Transferring Information across Image and Shape Collections

Niloy J. Mitra

UCL

GeometryProcessing@UCL

UCL

erc

SEVENTH FRAMEWORK PROGRAMME

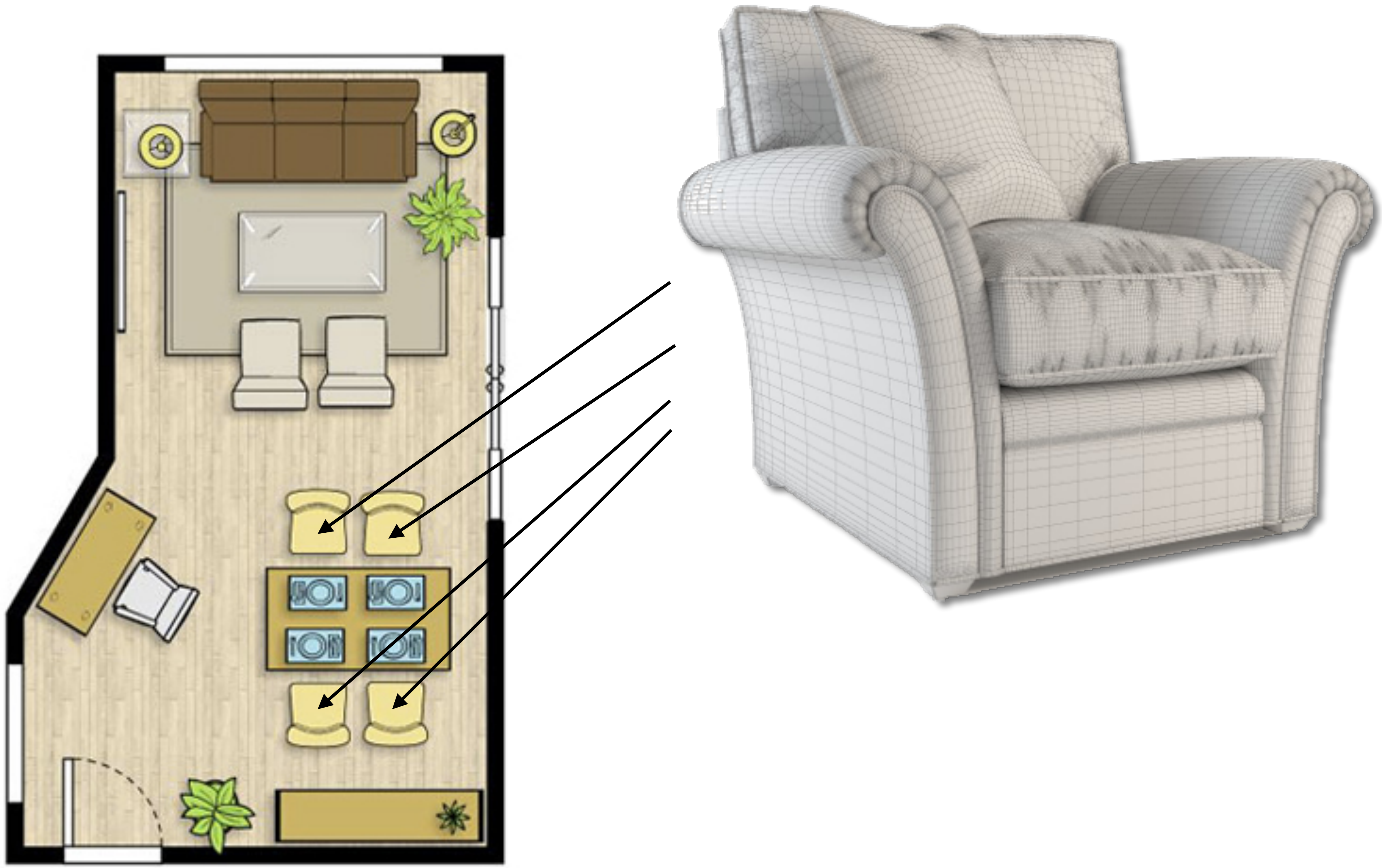
Adobe

Microsoft Research

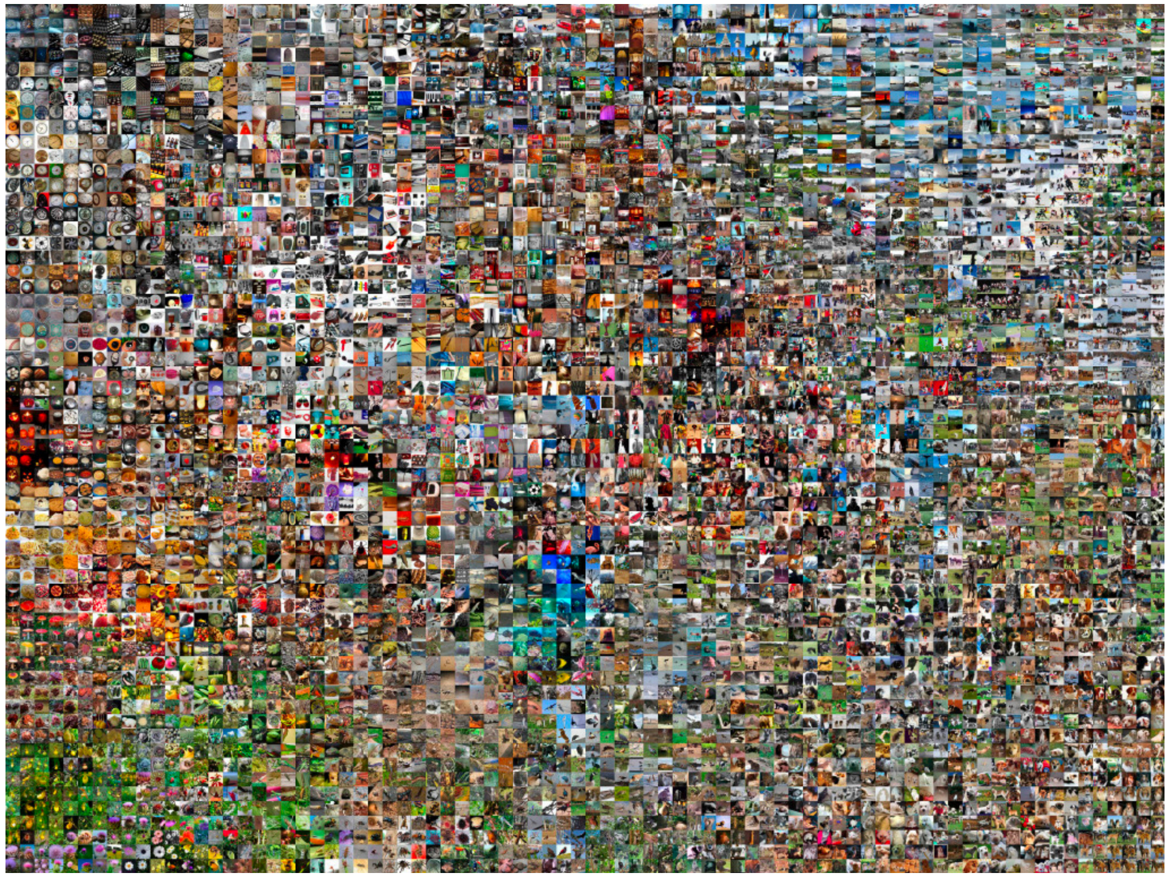
UCL ENGINEERING
Change the world


sofa
?

3D MockUp



Data Sources



images

3D models

Overview

UCL

Observations
(images)

dynamics
(actors, physics, etc.)

Objects
(3D geometry)

Illumination
(light models)

Priors
(shape manifold,
illumination prior,
room priors, etc.)

Appearance
(diffuse, specular, texture)

Overview

Observations
(images)

dynamics
(actors, physics, etc.)

Objects
(3D geometry)

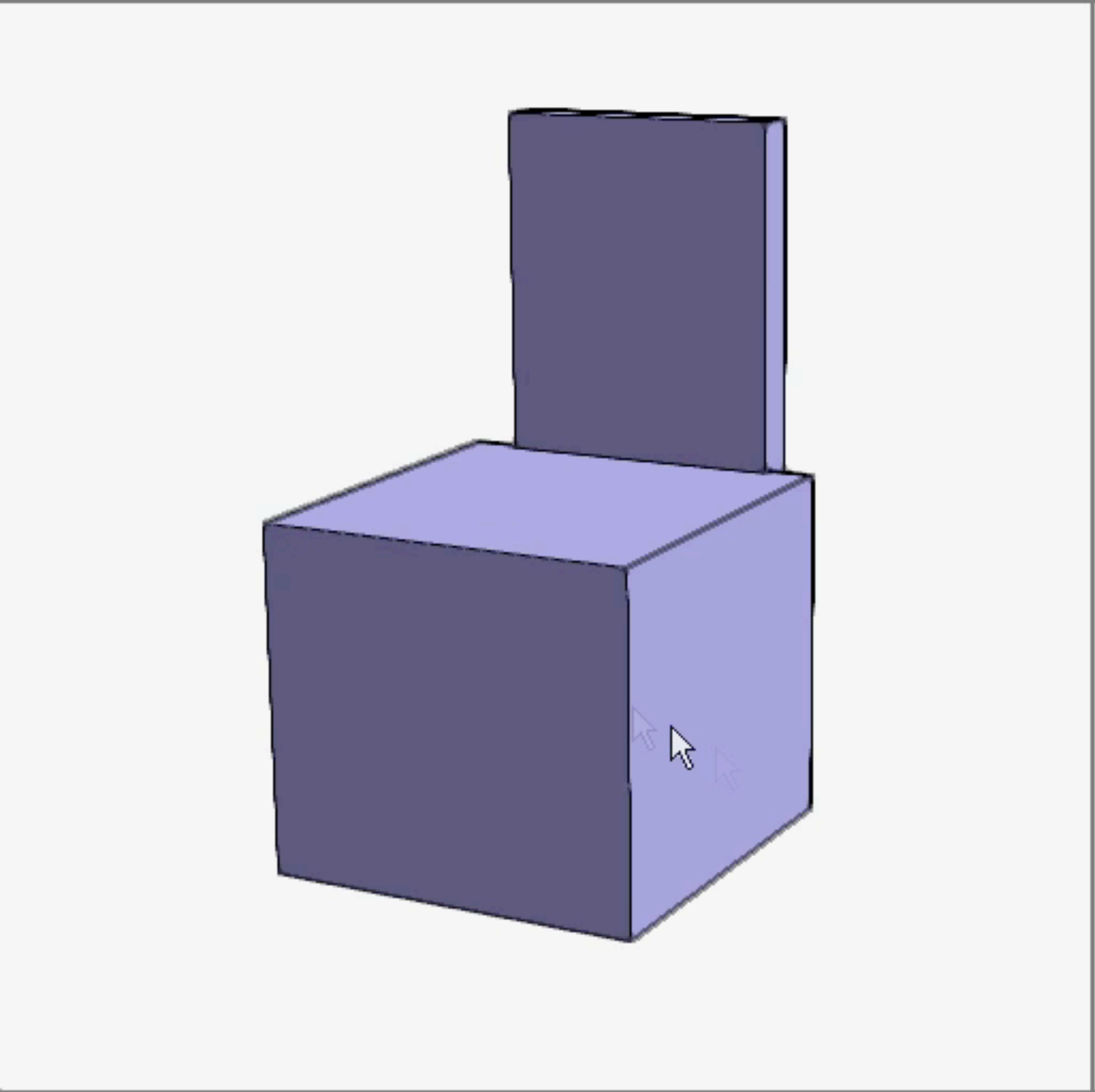
Illumination
(light models)

Appearance
(diffuse, specular, texture)

**Priors
(shape manifold,
illumination prior,
room priors, etc.)**

Encoding Shape Manifold (Prior)

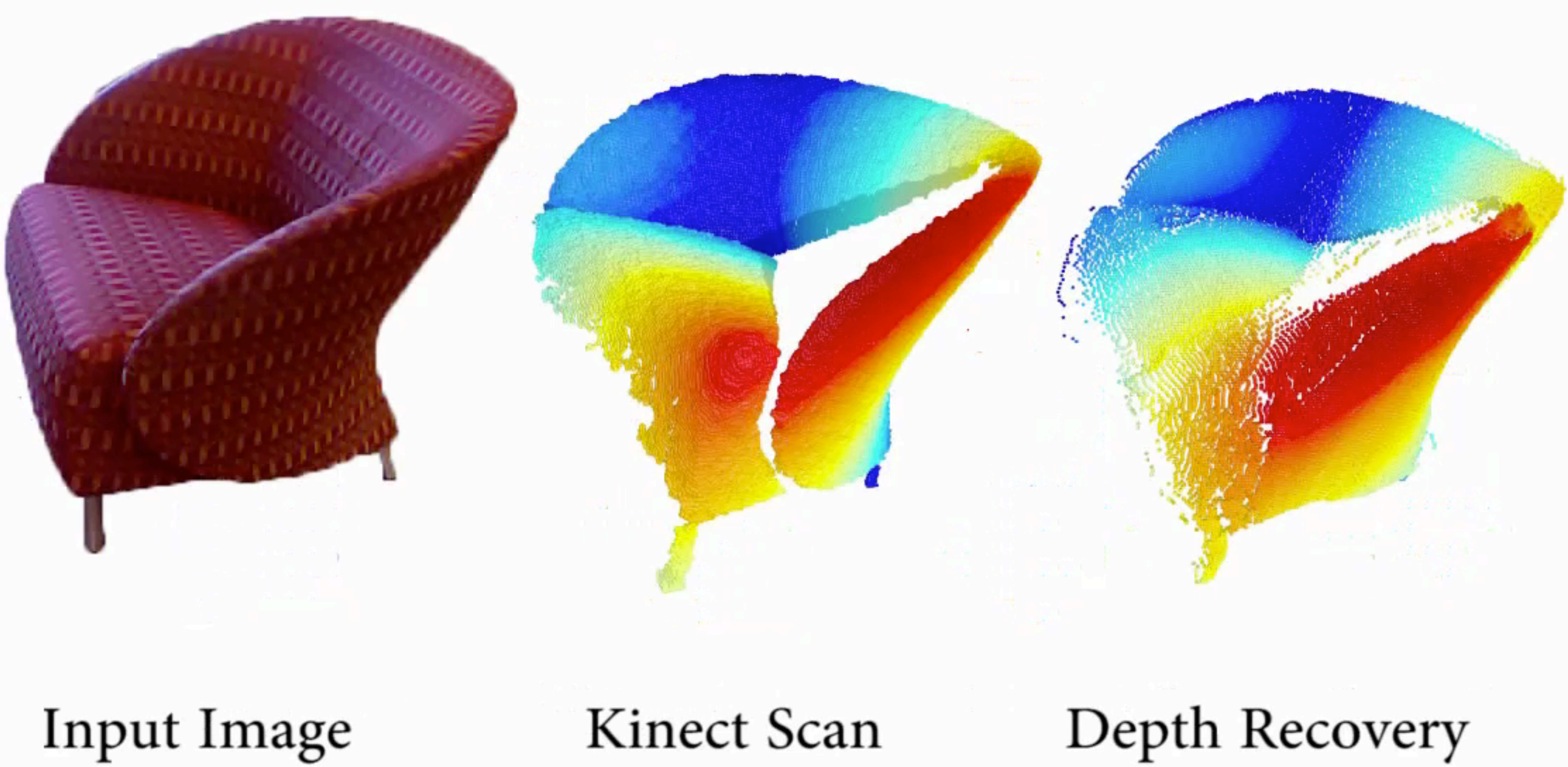
[SIGGRAPH 2011, 2013]



low DOF shape variation $\Longrightarrow$ smooth curves in signature space

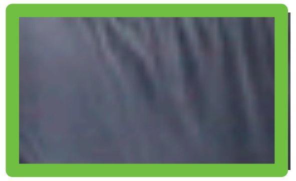
Single Image ⇒ (fused) PCD

[SIGGRAPH 2014]



retrieve shapes/views ⟹ create partial PCD






Overview

UCL

Observations (images)

dynamics (actors, physics, etc.)

Objects (3D objects)

Illumination (light models)

Priors (shape manifold, illumination prior, room priors)

Appearance (diffuse, specular, texture)

CrossLink
images $\Leftrightarrow$ models

Image search "airplane"

UCL

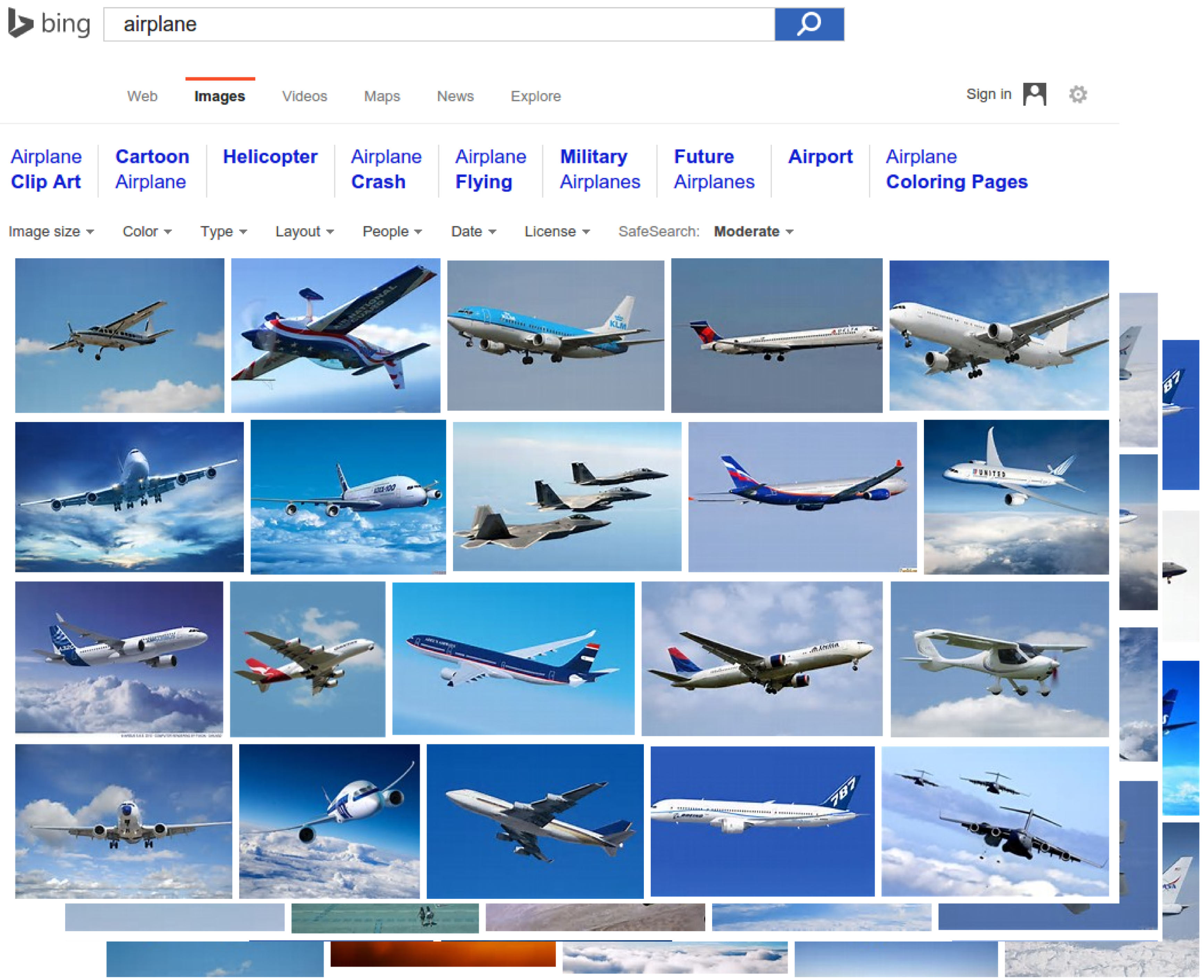


Image search “airplane”

