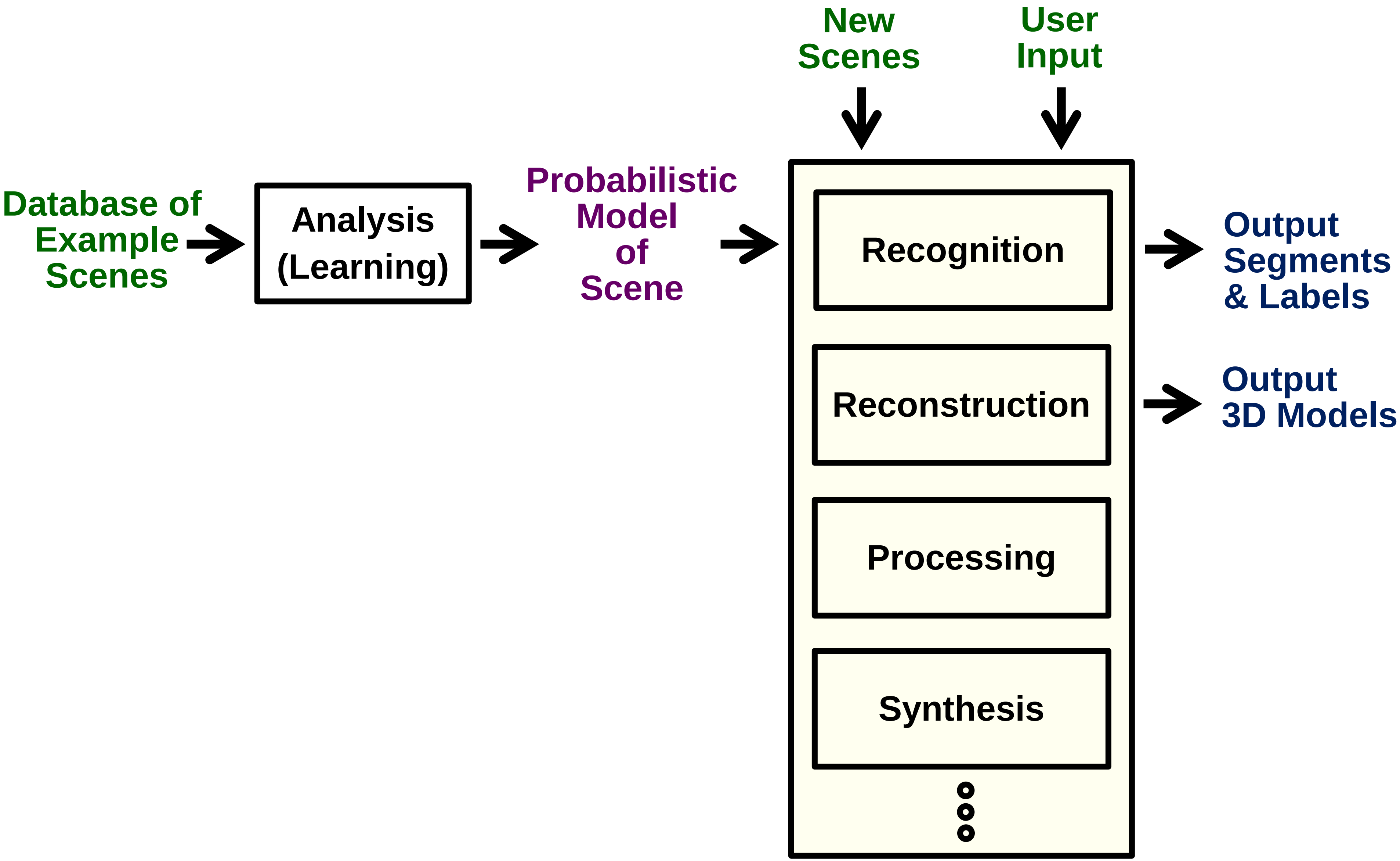
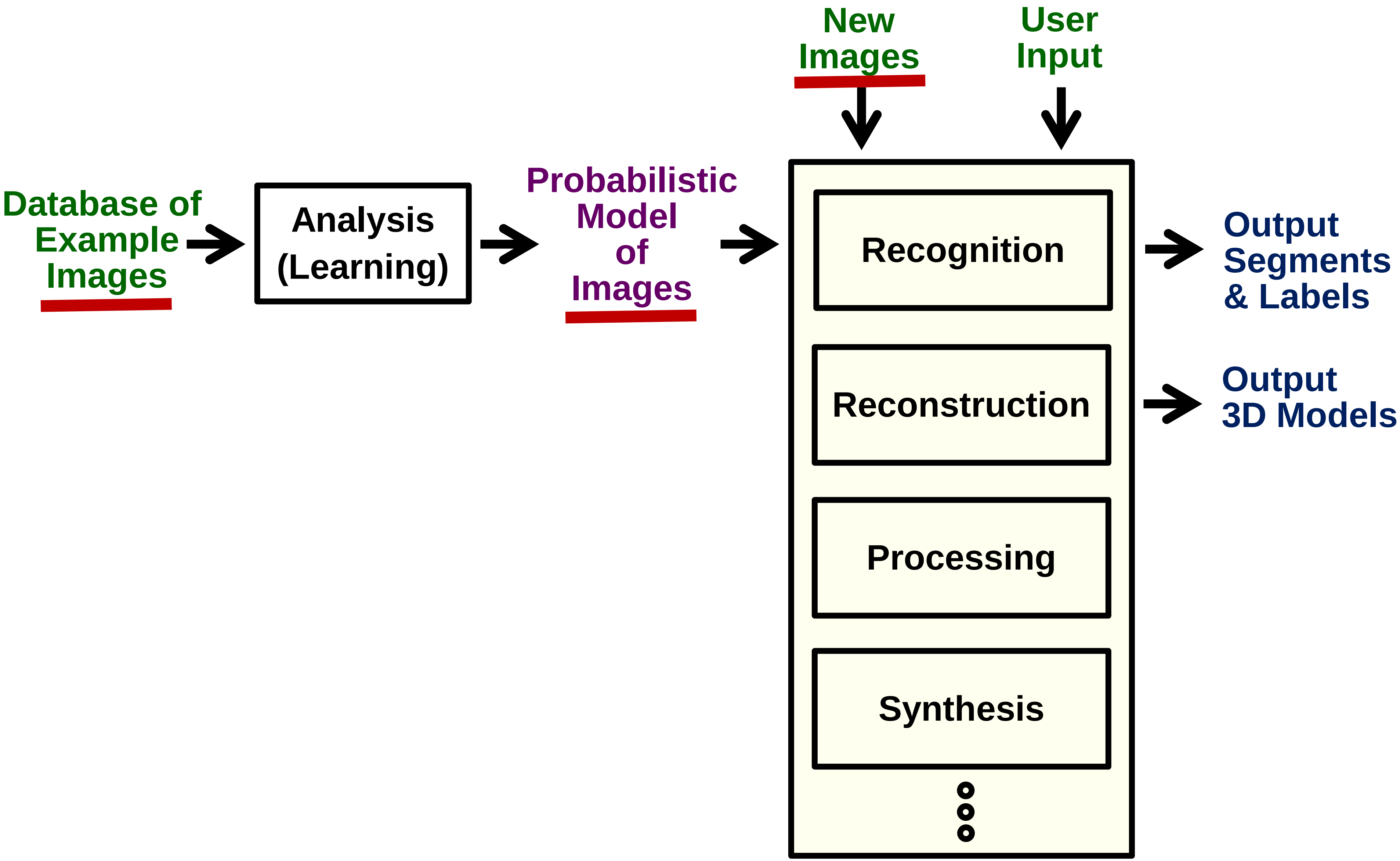
Learning 3D Models for Scene Understanding

Thomas Funkhouser
Princeton University

Scene Understanding



Traditional Computer Vision



Traditional Computer Vision

New Images

User Input

Database of Example Images → Analysis (Learning) → Probabilistic Model of Images →

Recognition

Reconstruction

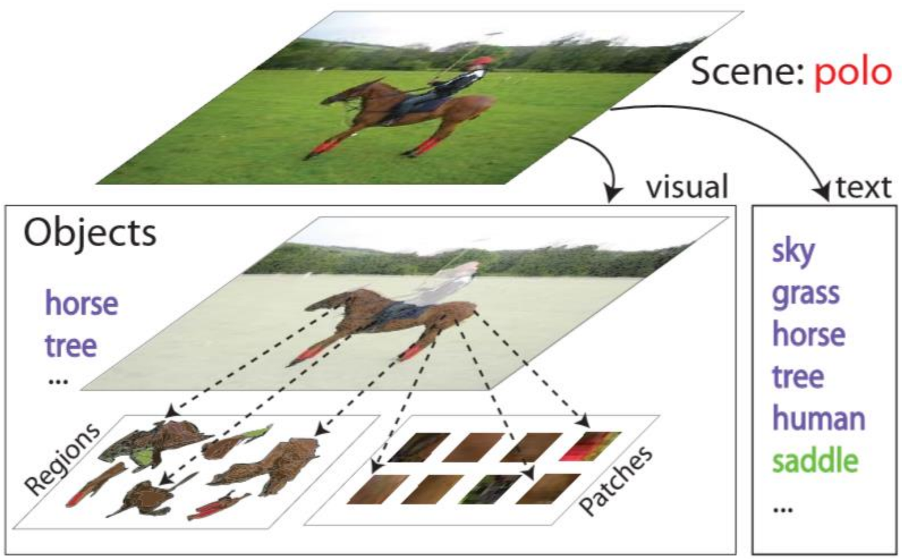
Processing

Synthesis

⋮

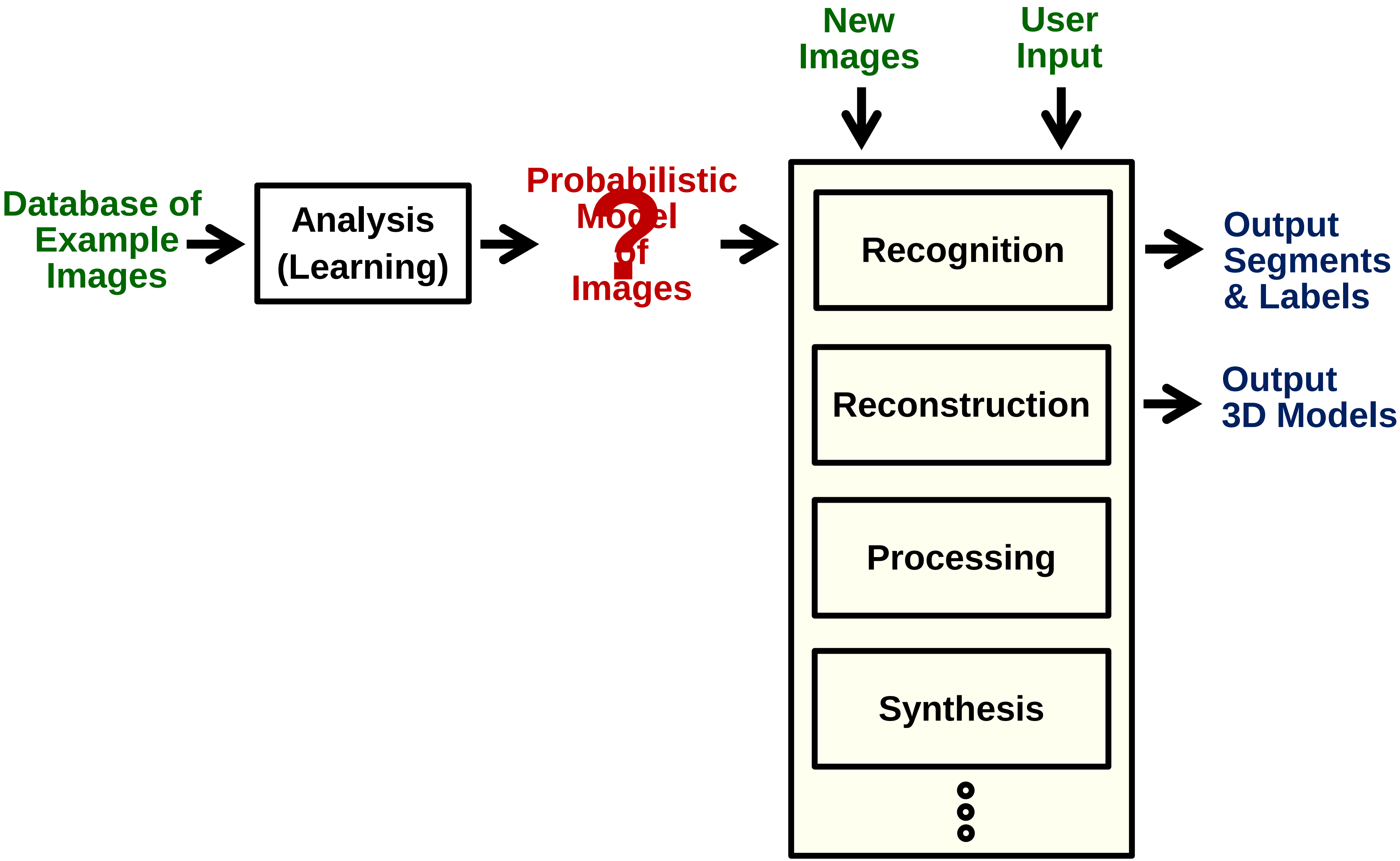
→ Output Segments & Labels

→ Output 3D Models



[Li, Socher, Fei-Fei, 2009]

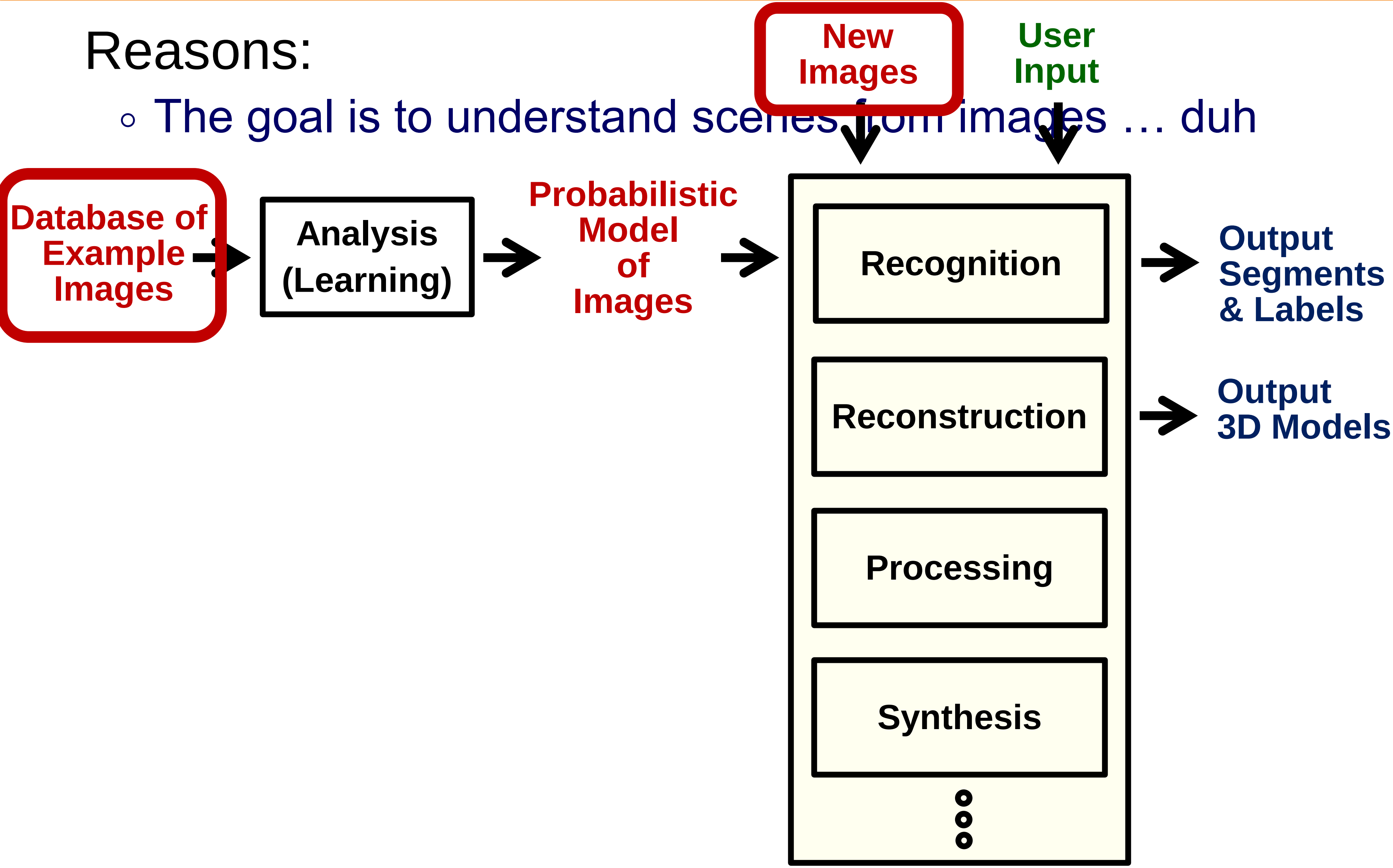
Why Learn Models of Images?



Why Learn Models of Images?

Reasons:

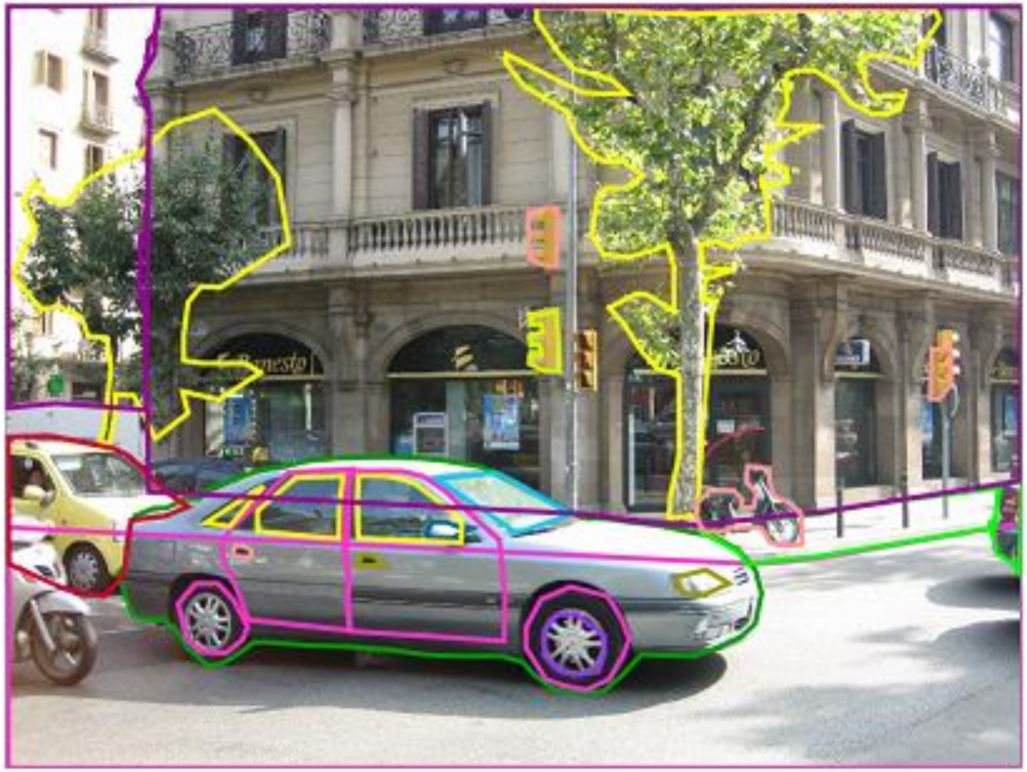
- The goal is to understand scenes from images … duh



Why Learn Models of Images?

