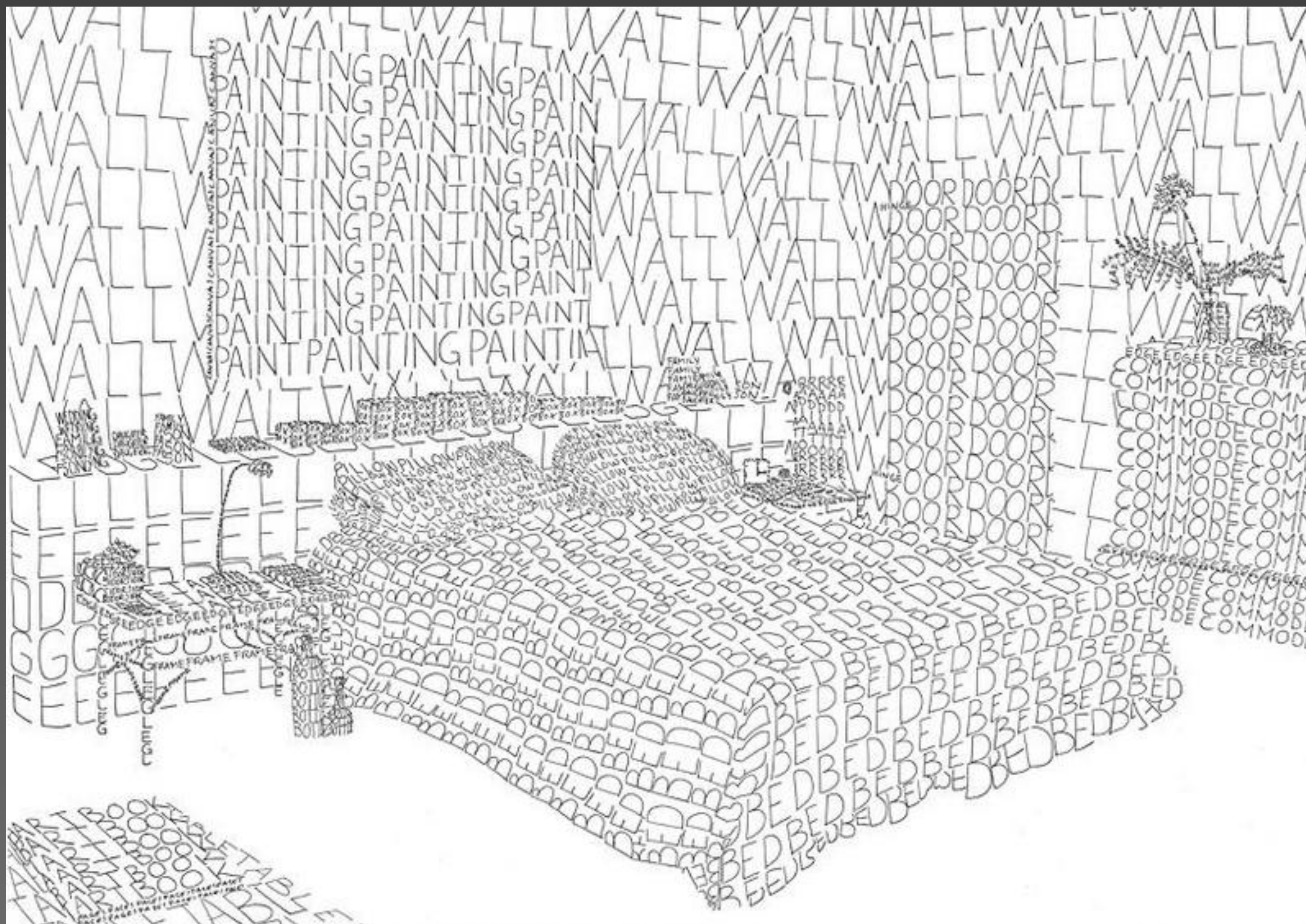
Exploiting the Latent Structures of 3D Geometry and Appearance

Ko Nishino
Dept. of Computer Science
Drexel University

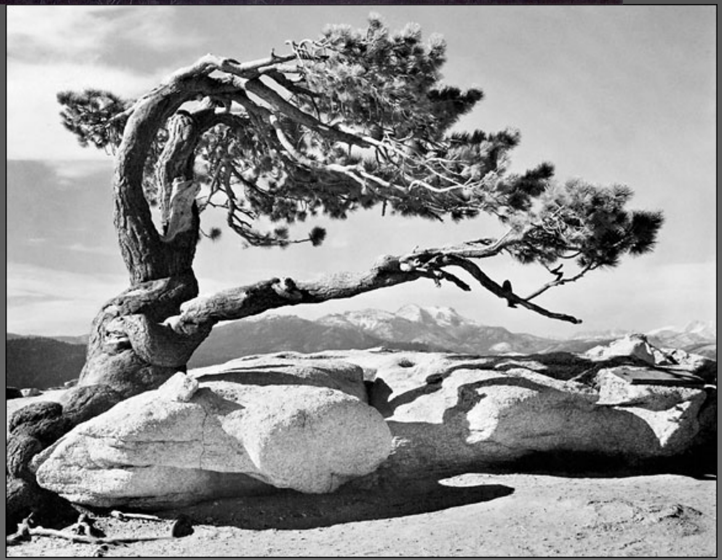
Drexel UNIVERSITY

Our Visual World is Intricate

But Our World is Structured



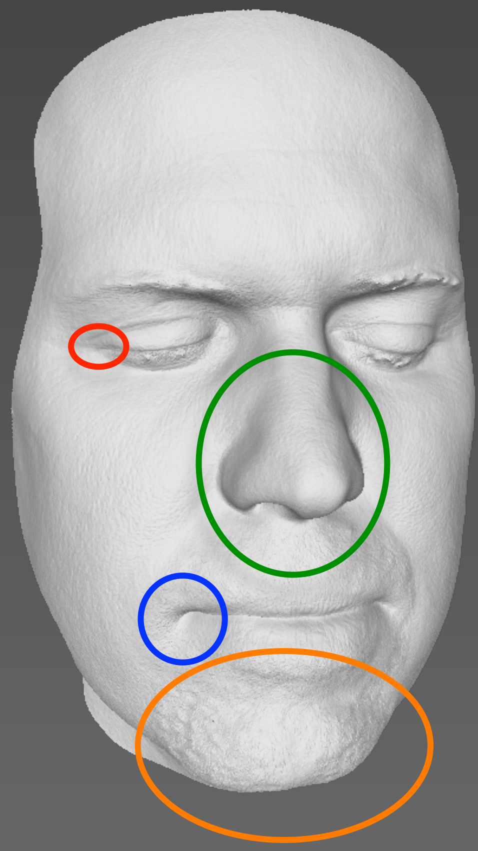
In the Geometry



Ansel Adams

Geometric Scale Variability

- Scales of local 3D geometric structures
  - Natural support sizes of "structures"
  - Relative variation within an object
  - Scales that are relevant (observable)
- Hidden dimension of 3D geometry
  - Characterizes the overall structure
  - Reveals hierarchical structure

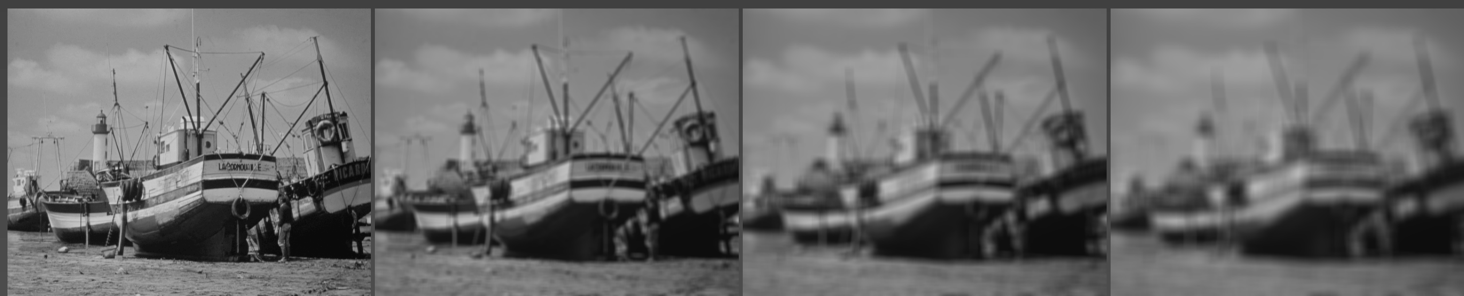


3D model courtesy of USC ICT Graphics Lab

Related Work

- Multi-scale features and descriptors
  - [Li & Guskov 05] [Gelfand et al. 05] [Lalonde et al. 05] [Dinh & Kropac 06] [Pauly et al. 06] [Skelly & Sclaroff 07] …
- Mesh smoothing
  - [Taubin 95] [Desbrun et al. 99] [Eck et al. 02] [Jones et al. 03]
- Range image characterization
  - [Ponce & Brady 85] [Morita 99] [Mokhtarian 01]
- Mesh saliency
  - [Lee, Varshney, and Jacobs 05]

Image Scale-Space

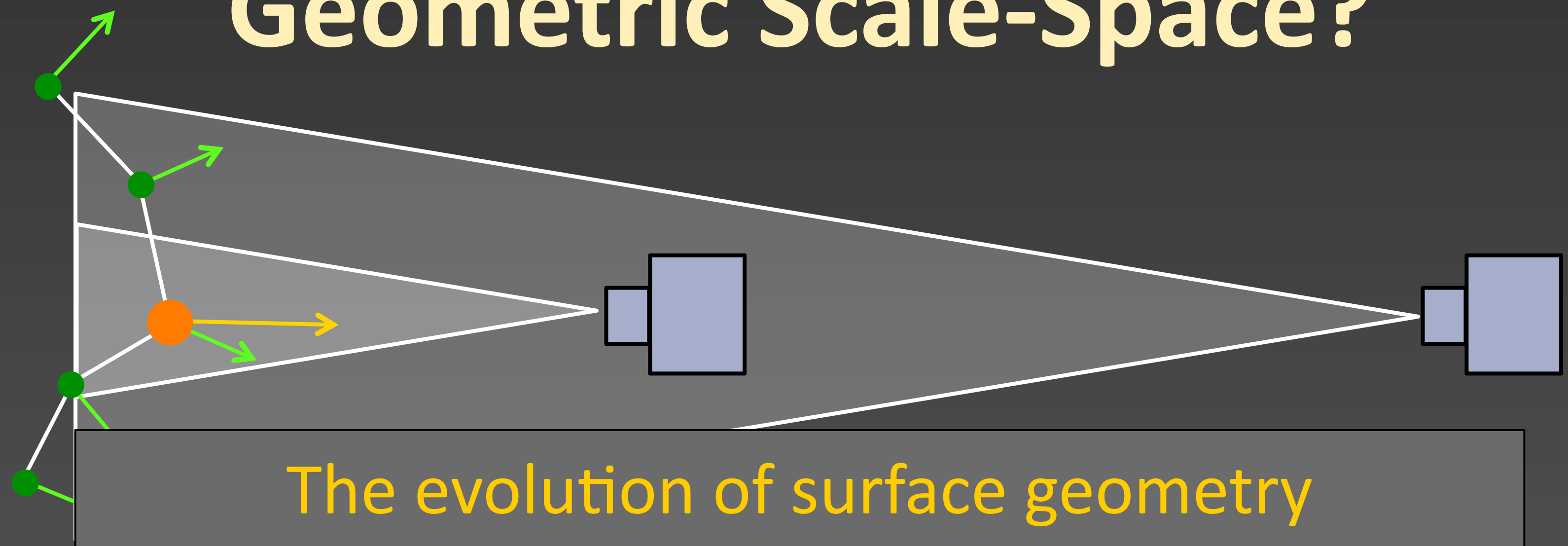


- Simulate how the scene would look like at different distances (w/o the subsampling)
- Gaussian scale-space [Lindeberg 90]…
  - Diffusion equation

$$\frac{\partial I}{\partial t} = \frac{1}{2}\nabla^2 I$$

  - Scale-space axioms, esp., the causality assumption

Geometric Scale-Space?



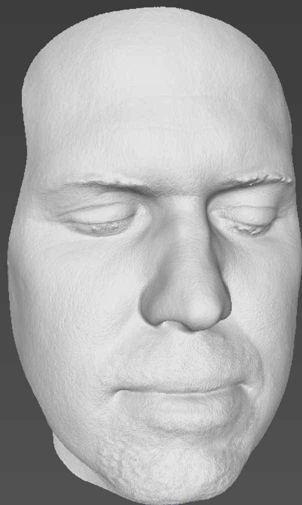
The evolution of surface geometry
as its high-frequency variations are suppressed

- 3D points define the sampling not the signal
  - The actual geometry should not change
  - Evolution on the surface not the embedding

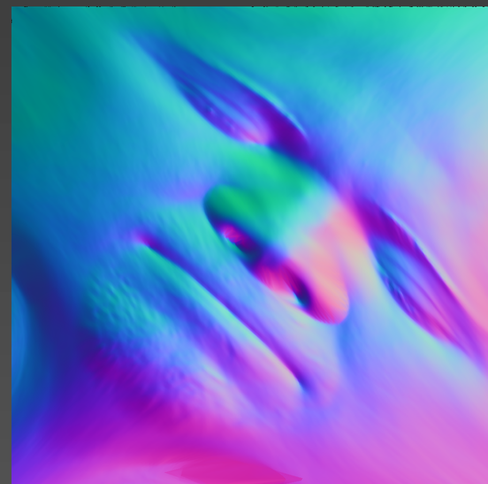
cf. [Lee, Varshney, and Jacobs 05]

- Surface geometry in its rawest form
  - Surface normals inherently intrinsic to the surface
  - Distances measured on the surface

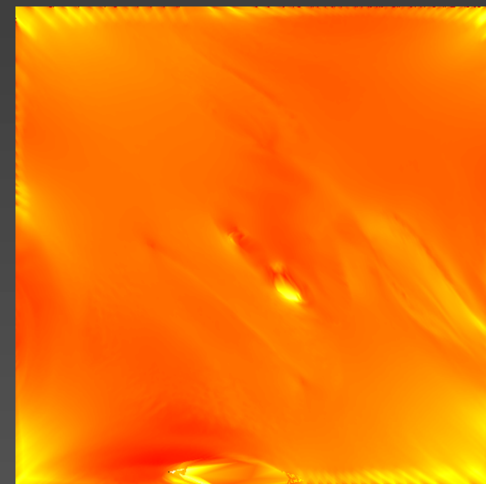
Representing Surface Geometry



$\mathcal{M}$



$\mathcal{D}$



$\mathcal{E}$

- 2D surface normal map
  - Reparametrize vertices and interpolate

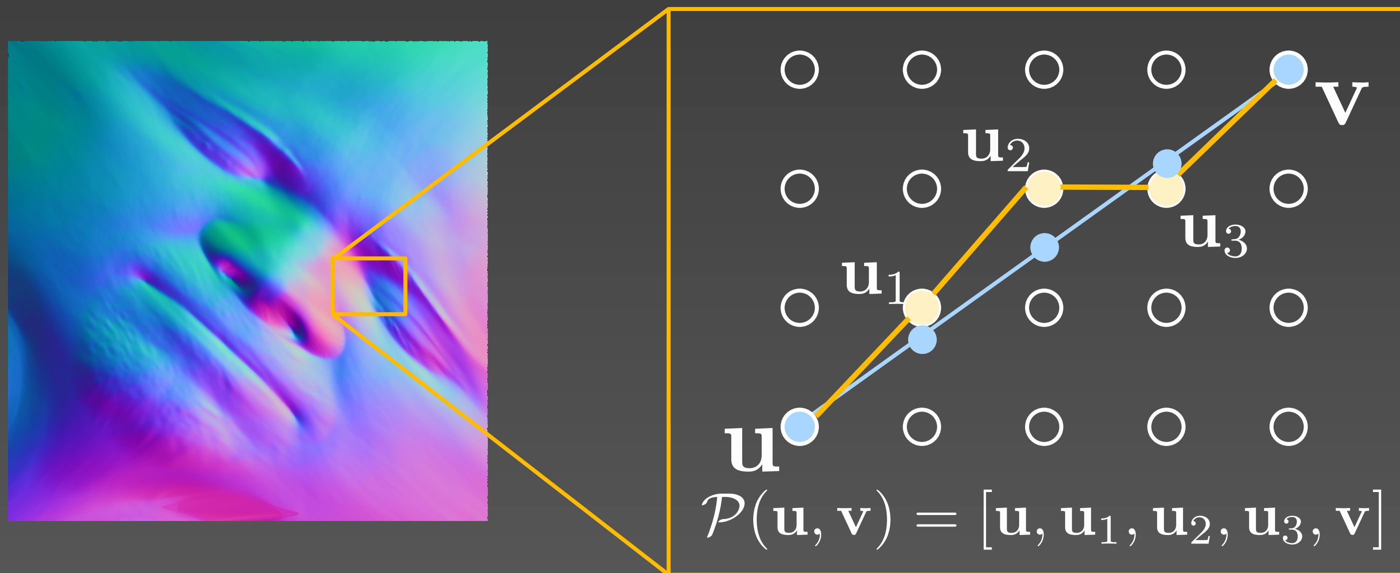
$$\phi : \mathcal{D} \rightarrow \mathcal{M}$$

[Floater et al. 05, Yoshizawa et al. 04]

  - Distortion

$$\varepsilon(\mathbf{u}) = \frac{1}{|\mathcal{A}(\mathbf{u})|} \sum_{\mathbf{v} \in \mathcal{A}(\mathbf{u})} \frac{\| \mathbf{u} - \mathbf{v} \|}{\| \phi(\mathbf{u}) - \phi(\mathbf{v}) \|} \quad \mathbf{u} = (s, t) \in \mathcal{D}$$

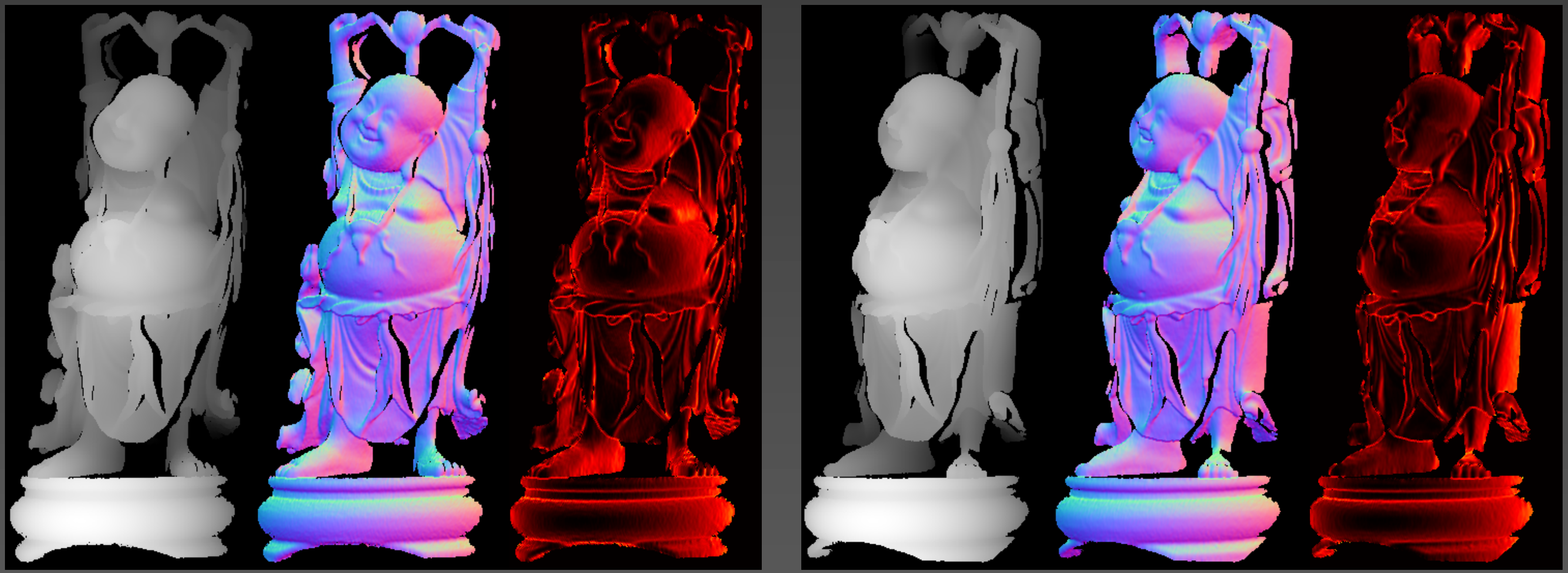
Distance Metric



- Approximate geodesic distance

$$d(\mathbf{u}, \mathbf{v}) \approx \sum_{\mathbf{u}_i \in \mathcal{P}(\mathbf{u},\mathbf{v}), \neq \mathbf{v}} \frac{\varepsilon(\mathbf{u}_i)^{-1} + \varepsilon(\mathbf{u}_{i+1})^{-1}}{2} \parallel \mathbf{u}_i - \mathbf{u}_{i+1} \parallel$$

Range Images



- Main form of geometric data in computer vision
- Already 2D embeddings
  - Perspective/Orthographic projection

Geometric Scale-Space Operator

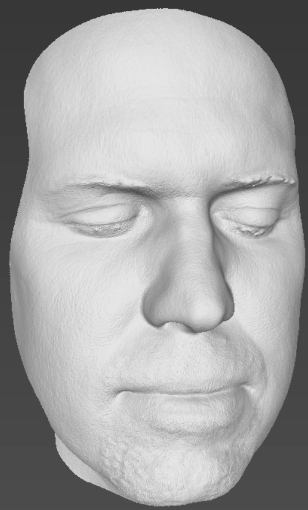
$$\min_{\mathbf{N}:\mathbb{R}^2 \to \mathbb{S}^2} \int\int_D \| \nabla \mathbf{N} \|^2 \, ds dt$$

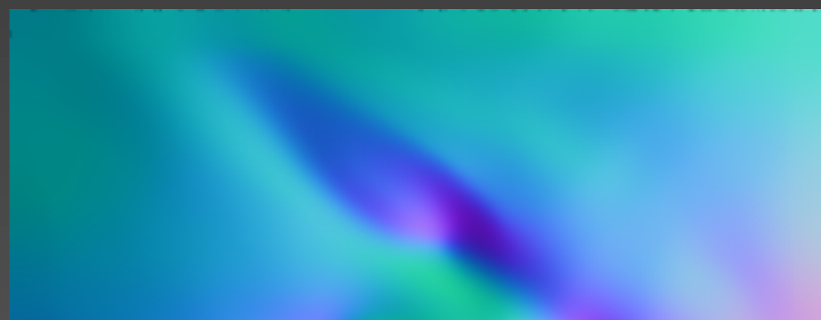
- 2-harmonic flow [Tang et al. 00]

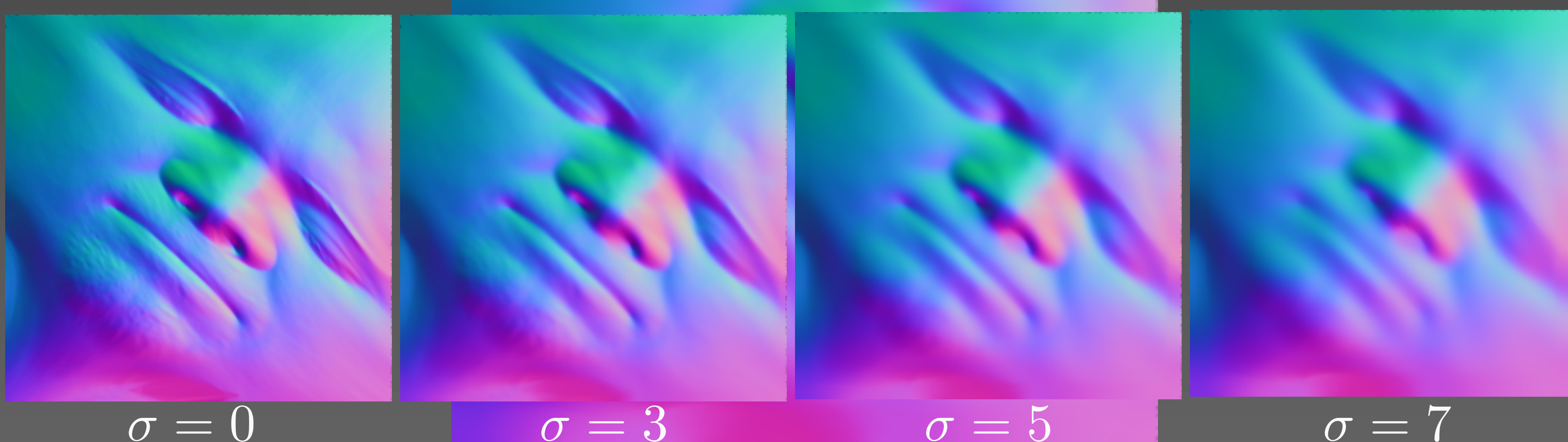
$$\frac{\partial N_i}{\partial t} = \nabla^2 N_i + N_i \| \nabla \mathbf{N} \|^2 \quad (i = x, y, z)$$

