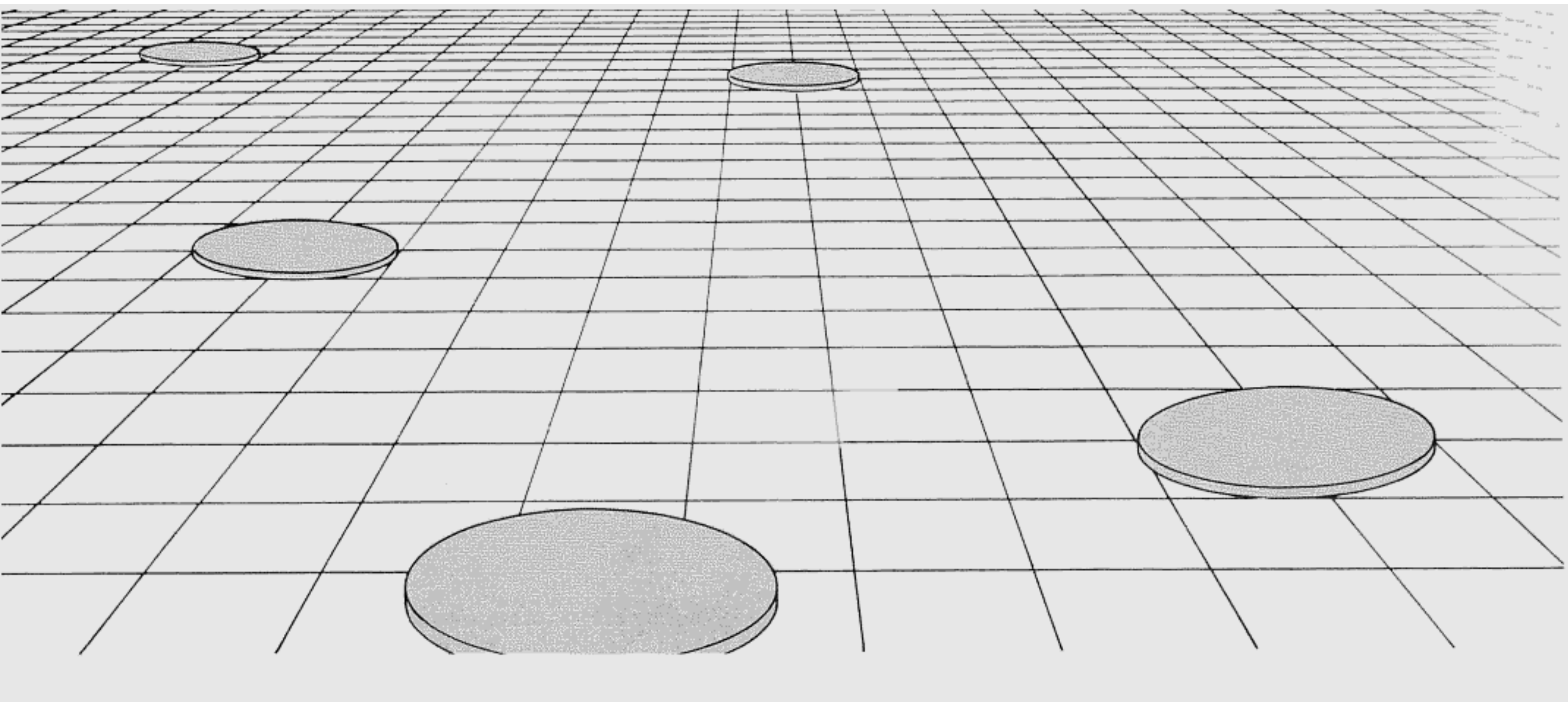
Three-Dimensional Perception from a Single Image