Reasons:

- The goal is to understand scenes from images … duh
- Some labeled examples, lots of unlabeled examples



LabelMe [Russell 2005]

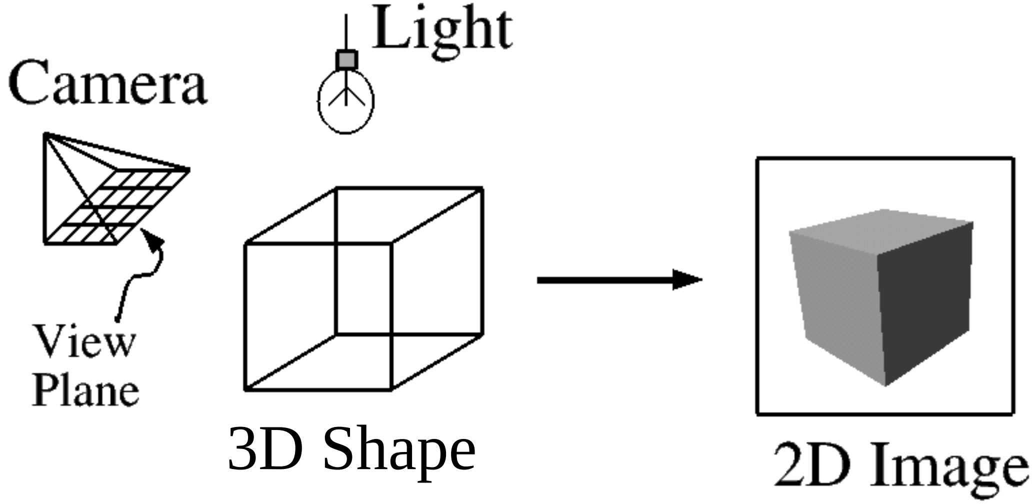
Why Learn Models of Images?

Reasons:

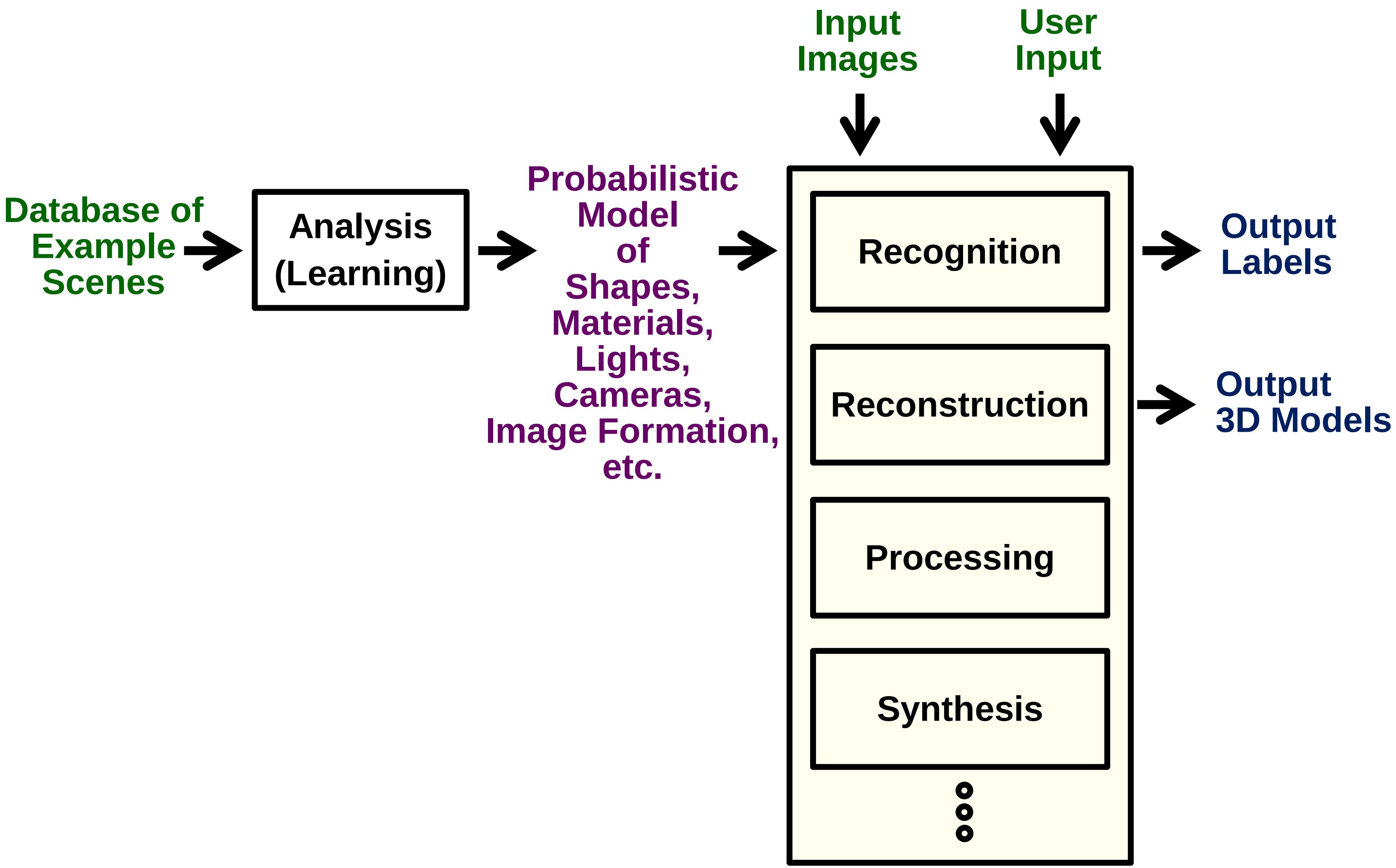
- The goal is to understand scenes from images … duh
- Some labeled examples, lots of unlabeled examples

Problems:

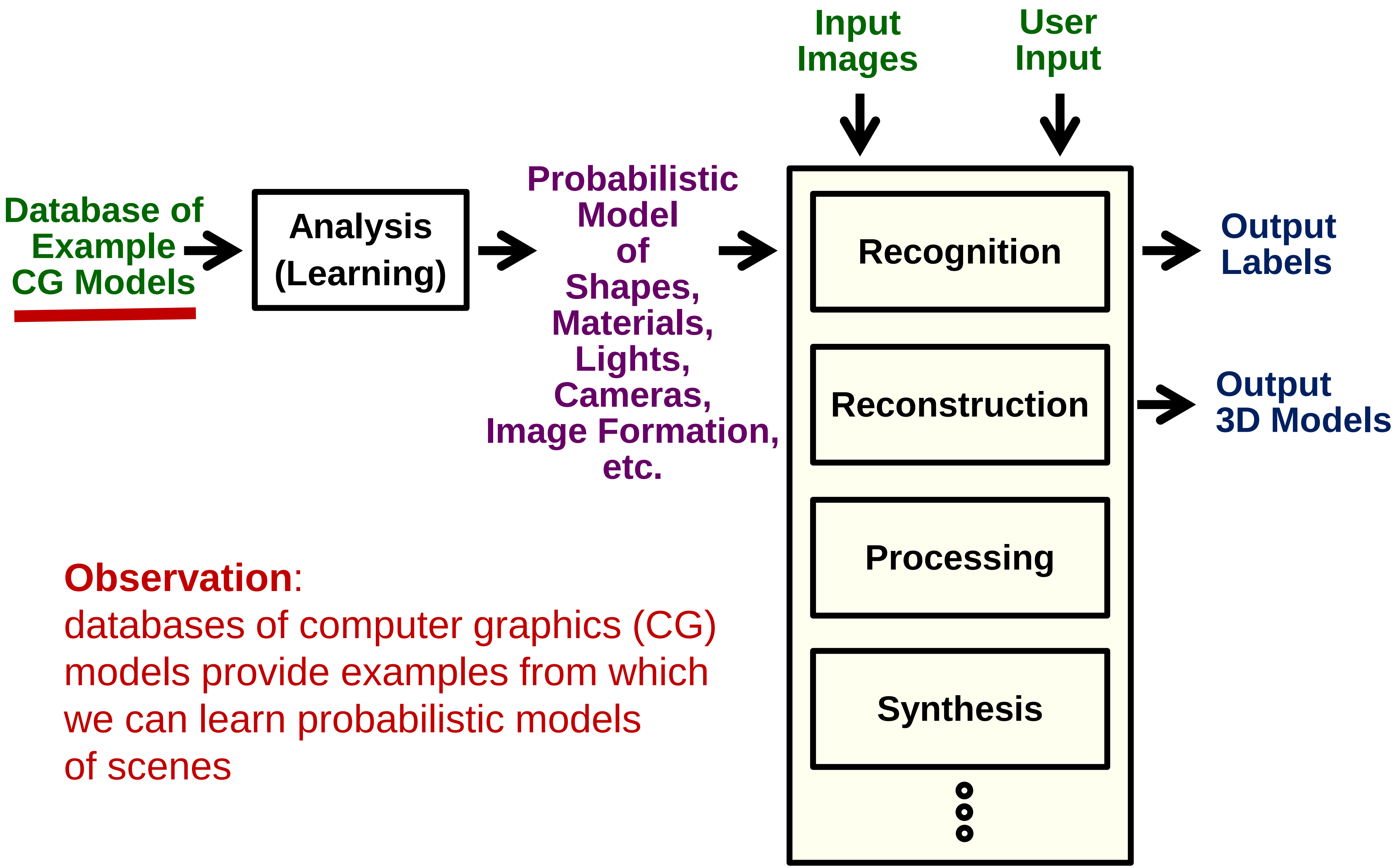
- Shape
- Materials
- Lighting
- Viewpoint
- Perspective
- Occlusions
- Light transport
- Segmentation
- Noise



Shouldn't We Learn Models of Scenes?



Shouldn't We Learn Models of Scenes?

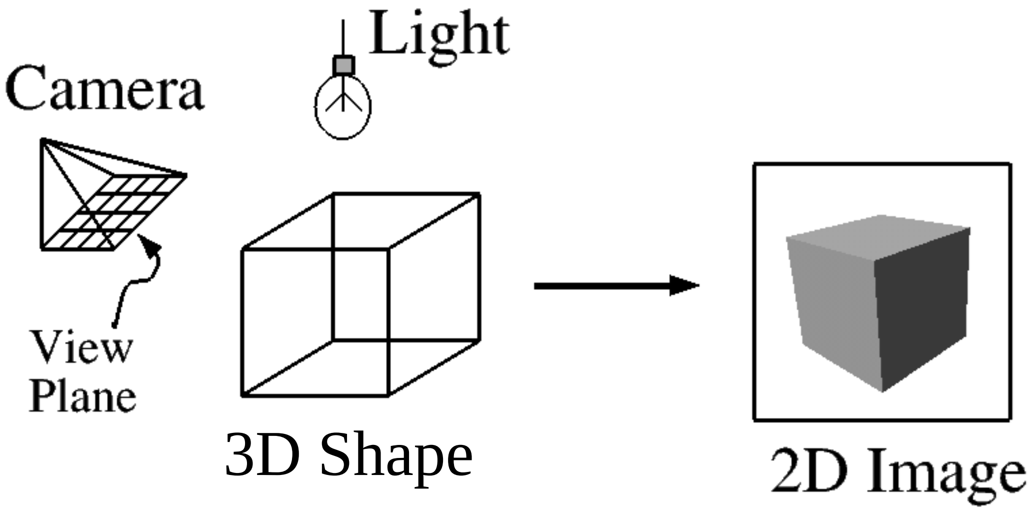


Observation:
databases of computer graphics (CG) models provide examples from which we can learn probabilistic models of scenes

Why Learn from CG Models?

CG models provide …

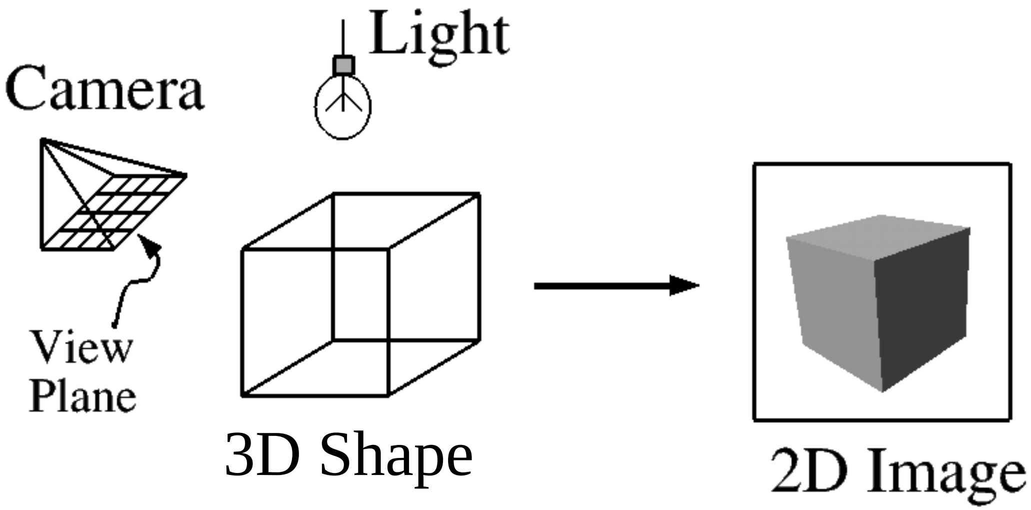
- Shape
- Materials
- Lighting
- Viewpoint
- Perspective
- Occlusions
- Light transport
- Segmentation
- No noise



Why Learn from CG Models?

CG models provide …

- Shape
- Materials
- Lighting
- Viewpoint
- Perspective
- Occlusions
- Light transport
- Segmentation
- No noise



Issues:

- Enough examples?
- Quality?

Why Learn from CG Models?

CG models provide …

- Shape
- Materials
- Lighting
- Viewpoint
- Perspective
- Occlusions
- Light transport
- Segmentation
- Noise

Issues:

- ➢Enough examples?
- Quality?

Trimble 3D Warehouse

Why Learn from CG Models?

CG models provide …

- Shape
- Materials
- Lighting
- Viewpoint
- Perspective
- Occlusions
- Light transport
- Segmentation
- Noise

Issues:

- Enough examples?
- ➢Quality?

Ikea

Related Work

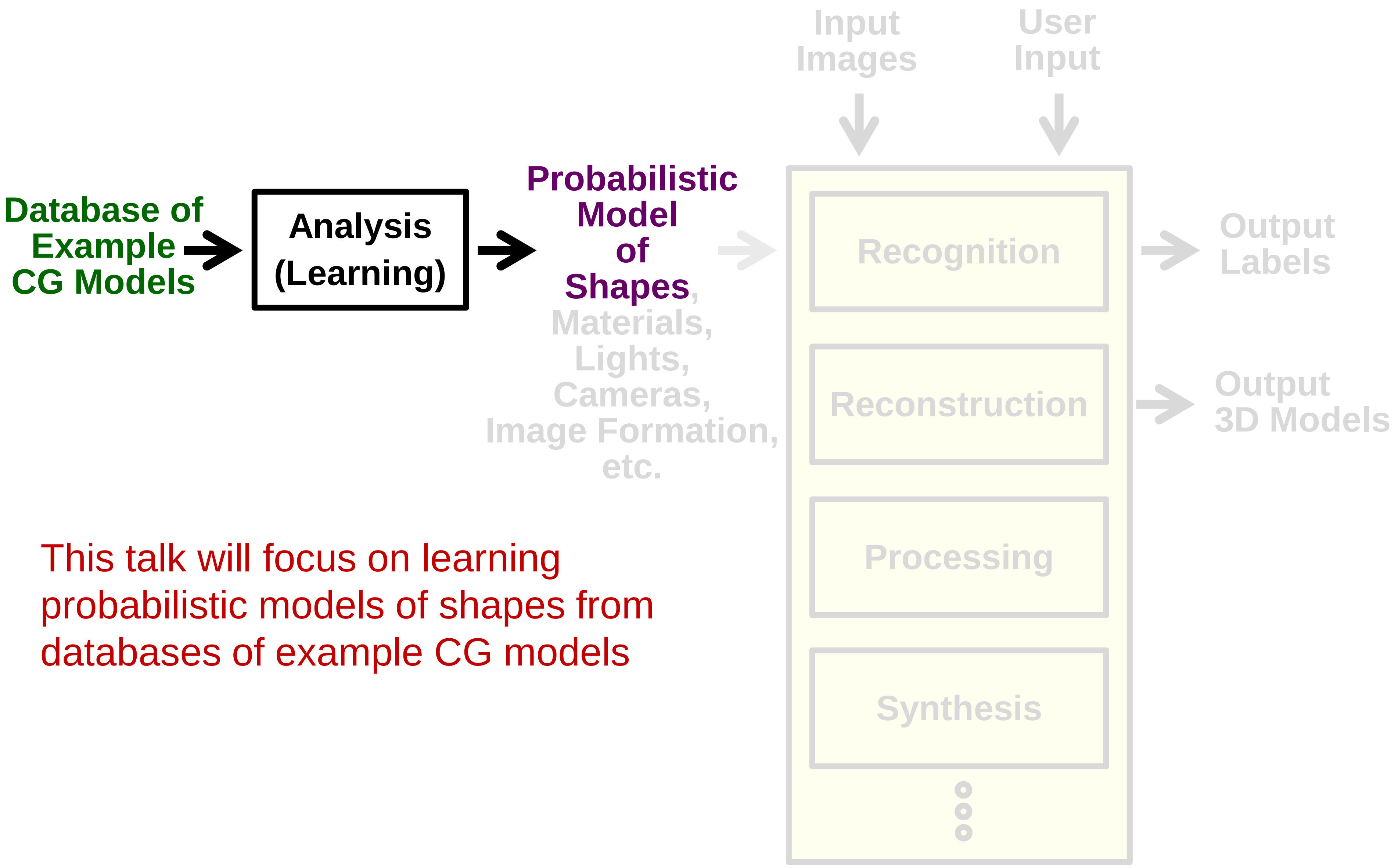
Using CG models for scene understanding

- Fitting CG models to images
  - Lai 2009, Xu 2011, Satkin 2013, Aubry 2014, etc.
- Fitting CG models to range scans
  - Nan 2012, Shen 2012, Kim 2012, Song 2014, etc.
- Using CG models to learn parameters
  - Zhao 2013, etc.

Analyzing databases of CG models

- Consistent segmentation, labeling, correspondence, …
  - Golovinskiy 2009, Sidi 2011, Kim 2013, Mitra 2013, etc.
- Learning probabilistic models
  - Chaudhuri 2010, Kalogerakis 2012, Fisher 2012, Kim 2013, etc.

Focus of This Talk



This talk will focus on learning probabilistic models of shapes from databases of example CG models

Outline of Talk

Introduction

Learning probabilistic models from CG collections

- Object templates
- Contextual model
- Hierarchical grammar

Conclusions

Outline of Talk

Introduction

Learning probabilistic models from CG collections

- Object templates
- Generative model
- Hierarchical grammar

Conclusions

Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. "Learning Part-based Templates from Large Collections of 3D Shapes," SIGGRAPH 2013.

