

Intrinsic Scene Properties from a Single RGB-D Image



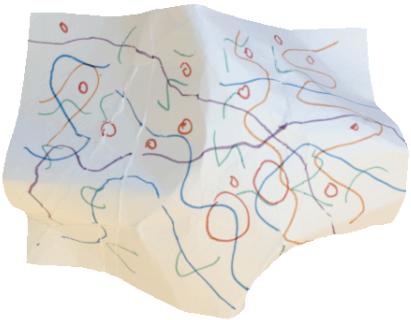
Jonathan Barron & Jitendra Malik
UC Berkeley
CVPR 2013

SIRFS

Shape, Illumination and Reflectance from Shading

Barron & Malik 2011, 2012A, 2012B, 2013

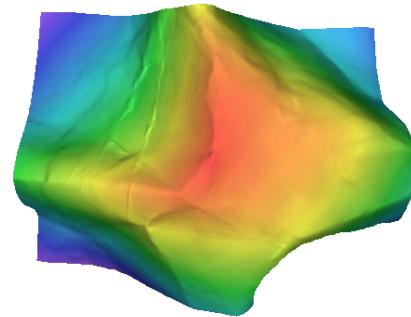
Input



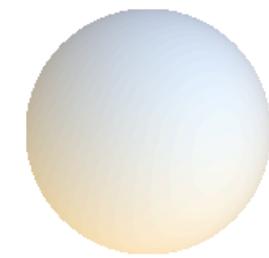
Image



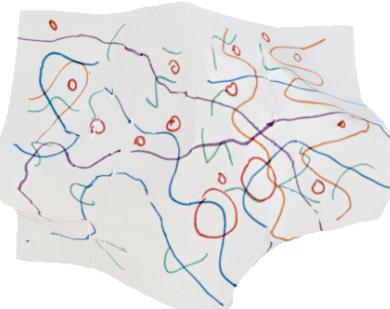
Output



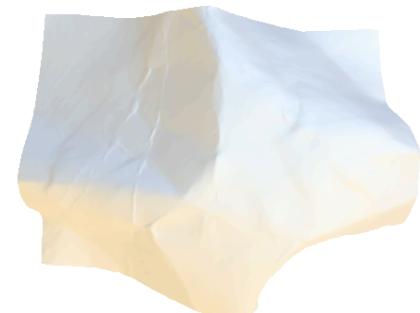
Shape



Global Illumination



Reflectance



Shading

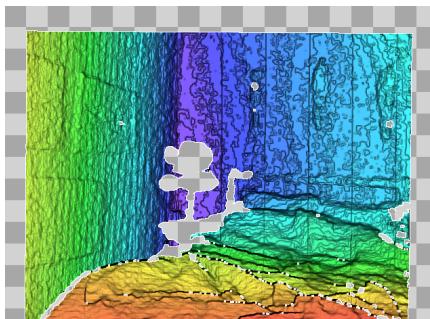
Scene-SIRFS

(this paper)