- Geodesic Gaussian smoothing
  - Gaussian smoothing of normals in the 2D domain
  - Renormalization after each step
- Causality not guaranteed but rarely violated

Geometric Scale-Space

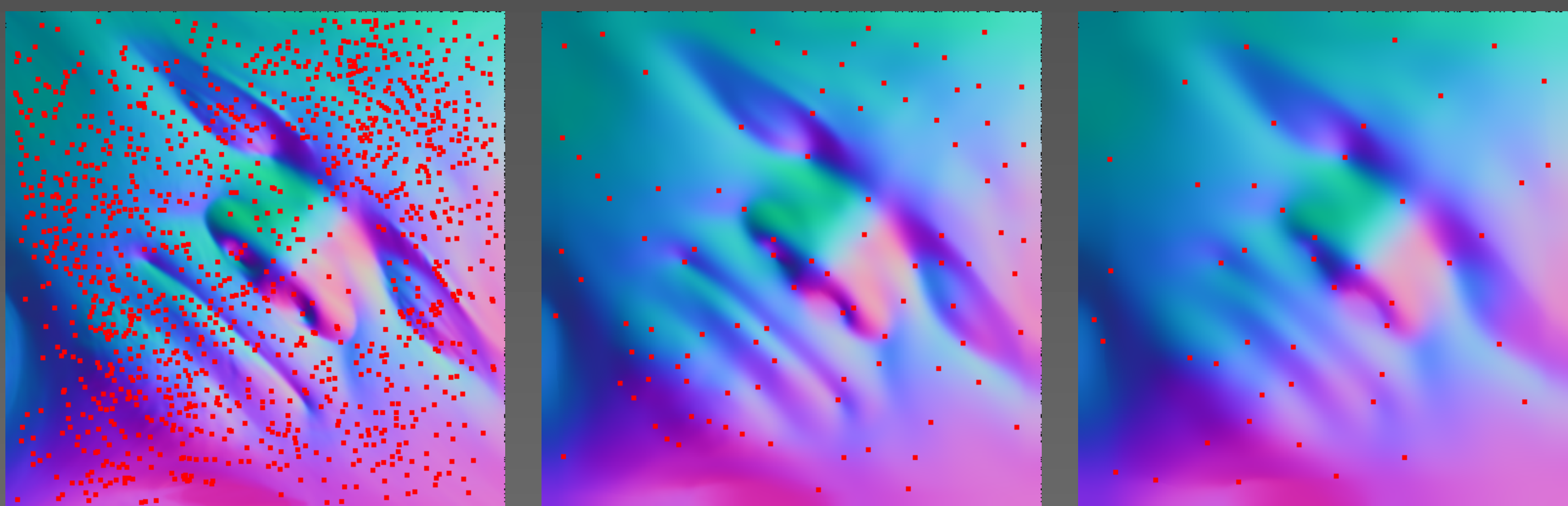






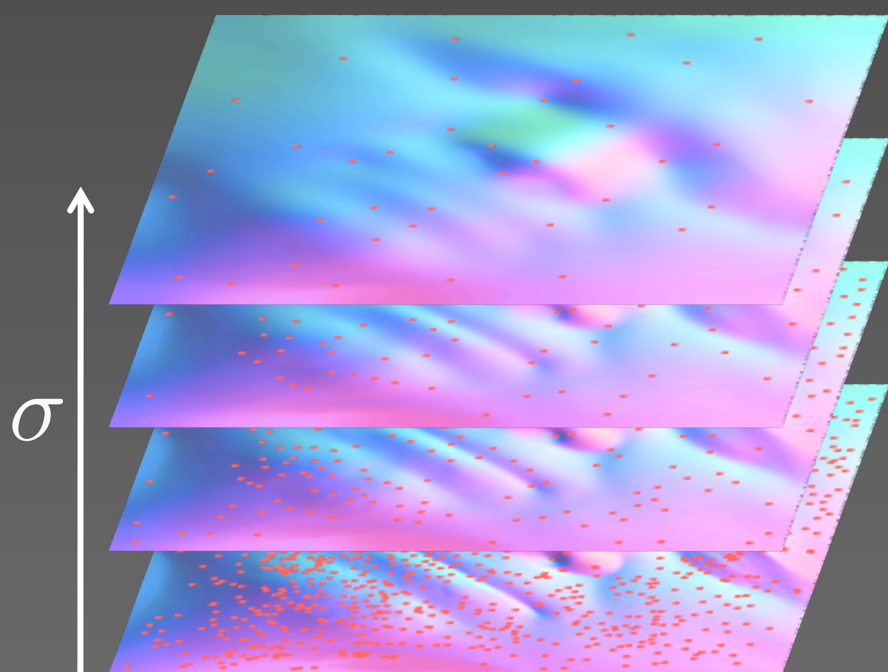
Features: Corners

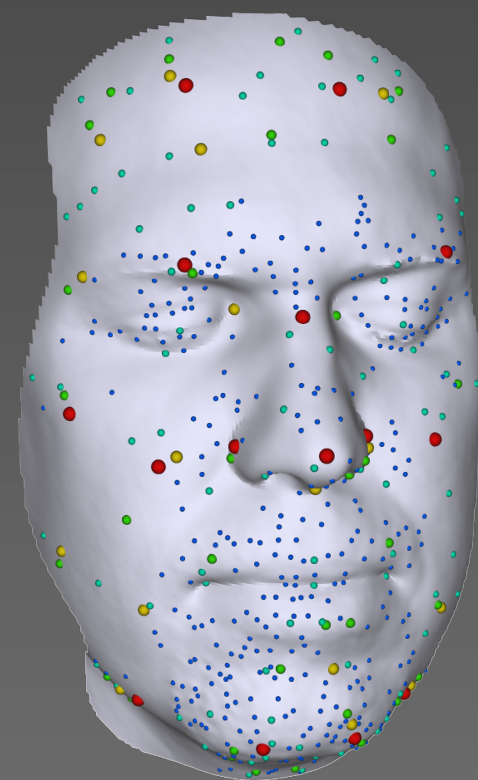
- Gram matrix of gradients of the normal map
  - Gradients derived based on normal curvature in s, t
  - First eigenvalue as the corner response
  - Corners on edges pruned using 2nd-order deriv's



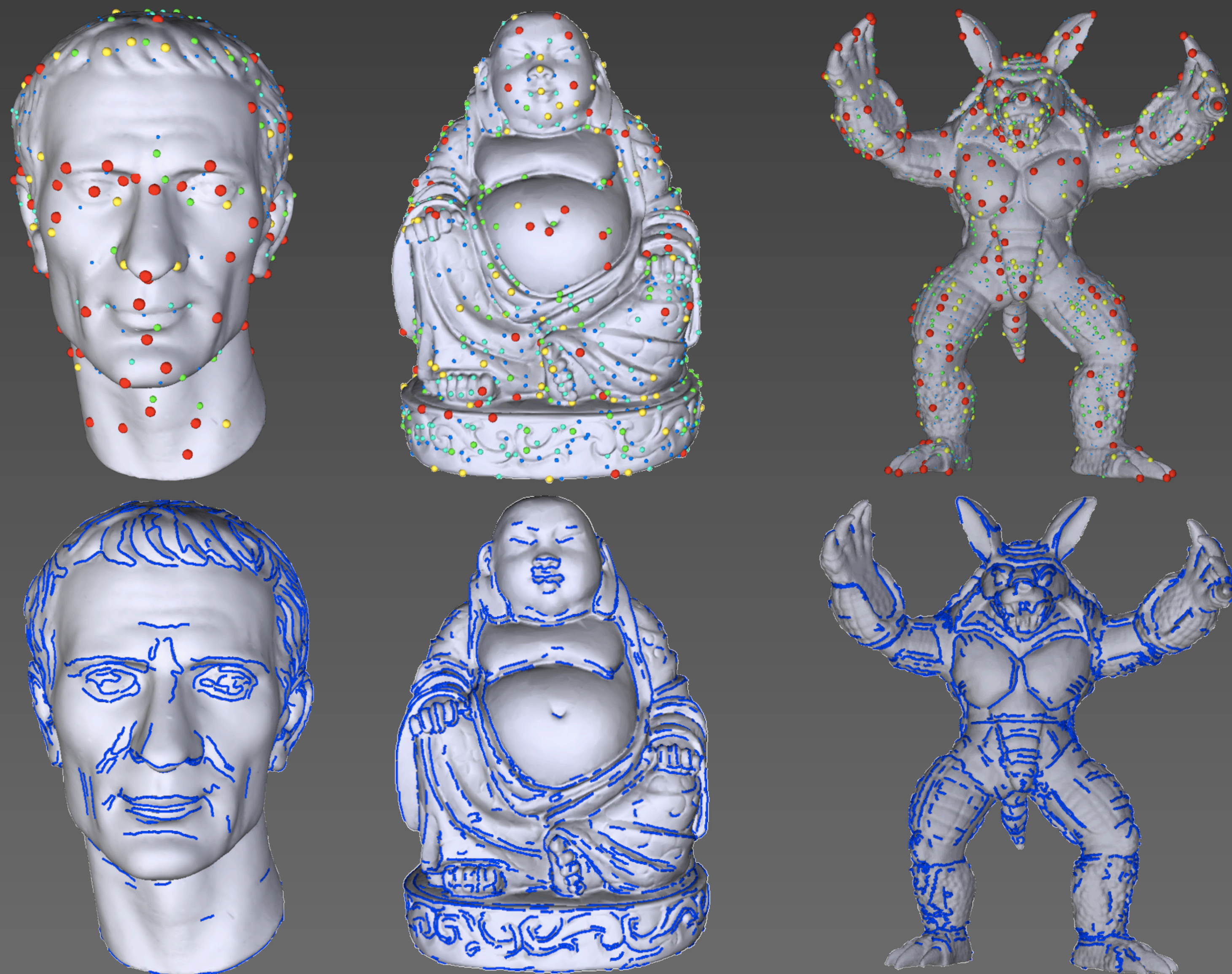
Scale Selection (cf. [Lindeberg 98])

- Identify the "natural scale" of each feature
  - Normalize derivatives by weighting with σ^{γ} and $\sigma^{2\gamma}$
  - Maximum feature response across all scales

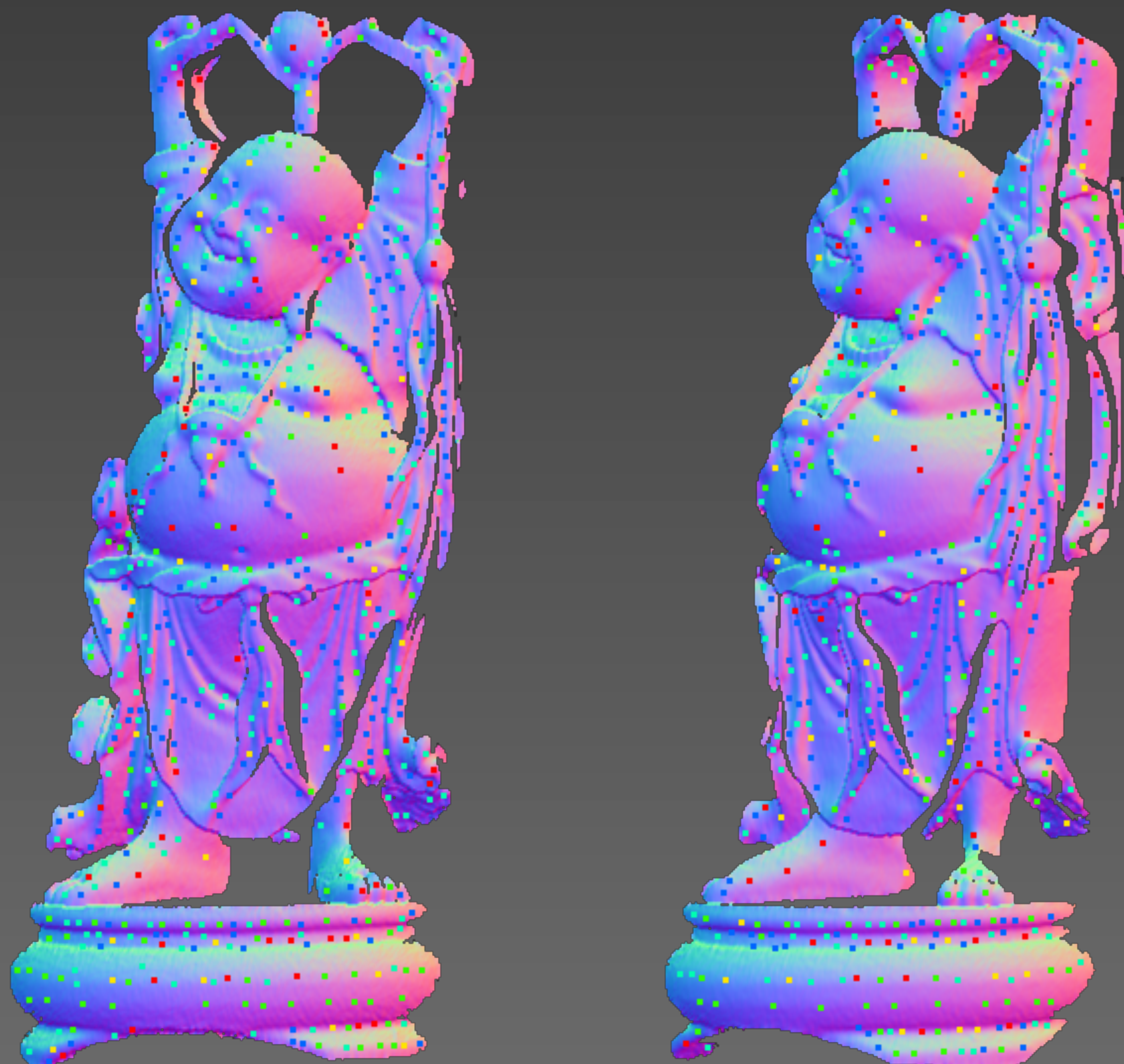




Scale-Dependent Features

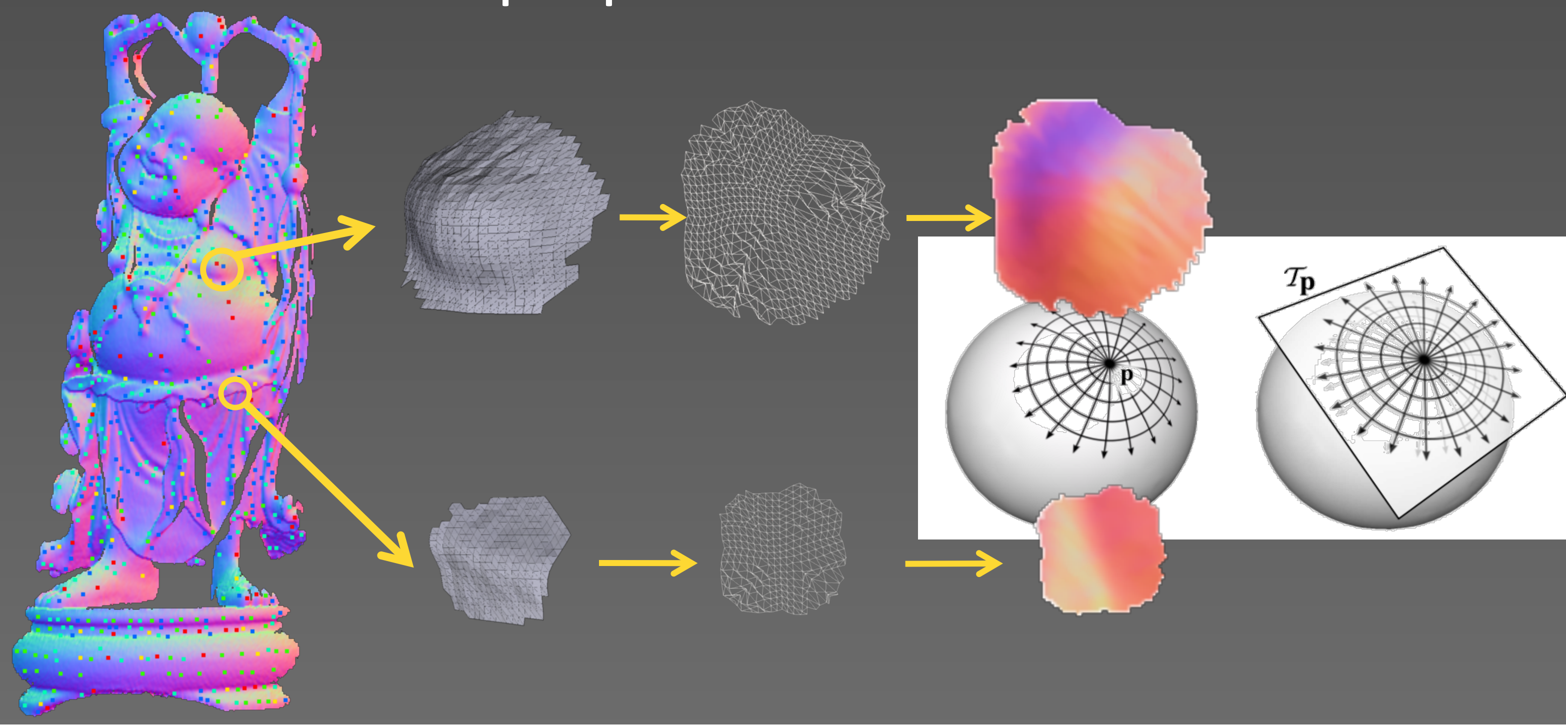


Scale-Dependent Features

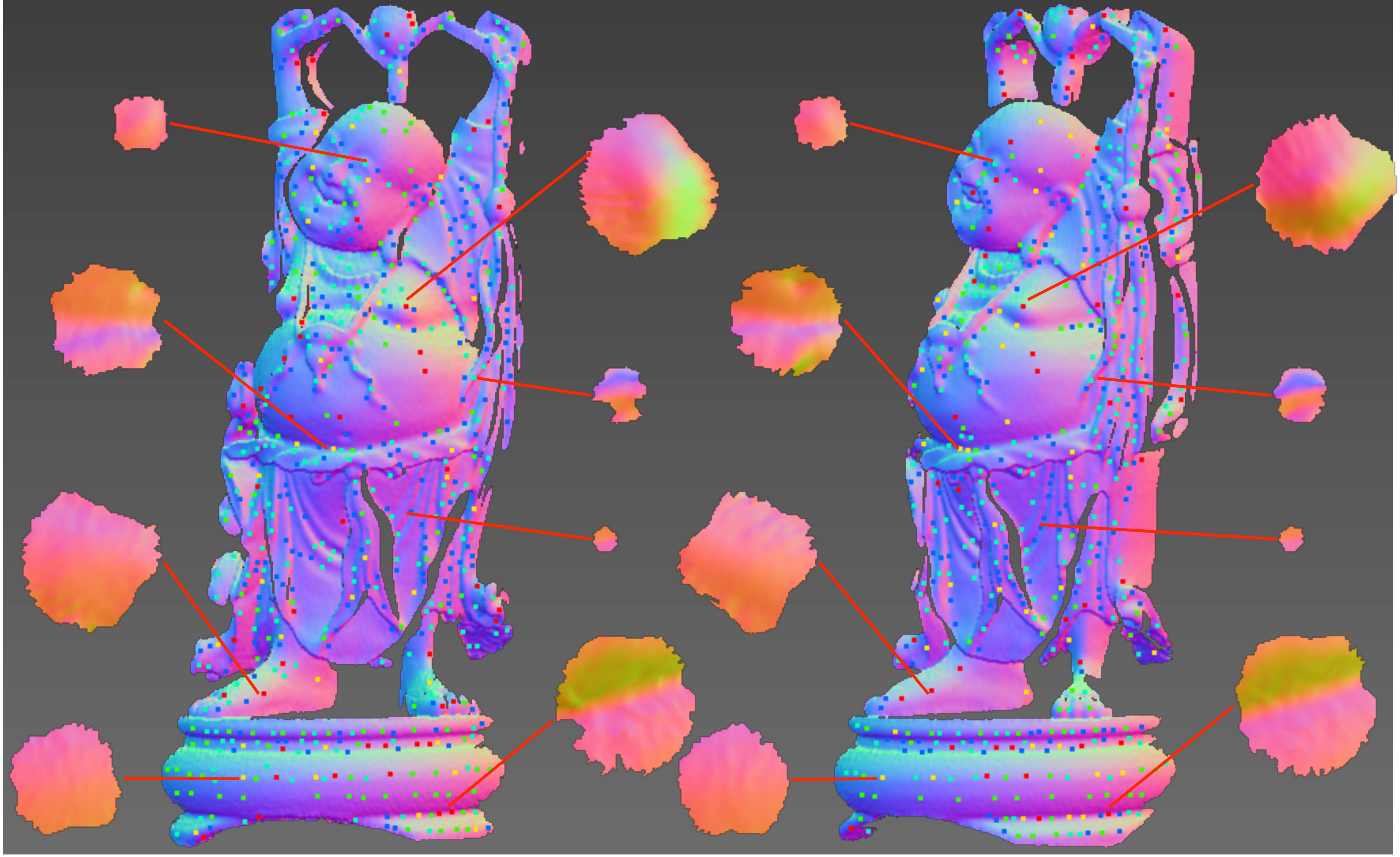


Scale-Dependent Local Shape Descriptors

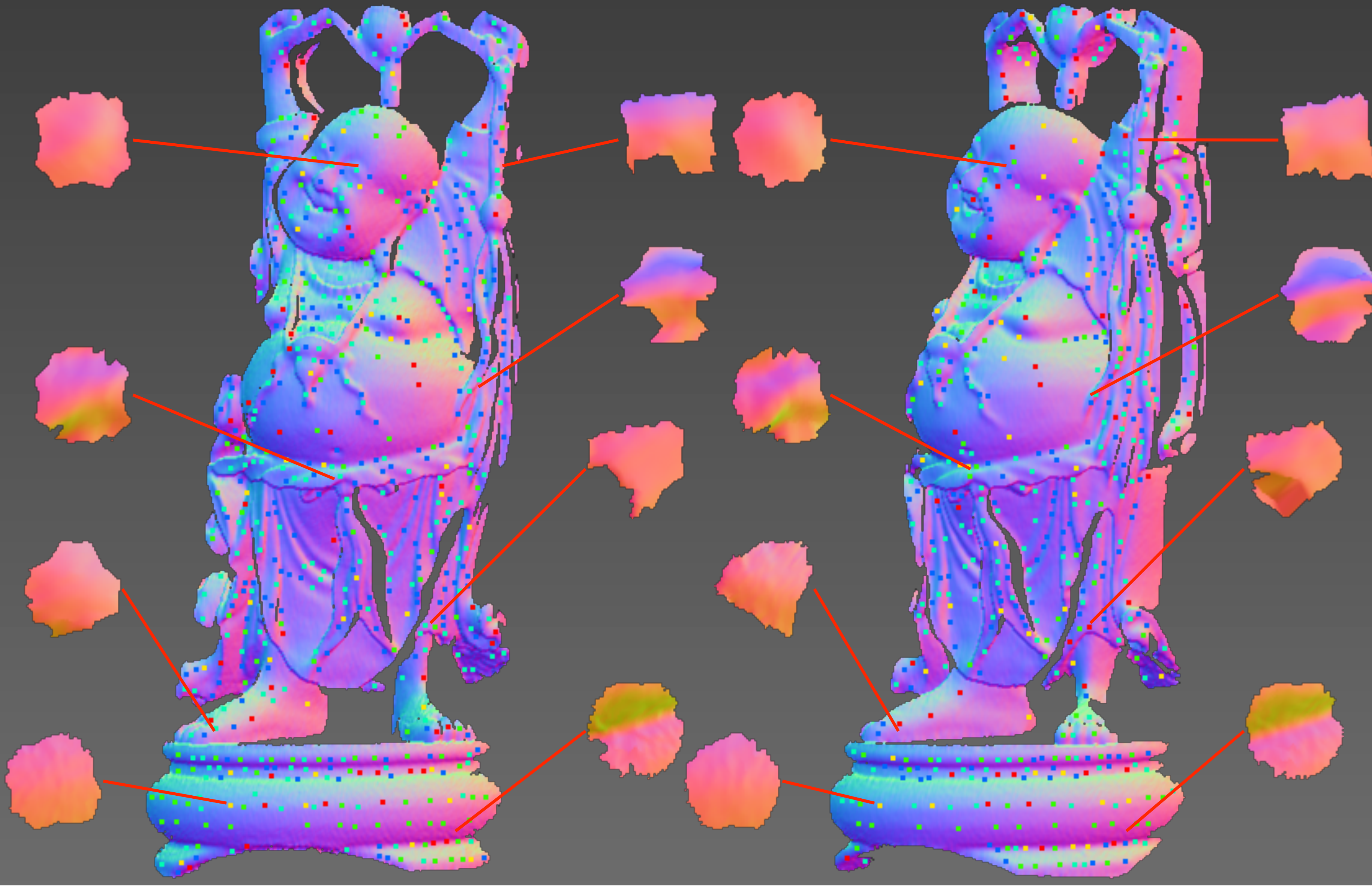
- Encode the local geometric structure that give rise to each scale-dependent corner
  - Exponential map to encode local normals
  - Geodesic radius proportional to scale of corner