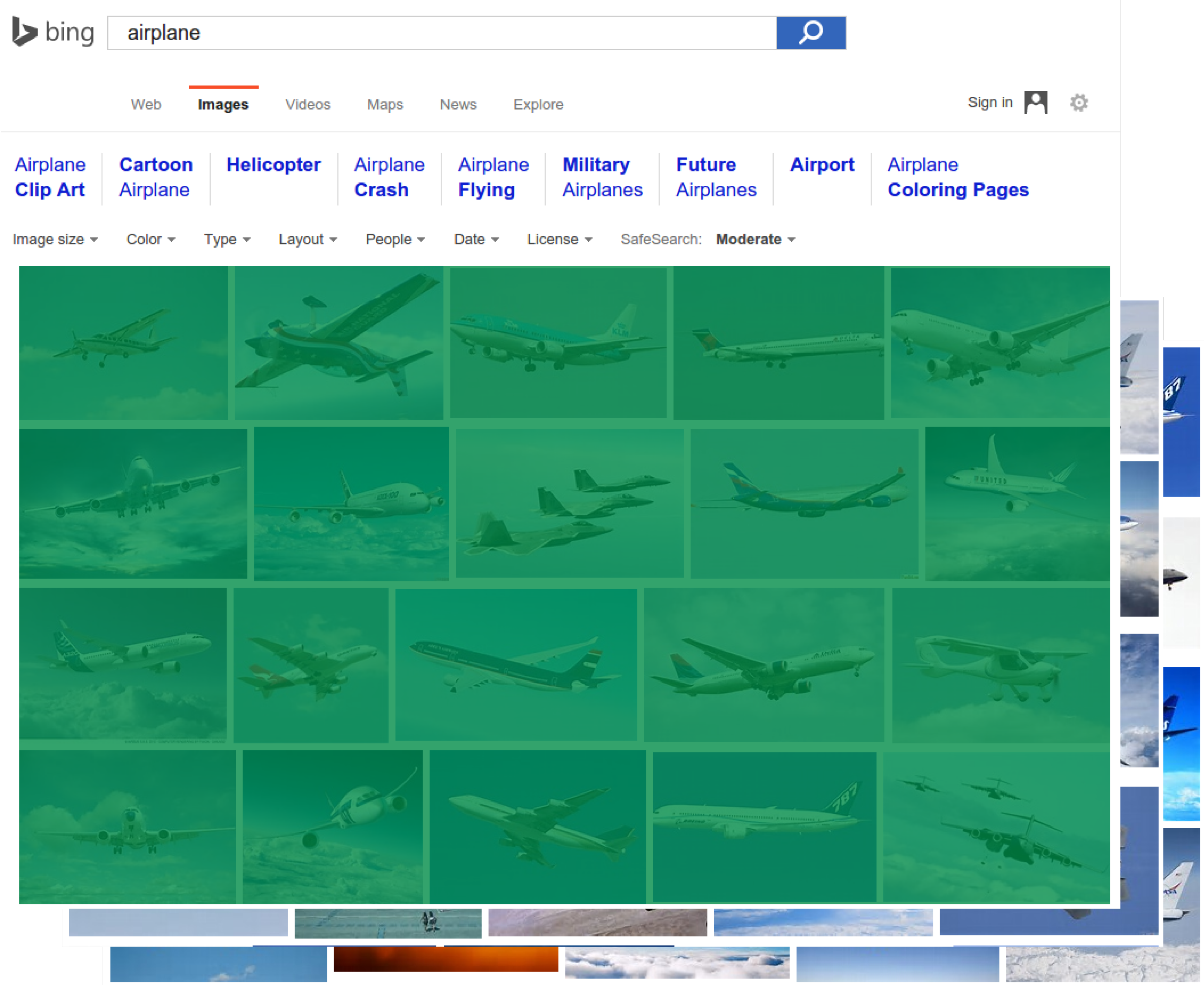
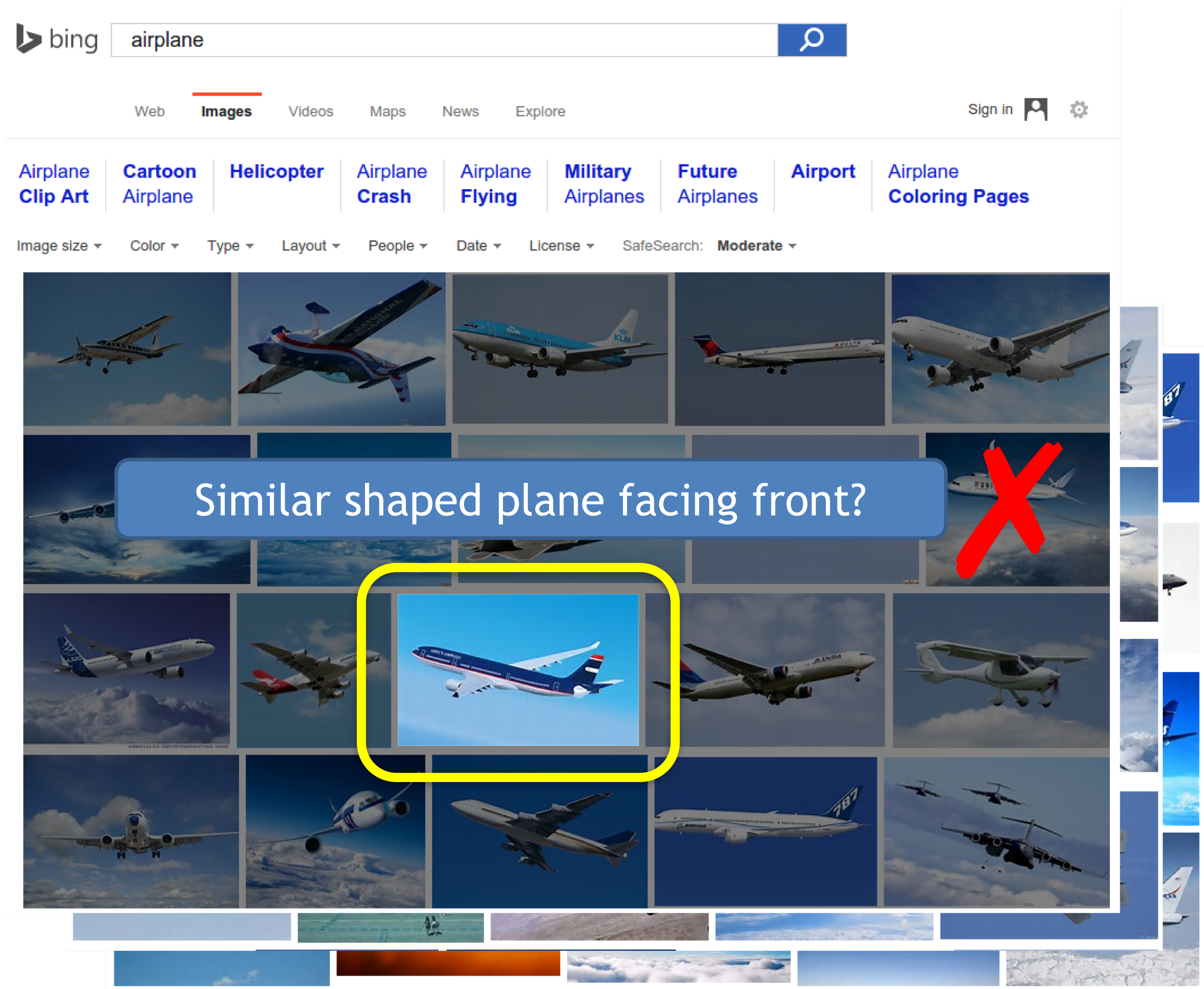
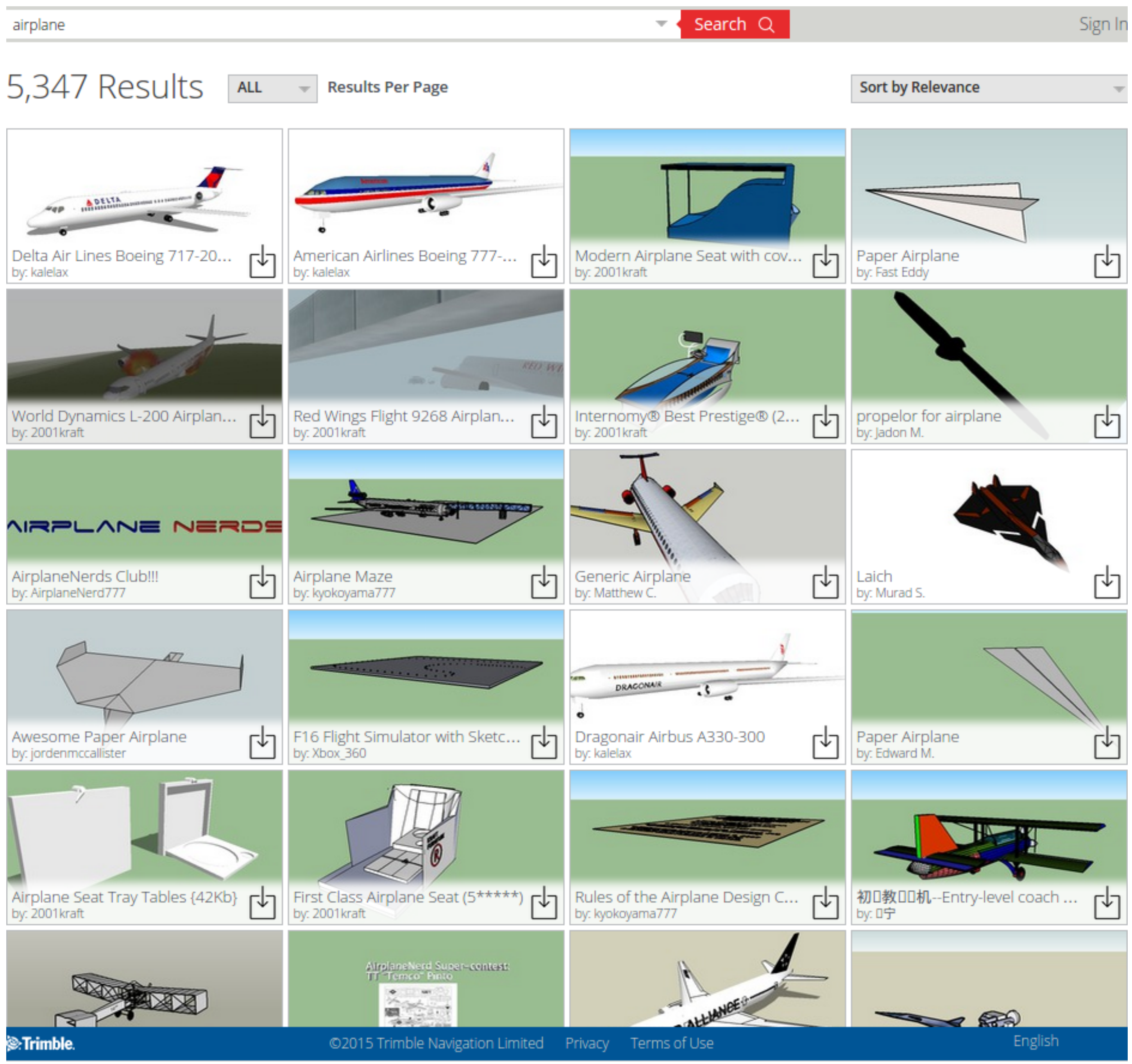


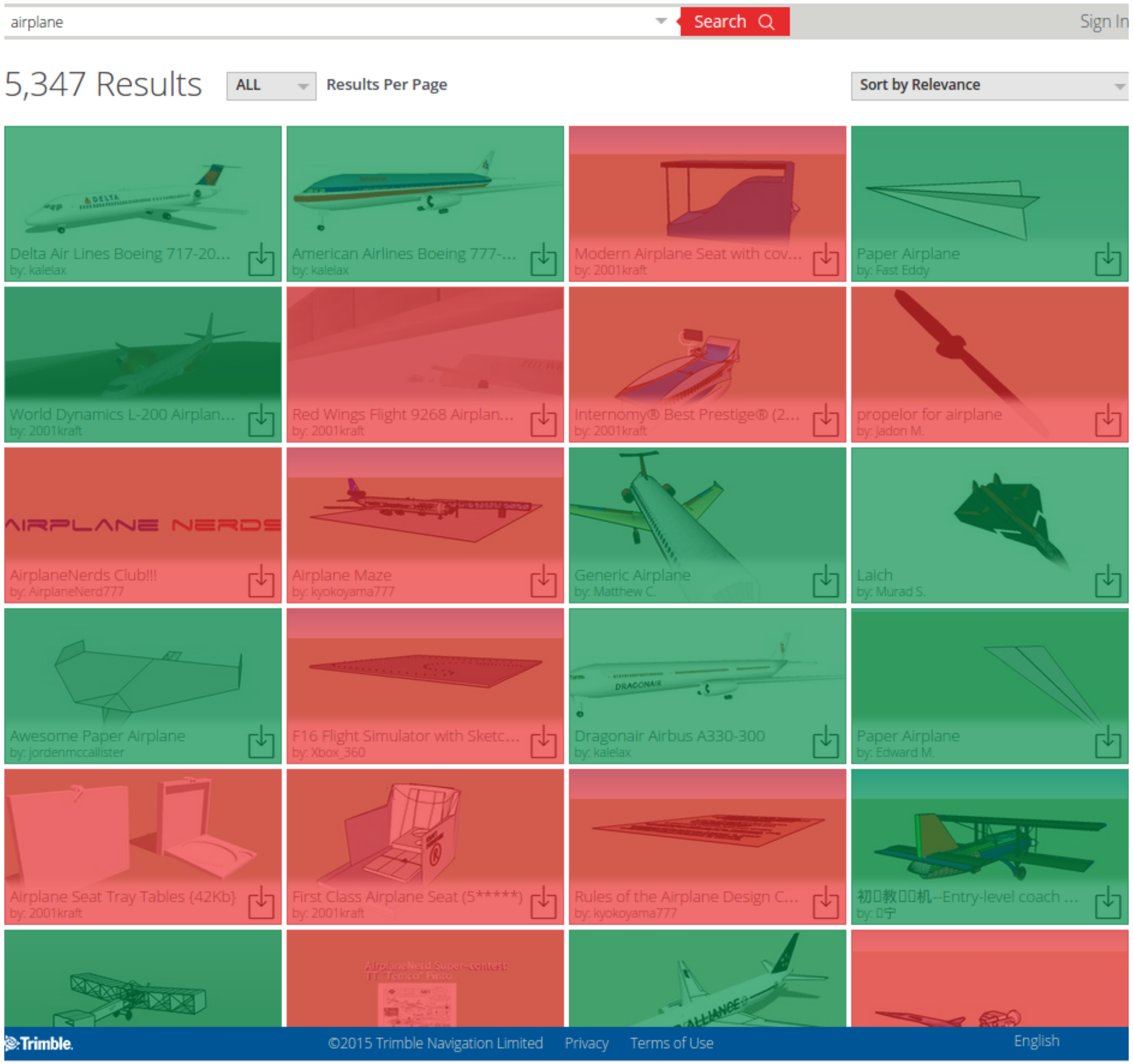
Image search "airplane"



3D model search "airplane"



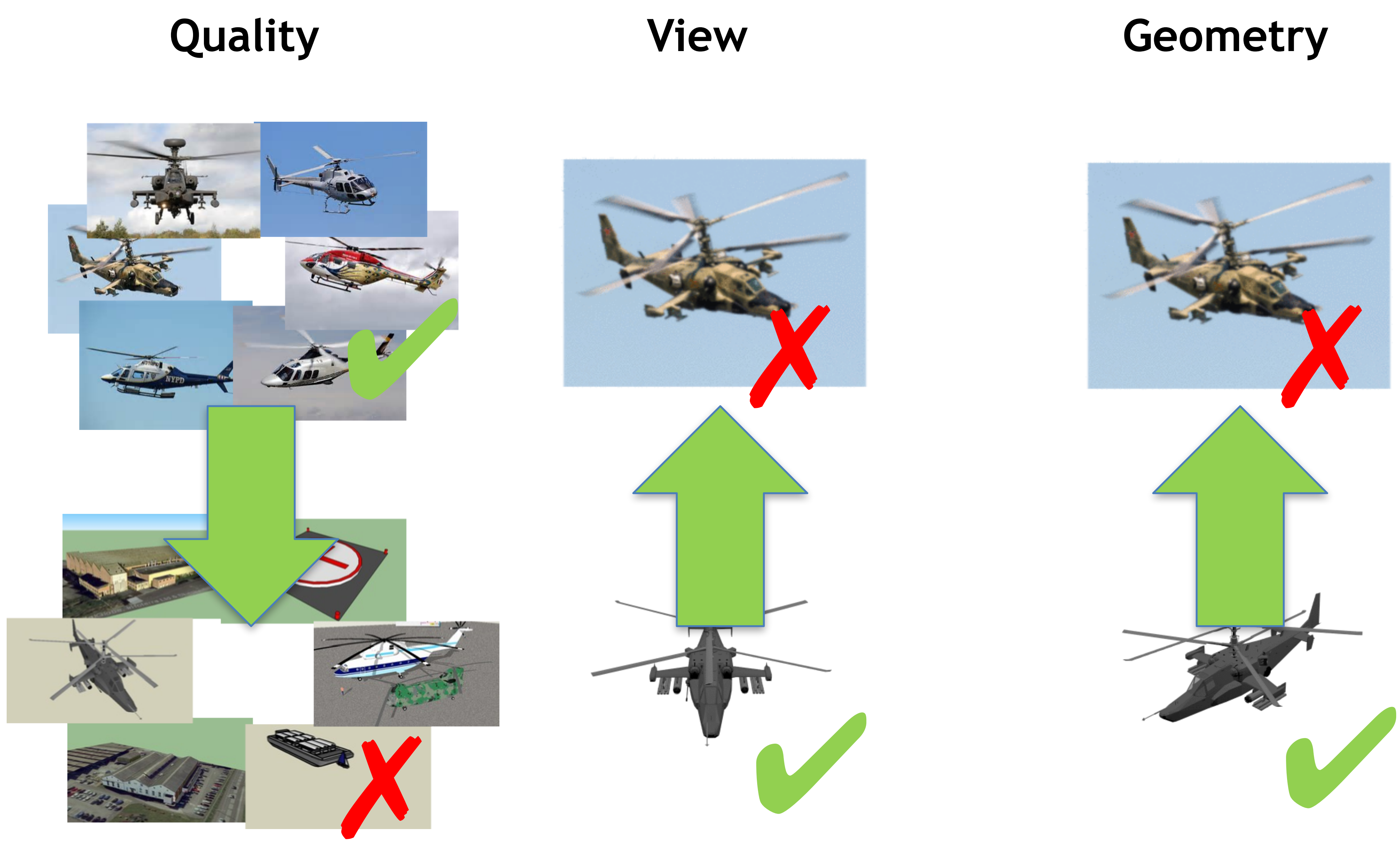
3D model search “airplane”



To Summarize

	Quality	View	Geometry
2D data			
3D data			

Link the Data Sources



Related work – search

Image search

- Both supervised and unsupervised (surveys [Datta et al. 2008; Zhang and Rui 2013])
- Linked with text [Weston et al. 2001; Masci et al. 2014; Pereira et al. 2014]

Shape search

- Text-based [Min et al. 2004]
- Content-based [Eitz et al. 2012; Li et al. 2015]

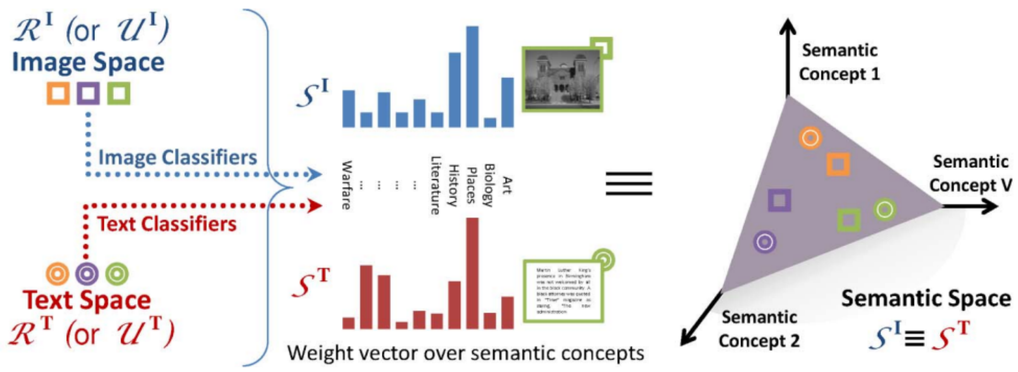


Fig. 6. Semantic matching. Text and images are mapped into a common semantic space, using the posterior class probabilities produced by a multiclass text or image classifier.