Goal for This Project

Database of 3D meshes
representing an object class

Probabilistic
Model of Shape

Consistent part segmentations,
labels, and correspondences

Goal for This Project

Database of 3D meshes representing an object class

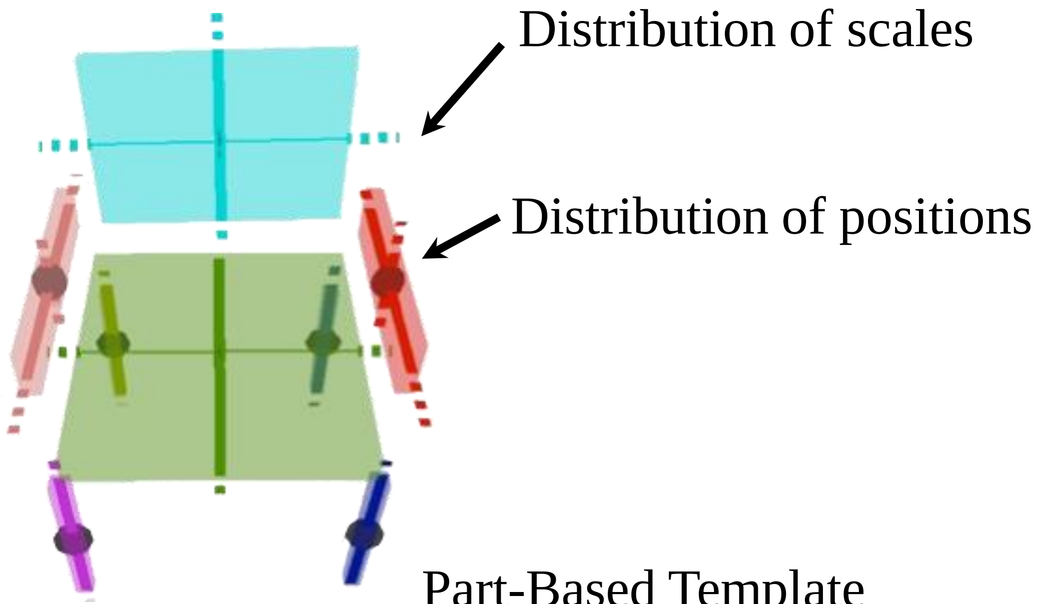
Probabilistic Model of Shape

Consistent part segmentations, labels, and correspondences

Challenge
Need to discover segmentations, labels, correspondences, and deformation modes all together

Object Templates

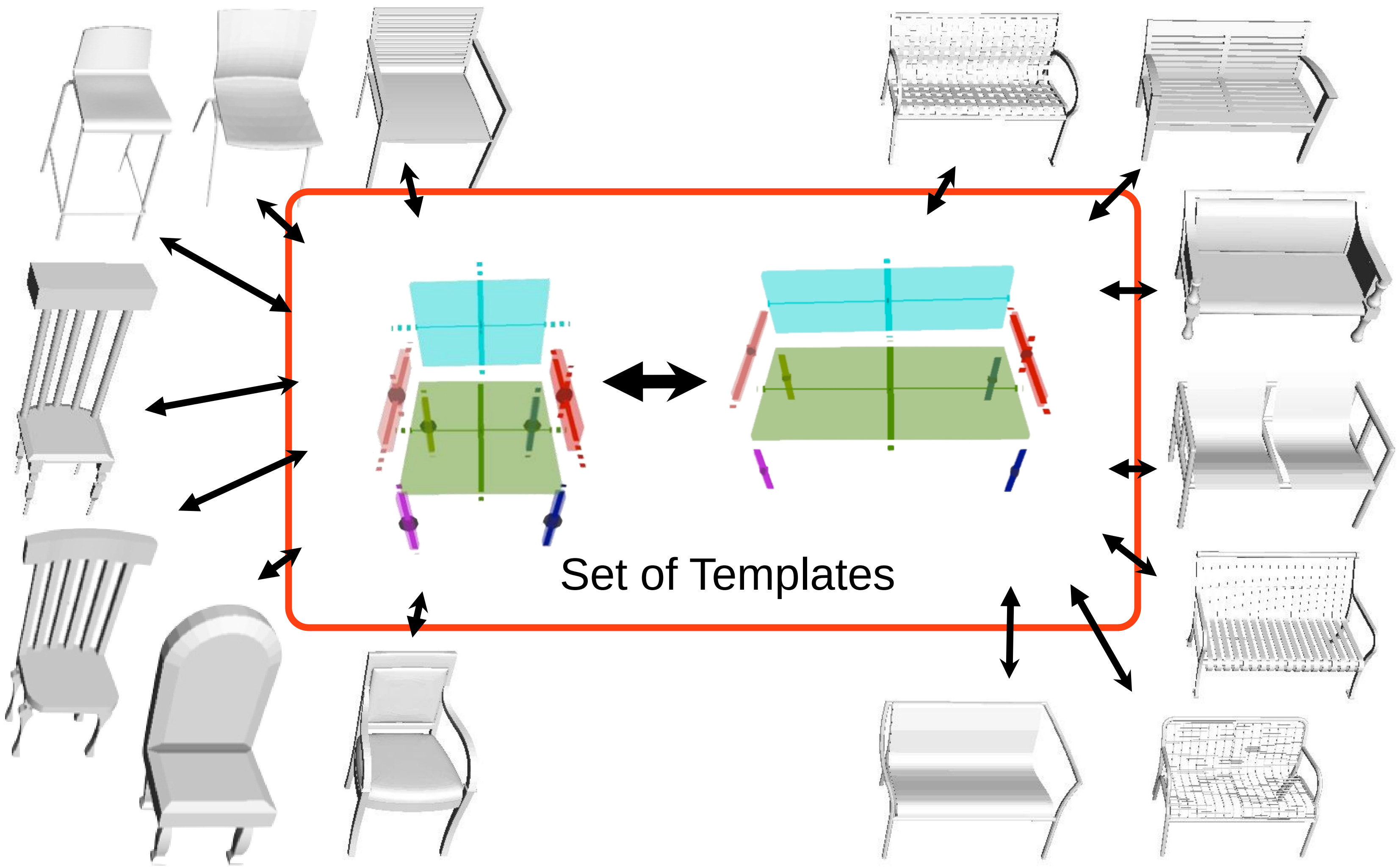
Represent object class by part-based templates where each template has a set of parts, and each part has probability distributions for its shape, position, and anisotropic scales



Part-Based Template

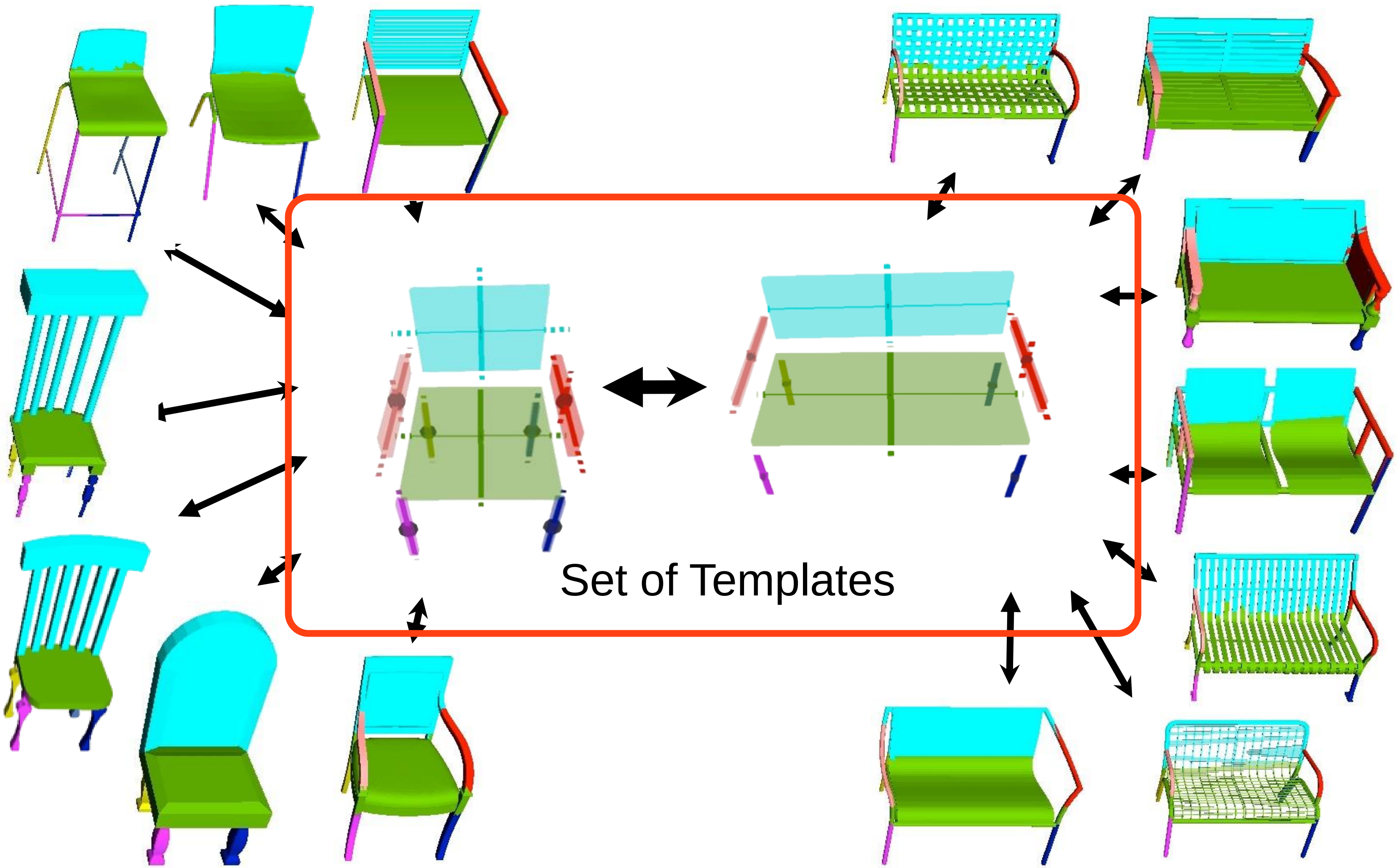
Template Learning and Fitting

Aim to learn a set of corresponding templates that provides a good fit to every mesh in the database



Template Learning and Fitting

Aim to learn a set of corresponding templates that provides a good fit to every mesh in the database

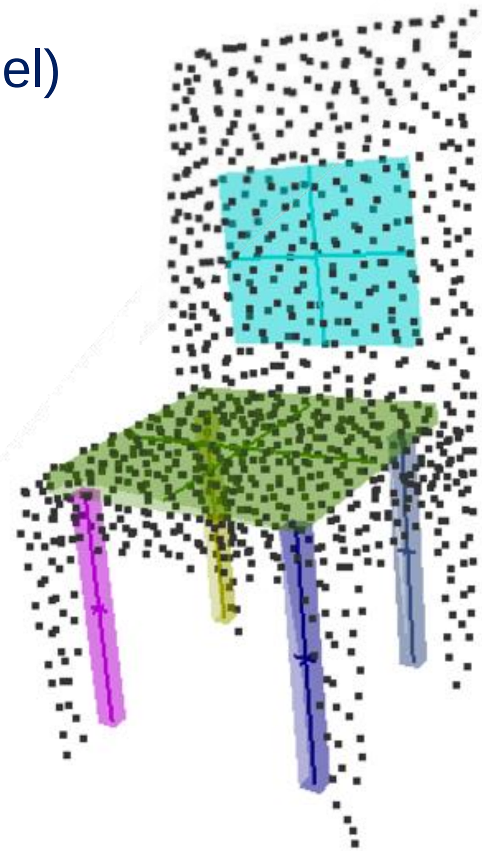


Template Fitting Problem

For a given template and mesh, aim to minimize:

$$E = E_{\text{data}} + \gamma E_{\text{deform}} + \beta E_{\text{smooth}}$$

- E_{data} (template ↔ shape distance + local shape features)
- E_{deform} (plausibility of template deformation)
- E_{smooth} (close & similar regions get same label)



Unknowns are:

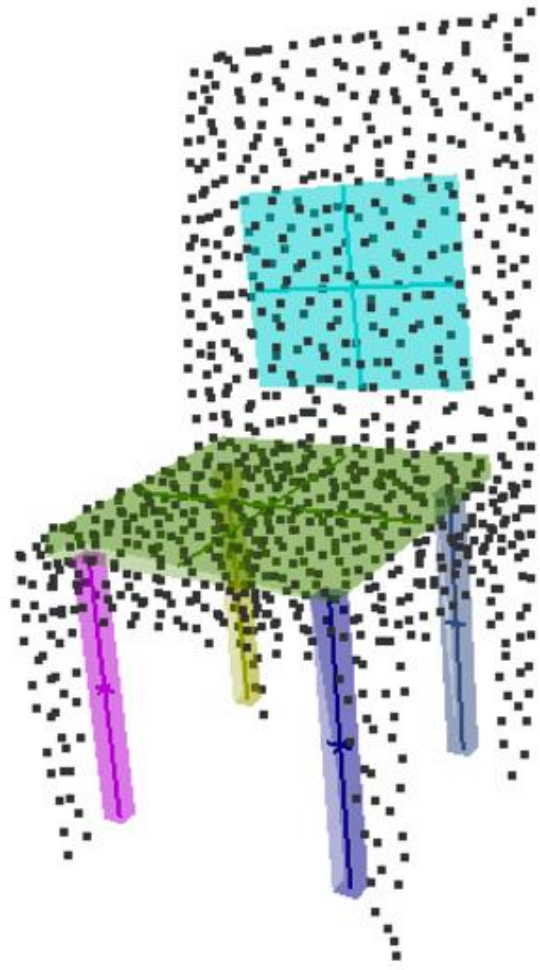
- Point segmentations and labels
- Point correspondences
- Part center positions
- Part anisotropic scales

Template Fitting Algorithm

Solve by iteratively minimizing different energy terms:

- Segmentation and labeling
- Point correspondence
- Part-aware deformation

$$E = E_{\text{data}} + \gamma E_{\text{deform}} + \beta E_{\text{smooth}}$$

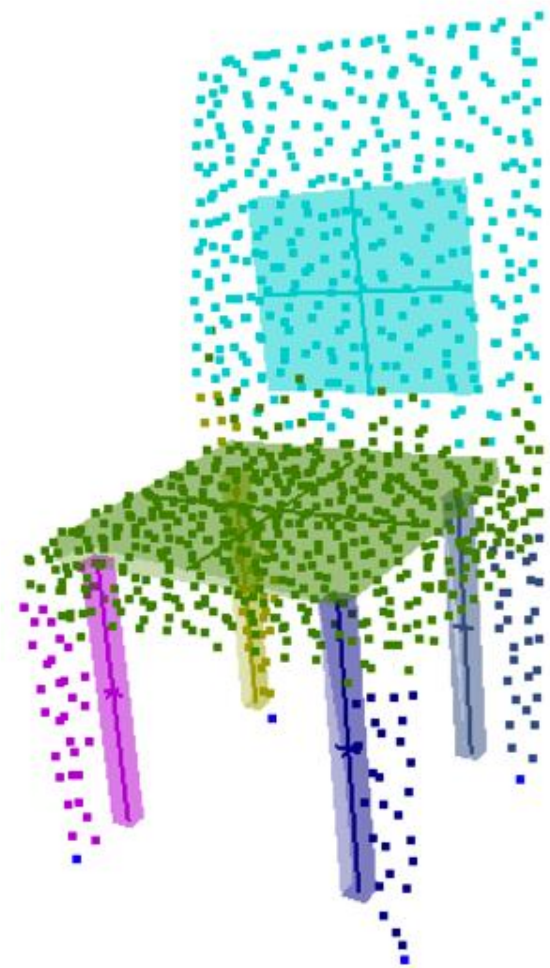


Template Fitting Algorithm

Solve by iteratively minimizing different energy terms:

- Segmentation and labeling
- Point correspondence
- Part-aware deformation

$$E = \underline{E_{\text{data}}} + \gamma E_{\text{deform}} + \underline{\beta E_{\text{smooth}}}$$



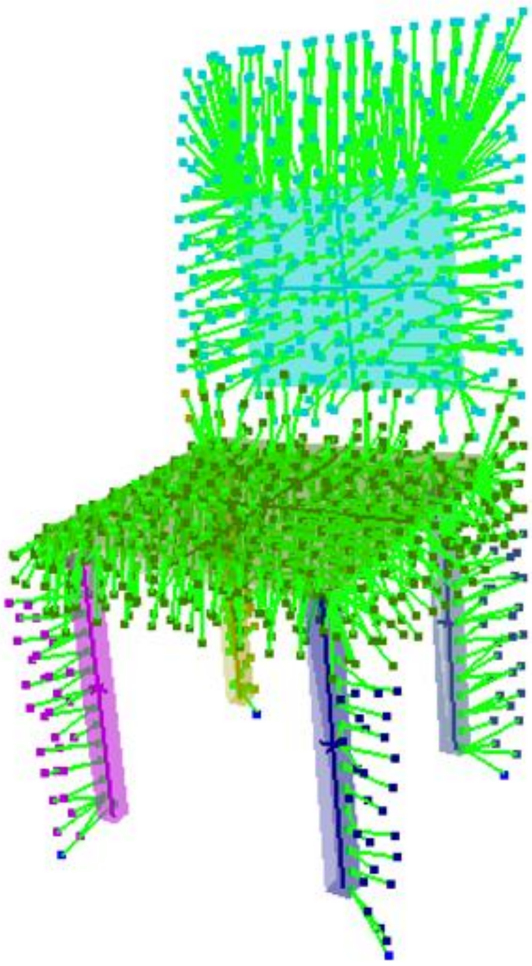
Solve with graph cut [Boykov 2001]

Template Fitting Algorithm

Solve by iteratively minimizing different energy terms:

- ◦ Segmentation and labeling
- ➢Point correspondence
- ◦ Part-aware deformation

$$E = E_{\text{data}} + \gamma E_{\text{deform}} + \beta E_{\text{smooth}}$$



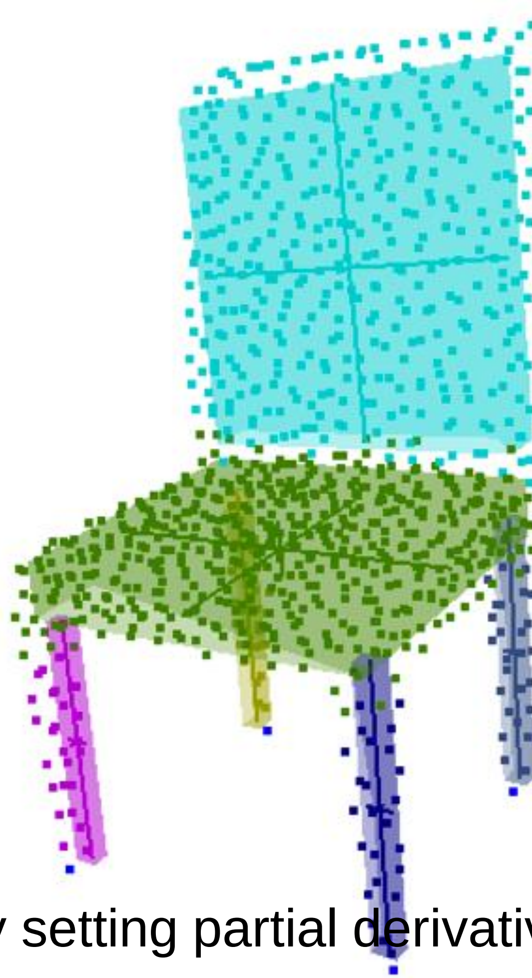
Solve with part-aware closest points

Template Fitting Algorithm

Solve by iteratively minimizing different energy terms:

- ◦ Segmentation and labeling
- ◦ Point correspondence
- ➢ Part-aware deformation