Input



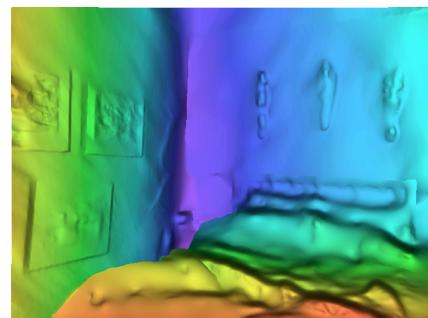
Image



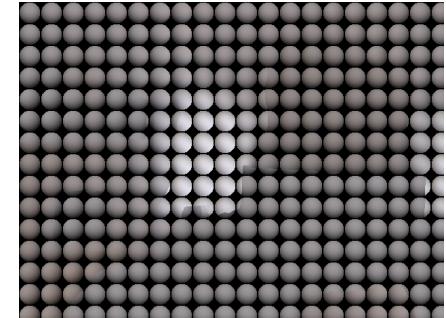
Rough Shape



Output



Improved Shape



Per-Pixel Illumination



Reflectance

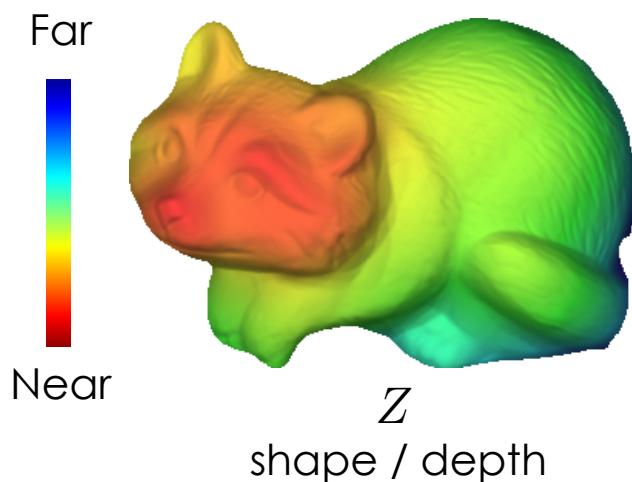


Shading

Outline

- 1) Review **SIRFS**
- 2) Present **Scene-SIRFS**

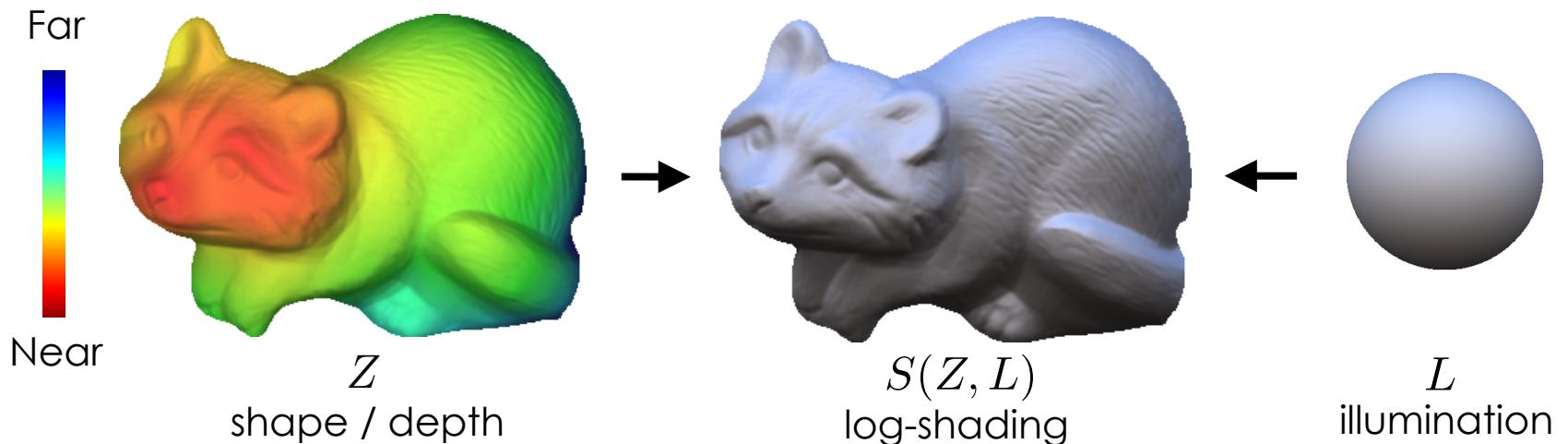
Forward Optics



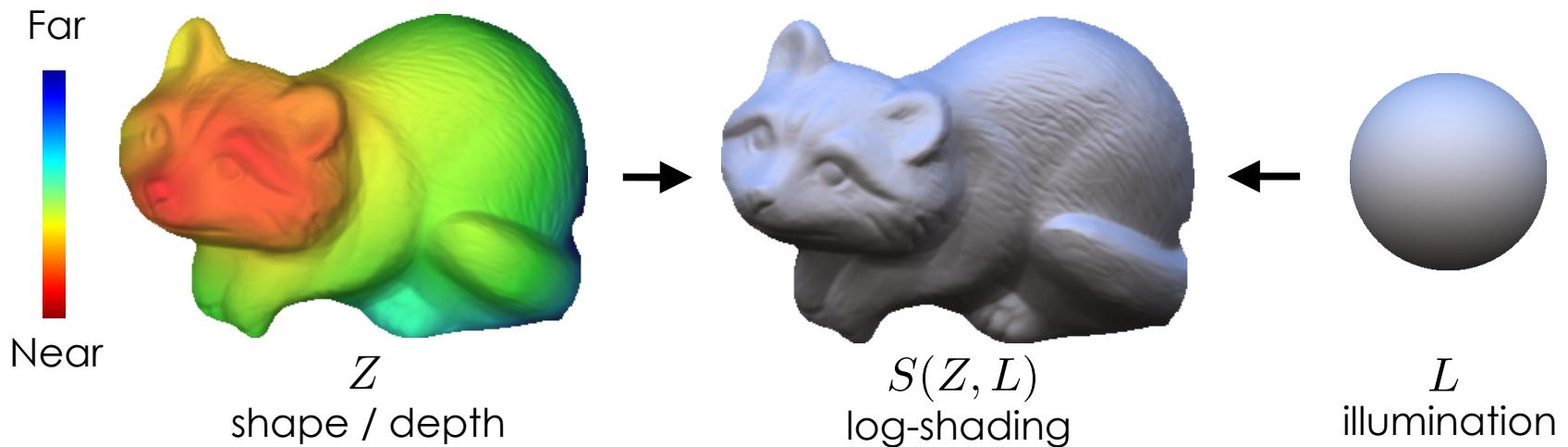
Forward Optics



Forward Optics

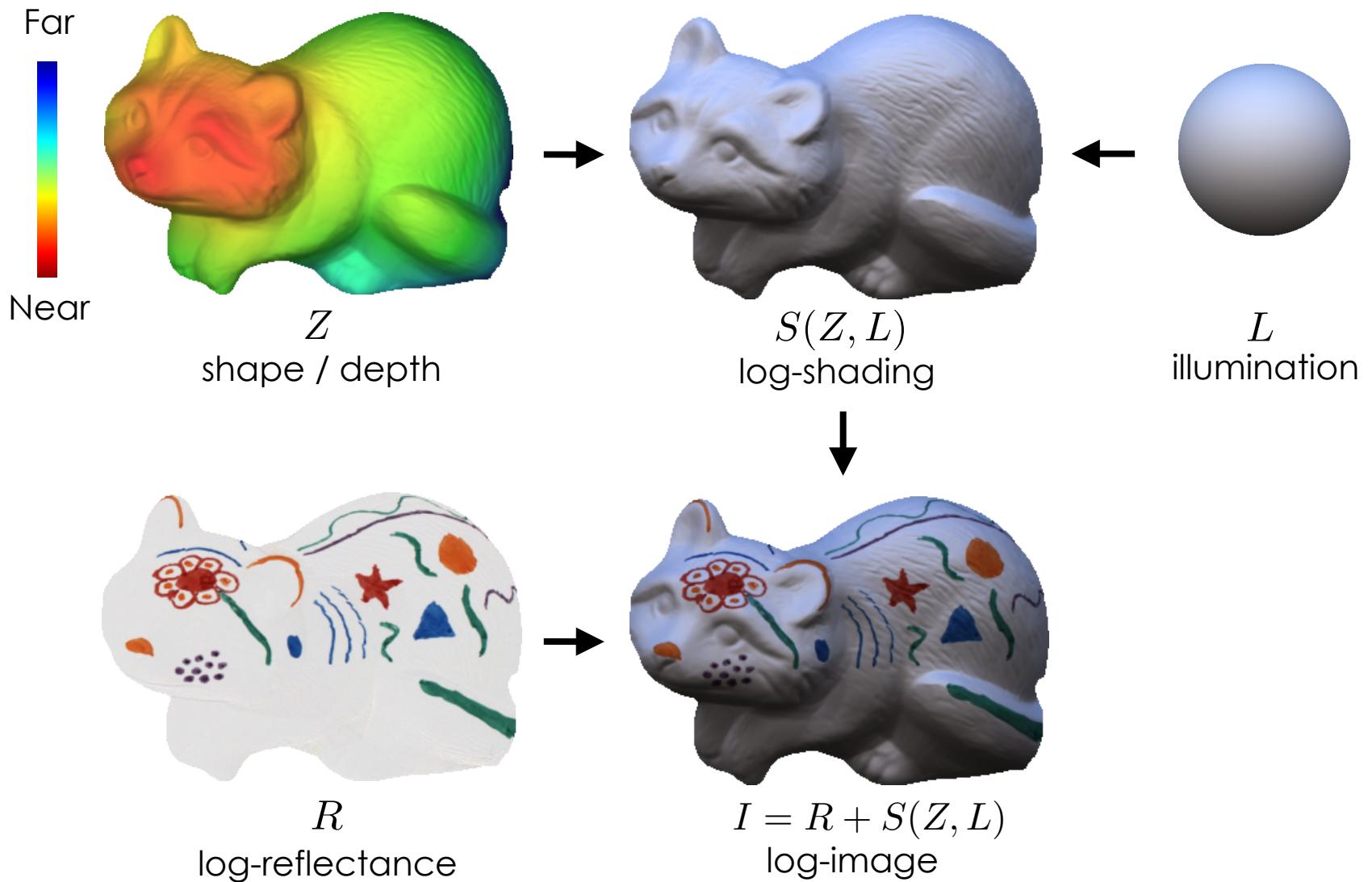


Forward Optics

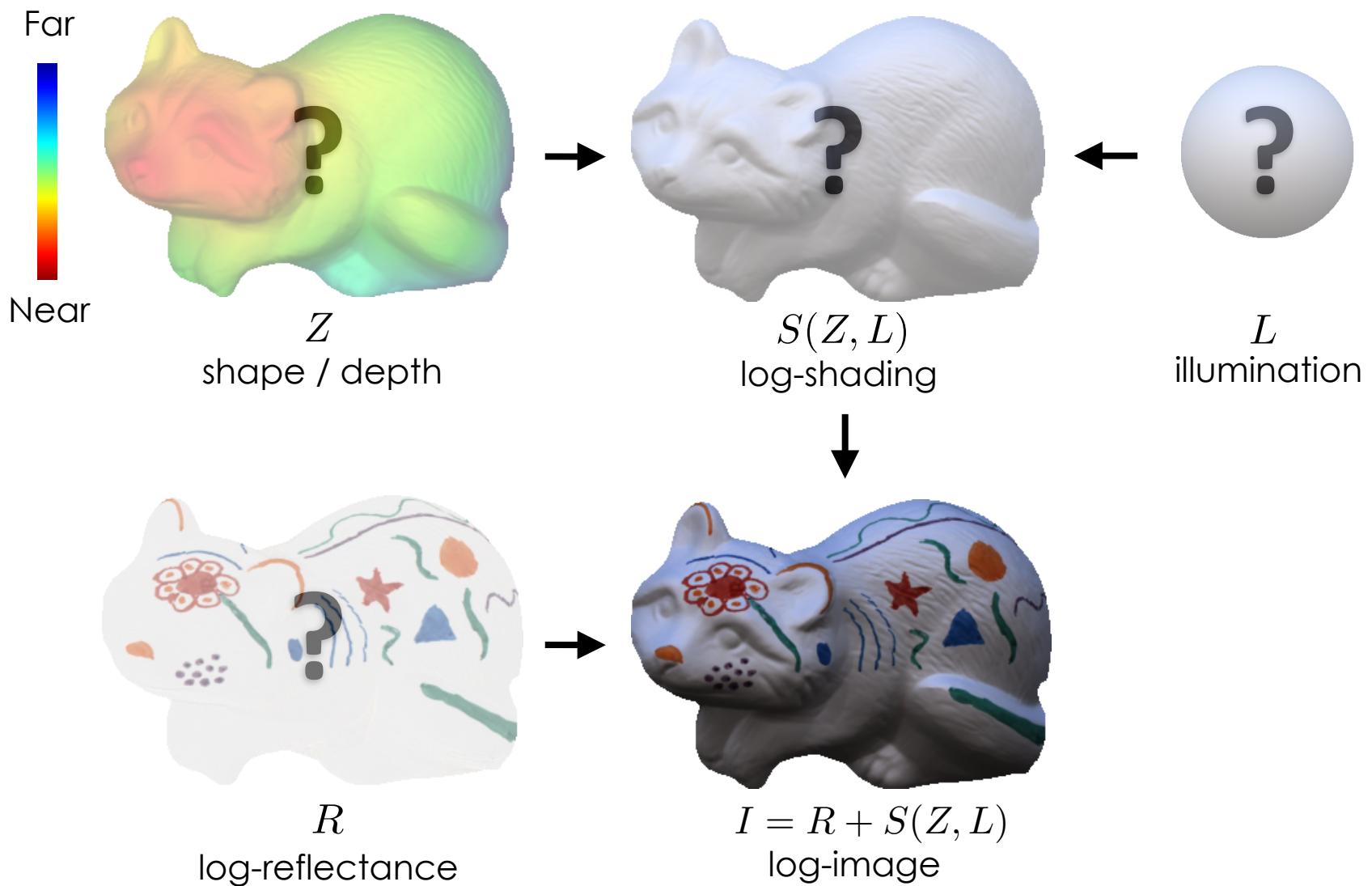


R
log-reflectance

Forward Optics



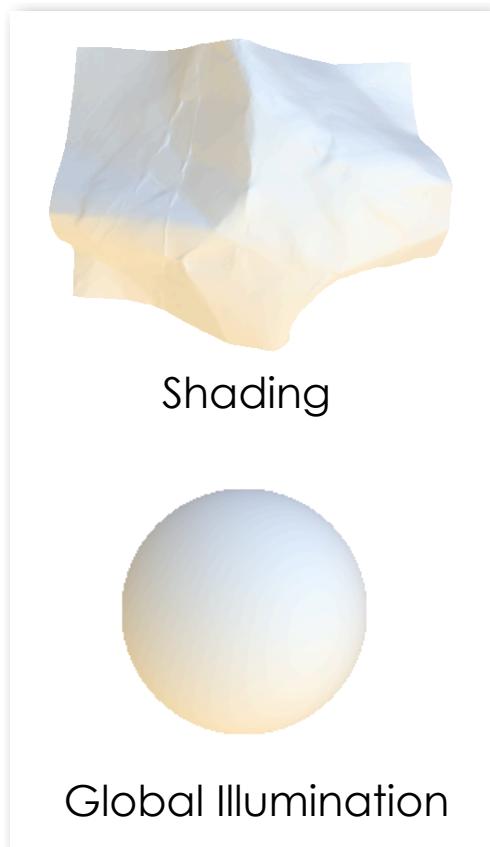
Inverse Optics



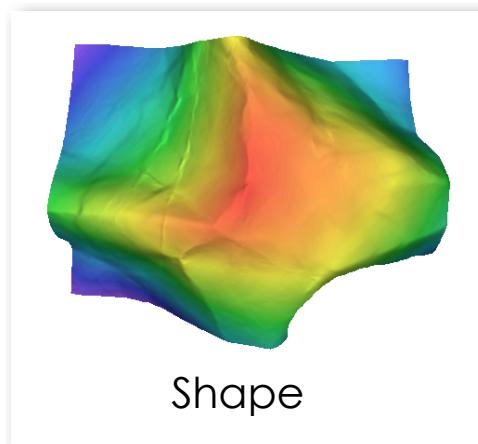
Shape from Shading

Horn 1970, 1975, 1989, Ikeuchi & Horn 1981, Belhumeur et al. 1999, Johnson & Adelson 2011

Input



Output

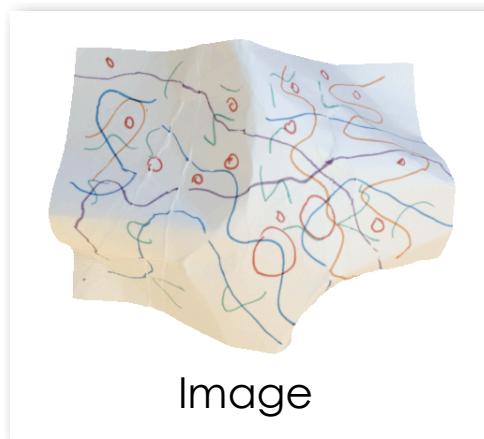


Problematic Assumption:
illumination and
reflectance are
assumed to be known

Intrinsic Images

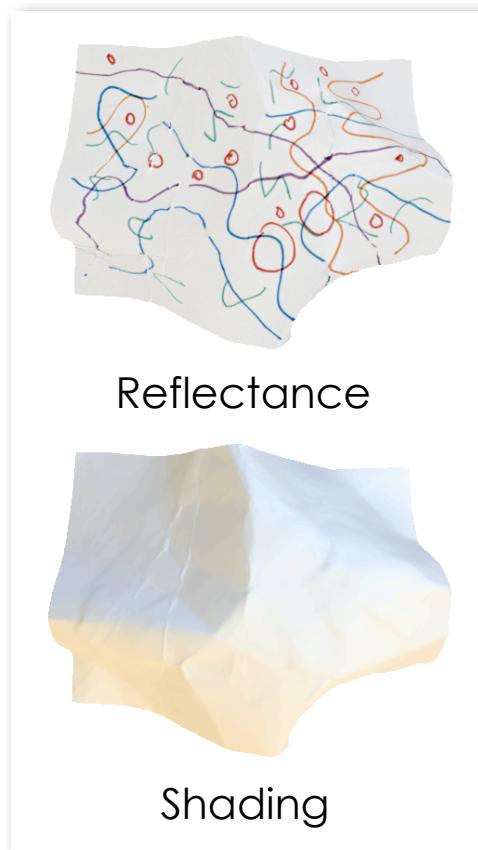
Land & McCann 1971, Horn 1974, Barrow & Tenenbaum 1978, Tappen et al. 2005, Gehler et al. 2011

Input



Image

Output



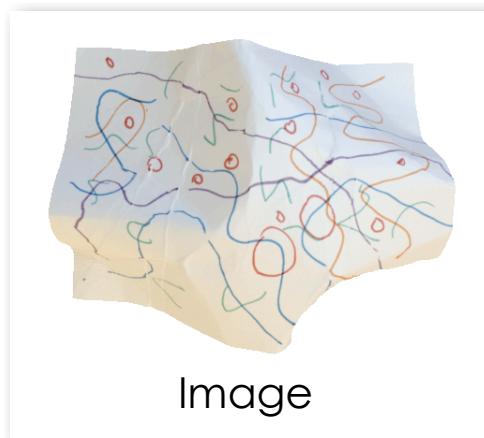
Problematic Assumption:
shape and illumination
are ignored