Jitendra Malik

UC Berkeley

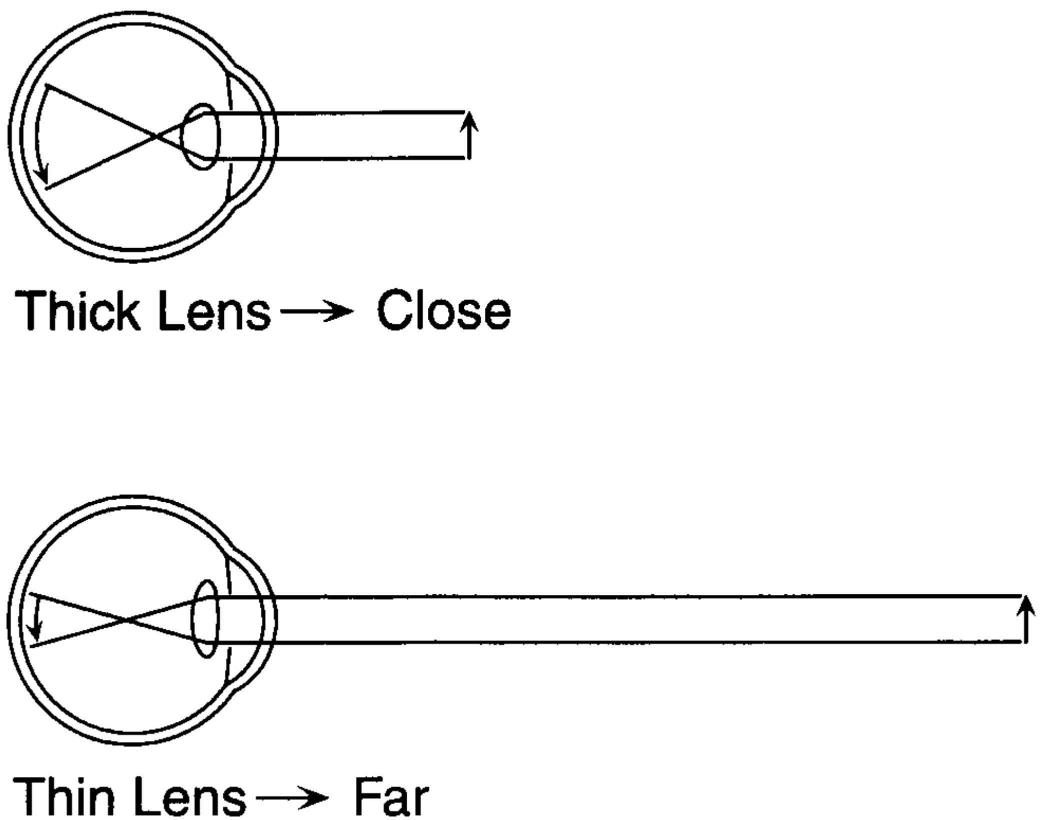
We can perceive depth in a single picture