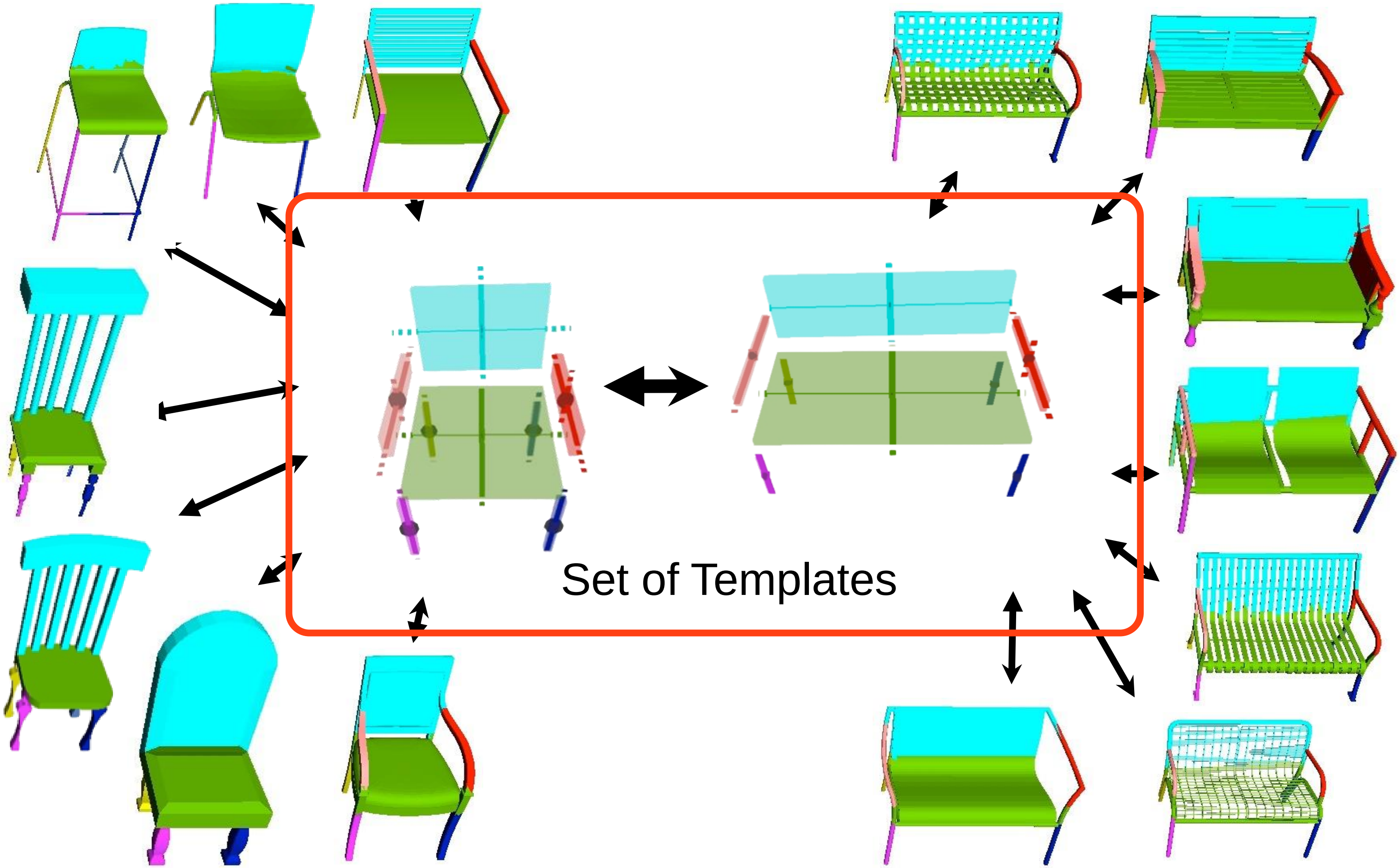
$$E = E_{\text{data}} + \gamma E_{\text{deform}} + \beta E_{\text{smooth}}$$



Solve for positions and scales of each part by setting partial derivatives to zero.

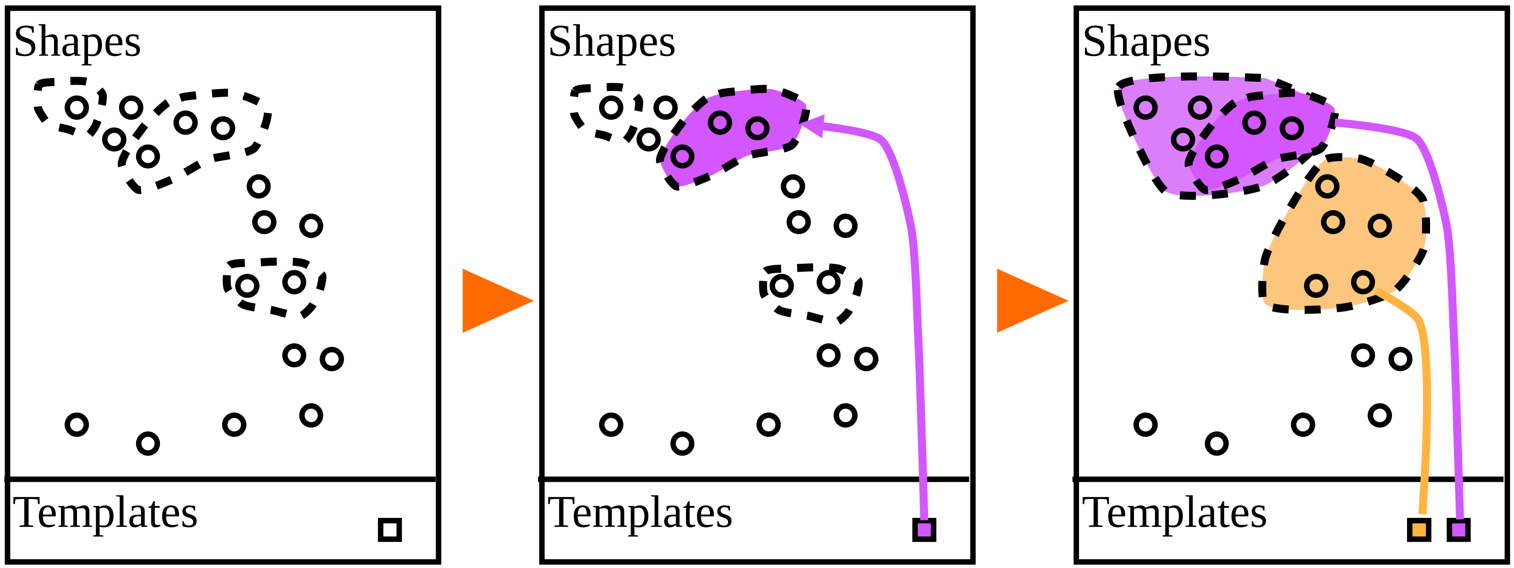
Template Learning Problem

Aim to learn a set of corresponding templates that provides a good fit to every mesh in the database



Template Learning Algorithm

Iteratively grow a set of templates with each optimized to fit a different cluster of meshes



