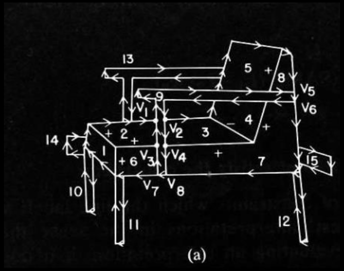
Data-Driven 3D

David Fouhey

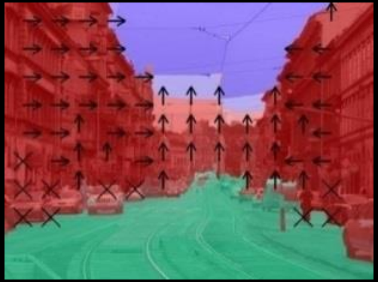
Recap

Martial

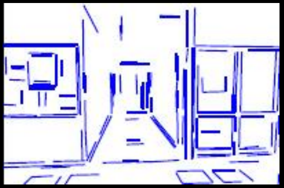


Introduction, Applications, History

Derek

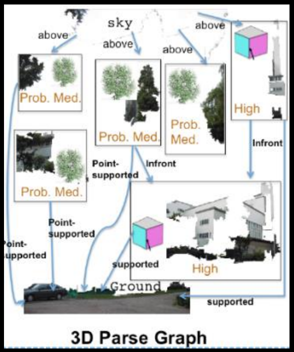


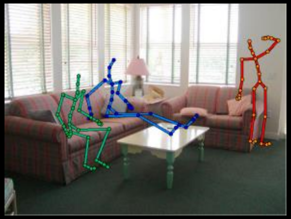
Region labels +Boundaries +Objects



Stronger geometric constraints

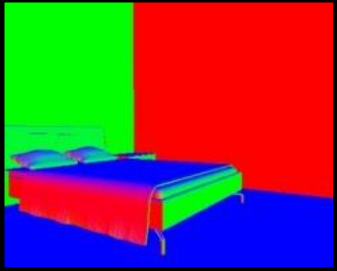
Abhinav

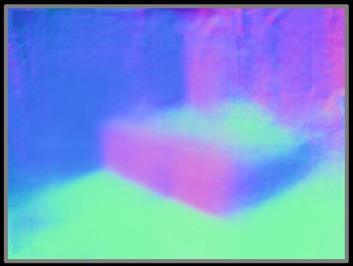




Volumetric + Functional Constraints

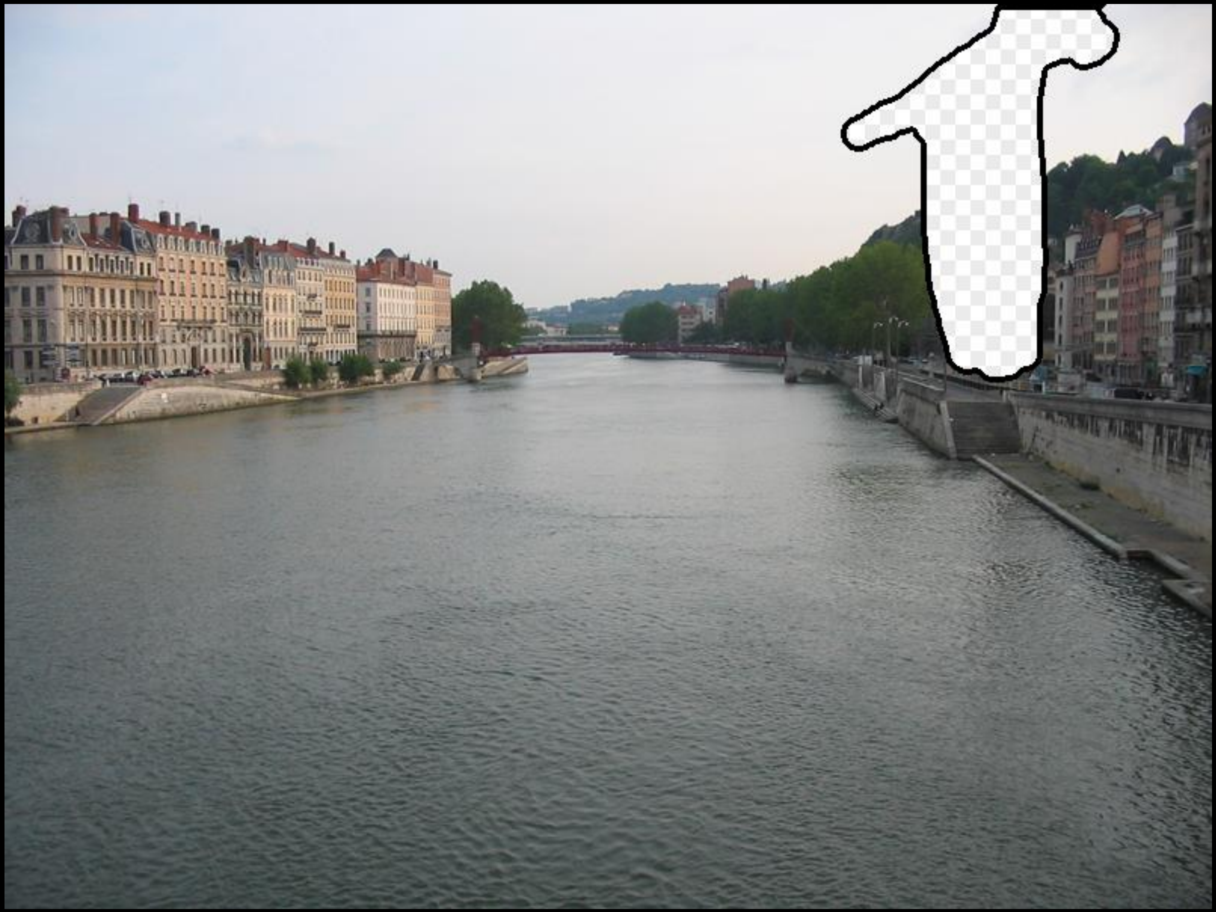
David





Data-Driven 3D

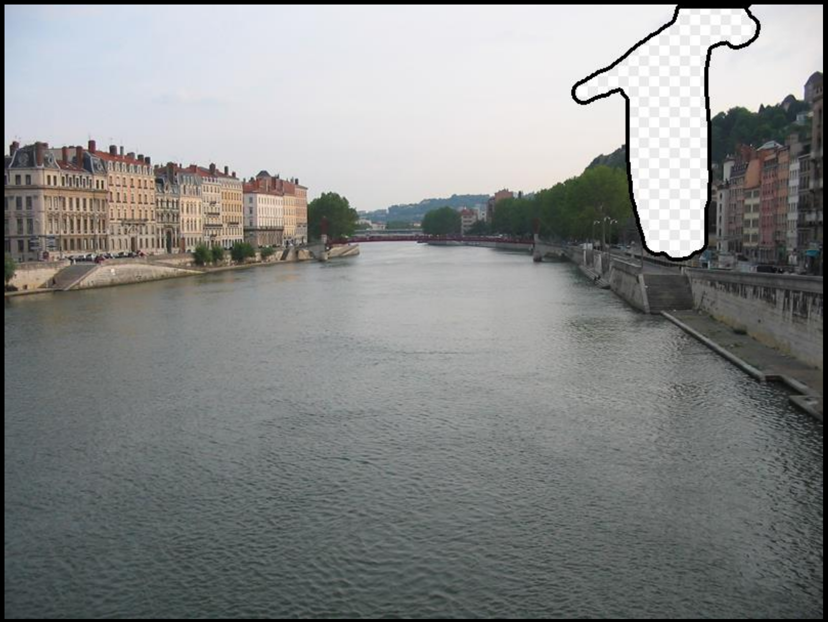
Data-Driven Interpretation



Every image that can be seen has been seen before (approximately)

Hays and Efros 2007

Data-Driven Interpretation

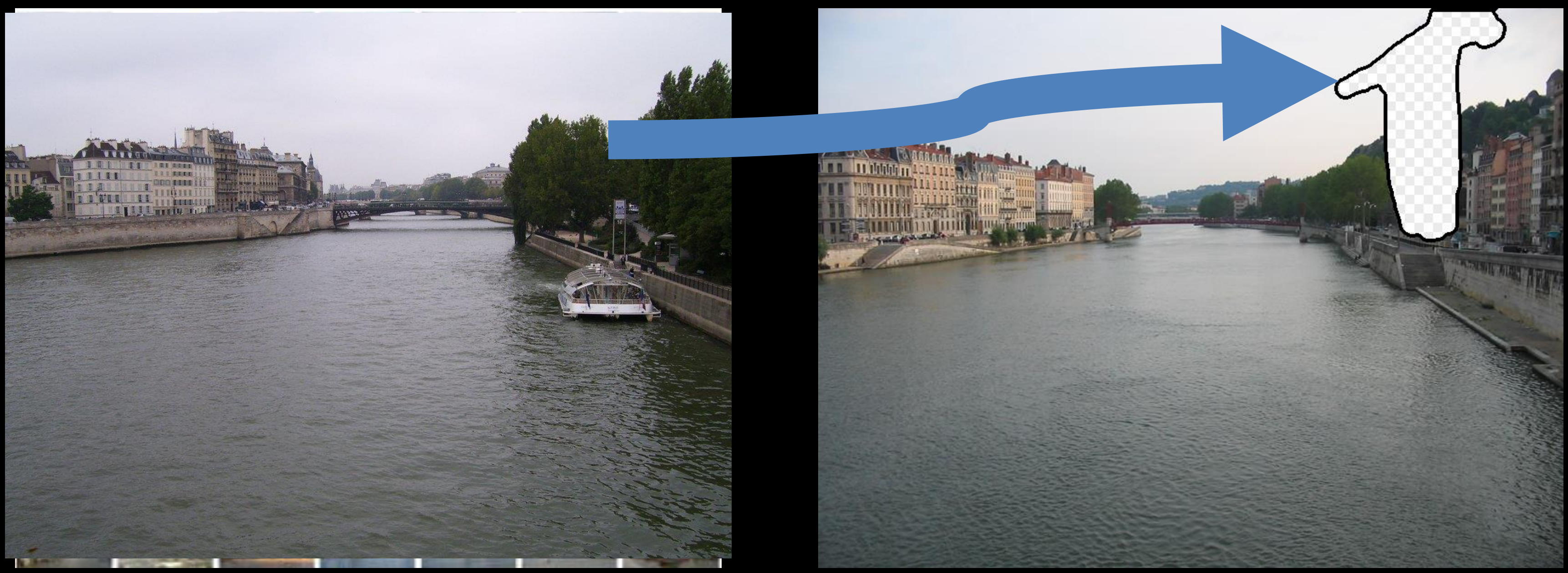


...

Every image that can be seen has been seen before (approximately)

Hays and Efros 2007

Data-Driven Interpretation



. . .

Every image that can be seen has been seen before (approximately)

Hays and Efros 2007

Data-Driven Interpretation

. . .

Every image that can be seen has been seen before (approximately)

Hays and Efros 2007

Data-Driven Interpretation

Works well where parametric modeling is hard but where there's data

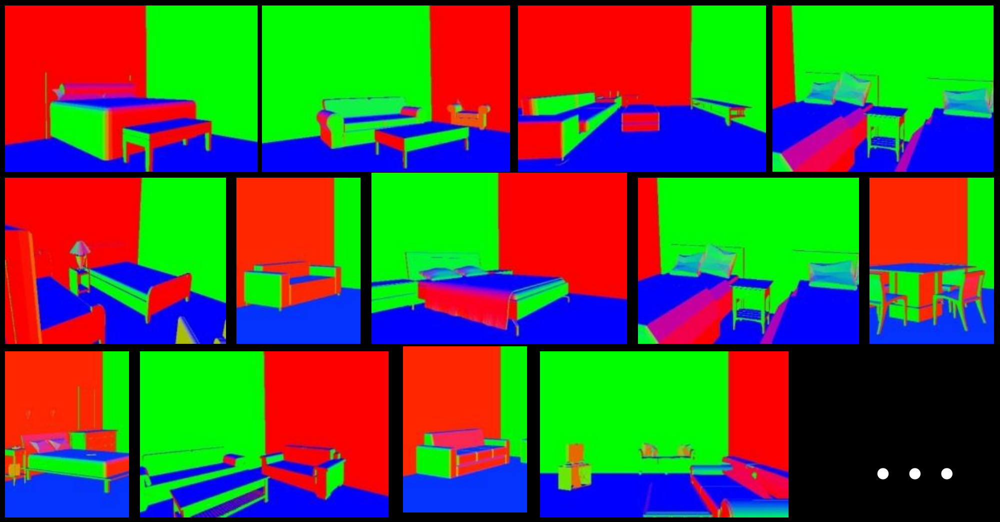
Advantages

Volumetric Interpretation

Interpretation by 3D Models

Sources

3D Model Databases



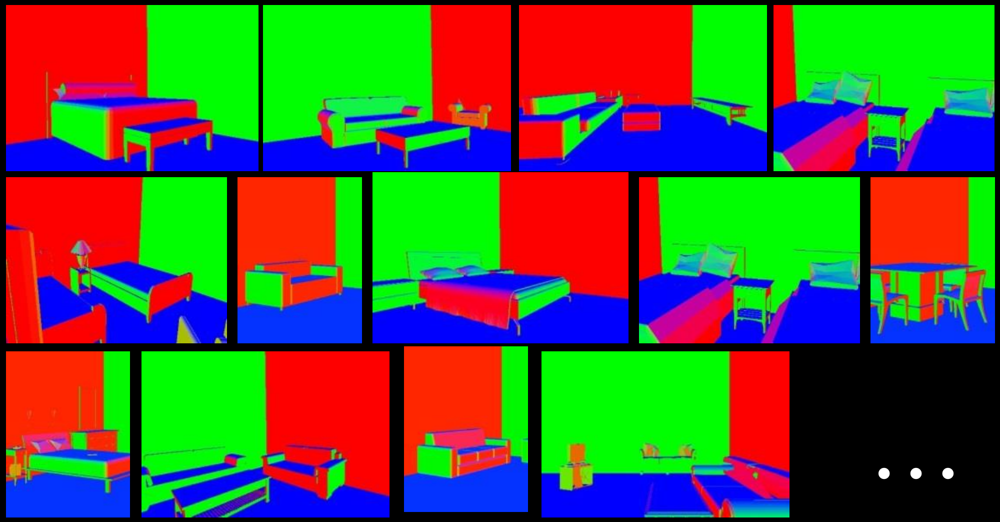
• • •

Kinect Databases

• • •

Goal

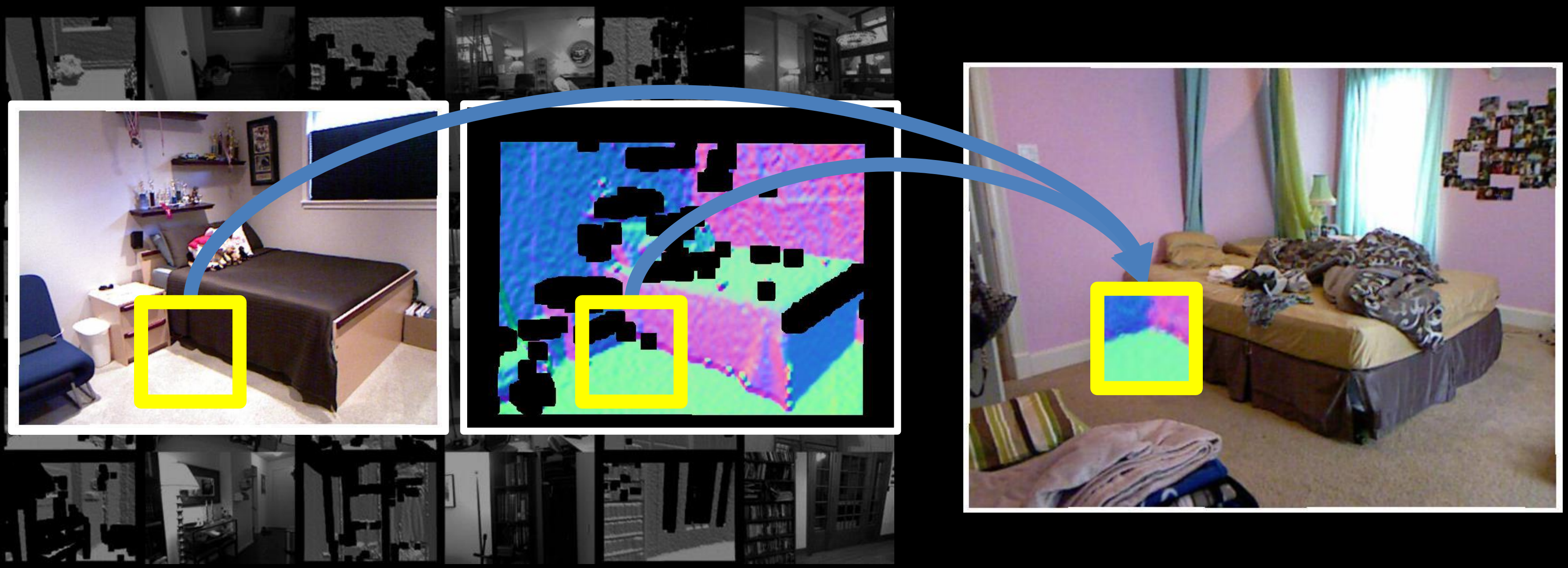
Goal



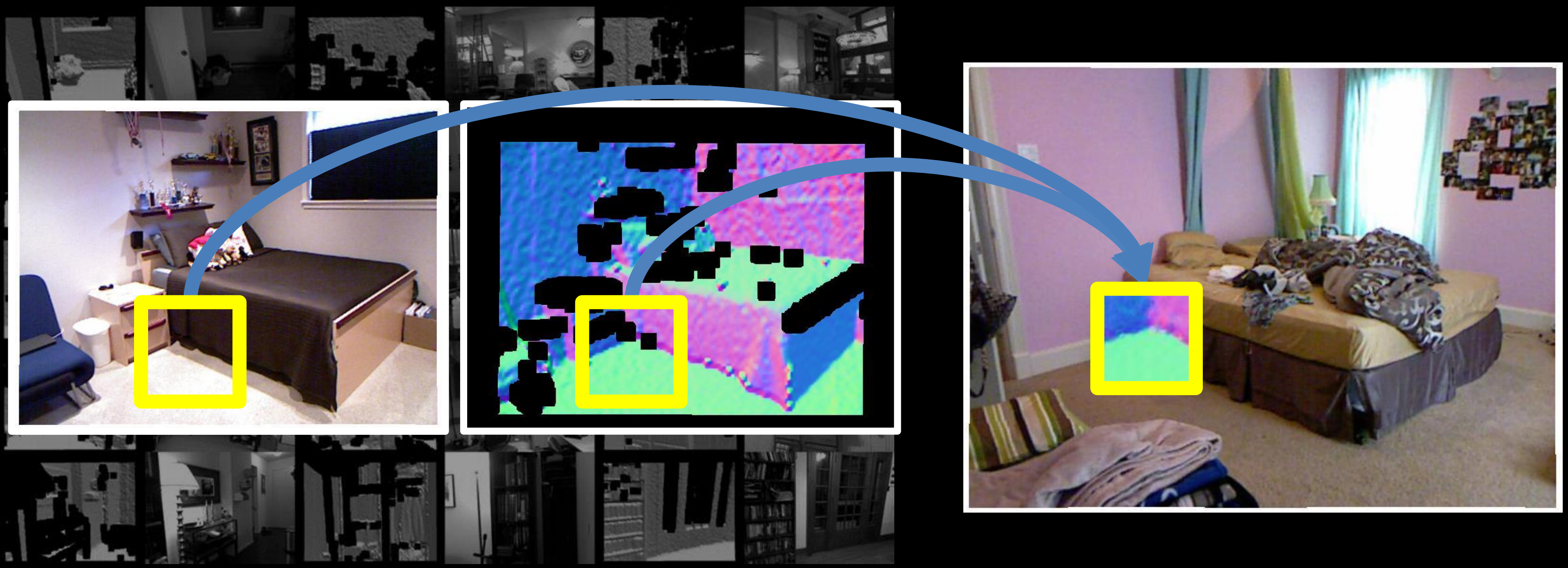
Goal

Goal

Goal



Goal



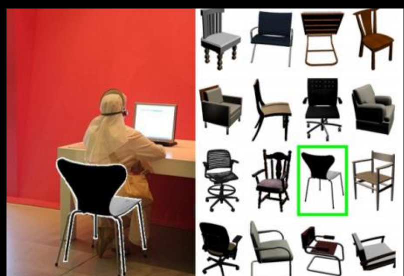
How do you:

(a) establish correspondence?

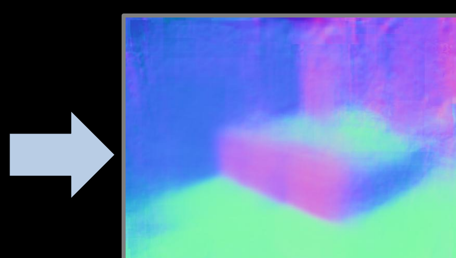
(b) transfer representations?