SIRFS

Shape, Illumination and Reflectance from Shading

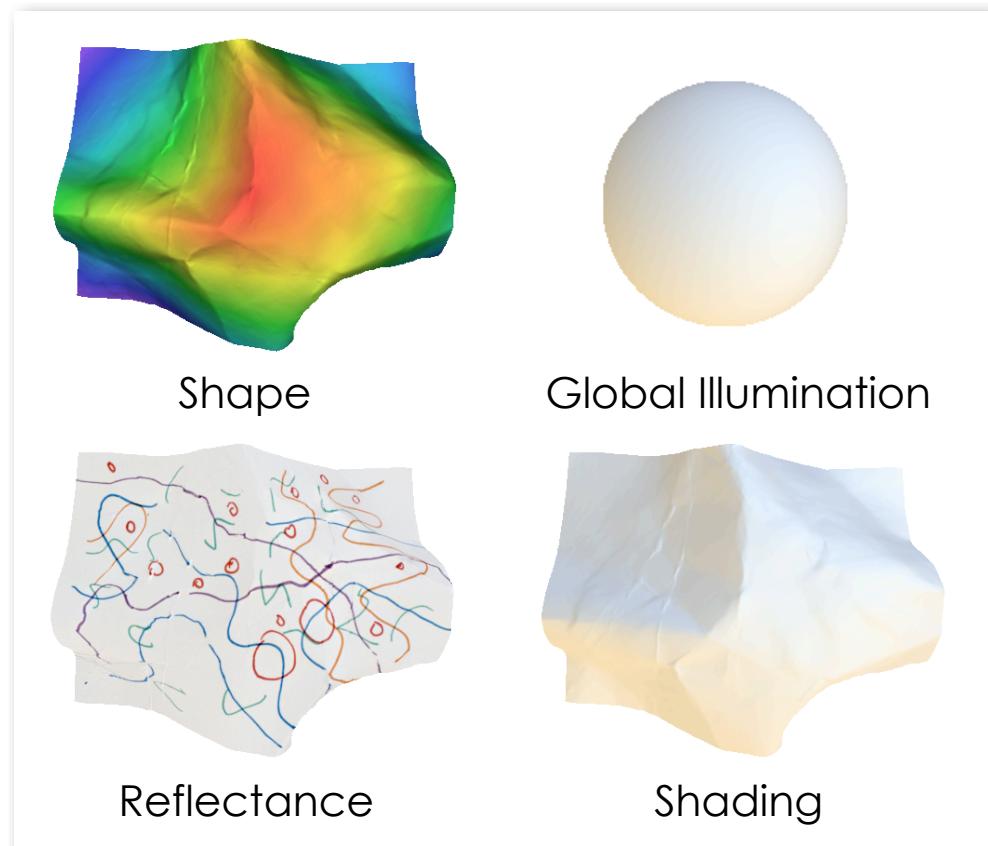
Barron & Malik 2011, 2012A, 2012B, 2013

Input



Image

Output



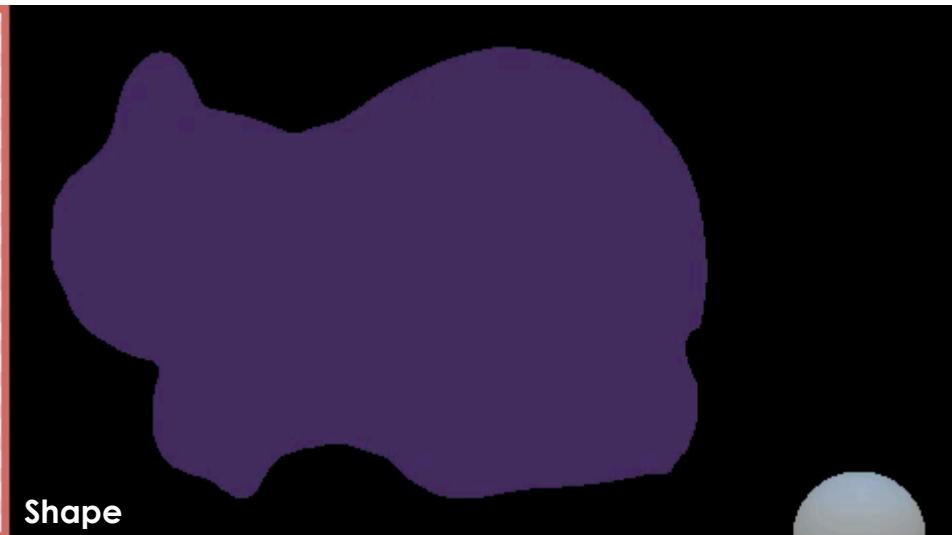
Optimization Demo

Image

Optimization Demo



Image



Shape



Reflectance

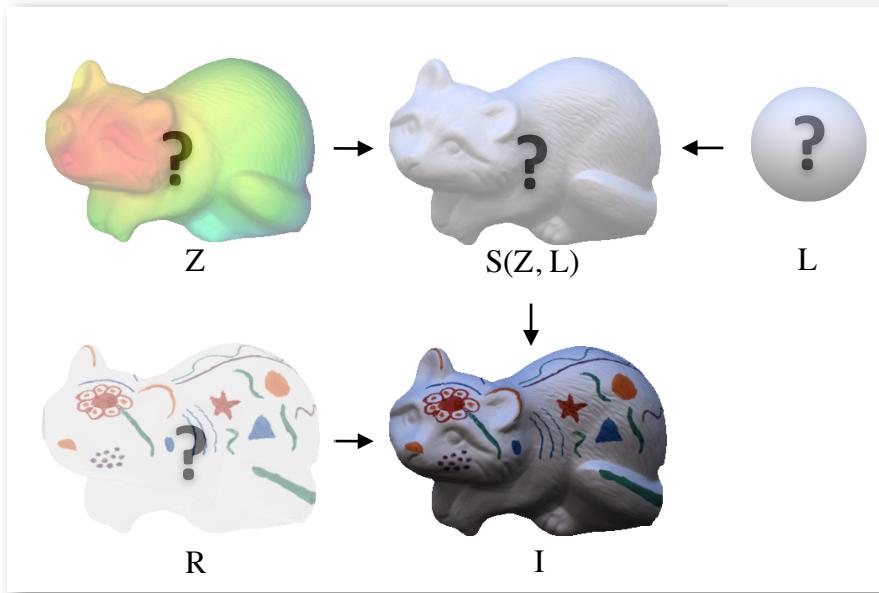


Shading



Illumination

Problem Formulation



maximize
 Z, R, L

$$P(R)P(Z)P(L)$$

subject to

$$I = R + S(Z, L)$$

“Search for the most likely world
that exactly recreates the input image”

Real World Images



Image

Shape

Normals

Refl.

Shading Light

Rotated

Relit

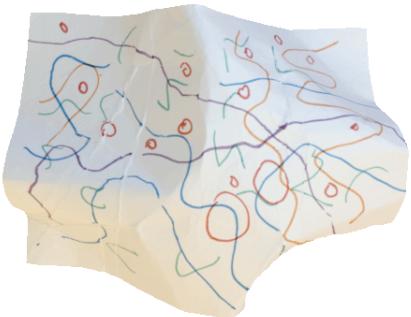
Graphics Applications

Graphics Applications



SIRFS

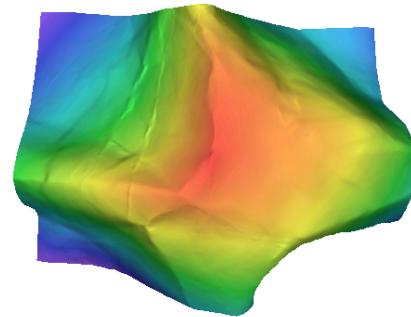
Input



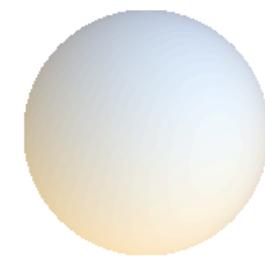
Image



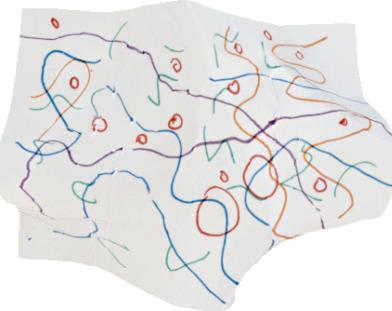
Output



Shape



Global Illumination



Reflectance



Shading

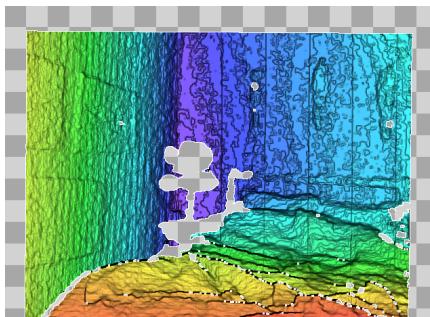
Works well on
single objects, but
what about
scenes?

