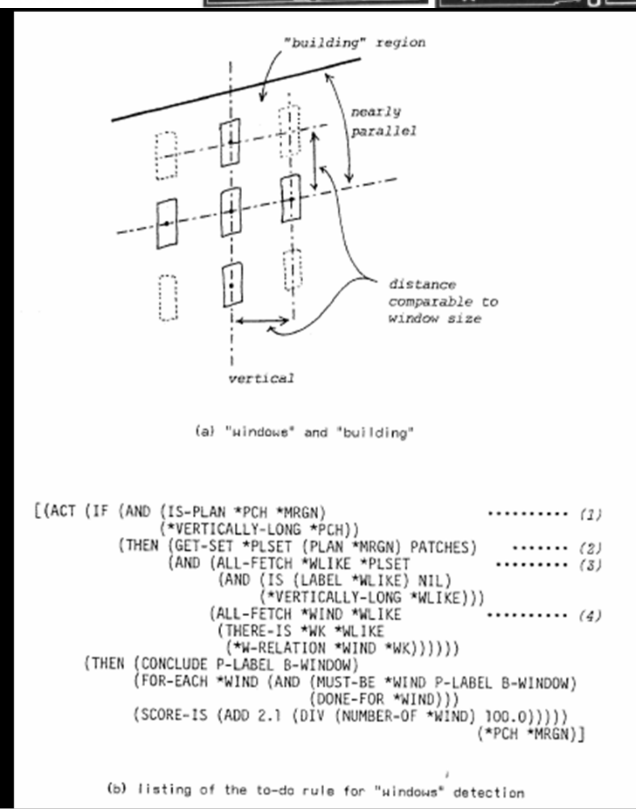


Scene parsing



[Ohta & Kanade 1978]

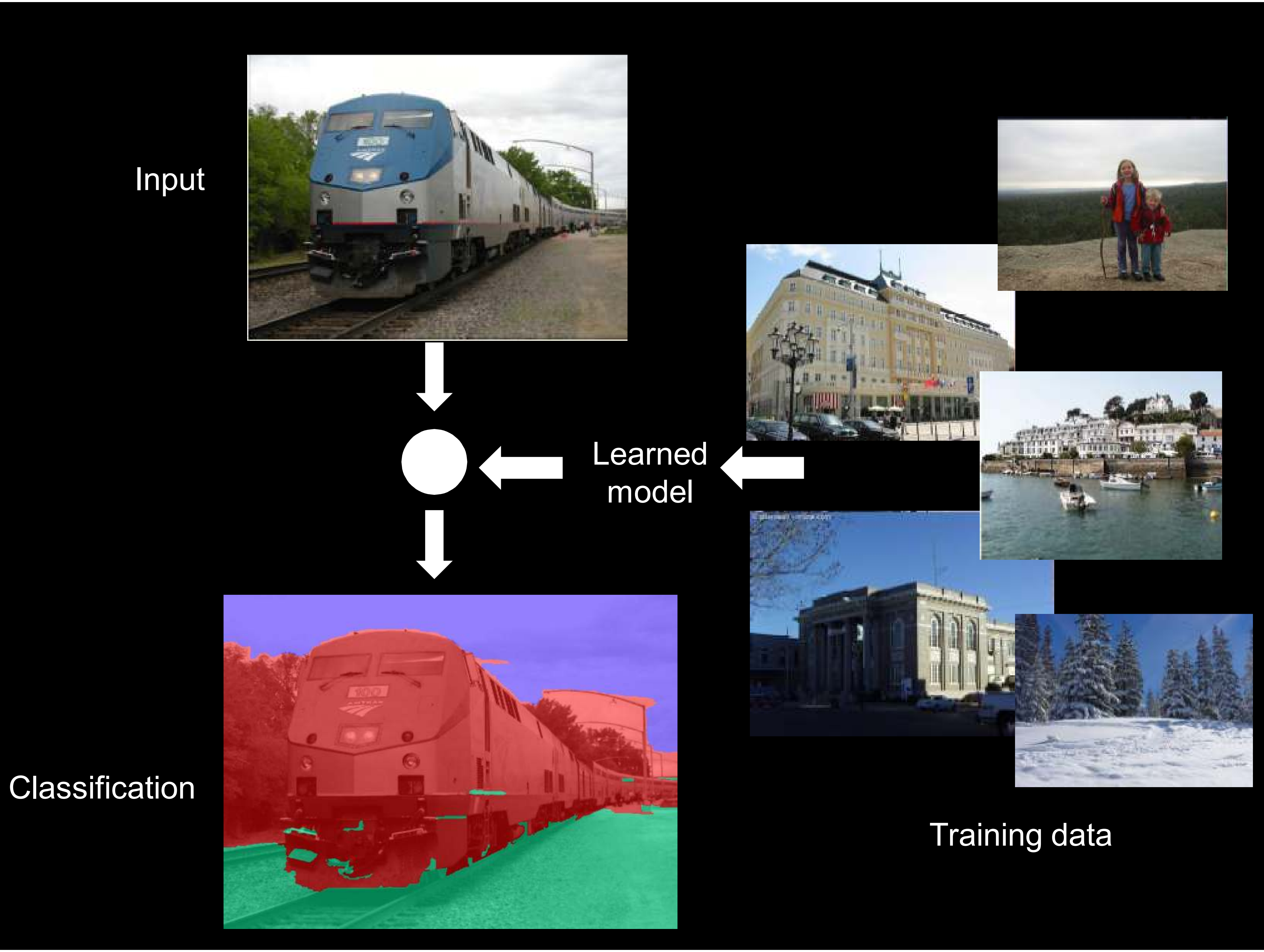
- Guzman (*SEE*), 1968
- Yakimovsky & Feldman, 1973
- Hansen & Riseman (*VISIONS*), 1978
- Barrow & Tenenbaum 1978
- Brooks (*ACRONYM*), 1979
- Ohta & Kanade, 1978