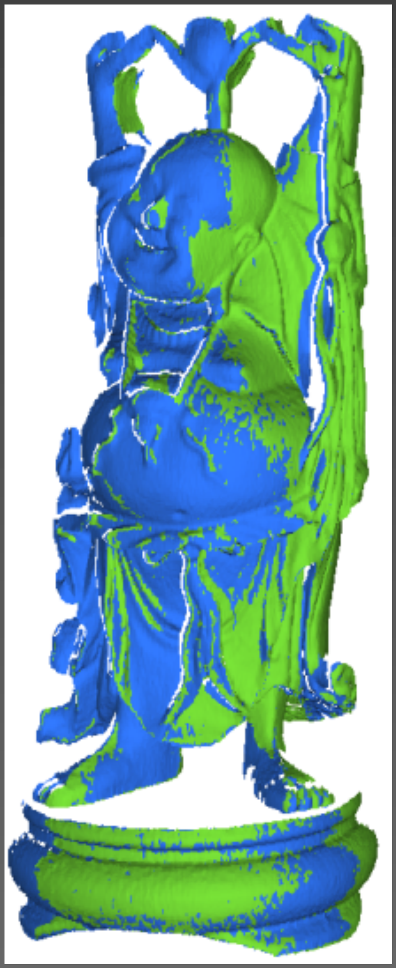
Scale-Dependent Local Shape Descriptors



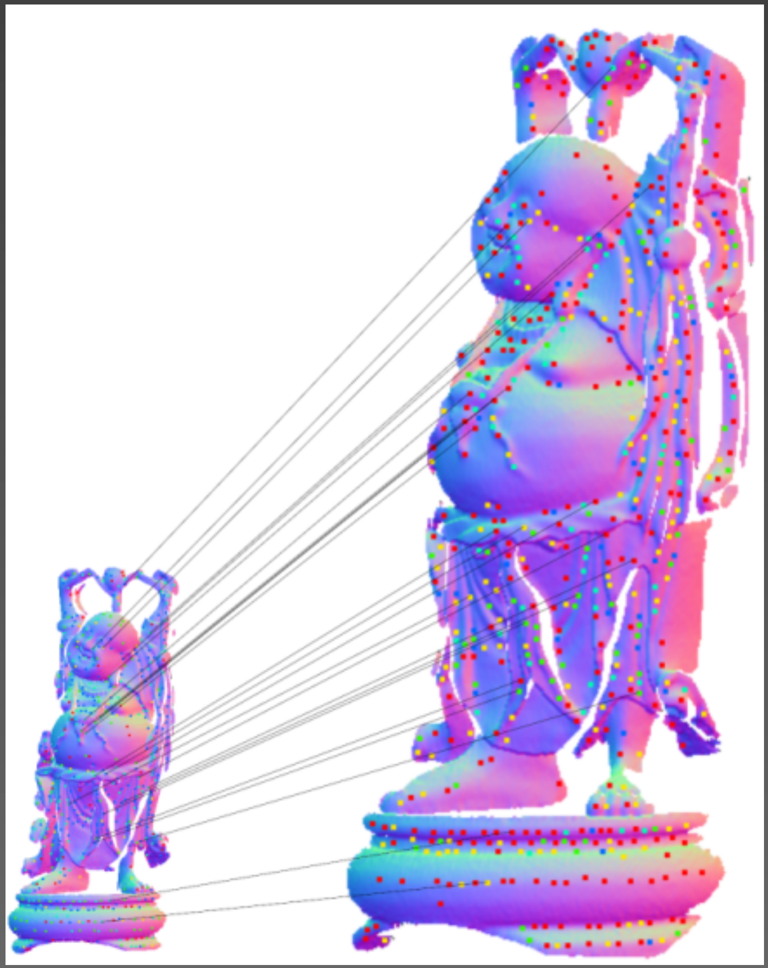
Scale-Invariant Local Shape Descriptors

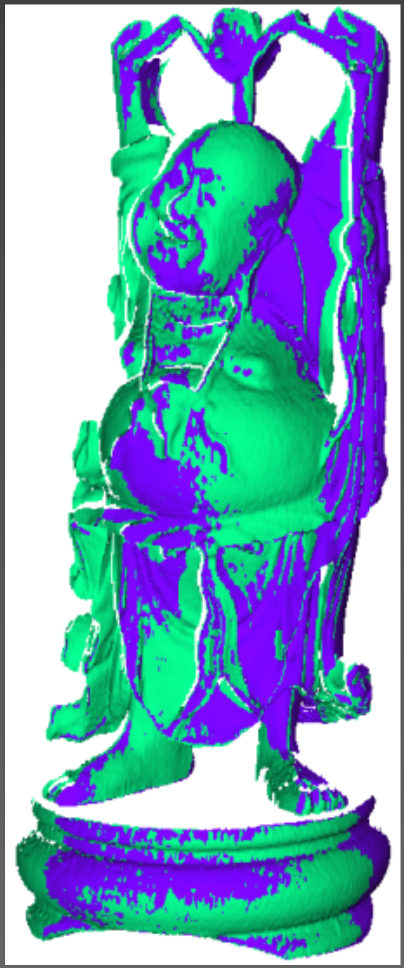


Scale-Hierarchical Matching

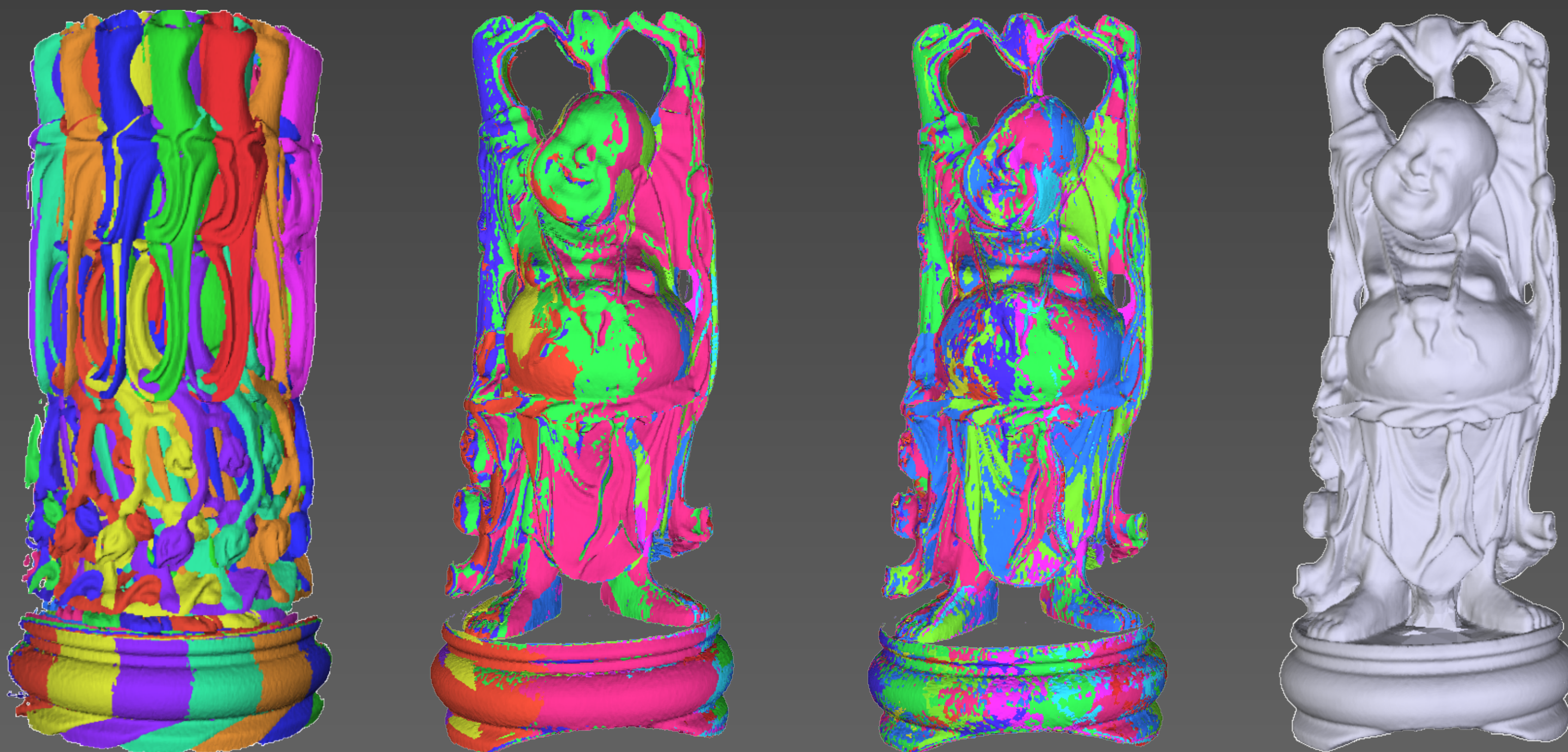


Scale-Invariant Matching



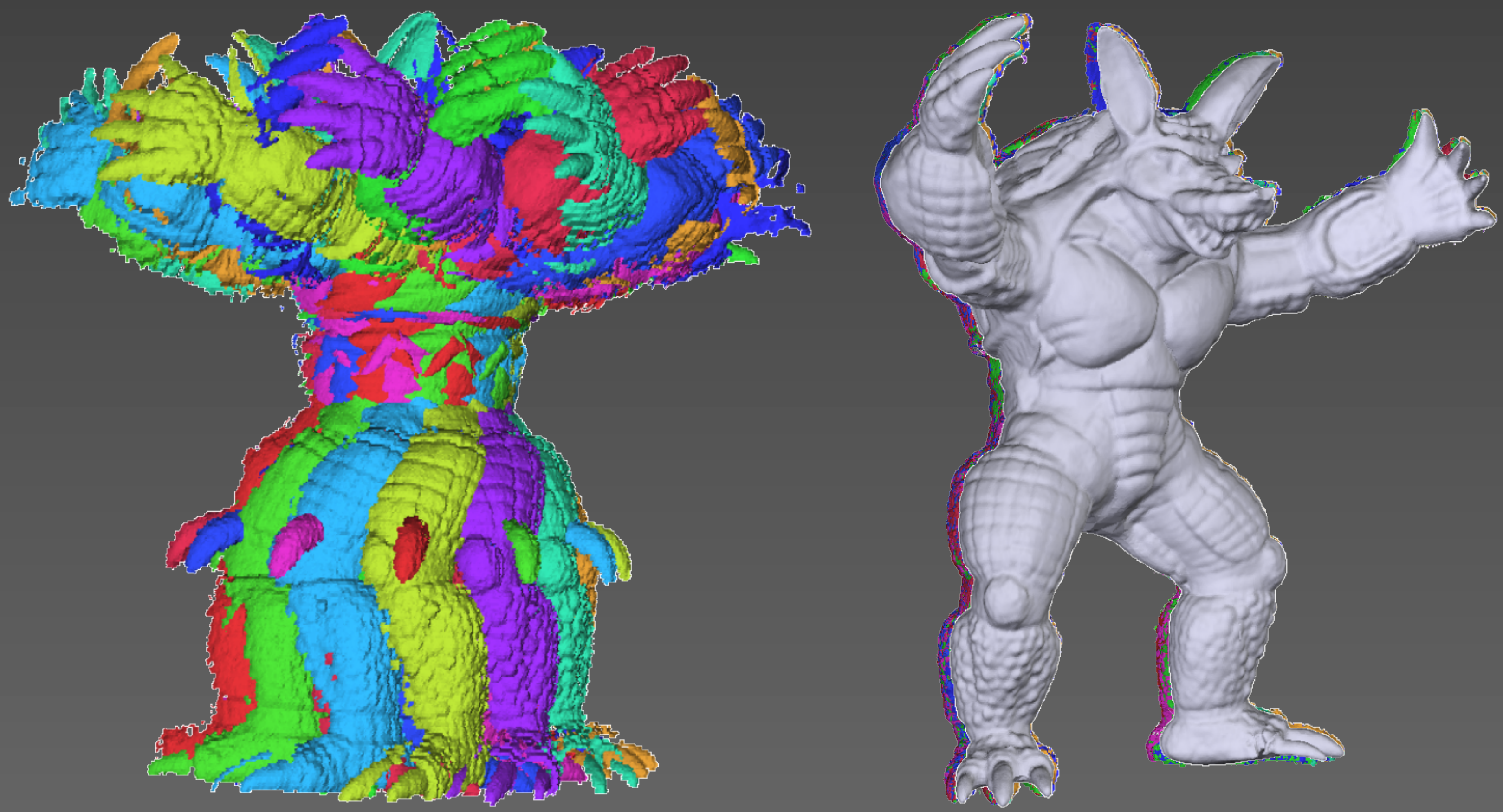


Fully-Automatic Multi-View Registration



Input | Our Result cf. [Huber and Hebert 03] | After ICP [Nishino et al. 02] | After Surface Recon. [Kazhdan et al. 06]

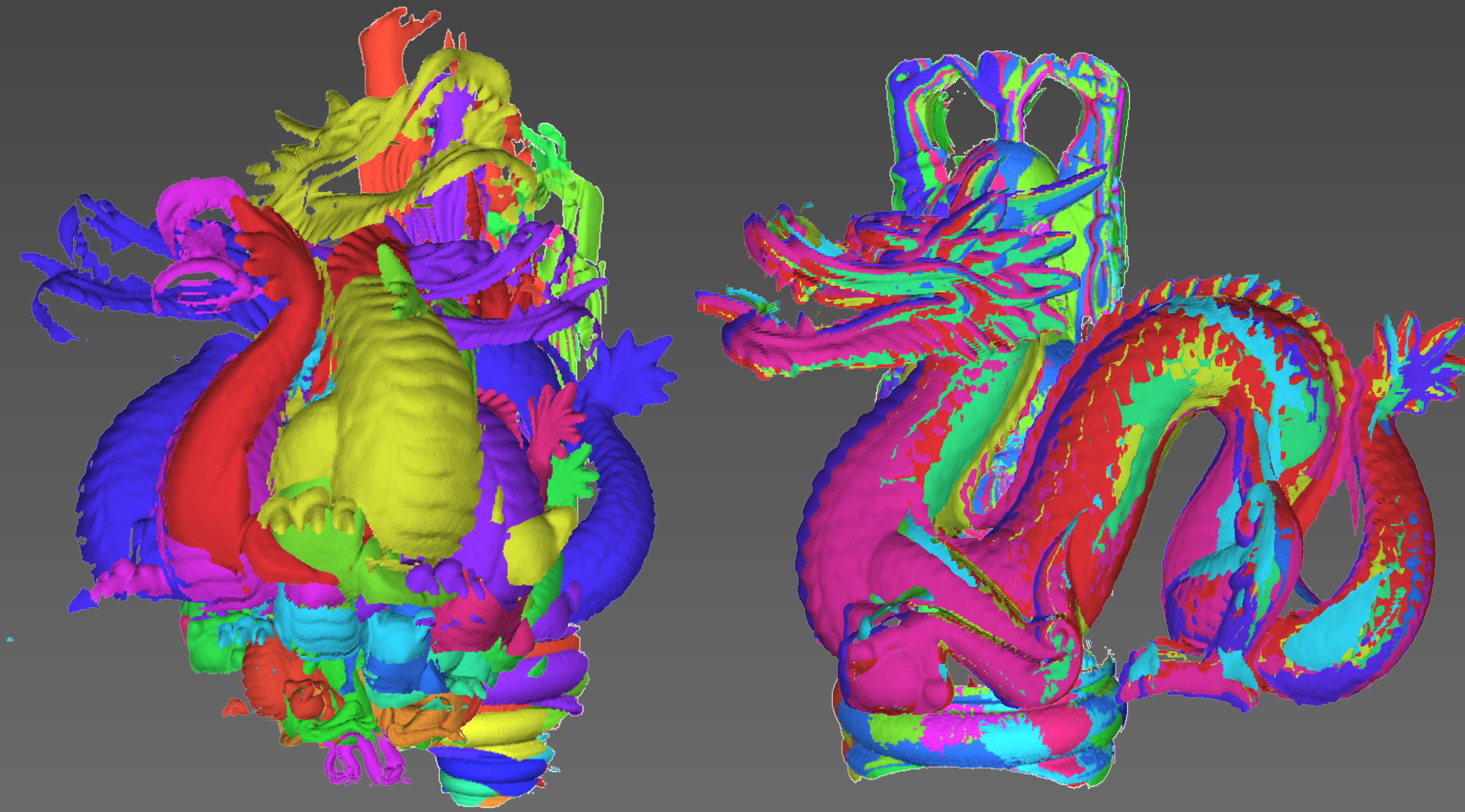
Fully-Automatic Multi-View Registration



cf. [Huber & Hebert 03] [Makadia et al. 06]

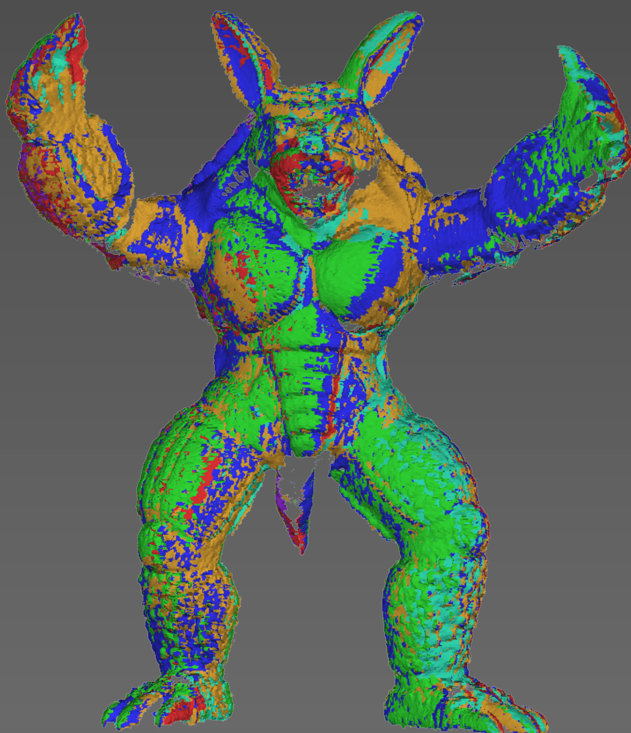
Scale-Invariant
Fully-Automatic Multi-View Registration

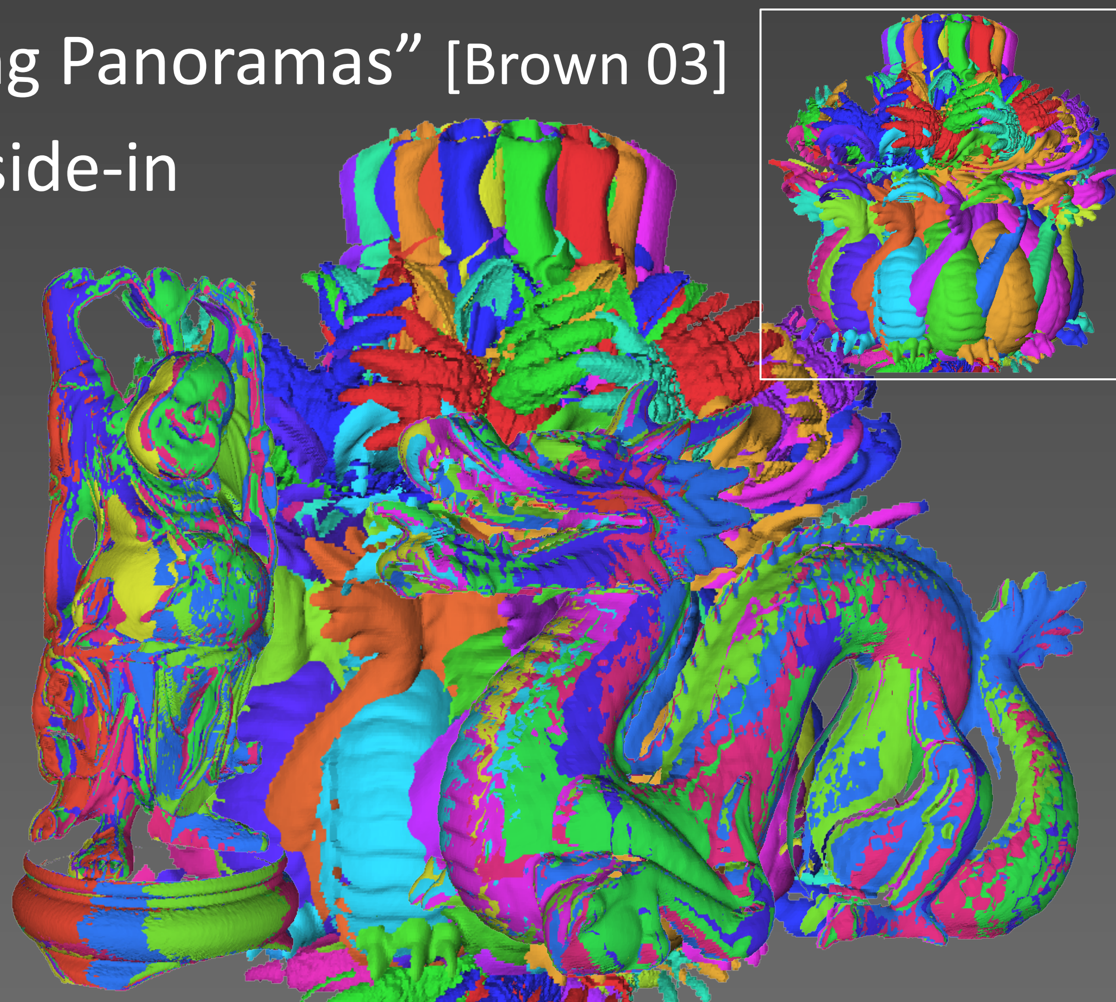
- Range images with different global scales