Scene-SIRFS

Input



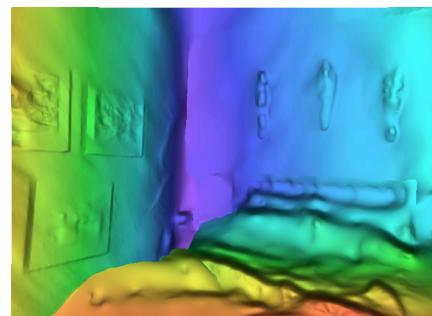
Image



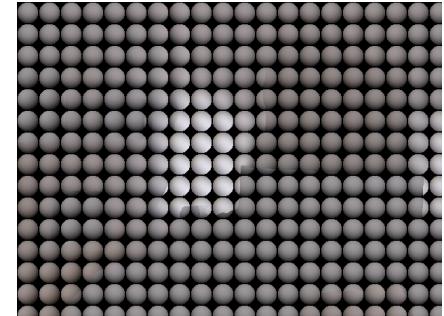
Rough Shape



Output



Improved Shape



Per-Pixel Illumination

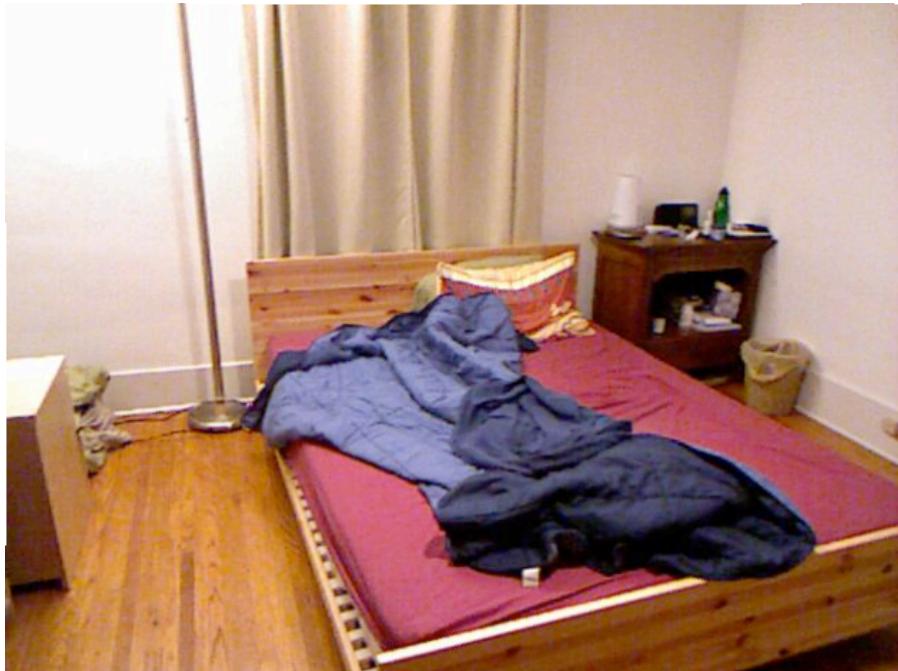


Reflectance

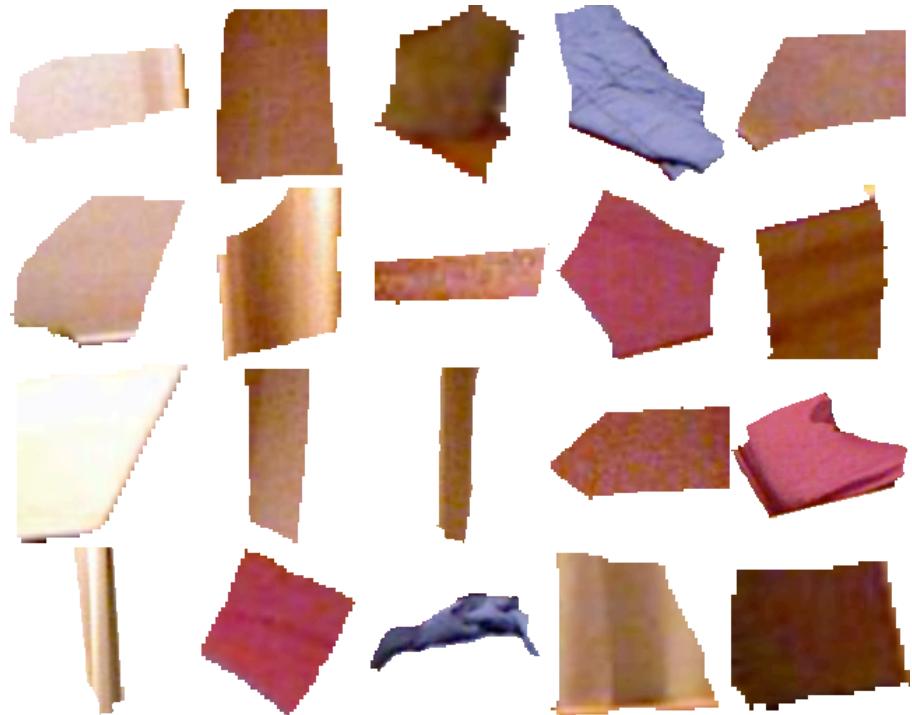


Shading

Segmentation + SIRFS Doesn't Work



Input Image



Segments

- Segments rarely correspond to individual objects or individual illumination.
- Inference needs to share information across segments.

Scene-SIRFS

Goals: 1) Generalize SIRFS from objects to scenes

Scene-SIRFS

Goals:

- 1) Generalize SIRFS from objects to scenes
 - 1a) 2.5D shape mixture (occlusion, depth discontinuities)

Scene-SIRFS

Goals:

- 1) Generalize SIRFS from objects to scenes
 - 1a) 2.5D shape mixture (occlusion, depth discontinuities)
 - 1b) Spatially varying illumination (shadows, attenuation)

Scene-SIRFS

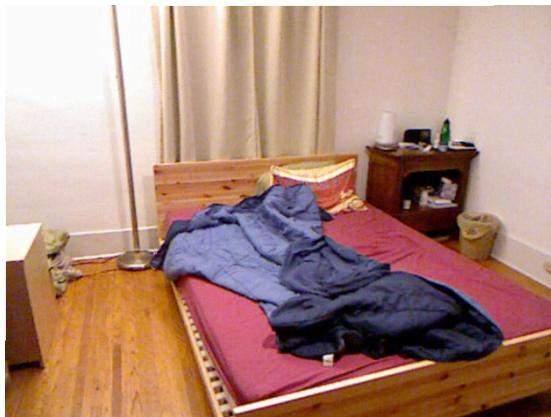
- Goals:
- 1) Generalize SIRFS from objects to scenes
 - 1a) 2.5D shape mixture (occlusion, depth discontinuities)
 - 1b) Spatially varying illumination (shadows, attenuation)
 - 2) Incorporate external shape information (Kinect)