Issues

- Assumed “good” (perfect?) geometric elements inferred from the image.
- Limitations on computation, data, inference techniques prevented practical estimation of geometric primitives.

Second era: Statistical machine learning


Input
Learned
model
Classification
Training data

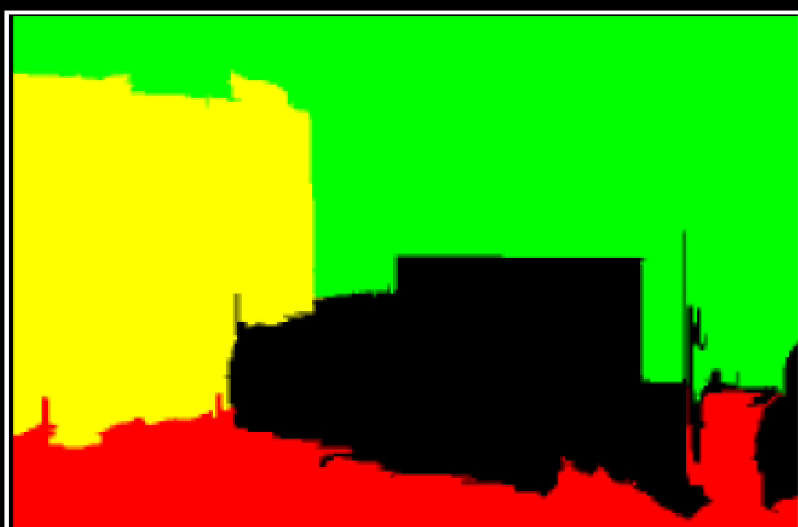
- Now we have the opposite problem:
  - Powerful tools to estimate low-level geometric cues (e.g., surface labels)
  - Does not incorporate high-level geometric constraints (e.g., orthogonality, intersections, etc.)
  - Does not incorporate higher-level reasoning

Now (Part I): learning+reasoning

Structured prediction tools

$$\min_{\boldsymbol{w}} \|\boldsymbol{w}\|^2 + C\sum_{n=1}^{l} \max_{y\in Y}\big(\underbrace{\Delta(y_n, y)}_{\text{Loss}} + \boldsymbol{w}(\psi(x_n, y) - \underbrace{\psi(x_n, y_n)}_{\text{Margin}})\big)$$

Orientation maps (Lee et al. 2009)