Pereira et al. 2014

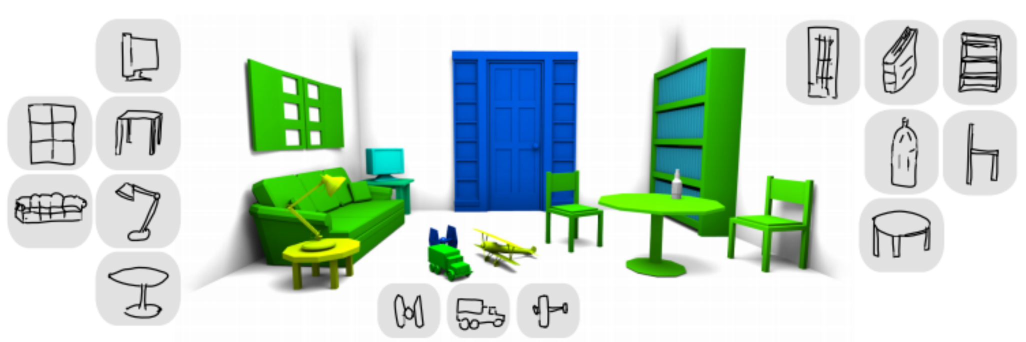


Figure 1: *A complete scene with objects retrieved using our sketch-based system in a total time of about two minutes.*

Eitz et al. 2012

Key Ideas

1. Push image ordering to **reorder** 3D models

2. 3D model **coalignment**

   Exploit loop structure and pose as a reshuffling problem

3. Push 3D shape renderings to **view-classify** images

   Map view variation to SVM changes

4. Push 3D **shape attributes** to images

no p2p correspondence

no direct links

no background subtraction

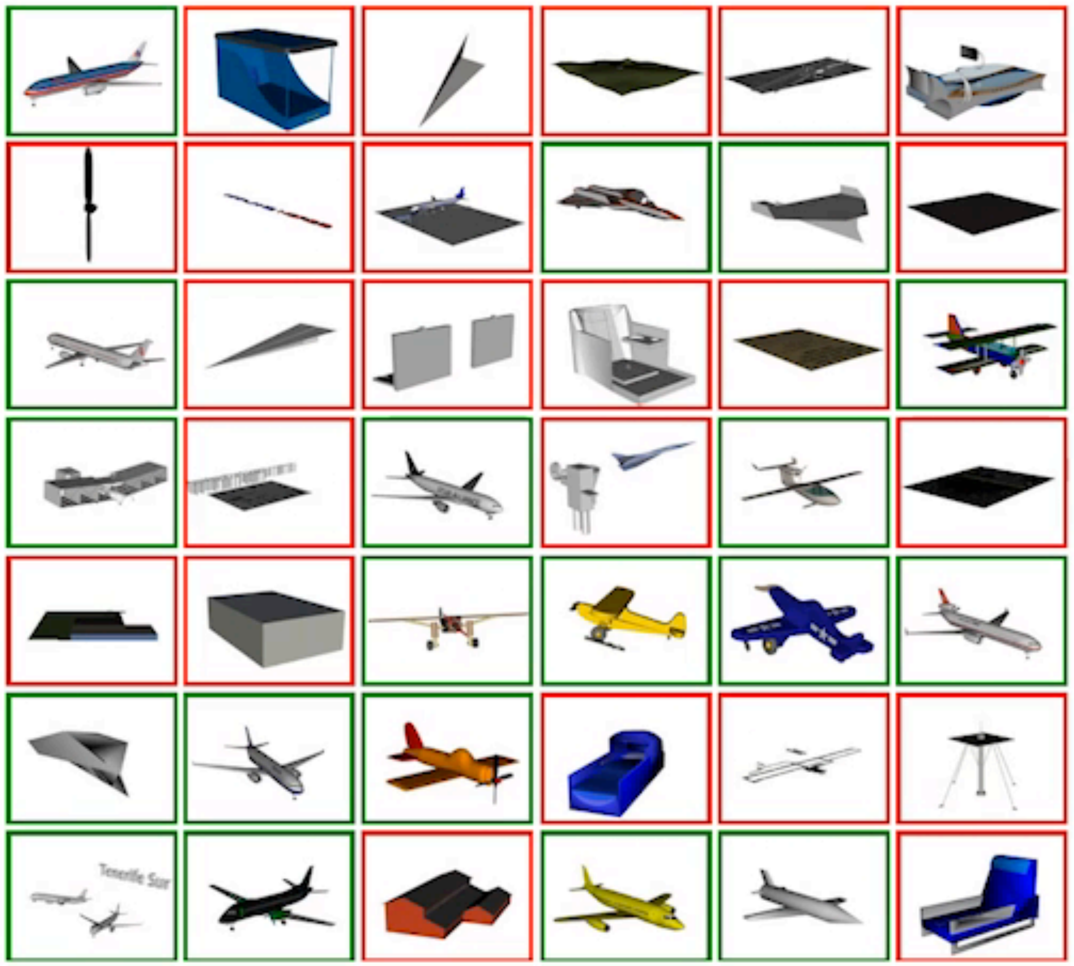
Top 2D results

CrossLink

Top 3D results

Filter → Align →

Shape ← Align ←



Key Ideas

1. Push image ordering to **reorder** 3D models

2. 3D model **coalignment**

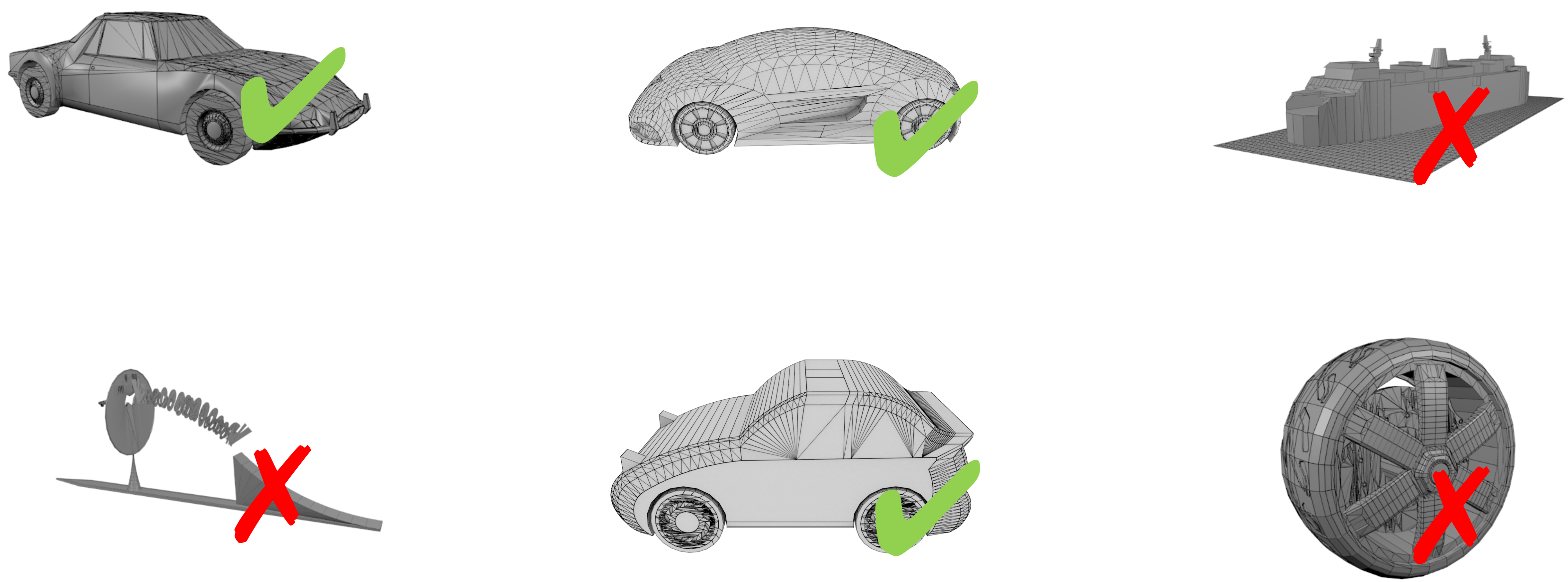
   Exploit loop structure and pose as a reshuffling problem

3. Push 3D shape renderings to **view-classify** images

   Map view variation to SVM changes

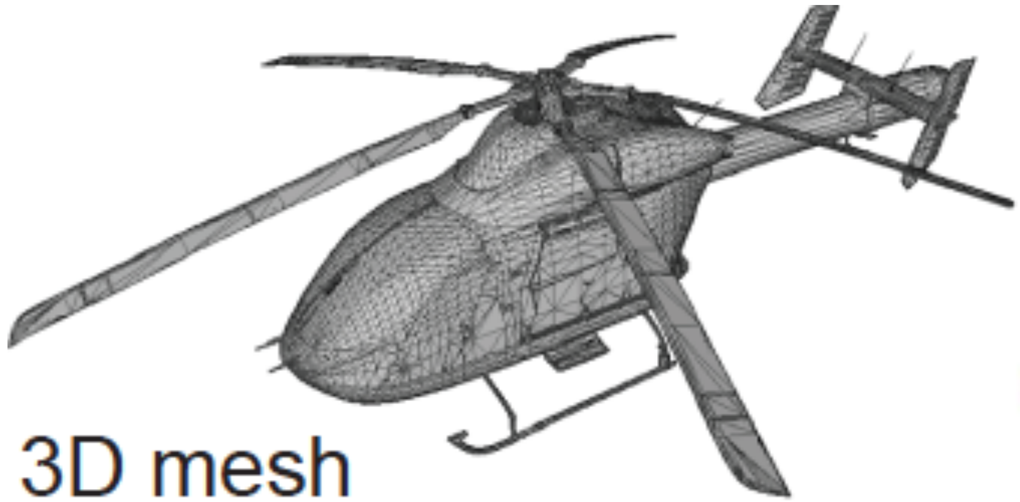
4. Push 3D **shape attributes** to images

Filter 3D using 2D - input



1. Filtering 3D models using 2D images

Render Images



Images Features (HOG or CNN)

rendering

$$\text{image} \rightarrow \mathbb{R}^d$$

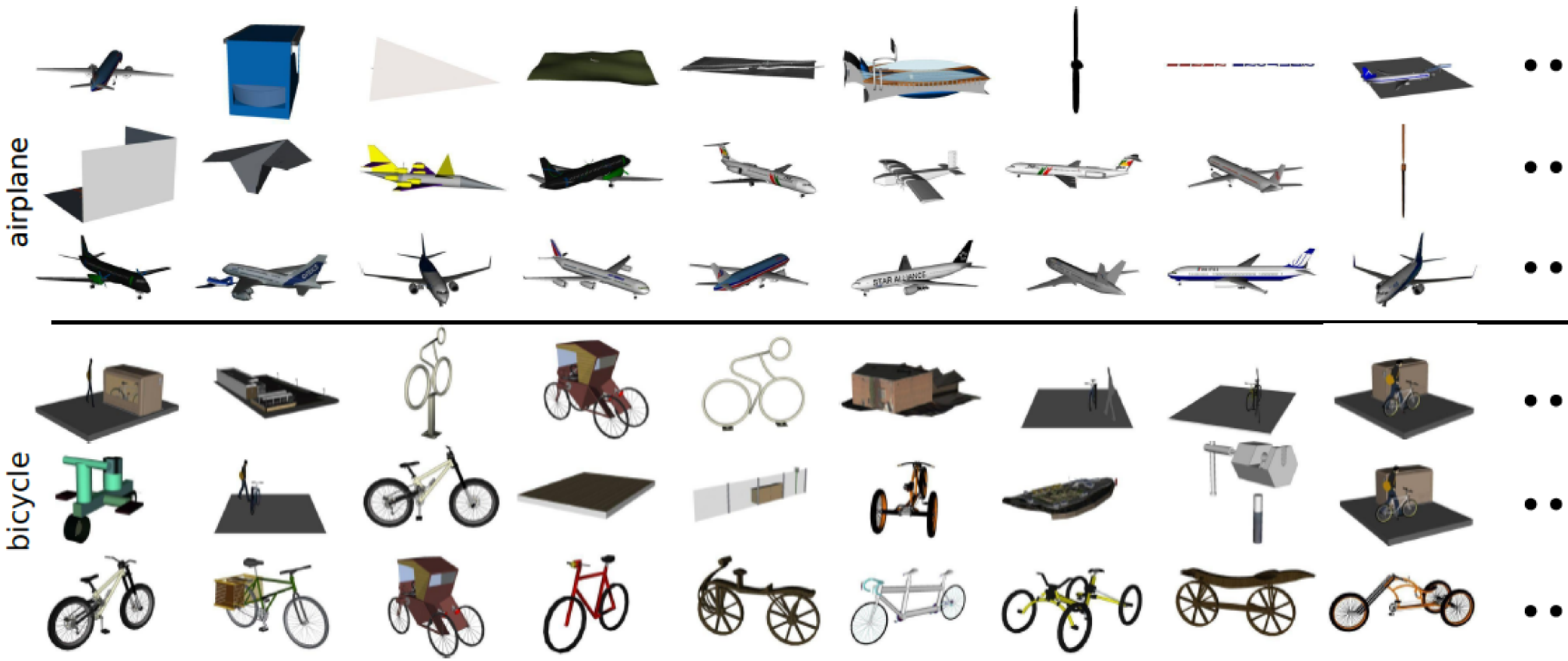
Positive/Negative Images

- **In class**: models + different views
- **Out of class**: other models + random views
- (background effects *not modelled*)

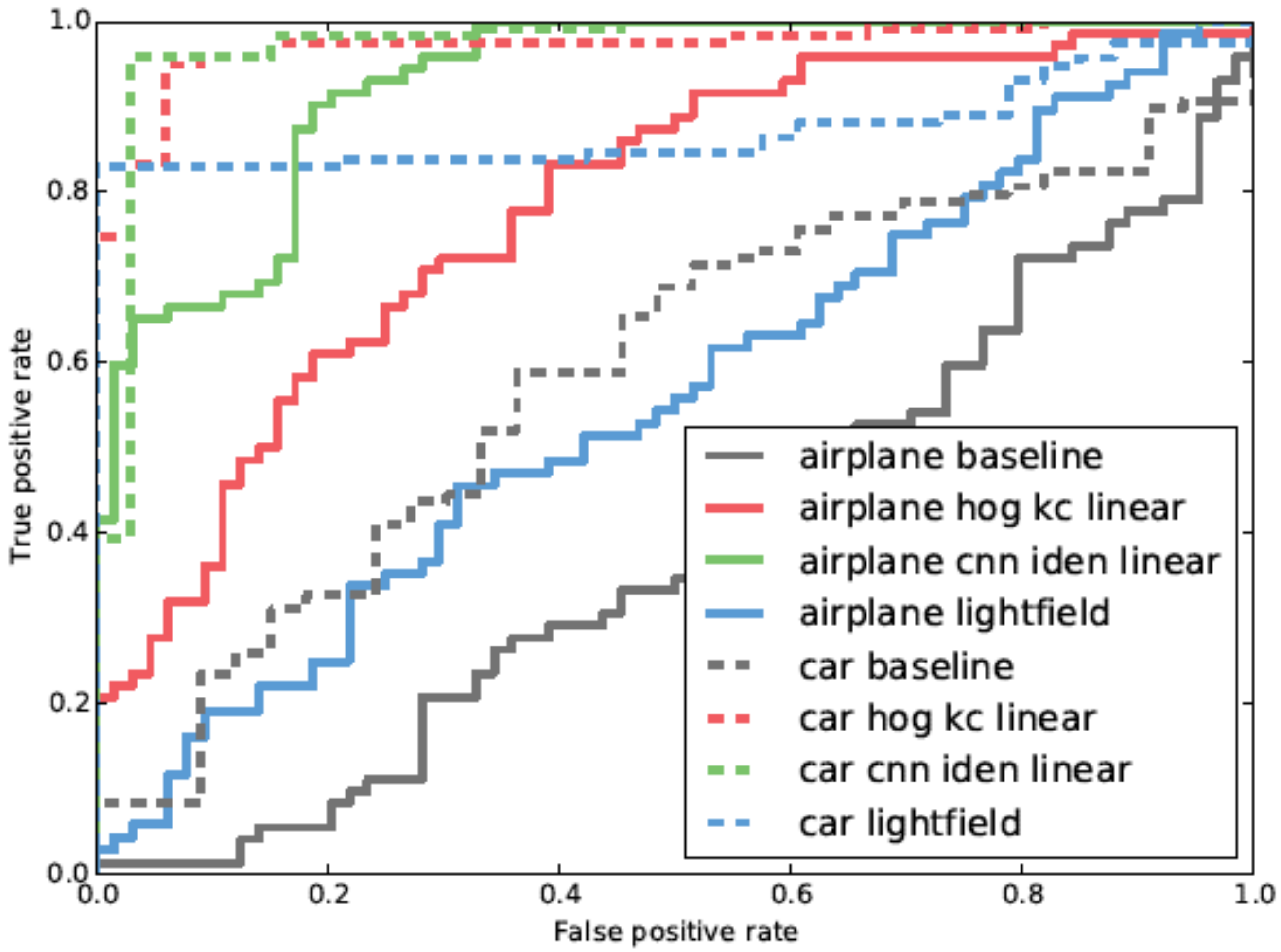
Train SVM classifier

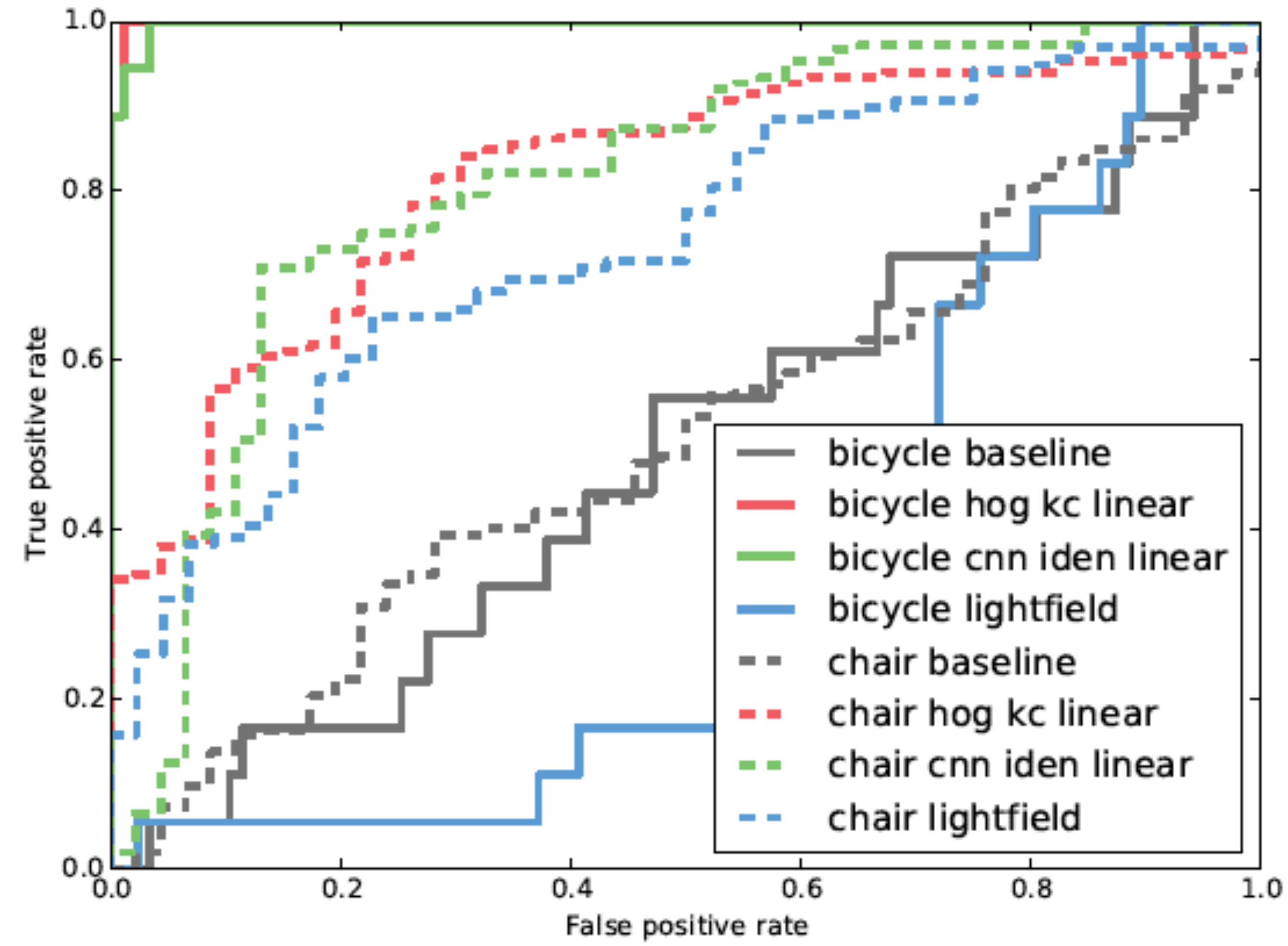
Resorting Results

Comparison with Shape Sig.