Multiple 3D Objects from A Mixed Set of Range Images

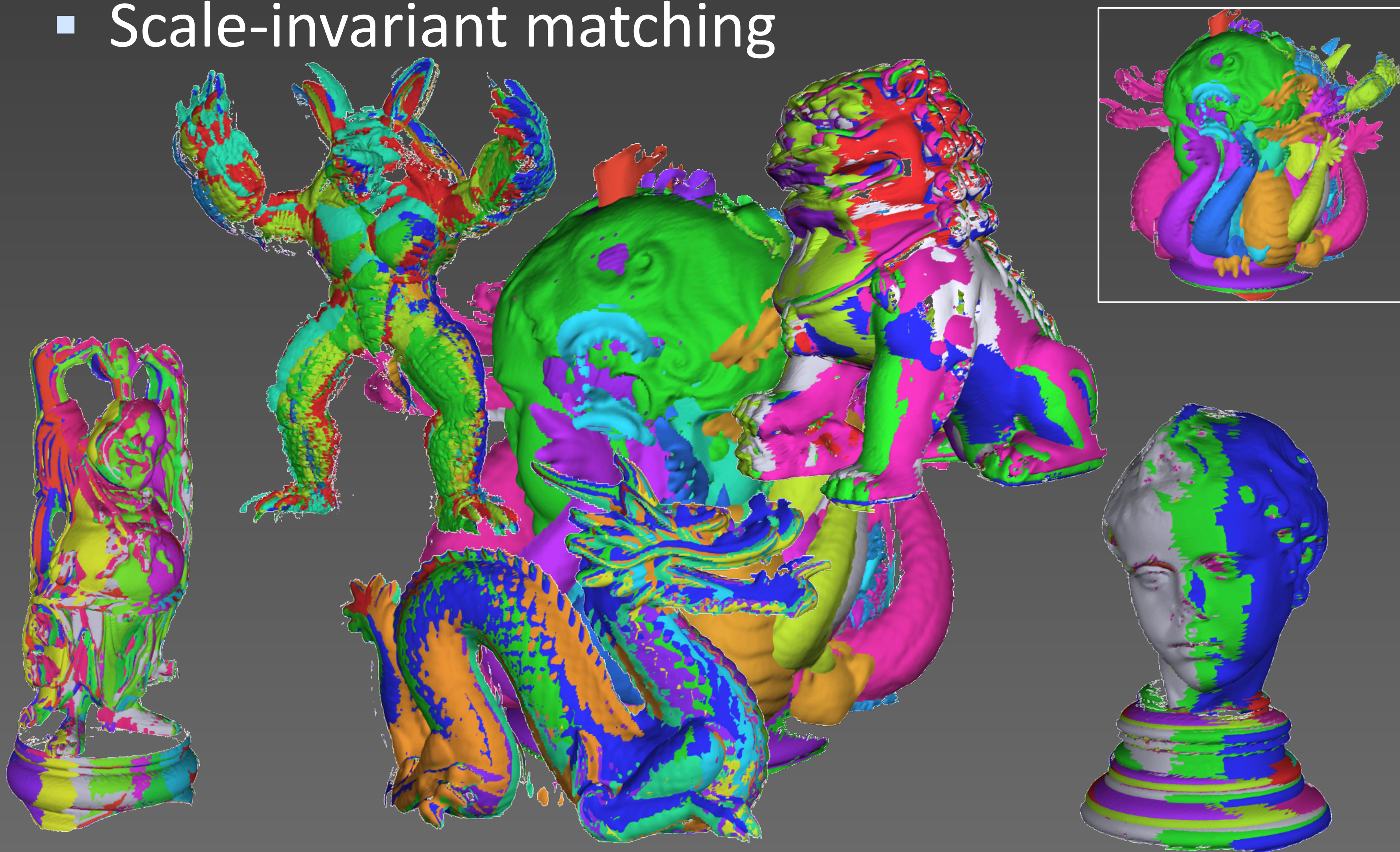
- Automatically discovers 3D object from a pile
  - A la "Recognising Panoramas" [Brown 03]
  - But 3D and outside-in



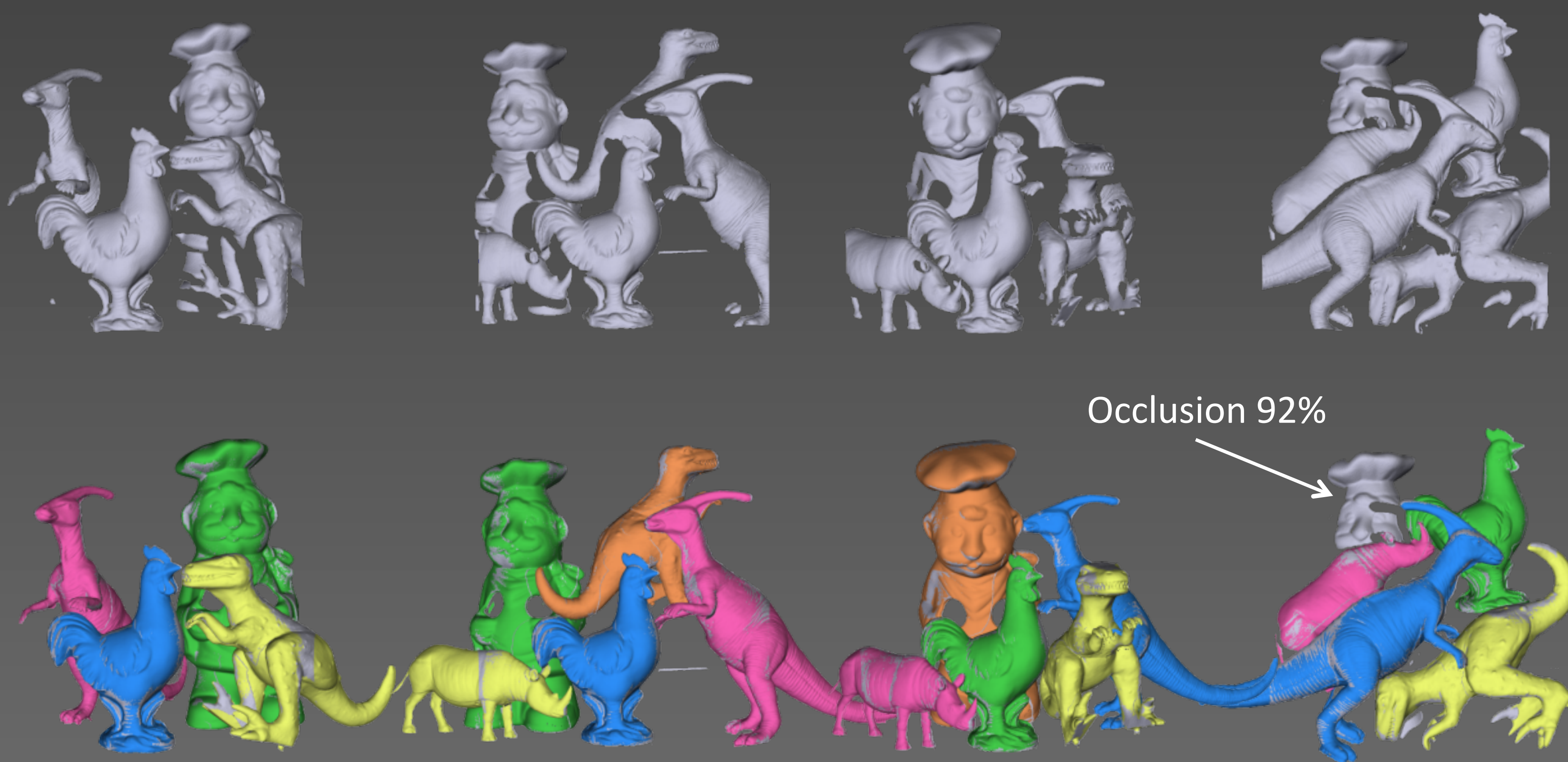


Multiple 3D Objects from A Mixed Set of Range Images

- Scale-invariant matching

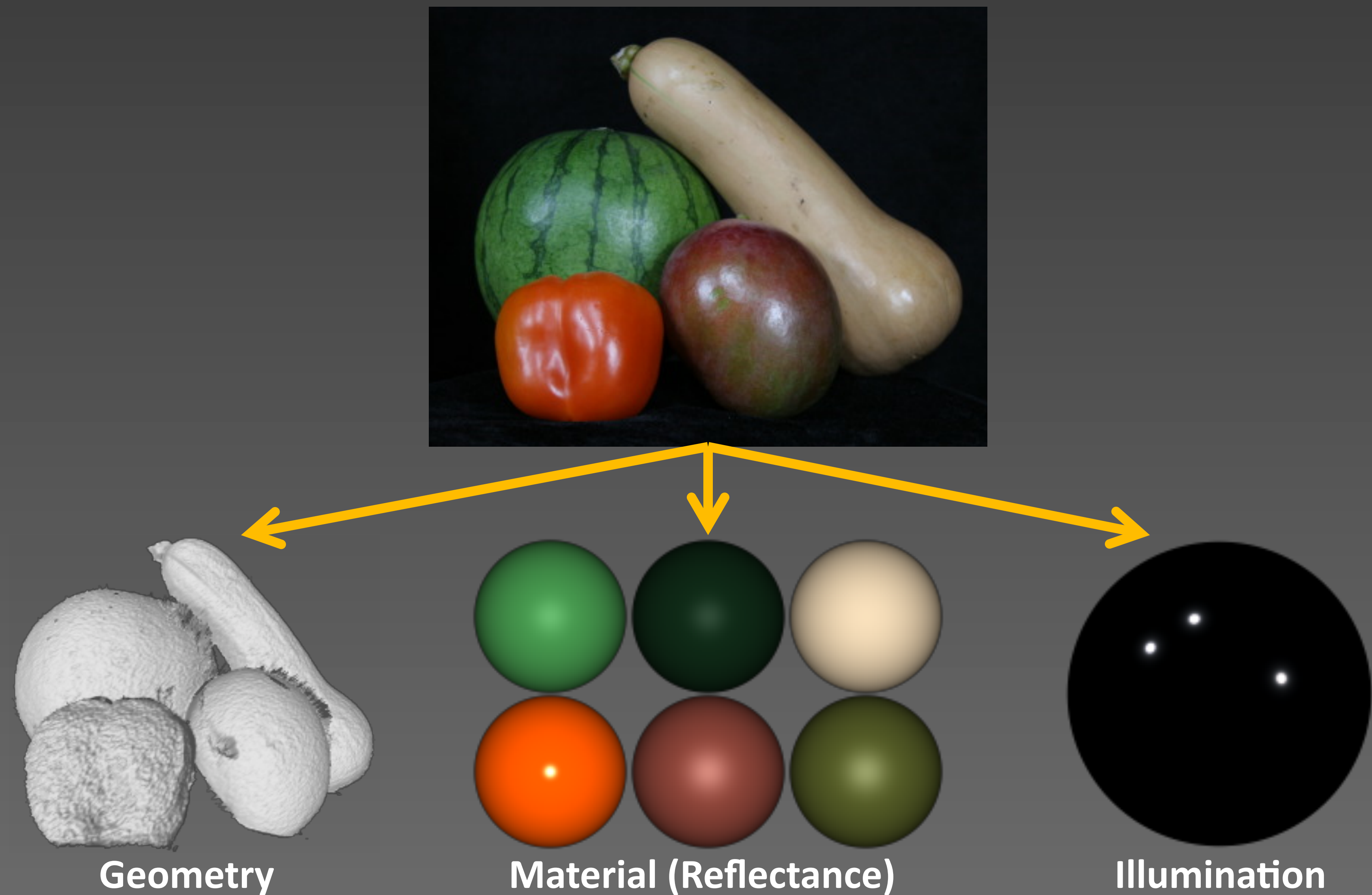


Scale-Hierarchical 3D Object Recognition



In the Appearance

Radiometric Scene Decomposition



The Role of Reflectance

$$\text{image} = f_{\textbf{material}}(\text{illumination}, \text{geometry})$$

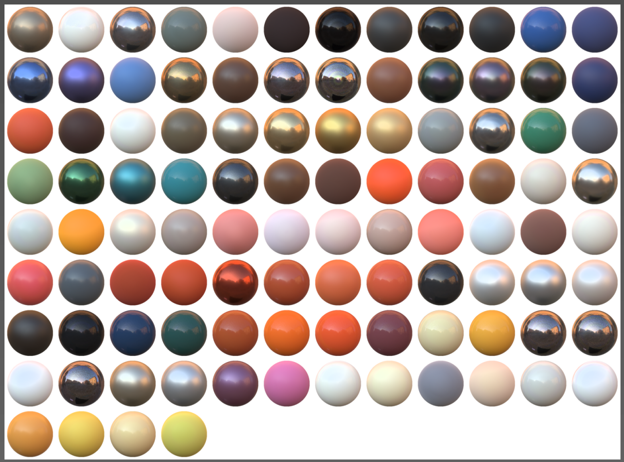
- Complex interplay of light with the surface
- Material determines the interaction (**reflectance**)

$$f^{-1}_{\textbf{material}}(\text{image}) = \{\text{illumination}, \text{geometry}\}$$

- Object recognition based on materials
- Tracking and navigation under varying illumination
- Geometry reconstruction of arbitrary objects
- Image synthesis of complex real-world scenes
- …

Representing Reflectance

- Lambertian (occasionally with Torrance-Sparrow)
  - Prevalent in all radiometric decomposition methods
  - Current tradeoff of accuracy vs. analytical simplicity
- Parametric models
  - Low-dimensional analytic form
  - Limited expressiveness
- Non-parametric models
  - Great for accuracy
  - Cursed by its high dimensionality

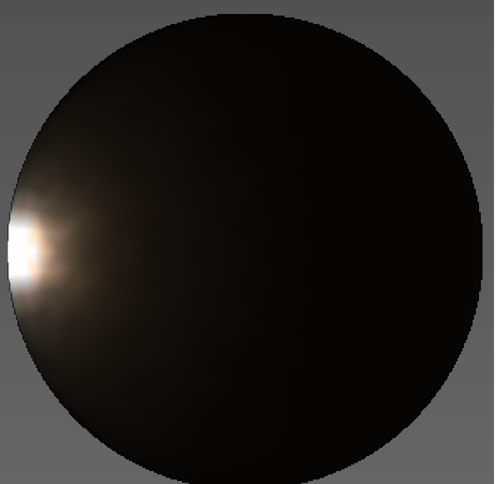


MERL Isotropic BRDF Database

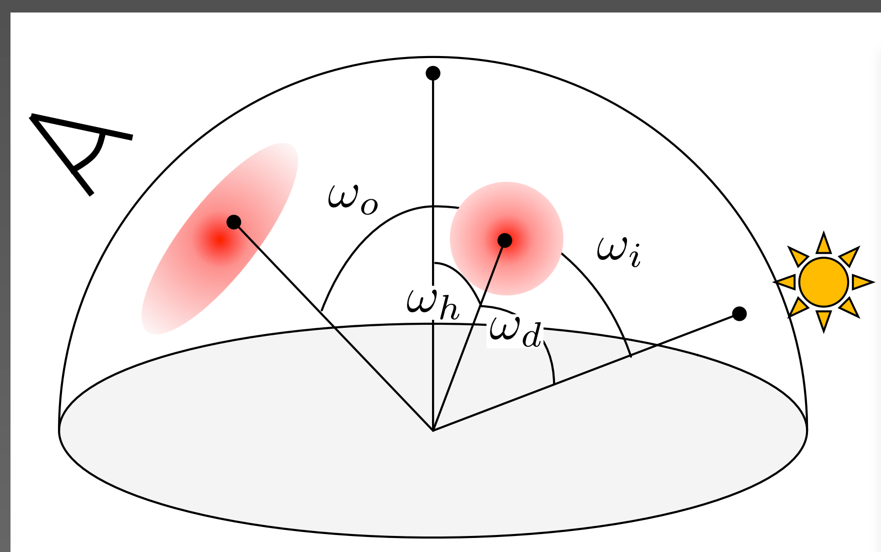
Enable exploitation of the intrinsic structure of the space of materials

BRDF As A Directional Statistics Dist.

- BRDF $$f_r(\theta_o, \phi_o; \theta_i, \phi_i) = \frac{dL(\theta_o, \phi_o)}{dE(\theta_i, \phi_i)}$$
- Isotropy $$f_r(\theta_o, \theta_i, |\phi_o - \phi_i|)$$
- Half-way vector reparametrization [Rusinkiewicz 98]



$f_r(\theta_o, \phi_o | \theta_i)$



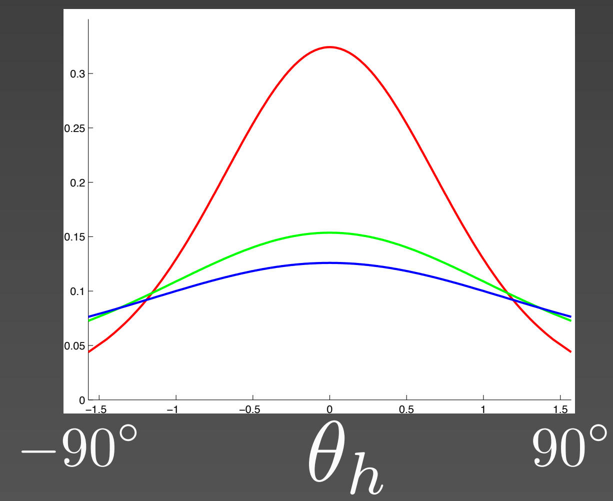
$\omega_o = (\theta_o, \phi_o), \omega_i = (\theta_i, \phi_i)$

$f_r(\theta_h, \phi_h | \theta_d)$

BRDF As A Directional Statistics Dist.

- Conventional dir. stat. dists.
  - Defined on a unit sphere
  - Von Mises-Fisher

$$\frac{\kappa}{4\pi \sinh \kappa} \exp\left[\kappa \cos \theta_h\right]$$



- Hemispherical Exponential Power Distribution

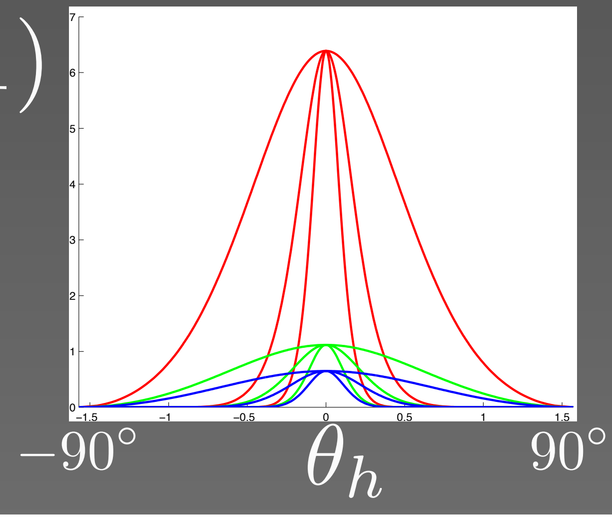
$$C(\kappa, \gamma)\left(\exp\left[\kappa \cos^{\gamma} \theta_h\right] - 1\right)$$

Scale parameter (albedo)

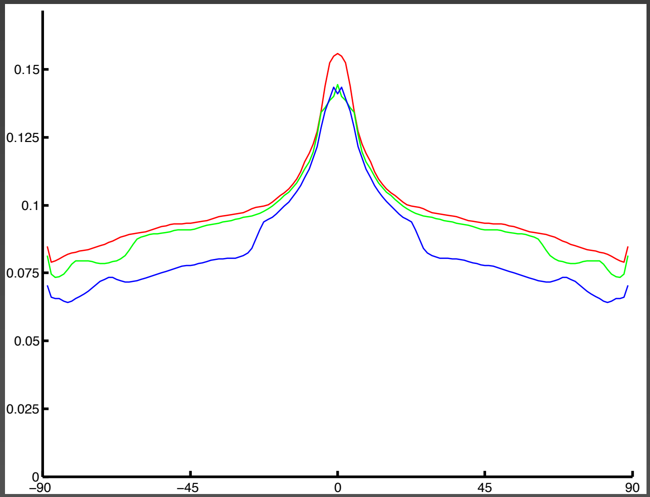
Shape parameter (kurtosis)

$\gamma \rightarrow 0$: Lambertian

$\gamma \rightarrow \infty$: Perfect mirror



Directional Statistics BRDF Model



- Real-world reflectance exhibit compound dists.
- Model with a Hemi-EPD Mixture Model

$$f_r(\theta_h, \phi_h | \theta_d) = \sum_{k=1}^{K} \exp\left[\kappa^{(k)} \cos^{\gamma^{(k)}} \theta_h\right] - 1$$

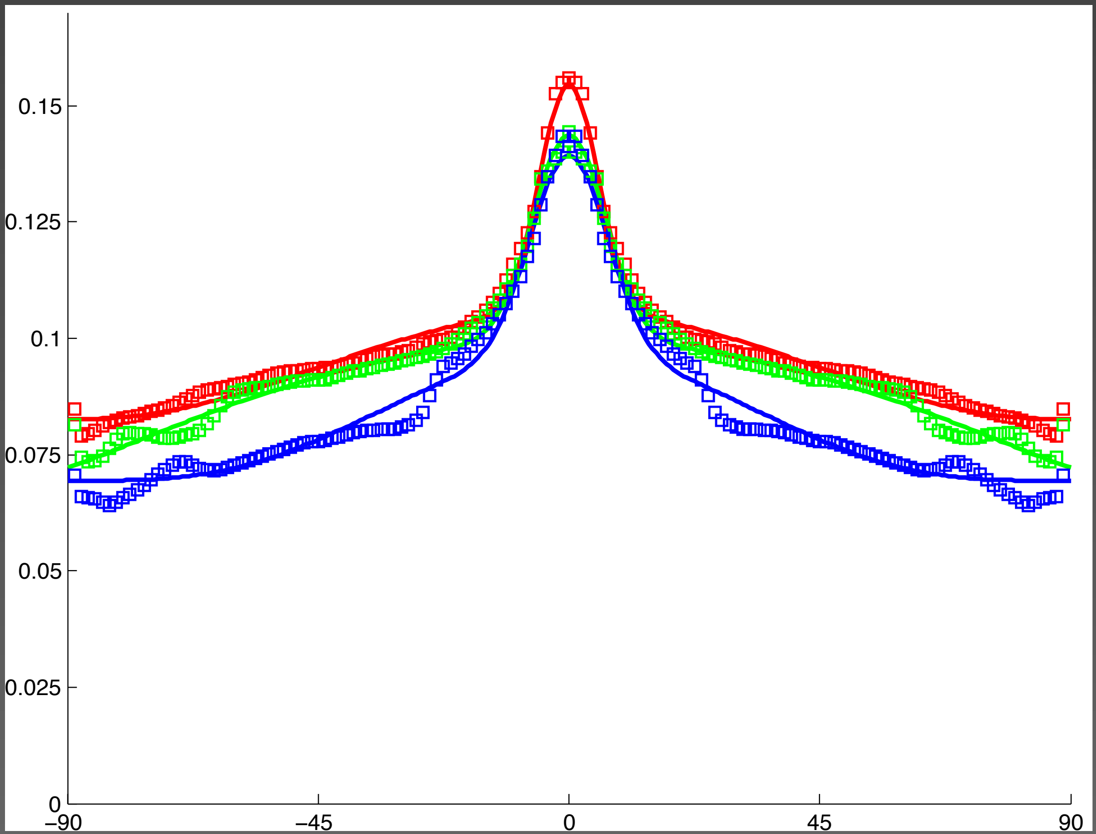
  - Unnormalized to model measured data

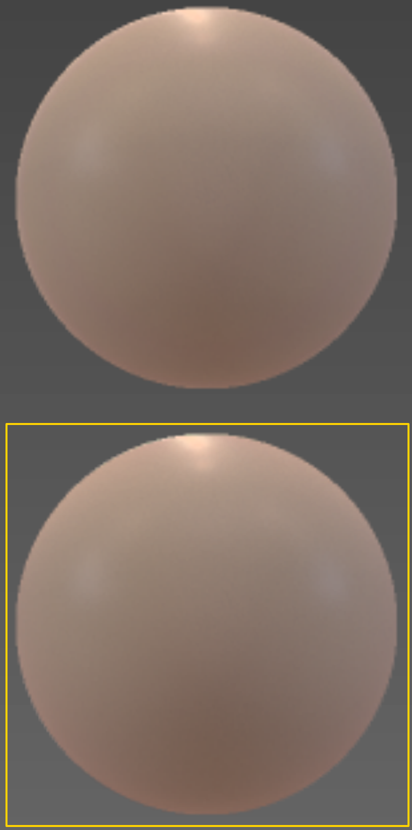
Fitting DSBRDF

$$f_r(\theta_h, \phi_h | \theta_d) = \sum_{k=1}^{K} \exp\left[\kappa^{(k)} \cos^{\gamma^{(k)}} \theta_h\right] - 1$$

- Canonical est. algorithm based on EM
  - E-step: Estimate responsibilities (relative mixture weights)
  - M-step: Maximize likelihood of each lobe
- Determination of the # of lobes
  - Student's *t*-test