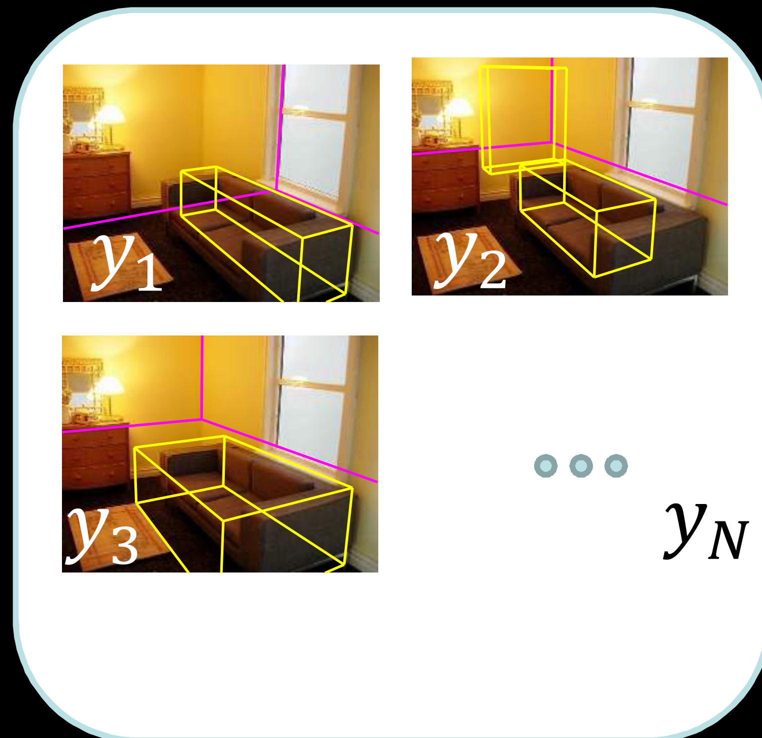


Geometric Context (Hoiem et al. 2007)

[Schwing/Urtasun 2012]

Structured prediction tools+ search

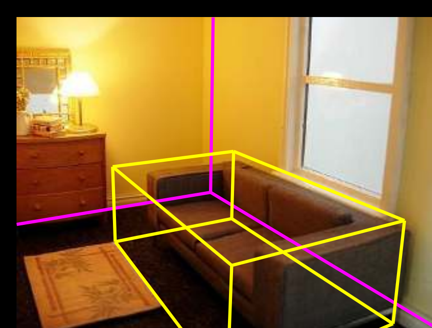
Input image features x



Generate hypotheses
Search through hypotheses to pick the best one
“Best” = maximum score

$$\max_y w^T \varphi(x, y)$$

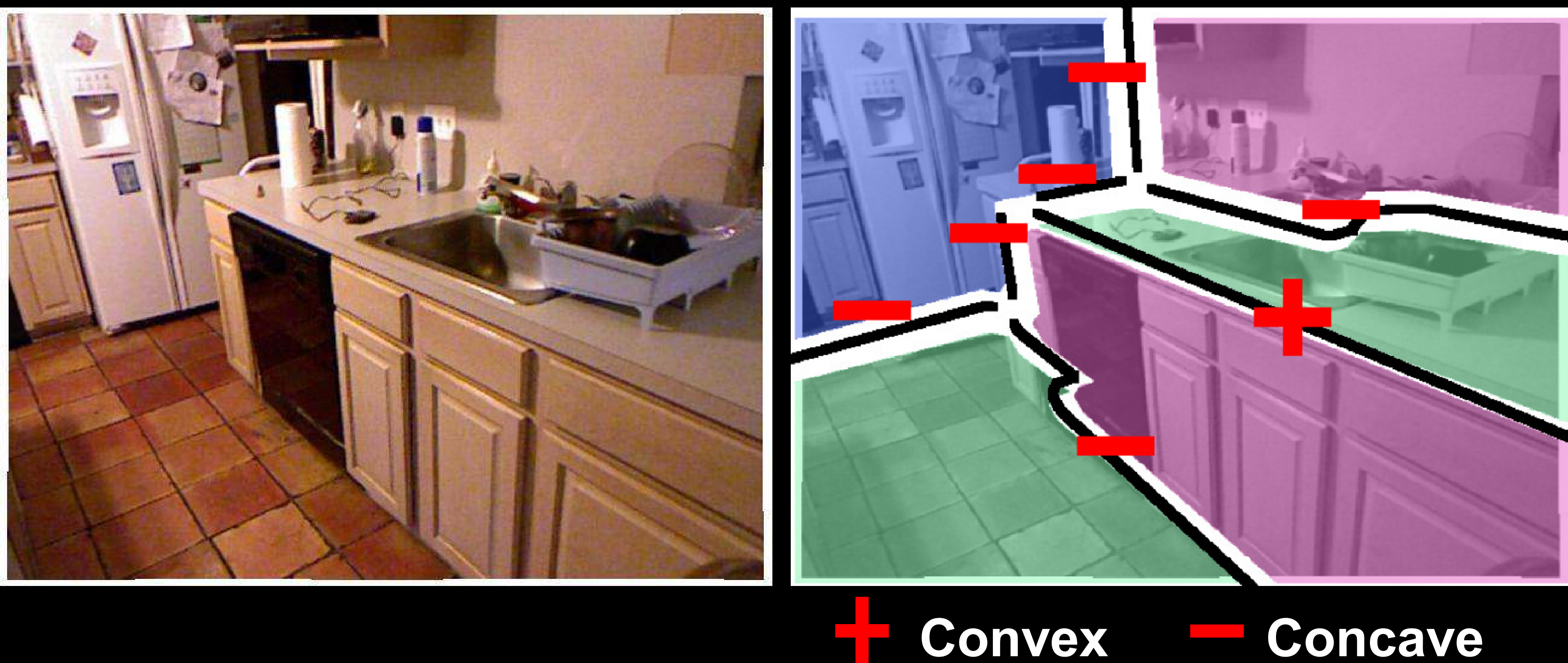
Score computation learned from data using structured prediction tools