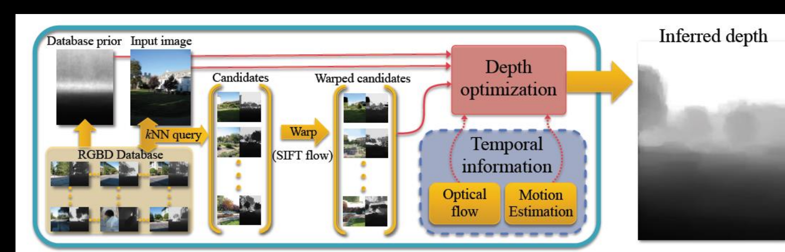
Overview

1. How to use 3D models



2. How to use the Kinect





Why 3D Models

Object Detector

Segmentation

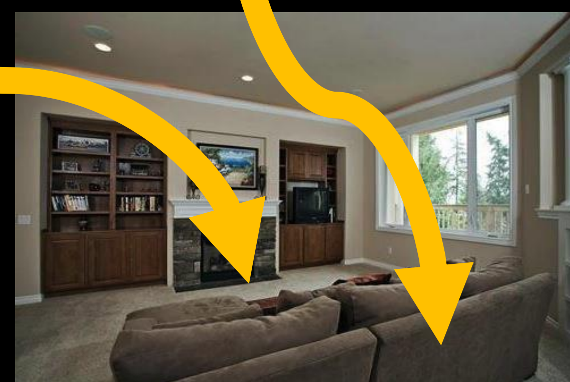
3D Model

Aubry et al. 2014

Why 3D Models

Input

Top 2D (GIST) Match



Top 3DNN Match

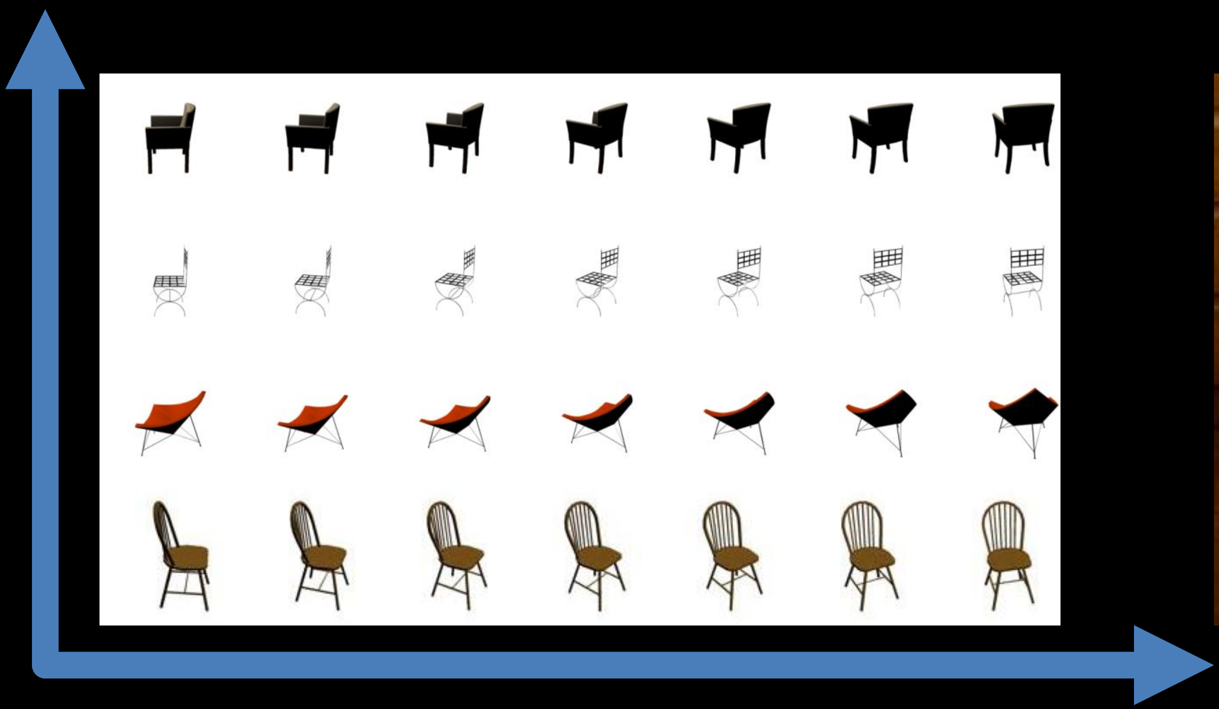
3D Models

- Advantages:
  - Full 3D – can be rendered and modified
  - Precise models may exist (e.g., IKEA)

- Disadvantages:
  - No corresponding natural color image (untextured or missing)

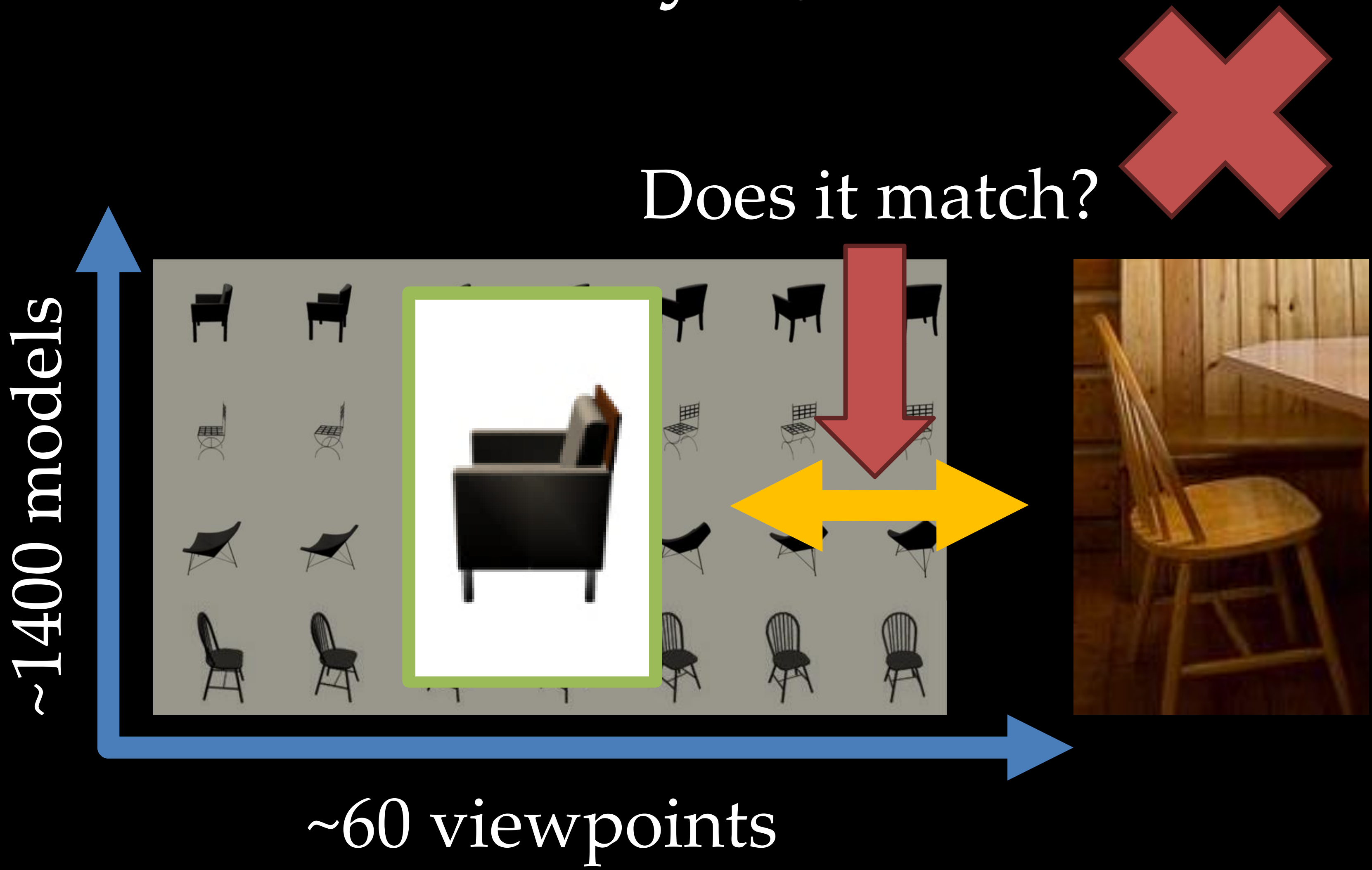
General Approach

Search over model and viewpoint



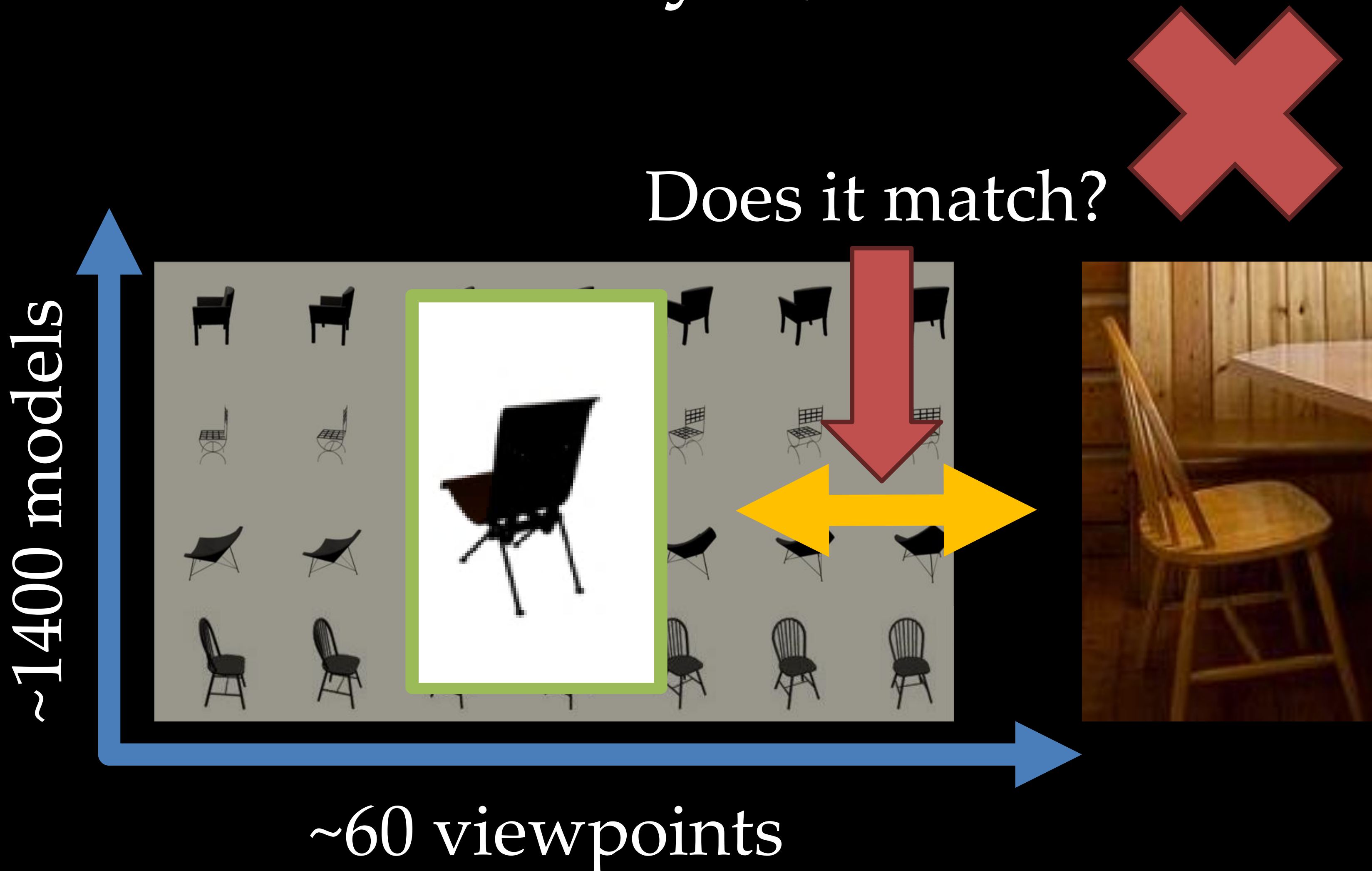
Chairs from Aubry et al. 2014

Primary Question



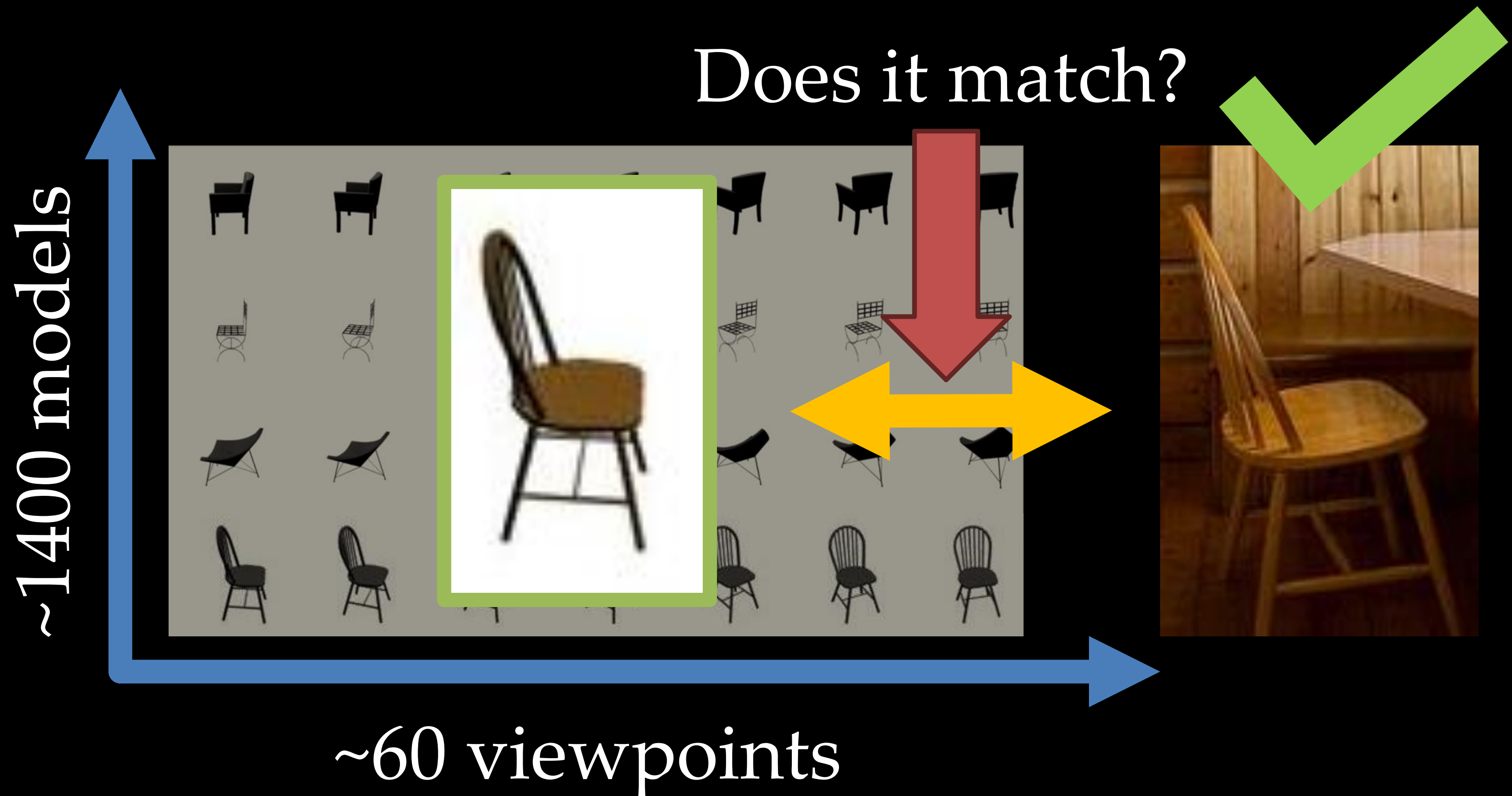
Chairs from Aubry et al. 2014

Primary Question



Chairs from Aubry et al. 2014

Primary Question



Chairs from Aubry et al. 2014

Difficulties

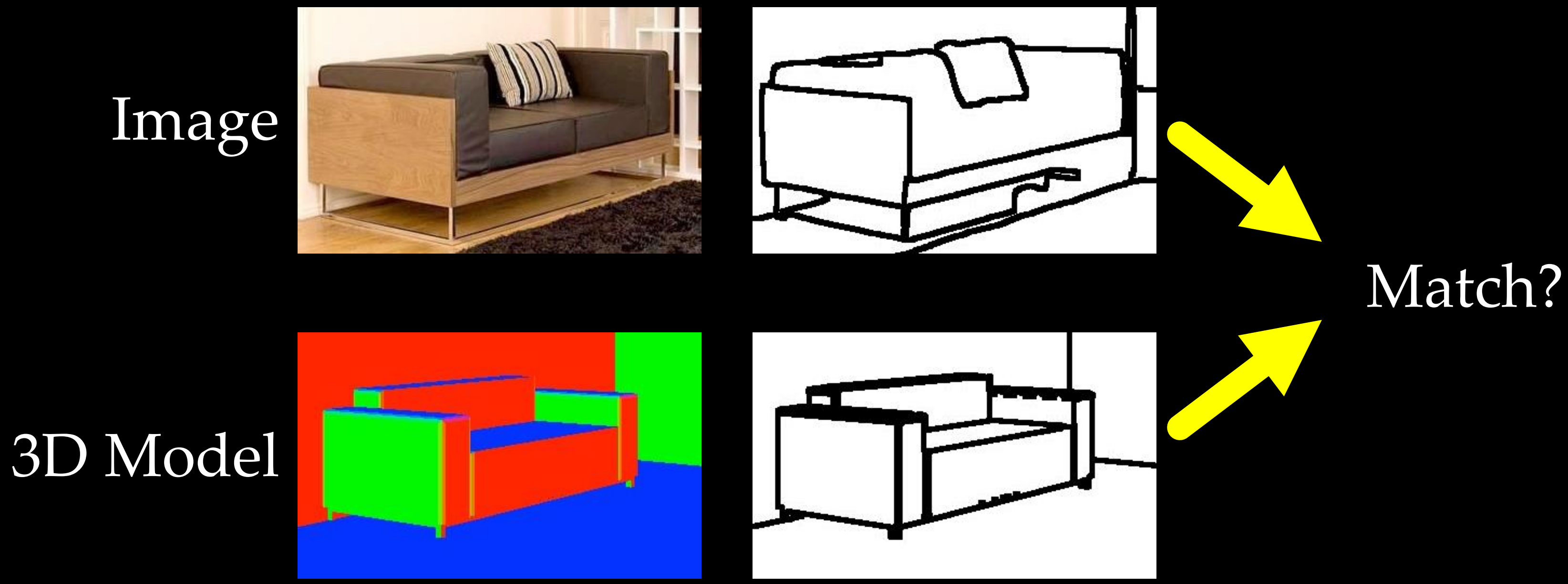
	Rendered	Natural
		
Texture	NO	YES
Occlusion	NO	YES
Background	**Fake**	**Natural**

Cross-Domain Matching

Goal: bring image and model into common representation

Chamfer Matching

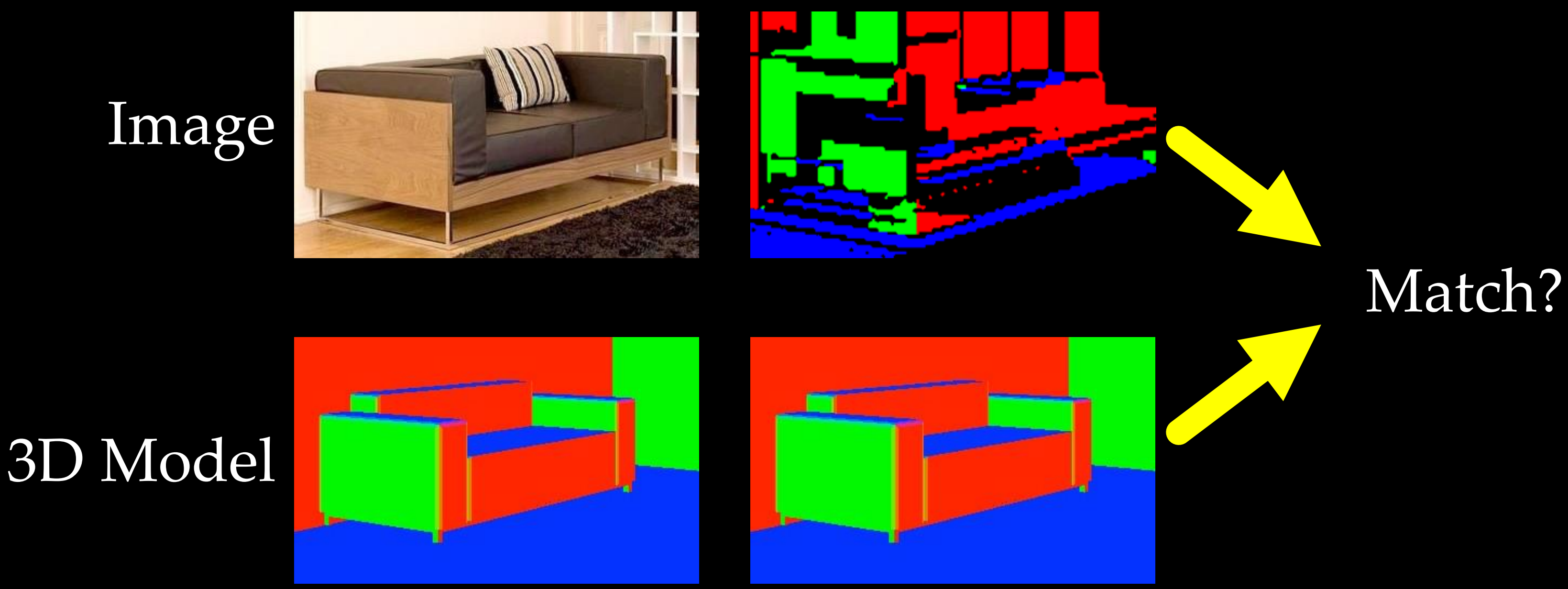
Assumption: edges in 3D are edges in 2D



Satkin et al., 2012, 2013, 2014; Lim et al., 2013; Ramnath et al., 2014;