ROC Curves


1.0
0.8
0.6
0.4
0.2
0.0
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
airplane baseline
airplane hog kc linear
airplane cnn iden linear
airplane lightfield
car baseline
car hog kc linear
car cnn iden linear
car lightfield


1.0
0.8
0.6
0.4
0.2
0.0
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
bicycle baseline
bicycle hog kc linear
bicycle cnn iden linear
bicycle lightfield
chair baseline
chair hog kc linear
chair cnn iden linear
chair lightfield

Key Ideas

1. Push image ordering to **reorder** 3D models

2. 3D model **coalignment**

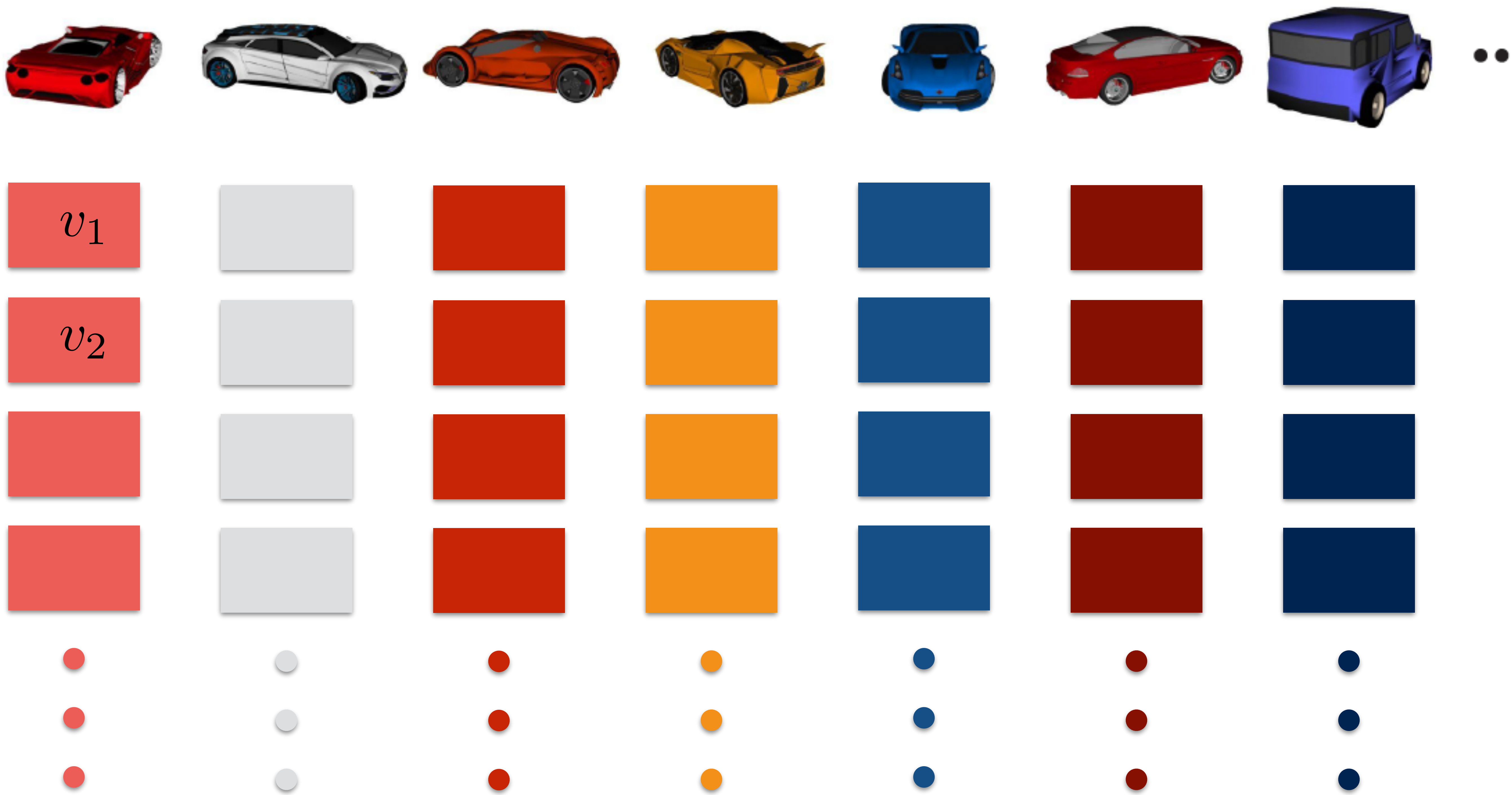
   Exploit loop structure and pose as a reshuffling problem

3. Push 3D shape renderings to **view-classify** images

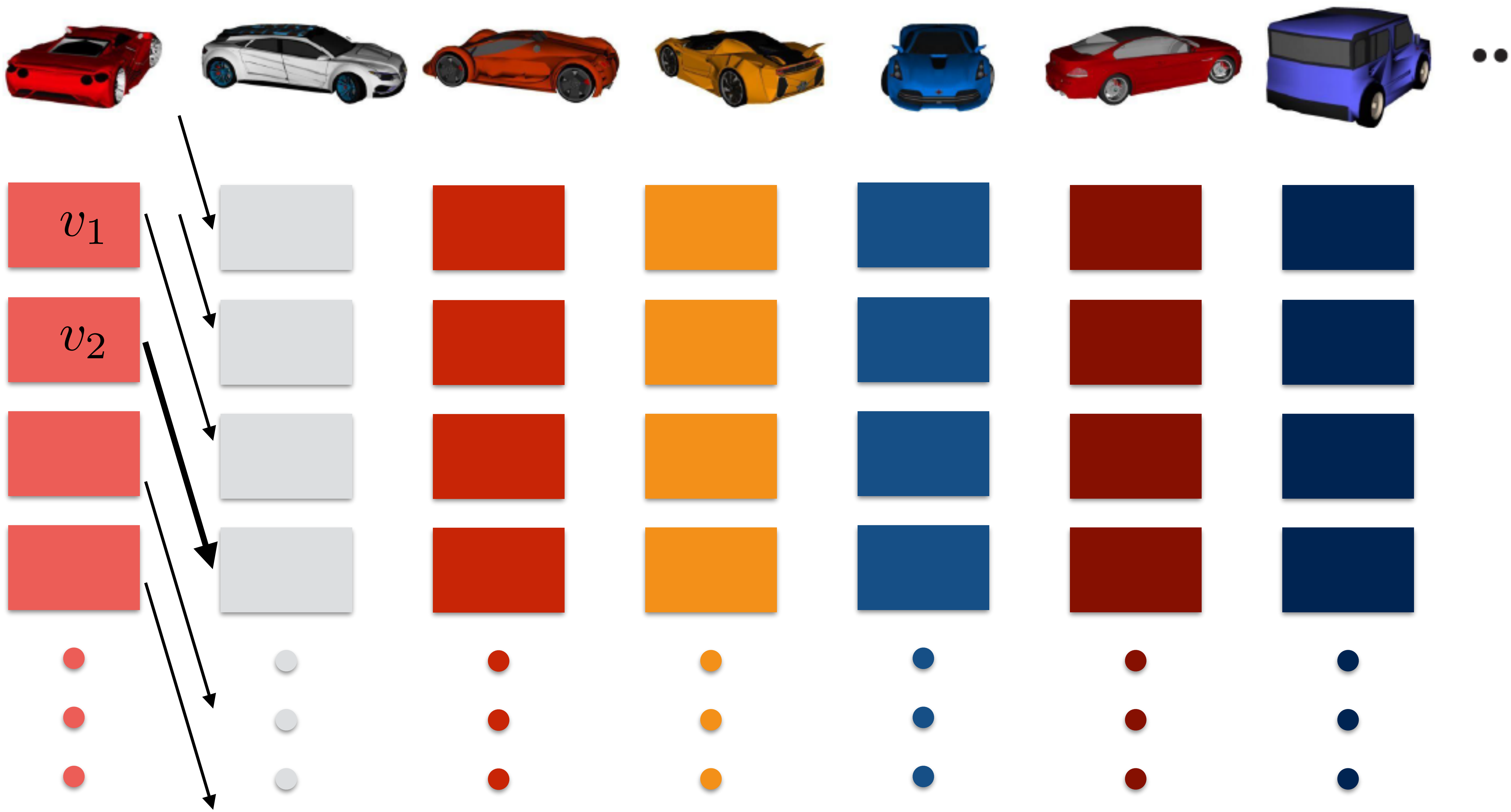
   Map view variation to SVM changes

4. Push 3D **shape attributes** to images

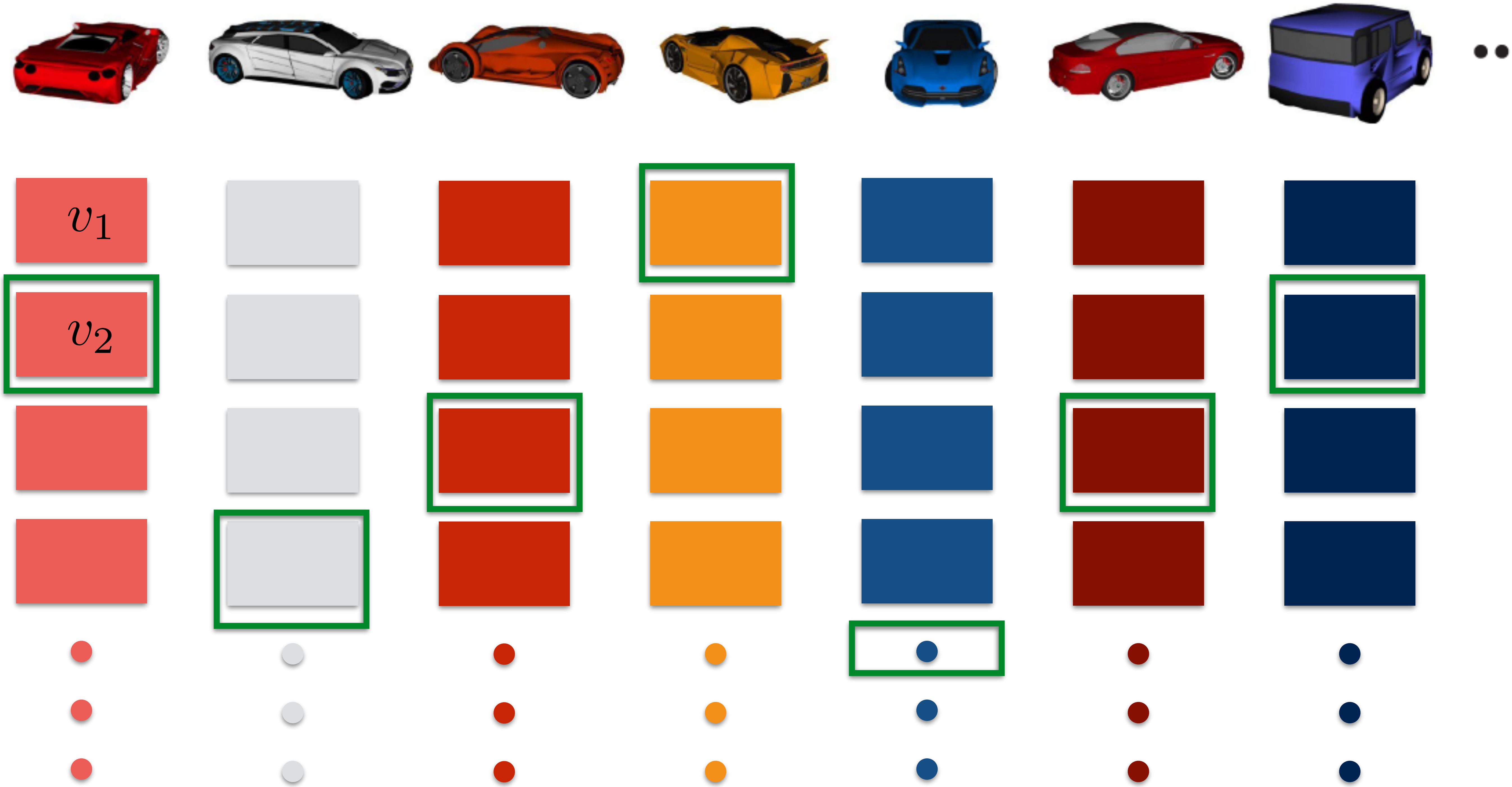
Model Coalignment



Model Coalignment



Model Coalignment



Coalignment Formulation

$$E(V) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=0}^{35} \| M_i^{V_i + k \bmod 36} - M_j^{V_j + k \bmod 36} \|_2^2$$

$$\arg\min_{\{V_i\}} E(V)$$

Key Ideas

1. Push image ordering to **reorder** 3D models

2. 3D model **coalignment**

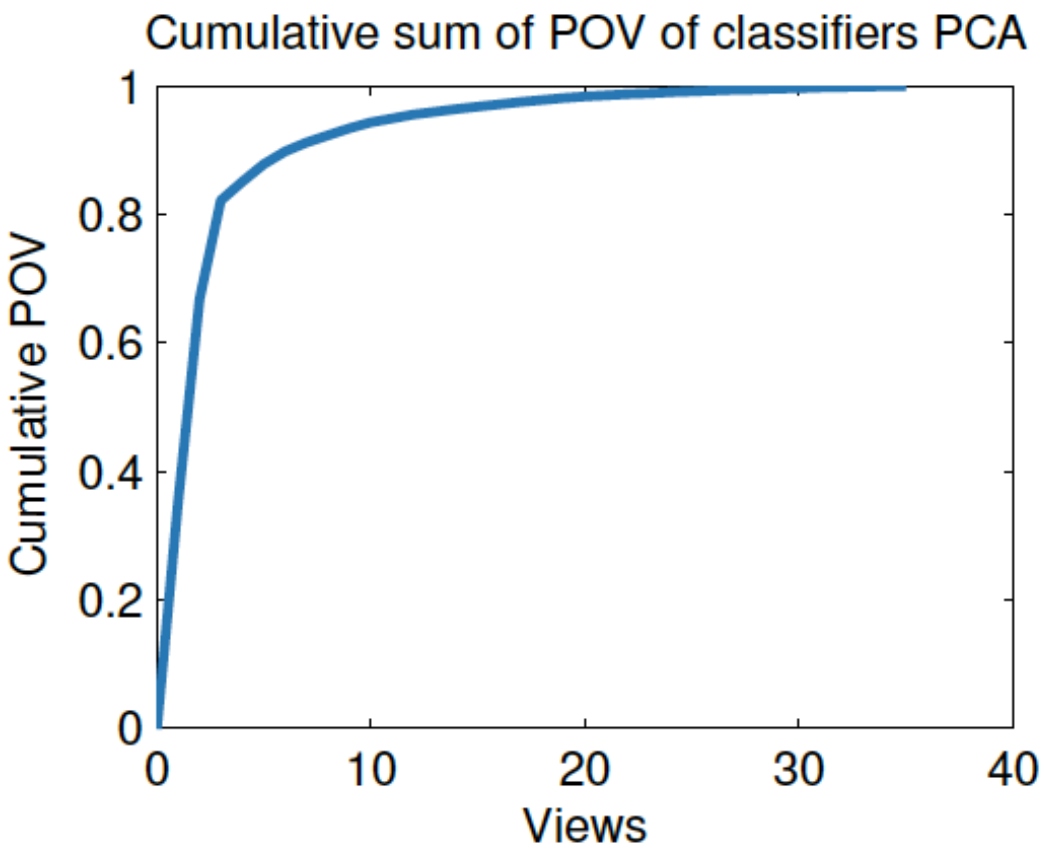
   Exploit loop structure and pose as a reshuffling problem

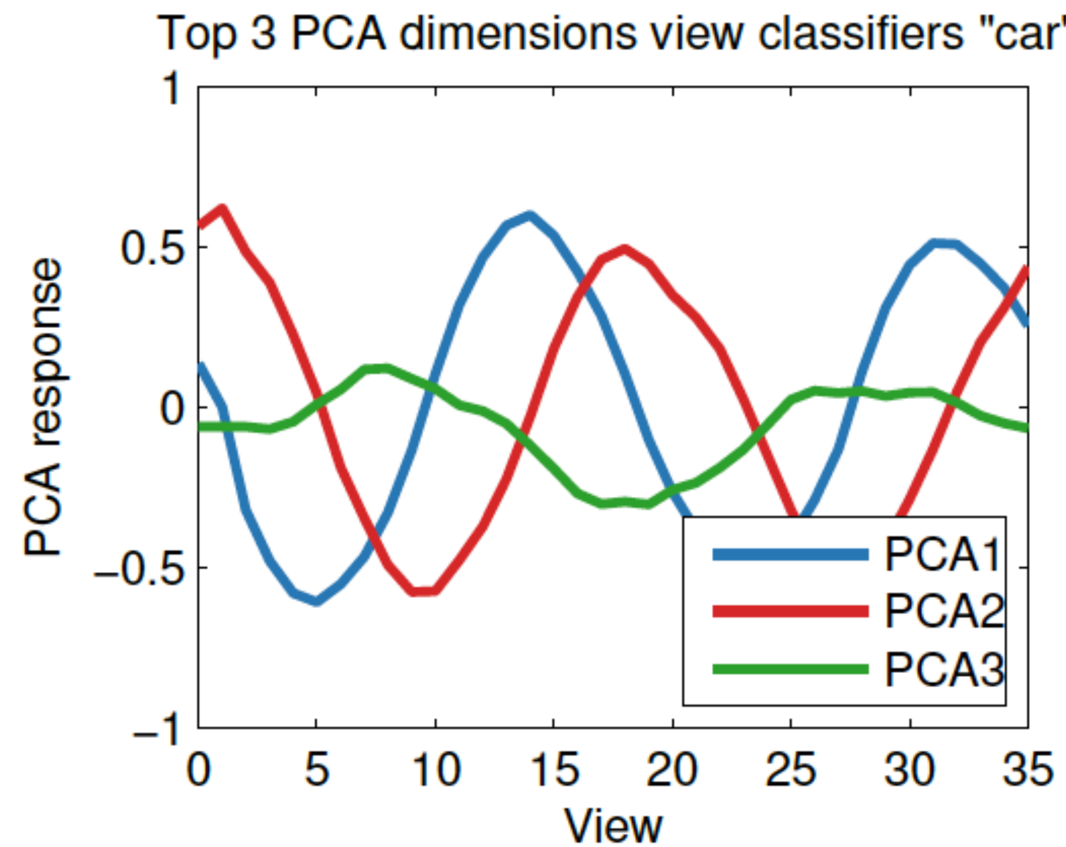
3. Push 3D shape renderings to **view-classify** images