Final scene configuration

Optimization tools

$$\arg\max_{\mathbf{x}\in\{0,1\}^n} \mathbf{c}^T\mathbf{x} + \mathbf{x}^T\mathbf{H}\mathbf{x} \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{1}$$



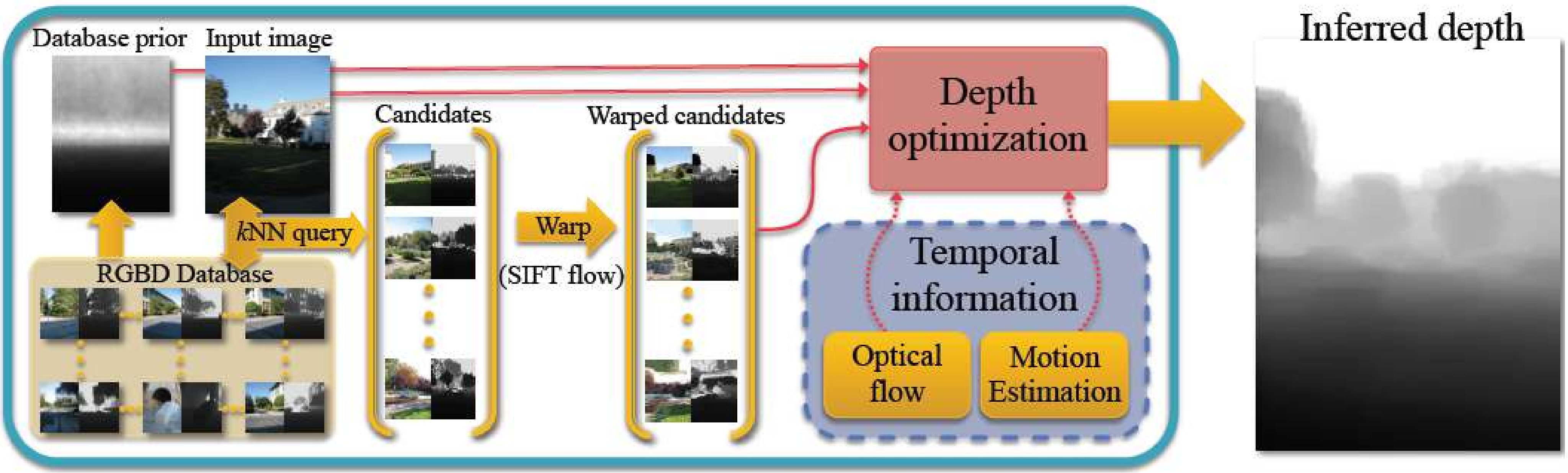
[Fouhey et al. ECCV 2014]

+ Richer representations, including reasoning about geometric primitives (e.g., relative placement of surfaces, contact relationships, etc.)