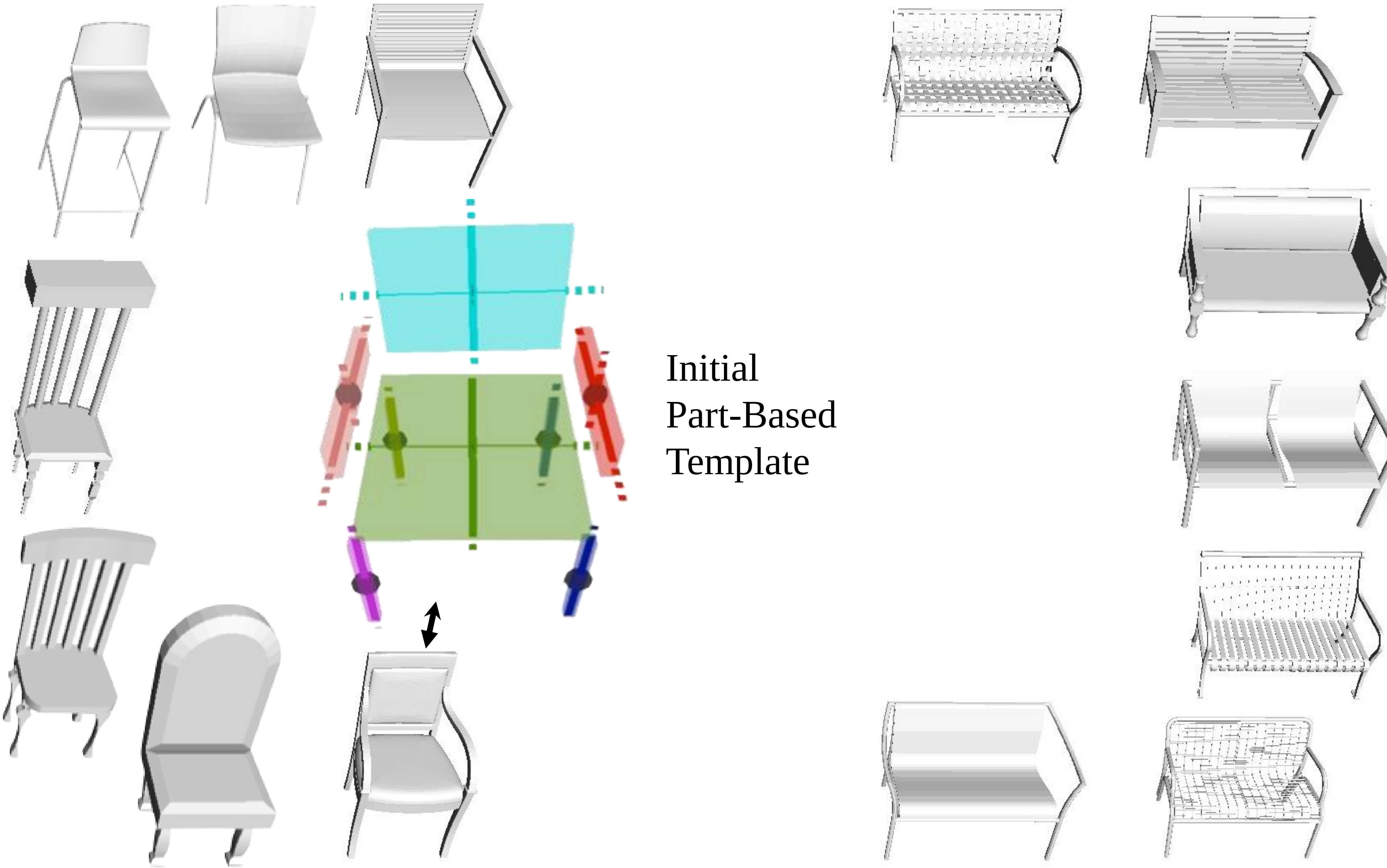
Template Initialization

Template Fitting

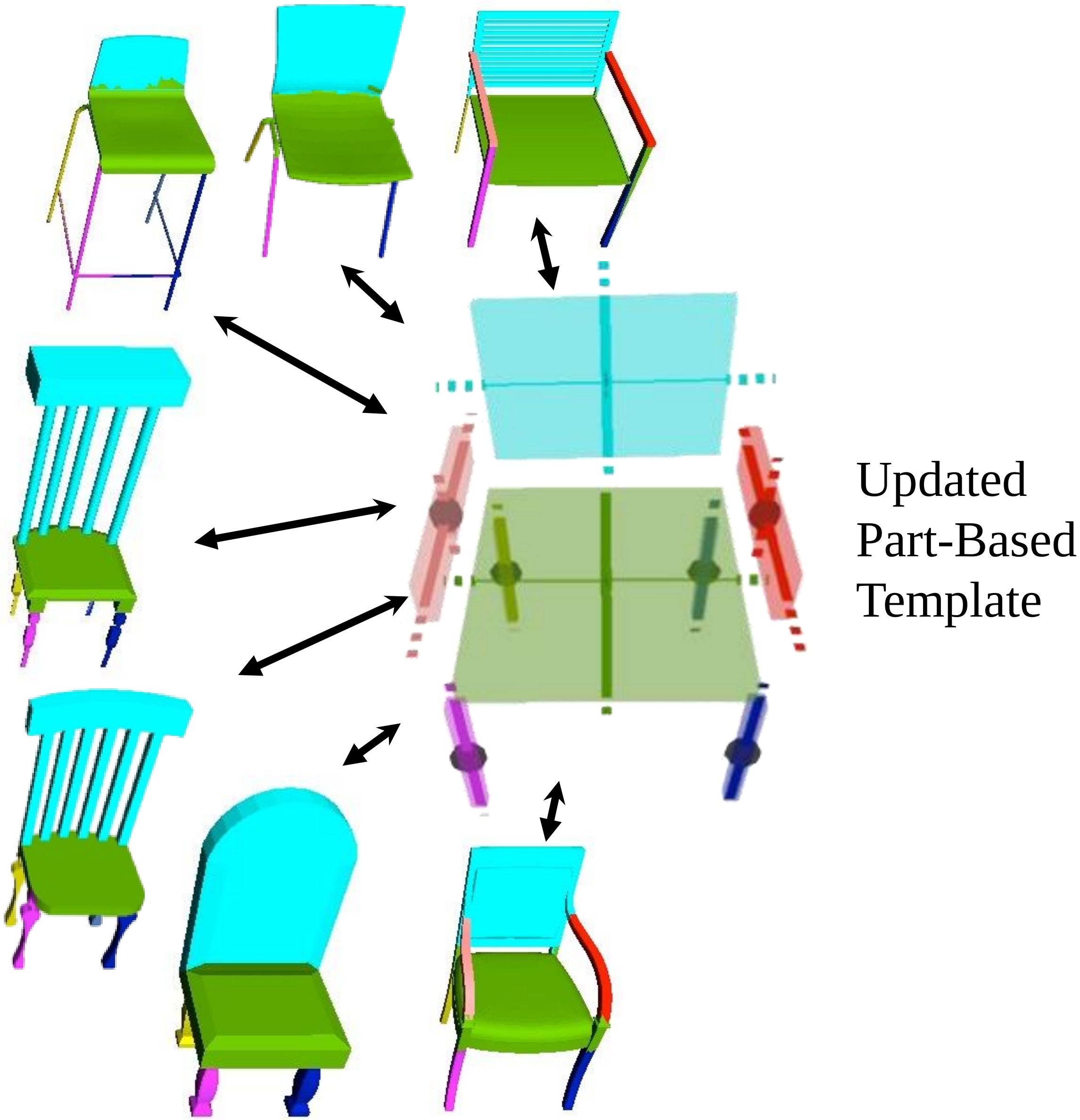
Template Refinement

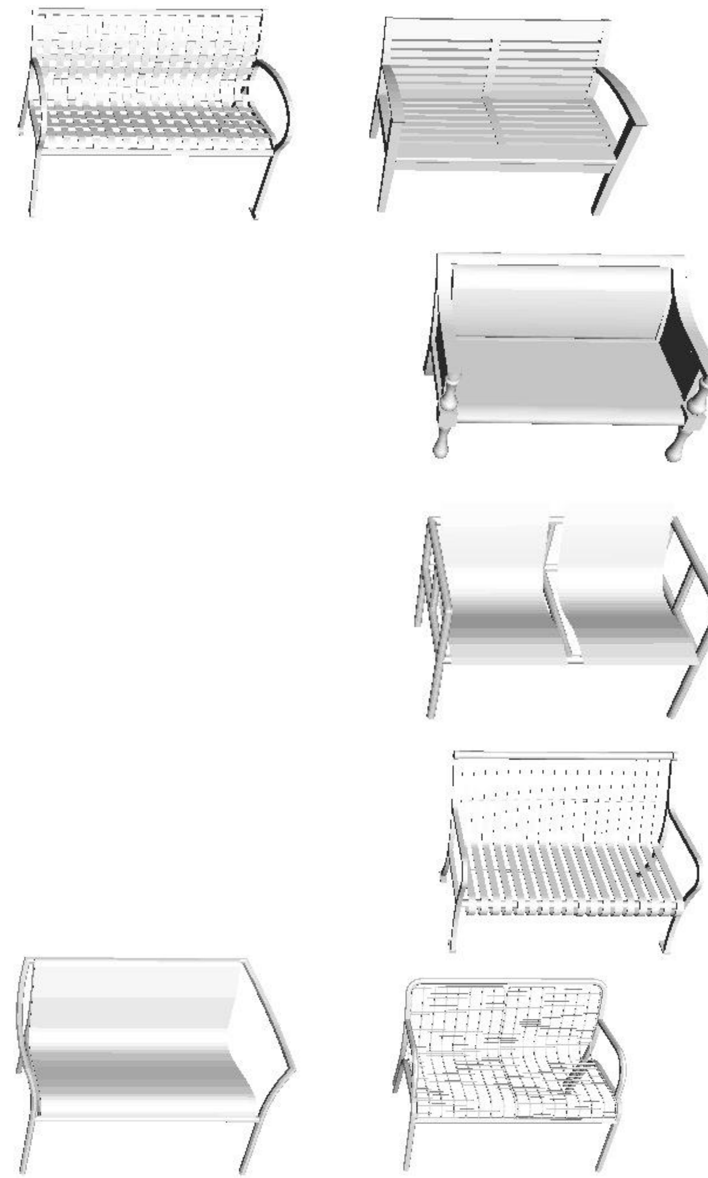
repeat until convergence

Template Learning Example

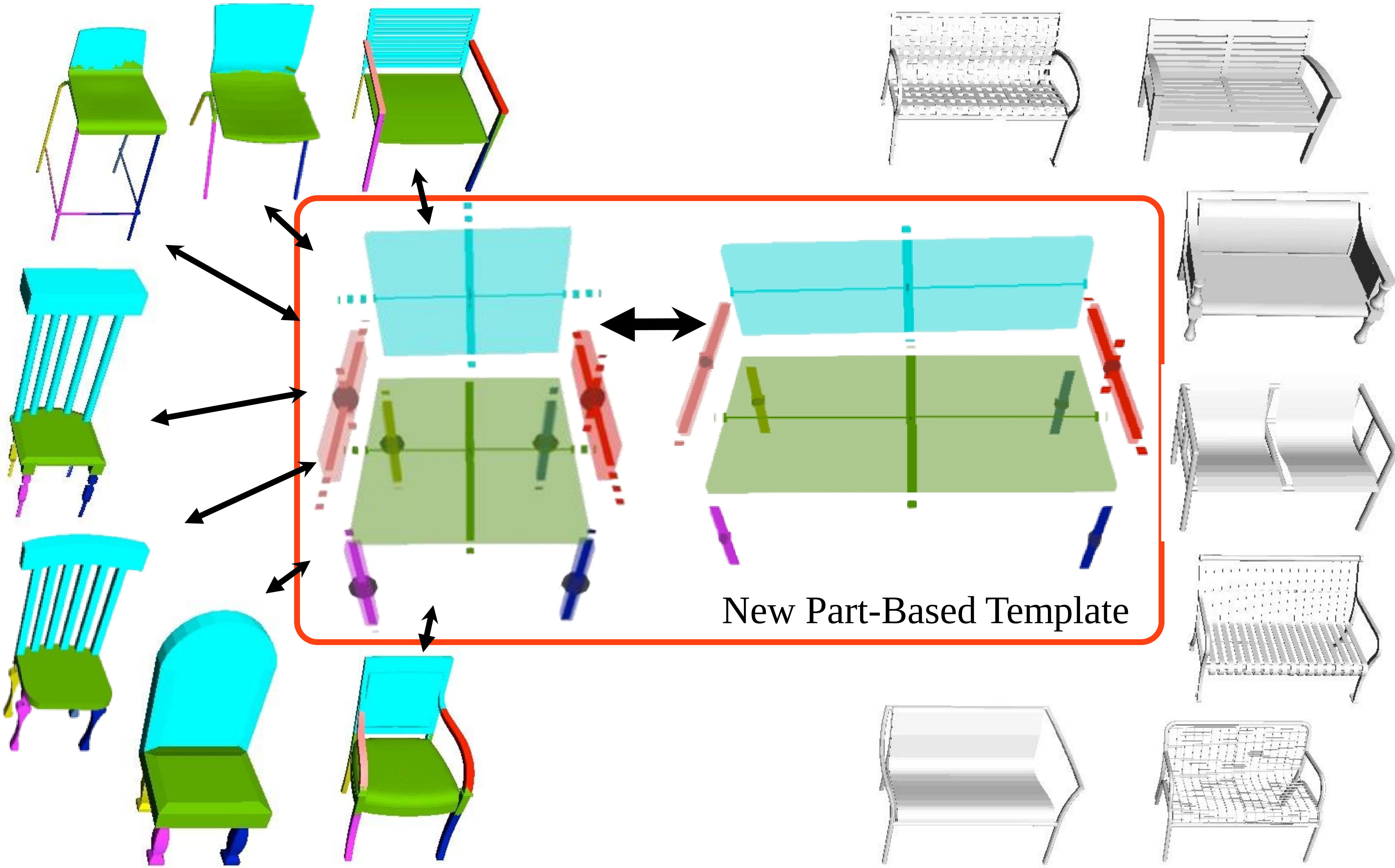


Template Learning Example

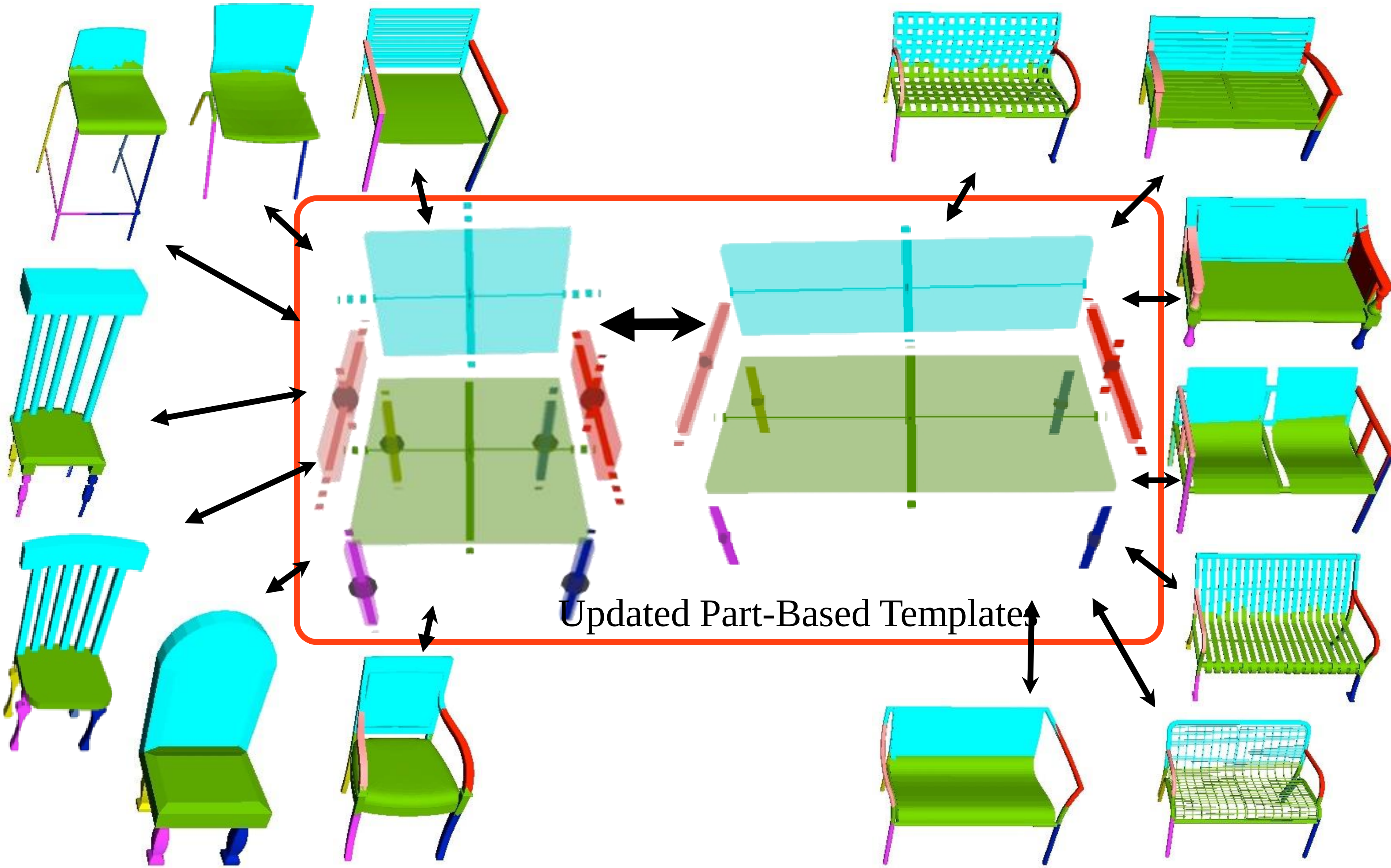




Template Learning Example



Template Learning Example



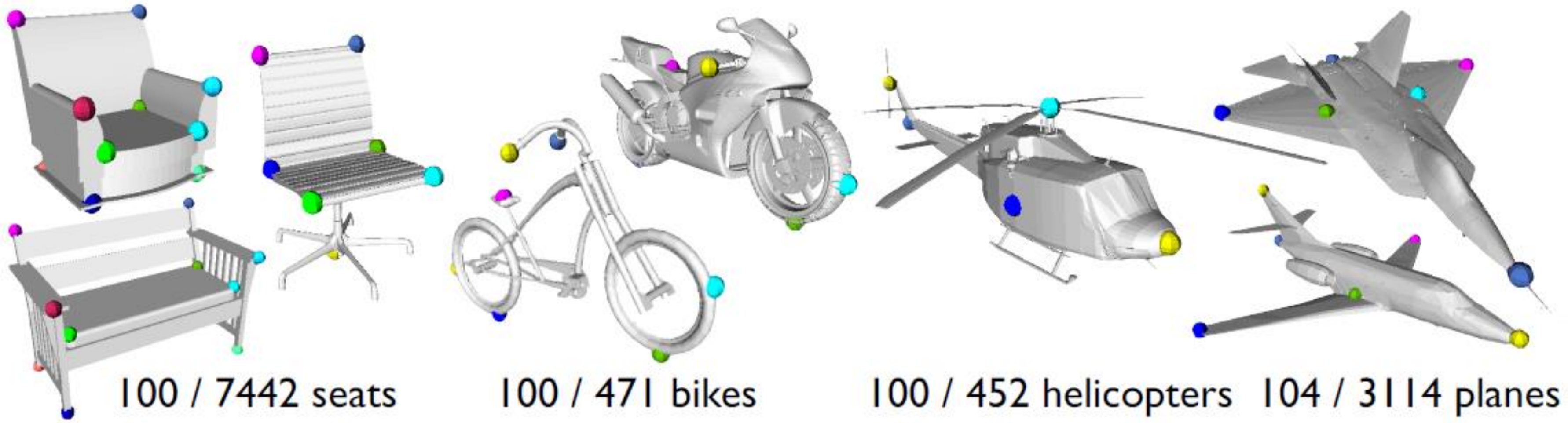
Template Learning and Fitting Results

Data sets:

- Crawl SketchUp Warehouse for collections by keyword
- Eliminate outliers with Mechanical Turk
- Specify manual correspondences for subset of models

Experiments:

- Solve for part-based templates for collection
- Evaluate correspondences & segmentations



Template Learning and Fitting Results

2 Templates

3113 Airplanes

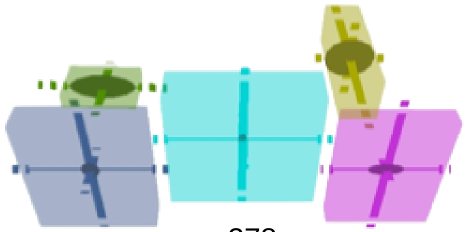
Template Learning and Fitting Results

2 Templates

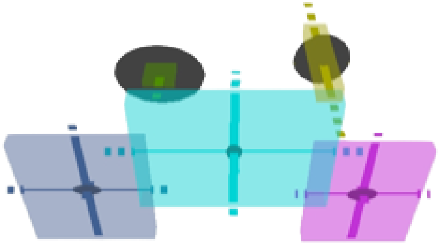
3113 Airplanes

Template Learning and Fitting Results

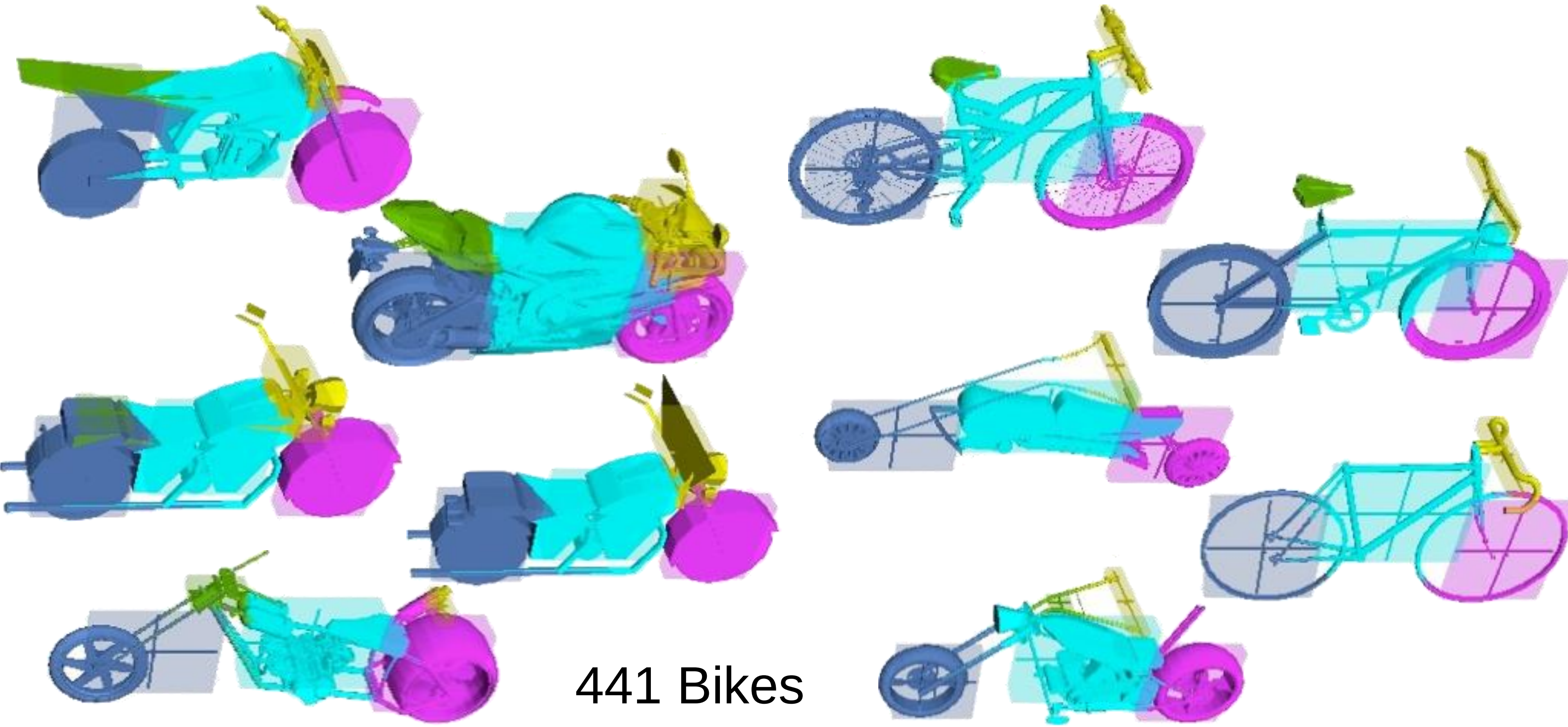
2 Templates



378

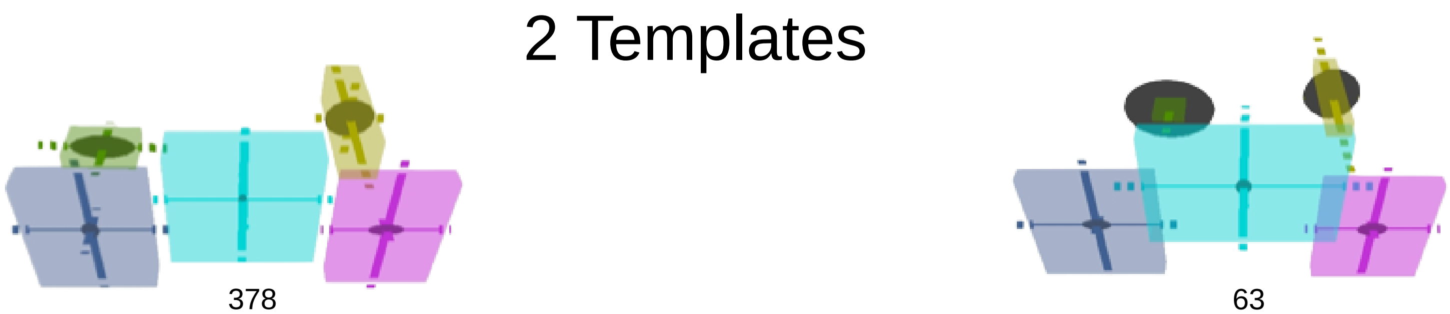


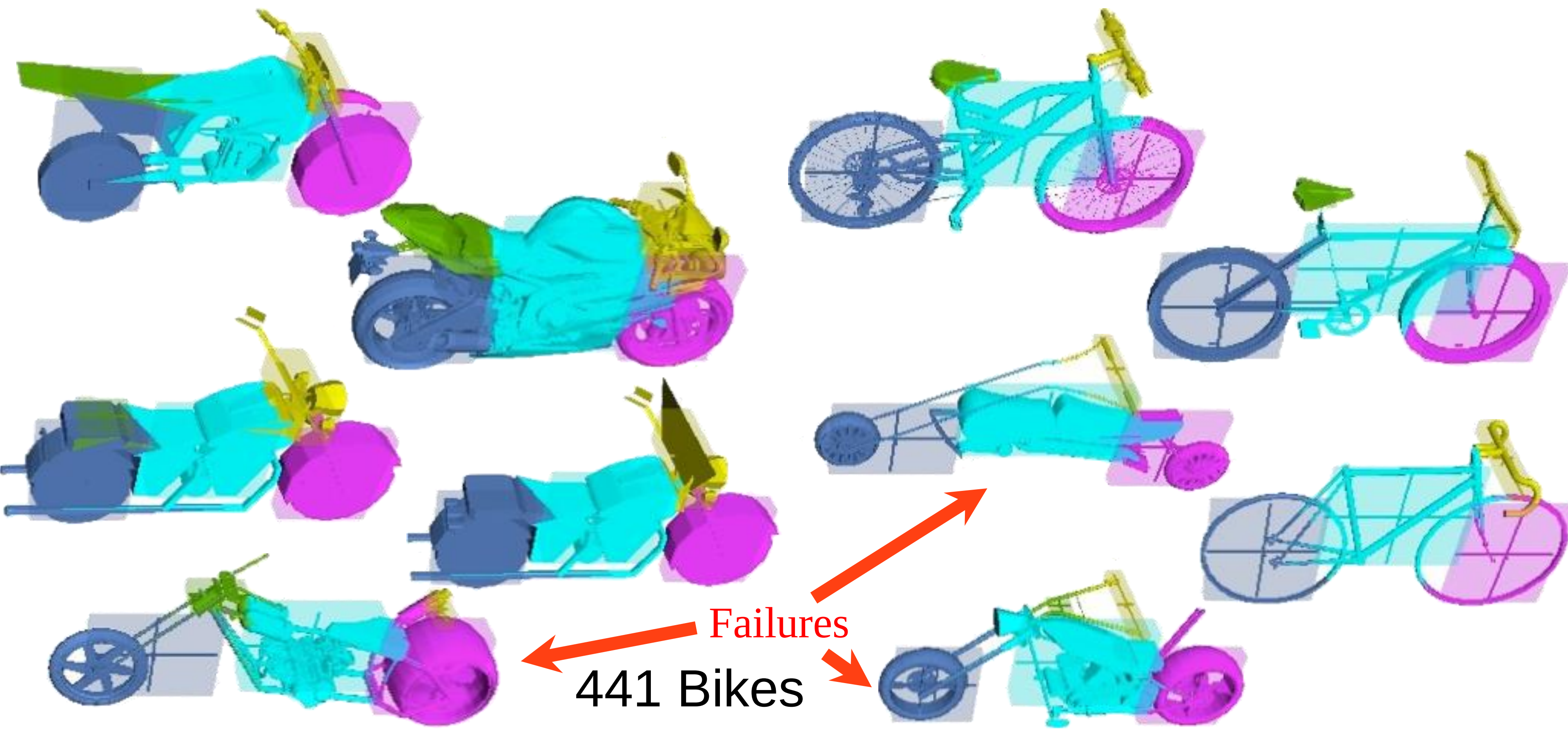
63



441 Bikes

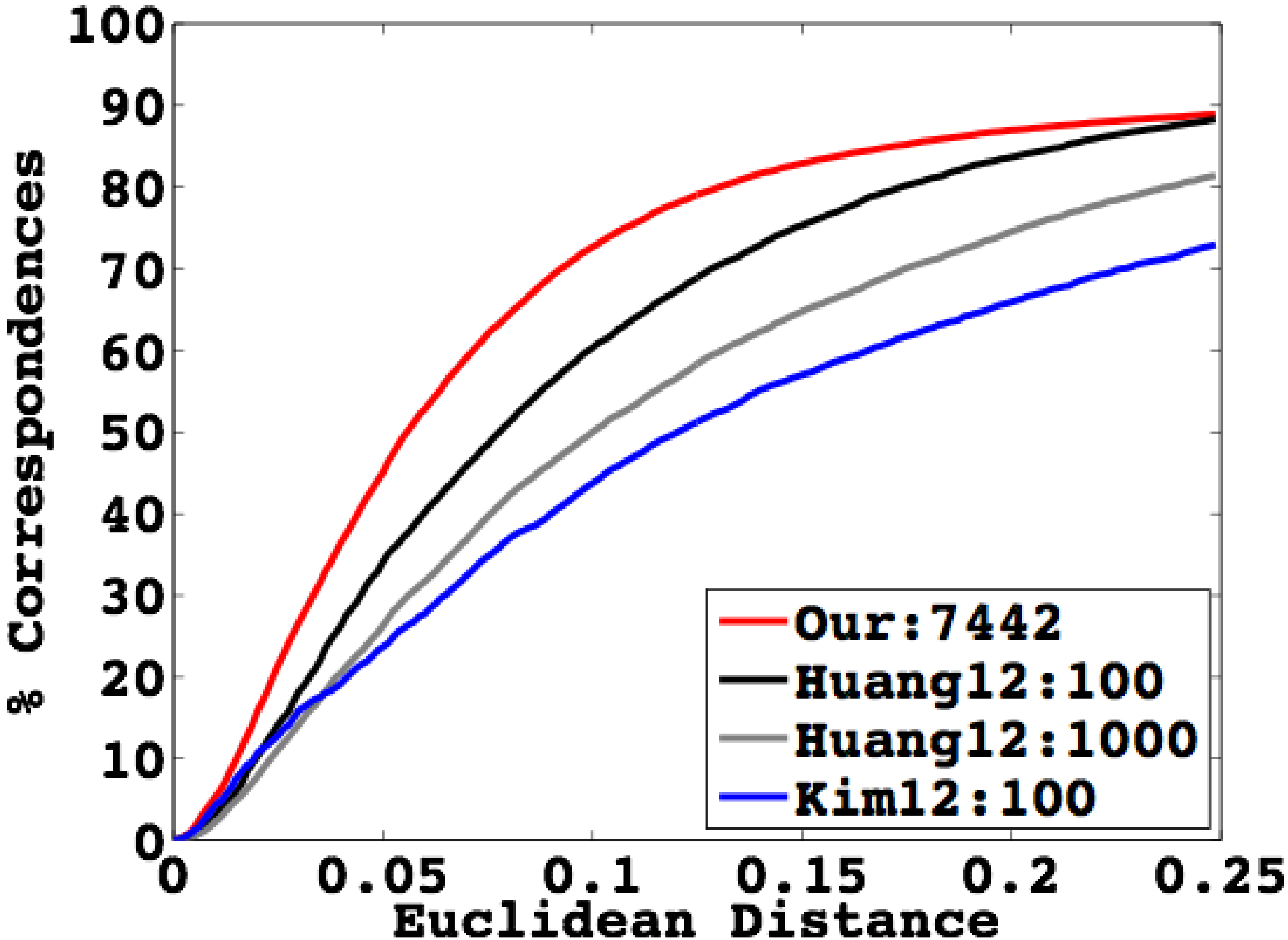
Template Learning Results





Surface Correspondence Results

Correspondence benchmark (7442 seats)



Surface Segmentation Results

Co-segmentation benchmark [Sidi et al, 2011]

Class	Hu	Our
Chairs	89.6	$\boxed{97.6}$
Lamps	90.7	$\boxed{95.2}$
FourLegged	88.7	86.9
Goblets	99.2	97.6
Vase	80.2	81.3
Guitars	98.0	88.5
Candelabra	93.9	82.4

within 2%
or ours
is better

Outline of Talk

Introduction

Learning probabilistic models from CG collections

- Objet templates
- Contextual model
- Hierarchical grammar

Conclusions

Matthew Fisher, Daniel Ritchie, Manolis Savva, Thomas Funkhouser, and Pat Hanrahan,
Example-based Synthesis of 3D Object Arrangements, SIGGRAPH Asia, 2012.