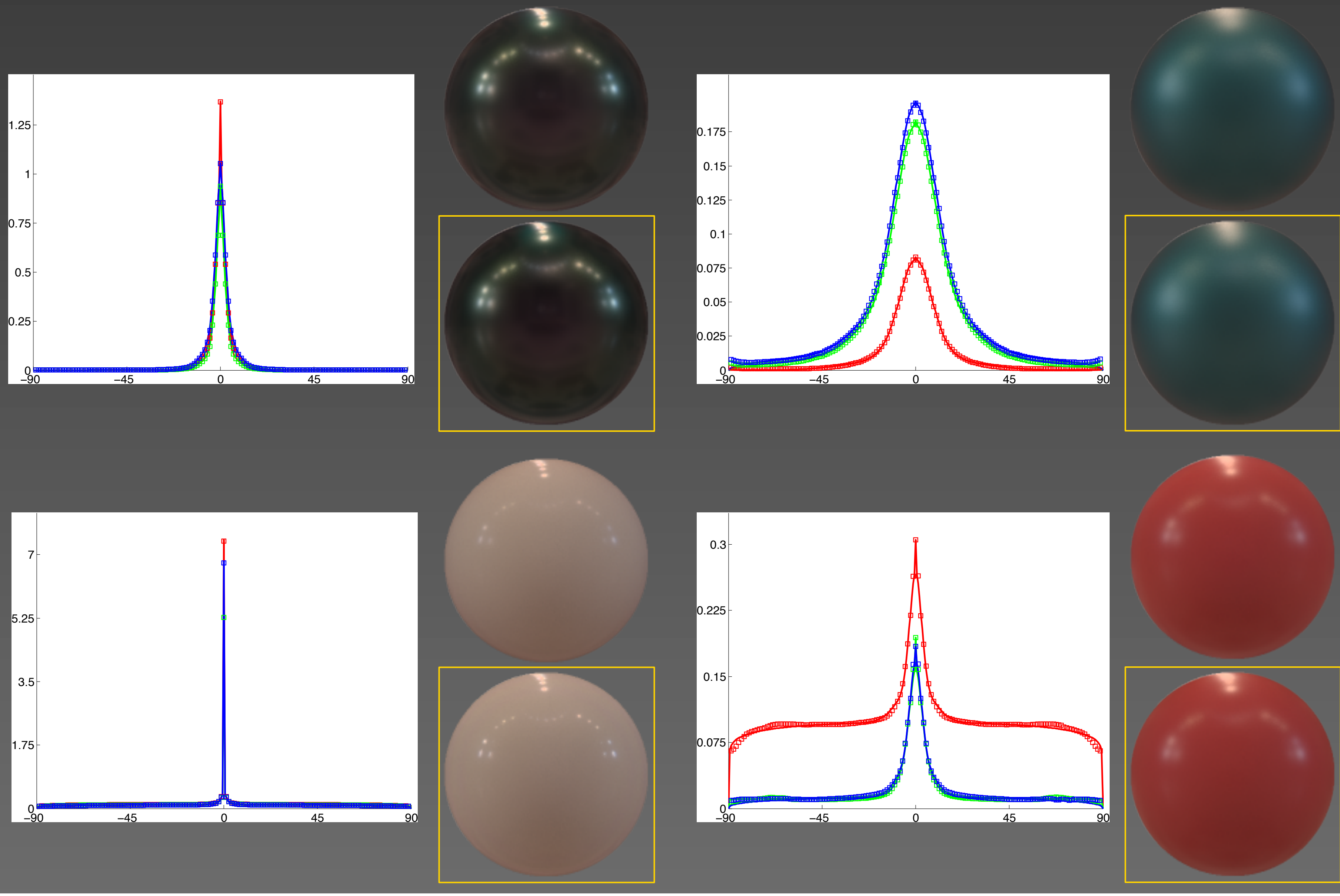
Modeling Real-World Isotropic BRDFs



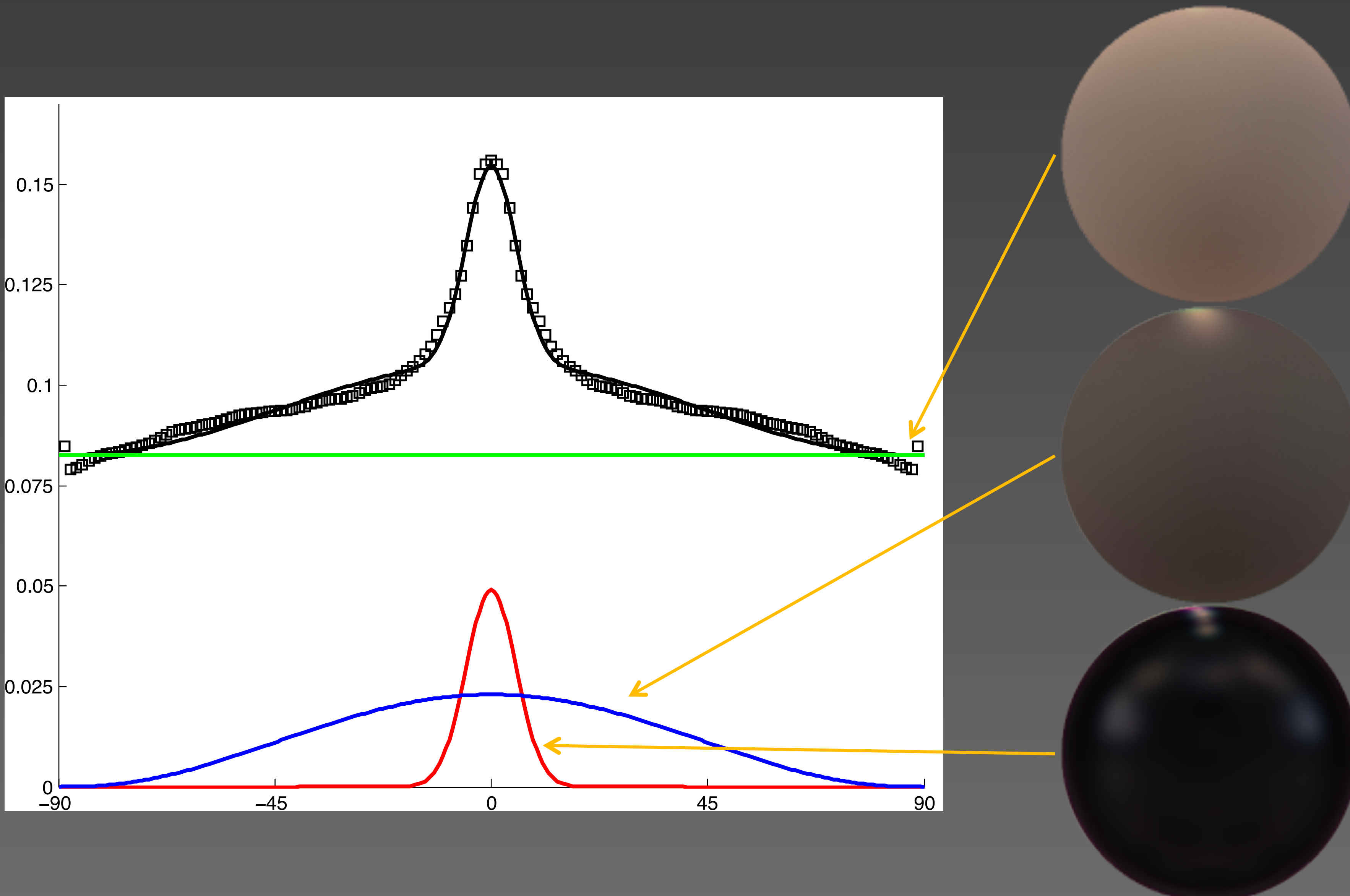


-90° $p(\theta_h, \phi_h = 0 | \theta_d = 0)$ 90°

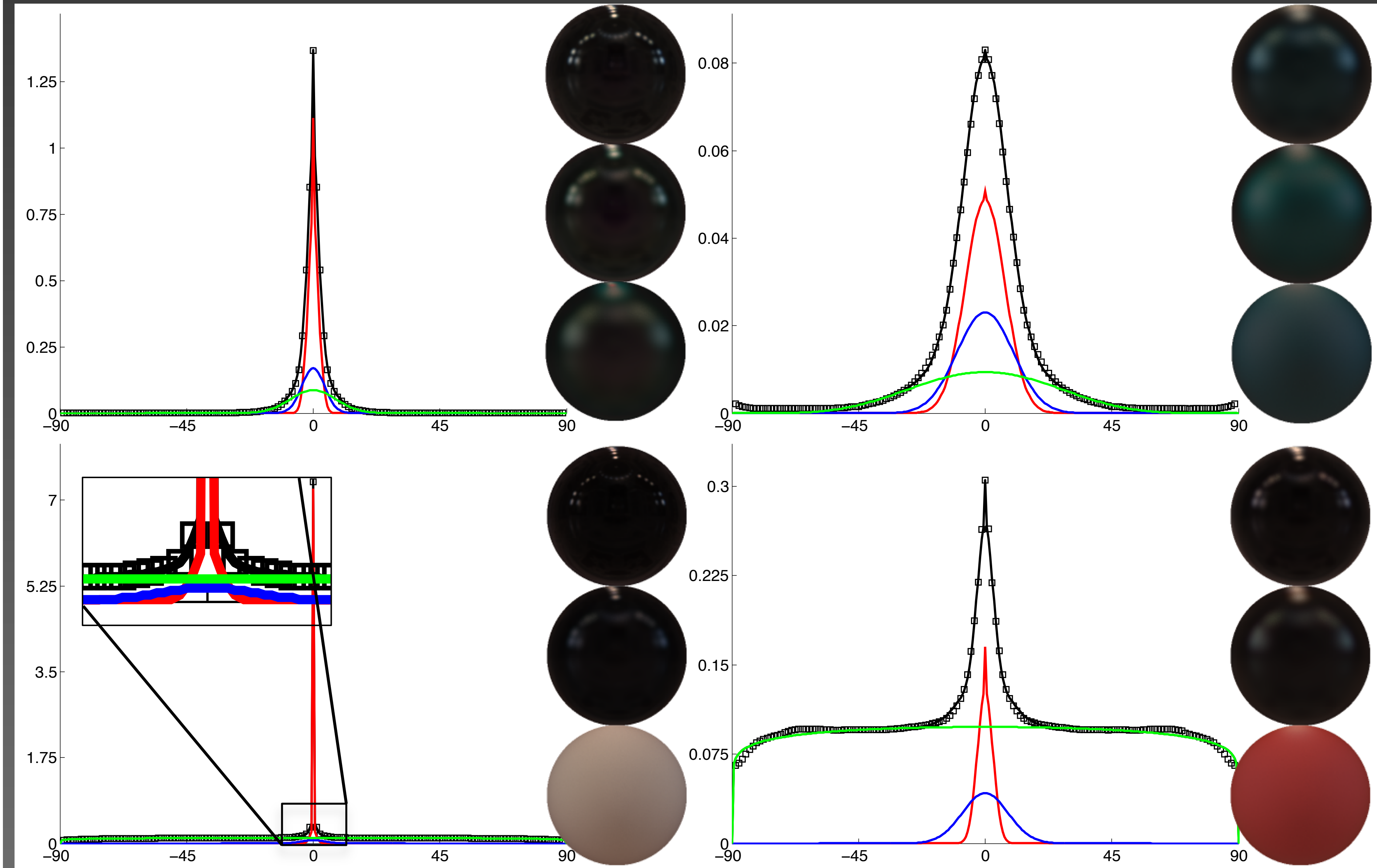
Modeling Real-World Isotropic BRDFs



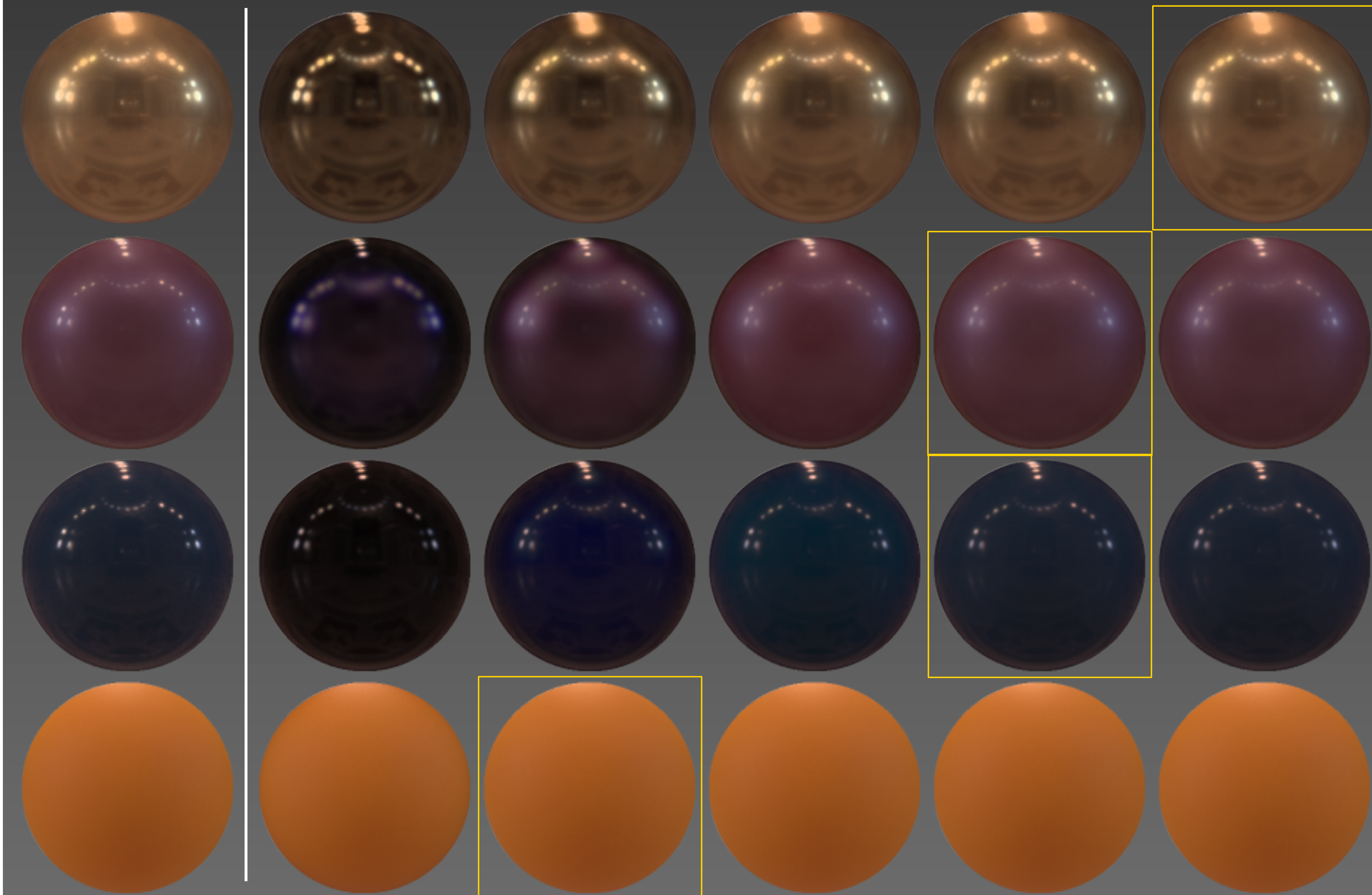
Reflectance Lobe Decomposition



Reflectance Lobe Decomposition

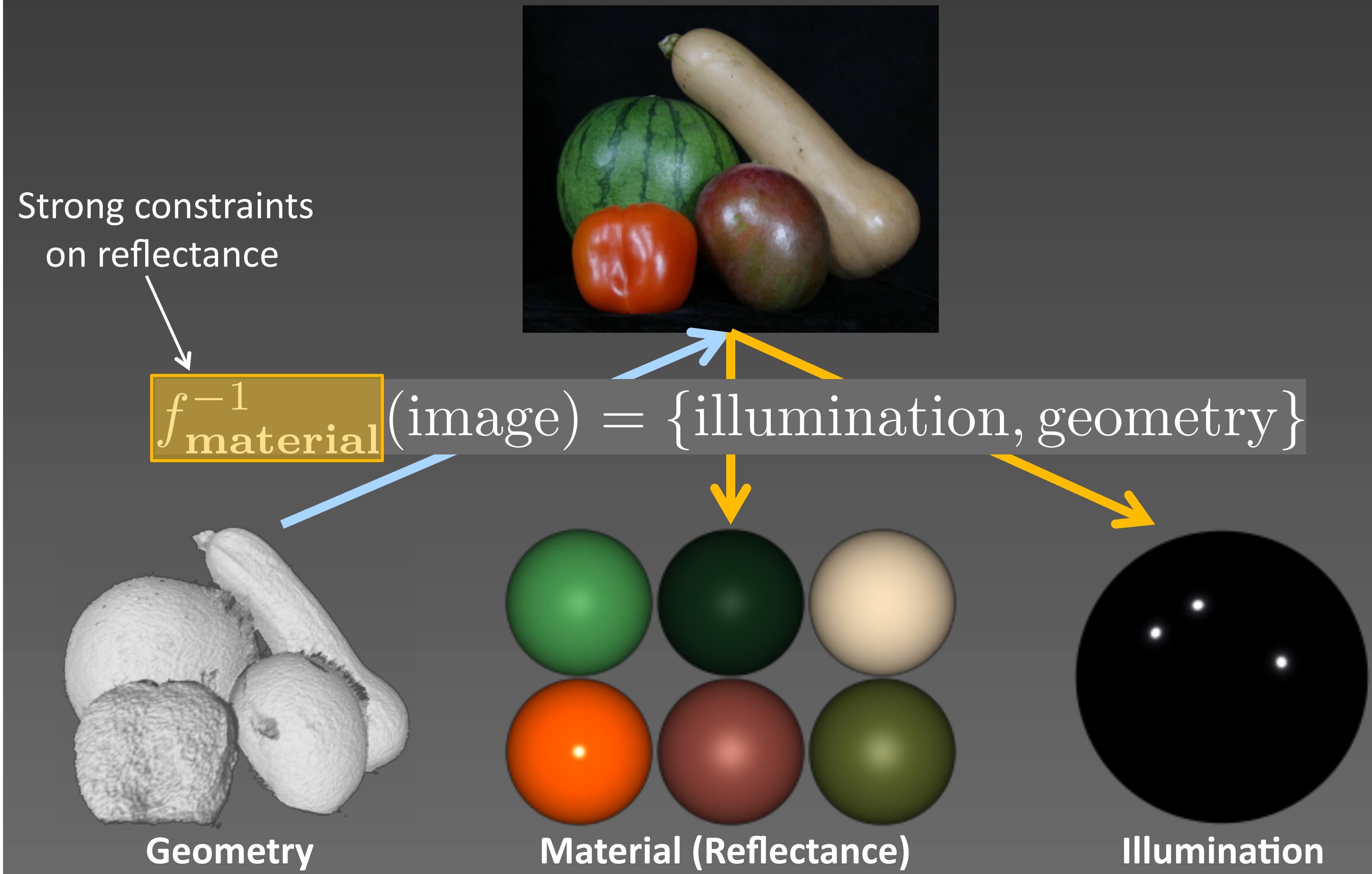


Optimal Number of Lobes

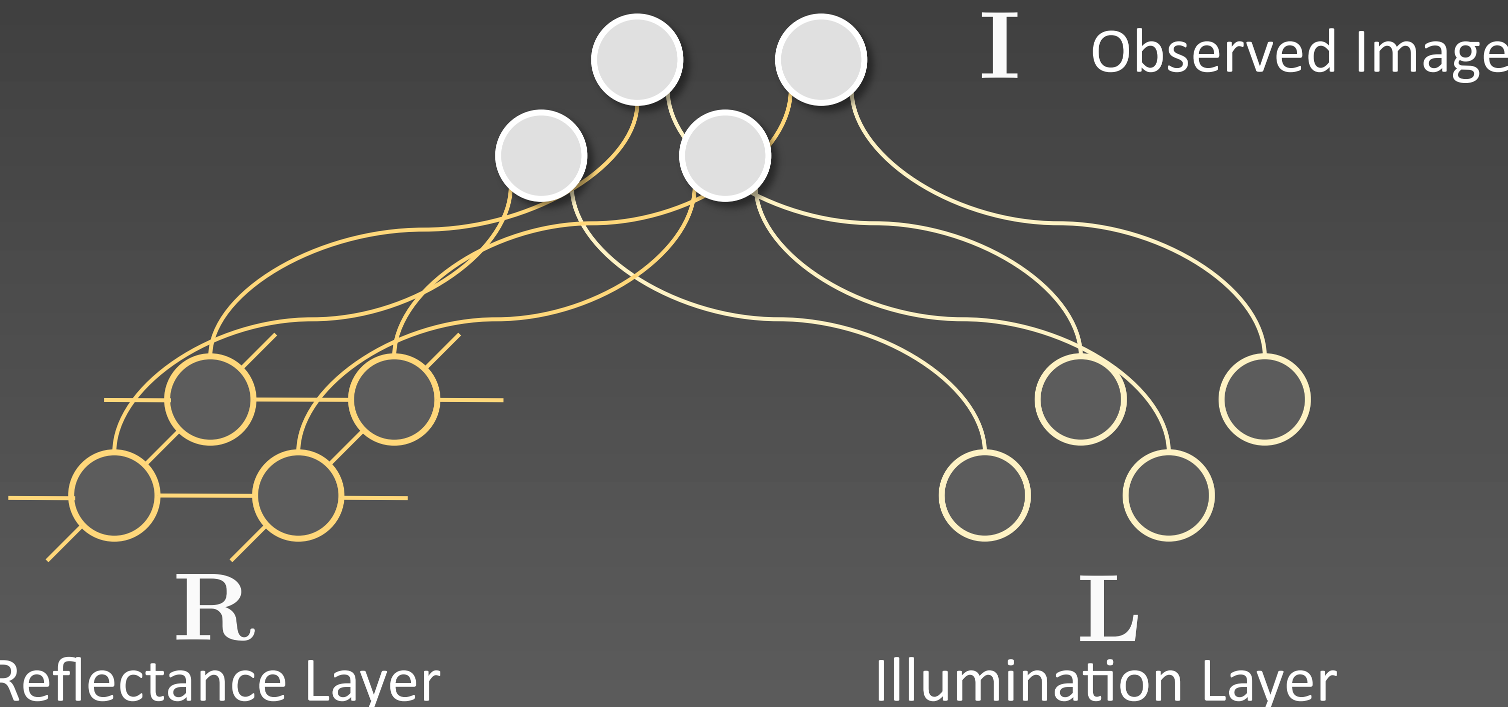


The Space of Isotropic BRDFs

Joint Estimation of Reflectance and Illumination



A Probabilistic Formulation

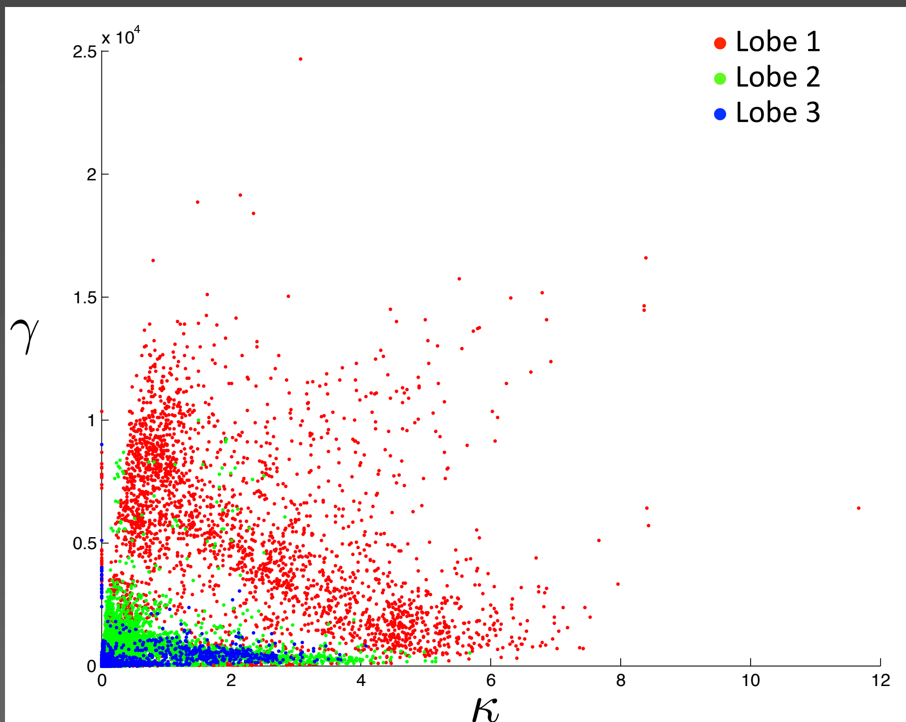


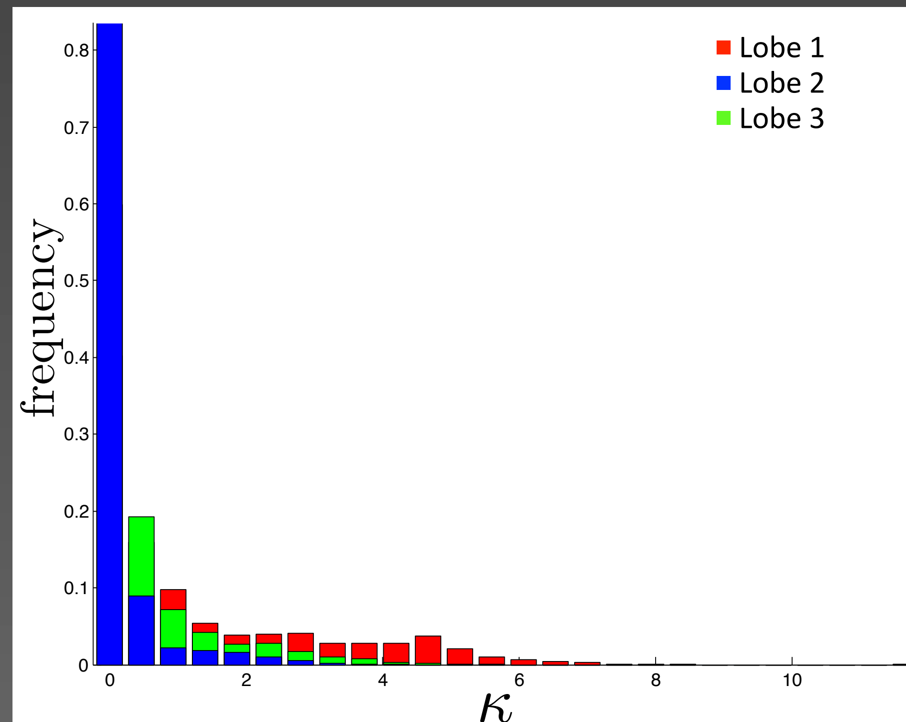
Factorial Markov random field

$$\underset{\mathbf{R},\mathbf{L}}{\operatorname{argmax}}\, p(\mathbf{R},\mathbf{L}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{R},\mathbf{L})\,p(\mathbf{R})\,p(\mathbf{L})$$