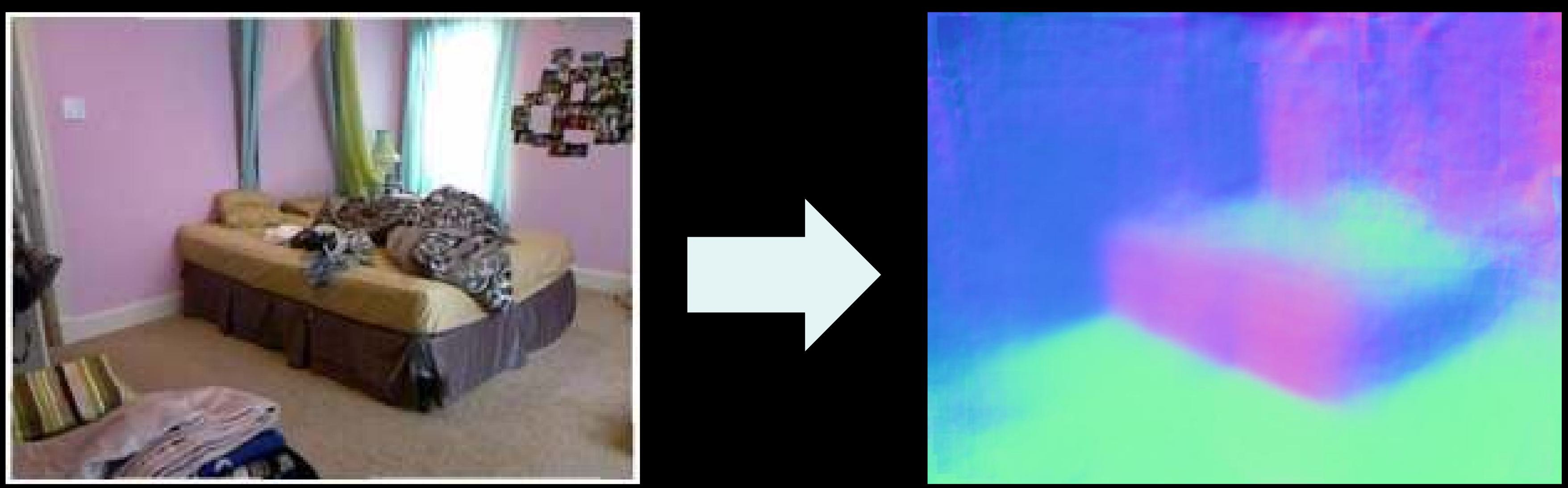
- Taming the combinatorics: How to generate and search hypothesis space efficiently?
- Summarizes a large amount of training data into a “simple” model
- Difficult to capture the richness of big data

Now (Part II): Data-driven interpretation

Label transfer

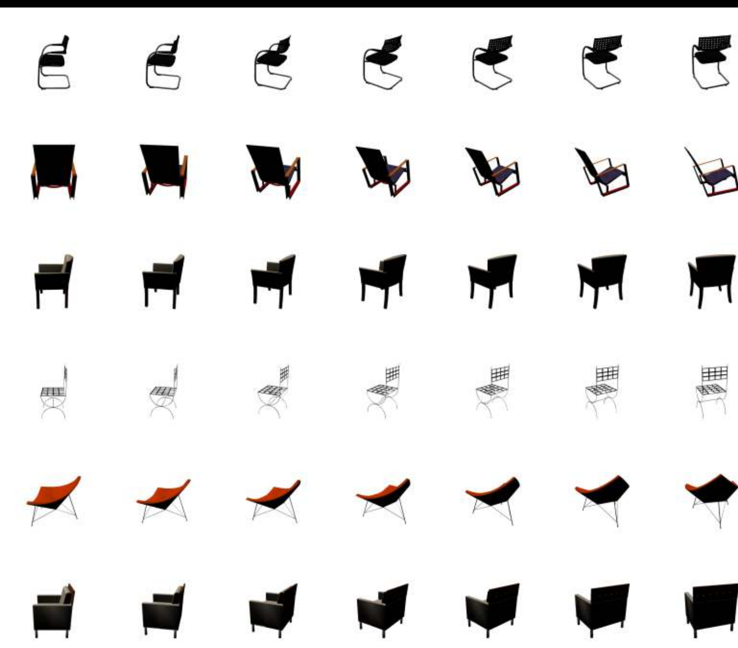


Karsch, Liu, Kang. *Depth Extraction from Video Using Non-parametric Sampling*. ECCV 2012.