Domain-Invariant

Assumption: can estimate 3D property from 2D



Satkin et al., 2012, 2013, 2014;

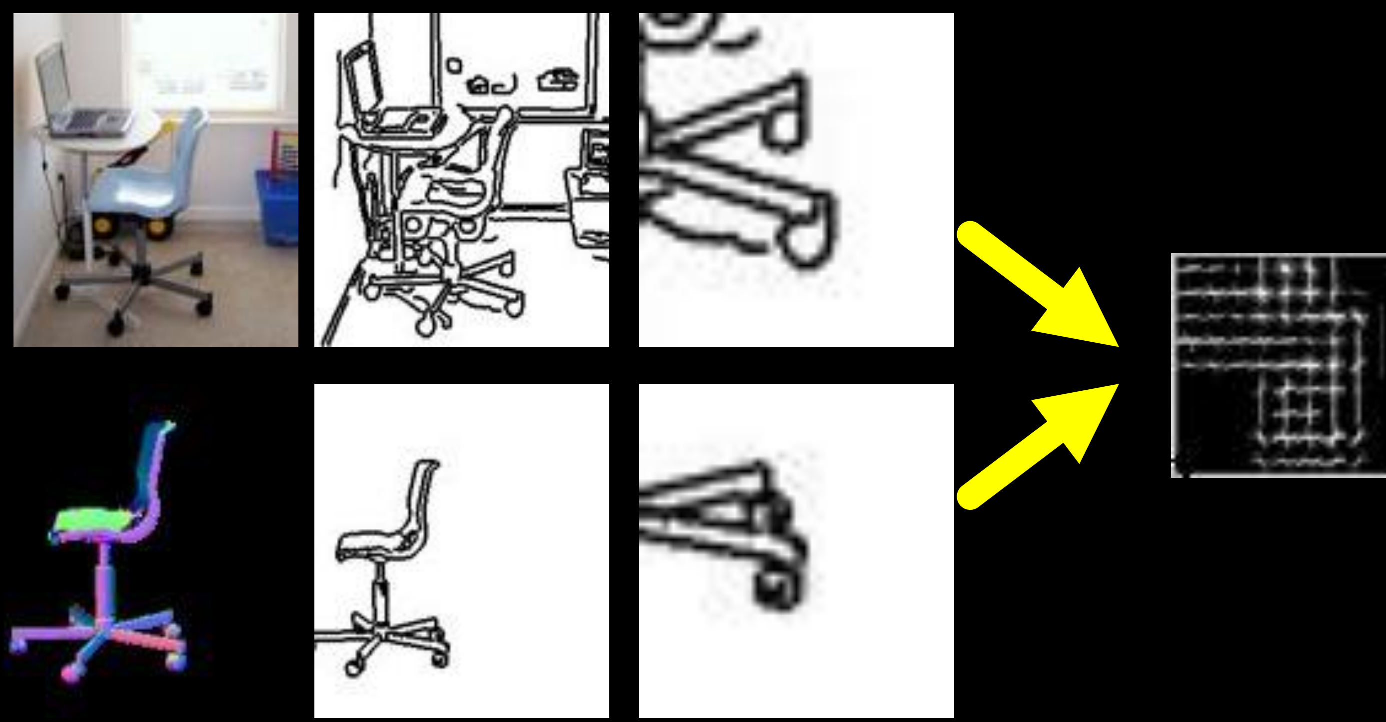
Domain-invariant "Images"

Assumption: edges in 3D are edges in 2D

Apply standard features/techniques

Image

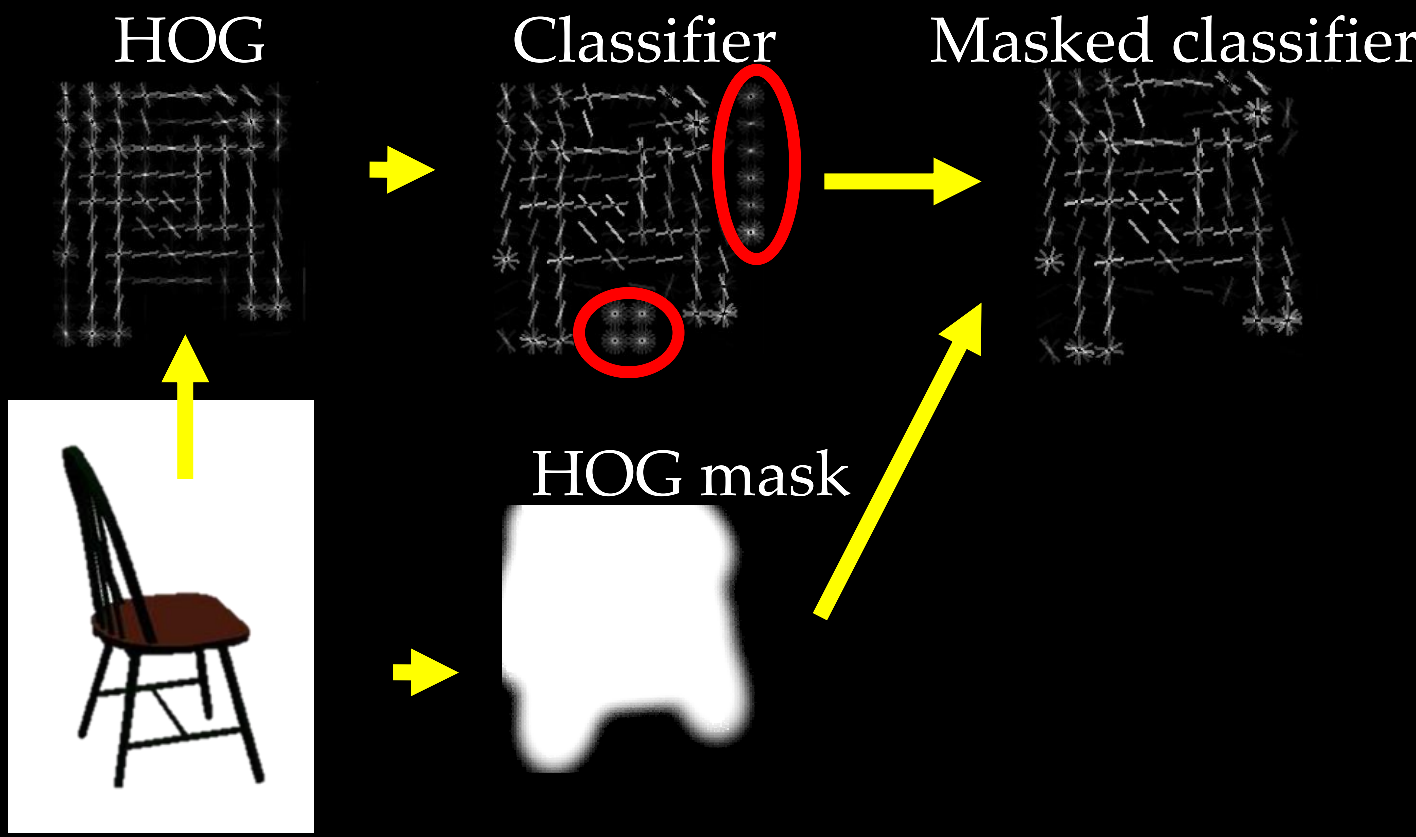
3D Model



Lim et al., 2013

Masking Features

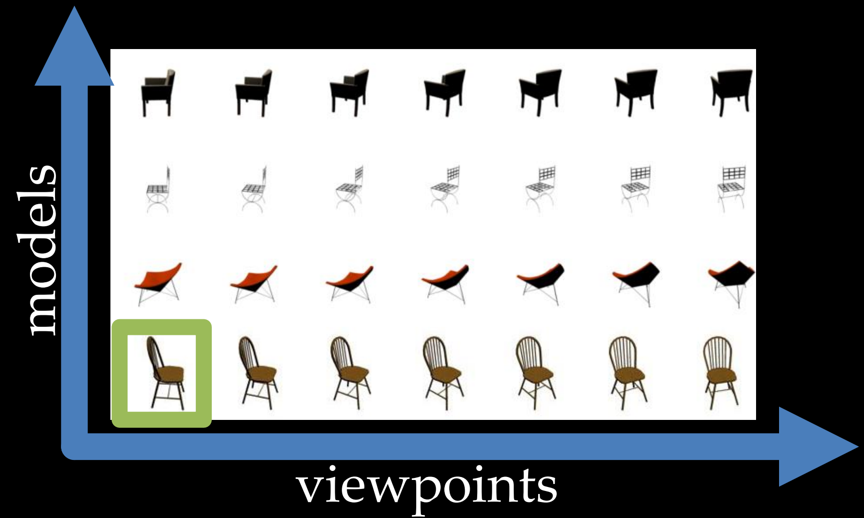
Assumption: only issue is background



Aubry et al., 2014; see also Shrivastava et al., 2011

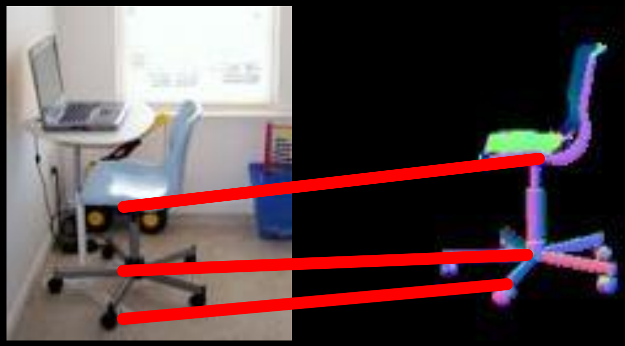
Searching Hypotheses

Render object parts



Aubry et al., 2014

Matches generate proposals



Lim et al., 2013

Results

Results

Results

Issues

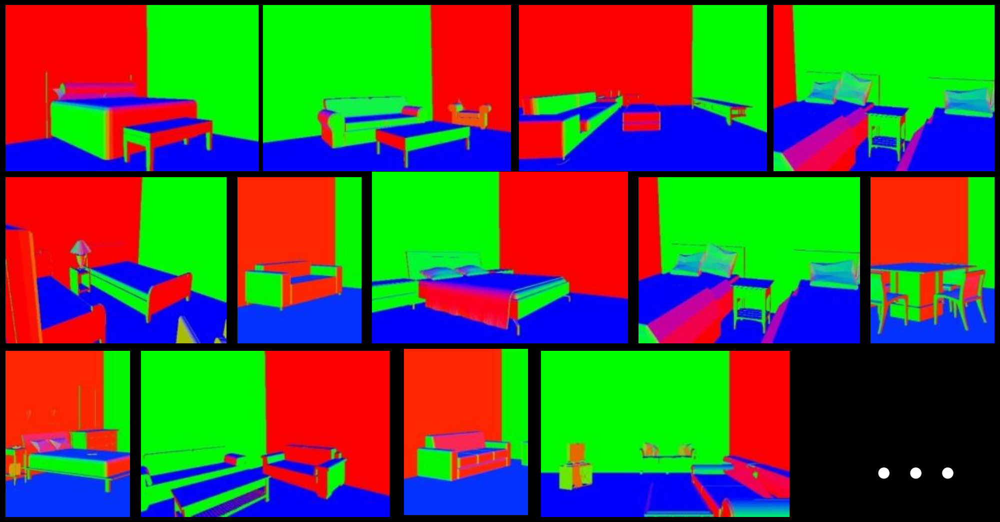
What's this?

Issues

Recognition and pose estimation is <u>hard,</u> but made easier by seeing the rest of the room.

2D-3D Scene Matching

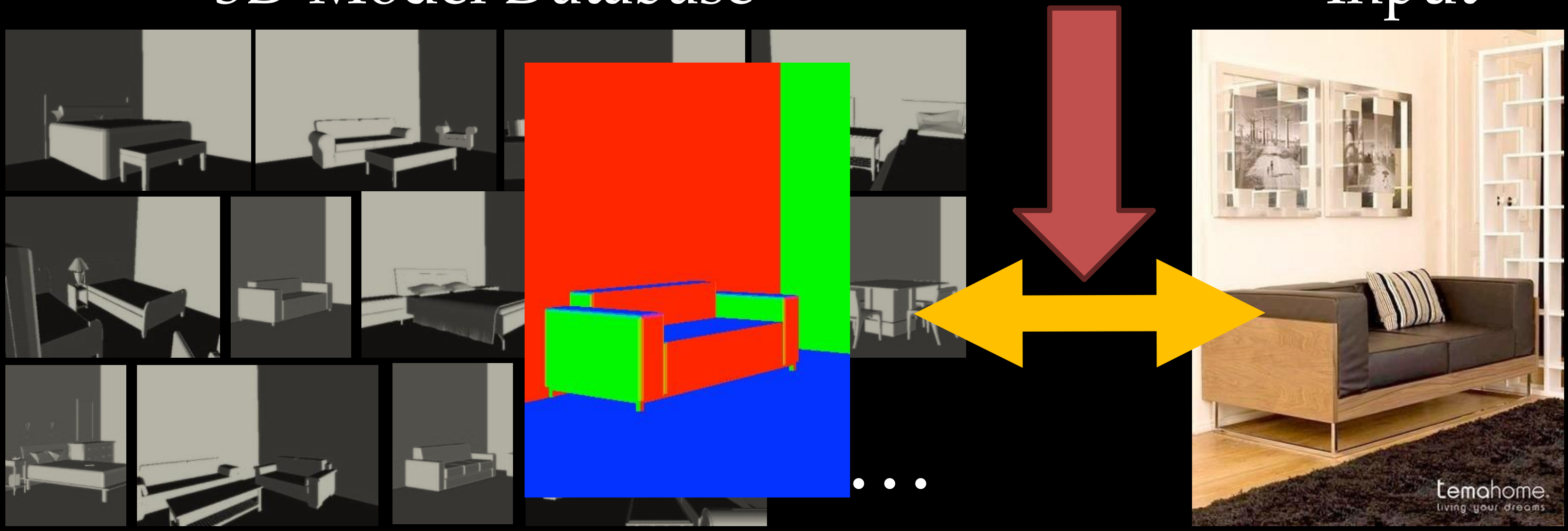
3D Model Database



Input

Satkin et al., 2012,2013,2014

2D-3D Scene Matching



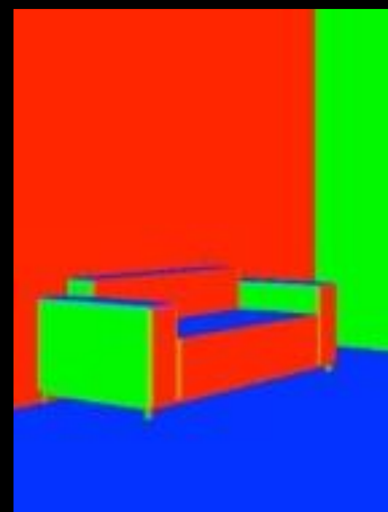
Satkin et al., 2012,2013,2014

Naïve 2D-3D Scene Matching

1K Models

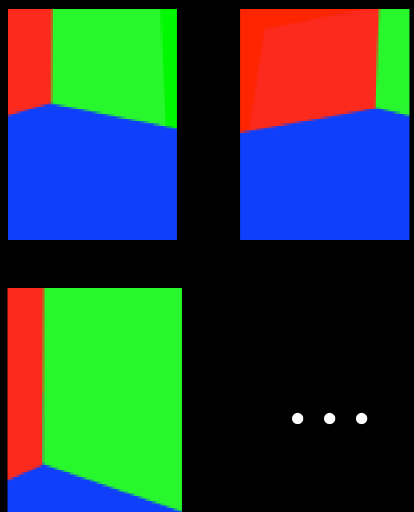


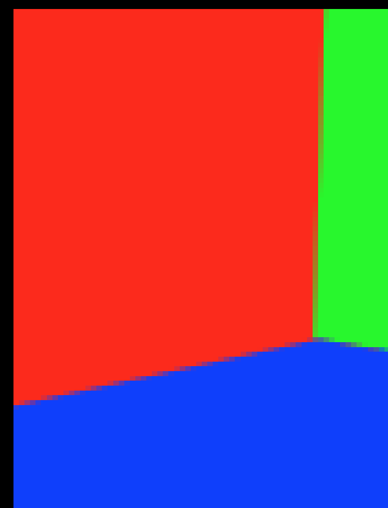
. . .