Unary Reflectance Prior

$$p(\mathbf{R}) = \prod p(\mathbf{r_x}) \prod p(\mathbf{r_x}, \mathbf{r}_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})})$$





$$p(\mathbf{r_x}) = p(\kappa_\mathbf{x}, \gamma_\mathbf{x})$$

$$p(\mathbf{r_x}) = p(\kappa_\mathbf{x}) p(\gamma_\mathbf{x})$$

Clique Reflectance Prior

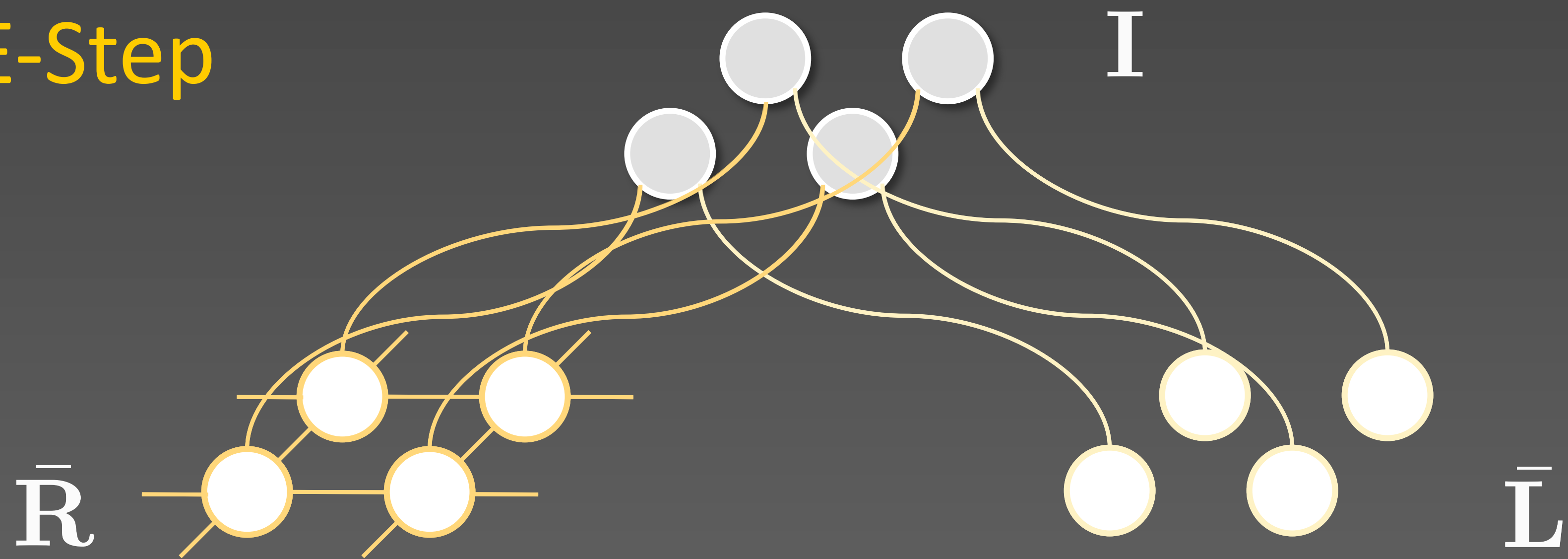
$$p(\mathbf{R}) = \prod p(\mathbf{r}_{\mathbf{x}}) \prod p(\mathbf{r}_{\mathbf{x}}, \mathbf{r}_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})})$$

- Conventional priors on DSBRDF parameters
  - Gaussian: smooth $\exp\left[\gamma_{\mathbf{x}} - \gamma_{\mathbf{x}'}\right]^2$
  - L1: piecewise smooth $\exp\left|\gamma_{\mathbf{x}} - \gamma_{\mathbf{x}'}\right|$
  - Potts model: piecewise constant $\delta\left(\gamma_{\mathbf{x}} - \gamma_{\mathbf{x}'}\right)$
- Separate prior on each reflectance lobe
  - E.g., Potts on 1st lobe and L1 for others

Probabilistic Factorization of Reflectance and Illumination

$$\operatorname*{argmax}_{\mathbf{R},\mathbf{L}} p(\mathbf{R},\mathbf{L}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{R},\mathbf{L})p(\mathbf{R})p(\mathbf{L})$$

E-Step



$$\operatorname*{argmax}_{\mathbf{R}} p(\mathbf{I}|\mathbf{R},\bar{\mathbf{L}})p(\mathbf{R})$$

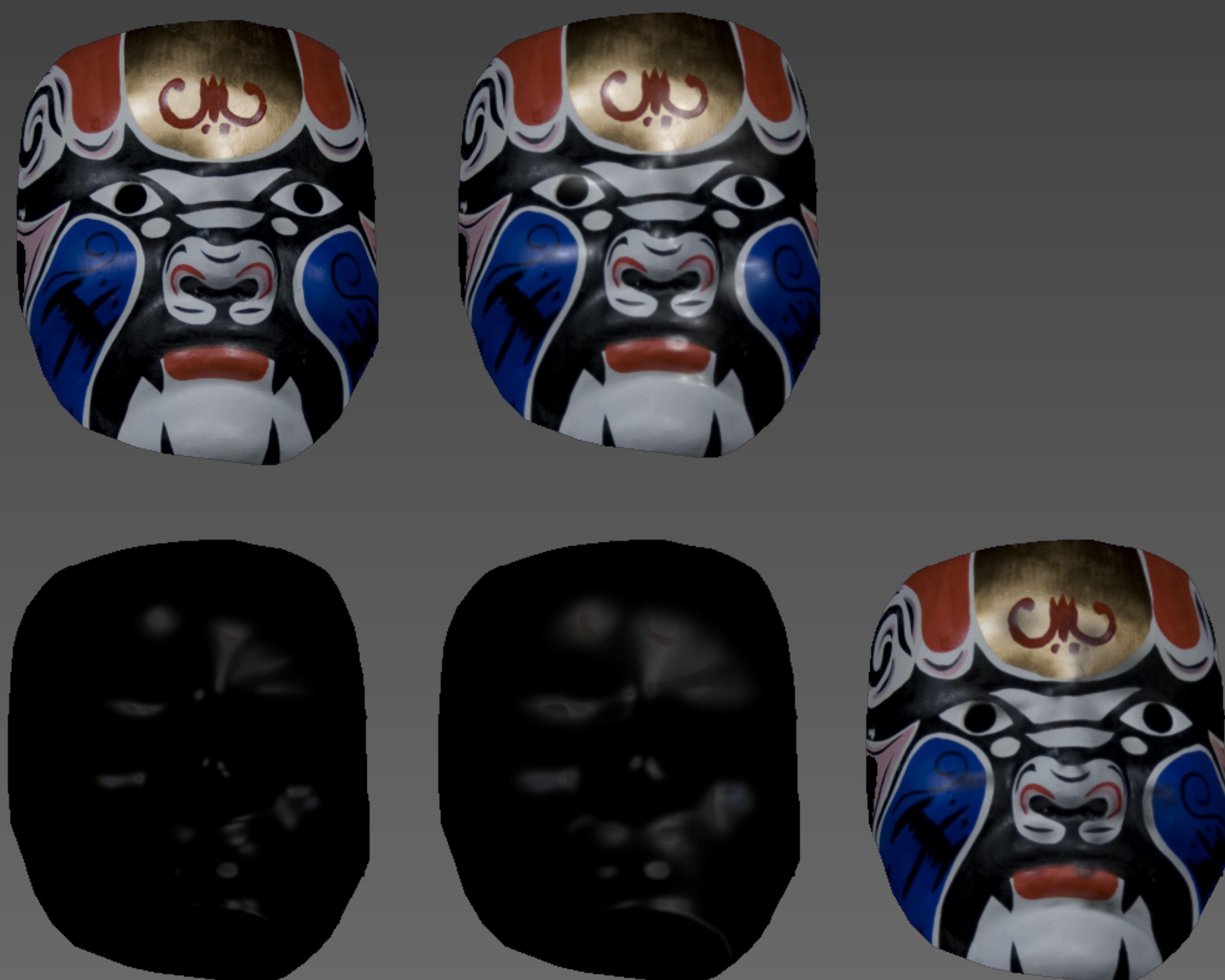
$$\operatorname*{argmax}_{\mathbf{L}} p(\mathbf{I}|\mathbf{L},\bar{\mathbf{R}})p(\mathbf{L})$$

M-Step

$$\operatorname*{argmax}_{\Sigma} p(\mathbf{I}|\bar{\mathbf{R}},\bar{\mathbf{L}})$$

c.f. [Kim & Zabih 02]

(Preliminary) Decomposition Results



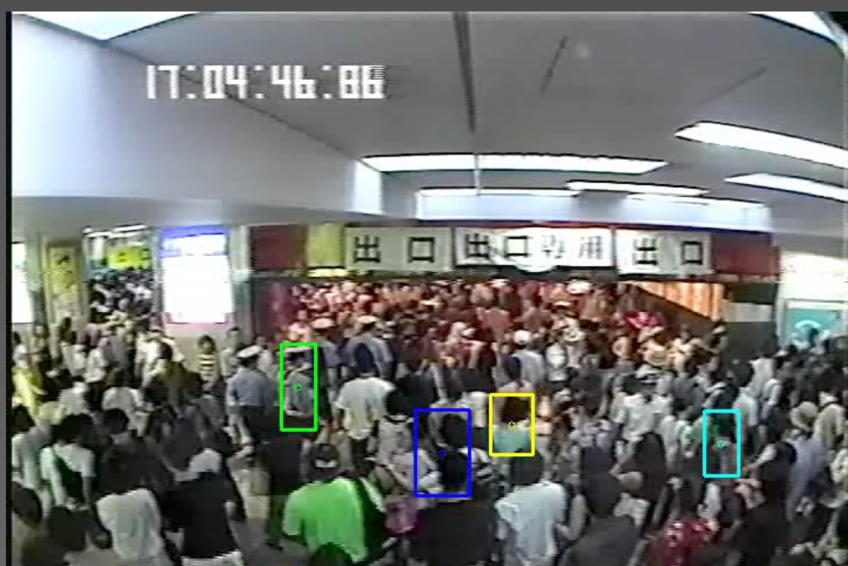
Summary

- Exploit latent structures of visual data!
  - Enables new applications
  - Provides new insights/approaches to long-standing problems
- The Scale Variability in Geometry
  - Geometric scale-space
  - Scale-dependent/invariant features and descriptors
  - Fully automatic multi-object registration
- The Space of Reflectance in Appearance
  - Directional statistics BRDF model
  - The space of reflectance and its stat. characterization
  - Probabilistic factorization of reflectance and illumination

Other Latent Structures

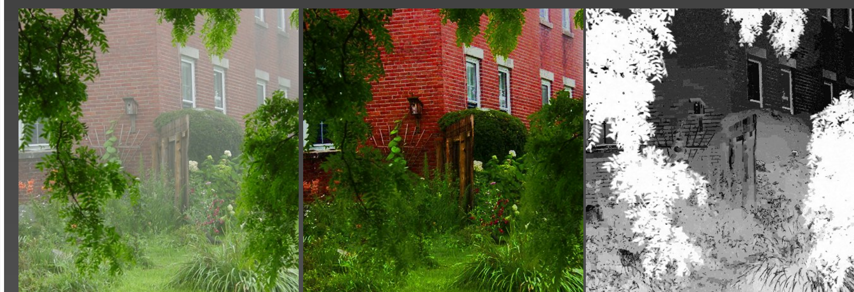
In the Video

Anomaly Detection in Crowds



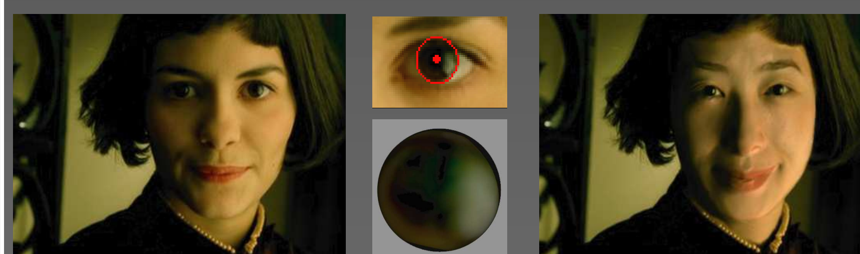
Tracking in Crowds

In a Single Image



Defogging

Membrane Nonrigid Registration



Eye

Acknowledgements

- Students/Collaborators
  - Louis Kratz, Prabin Bariya, Geoffrey Oxholm, Stephen Lombardi, and Ian Johnston @ Drexel
  - Kenji Hara @ Kyushu
- Support
  - National Science Foundation
  - Nippon Telephone and Telegraph

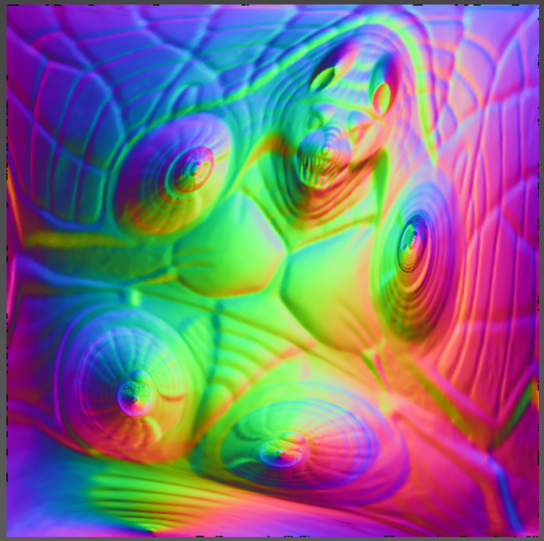
http://www.cs.drexel.edu/~kon/

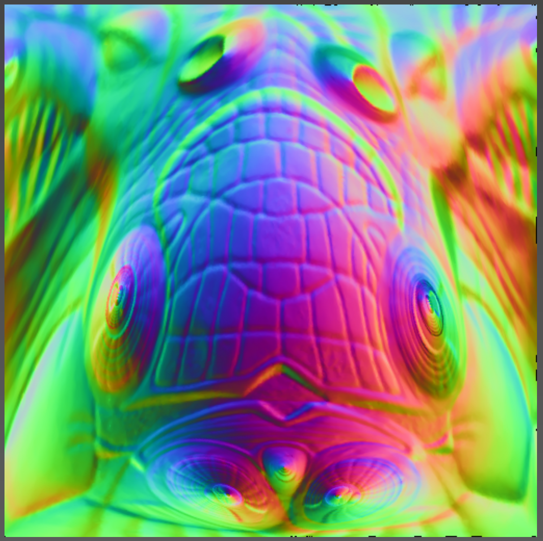
Structures Embedded in Visual Data

- We live in a structured world
  - Visual data are projections of those structures
  - Manifest beyond what is visible to the naked eyes
  - Not just those of the images
    - Natural image statistics, geometric context, etc.
- Exploit the structure in some way
  - Novel approaches to long-standing problems
  - Novel applications of visual data

What structure!?

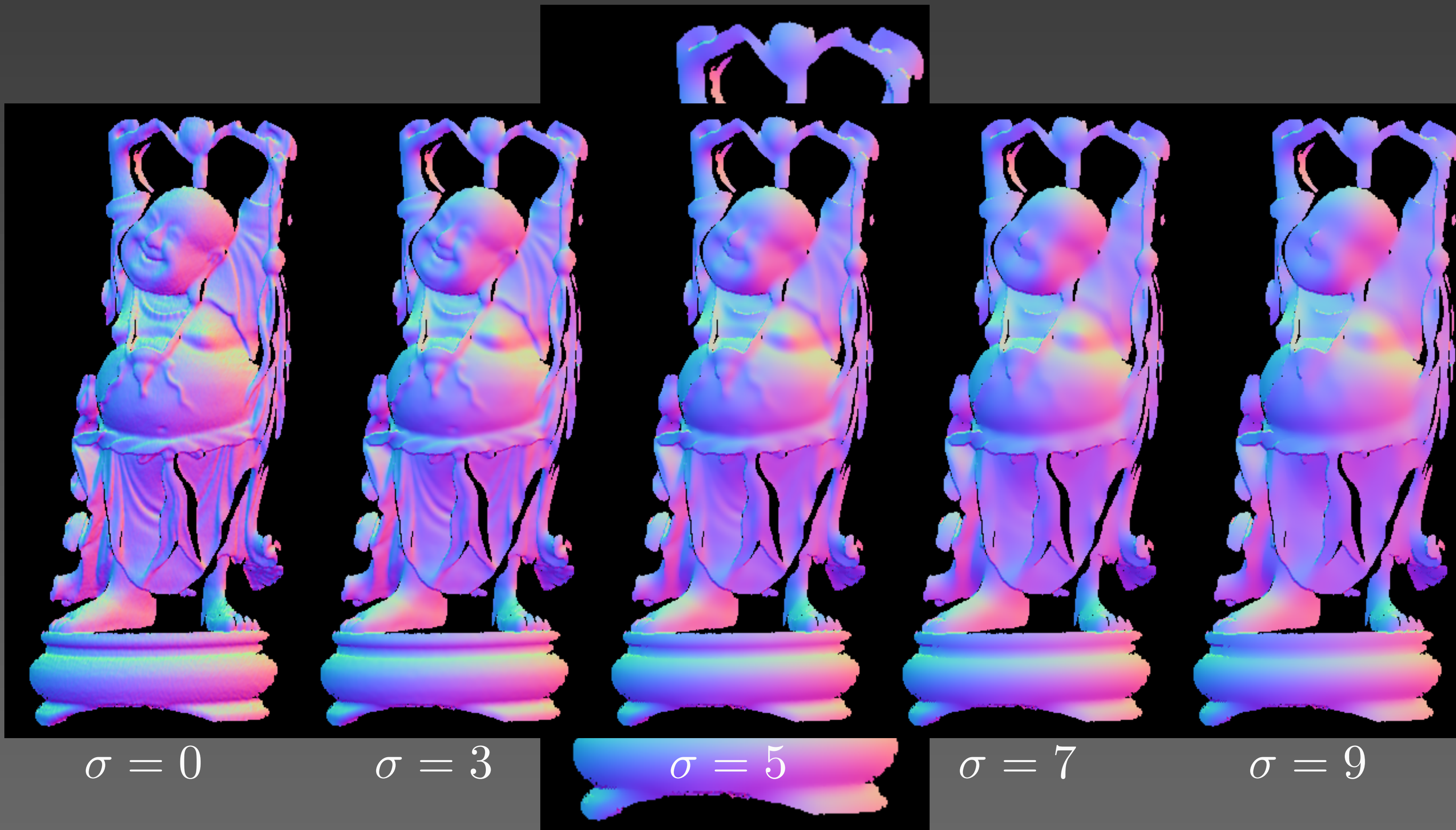
Beyond Disc Topology





- Cut and embed
  - Cut through featureless regions
  - Use complementary cuts to account for seam