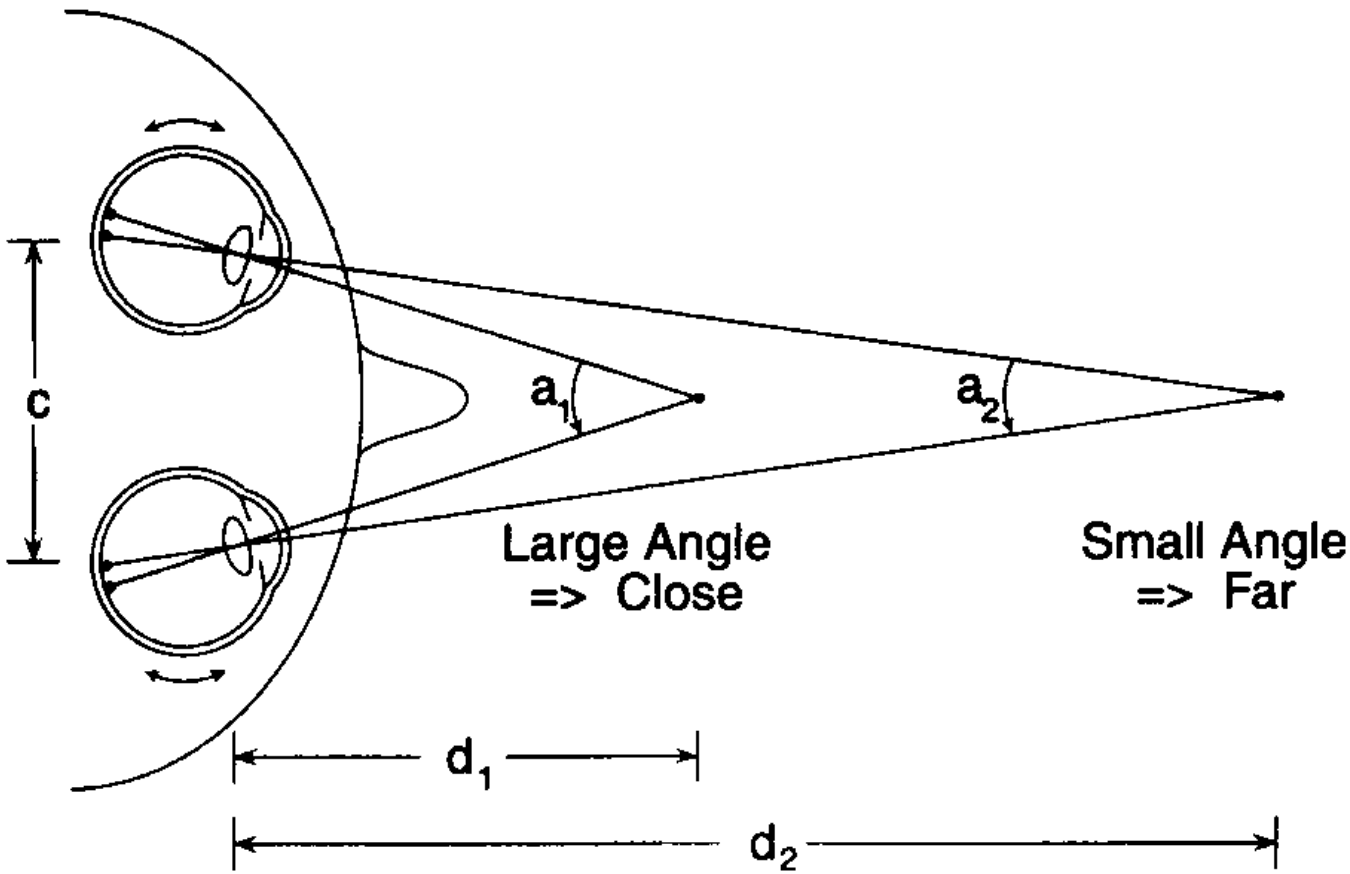
List of cues from Palmer's Vision book

INFORMATION SOURCE	Ocular/ Optical	Binocular/ Monocular	Static/ Dynamic	Relative/ Absolute	Qualitative/ Quantitative
Accommodation	ocular	monocular	static	absolute	quantitative
Convergence	ocular	binocular	static	absolute	quantitative
Binocular Disparity	optical	binocular	static	relative	quantitative
Motion Parallax	optical	monocular	dynamic	relative	quantitative
Texture Accretion/Deletion	optical	monocular	dyanmic	relative	qualitative
Convergence of Parallels	optical	monocular	static	relative	quantitative
Position relative to Horizon	optical	monocular	static	relative	quantitative
Relative Size	optical	monocular	static	relative	quantitative
Familiar Size	optical	monocular	static	absolute	quantitative
Texture Gradients	optical	monocular	static	relative	quantitative
Edge Interpretation	optical	monocular	static	relative	qualitative
Shading and Shadows	optical	monocular	static	relative	qualitative
Aerial Perspective	optical	monocular	static	relative	qualitative