Soft Segmentation

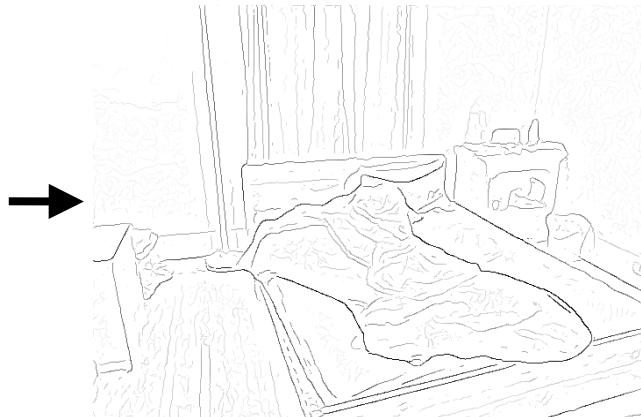


Input Image

Soft Segmentation



Input Image

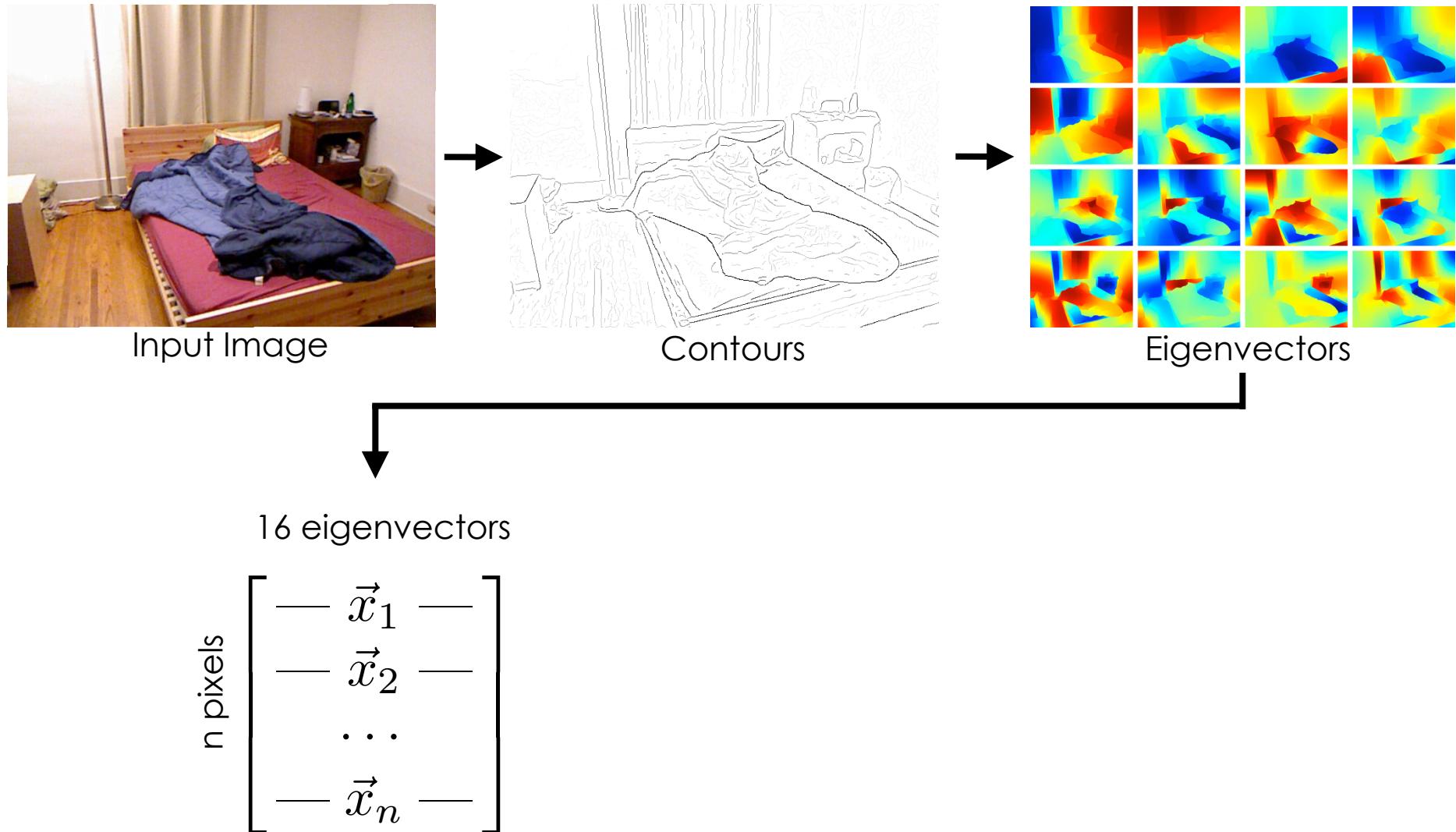


Contours

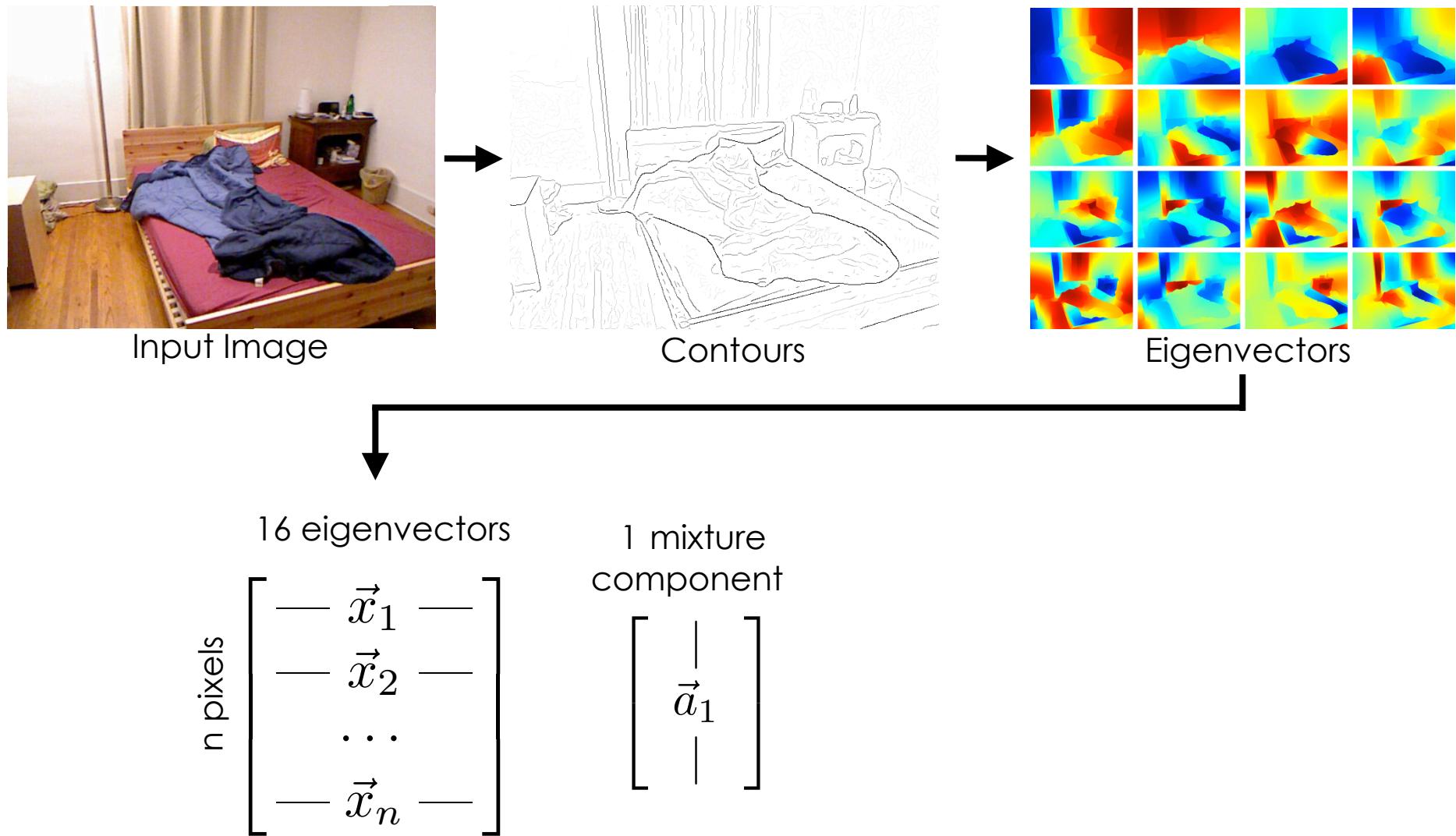
Soft Segmentation



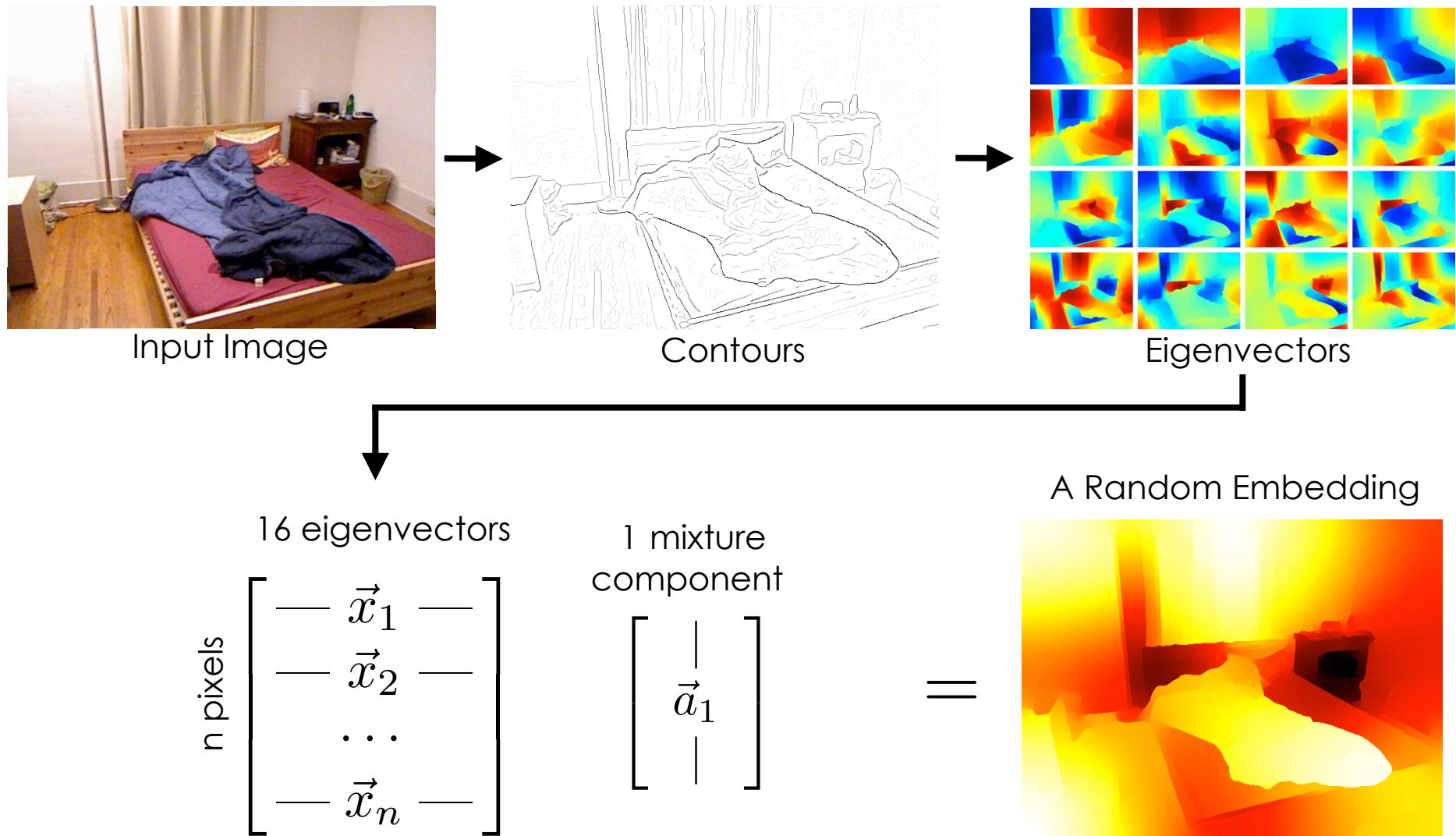
Soft Segmentation



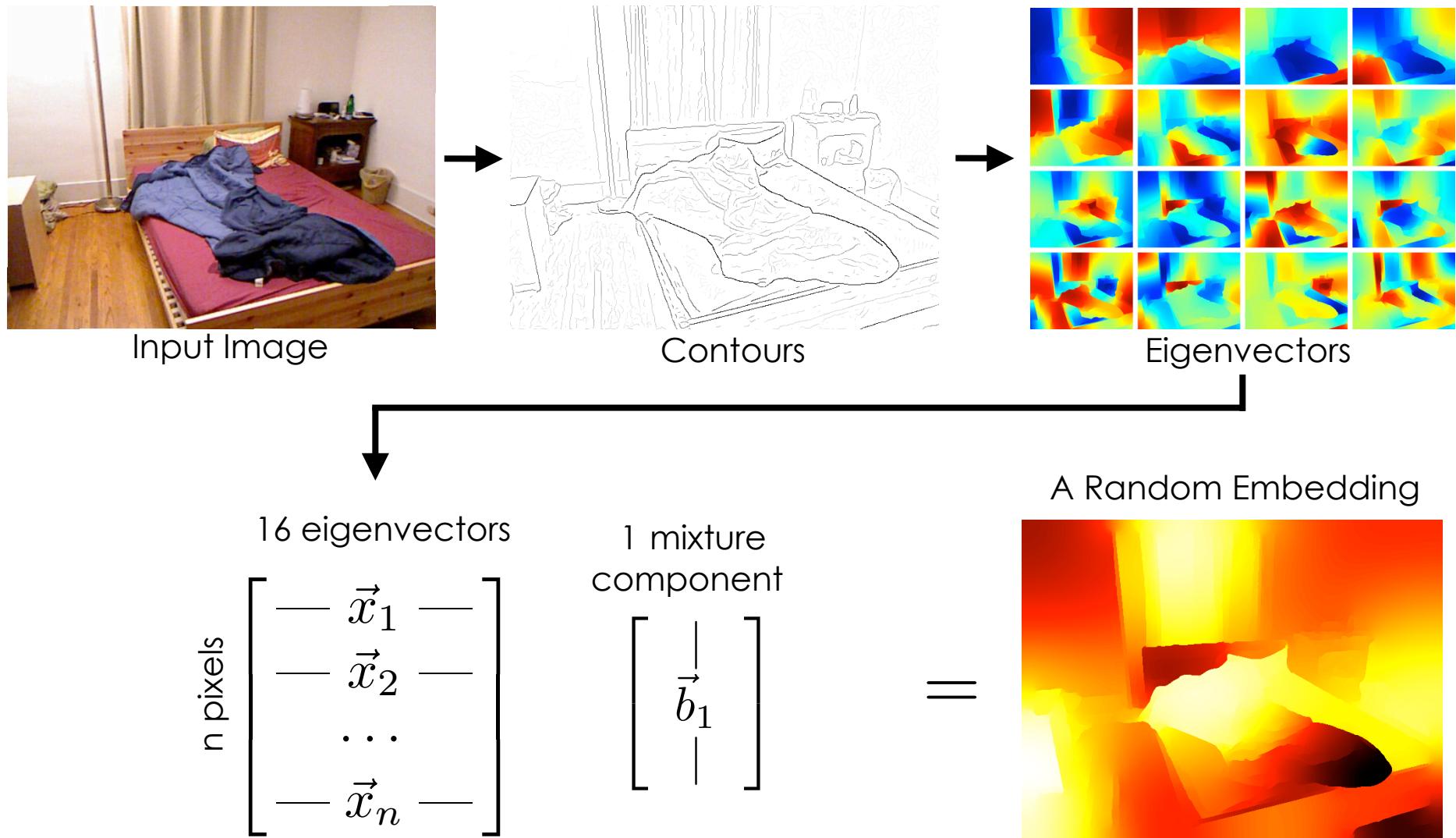
Soft Segmentation



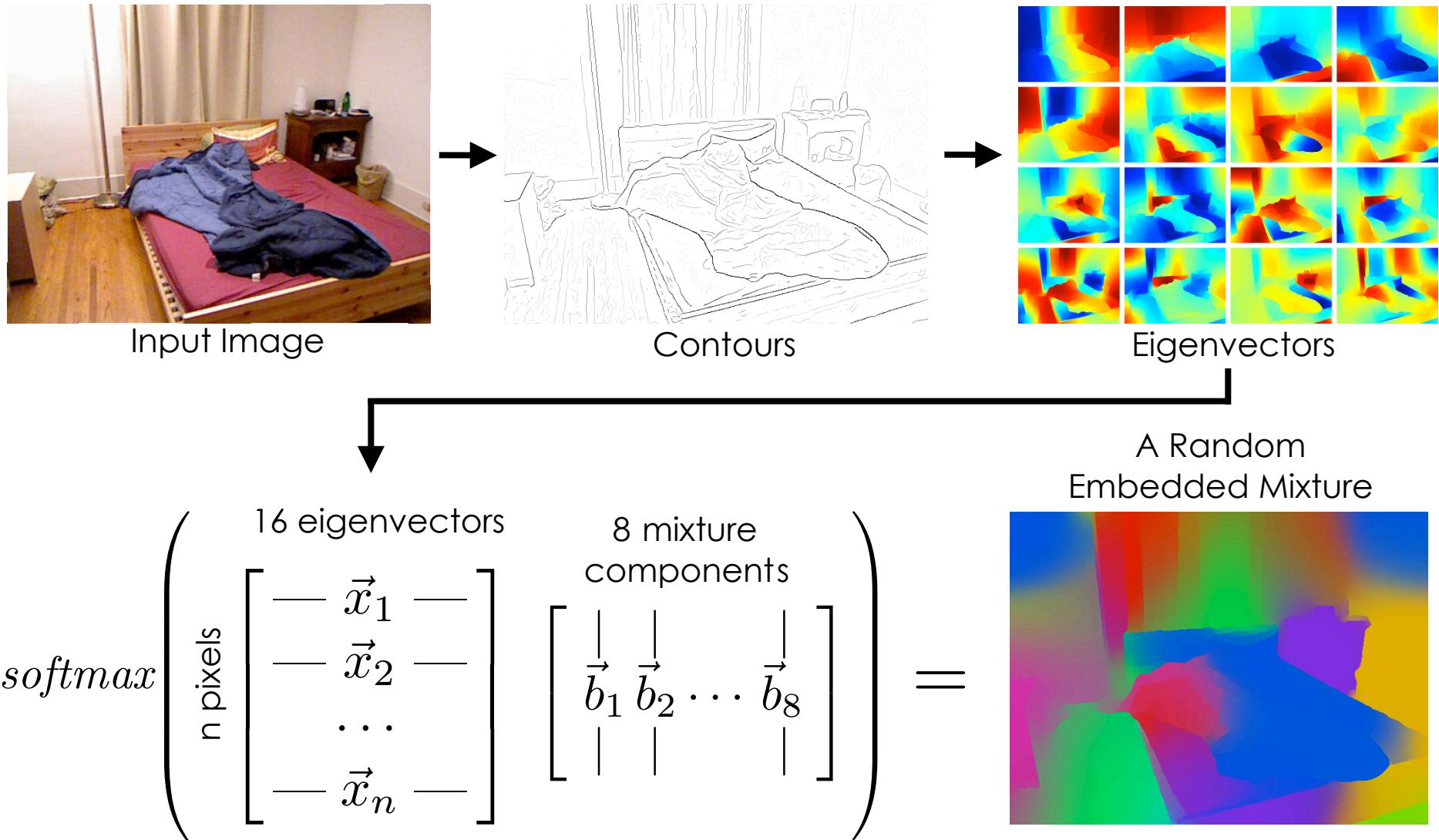
Soft Segmentation



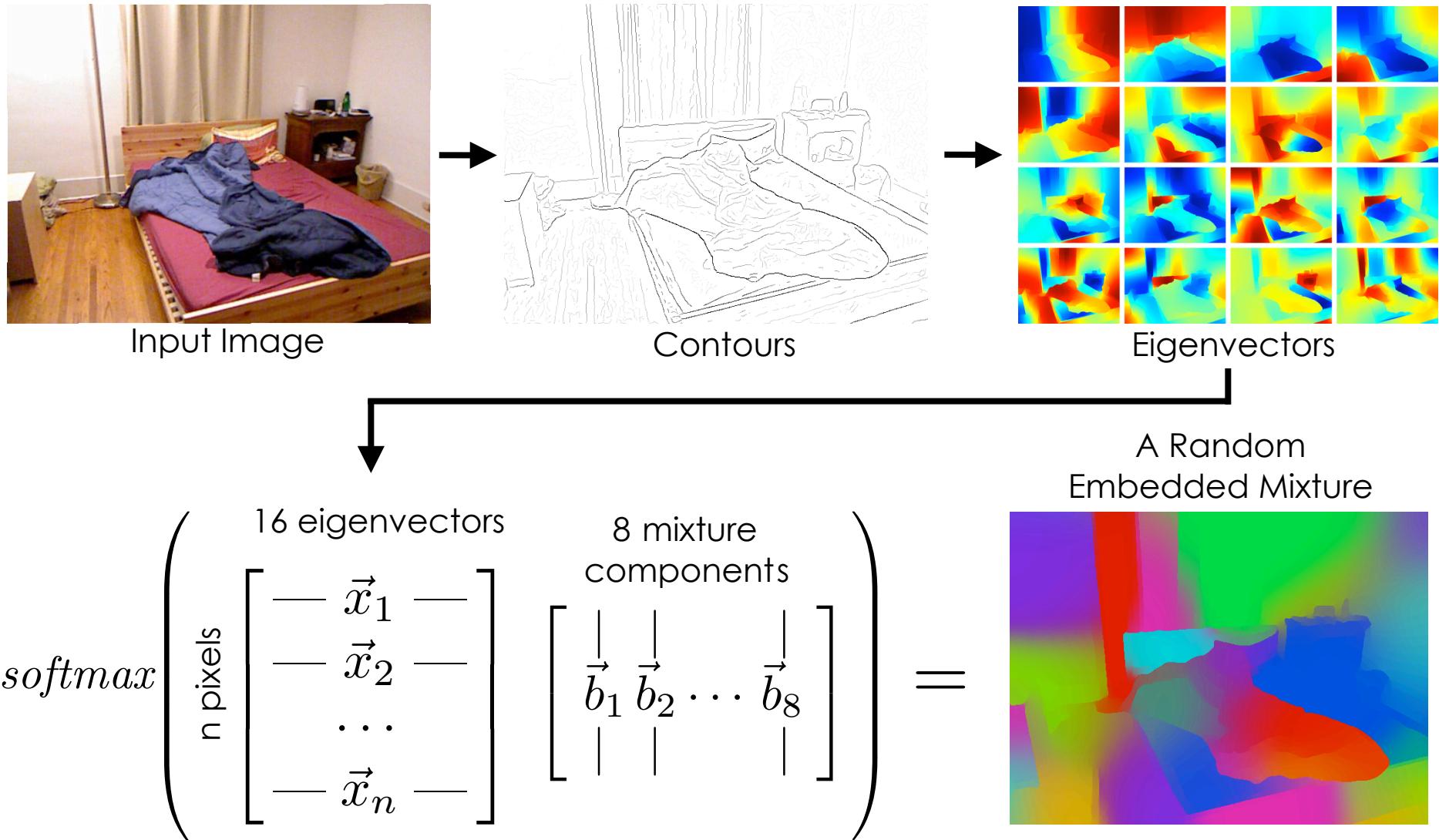
Soft Segmentation



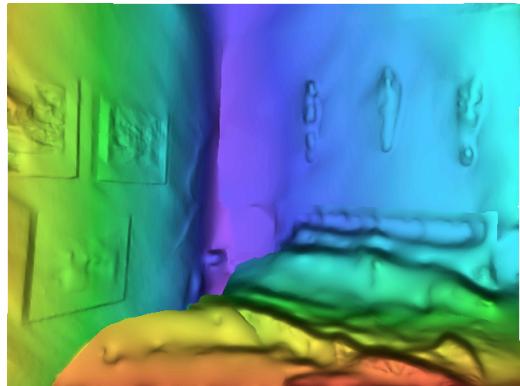
Soft Segmentation