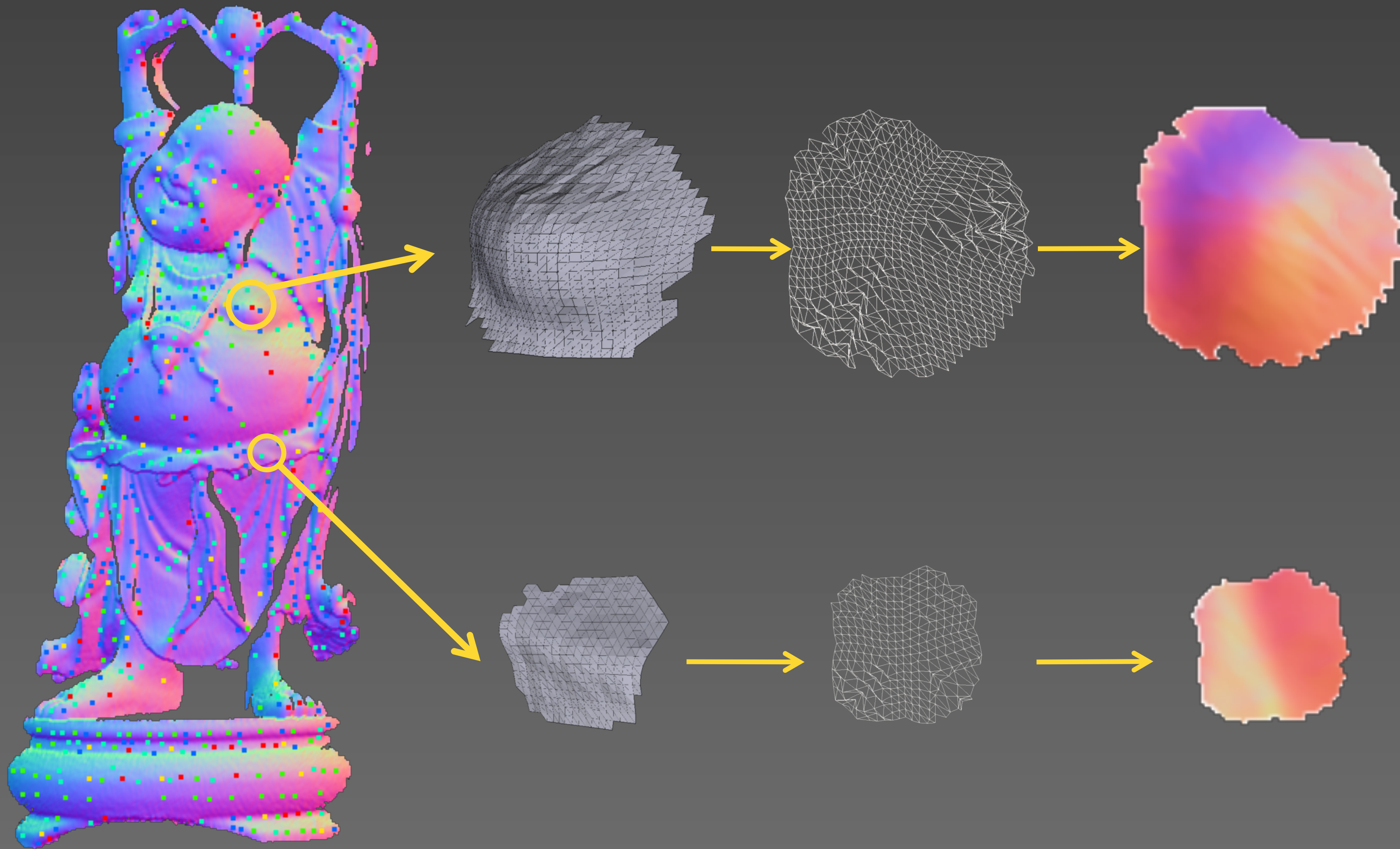
Geometric Scale-Space



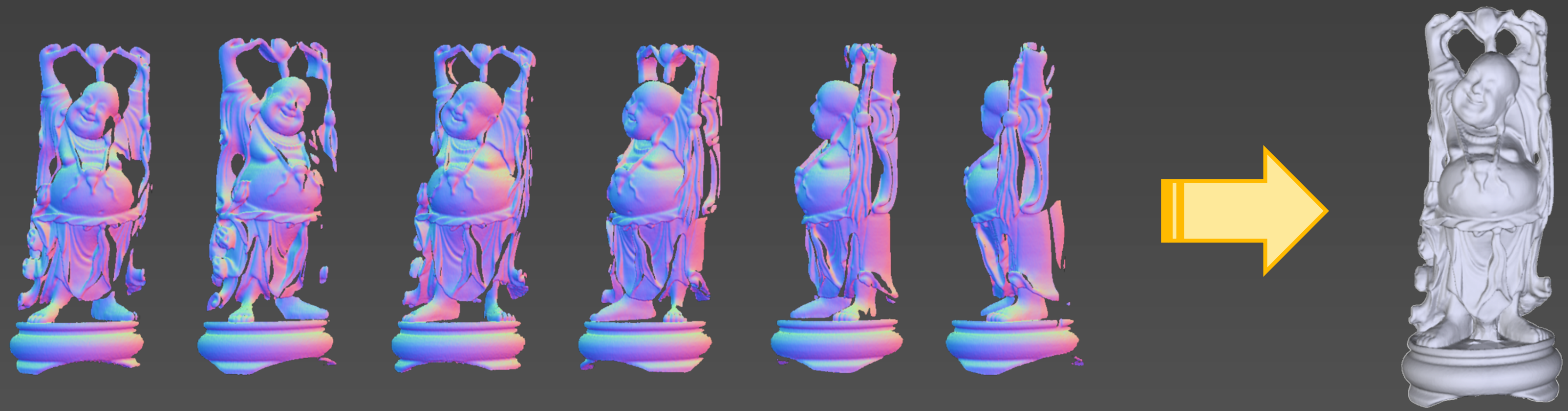
Features: Edges

- Zero-crossings of Laplacian
  - Prune edges on flat/slow regions using gradient magnitudes

Scale-Dependent Local Shape Descriptors



Fully-Automatic Multi-View Registration

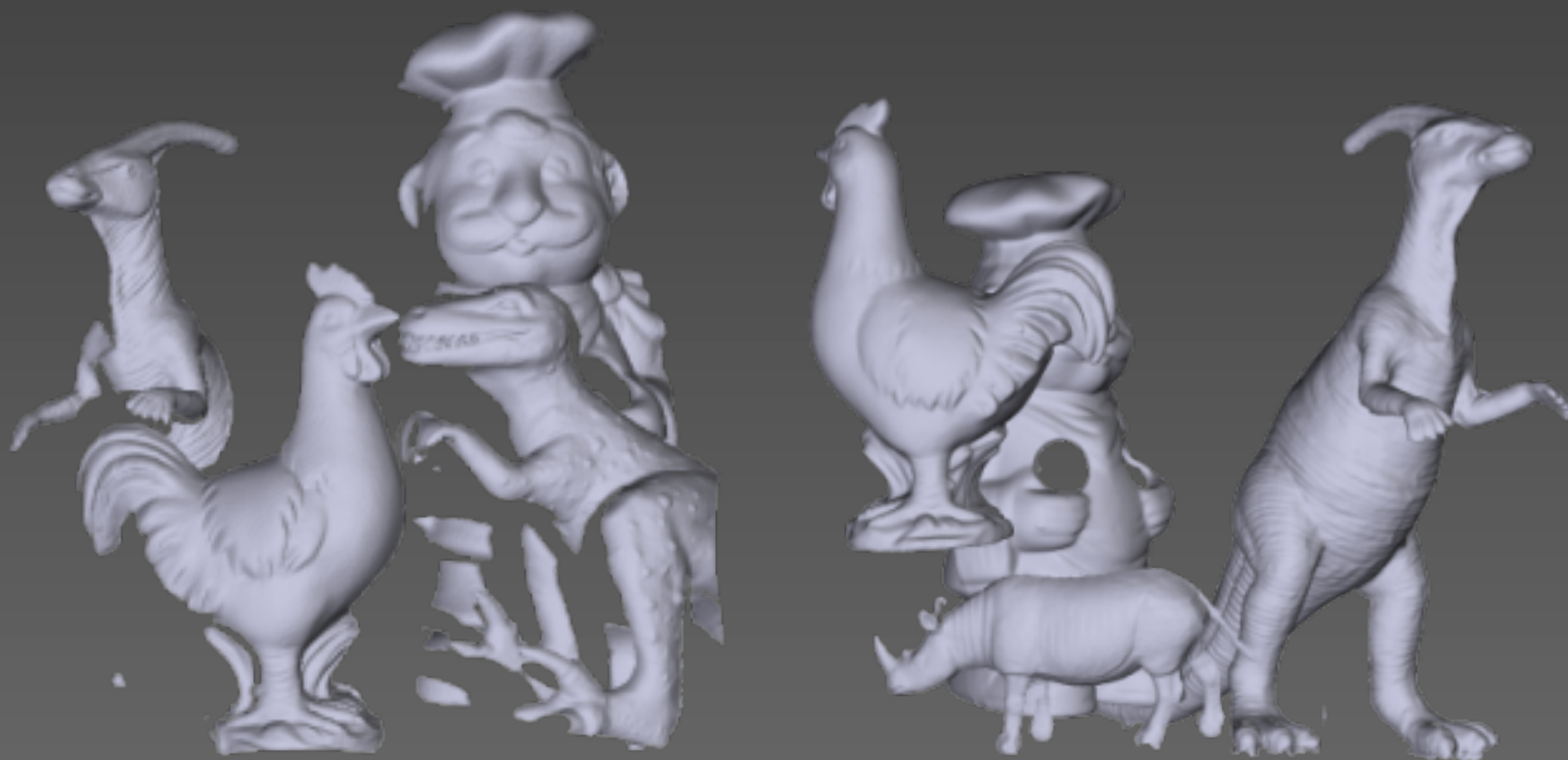


1. Scale-hierarchical pairwise alignments
2. MST on the resulting graph (cf. [Huber and Hebert 03])
3. Robust Multiview ICP to refine [Nishino et al. 02]
4. Possion surface reconstruction [Kazhdan et al. 06]

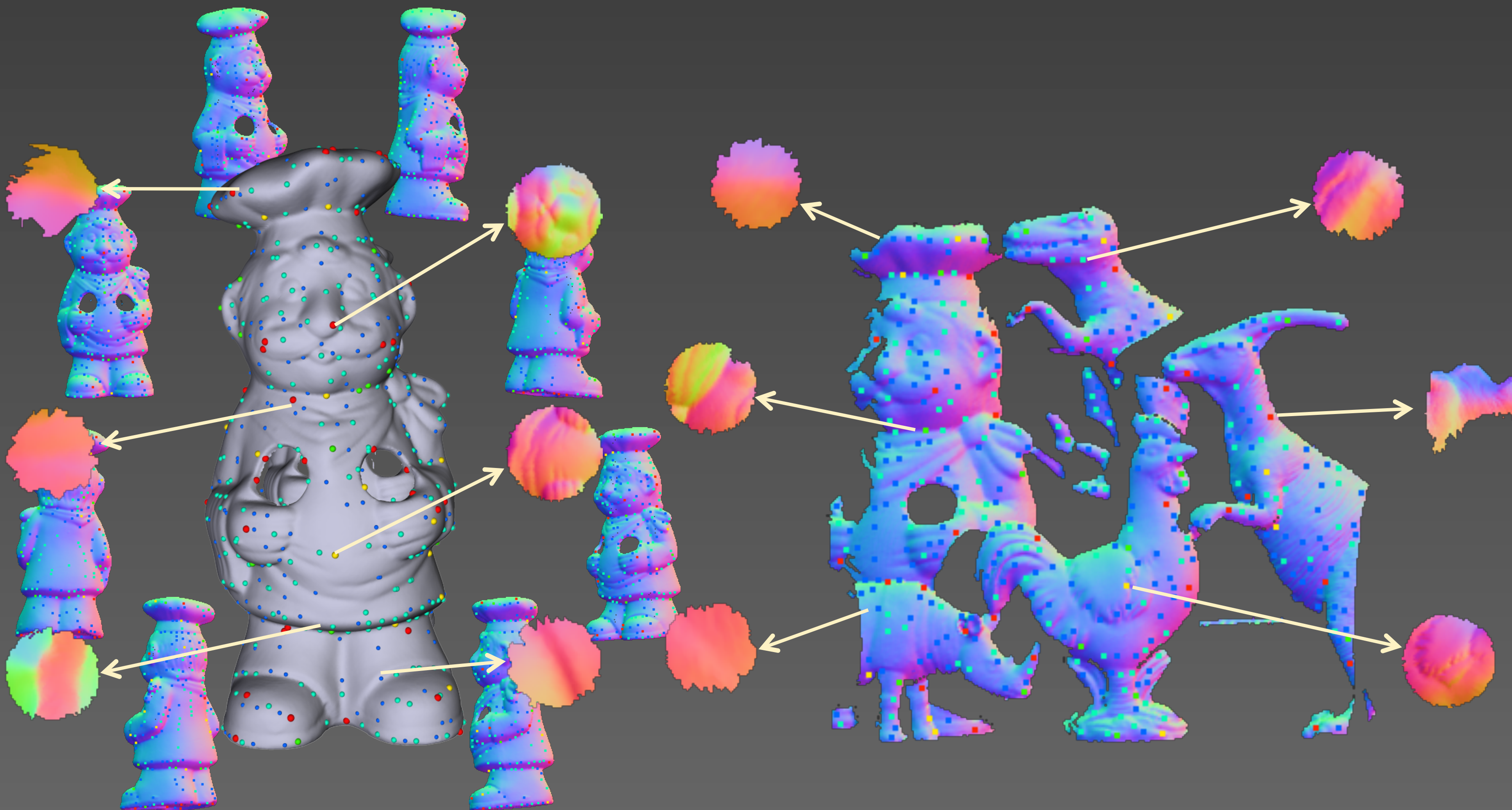
Scale-Dependent Matching

- Exploit the hierarchy induced by scale
  - Match from coarse to fine
  - Match between the same scales
  - Normalized cross-correlation as similarity metric
  - RANSAC at each scale w/ area of overlap as error measure
  - Bootstrap by taking in all matches that agree with the current transformation estimate when moving down another scale

3D Object Recognition



Model and Scene Representation



Scale-dependent/invariant shape descriptors consolidated from different views

Scale-Hierarchical Matching

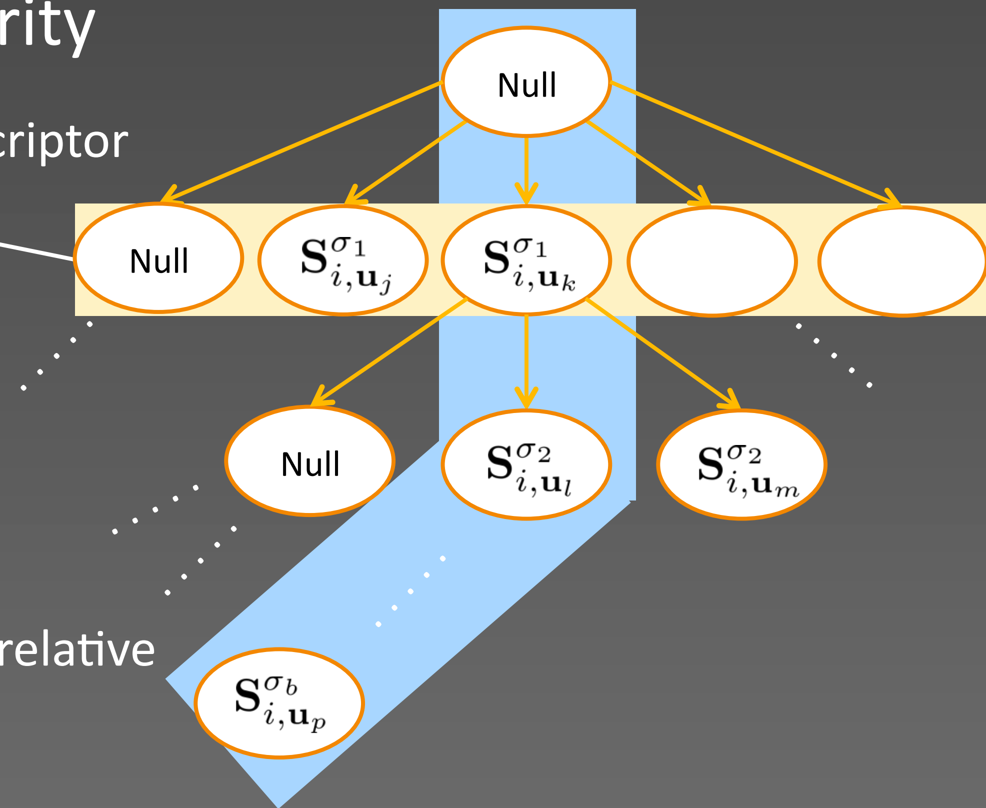
- Scale-constrained Interpretation Tree c.f. [Grimson et al.]
  - Matches restricted to same relative scale
  - Coarse-to-fine priority

Potential matches of one model descriptor

Prune based on
• Similarity
• Transformation error

Grow to pre-determined depth

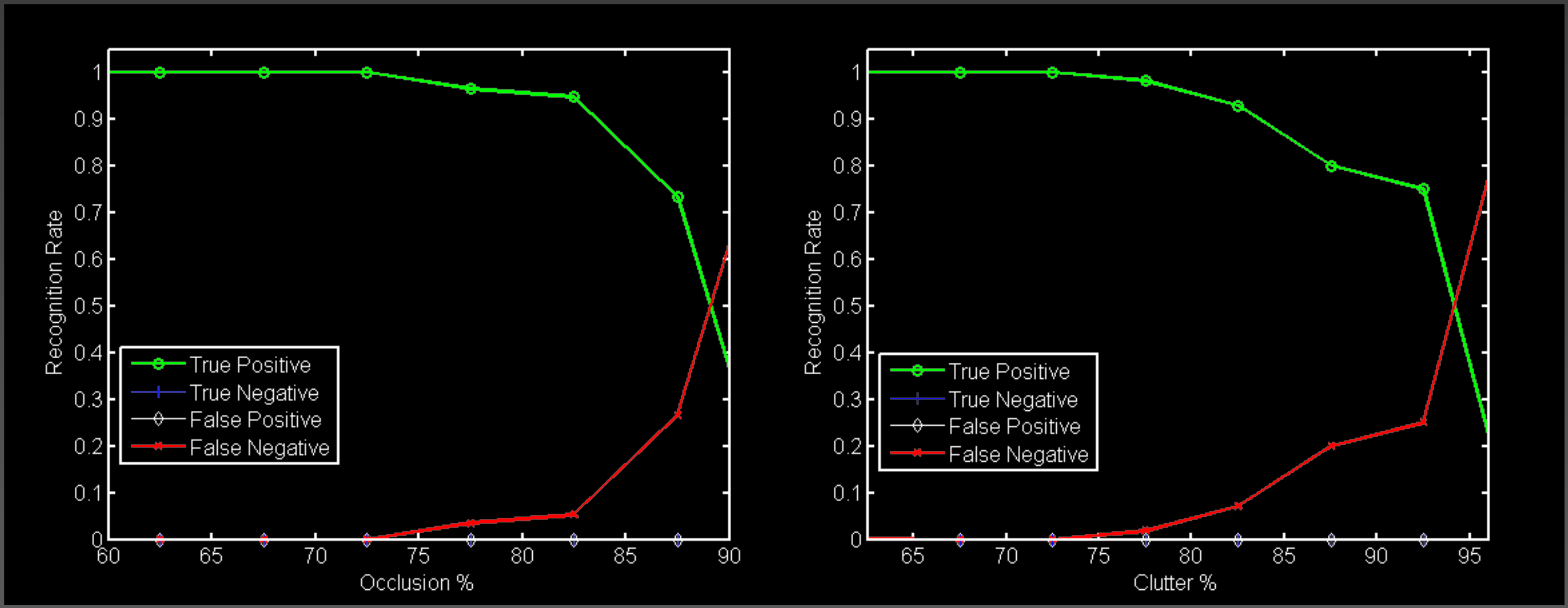
Pick the hypothesis with maximum relative overlap area



Scale-Hierarchical 3D Object Recognition

- Consistent global scales

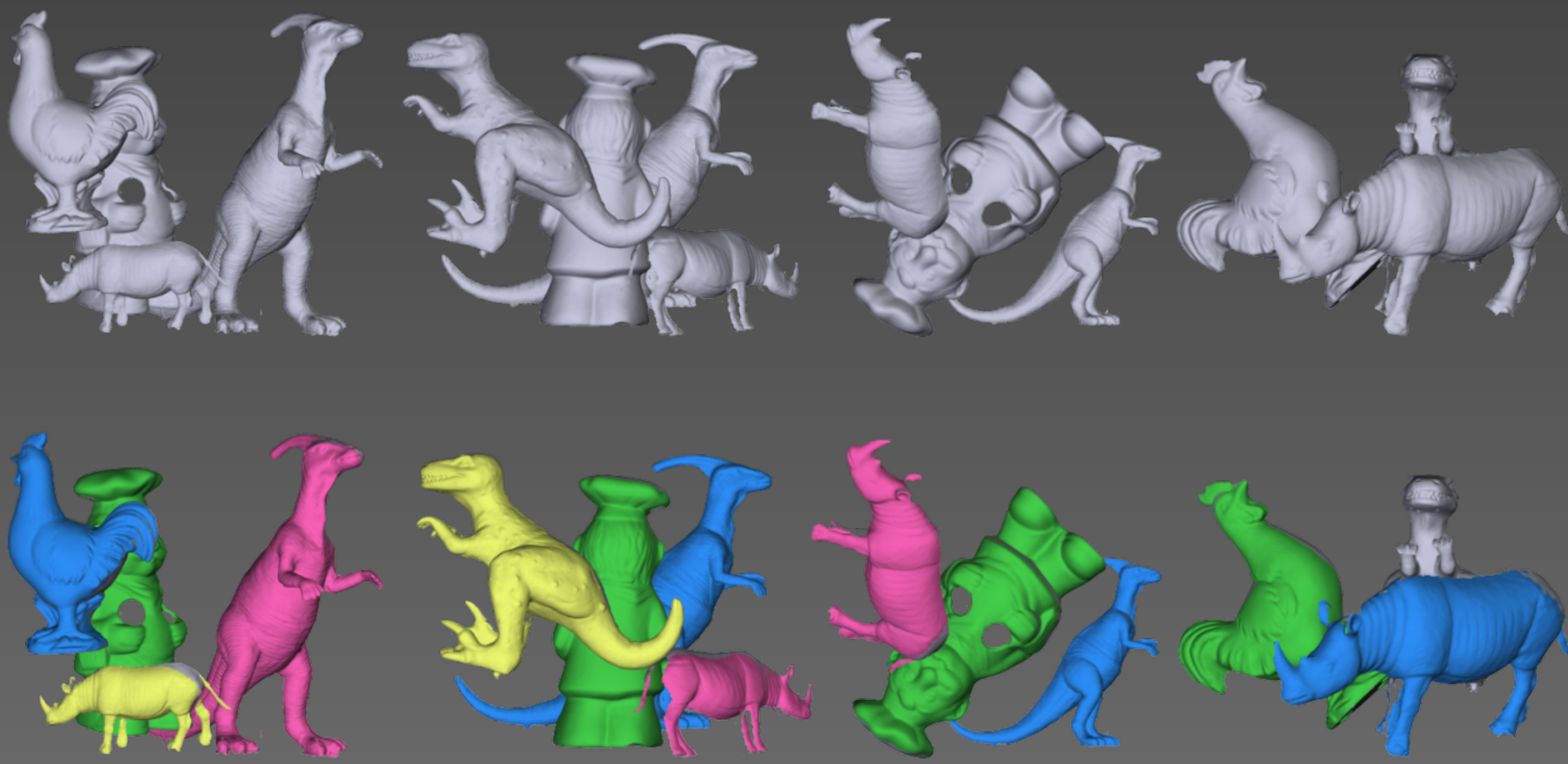
Scale-Hierarchical 3D Object Recognition



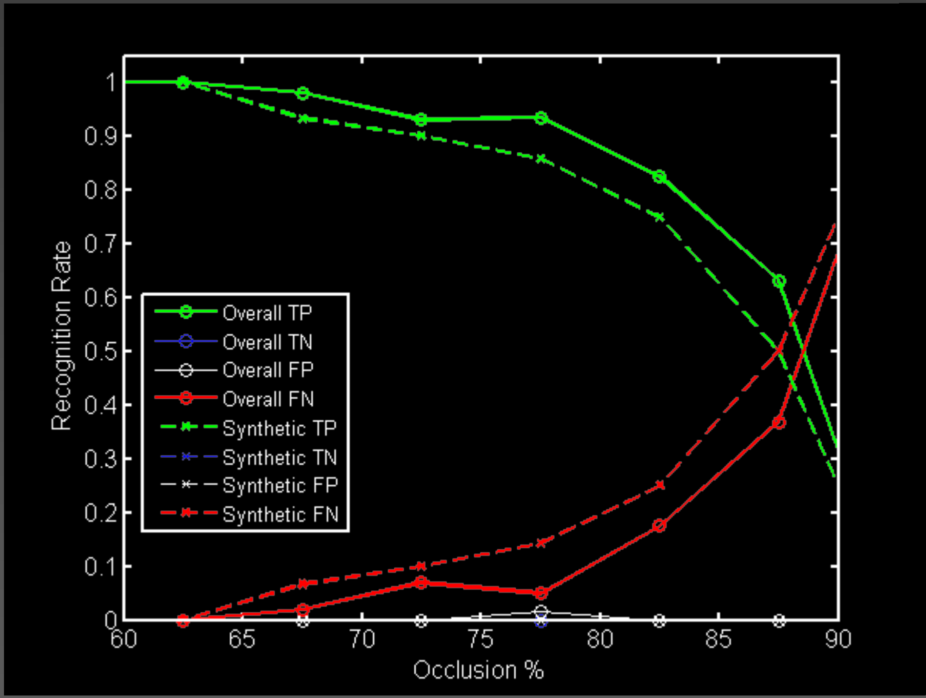
- 5 models 50 scenes with varying occlusion and clutter [Mian et al. 06]
- Over all recognition rate 93.58%
- For up to 84% occlusion
    1. Ours 97.5%
    2. Tensor Matching [Mian et al. 06] 96.6%
    3. Spin Images [Johnson and Hebert 99] 87.8%