   Map view variation to SVM changes

4. Push 3D **shape attributes** to images

Modeling Classifier Variation





rendered images with view variations

$\downarrow$

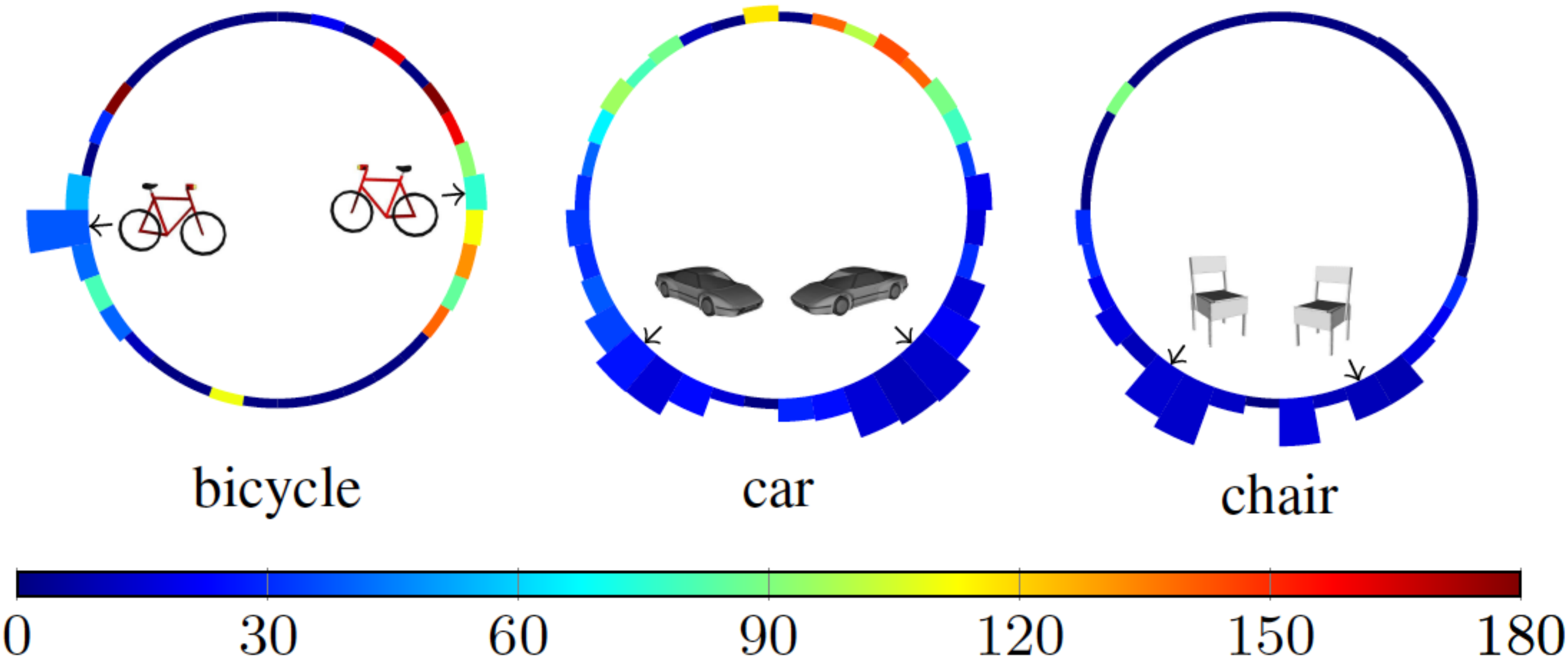
$\mathbb{R}^d$ under view variations

$\downarrow$

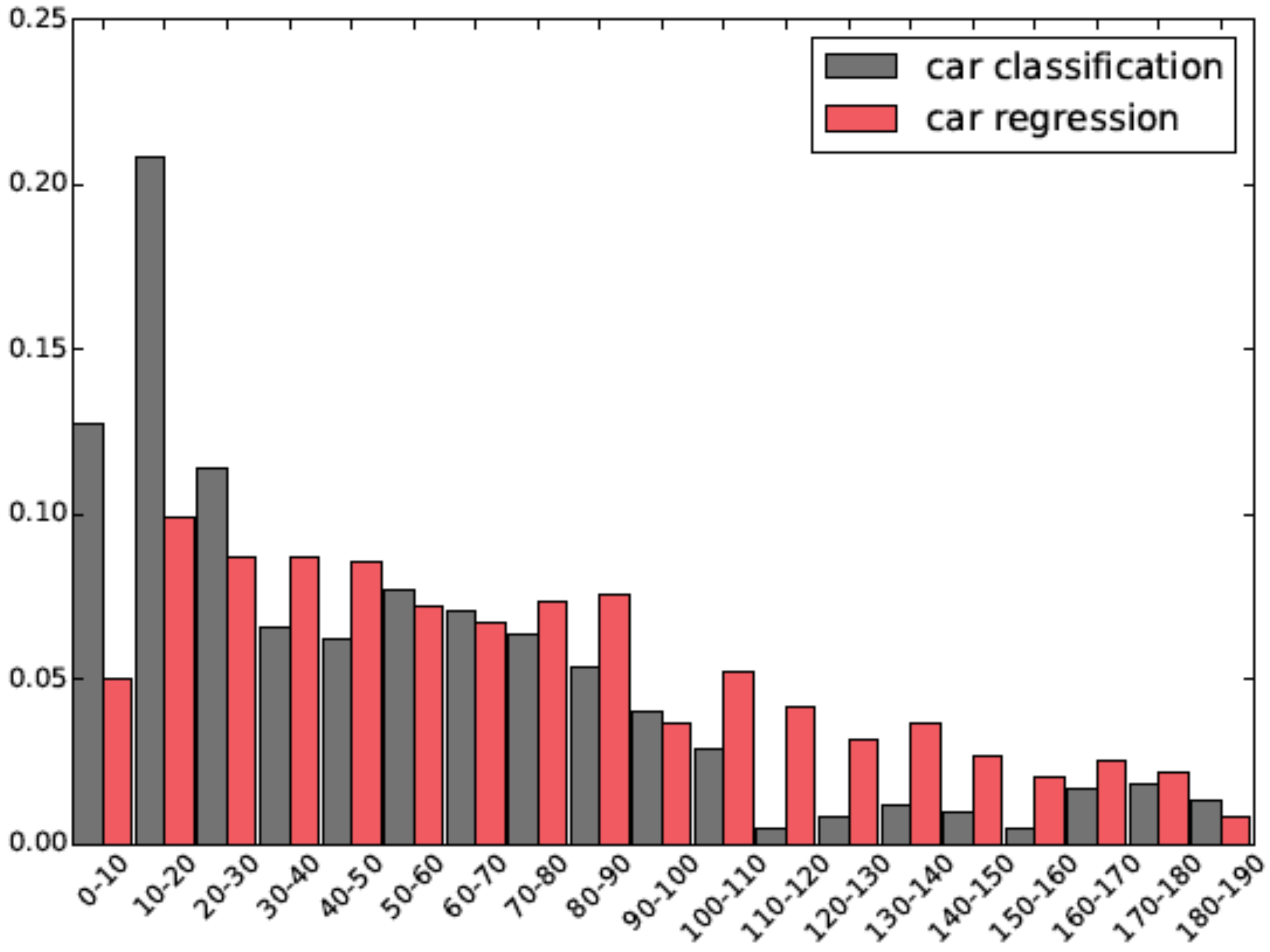
SVM separator parameters with view variations

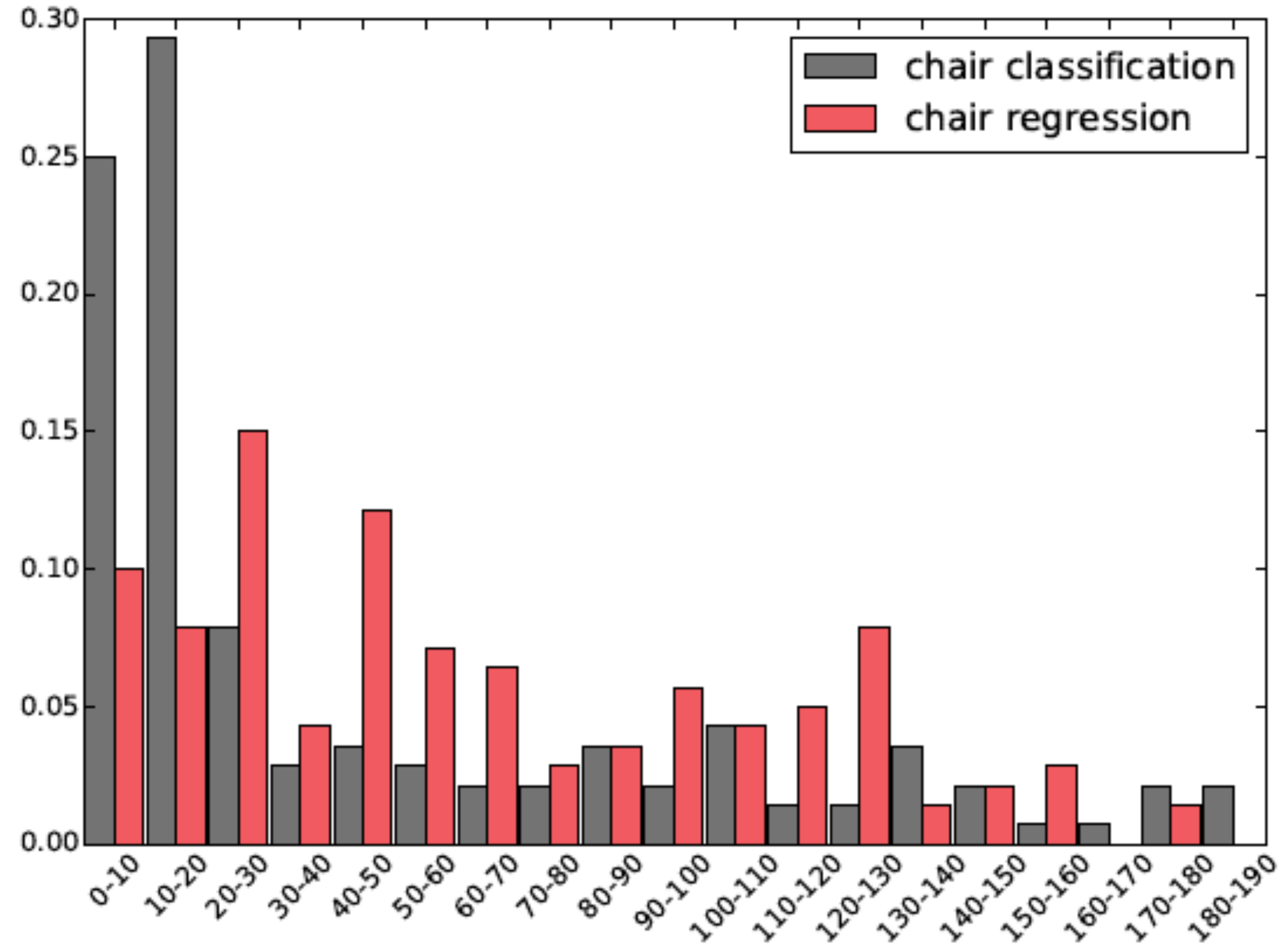
Camera Pose Estimation

View Classification



Comparison with Regression





Key Ideas

1. Push image ordering to **reorder** 3D models

2. 3D model **coalignment**

   Exploit loop structure and pose as a reshuffling problem

3. Push 3D shape renderings to **view-classify** images

   Map view variation to SVM changes

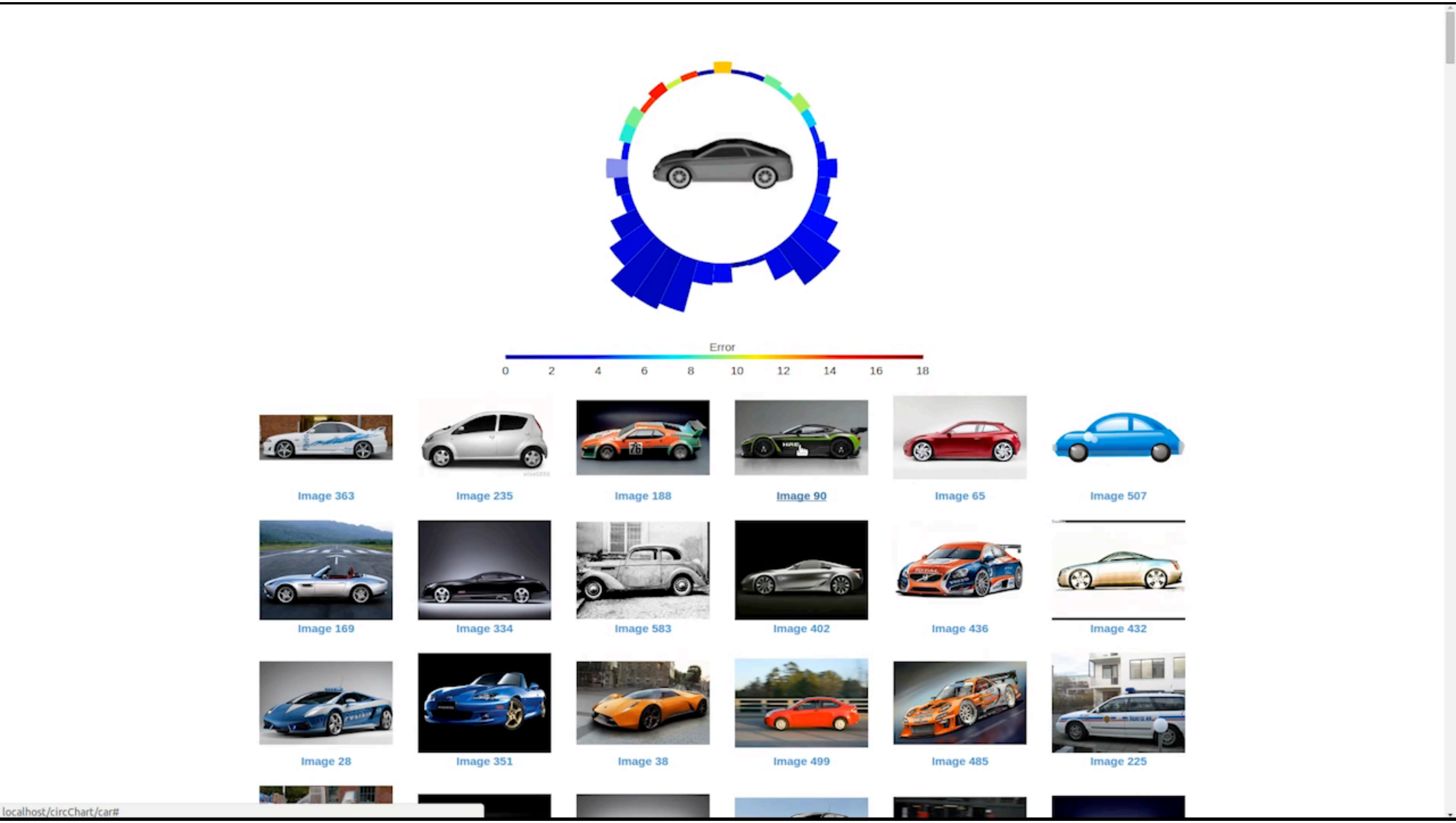
4. Push 3D **shape attributes** to images

Regressor for Shape Attributes

- Estimate shape attribute from 3D models
  - ratio of width/length

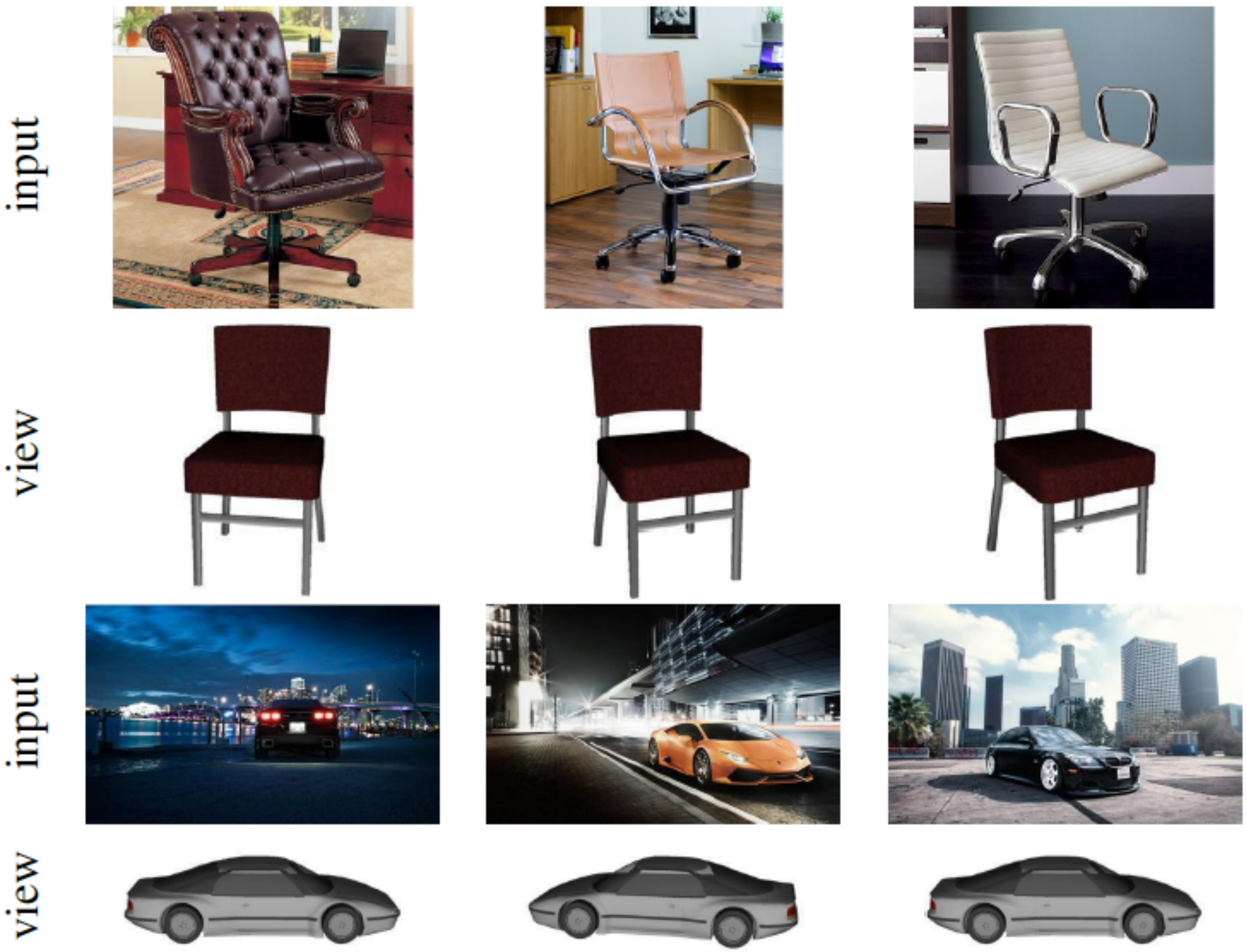
- Link attributes with view features using regressor

CrossLink



Error
0 2 4 6 8 10 12 14 16 18
Image 363
Image 235
Image 188
Image 90
Image 65
Image 507
Image 169
Image 334
Image 583
Image 402
Image 436
Image 432
Image 28
Image 351
Image 38
Image 499
Image 485
Image 225
localhost/circChart/car#

Limitations



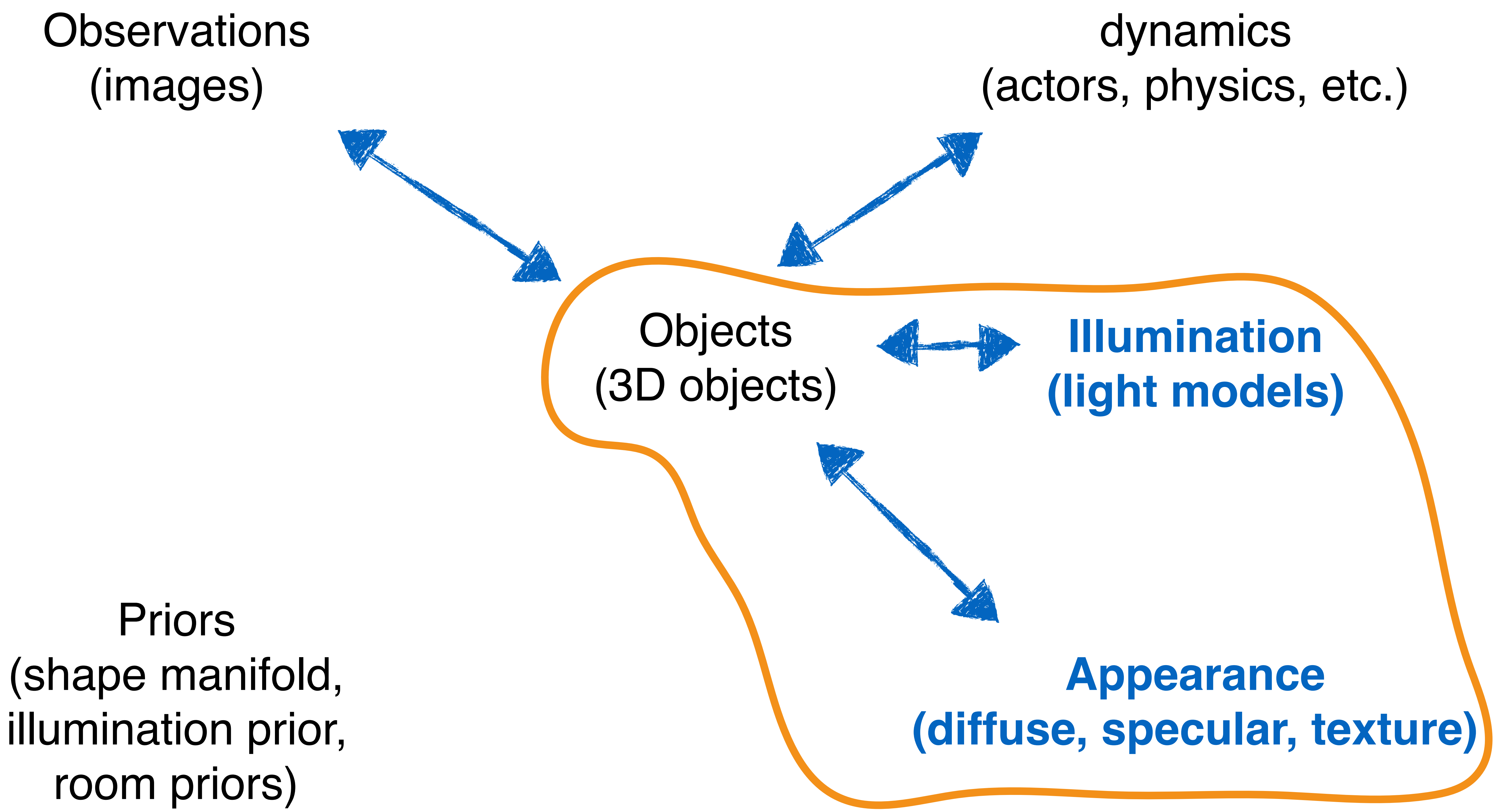
background clutter

limited shape attributes

no explicit links

only 1D view

Overview


Observations
(images)
dynamics
(actors, physics, etc.)
Objects
(3D objects)
Illumination
(light models)
Appearance
(diffuse, specular, texture)
Priors
(shape manifold,
illumination prior,
room priors)