Soft Segmentation



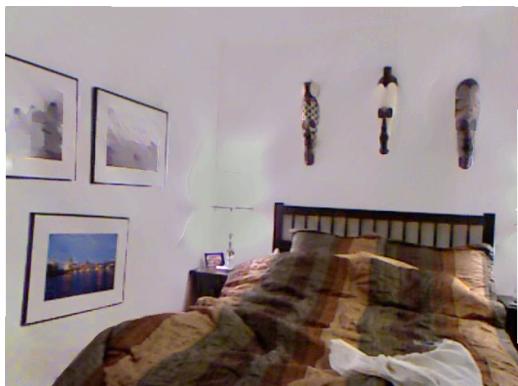
Scene-SIRFS



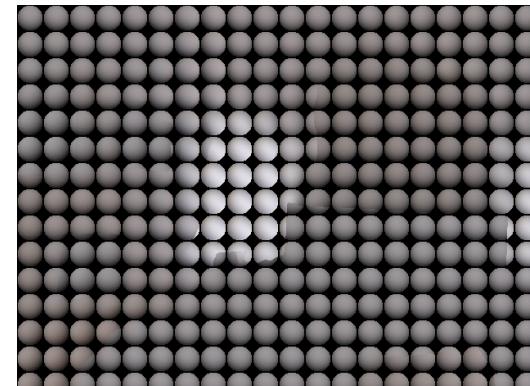
Z



$S(Z, L)$



R

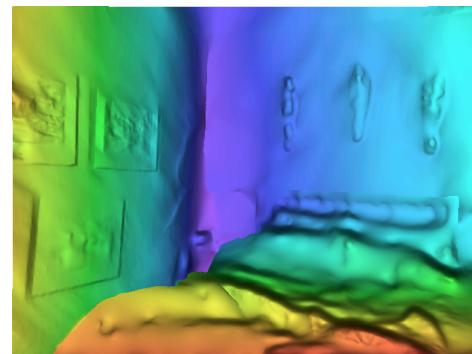


L

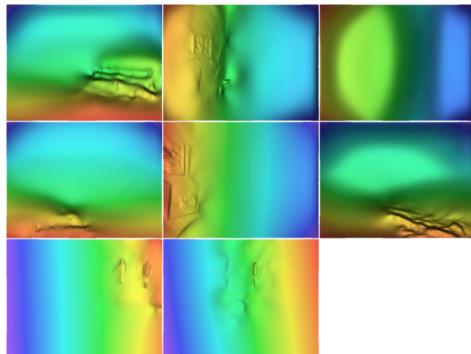


$I = R + S(Z, L)$

Scene-SIRFS



Scene-SIRFS



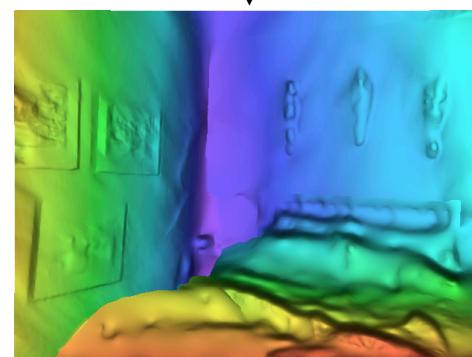
$$\{Z_i\}$$

Set of constituent shapes



$$\{U_i \propto \exp(B\psi_i)\}$$

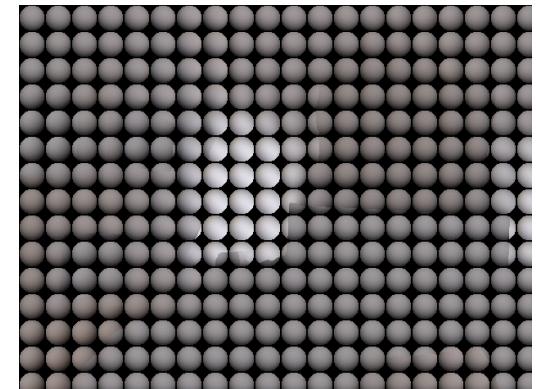
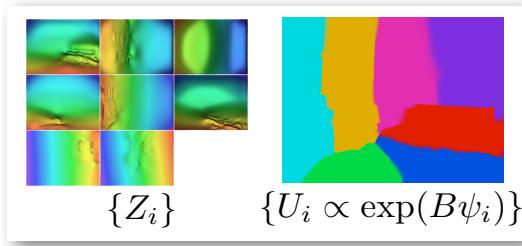
"Foreground" probability for each shape