Naïve 2D-3D Scene Matching

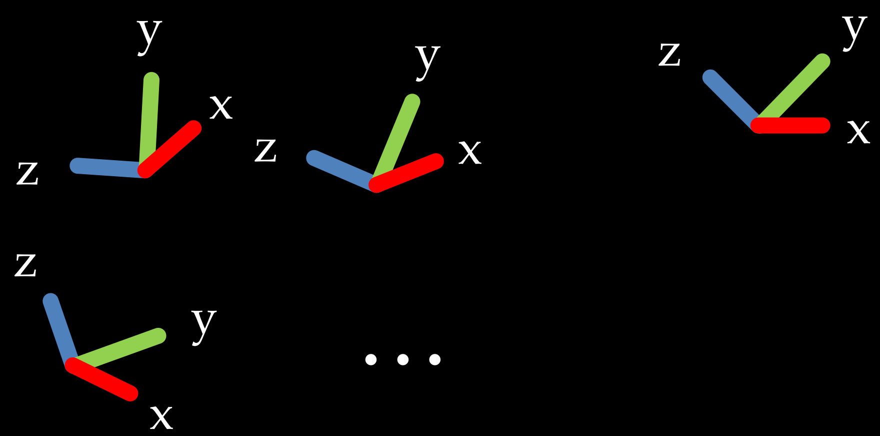
1K Models x 1K Layouts





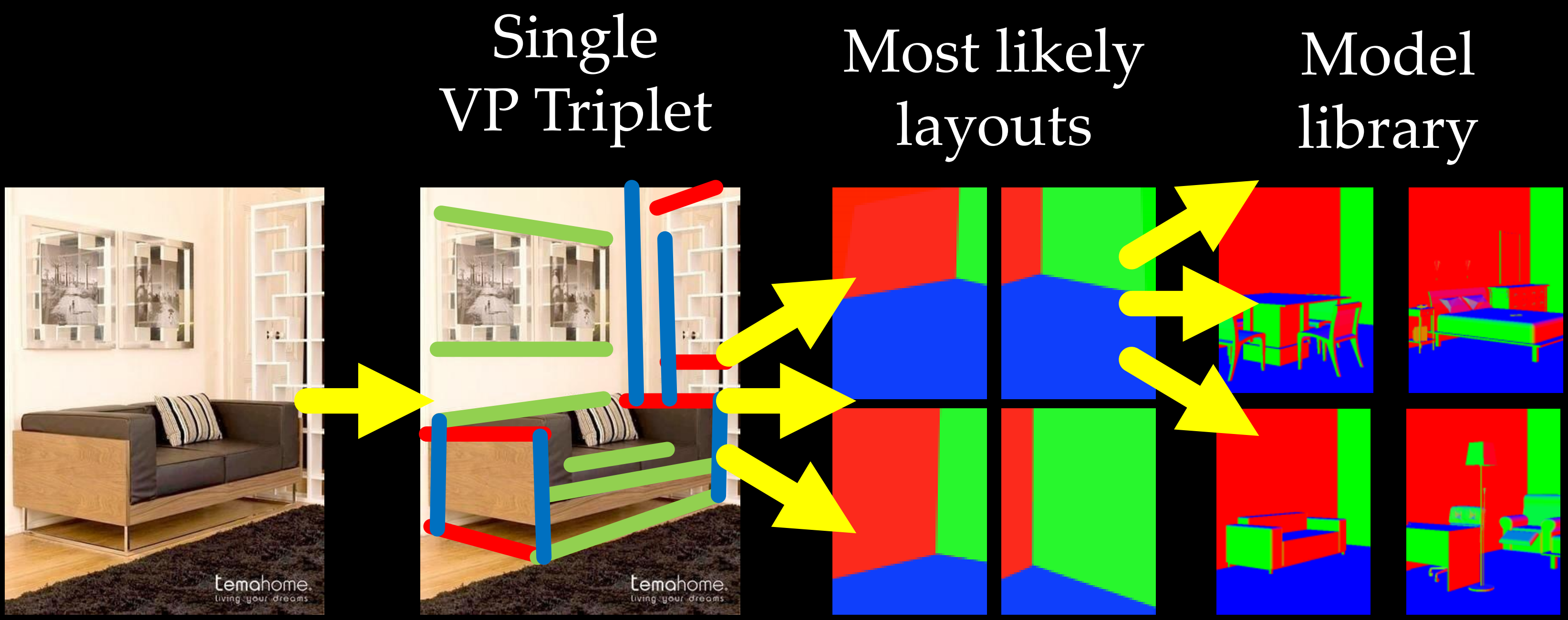
Naïve 2D-3D Scene Matching

1K Models x 1K Layouts x 100 rotations



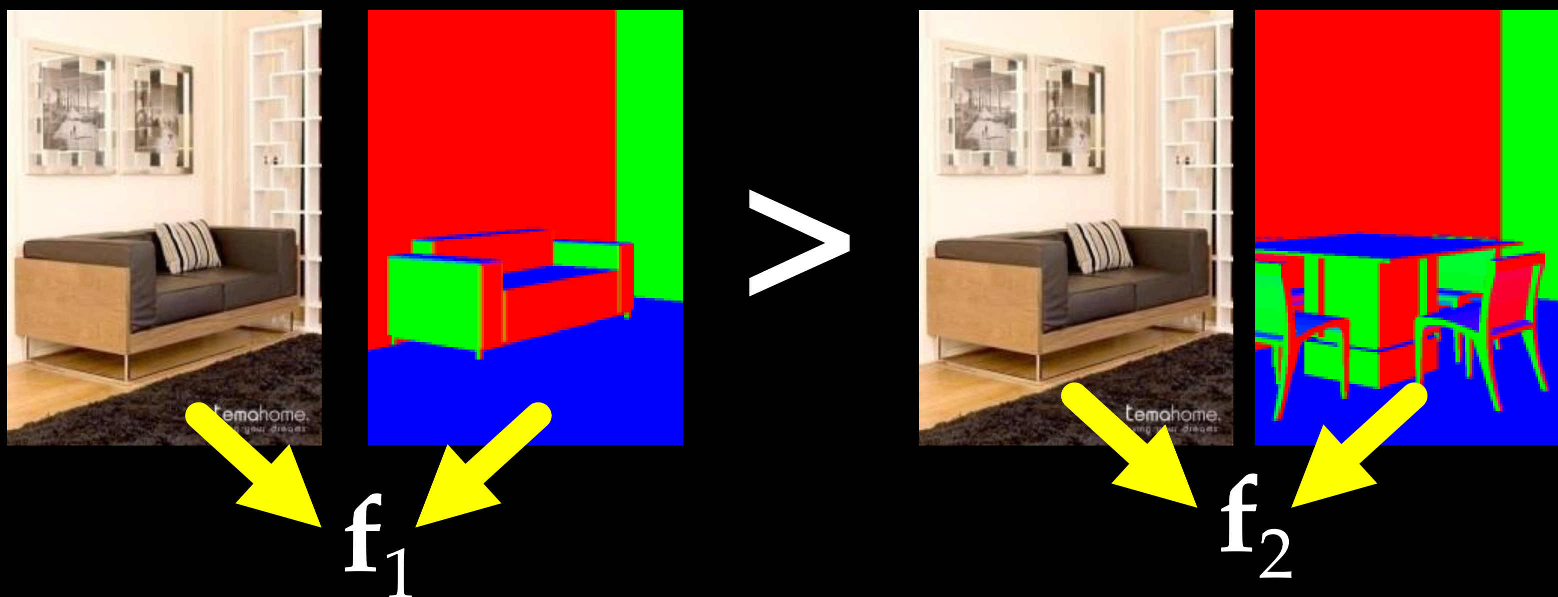
2D-3D Scene Matching

Instead: apply what we already know!



Satkin et al., 2012,2013,2014

2D-3D Scene Matching



Learn w to rank models using ranking svm

Pose and Object Sampling

Render+test enables search over hypotheses generated on the fly

Satkin et al., 2012,2013,2014

Pose and Object Sampling

On average: 5% gain in accuracy

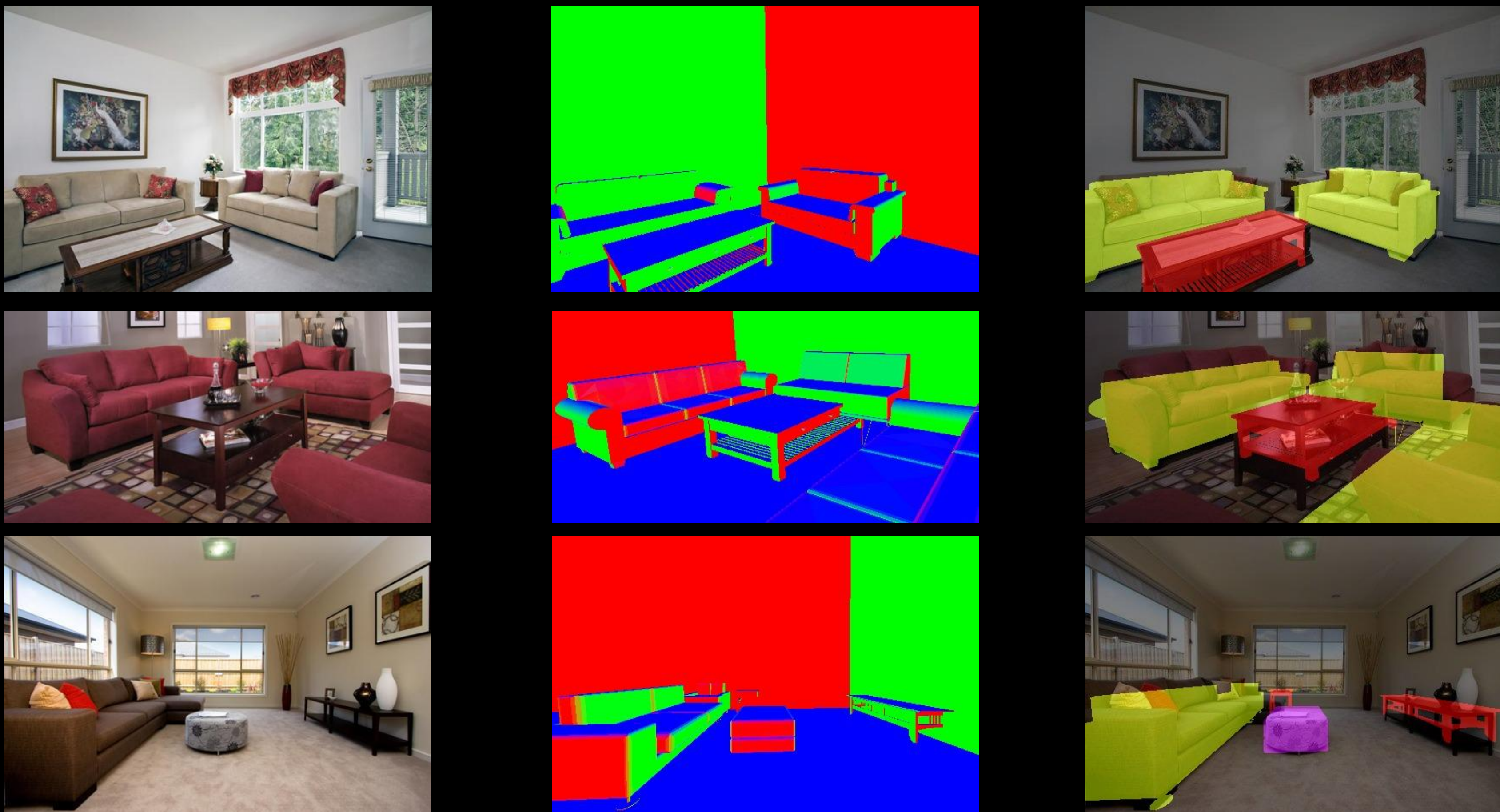
Initial Estimate

Final Estimate

Satkin et al., 2012,2013,2014

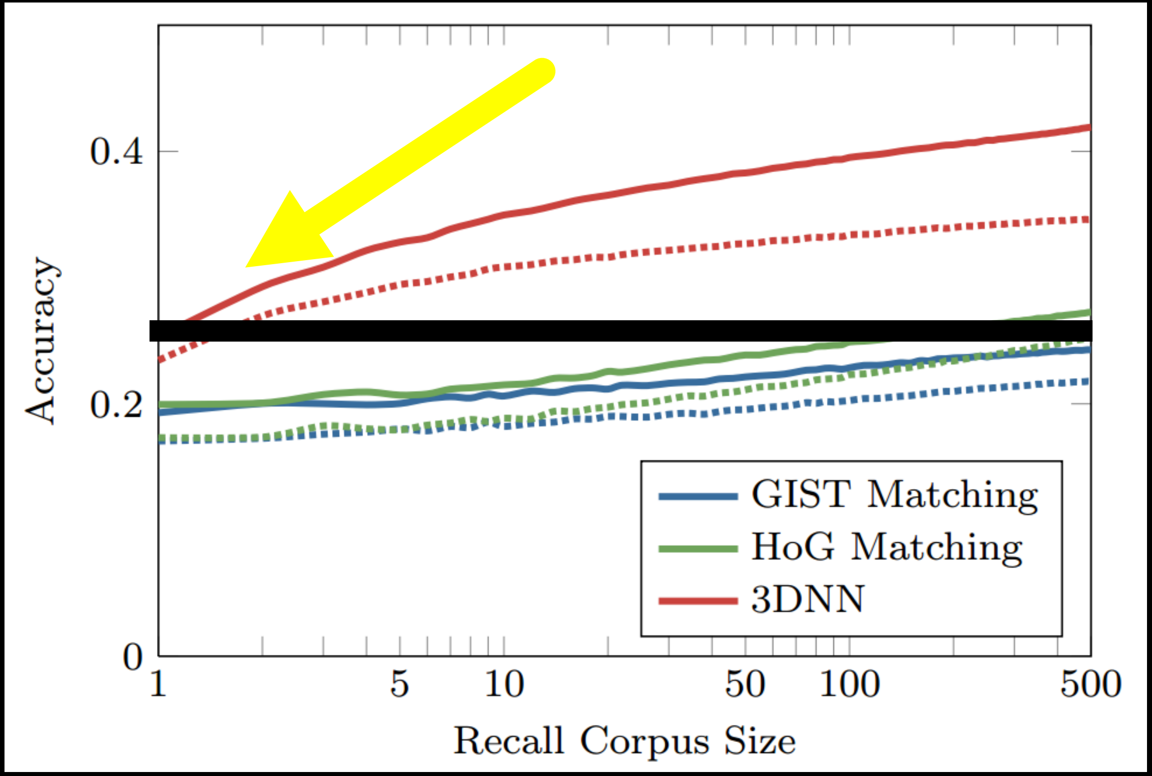
Results

Input | Normals | Semantics



Benefits of 3D

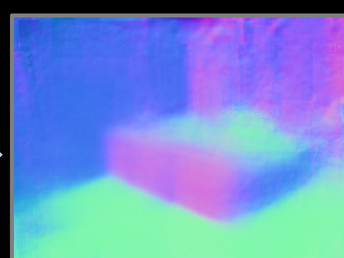
Don't need every viewpoint explicitly!

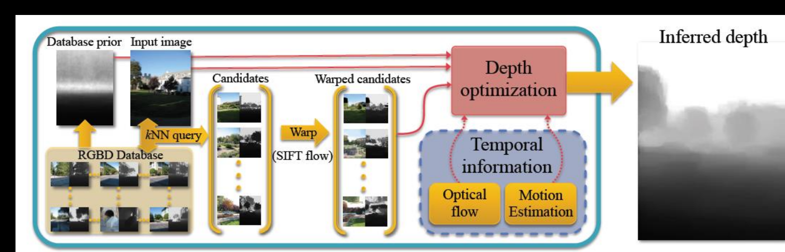


Overview

1. How to use 3D models

2. How to use the Kinect

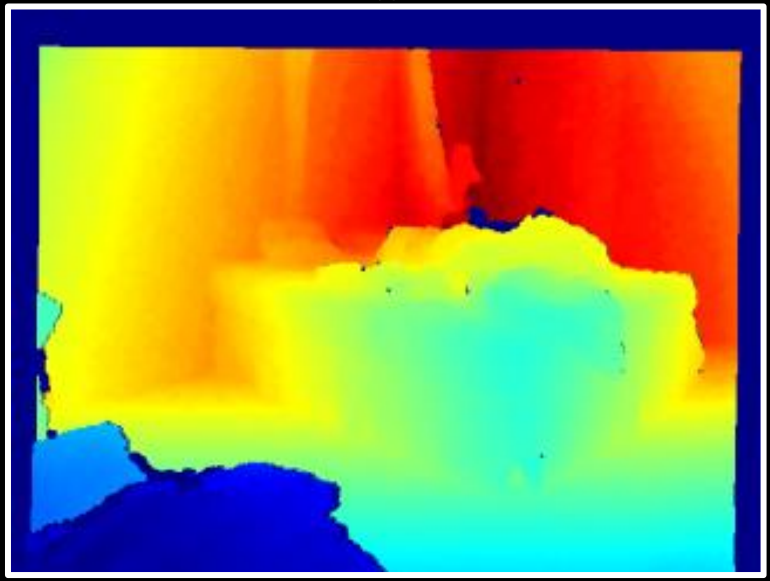




Kinect Data

RGB

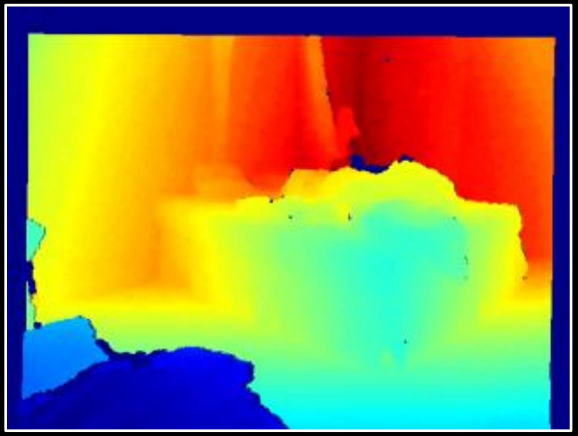
Depth



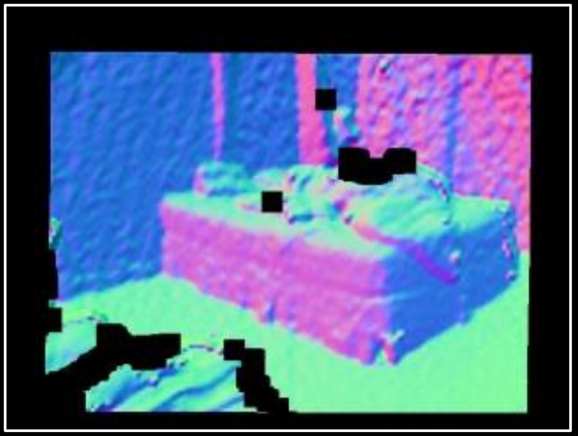
Kinect Data

RGB

Depth



Normals



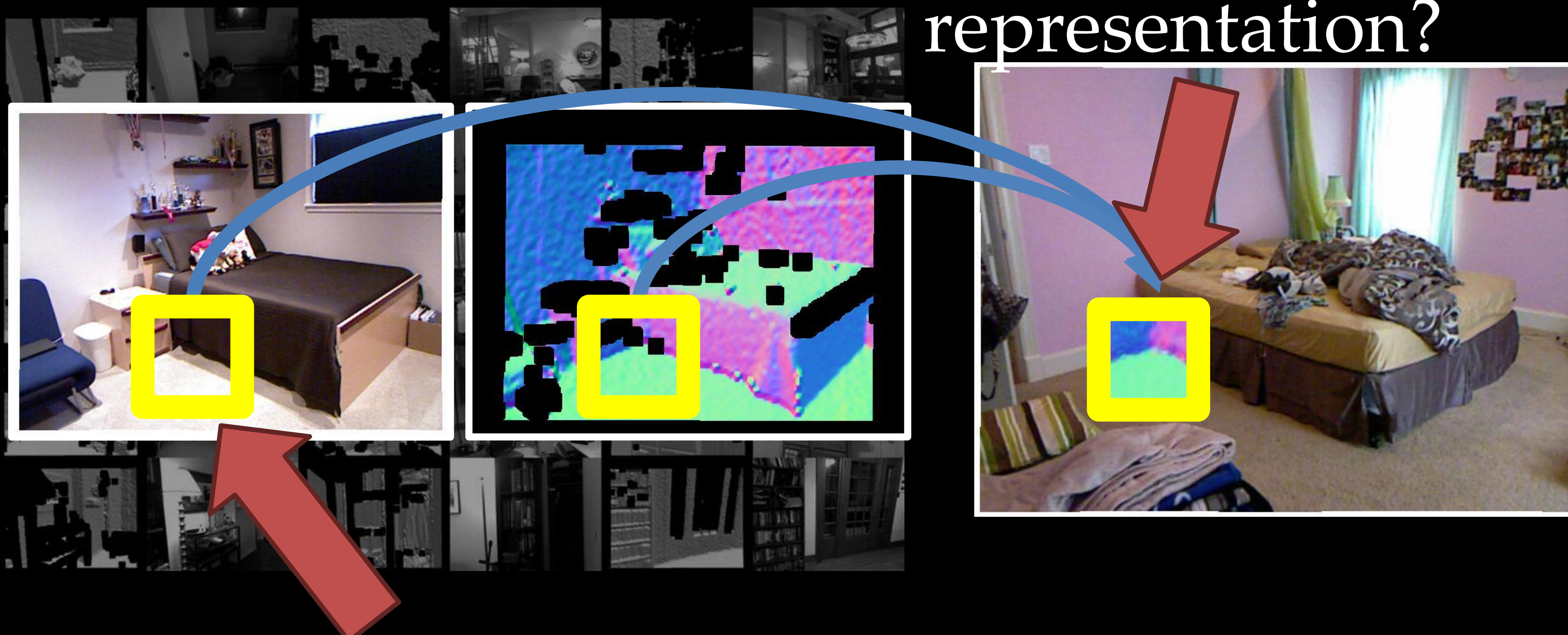
2.5D Data

- Advantages:
  – Corresponding natural color image

- Disadvantages:
  – 2.5D (can't render)
  – Missing data, noise
  – Representations can be difficult to transfer

General Approach

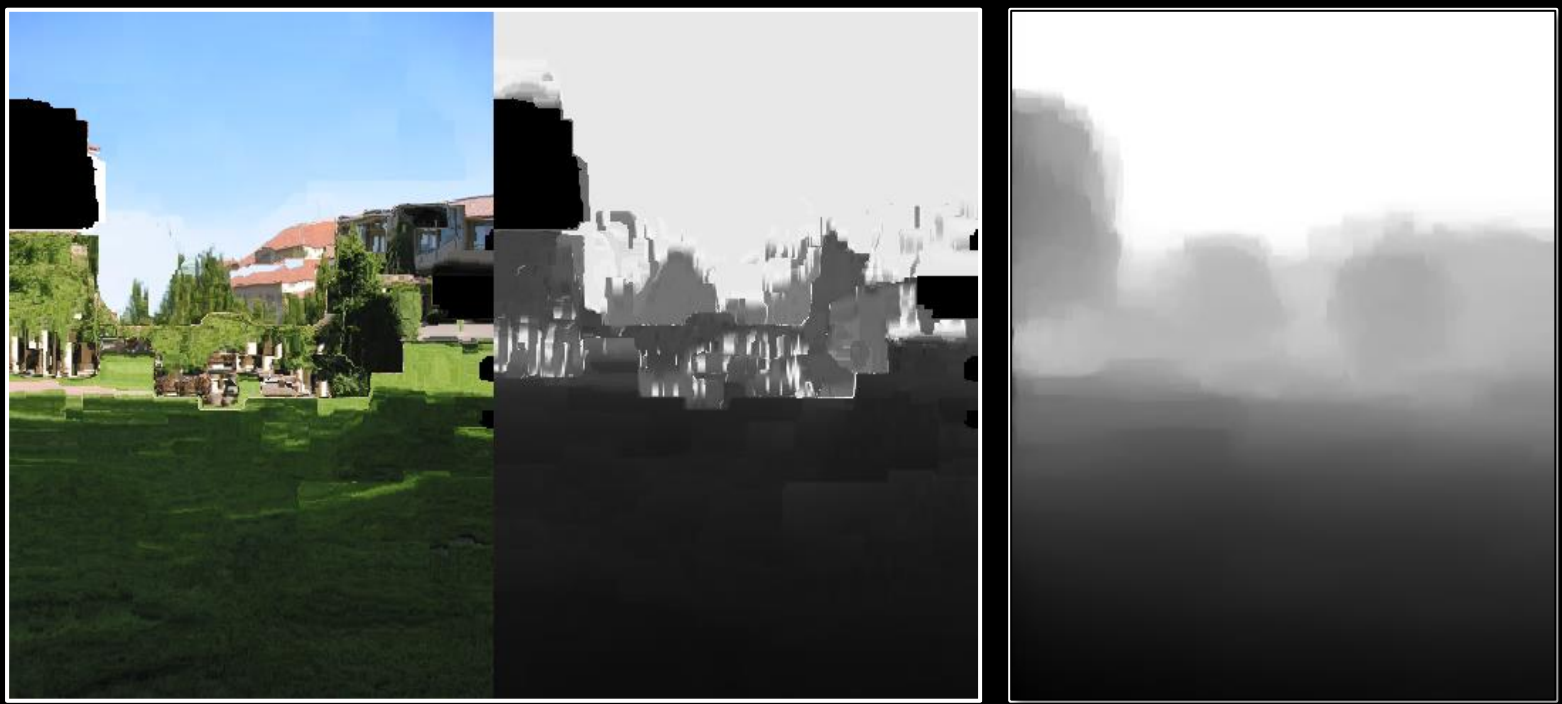
How to transfer representation?