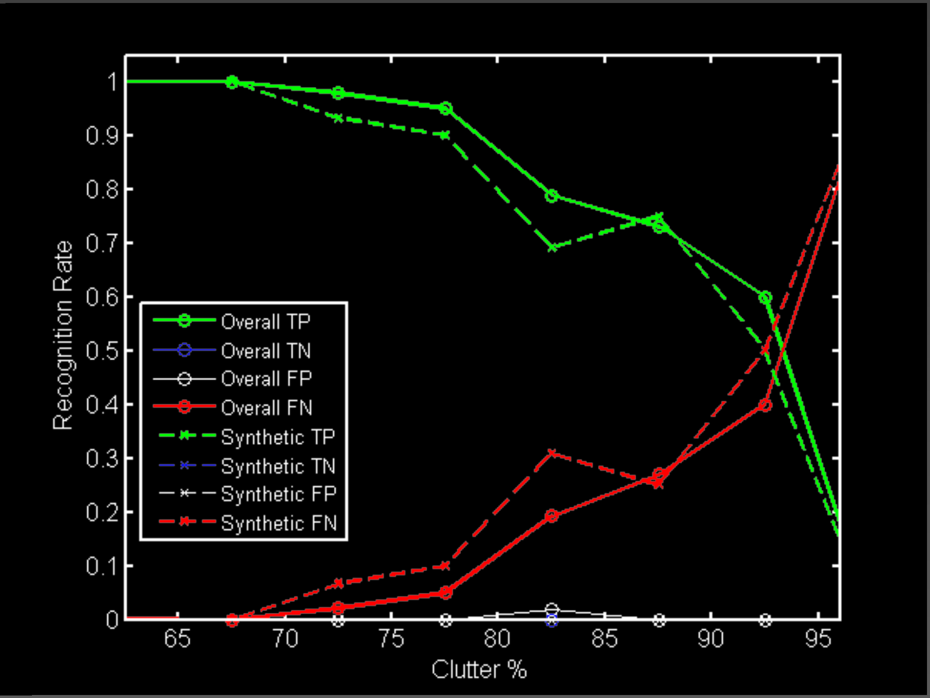
Scale-Hierarchical 3D Object Recognition

- Inconsistent global scales



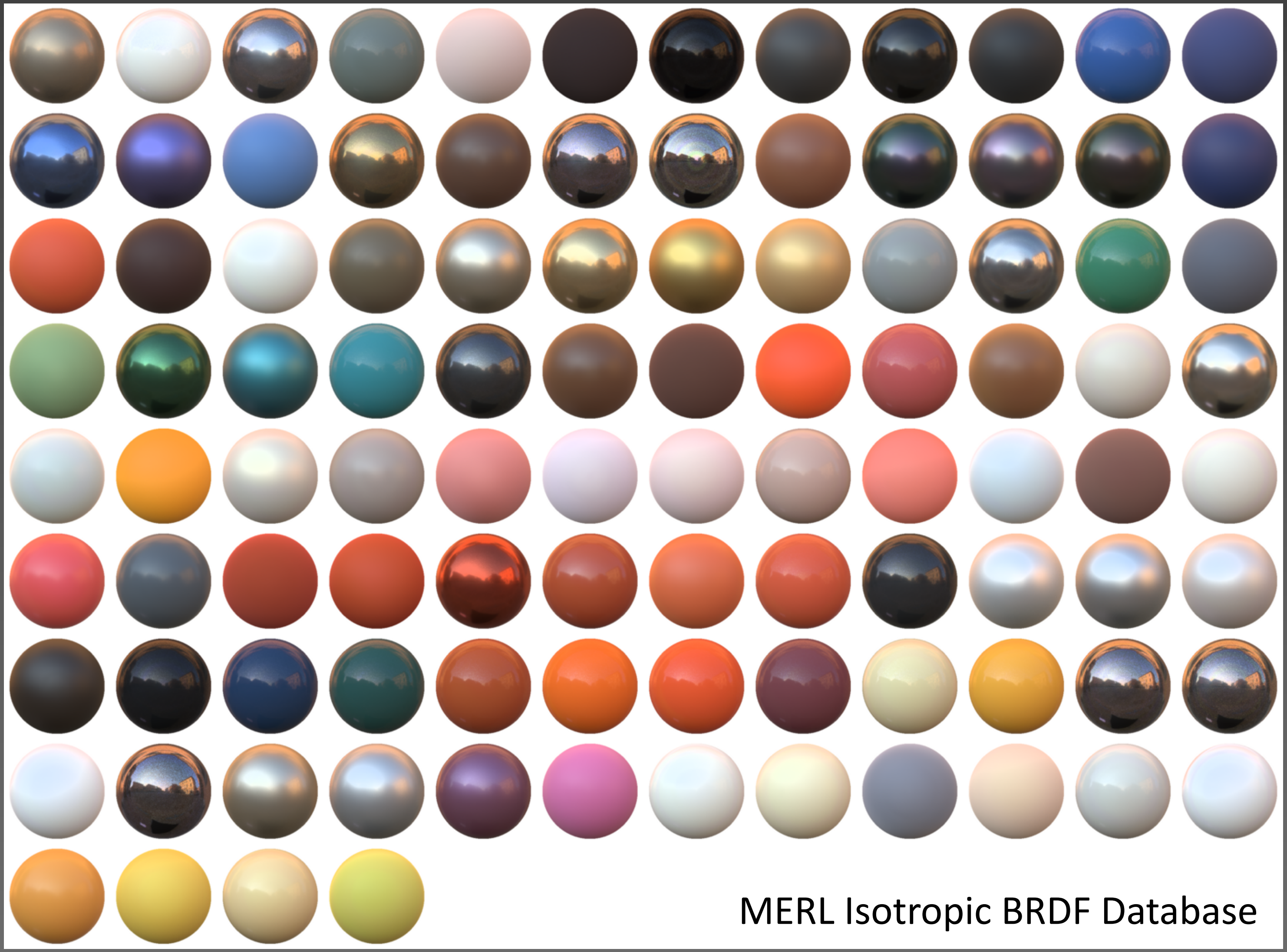
Scale-Hierarchical 3D Object Recognition



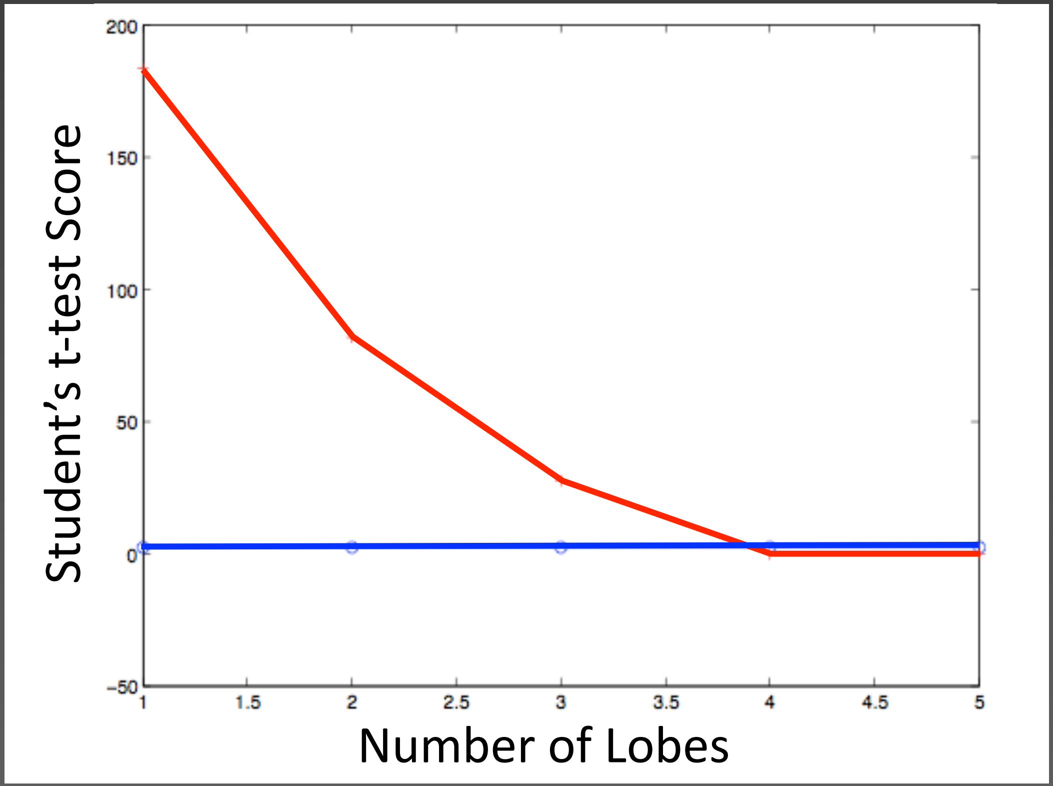


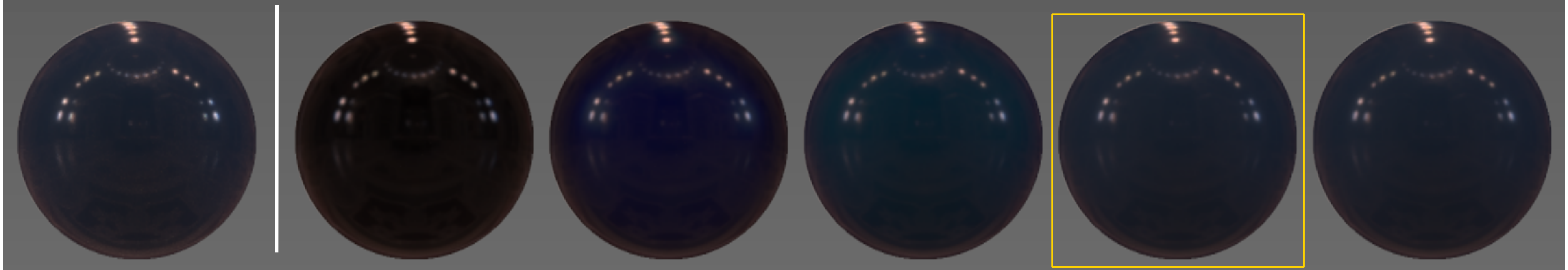
- 5 models scaled between 60-150% in 47 real +18 synthetic scenes
    - Finds similarity transformation
- Over all recognition rate 88.5%
- First systematic study of scale-invariant 3D object recognition

Real-World Reflectance is Rarely Lambertian plus Torrance-Sparrow

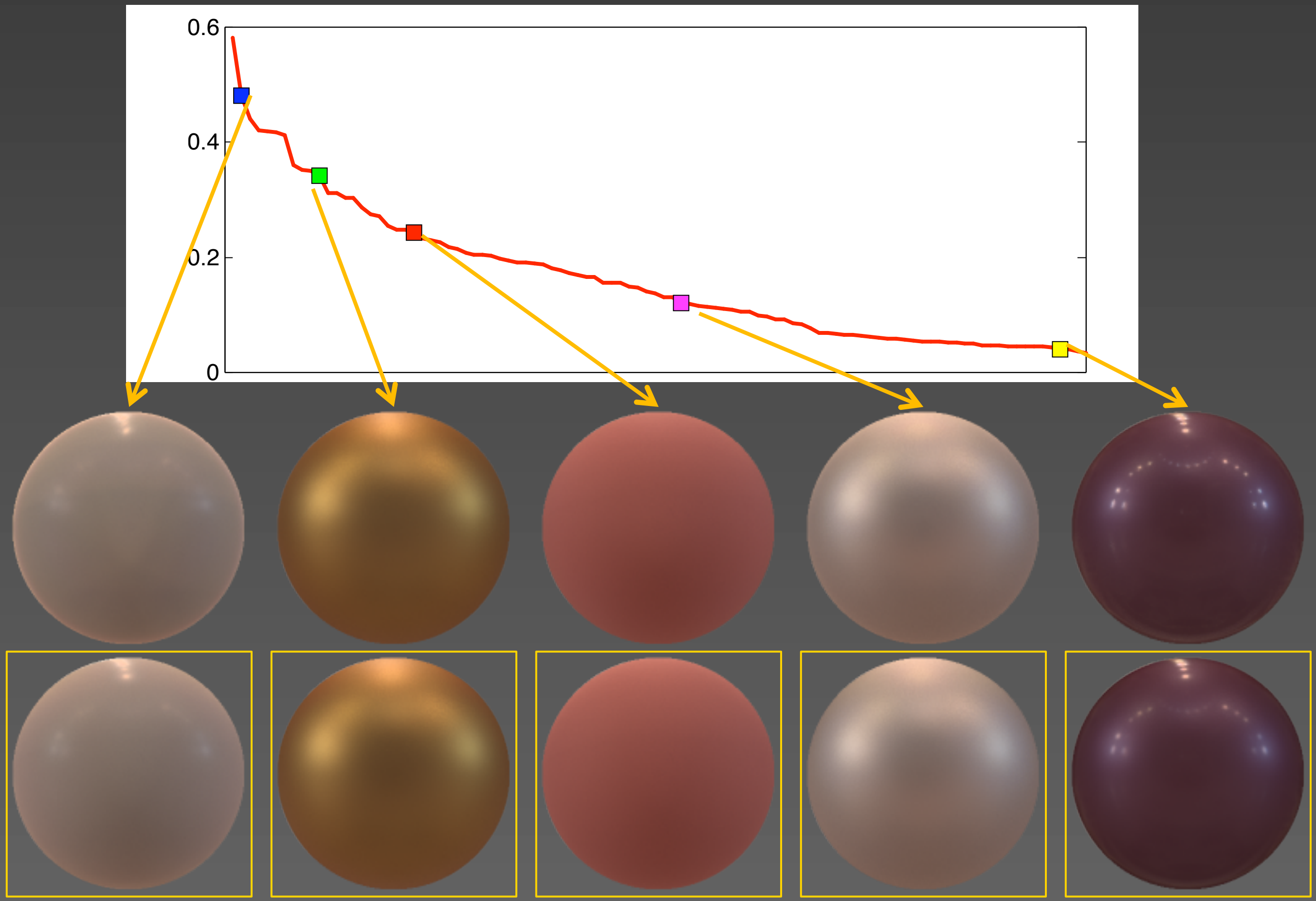


Optimal Number of Lobes





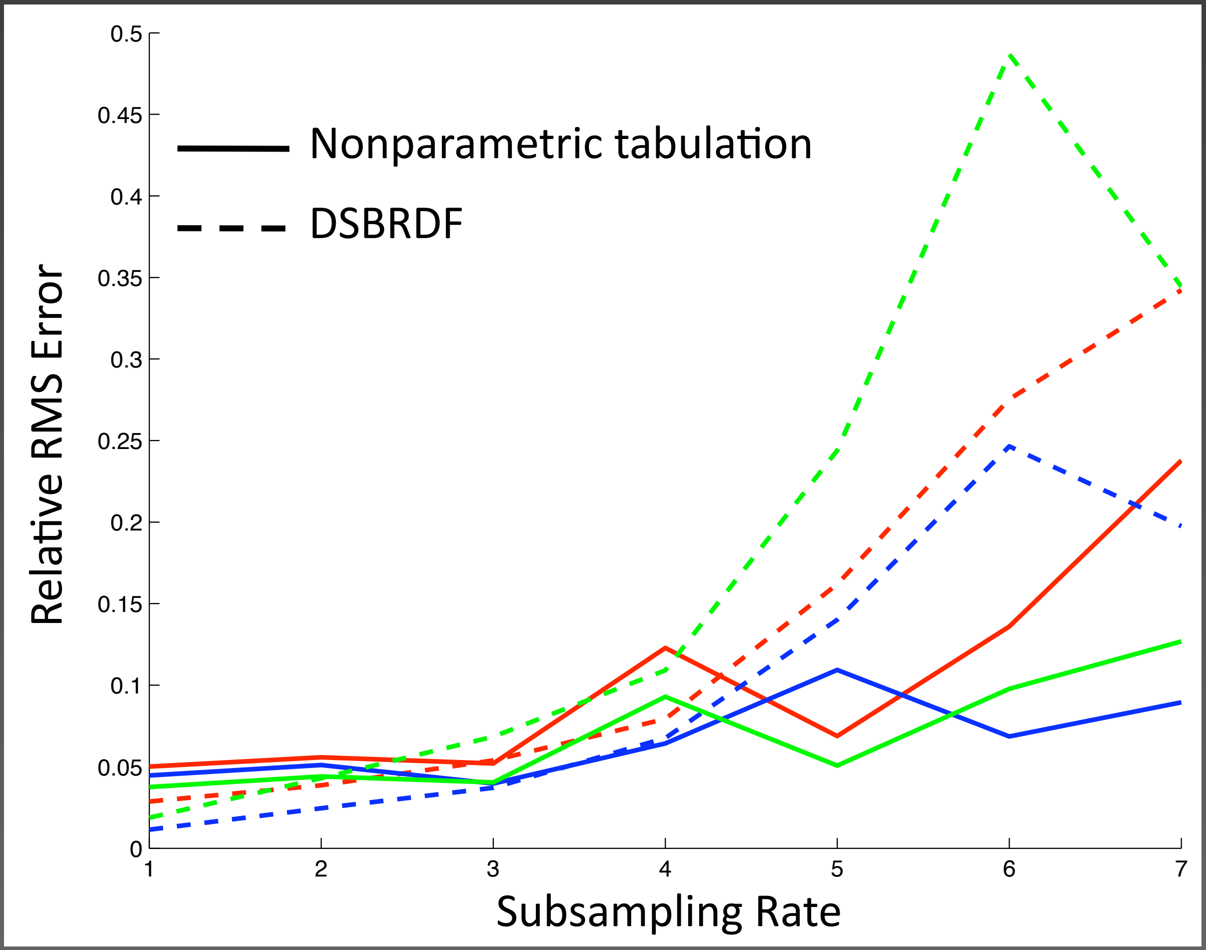
Relative RMS Error for 100 BRDFs



- Comparable to nonparametric model [Romeiro et al. 08]
  - With a much smaller footprint

Data courtesy of MERL Isotropic BRDF Database

Nonparametric vs. DSBRDF



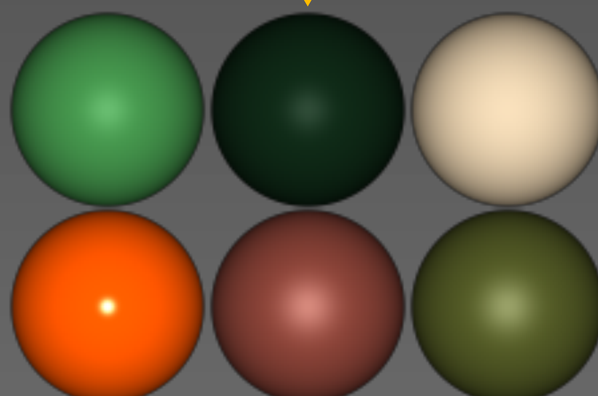
- Nonparametric representations are susceptible to reduced sampling

Radiometric Scene Decomposition

Strong constraints on reflectance

$$f^{-1}_{\mathbf{material}}(\text{image}) = \{\text{illumination}, \text{geometry}\}$$

Geometry

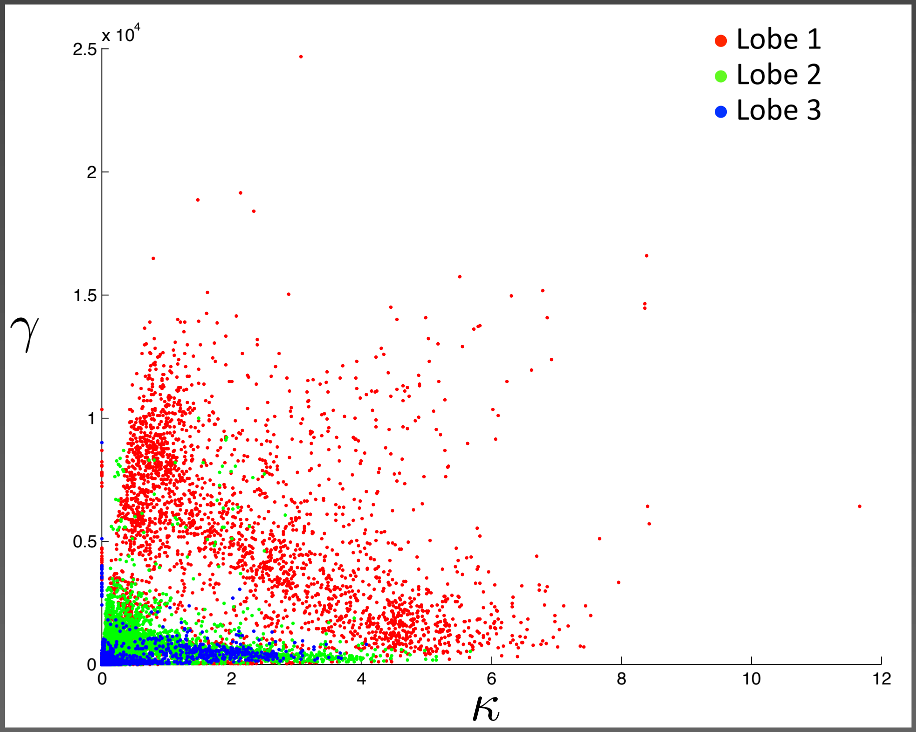


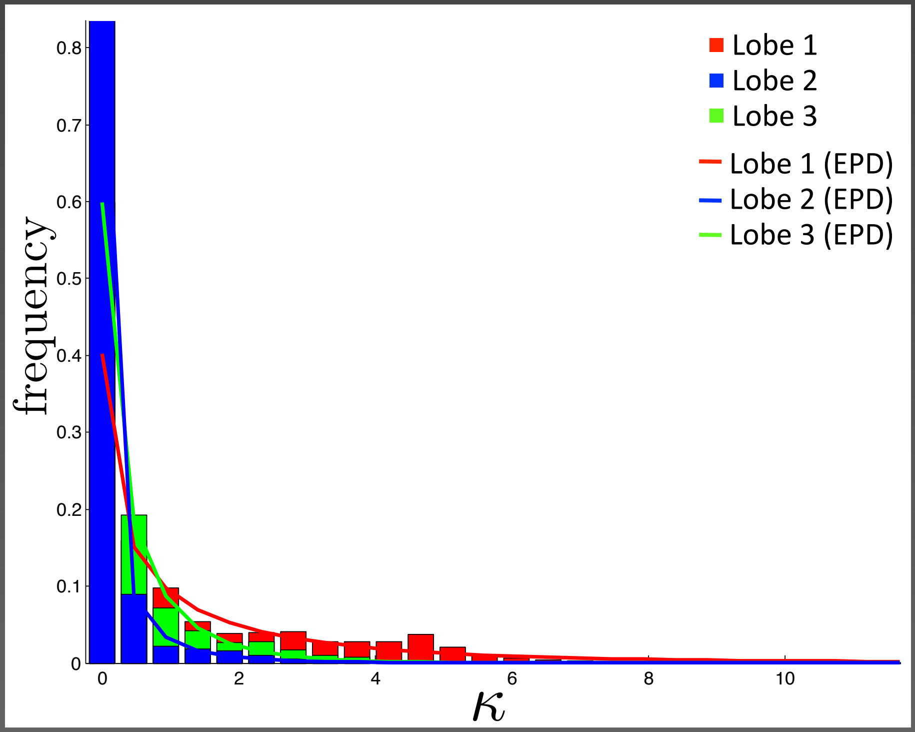
Material (Reflectance)

Illumination

Unary Reflectance Prior

$$p(\mathbf{R}) = \prod p(\mathbf{r_x}) \prod p(\mathbf{r_x}, \mathbf{r}_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})})$$





$$p(\mathbf{r_x}) = p(\kappa_\mathbf{x}, \gamma_\mathbf{x}) \qquad p(\mathbf{r_x}) = p(\kappa_\mathbf{x}) p(\gamma_\mathbf{x})$$

$$f_r(\theta_h, \phi_h | \theta_d) = \sum_{k=1}^{K} \exp\left[\kappa^{(k)} \cos^{\gamma^{(k)}} \theta_h\right] - 1$$

Probabilistic Factorization of Reflectance and Illumination

$$\underset{\mathbf{R},\mathbf{L}}{\text{argmax}}\, p(\mathbf{R},\mathbf{L}|\mathbf{I}) \propto p(\mathbf{I}|\mathbf{R},\mathbf{L})p(\mathbf{R})p(\mathbf{L})$$

- Likelihood $p(\mathbf{I}|\mathbf{R},\mathbf{L})$
  - DSBRDF model with Gaussian noise $\mathsf{N}(0,\Sigma)$
- Reflectance Priors $p(\mathbf{R})$
  - Unary: 1D/2D priors on DSBRDF parameters
  - Clique: Gaussian, L1, or Potts on DSBRDF parameters
- Illumination Prior $p(\mathbf{L})$
  - Clique: Gaussian, L1, or Potts for env. lighting