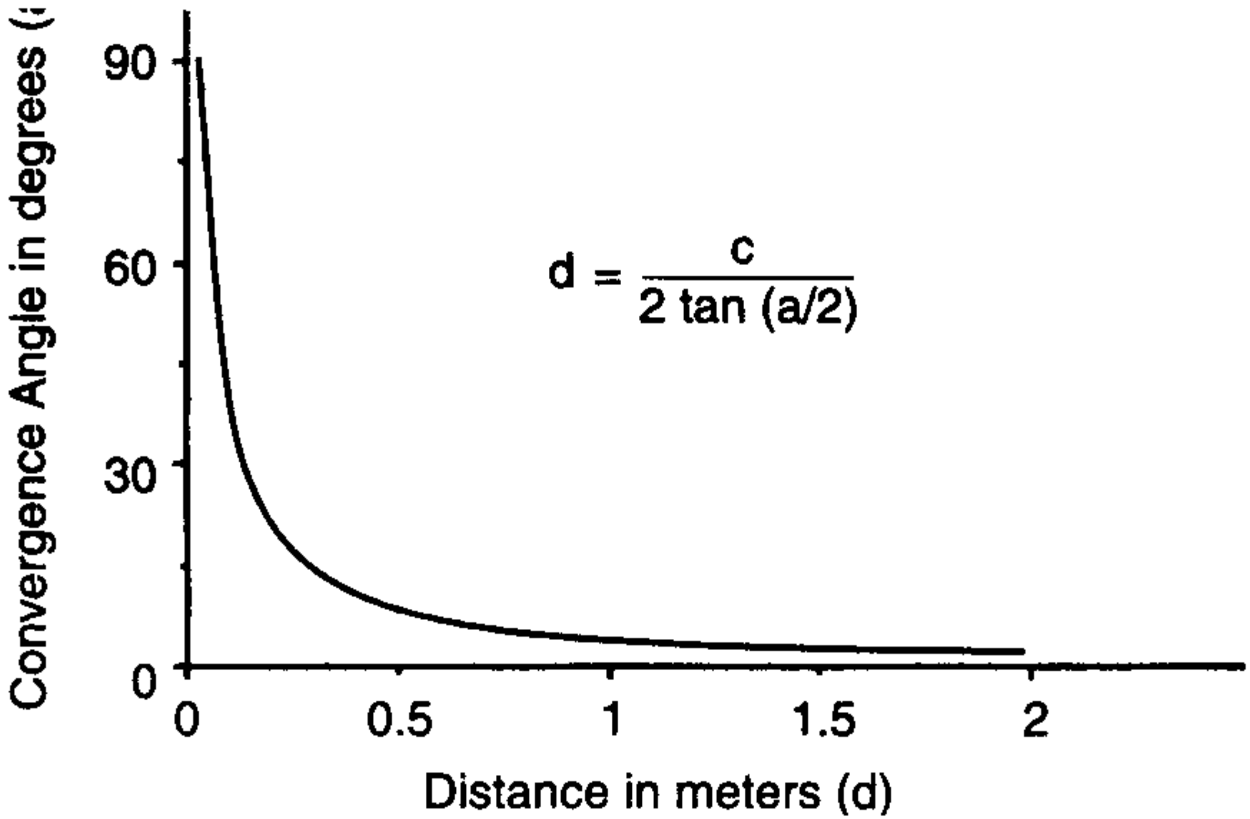
Accomodation/Depth of Focus



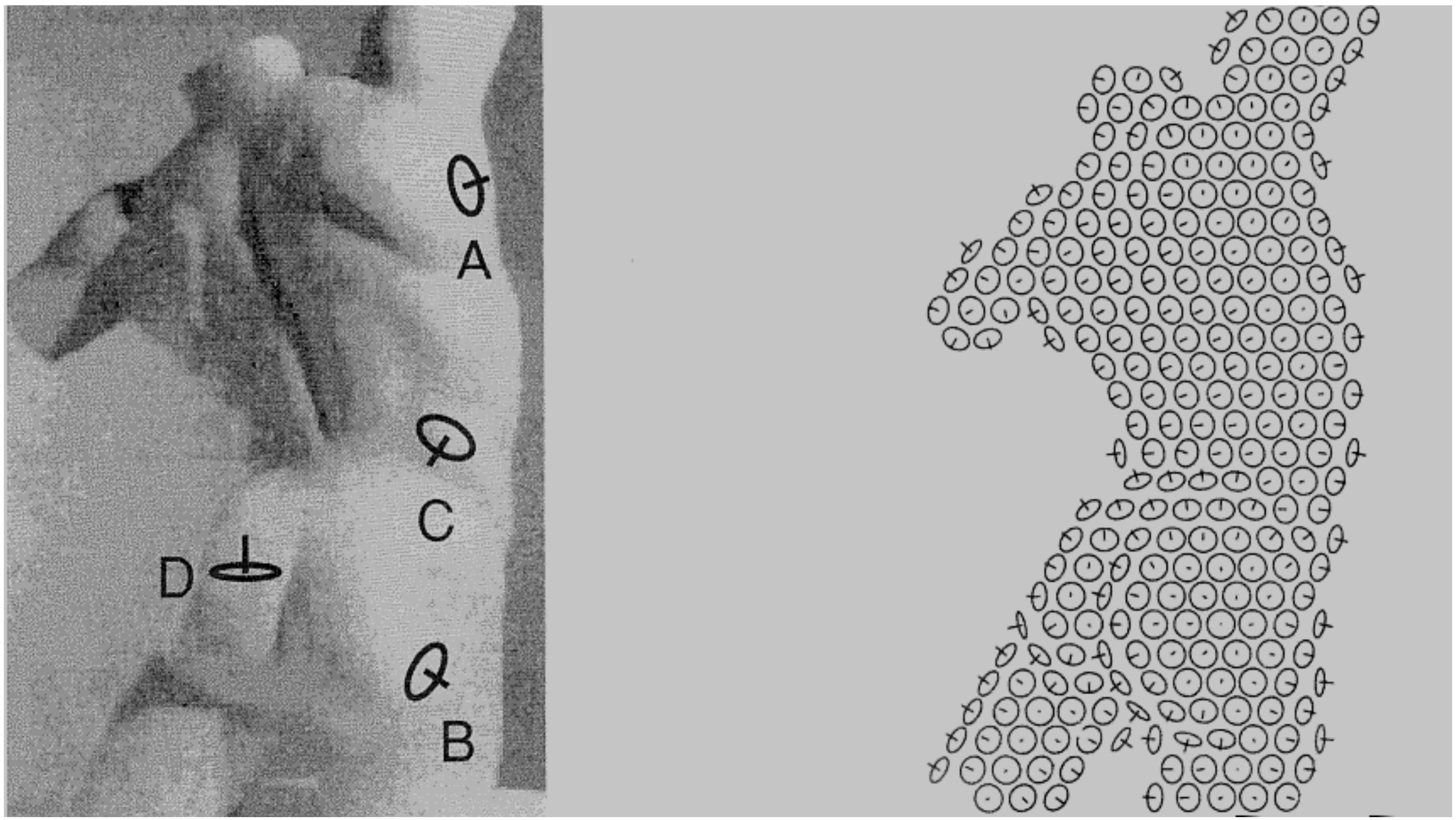
Convergence



Convergence angle vs. distance

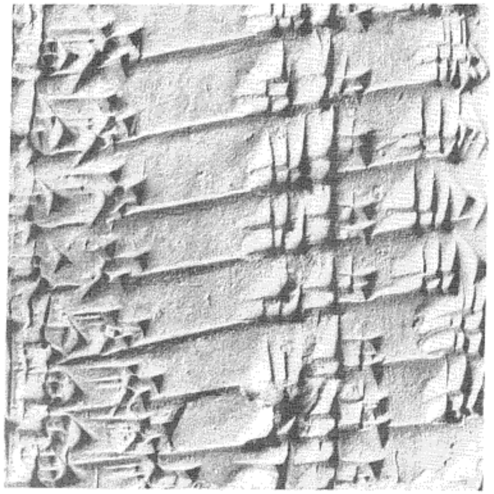


Humans perceive surface normals, not just depth, through a combination of various pictorial cues