How do we get this correspondence?

Two Approaches

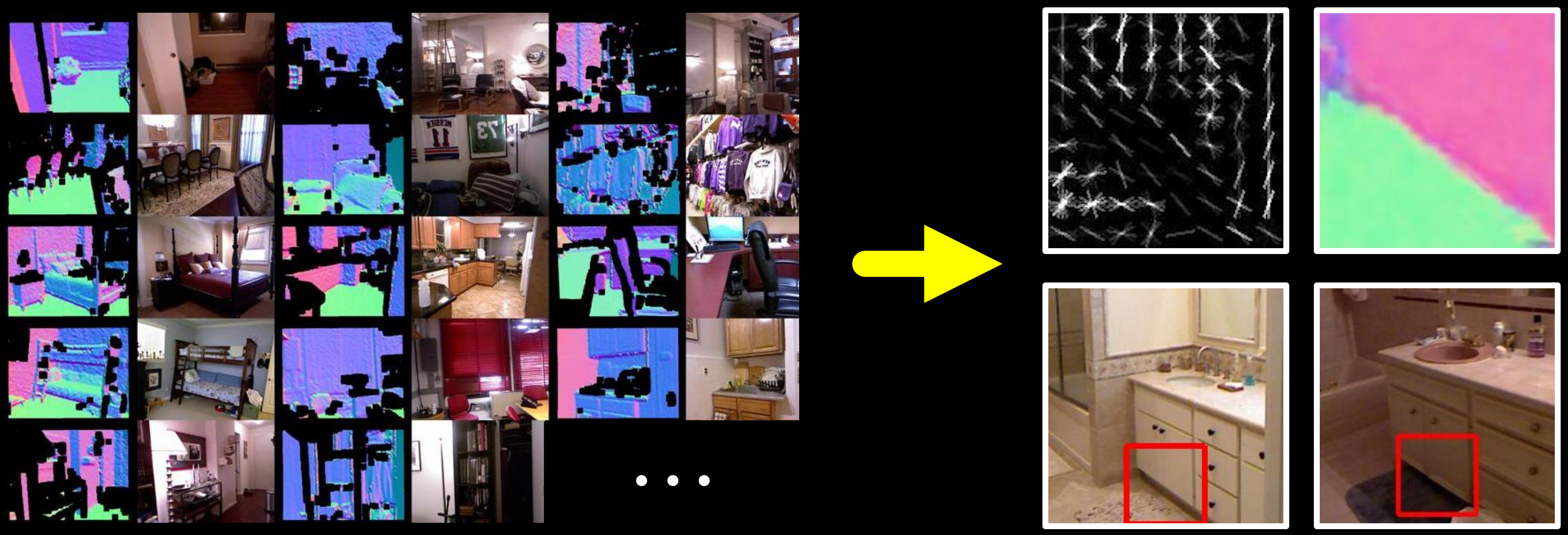
Data-Driven Alignment



Karsch et al., 2012

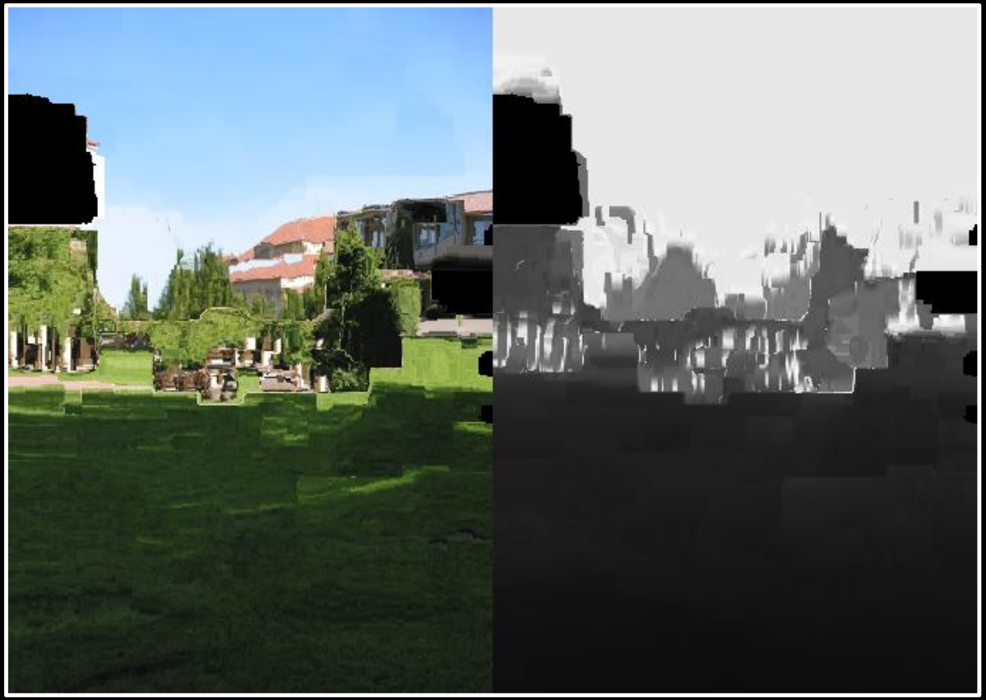
Two Approaches

Clustering + Detection



Fouhey et al., 2013

Data-Driven Alignment



Karsch et al., 2012

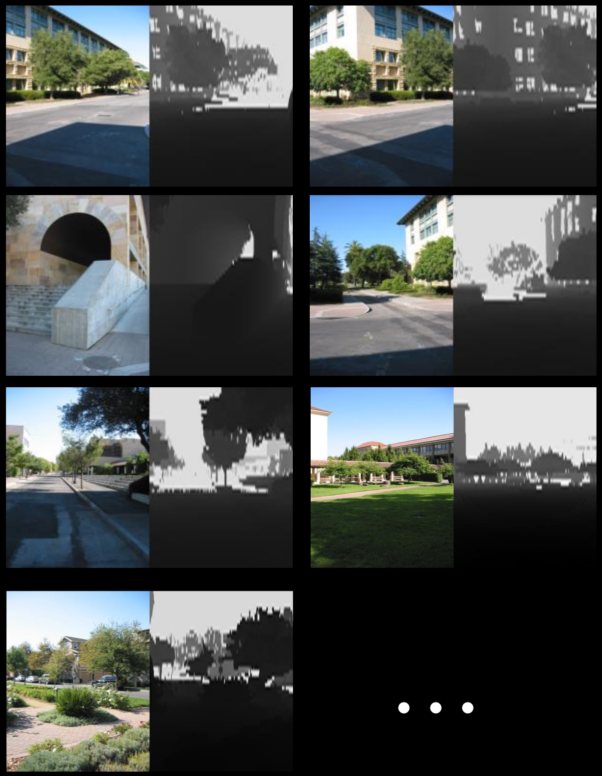
Finding Correspondences

Input

Karsch et al., 2012

Finding Correspondences

Training Set



Input

Karsch et al., 2012

Finding Correspondences

Training Set

Input

Karsch et al., 2012

Finding Correspondences

Training Set

Input

Karsch et al., 2012

Finding Correspondences

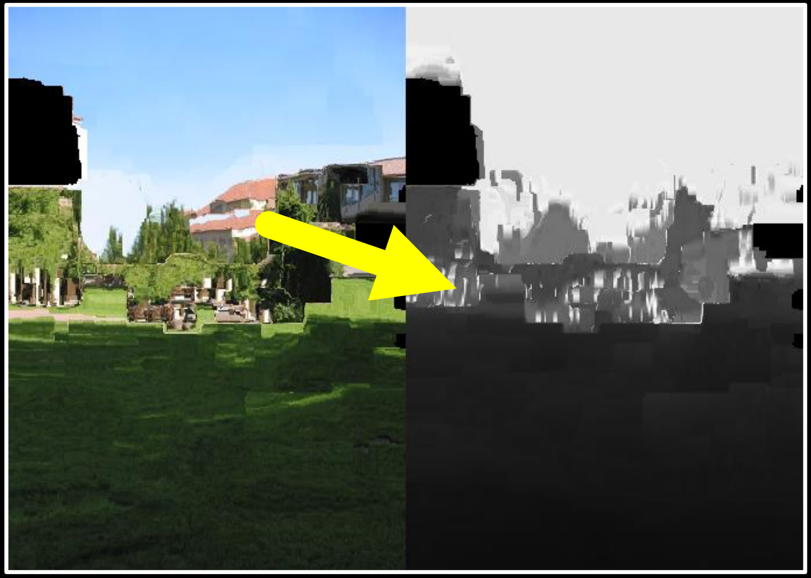
Training Set

Input

Karsch et al., 2012

Finding Correspondences

Candidate 1

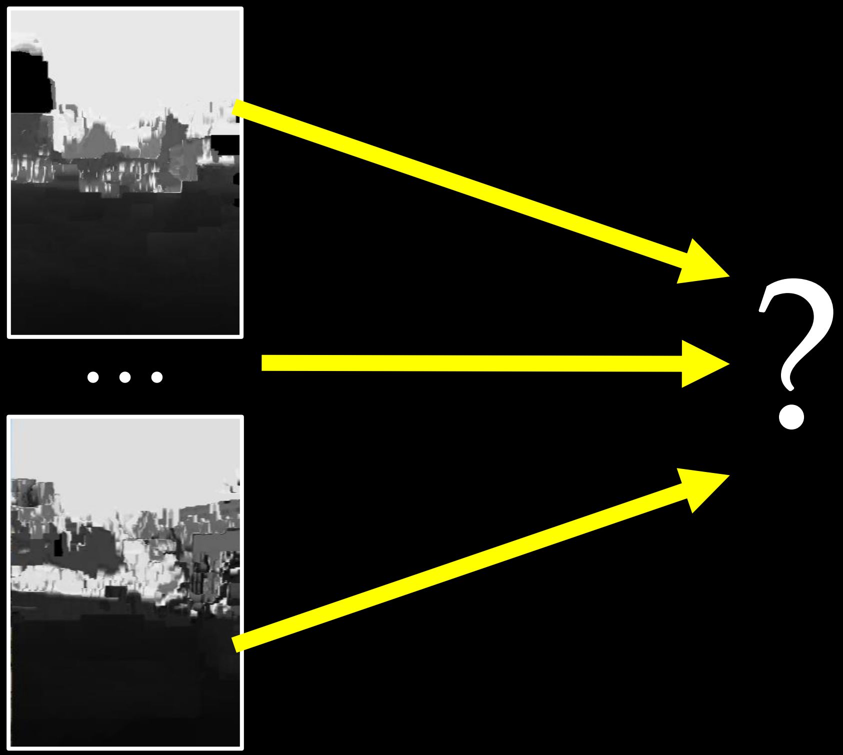


Candidate 2

Karsch et al., 2012

Finding Correspondences

Warped Depths



Karsch et al., 2012; see alternate approach from Liu et al., 2014

Optimizing Depthmaps

$$\sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right.$$

$$\left. + \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1 \right]$$

D_i -Depth being optimized

C_i -Warped depth candidate

Karsch et al., 2012

Optimizing Depthmaps

$$\sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right]$$

$$+ \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1$$

...

Karsch et al., 2012

Optimizing Depthmaps

Enforce depth to match candidates

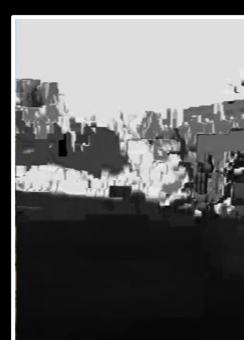
$$\sum_{i\in\text{pixels}} \left[\sum_{C\in\text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma|\nabla D_i - \nabla C_i|_1\right) \right]$$

$$+ \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1$$

Absolute depth

Relative depth

. . .



Karsch et al., 2012

Optimizing Depthmaps

$$\sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right.$$

$$\left. + \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1 \right]$$

Spatial smoothness

Karsch et al., 2012

Optimizing Depthmaps

$$\sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right.$$

$$\left. + \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1 \right]$$

Match the prior

Karsch et al., 2012

Results

Input

True depth

Inferred depth

Karsch et al., 2012

Results

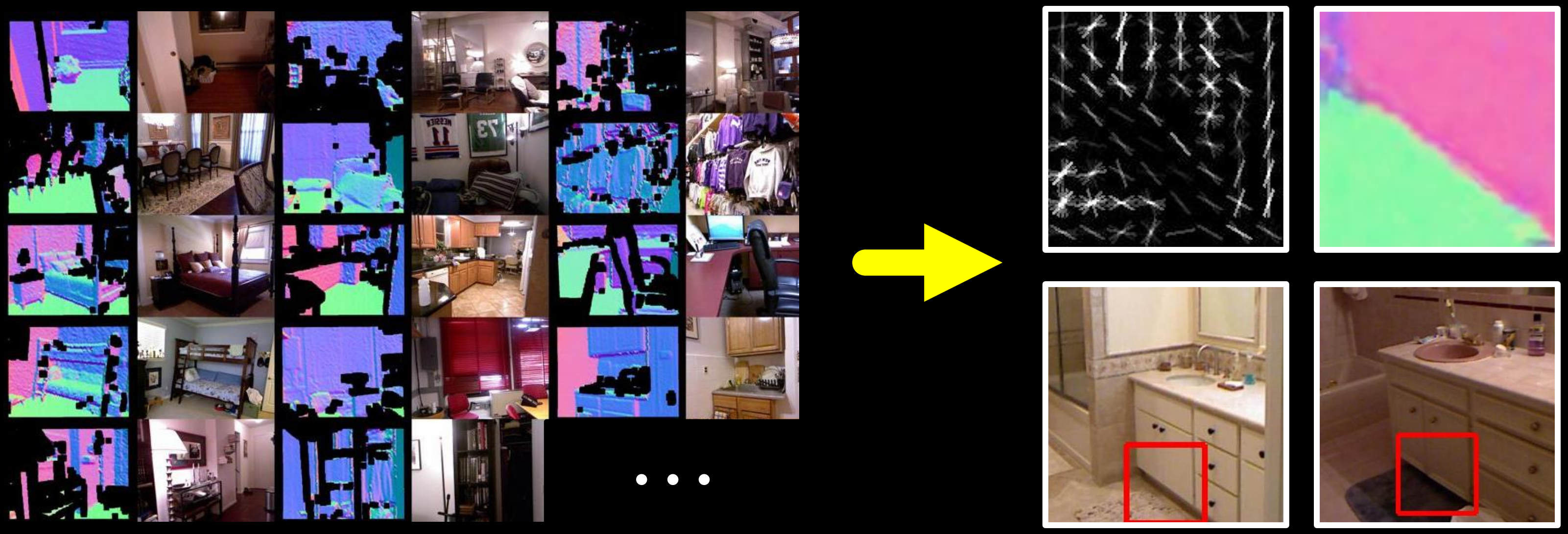
Input

True depth

Inferred depth

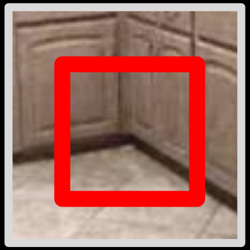
Karsch et al., 2012

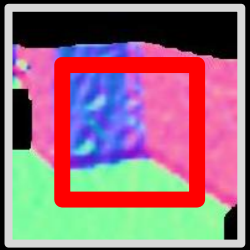
Discriminative Clustering + Detection



Fouhey et al., 2013

Goal





Visually Discriminative

Image

Geometrically Informative

Surface Normals

Fouhey et al., 2013

Goal

Learn from large-scale RGBD Data

NYU Depth v2, Silberman et al., 2012

Approach

Train time: discriminative clustering w/3D

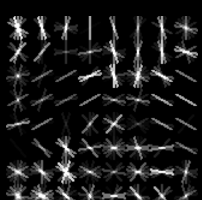
Fouhey et al., 2013

Objective

$$\min_{\mathbf{y},\mathbf{w},\mathbf{N}} R(\mathbf{w}) + \sum_{i=1}^{m} \Big[c_2 L(\mathbf{w}, \mathbf{N}, \mathbf{x}_i^A, y_i) + c_1 y_i \Delta(\mathbf{N}, \mathbf{x}_i^G) \Big]$$

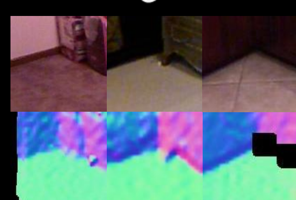
Primitive

$\mathbf{w}$



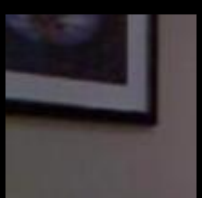
$\mathbf{N}$

$\mathbf{y}$



Patch

$\mathbf{x}_i^A$



$\mathbf{x}_i^G$

Fouhey et al., 2013

Objective

Misclassification loss

$$\min_{\mathbf{y},\mathbf{w},\mathbf{N}} R(\mathbf{w}) + \sum_{i=1}^{m} \left[c_2 L(\mathbf{w}, \mathbf{N}, \mathbf{x}_i^A, y_i) + c_1 y_i \Delta(\mathbf{N}, \mathbf{x}_i^G) \right]$$

Primitive

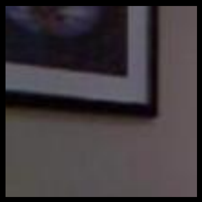
$\mathbf{w}$

$\mathbf{N}$

$\mathbf{y}$

Patch

$\mathbf{x}_i^A$



$\mathbf{x}_i^G$

Fouhey et al., 2013

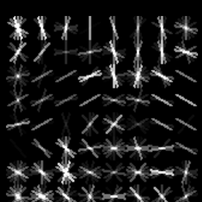
Objective

Regularization

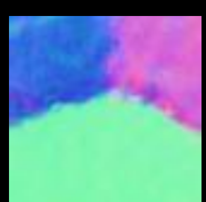
$$\min_{\mathbf{y},\mathbf{w},\mathbf{N}} R(\mathbf{w}) + \sum_{i=1}^{m} \left[c_2 L(\mathbf{w}, \mathbf{N}, \mathbf{x}_i^A, y_i) + c_1 y_i \Delta(\mathbf{N}, \mathbf{x}_i^G) \right]$$

Primitive

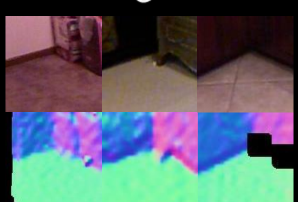
$\mathbf{w}$



$\mathbf{N}$

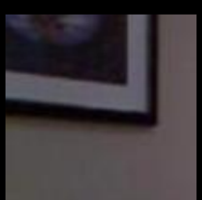


$\mathbf{y}$



Patch

$\mathbf{x}_i^A$



$\mathbf{x}_i^G$

Fouhey et al., 2013

Objective

Ensure geometric consistency

$$\min_{\mathbf{y},\mathbf{w},\mathbf{N}} R(\mathbf{w}) + \sum_{i=1}^{m} \left[c_2 L(\mathbf{w}, \mathbf{N}, \mathbf{x}_i^A, y_i) + c_1 y_i \Delta(\mathbf{N}, \mathbf{x}_i^G) \right]$$