$$\sum Z_i U_i$$

Piecewise-contiguous shape

Scene-SIRFS



L

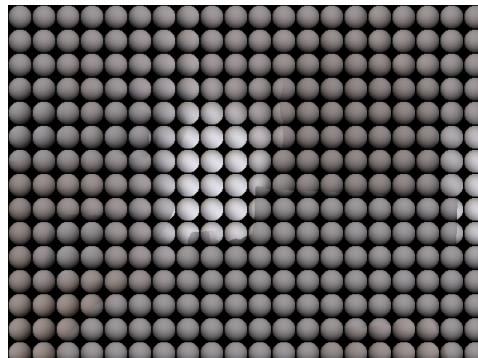
$$S \left(\sum Z_i U_i, L \right)$$



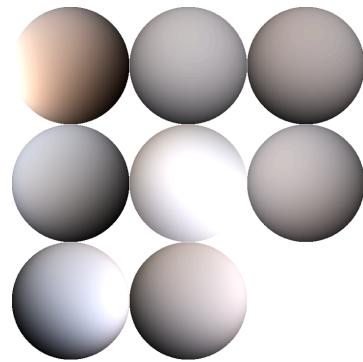
R

$$I = R + S \left(\sum Z_i U_i, L \right)$$

Scene-SIRFS

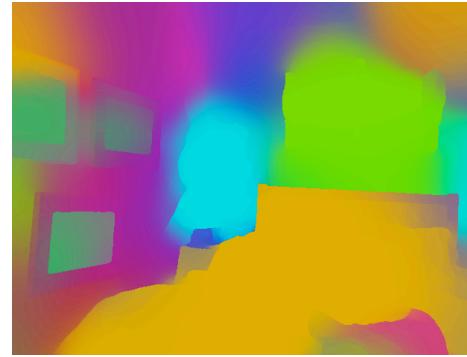


Scene-SIRFS



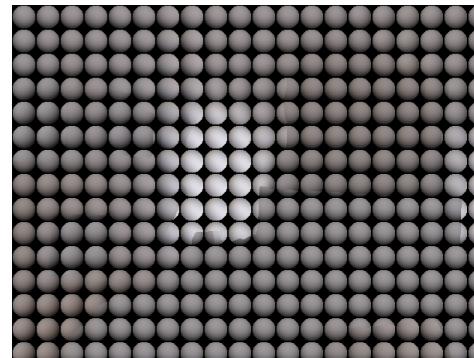
$$\{L_i\}$$

Set of constituent lights



$$\{V_i \propto \exp(B\omega_i)\}$$

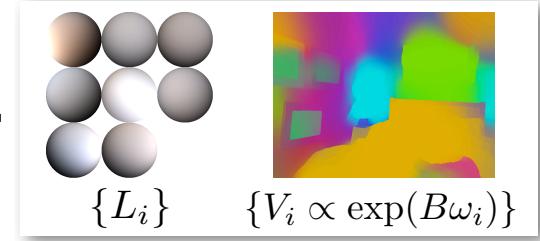
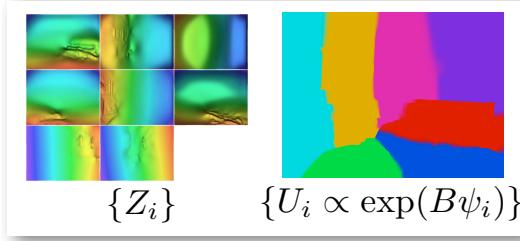
“Ownership” probability for each light



$$\sum L_i V_i$$

Spatially-varying illumination

Scene-SIRFS

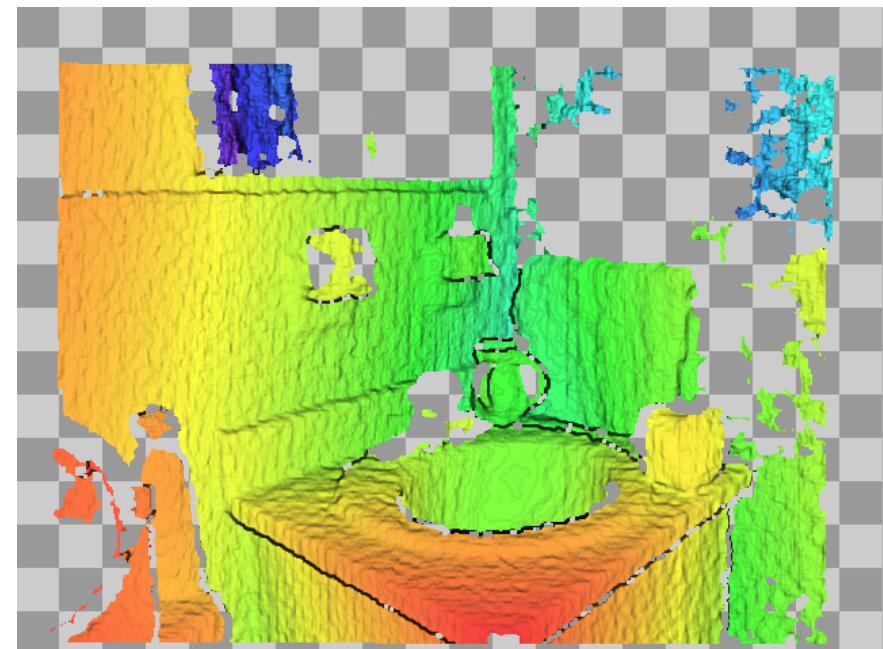


$$I = R + S \left(\sum Z_i U_i, \sum L_i V_i \right)$$

the Kinect



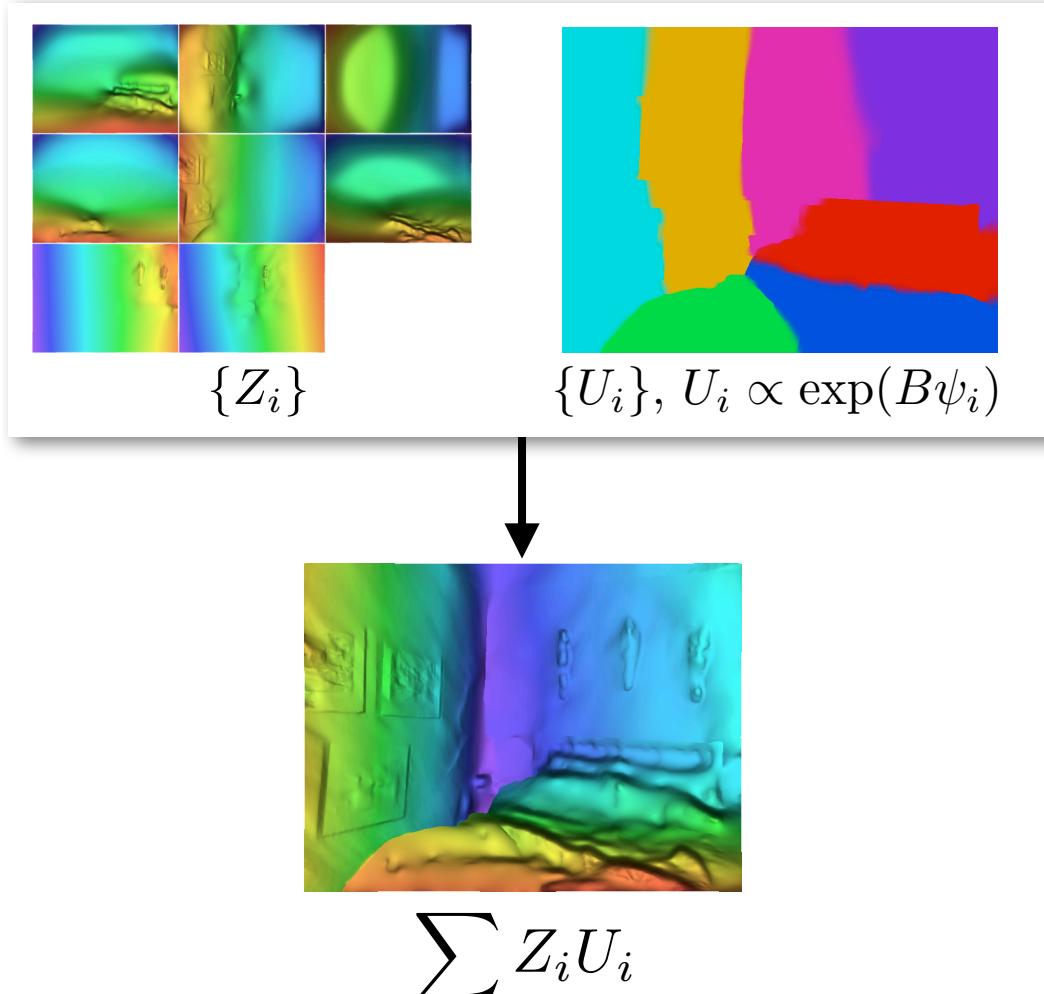
RGB



D

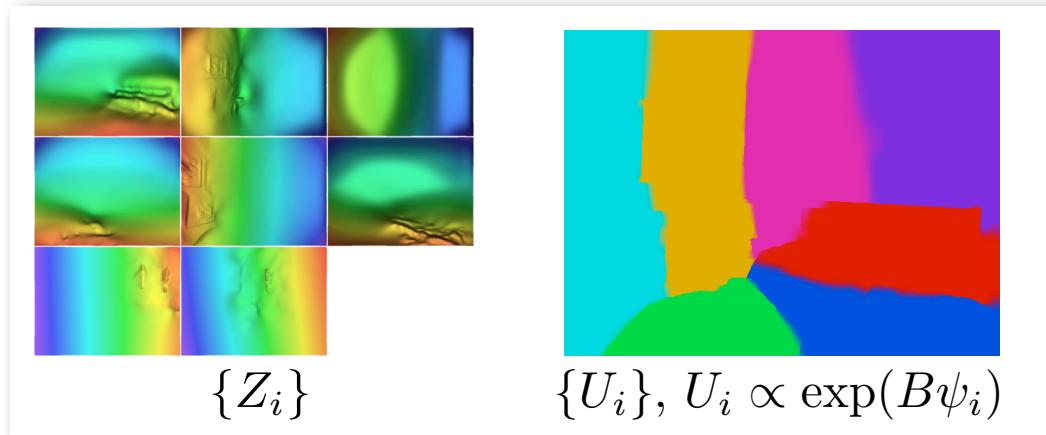
Noise due to:
occlusion of structured light, dark objects,
sensor noise, alignment errors, time
synchronization, quantization, inherent limits
of binocular disparity, etc

the Kinect



Expected “foreground” shape

the Kinect



A large downward arrow points from the top row to the bottom row. The bottom row consists of two side-by-side depth maps separated by a minus sign. The left depth map is labeled $\sum Z_i U_i$. The right depth map is labeled \hat{Z} . To the right of the right depth map is a double vertical bar symbol followed by the letter p , representing the Kinect depth.