Goal for This Project

Exemplar scenes

+

Database of Scenes

Probabilistic Model of Shape

Synthesized novel scenes

Goal for This Project

Exemplar scenes

+

...

Database of Scenes

Probabilistic Model of Shape

Synthesized novel scenes

Challenge

Need to learn a model with great generality from few examples

Contextual Object Categories

Define categories of objects based on their contexts in a scene rather than basic functions

- Learned from examples by clustering of objects with similar spatial neighborhoods

Some Contextual Object Categories

Contextual Model

Represent the probability of a scene S by a generative model based on category cardinalities (c), support hierarchy topology relationships (t), and spatial arrangement relationships (a)

$$P(S) = P(c,t,a) = P(a|t,c)\, P(t|c)\, P(c)$$

Exemplar scenes

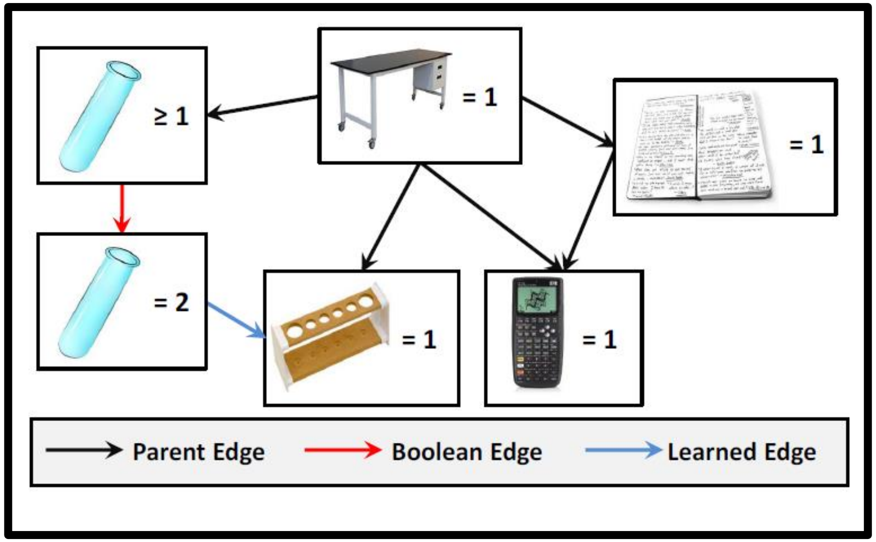
Contextual Model Details

Category cardinalities: $P(c)$

- Represent with Bayesian network
- Boolean random variables (# desks > 1?)
- Add support surface constraints

Lab table	Test tube rack	Test tube	Notebook	Calculator
1	1	2	0	0
1	0	1	0	1
1	0	0	1	1

Object frequencies in target scenes
+ support constraints



Bayesian network

Contextual Model Details

Support relationships: $P(t|c)$

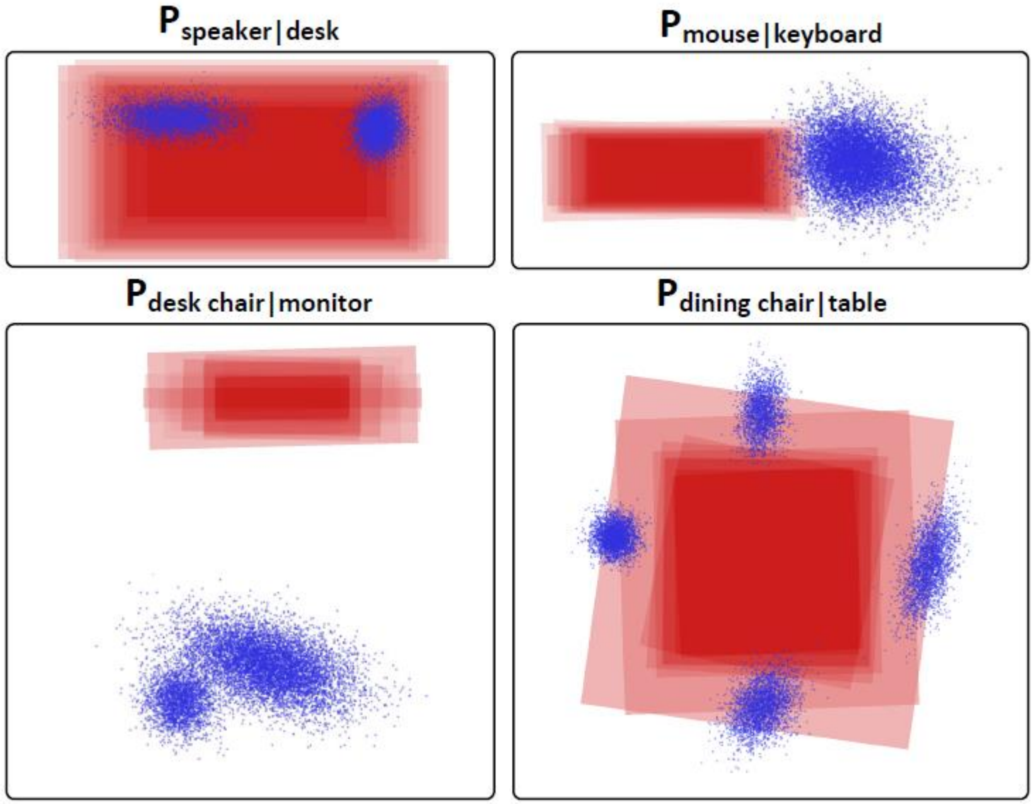
- Boolean random variables (desk supports keyboard?)
- Learn frequencies for pairs of categories
- Total probability is product over all objects in scene

$$P(t|c) = \prod_{o} P(C(o), C(support(o)))$$

Contextual Model Details

Spatial arrangements: $P(a|t,c)=\color{red}{R(a,t,c)}S(a,t,c)$

- Random variables for relative positions and orientations
- Pairwise distributions of spatial relationships

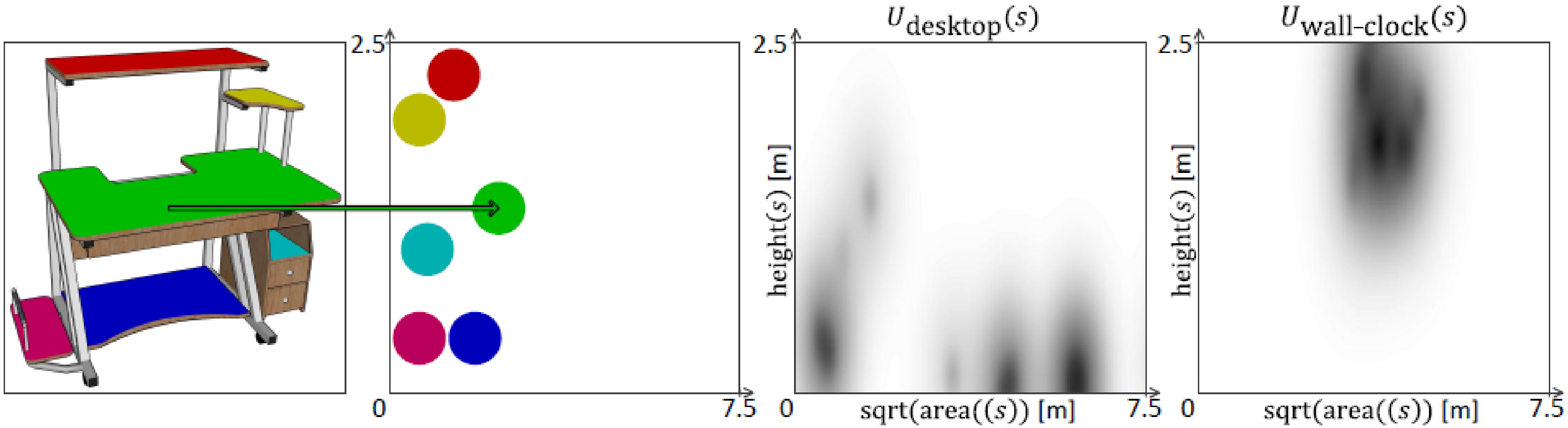


Distributions of spatial relationships for pairs of object categories

Contextual Model Details

Spatial arrangements: $P(a|t,c)=R(a,t,c)S(a,t,c)$

- Random variables for relative positions and orientations
- Pairwise distributions of spatial relationships
- Feature distributions for positions on support surfaces



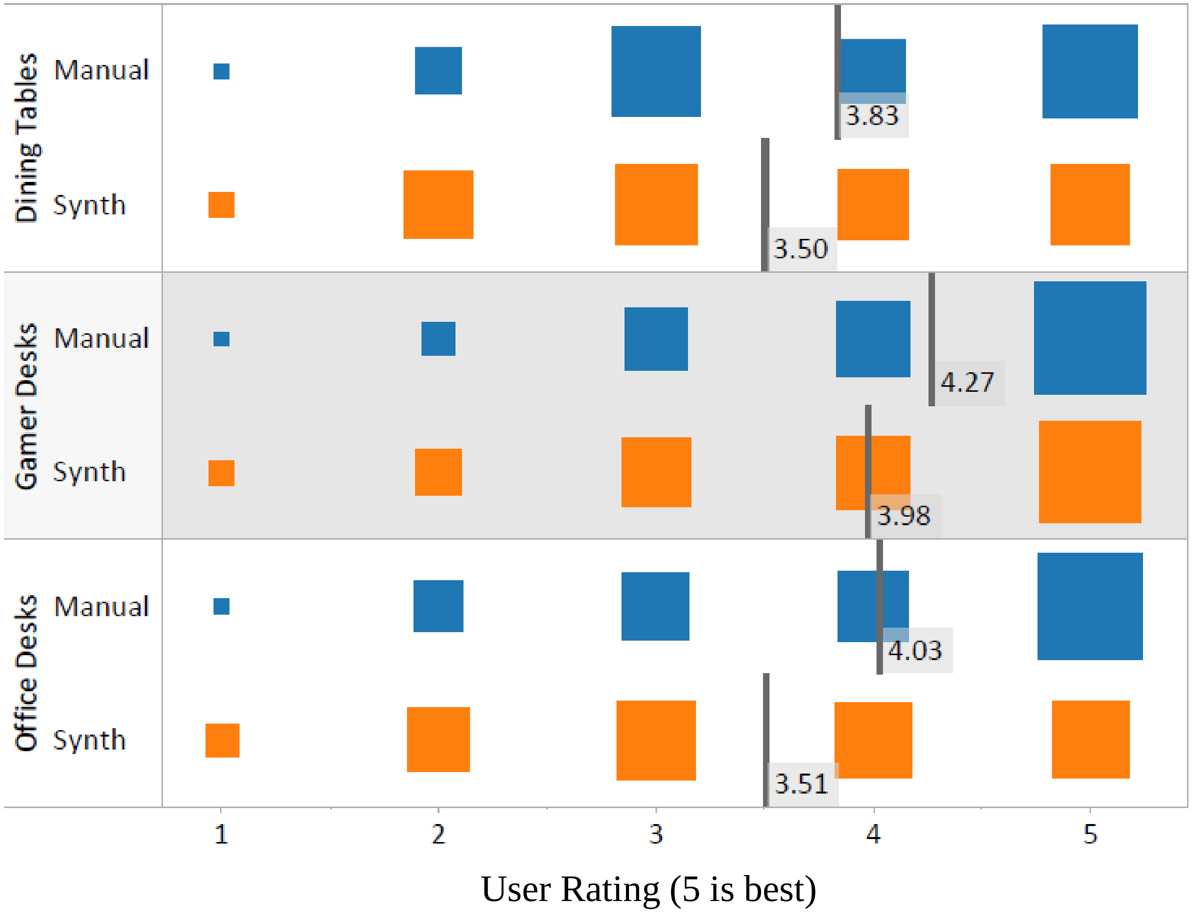
Distributions of geometric features of support surfaces

Scene Synthesis Results

Synthesized novel scenes

Scene Synthesis Results

User study suggests that people find our synthesized scenes almost as good as manually created ones



Outline of Talk

Introduction

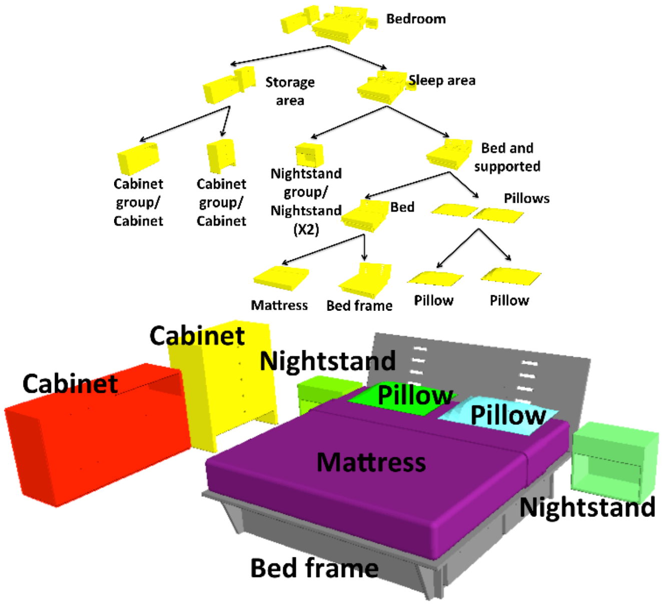
Learning probabilistic models from 3D collections

- Object templates
- Contextual model
- ➢ Hierarchical grammar

Conclusions

Tianqiang Liu, Sidhartha Chaudhuri, Vladimir Kim, Qixing Huang, Niloy Mitra, and Thomas Funkhouser,
Creating Consistent Scene Graphs Using a Probabilistic Grammar, SIGGRAPH Asia, 2014.