Primitive

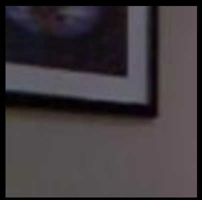
$\mathbf{w}$

$\mathbf{N}$

$\mathbf{y}$

Patch

$\mathbf{x}_i^A$



$\mathbf{x}_i^G$

Fouhey et al., 2013

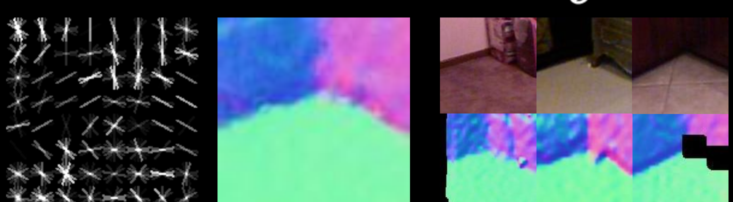
Objective

Solved with iterative method similar to block-coordinate-descent.
Include min-membership constraint

$$\min_{\mathbf{y},\mathbf{w},\mathbf{N}} R(\mathbf{w}) + \sum_{i=1}^{m} \left[c_2 L(\mathbf{w}, \mathbf{N}, \mathbf{x}_i^A, y_i) + c_1 y_i \Delta(\mathbf{N}, \mathbf{x}_i^G) \right]$$