Model + Texture

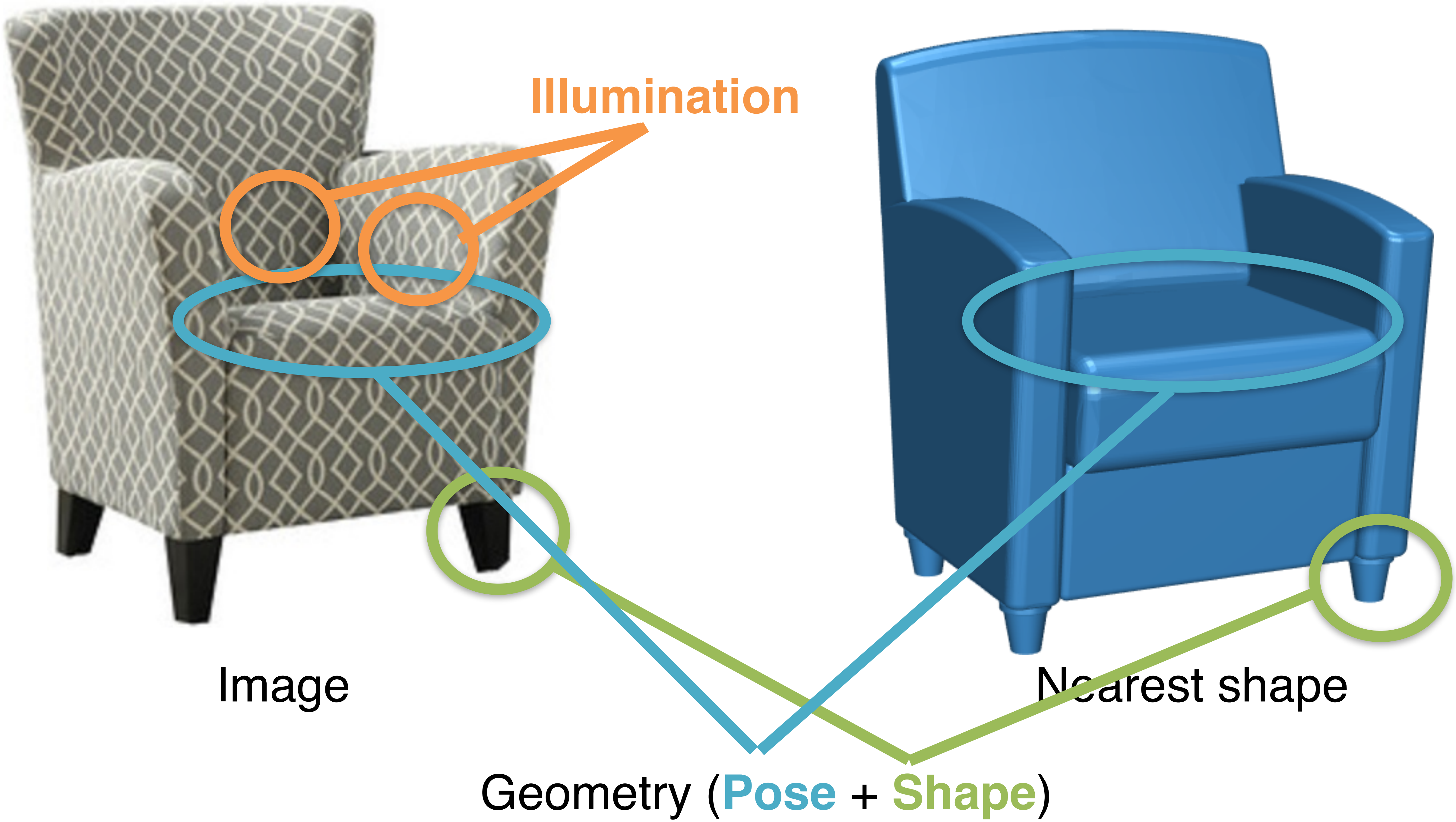
Images (Textured Objects)

Motivation

Real world image + Similar shape

Image to Textured 3D Models

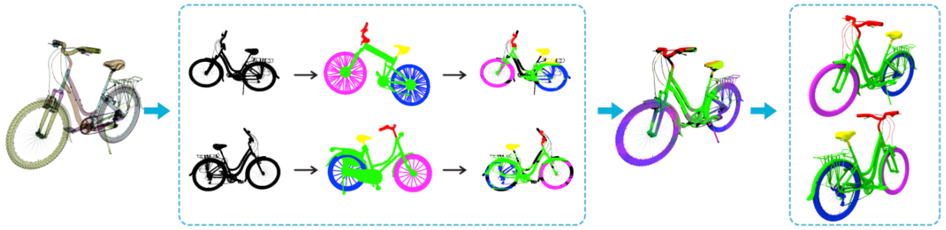
Challenges



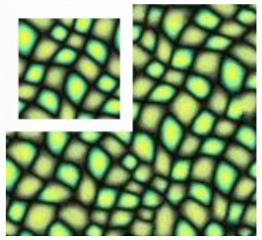
Related Work

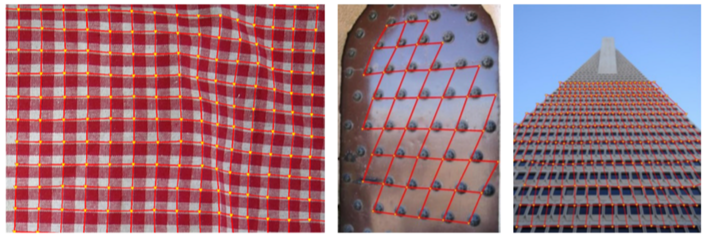
- **Intrinsic image decomposition**
  - [Barron and Malik 2015]

- **Joint image-shape analysis**
  - shape -> image [Hueting et al. 2015; Su et al. 2015; Lim et al. 2014]
  - image -> shape [Wang et al. 2013; Kholgade et al. 2014].

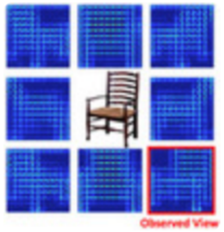


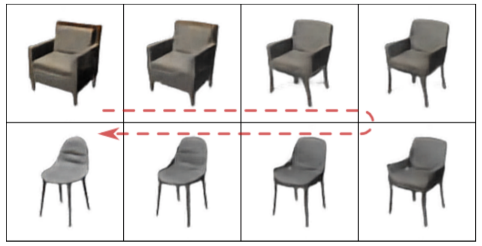
- **Texture**
  - representation [Wei et al. 2009]
  - structure detection [Liu et al. 2015]





- **Novel view prediction**
  - probabilisitic [Su et al. 2014b]
  - CNN [Dosovitskiy et al. 2015]





Assumptions

rectified texture (patch)
+
part-level UV layout

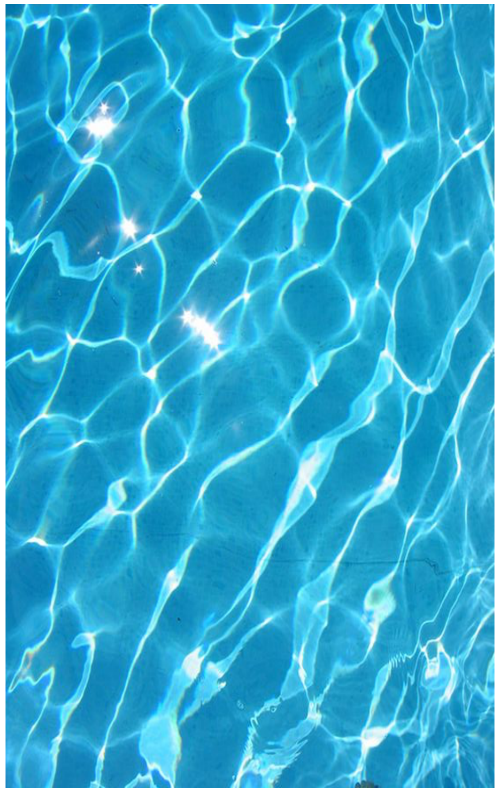
repetitive pattern structure

Regular

Irregular but repetitive

Non-supported Cases

repetitive pattern structure

Negative cases

Key Idea

rectified texture (patch)
+
part-level UV layout

rough geometric priors help!

Target

- **Appearance transfer**
  — understand texture pattern, texture orientation, illumination

- **Visual plausible**
  — no need to be exactly the same but highly visual realistic

- **Fully automatic**
  — feasible for large dataset

Pipeline

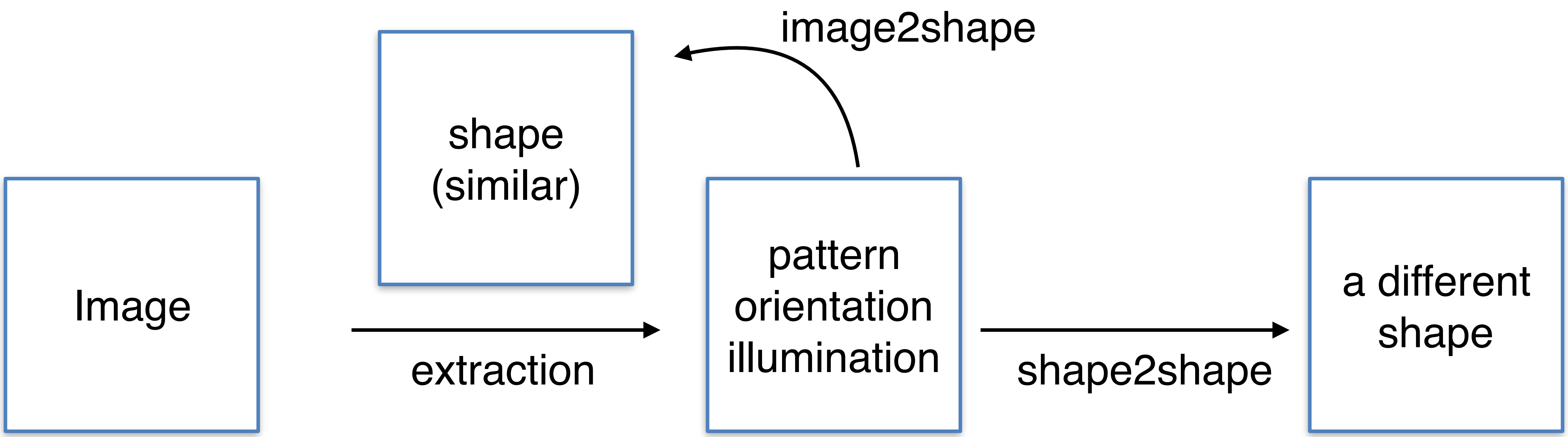


Image
shape (similar)
image2shape
extraction
pattern orientation illumination
shape2shape
a different shape

Pipeline

image2shape

shape (similar)

Image

extraction

pattern orientation illumination

shape2shape

a different shape

Pose Estimation

Pipeline

image2shape

shape (similar)

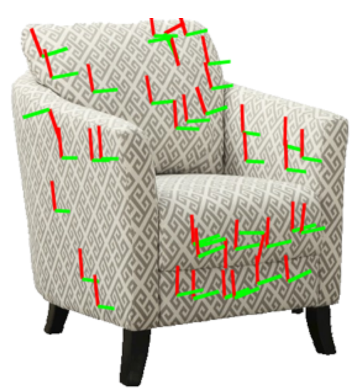
Image

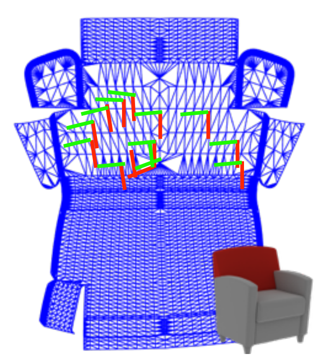
extraction

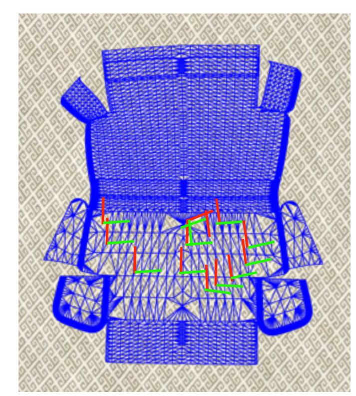
pattern orientation illumination

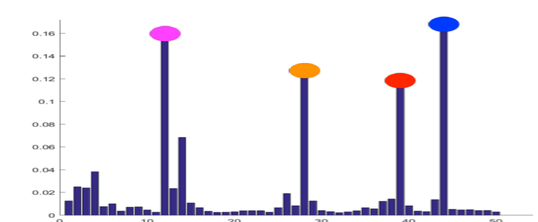
shape2shape

a different shape









Key Idea

depth map

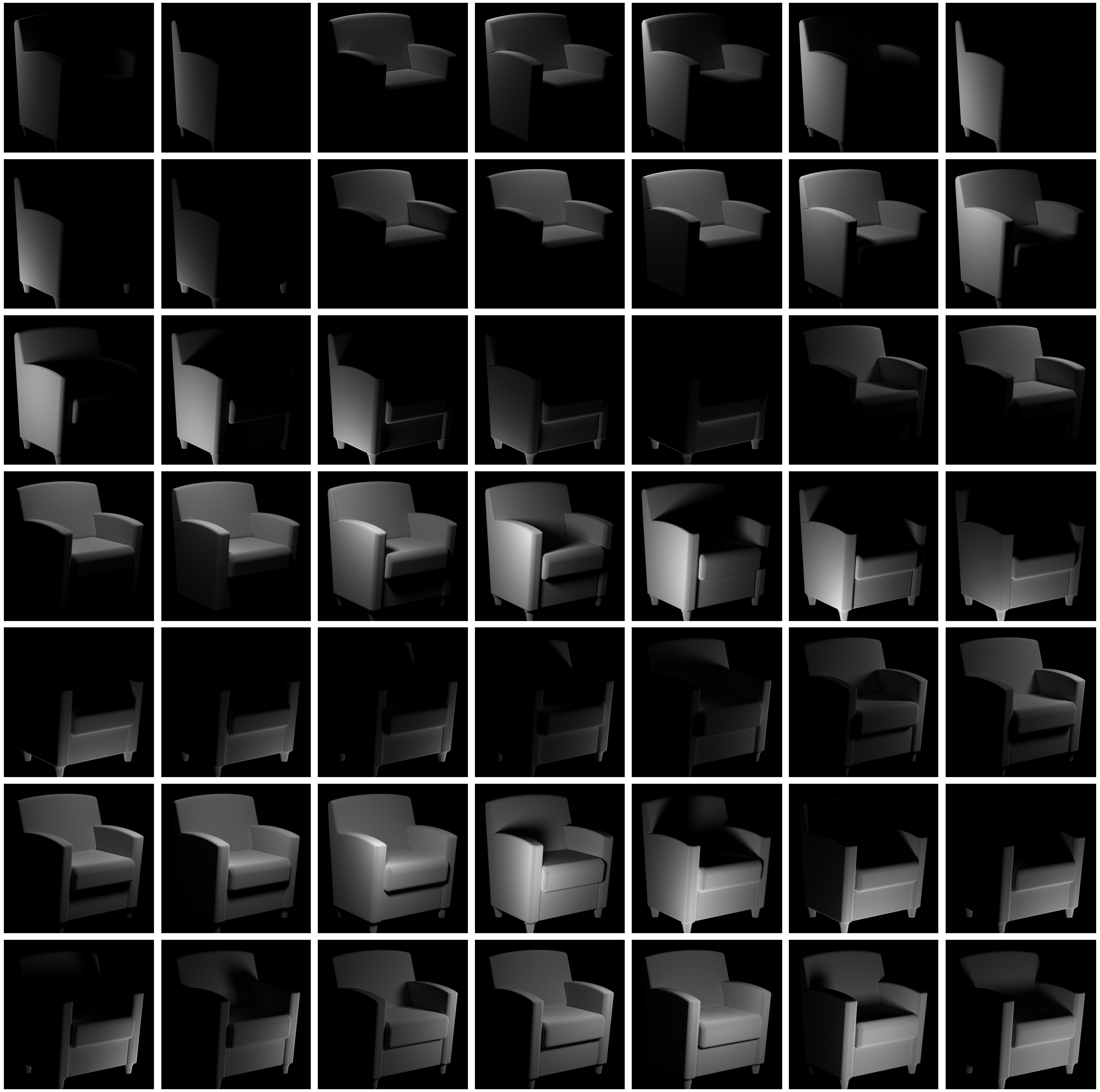
normal map

approximate geometric prior helps!!

Patch Extraction

Illumination Estimation

shading samples



Estimation

$$I_s := \arg\min_{I_s^*} \sum_{(i,j)\in\Pi} \|I_{r*}(l_i) - I_{r*}(l_j)\|$$

(a)

Texture Orientation

(a)

Pipeline

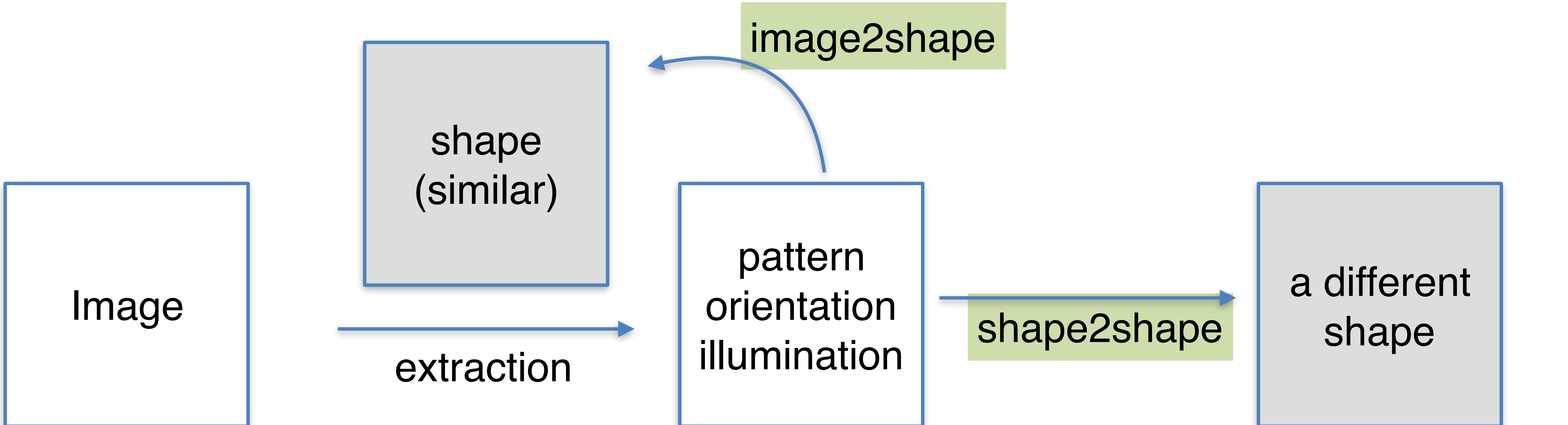
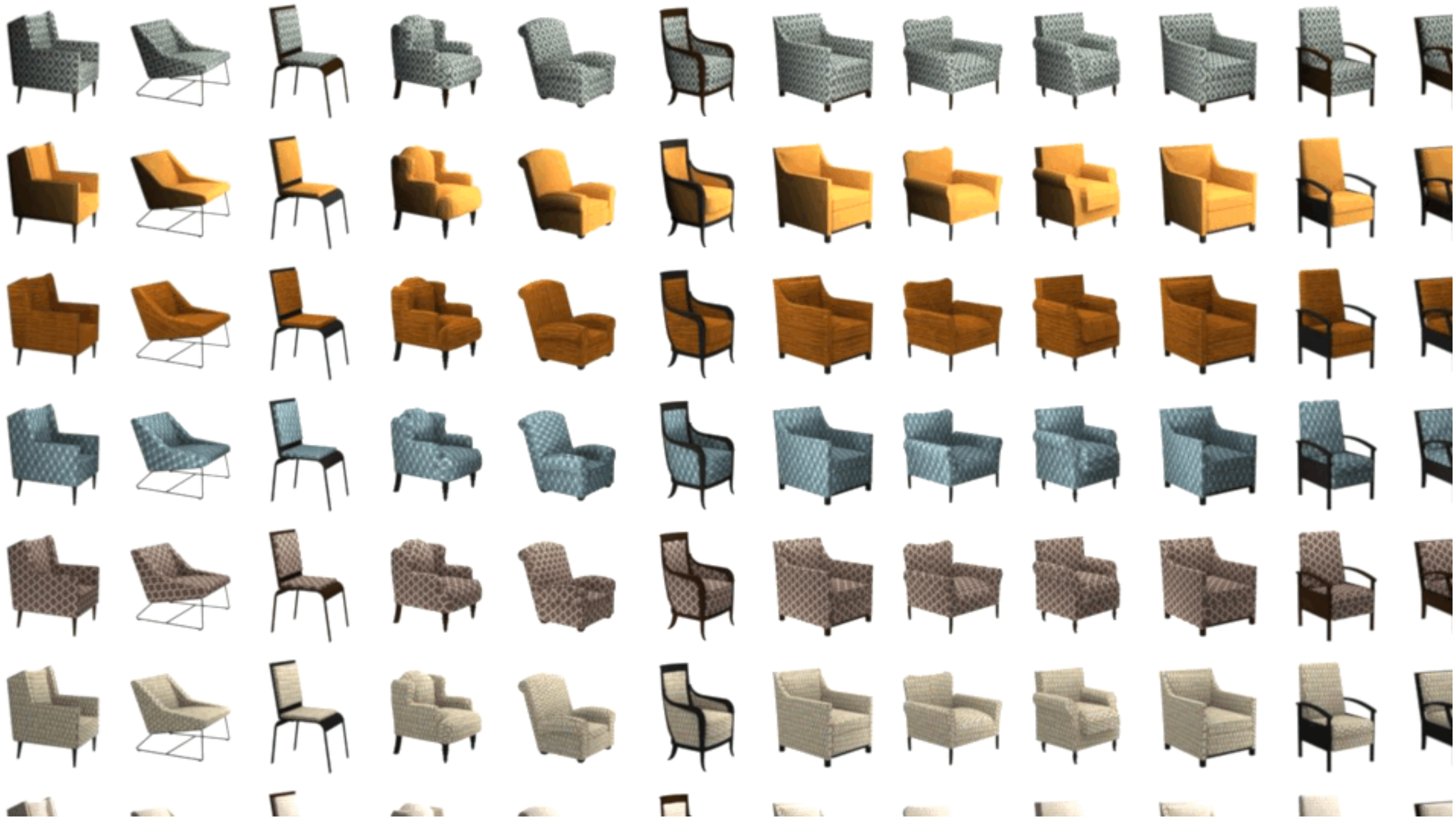