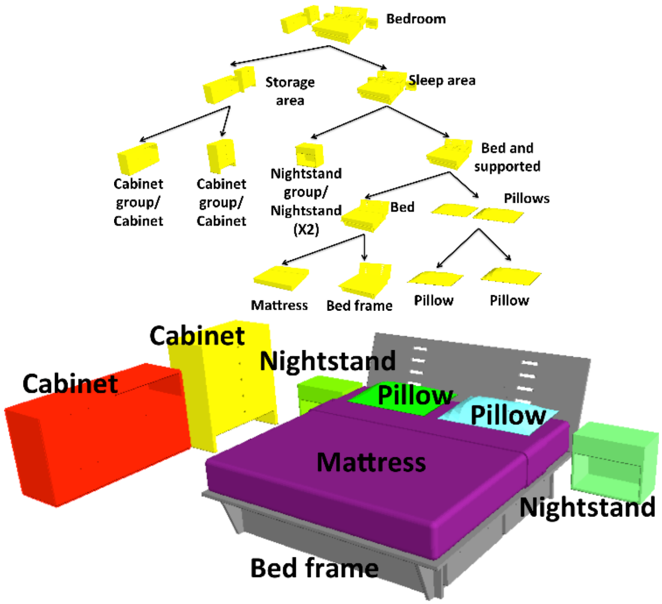
Goal for This Project



Training set of labeled scene graphs

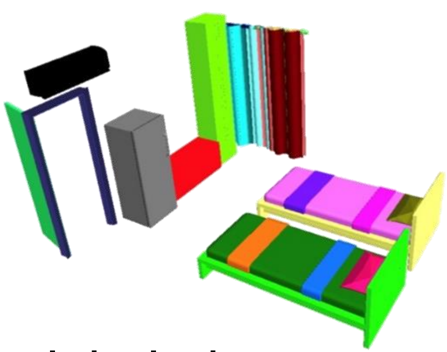
Probabilistic Model of Shape

Goal for This Project



Training set of labeled scene graphs

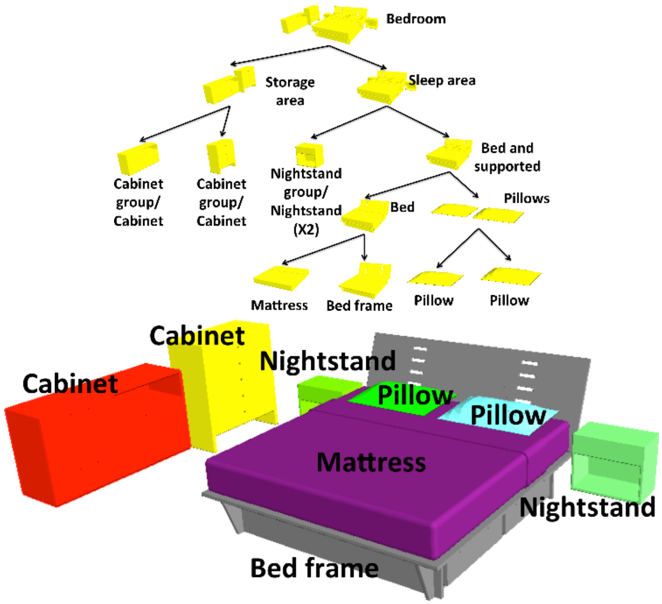
+



Unlabeled test scene

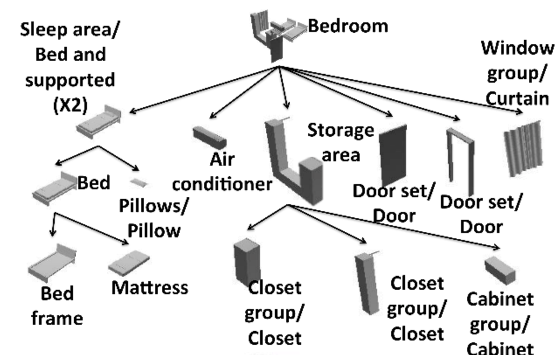
Probabilistic Model of Shape

Goal for This Project

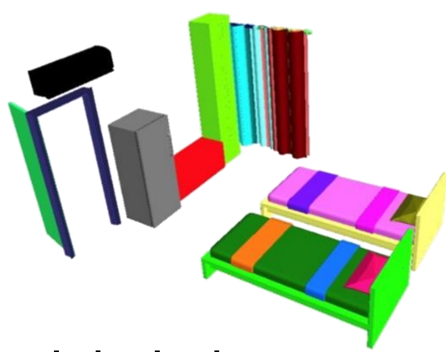


Training set of labeled scene graphs

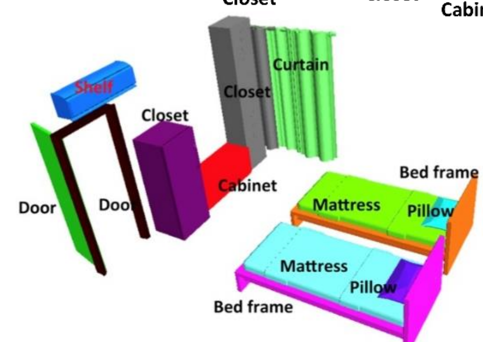
Probabilistic Model of Shape



+

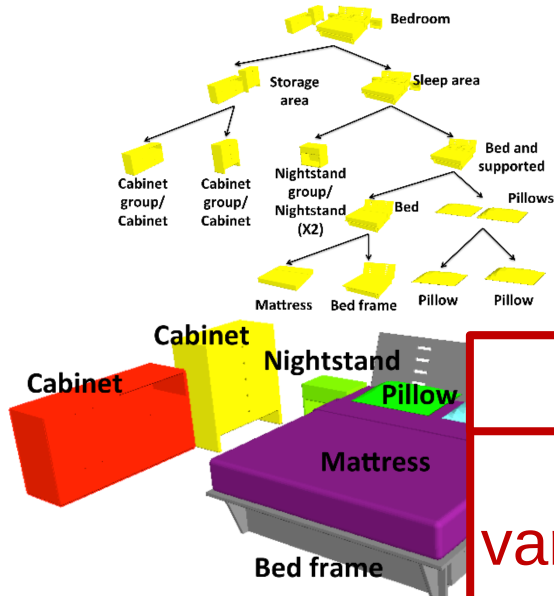


Unlabeled test scene



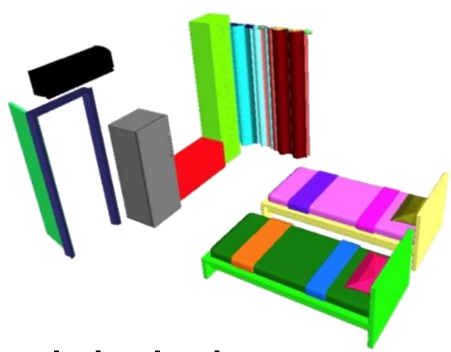
Labeled test scene graph

Goal for This Project



Training set of labeled

+

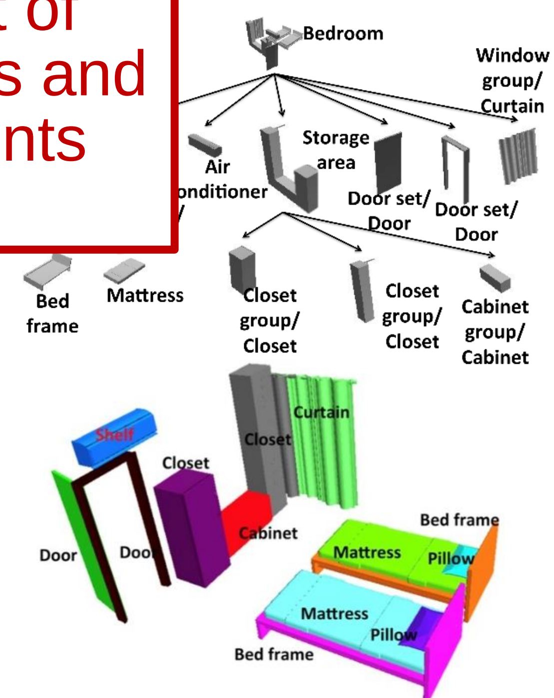


Unlabeled test scene

Probabilistic Model of Shape

Challenge

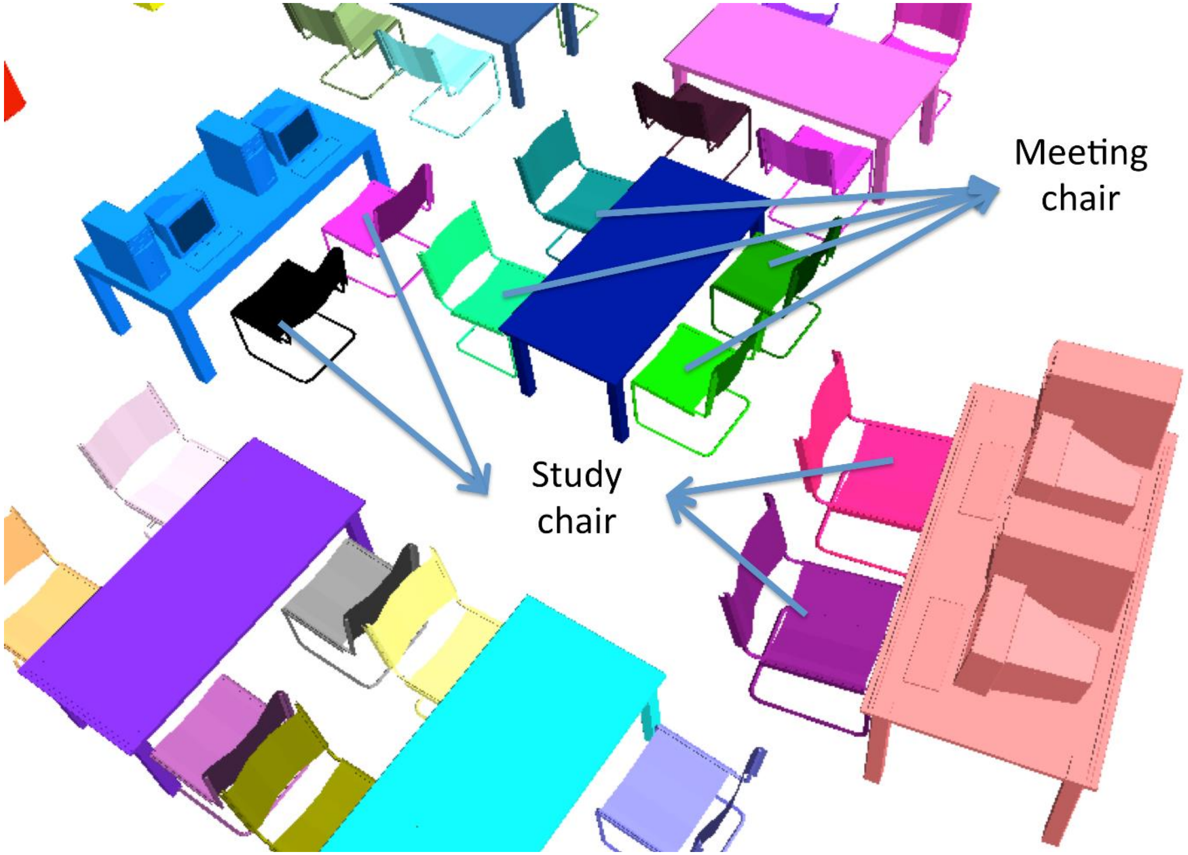
Scenes have a lot of variability in the types and spatial arrangements of objects

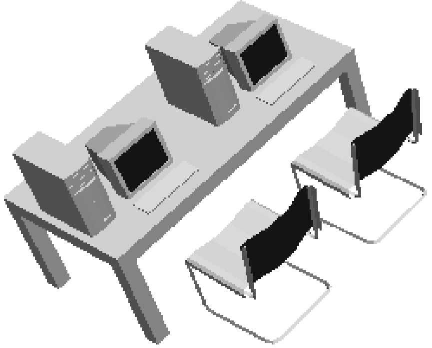


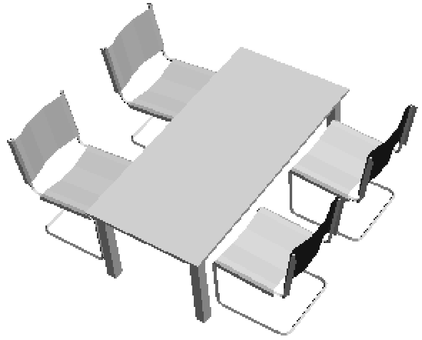
Labeled test scene graph

Observation

Semantic and functional relationships are often more prominent within hierarchical contexts

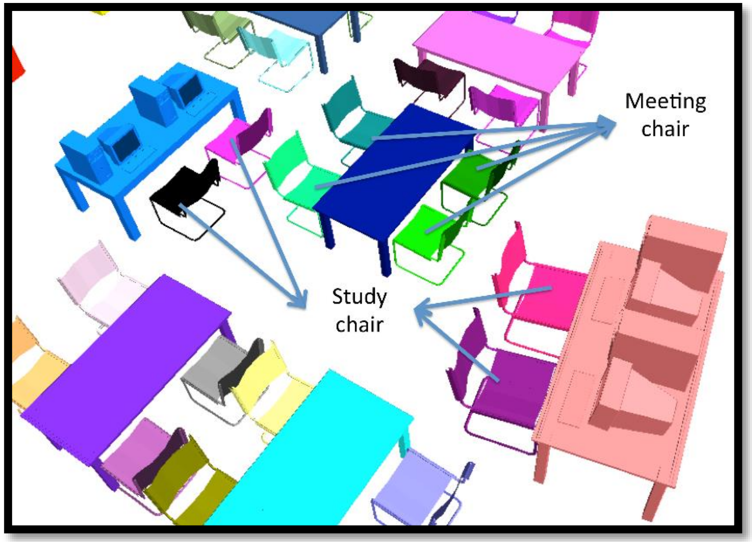


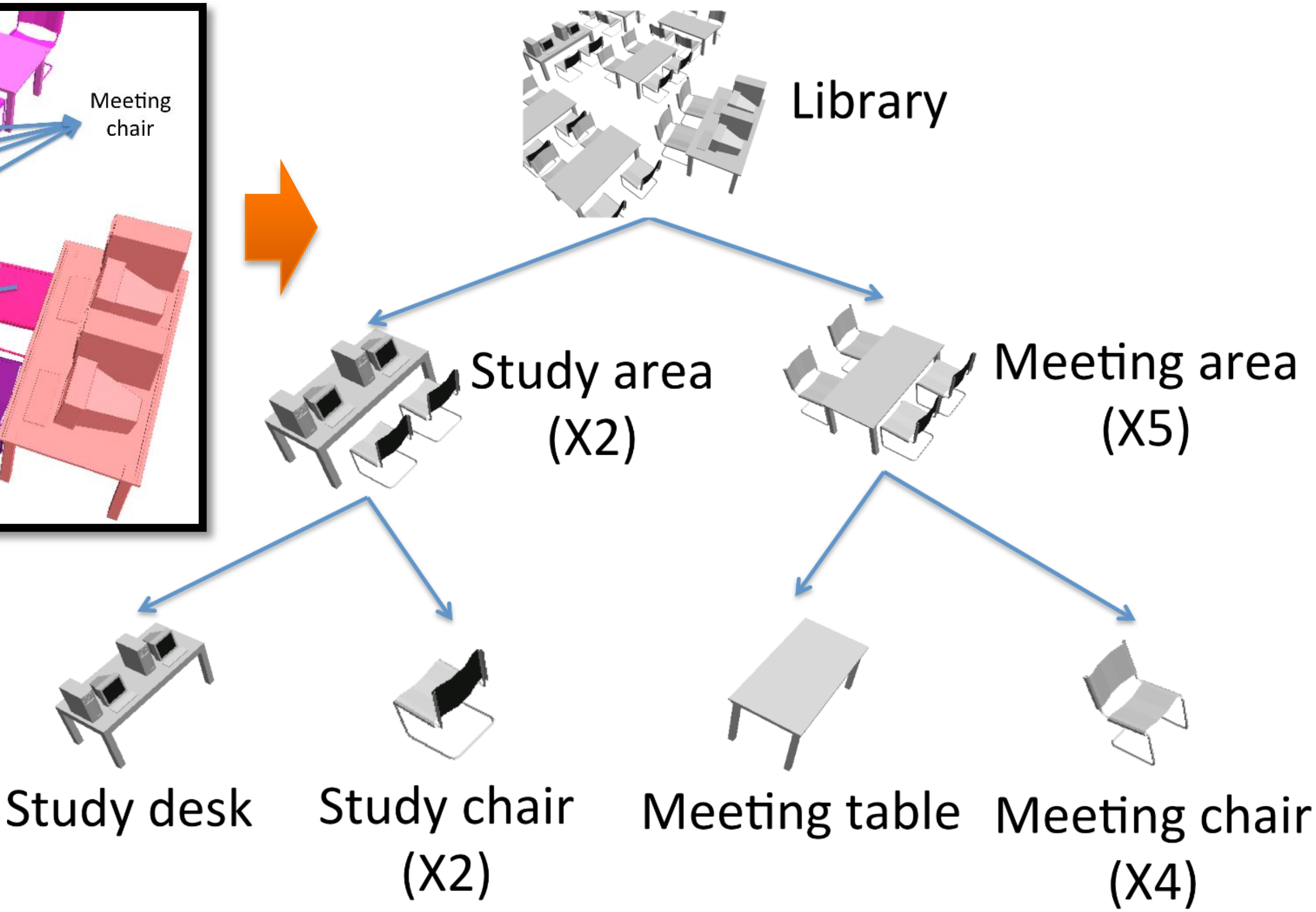




Hierarchical Grammar

We learn a hierarchical grammar from examples, and then use it to parse new test scenes





Hierarchical Grammar

Labels: object group, object category, object part

sleep area, bed, curtain piece

Rules: derivation from a label to a list of labels

bed $\rightarrow$ bed frame mattress

Hierarchical Grammar

Probabilities:

Derivation: $P_{nt}(rule \mid lhs)$

$$\text{bed} \xrightarrow[P\,=\,0.8]{} \text{frame} \quad \text{mattress}$$

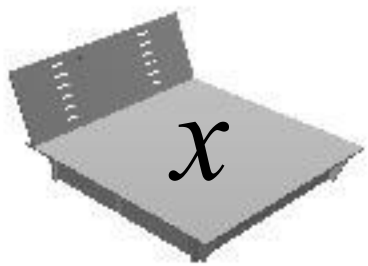
Cardinality distribution: $P_{Card}(\#, rhs \mid lhs)$

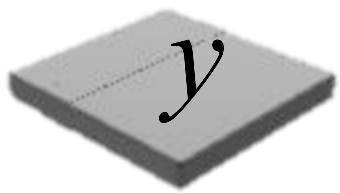
$$\text{sleep area} \rightarrow \text{bed} \quad \text{nightstand} \quad \text{rug} \quad \dots$$

$P_{card}(* \mid sleeparea)$	**0**	**1**	**2**	**3**	**4+**
bed	…	…	…	…	…
nightstand	0.3	0.3	0.4	0	0
rug	…	…	…	…	…

Hierarchical Grammar

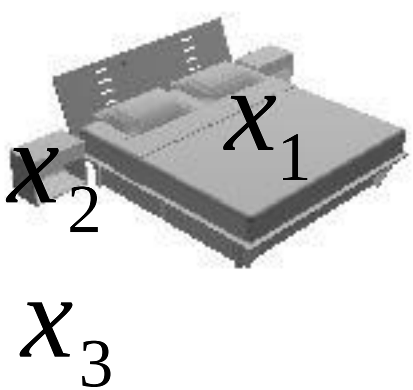
Shape descriptor probability: $P_g(x \mid label)$





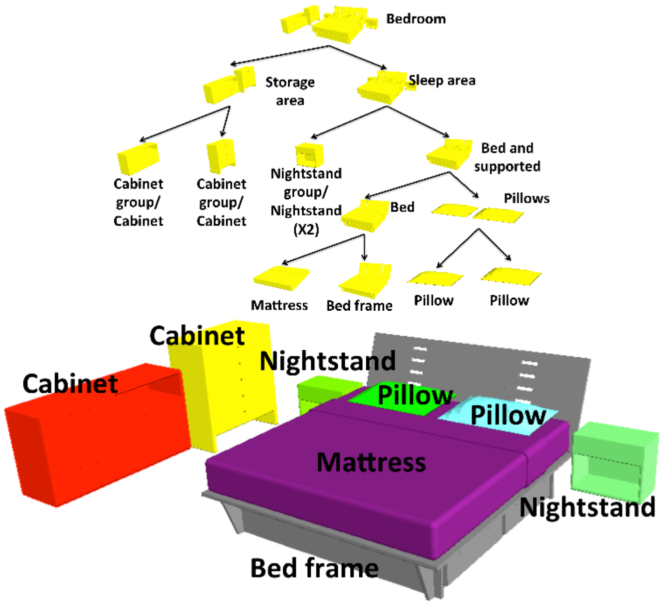
$$P_g(x \mid bedframe) > P_g(y \mid bedframe)$$

Spatial relationships: $P_g(v \mid lhs, rhs1, rhs2)$



$$P_s(x_1, x_2 \mid sleeparea, bed, nightstand) >$$
$$P_s(x_1, x_3 \mid sleeparea, bed, nightstand)$$