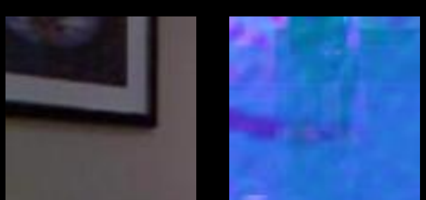
Primitive

$\mathbf{w}$ $\quad$ $\mathbf{N}$ $\quad$ $\mathbf{y}$



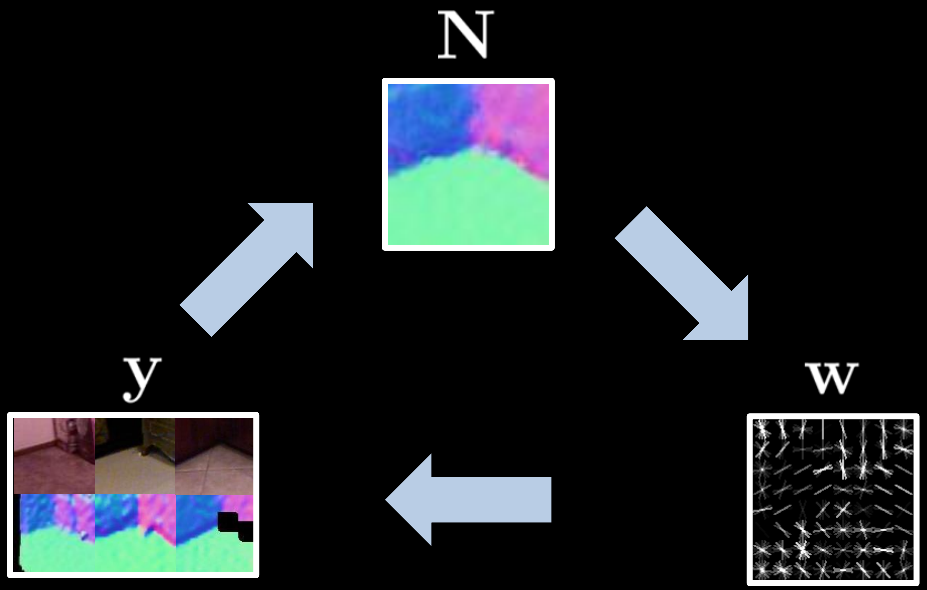
Patch

$\mathbf{x}_i^A$ $\quad$ $\mathbf{x}_i^G$



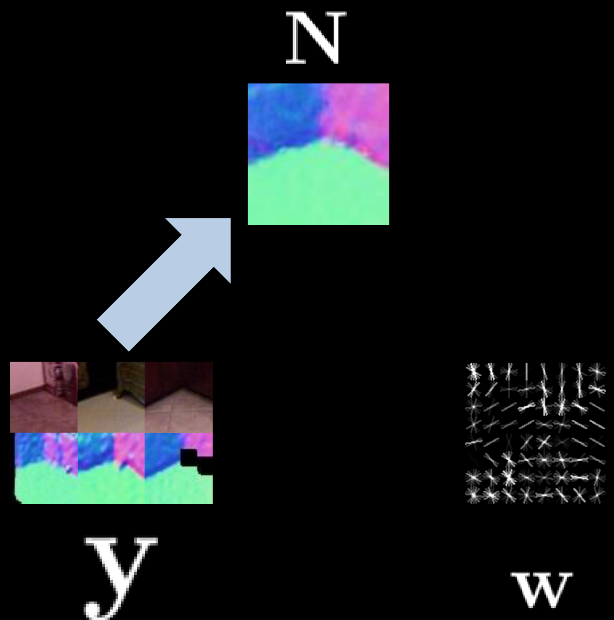
Fouhey et al., 2013

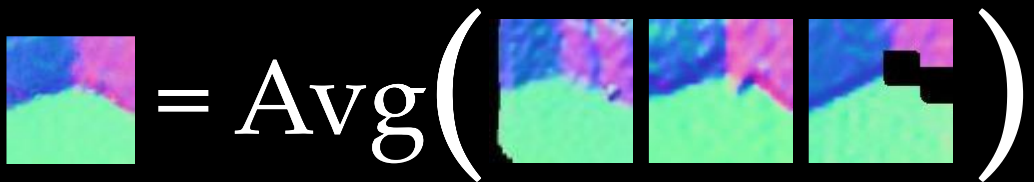
Iterative Procedure



Fouhey et al., 2013

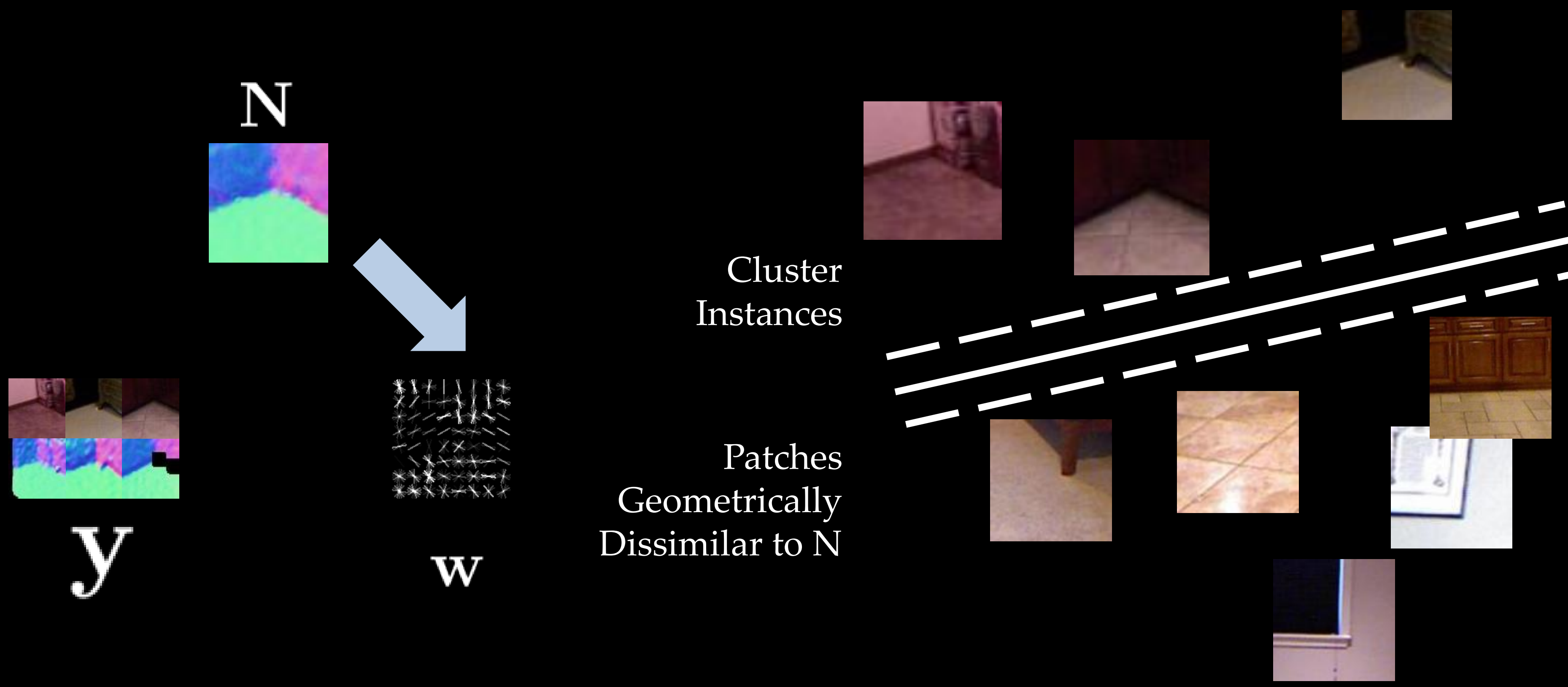
Iterative Procedure





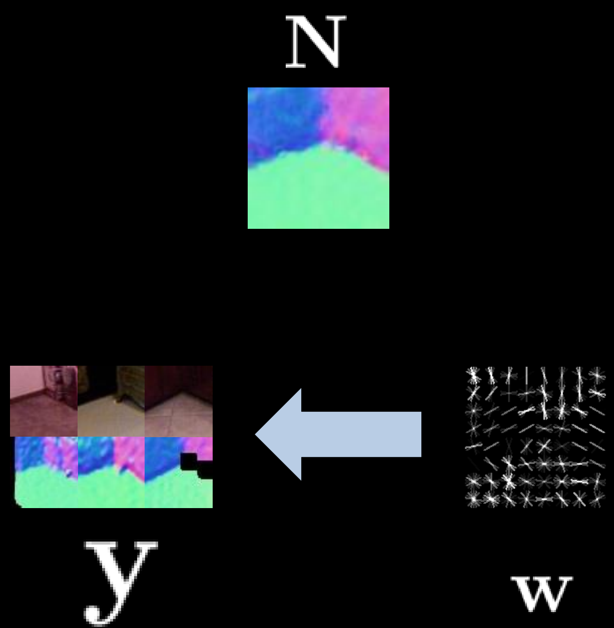
Fouhey et al., 2013

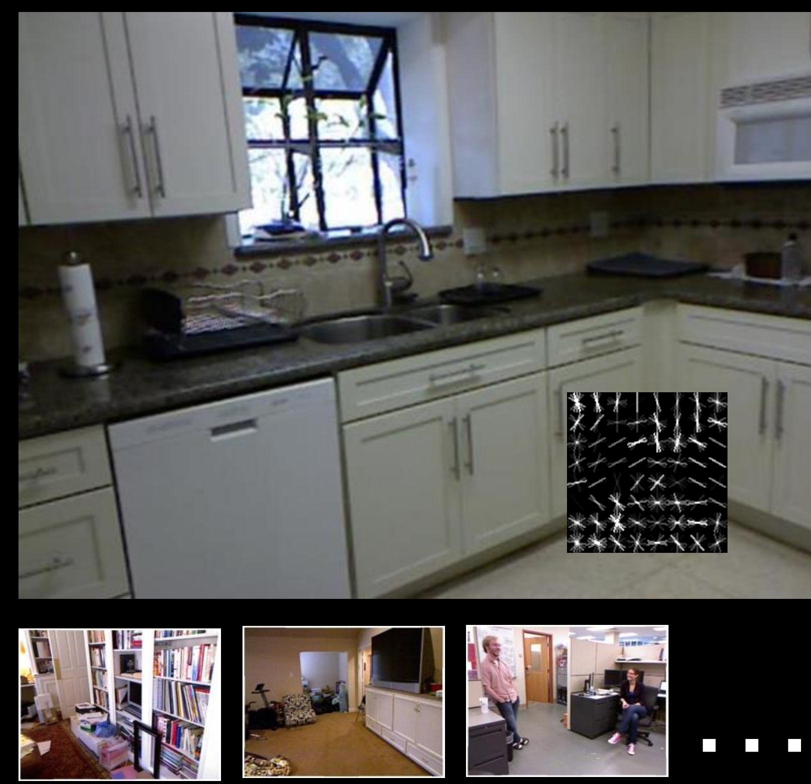
Iterative Procedure



Fouhey et al., 2013

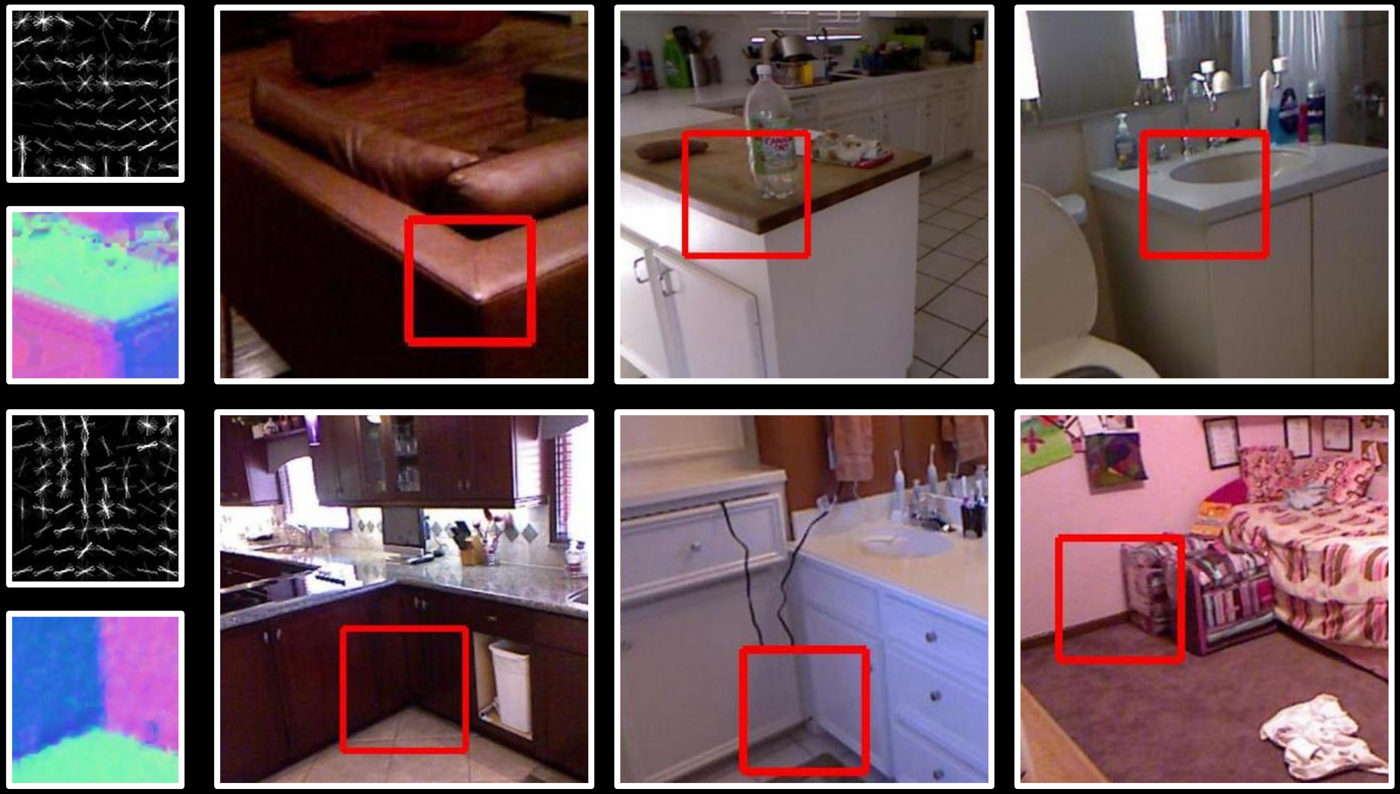
Iterative Procedure





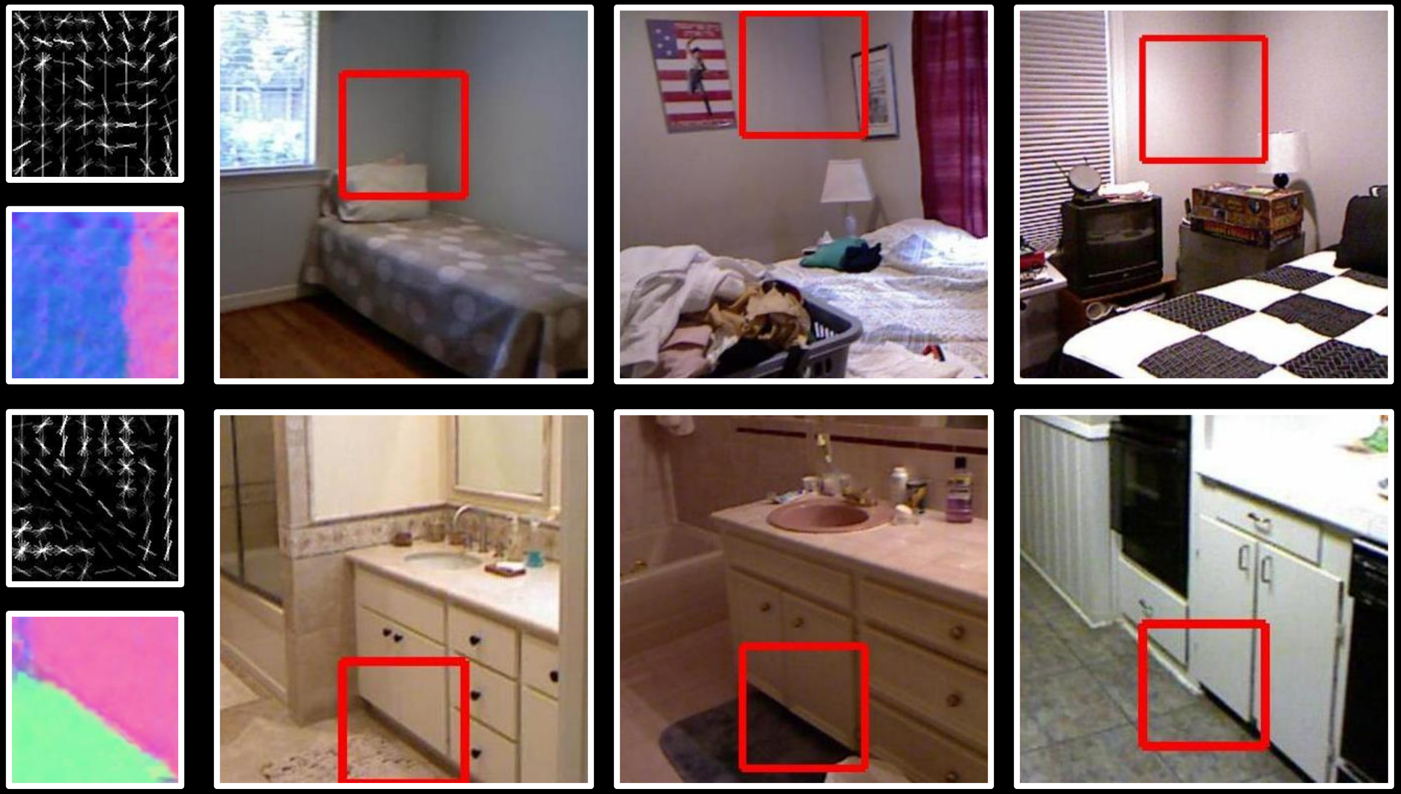
Fouhey et al., 2013

Primitives



Fouhey et al., 2013

Primitives



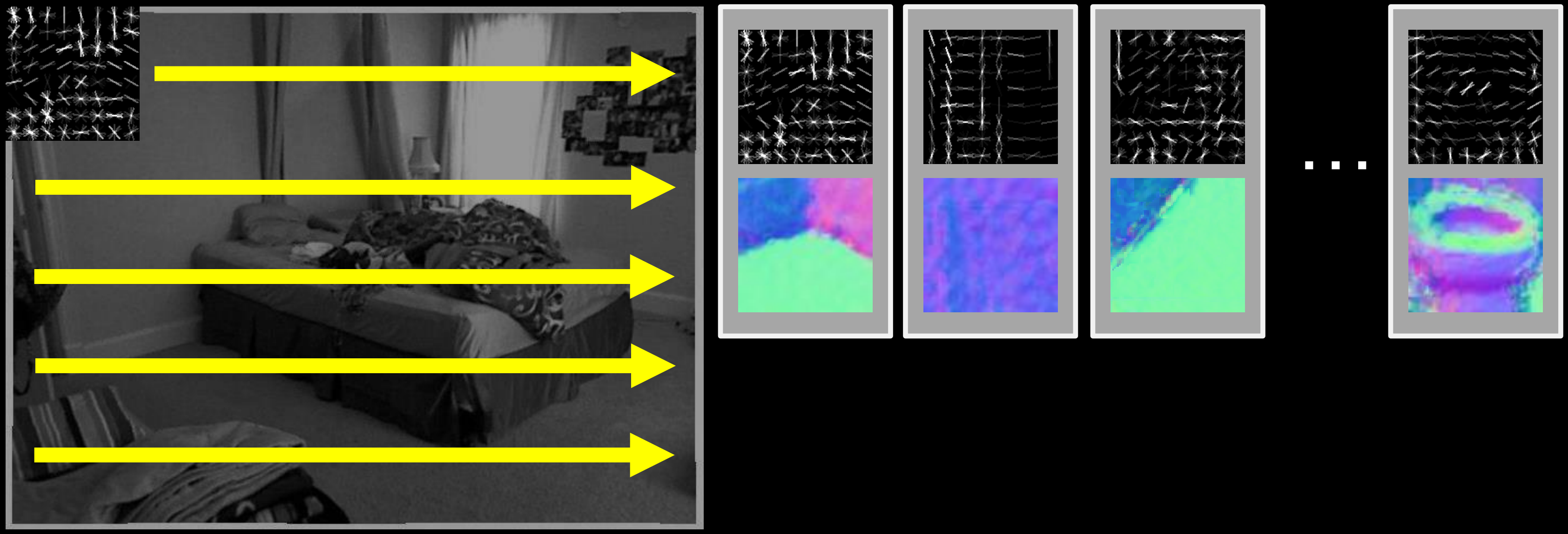
Fouhey et al., 2013

Primitives

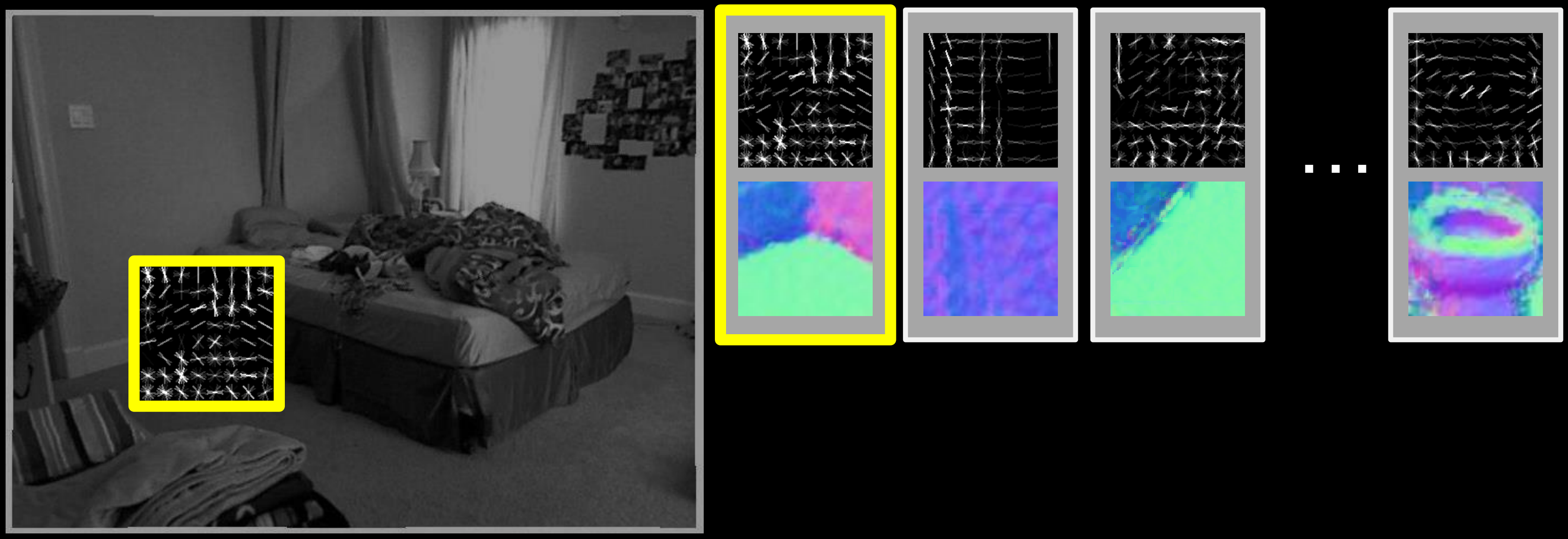
Fouhey et al., 2013

Test-time Correspondence

Correspondence via detection

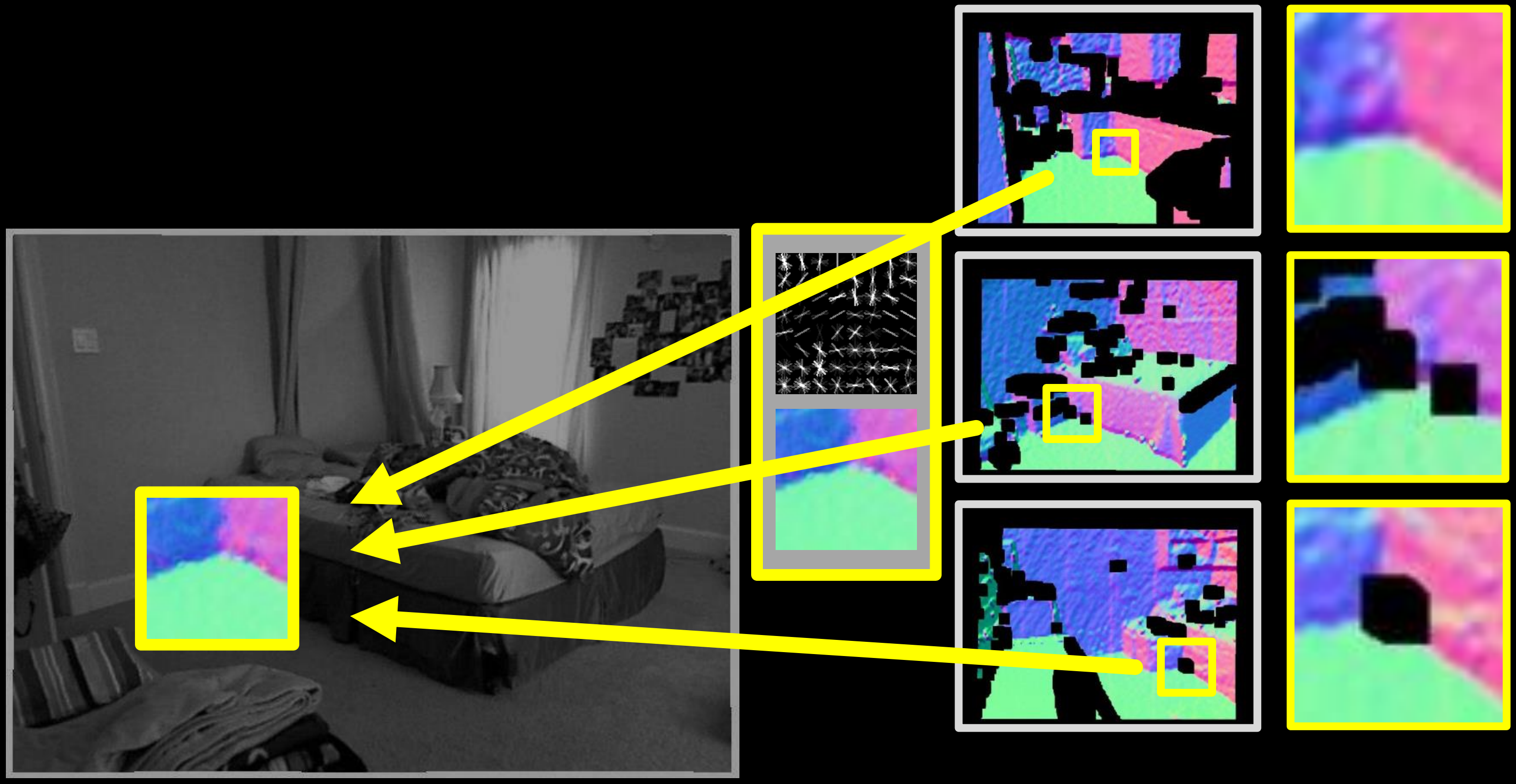


Representation Transfer



Fouhey et al., 2013

Representation Transfer

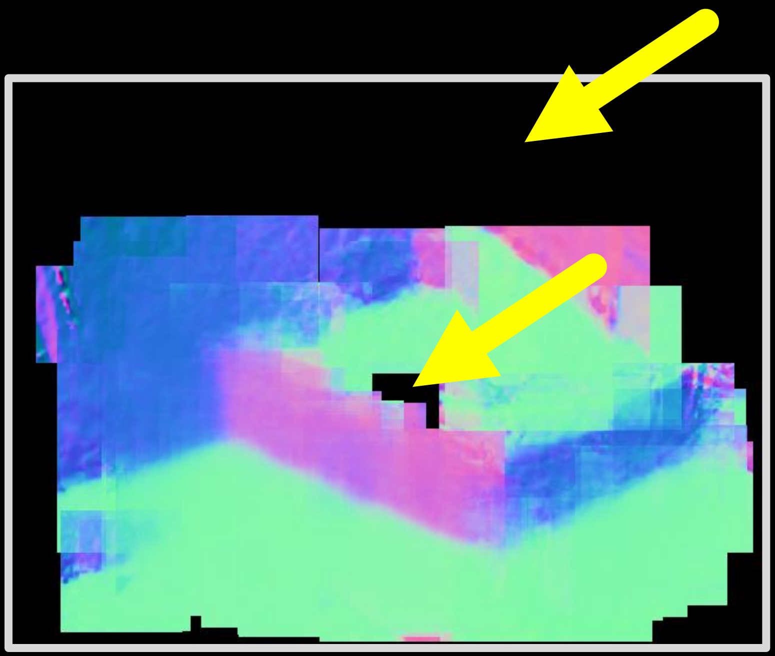


...