Image
extraction
shape (similar)
image2shape
pattern
orientation
illumination
shape2shape
a different shape

Results

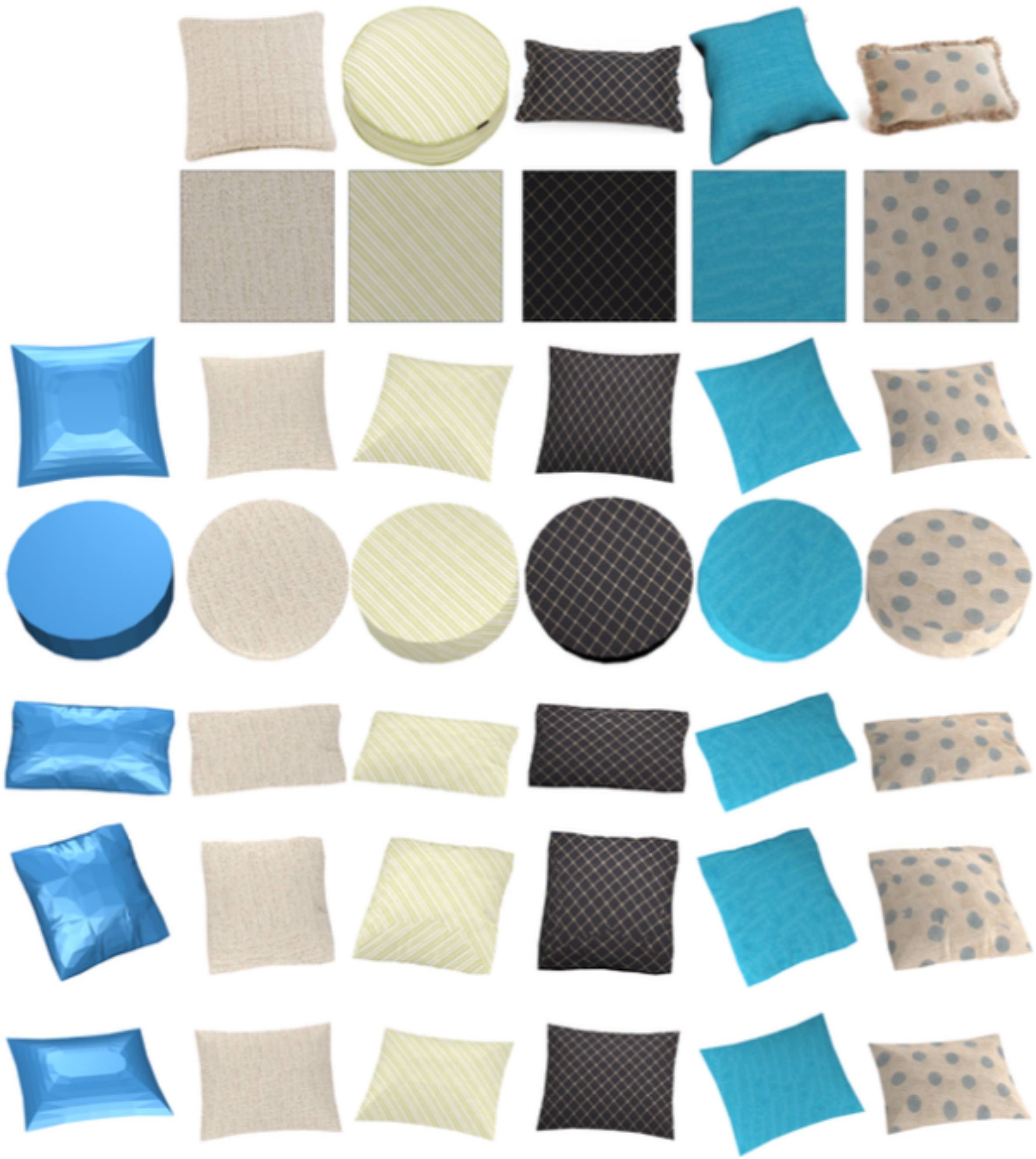
Chair dataset
(~2k textured objects)

Results



Chair dataset
(~2k textured objects)

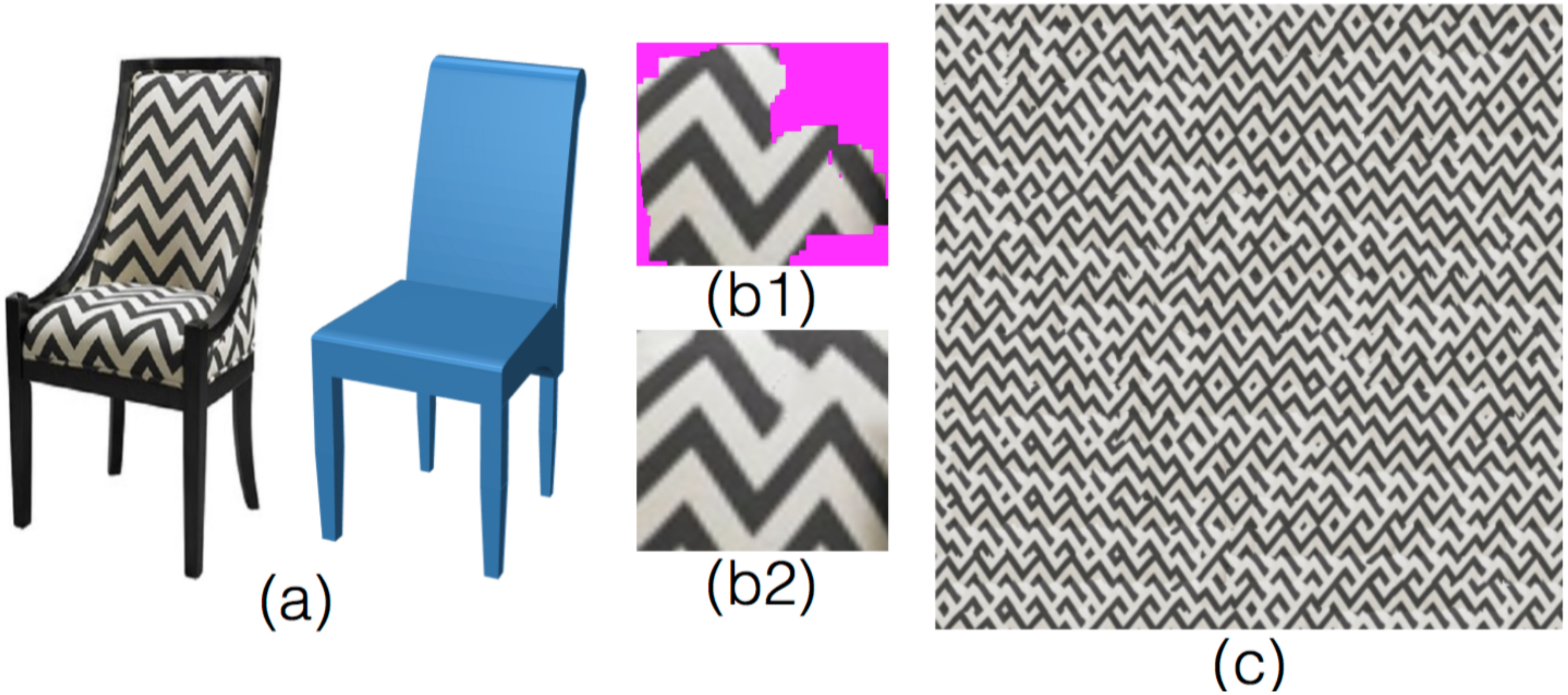
Results



Pillow dataset

Table dataset

Illumination Correction

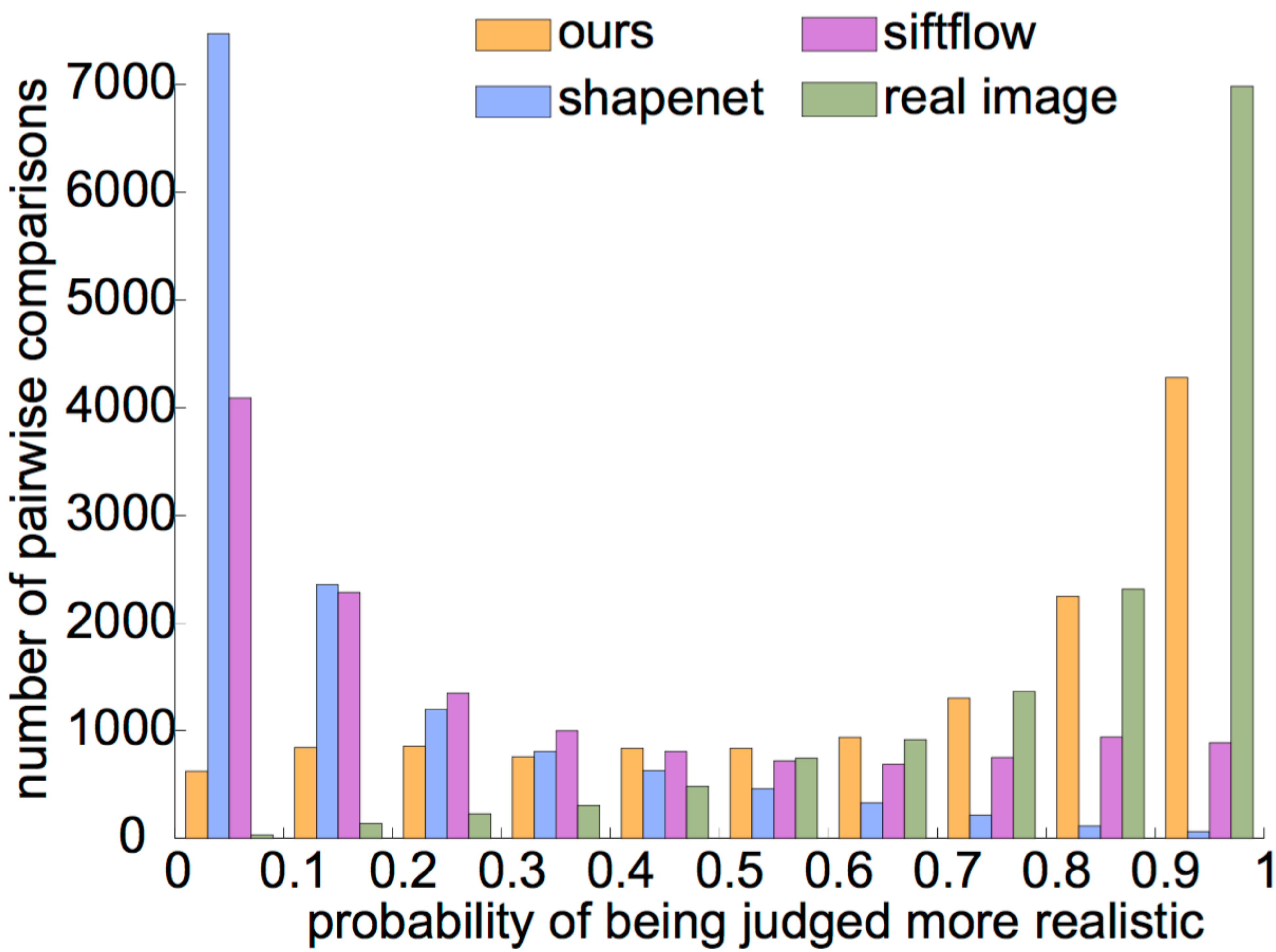


Dependence on Paths

(a)

Comparison

User Evaluation

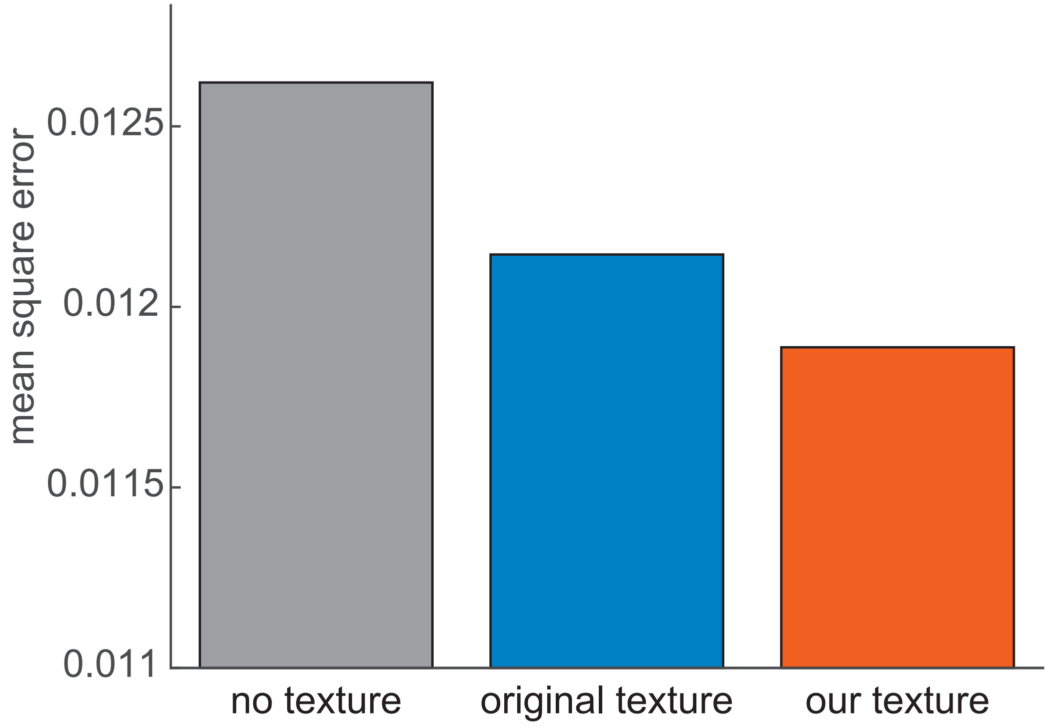


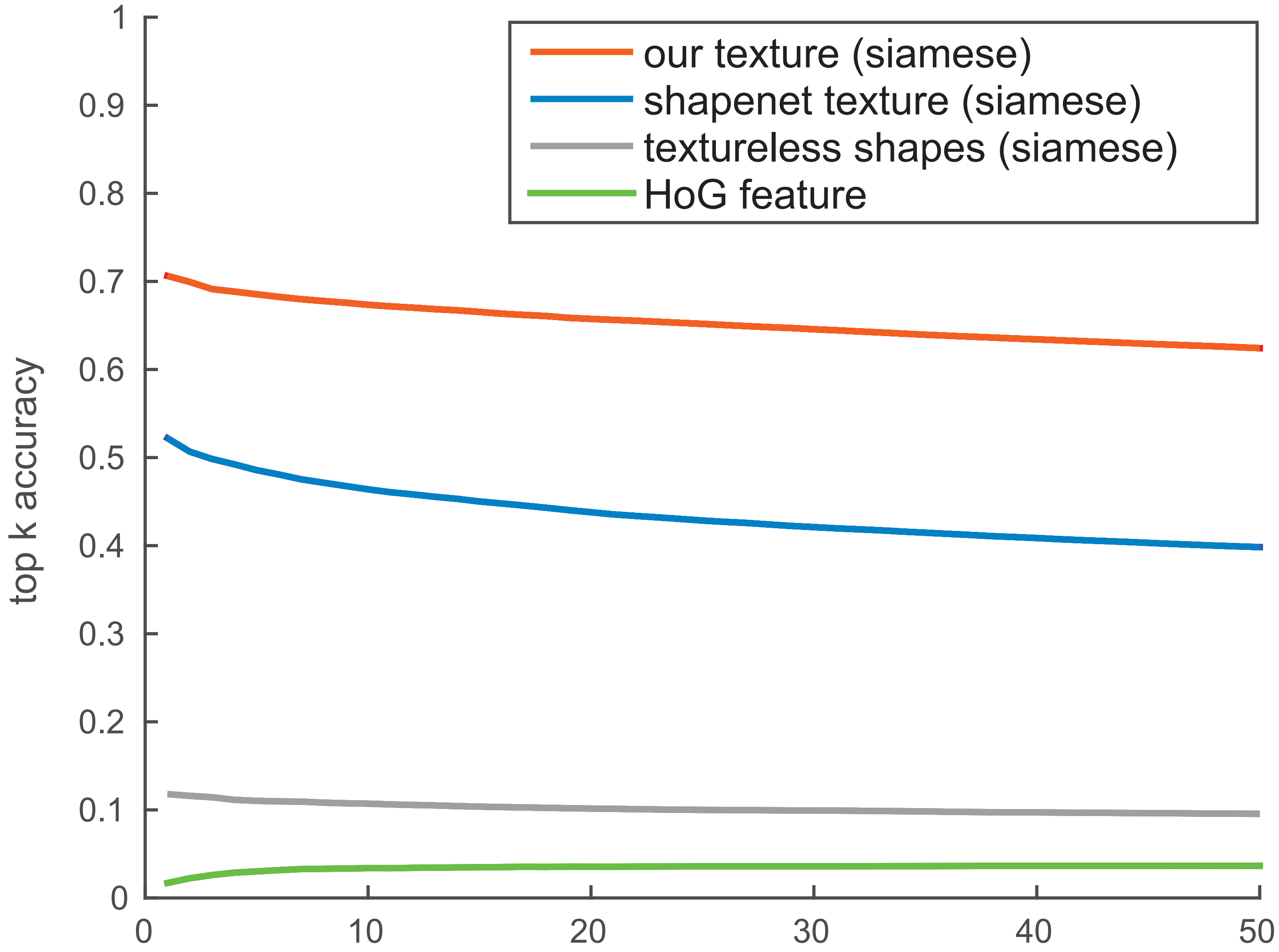
Application: Image Editing

Image

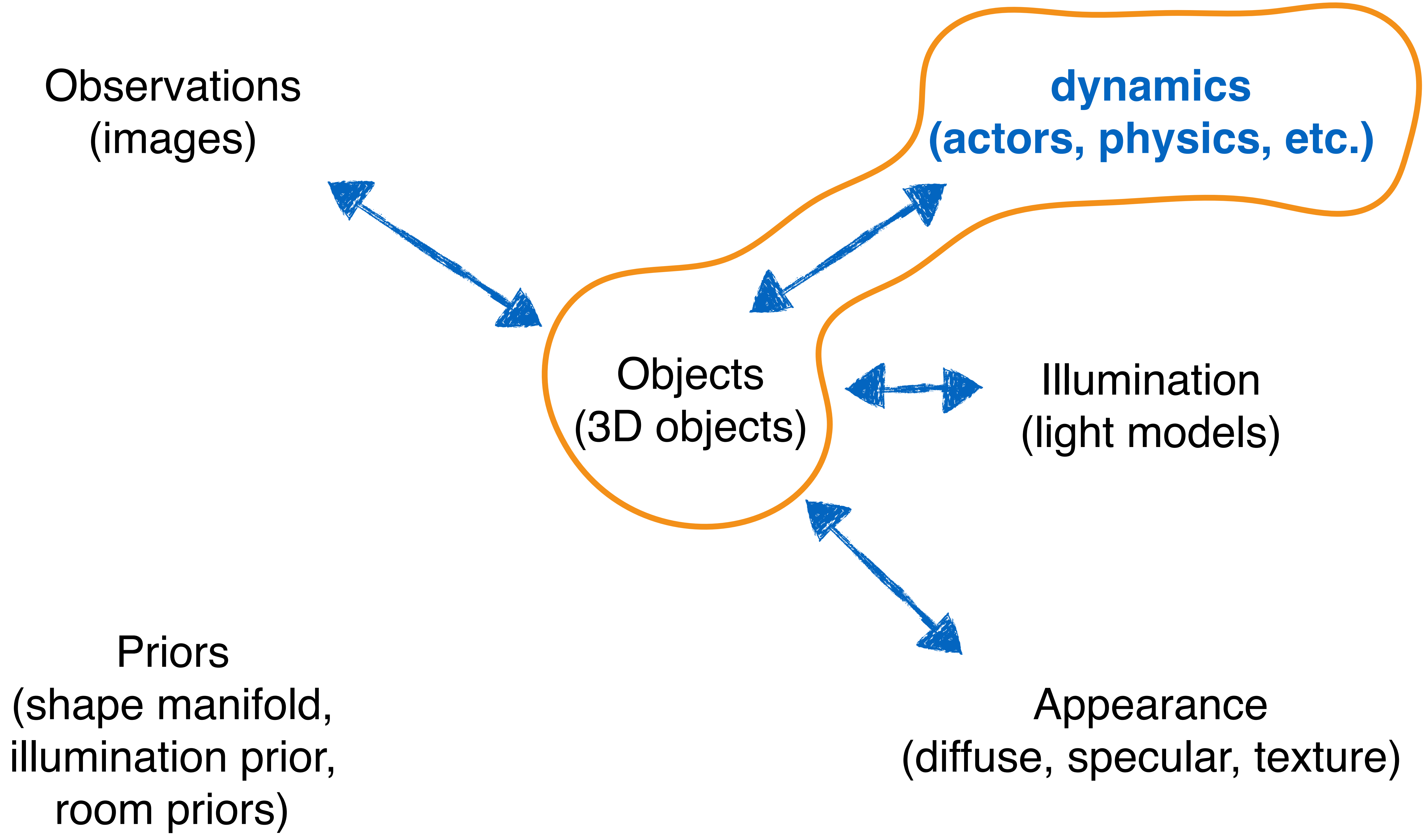
Shape

Better Training Data





Overview



Observations
(images)
dynamics
(actors, physics, etc.)
Objects
(3D objects)
Illumination
(light models)
Priors
(shape manifold,
illumination prior,
room priors)
Appearance
(diffuse, specular, texture)

[redacted]

[redacted]


Optimal Surface in Product Shape
Geometrically Consistent Elastic Matching of 3D Shapes: A Linear Programming Solution [Cremers et al. 11]
Minimal closed surface in a higher dimensional space.
Matching = finding an optimal surface in the product of the two shape surfaces.
minimizes physical deformation energy needed for deforming one shape into the other.