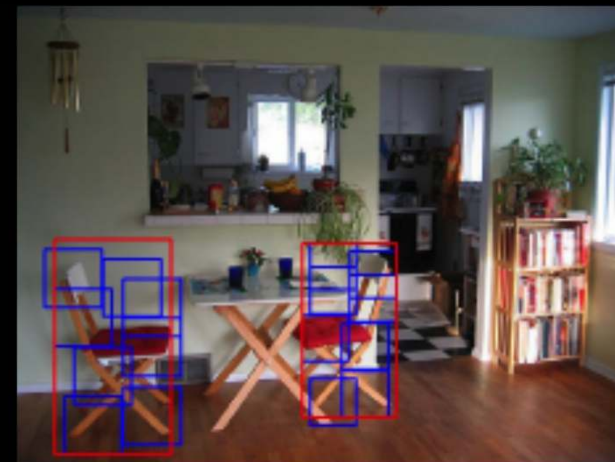
Fouhey, Gupta, Hebert. Data-Driven 3D Primitives for Single-Image Understanding. ICCV 2013.

Object transfer



Lots of object models

Input image



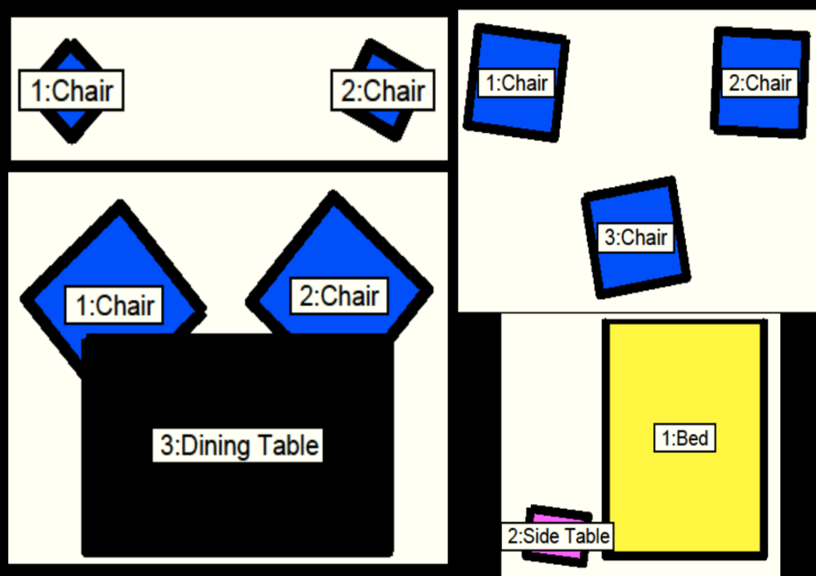
DPM output

Output

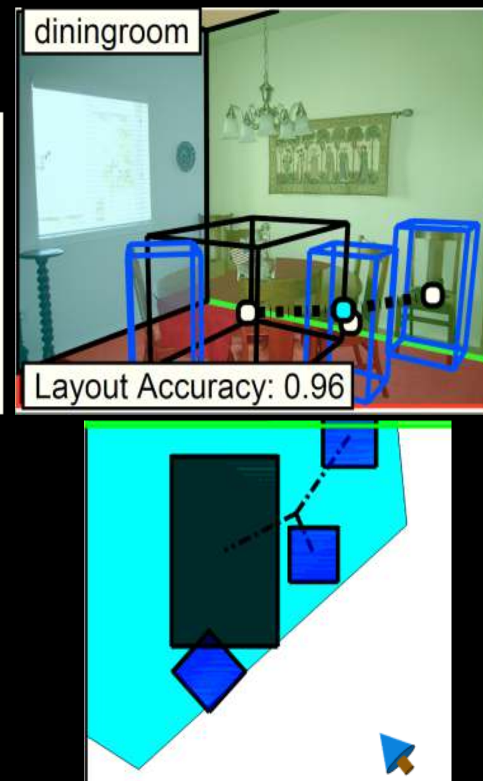
Matched models

Seeing 3D chairs: exemplar part-based 2D-3D alignment using a large dataset of CAD models. M. Aubry, D. Maturana, A. Efros, B. Russell and J. Sivic CVPR, 2014.