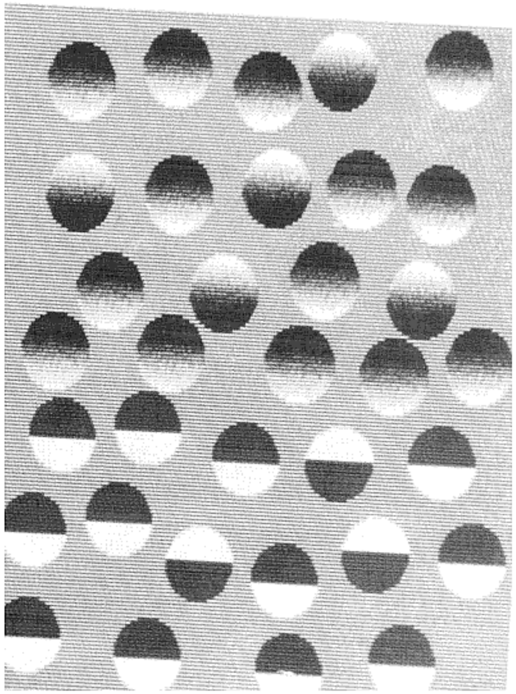


Koenderink, van Doorn and Kappers, 1992

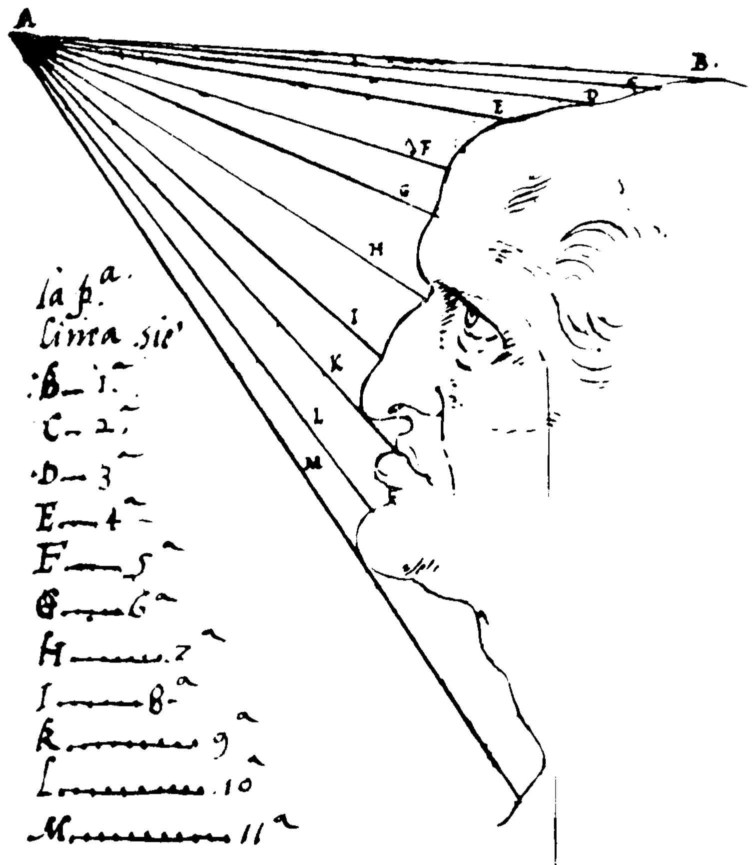
Shape from Shading



Grouping Based on Shape from Shading



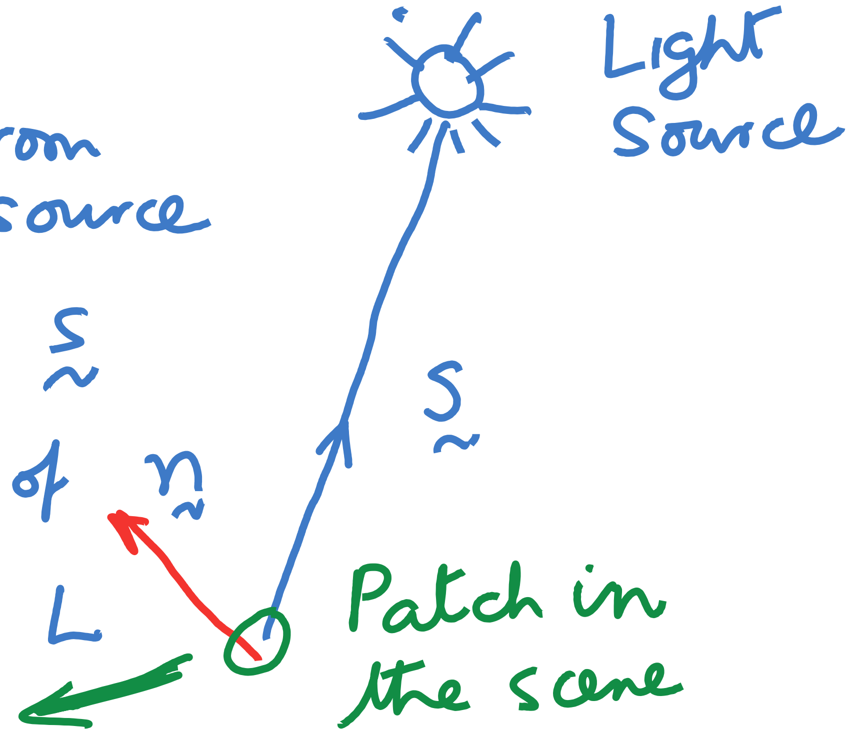
Leonardo thought of it first!