Grammar Learning and Parsing



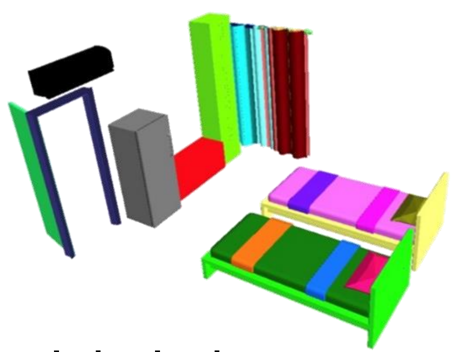
Training set of labeled scene graphs

Learn

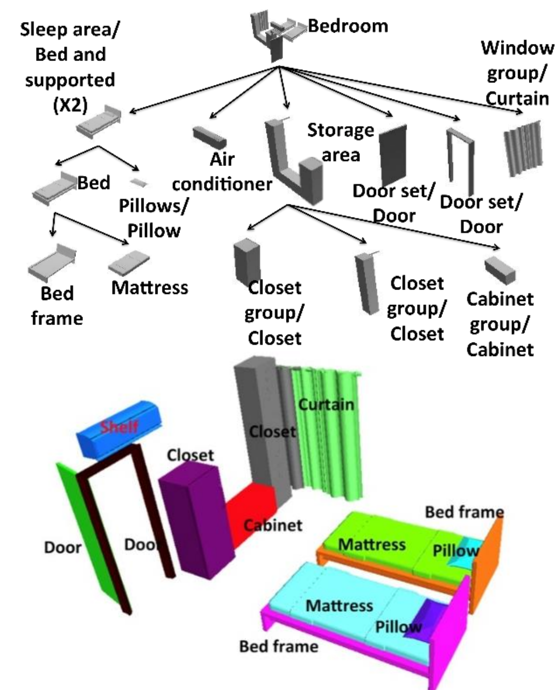
Probabilistic Hierarchical Grammar

Parse

+



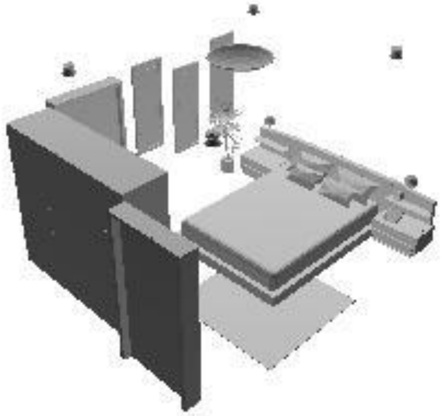
Unlabeled test scene



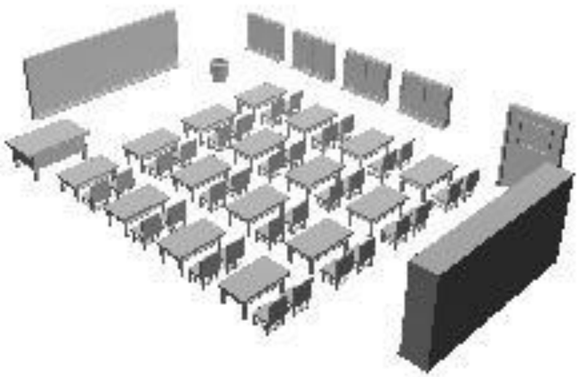
Labeled test scene graph

Hierarchical Grammar Results

Learned hierarchical probabilistic grammars from scenes in Trimble 3D Warehouse

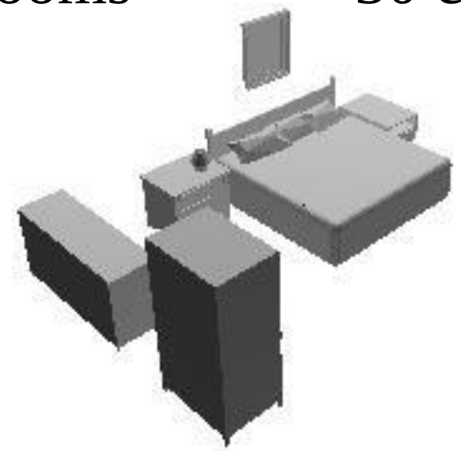


77 bedrooms

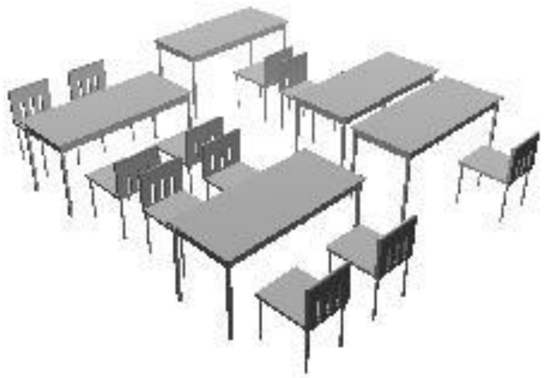


30 classrooms

8 libraries



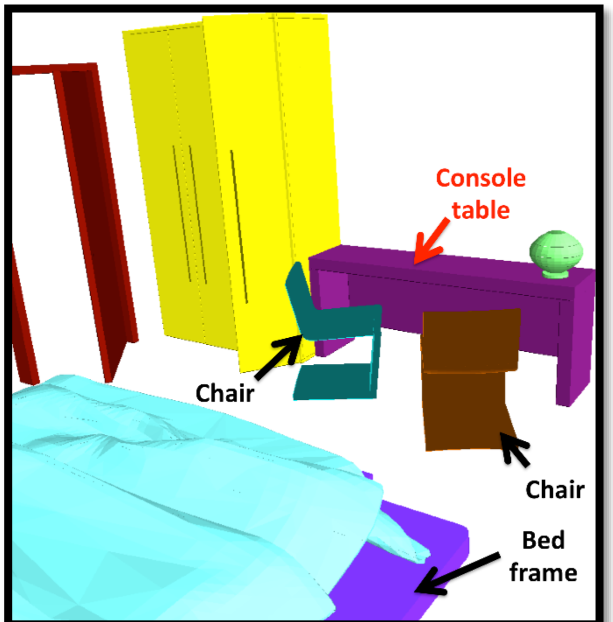
17 small bedrooms



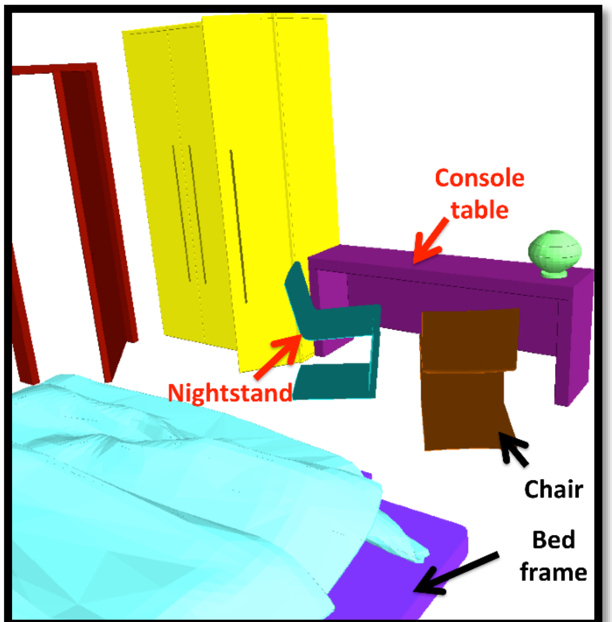
8 small libraries

Hierarchical Grammar Results

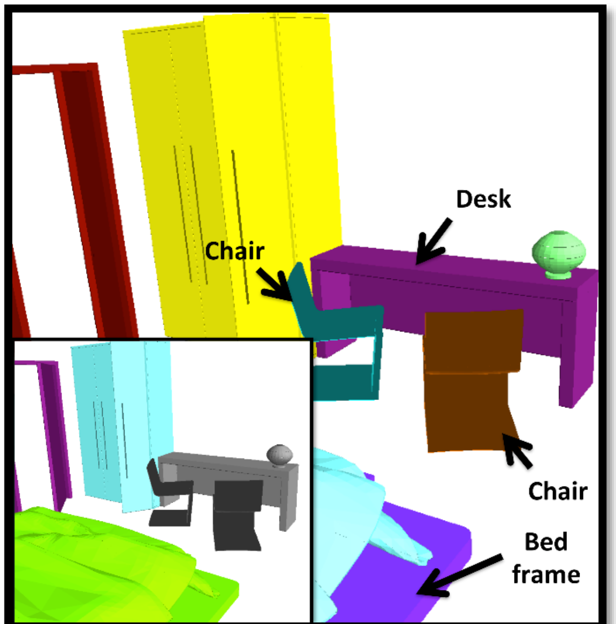
Parsed left-out scenes with learned grammar



Shape Only



Flat Grammar

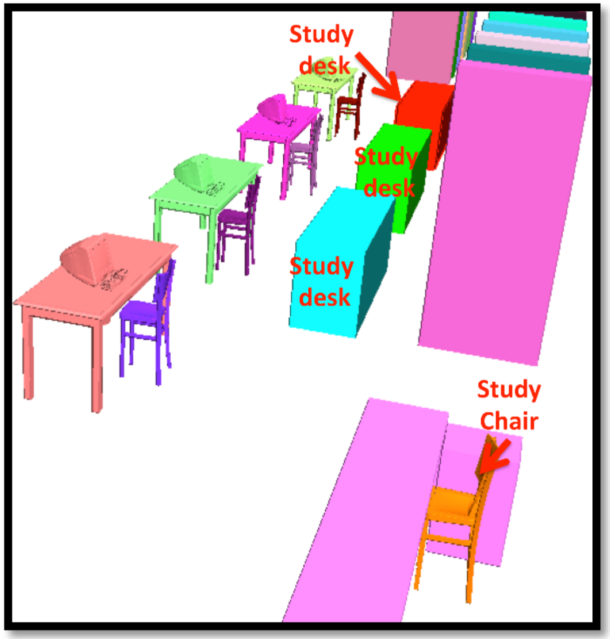


Our Hierarchical Grammar

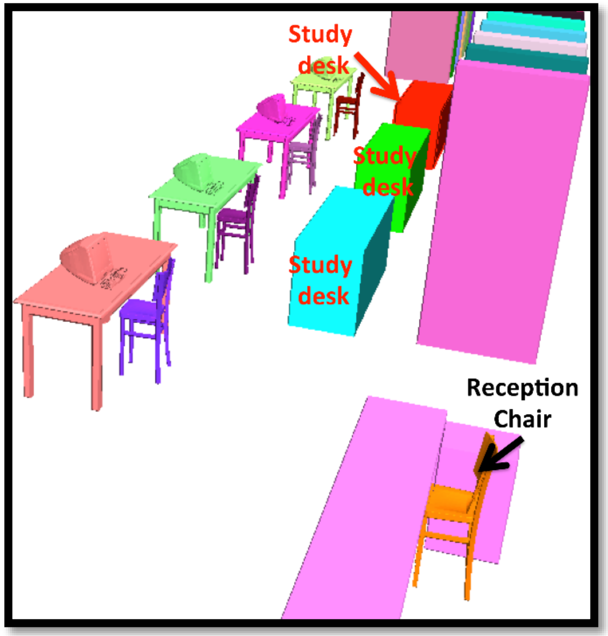
Comparison of our parsing results to other methods

Hierarchical Grammar Results

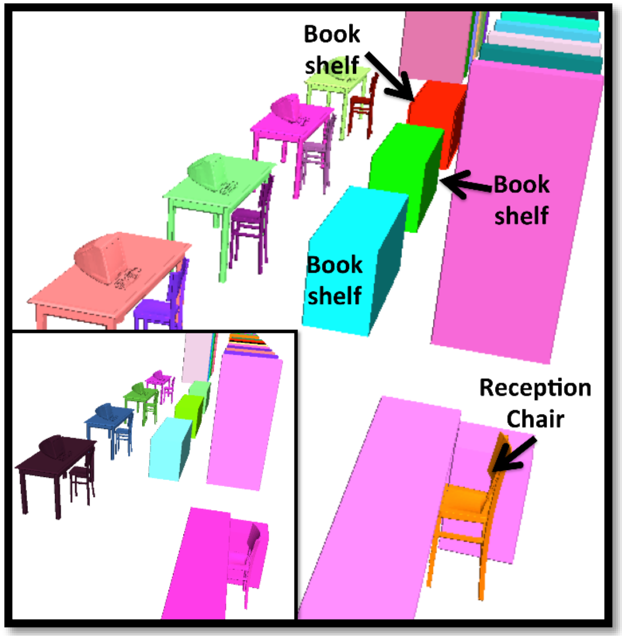
Parsed left-out scenes with learned grammar



Shape Only



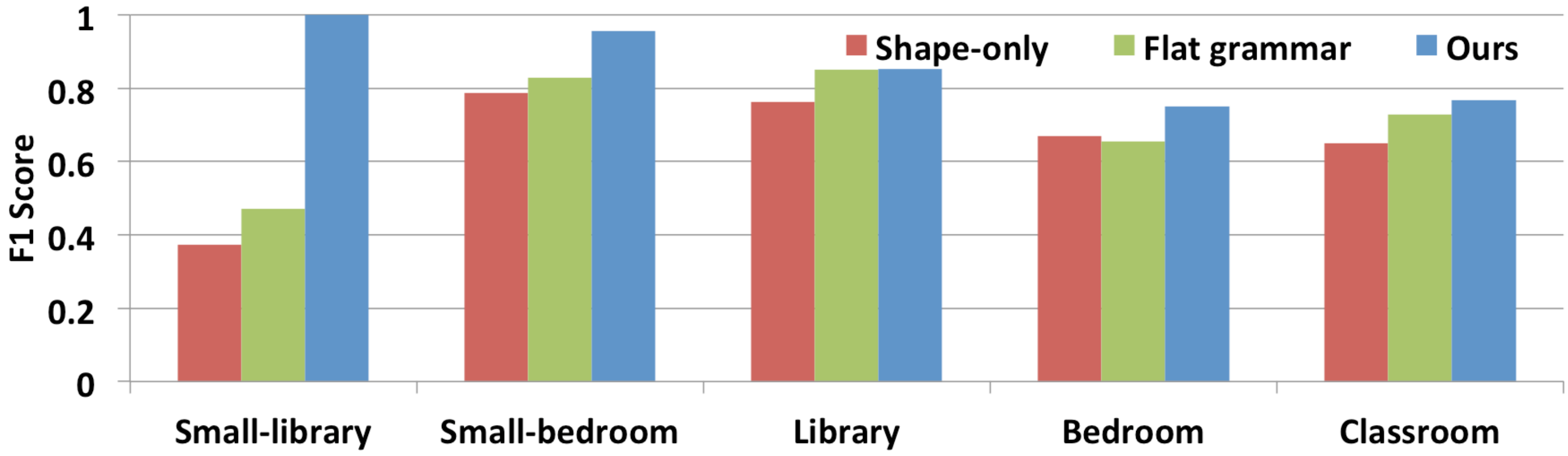
Flat Grammar



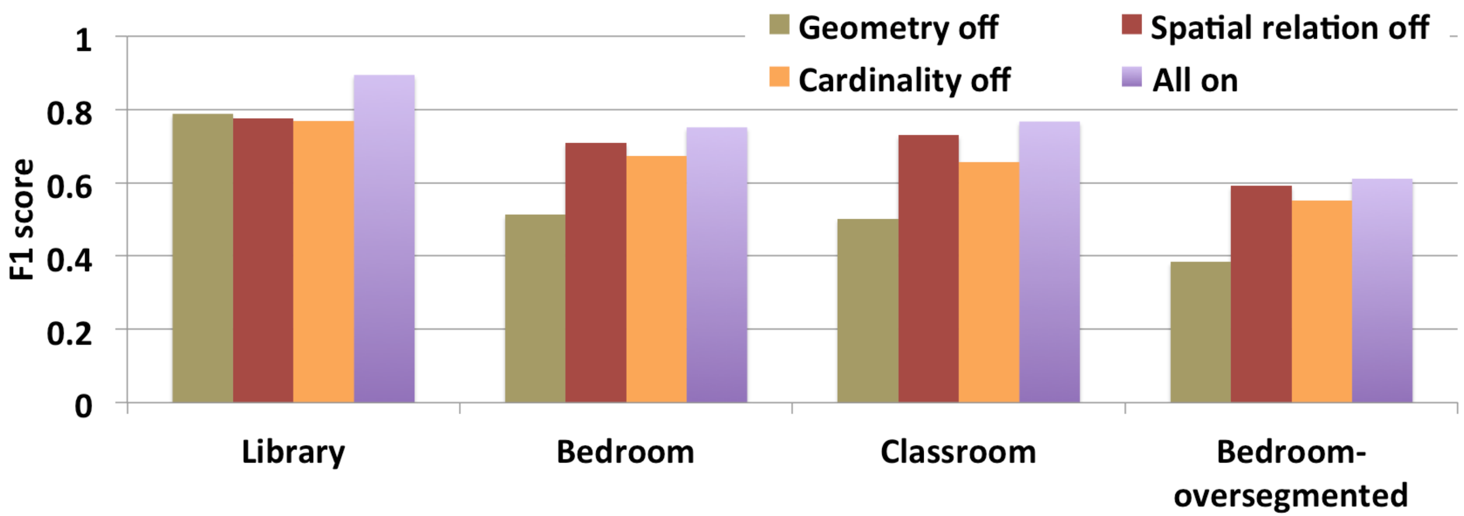
Our Hierarchical Grammar

Comparison of our parsing results to other methods

Hierarchical Grammar Results



Comparison of object classification



Impact of Individual Energy Terms

Outline of Talk

Introduction

Learning probabilistic models from 3D collections

- Part-based templates
- Generative model
- Hierarchical grammar

➢ Conclusions

Conclusions

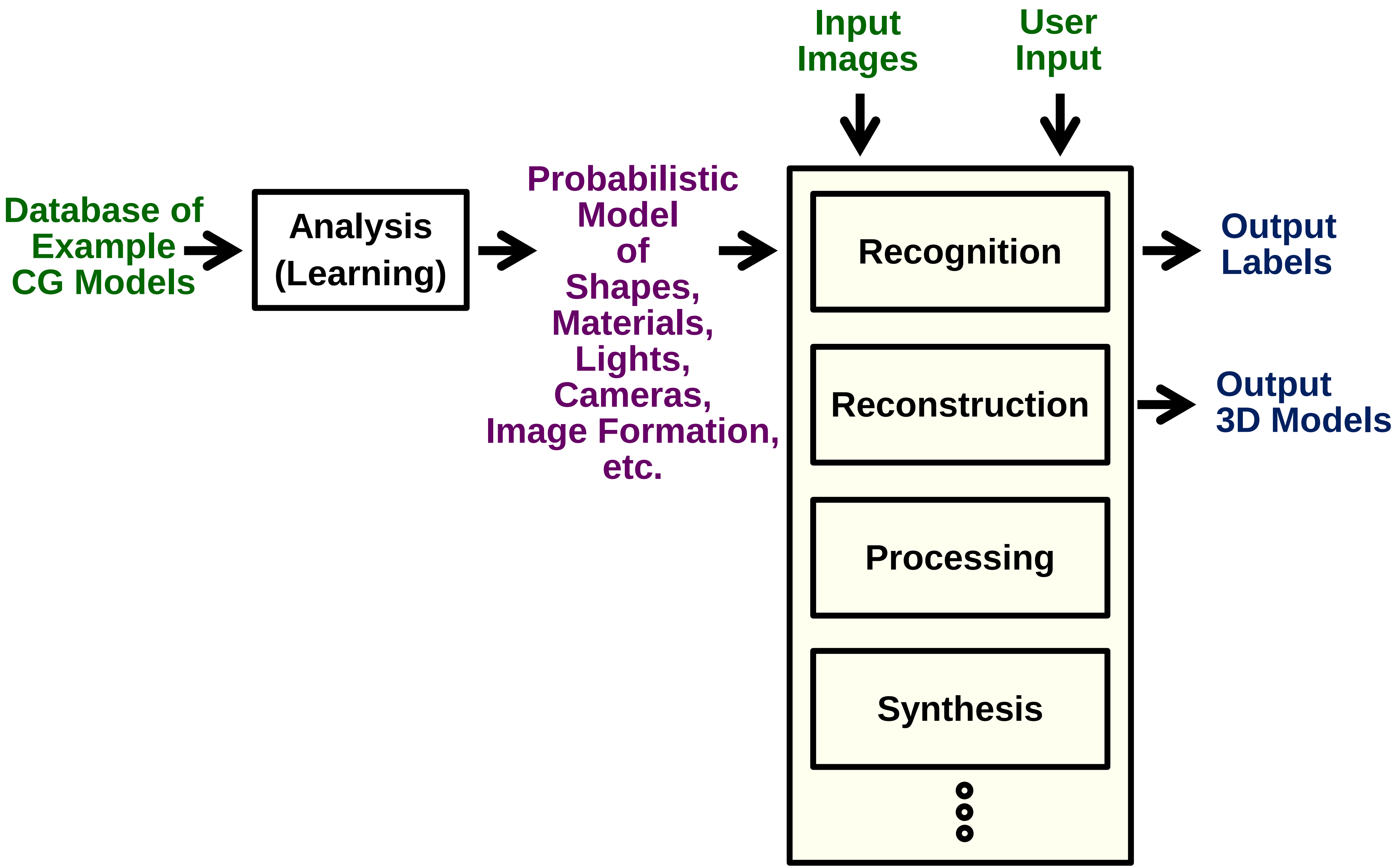
Main result:

- Probablistic models can be learned from collections of 3D meshes

Future work:

- Learn probabilistic models of lighting, materials, cameras
- Use these models for understanding scenes captured in scans and images

Conclusions



Acknowledgments

People:

- Sid Chaudhuri, Steve Diverdi, Matthew Fisher,
  Pat Hanrahan, Qixing Huang, Vladimir Kim, Wilmot Li,
  Tianqiang Liu, Niloy Mitra, Daniel Ritchie, Manolis Savva

Data sets:

- Trimble 3D Warehouse

Funding:

- NSF, Intel, Google, Adobe

Thank You!