Input video:

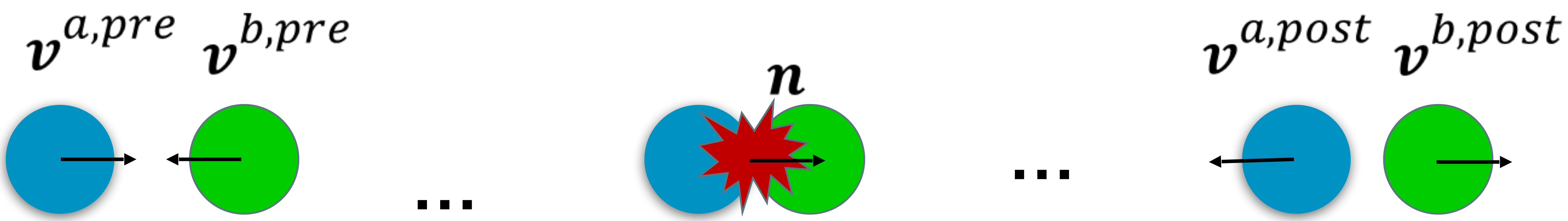
Target

- Reconstruction of **collision parameters** in 3D
  - Position, velocity
  - Orientation, angular velocity
  - Relative mass
  - Coefficient of restitution

At all times

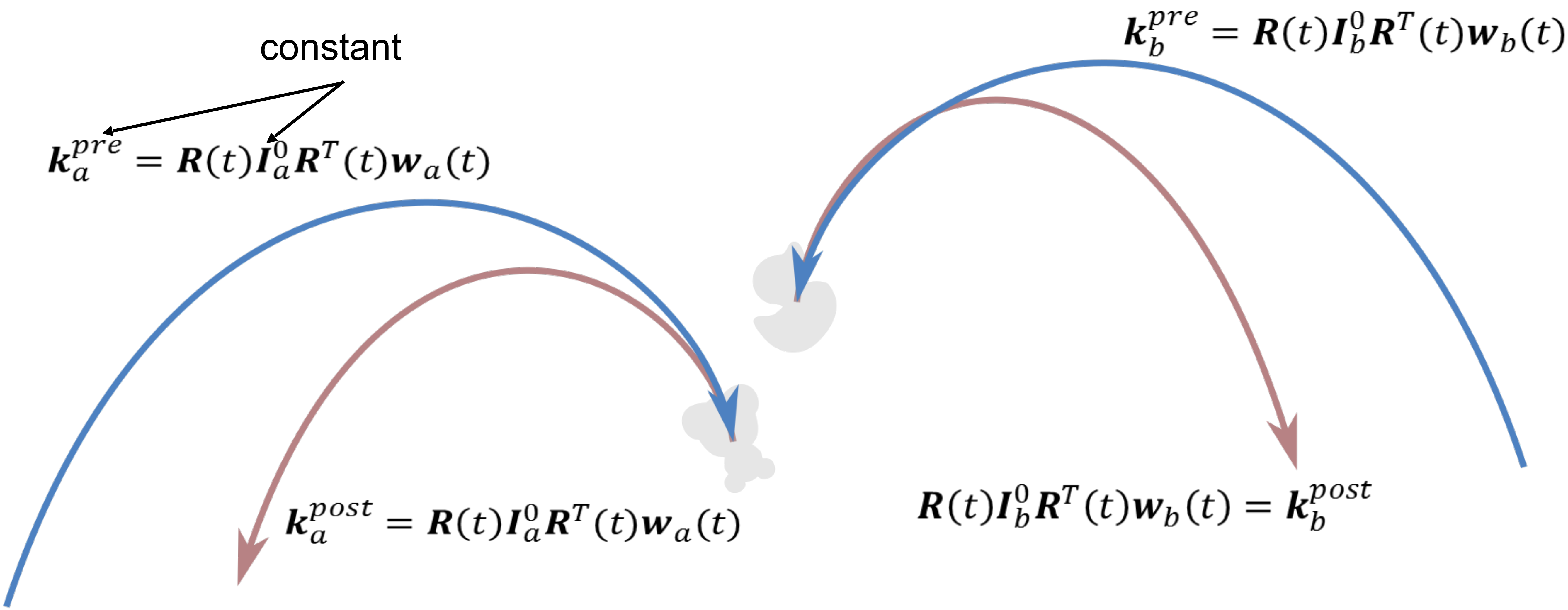
- Can be retrieved **without observing** the exact moment of collision ~~(collision detection)~~

Coefficient of Restitution



$$c = \frac{(\boldsymbol{v}^{b,post} - \boldsymbol{v}^{a,post}) \cdot \boldsymbol{n}}{(\boldsymbol{v}^{a,pre} - \boldsymbol{v}^{b,pre}) \cdot \boldsymbol{n}} = -\frac{\boldsymbol{v}_{rel}^{post} \cdot \boldsymbol{n}}{\boldsymbol{v}_{rel}^{pre} \cdot \boldsymbol{n}}$$

Away from Collision



At Moment of Collision

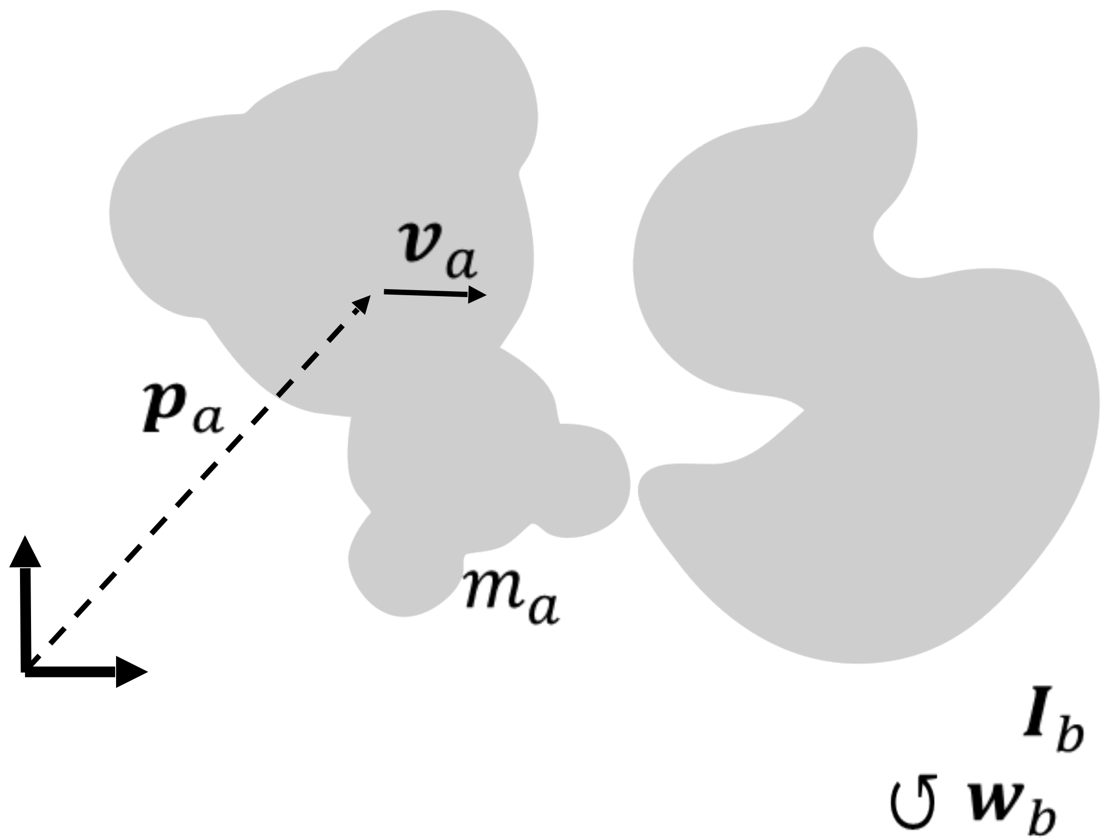
Linear momentum

$$\sum_{i\in\{a,b\}} m_i \boldsymbol{v}_i^{pre} = \sum_{i\in\{a,b\}} m_i \boldsymbol{v}_i^{post}$$

Angular momentum

$$\sum_{i\in\{a,b\}} \boldsymbol{I}_i \boldsymbol{w}_i^{pre} + \boldsymbol{p}_i \times (m_i \boldsymbol{v}_i^{pre})$$

$$= \sum_{i\in\{a,b\}} \boldsymbol{I}_i \boldsymbol{w}_i^{post} + \boldsymbol{p}_i \times (m_i \boldsymbol{v}_i^{post})$$

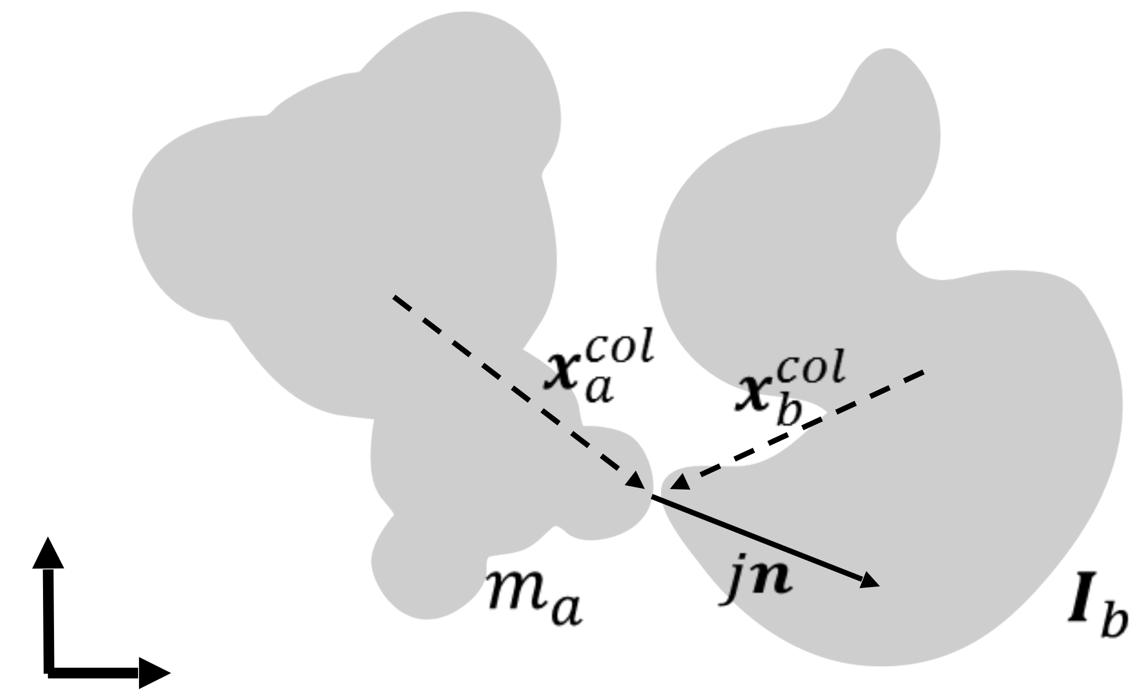


Note: $\boldsymbol{I}_i(t) = \boldsymbol{R}(t)\boldsymbol{I}_i^0\boldsymbol{R}^T(t)$,
and $\boldsymbol{w}_i$ in world coordinates

In Terms of Impulse

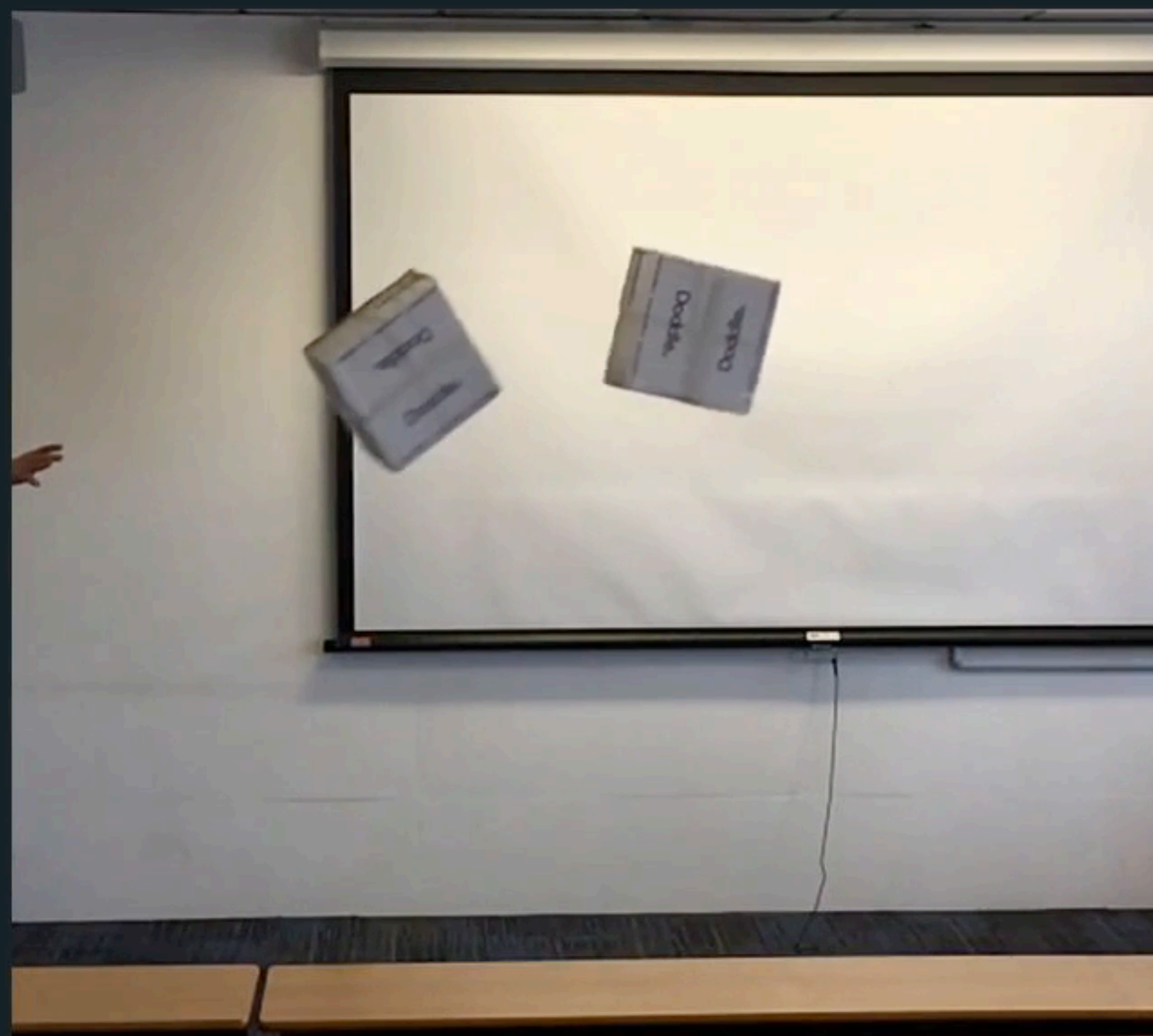
energy is not created

$$\left.\begin{aligned} m_a \boldsymbol{v}_a^{post} &= m_a \boldsymbol{v}_a^{pre} + \\ m_b \boldsymbol{v}_b^{post} &= m_b \boldsymbol{v}_b^{pre} - \\ \boldsymbol{I}_a \boldsymbol{w}_a^{post} &= \boldsymbol{I}_a \boldsymbol{w}_a^{pre} + \boldsymbol{x}_a^{col} \times \\ \boldsymbol{I}_b \boldsymbol{w}_b^{post} &= \boldsymbol{I}_b \boldsymbol{w}_b^{pre} - \boldsymbol{x}_b^{col} \times \end{aligned}\right. j\boldsymbol{n}$$



Method Overview

Input video sequence



Evaluation

Validation - Potential Energy

Comparison with coefficients of restitution calculated from potential energy loss

Evaluation

Results

Results

Input video:

Modeling/Generating Dynamic Env.

Observations
(images)

dynamics
(actors, physics, etc.)

Objects
(3D objects)

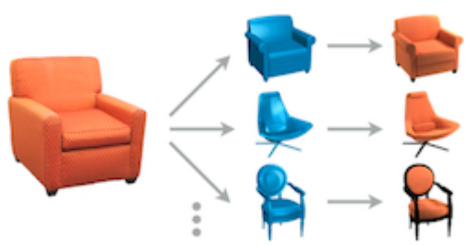
Illumination
(light models)

Priors
(shape manifold,
illumination prior,
room priors)

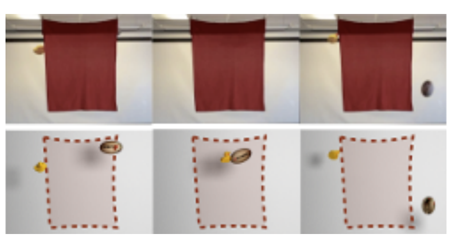
Appearance
(diffuse, specular, texture)

References

CrossLink: Joint Understanding of Image and 3D Model Collections through Shape and Camera Pose Variations,
Moos Heuting, Maks Ovsjanikov, Niloy J. Mitra
SIGGRAPH Asia 2015



Unsupervised Texture Transfer from Images to Model Collections
Tuanfeng Wang, Hao Su, Qixing Huang, Jingwei Huang, Leonidas Guibas, Niloy J. Mitra
SIGGRAPH Asia 2016



SMASH: Physics-guided Reconstruction of Collisions from Videos
Aron Monszpart, Nils Thuerey, Niloy J. Mitra
SIGGRAPH Asia 2016

http://geometry.cs.ucl.ac.uk/index.php

(code and data available)

Symposium on Geometry Processing 2017 London UK

Thank You

http://geometry.cs.ucl.ac.uk/index.php

(code and data available)