Object transfer



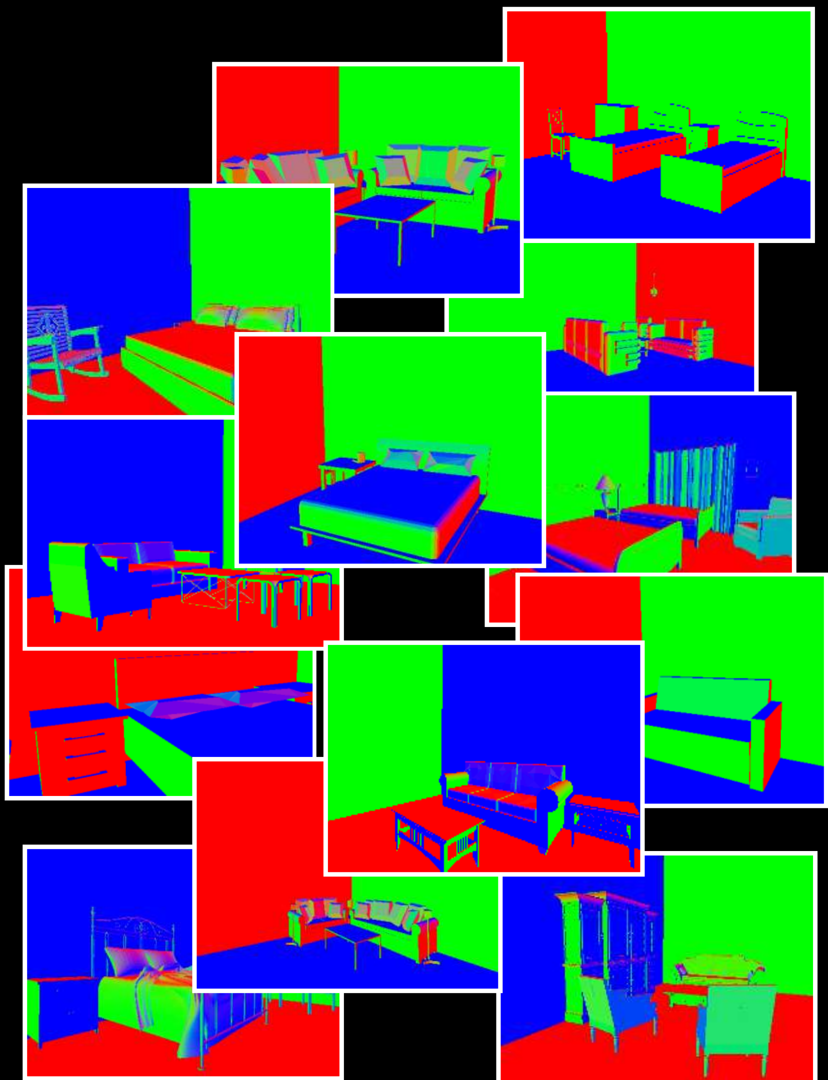
Training

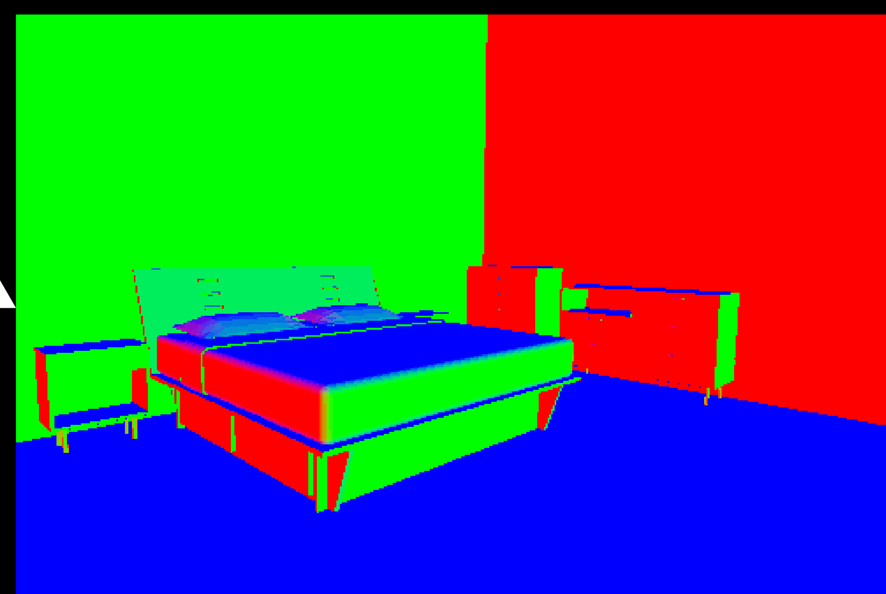


Testing

W. Choi, Y.-W. Chao, C. Pantofaru, and S. Savarese. Understanding Indoor Scenes using 3D Geometric Phrases. CVPR 2013.

Scene transfer





Nearest-neighbor search

Lots of 3D models

- (Arbitrarily) richer description: Transfer of semantics, 3D poses, segmentation, material properties, etc.