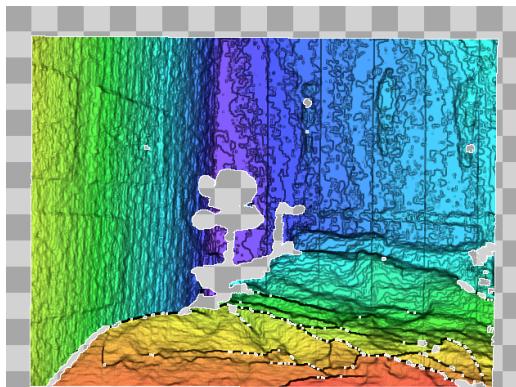
$$\sum Z_i U_i - \hat{Z} \| p$$

Expected “foreground” depth should resemble Kinect depth

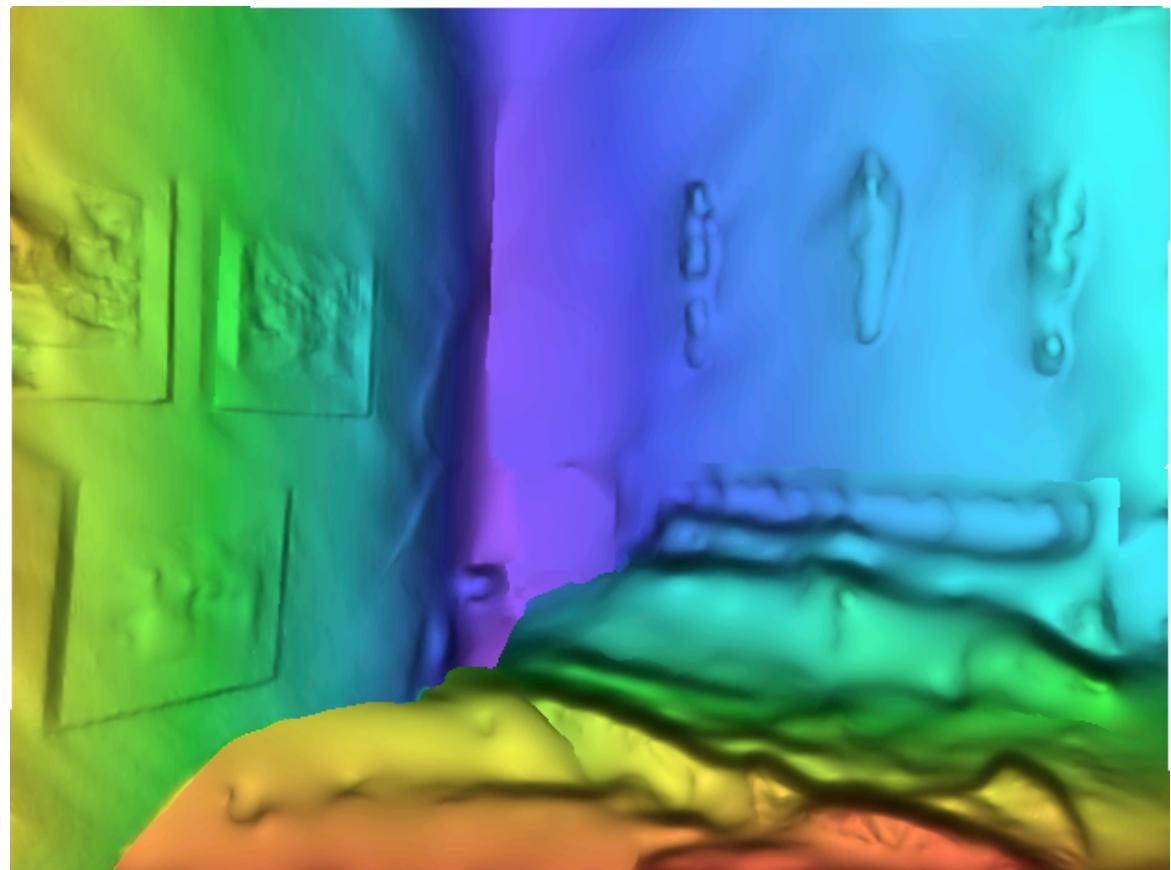
Output



Input Image



Input Shape

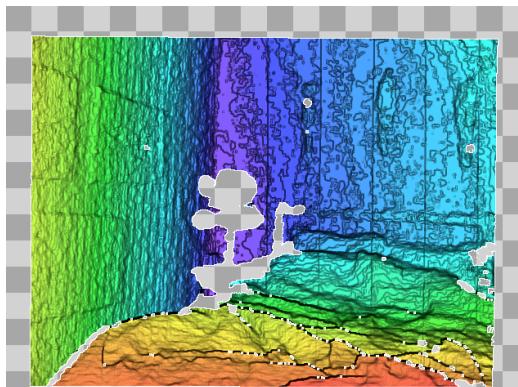


Output Depth

Output



Input Image



Input Shape

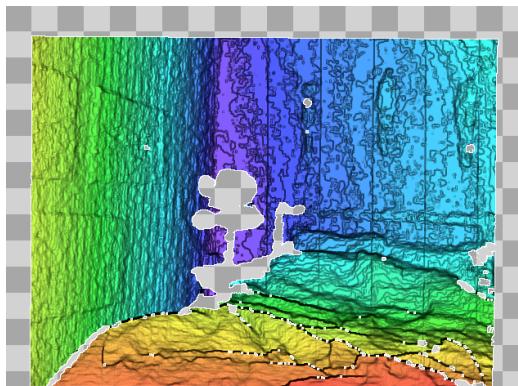


Output Depth Mixture

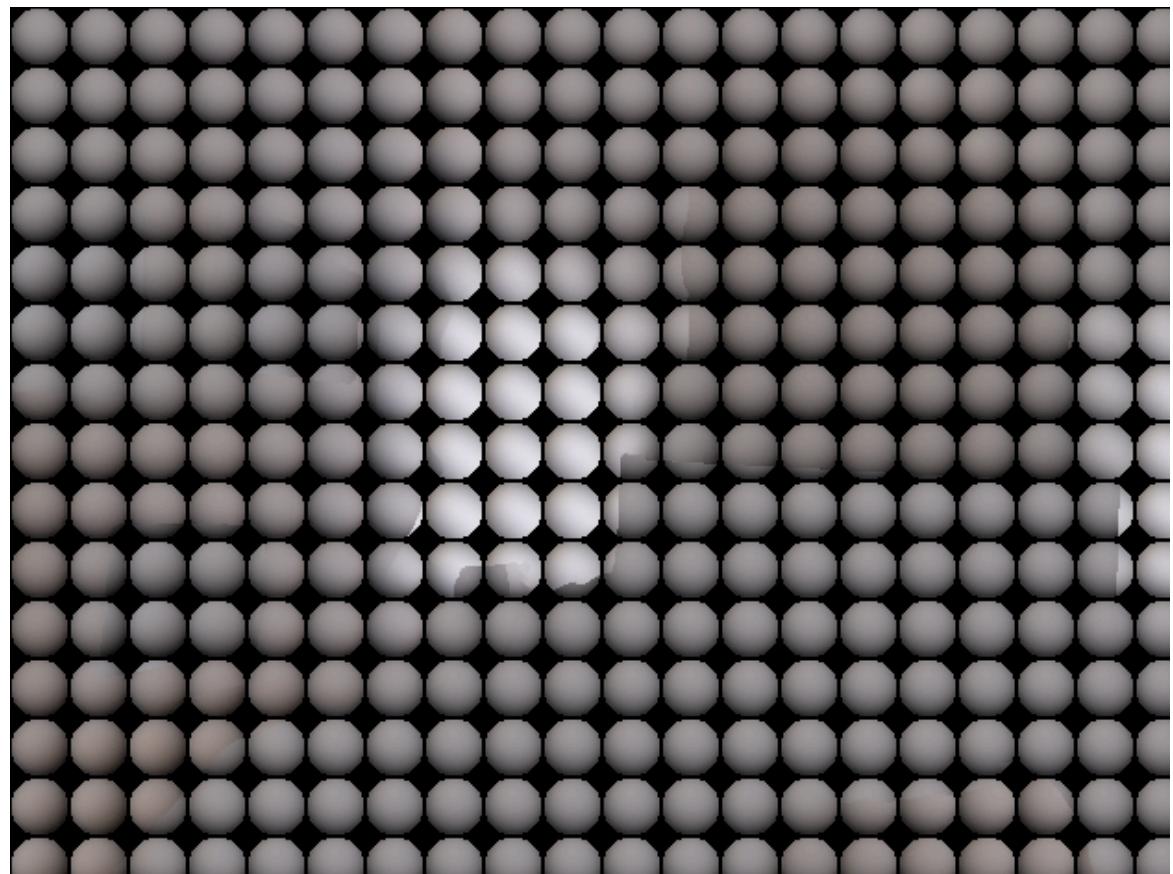
Output



Input Image



Input Shape

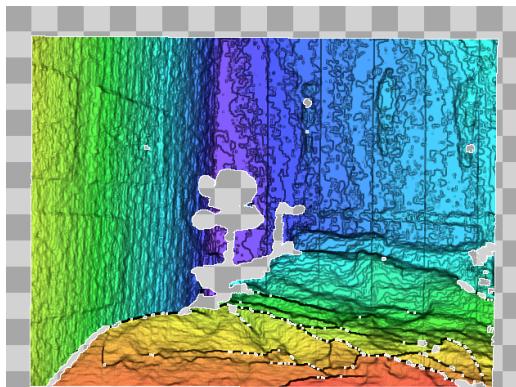


Output Illumination

Output



Input Image



Input Shape

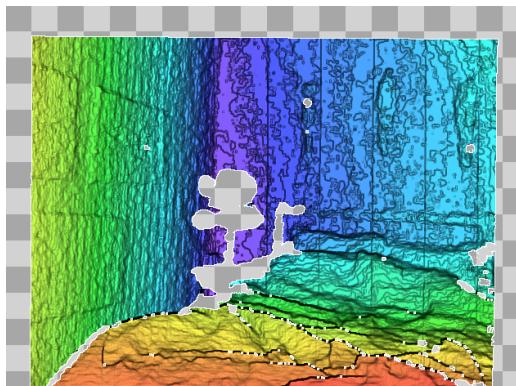


Output Illumination Mixture

Output



Input Image



Input Shape

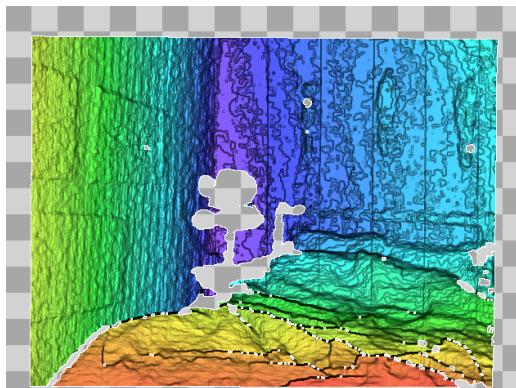


Output Shading

Output



Input Image



Input Shape



Output Reflectance

Input Image



Scene-SIRFS Rotation



Kinect-Only Rotation



Input Image



Scene-SIRFS Relighting



Kinect-Only Relighting



In Conclusion...



SIRFS
(single object)



Scene-SIRFS
(RGB-D scenes +
segmentation)

In Conclusion...



SIRFS
(single object)

Scene-SIRFS
(RGB-D scenes +
segmentation)

Thanks!