Fouhey et al., 2013

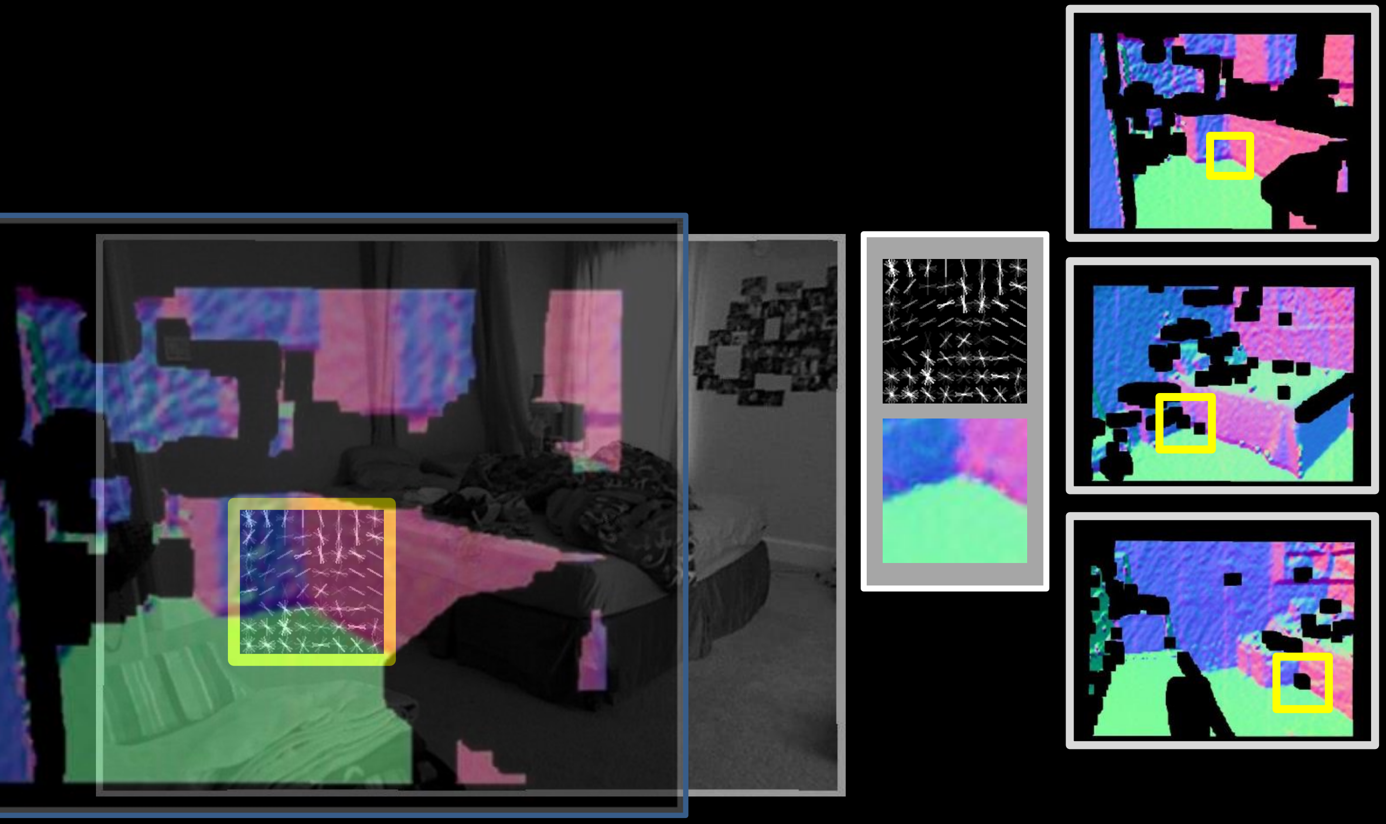
Representation Transfer

Overlaps resolved with averaging



Fouhey et al., 2013

Representation Transfer



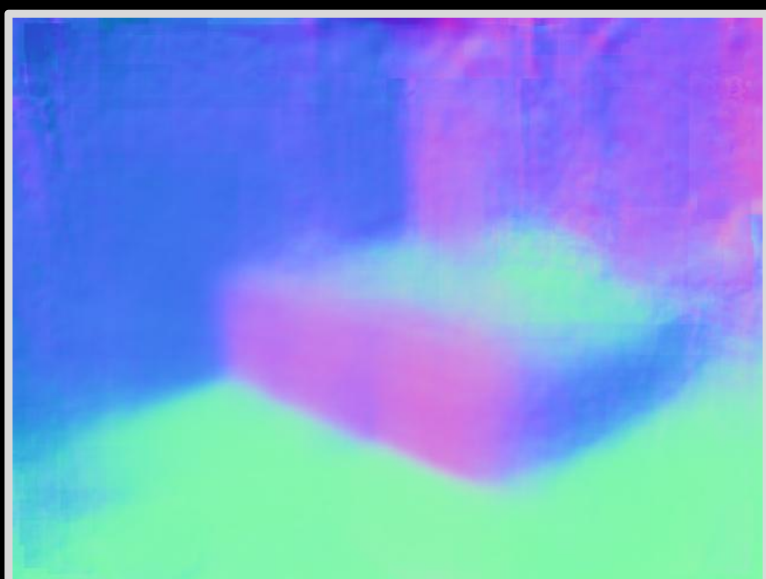
Fouhey et al., 2013

Representation Transfer

Fouhey et al., 2013

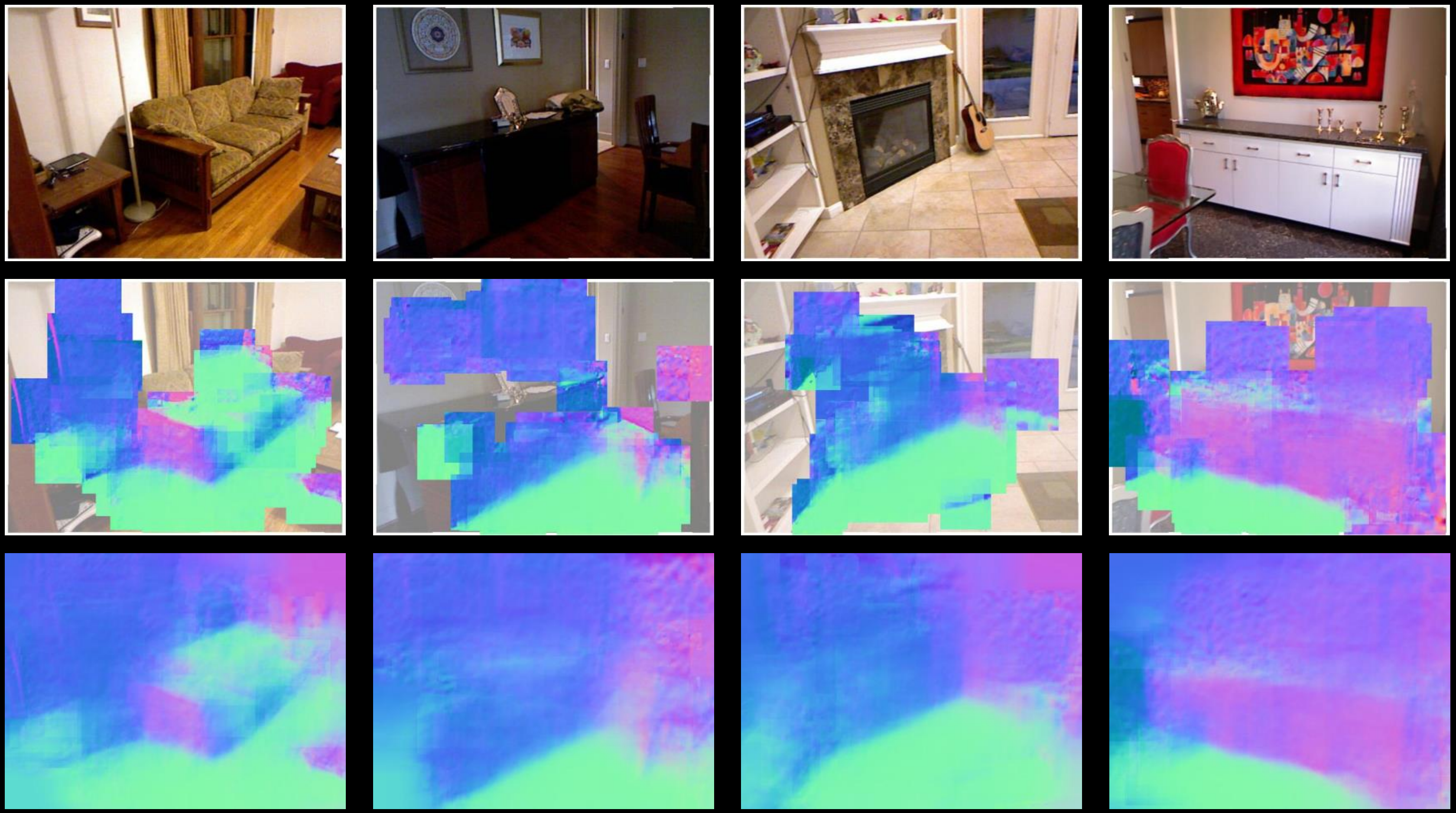
Representation Transfer

Overlaps resolved with averaging



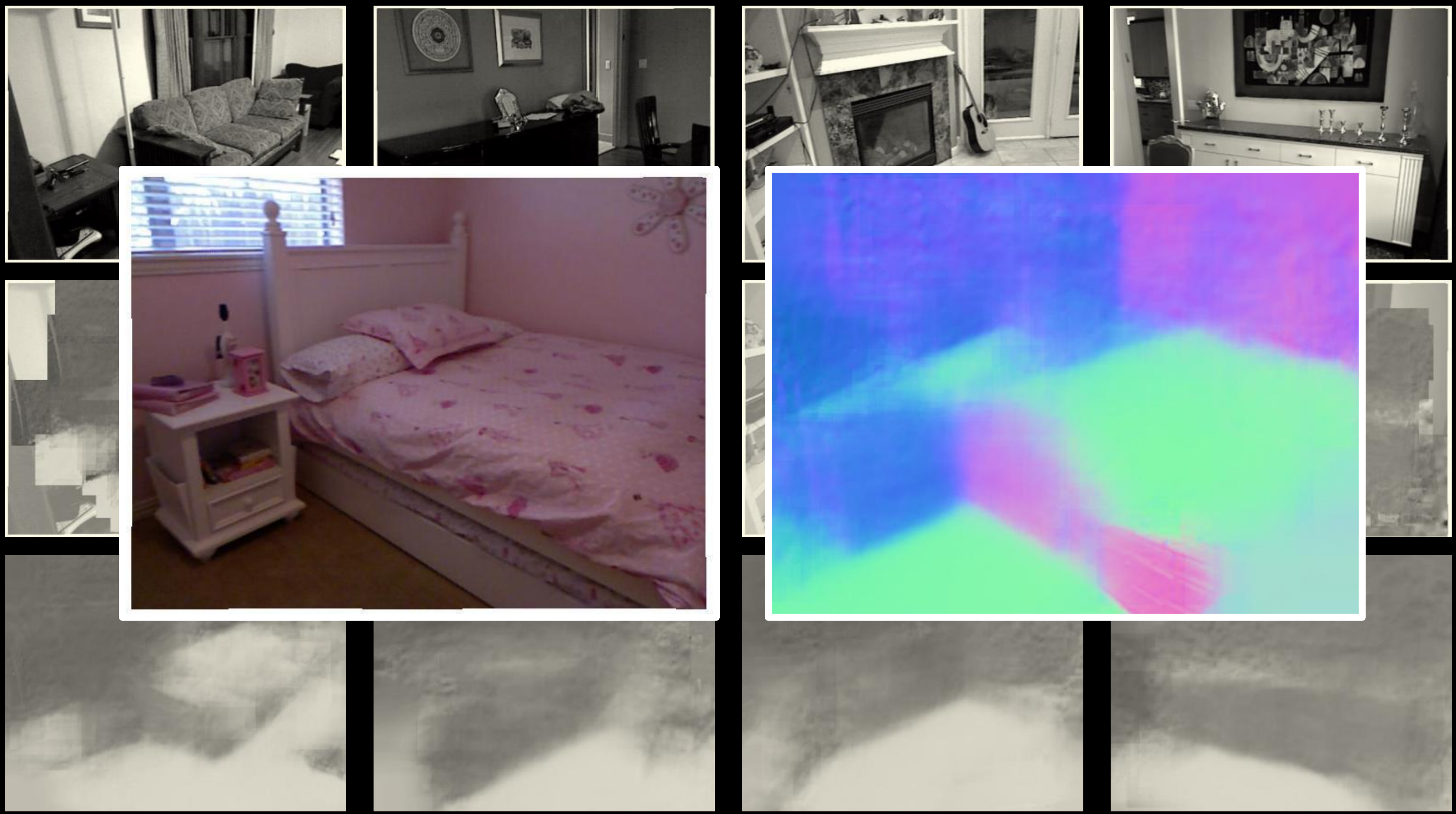
Fouhey et al., 2013

Results



Fouhey et al., 2013

Results

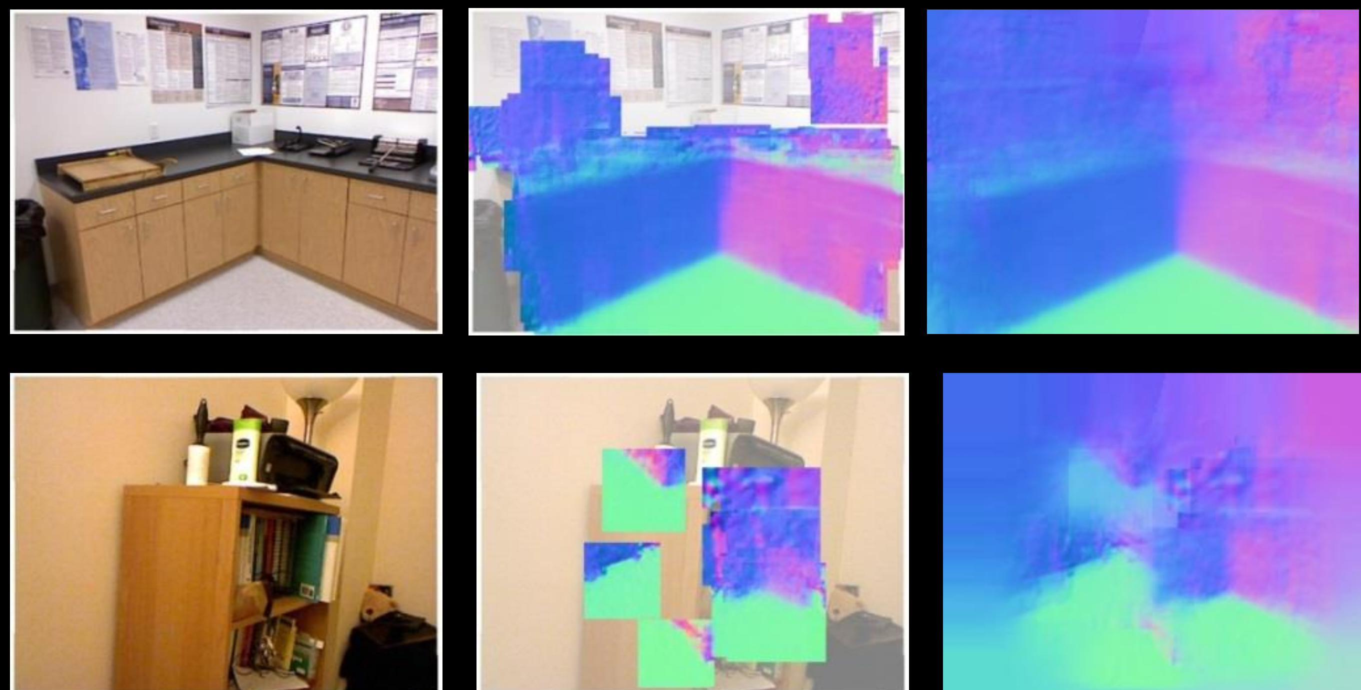


Fouhey et al., 2013

Confidences

Most Confident Result

Least Confident Result



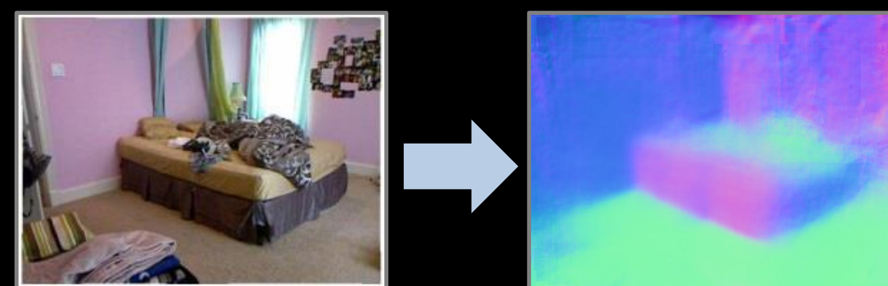
Fouhey et al., 2013

Conclusions

Introduced Data-Driven 3D Scene Understanding

Full 3D Models

RGBD Data



Two Main Problems:

1. Correspondence
2. Representation Transfer

Future Directions

- How do you get the best of 2.5D and 3D? (see Guo and Hoiem 2013)

- How do you incorporate constraints in data-driven techniques?

Resources

(See tutorial website for links + more data/code + slides)

Survey Books

- D. Hoiem, S. Savarese. *Representations and Techniques for 3D Object Recognition and Scene Interpretation*. Morgan & Claypool, 2011.

  (link on website)

Available Kinect Datasets

- RMRC (NYU + SUN3D)
- NYU v2:

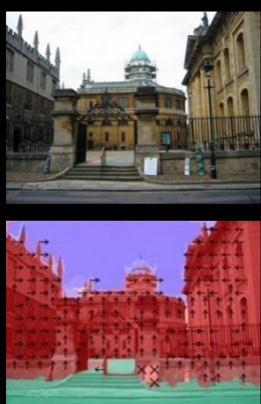
  1449 Pairs + semantic labels + raw videos
- SUN3D

  415 Sequences in large spaces + raw videos
- Berkeley 3D Object

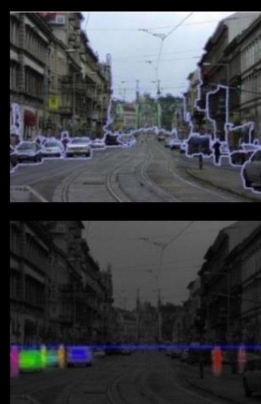
  849 images + bounding boxes
- MSR-V3D

  177 sequences

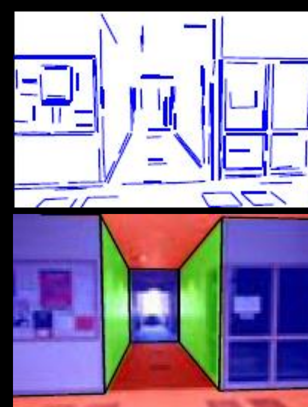
Available Code



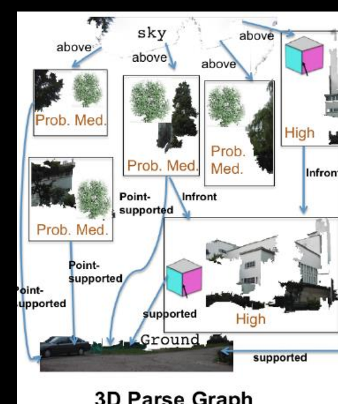
Region labels



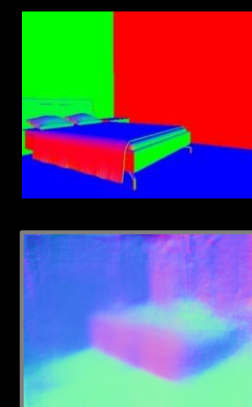
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



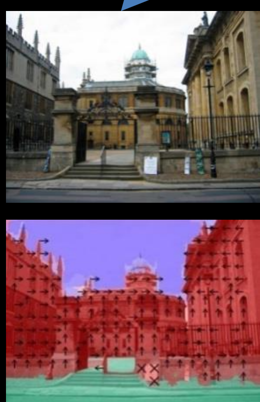
Volumetric + functional constraints



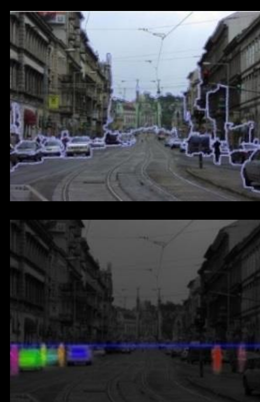
Data-driven 3D

Available Code

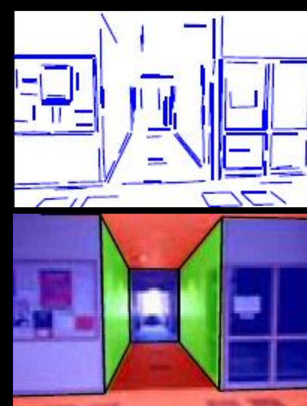
Hoiem et al., Geometric Context,
Saxena et al., Make 3D



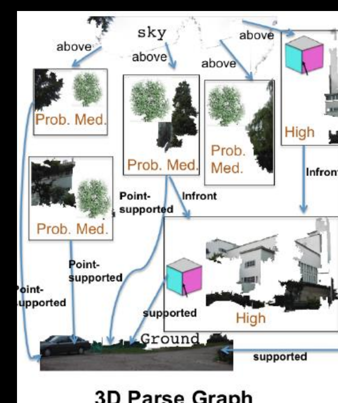
Region labels



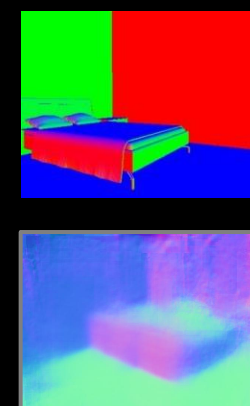
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



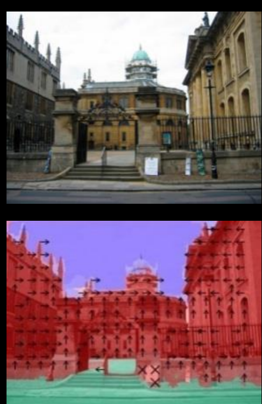
Volumetric + functional constraints



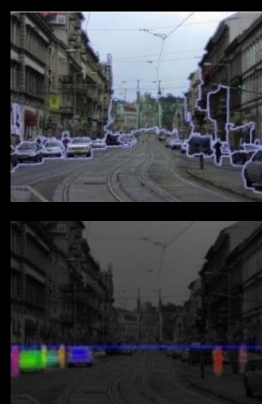
Data-driven 3D

Available Code

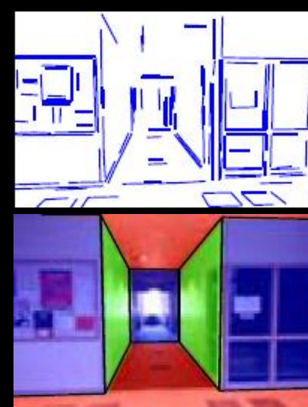
Hoiem et al., Occlusion boundaries
Hoiem et al., Putting objects in perspective



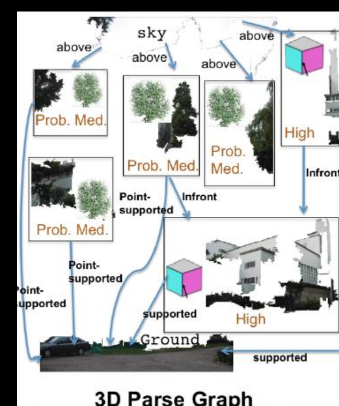
Region labels



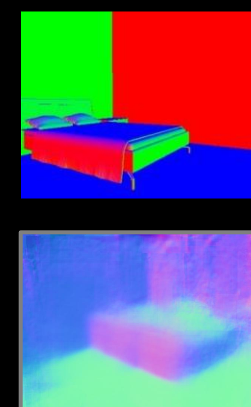
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



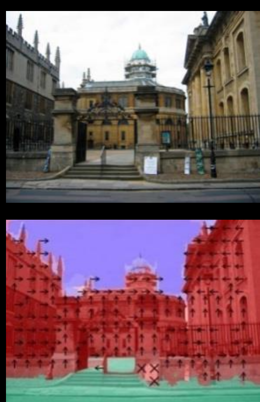
Volumetric + functional constraints



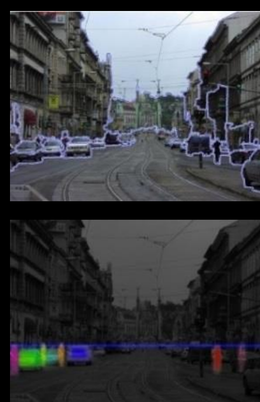
Data-driven 3D

Available Code

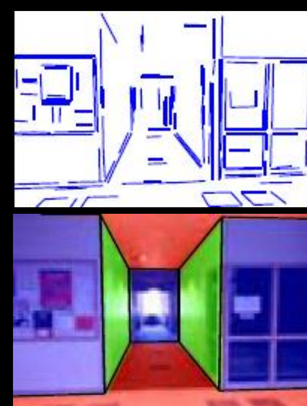
Lee et al., Orientation Maps
Hedau et al., Room-fitting



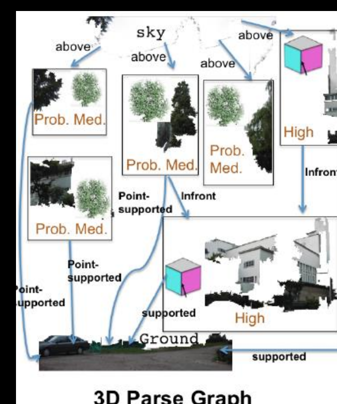
Region labels



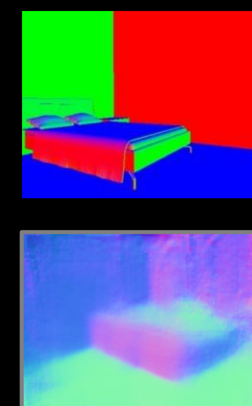
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



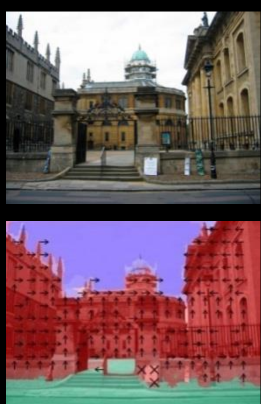
Volumetric + functional constraints



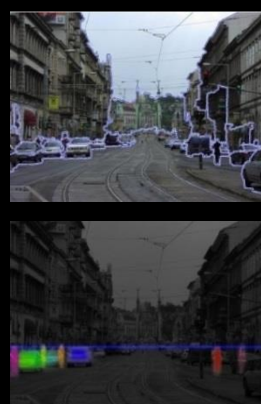
Data-driven 3D

Available Code

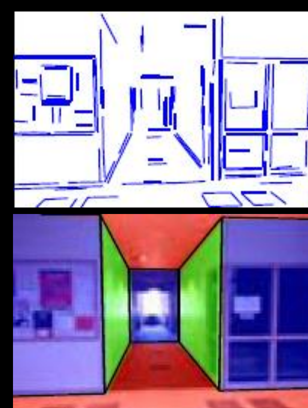
Gupta et al., Blocks World
Choi et al., Geometric Phrases



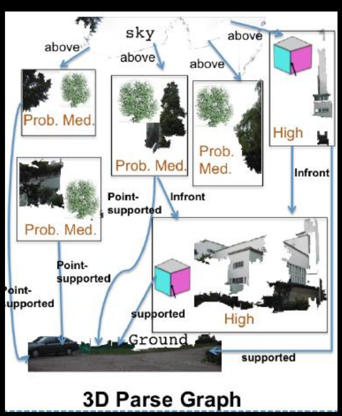
Region labels



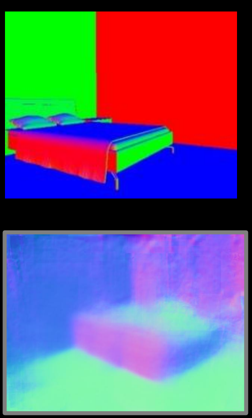
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



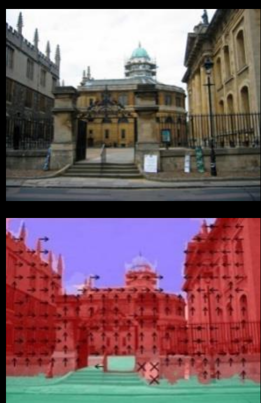
Volumetric + functional constraints



Data-driven 3D

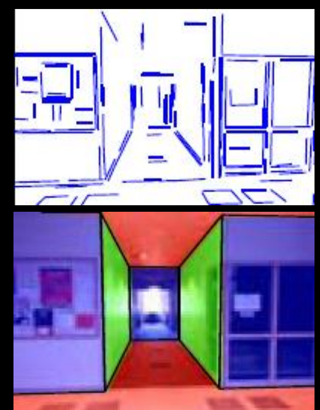
Available Code

Karsch et al., Depth-Transfer
Fouhey et al., Data-Driven 3D Primitives
Aubrey et al., Seeing 3D Chairs

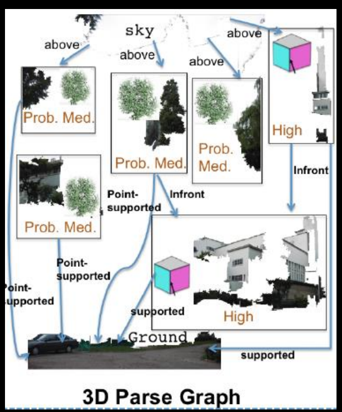


Region labels

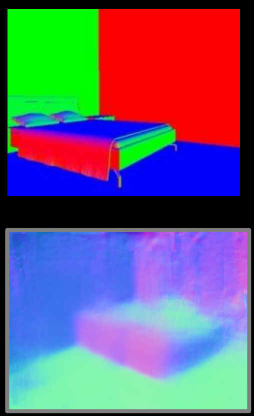
+ Boundaries and objects



Stronger geometric constraints from domain knowledge



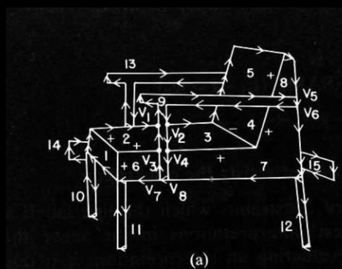
Volumetric + functional constraints



Data-driven 3D

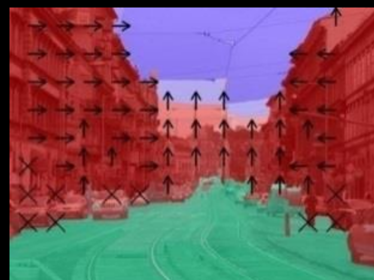
Thank You

Martial



Introduction,
Applications,
History

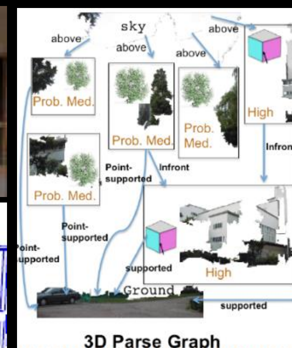
Derek

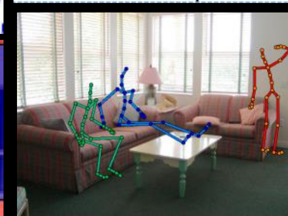


Region labels
+Boundaries
+Objects

Stronger
geometric
constraints

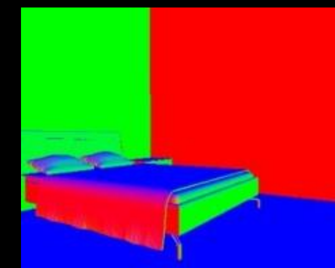
Abhinav

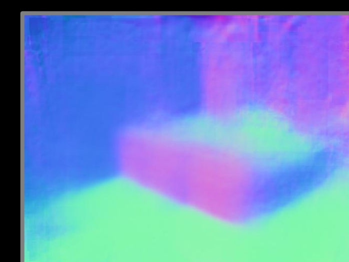




Volumetric +
Functional
Constraints

David





Data-Driven 3D