- How to relate 2D/appearance features to purely 3D geometric representations?
- What matching score/distance metric should be used?
- How to rank matches?

What will we talk about today?

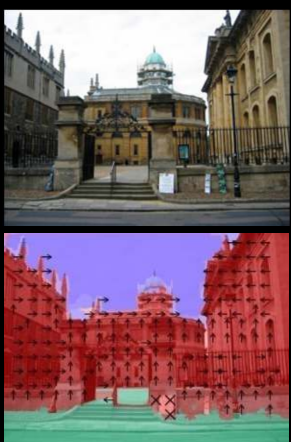
Tutorial Outline

Bottom up classifiers → More explicit constraint+reasoning

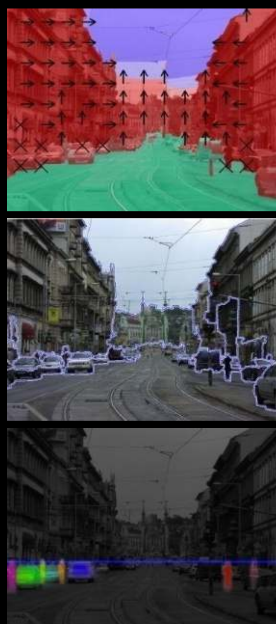
Qualitative → Explicit/Quantitative

Outline

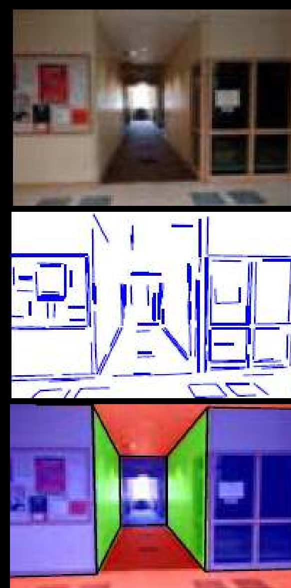
Part 1: Derek
Bottom-up Methods for Regions and Boundaries, Global Constraints



Region labels

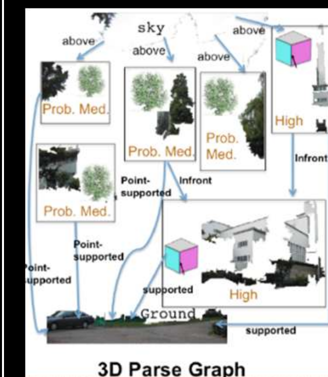


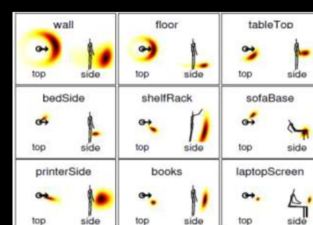
+ Boundaries and objects



Stronger geometric constraints from domain knowledge

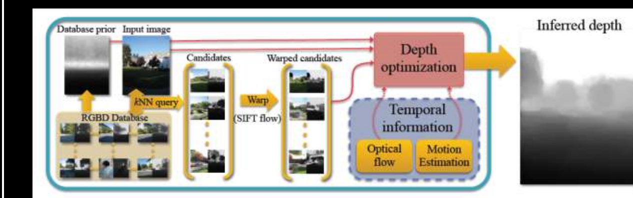
Part 2: Abhinav
Volumetric and Functional Constraints

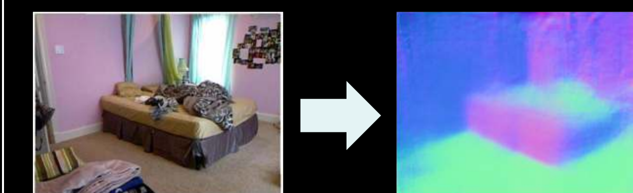


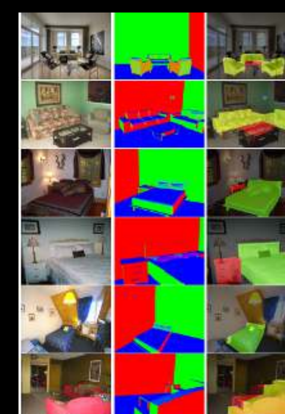


+ physical constraints +functional constraints

Part 3: David
Data-driven Models







Big Questions

- How to estimate geometric properties from an image?
- How to incorporate geometric constraints and which ones?
- How to combine reasoning tools with statistical classification/regression tools?
- How to use large-scale 3D data (3D models, kinect)

- How to combine with other 3D estimation methods?