47. Intensities of light on a face, based on Urb 219r.

What causes the outgoing radiance at a scene patch?

- Incoming radiance from the source
- Angle between $\underset{\sim}{n}$ and $\underset{\sim}{s}$
- Reflectance properties of the surface patch

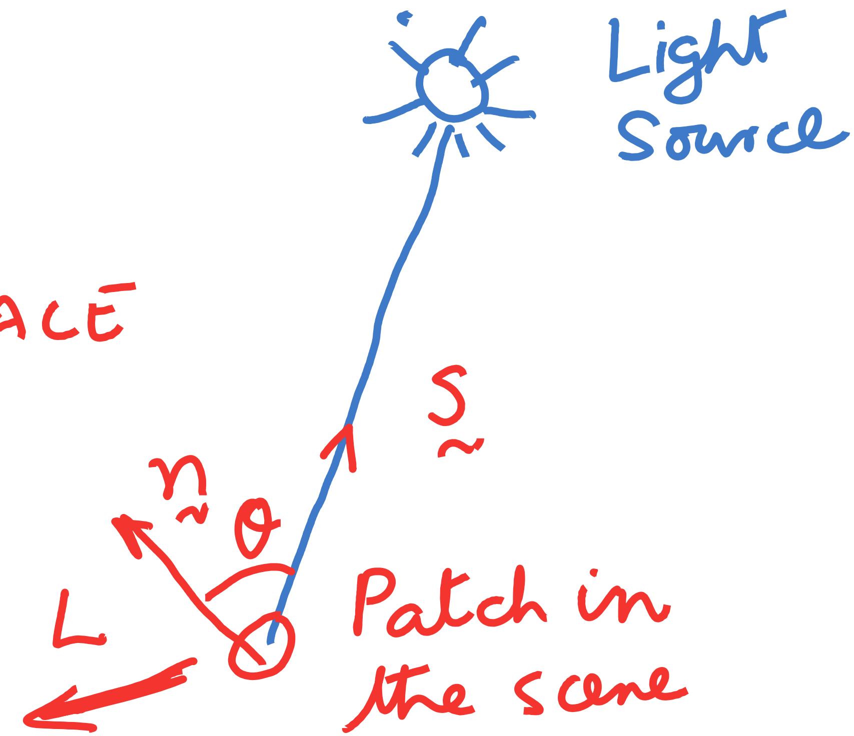


Two special cases:

- Specular surfaces - Outgoing radiance direction obeys angle of incidence=angle of reflection, and co-planarity of incident & reflected rays & the surface normal.
- Lambertian surfaces - Outgoing radiance same in all directions

The Lambertian model

$$L = \rho \lambda \, \underset{\sim}{n} \cdot \underset{\sim}{s}$$



ρ = ALBEDO OF SURFACE

0 $\leftrightarrow$ BLACK

1 $\leftrightarrow$ WHITE

λ depends on radiance of light source

$\underset{\sim}{n} \cdot \underset{\sim}{s} = \cos \theta$ corresponds to foreshortening

We often model reflectance by a combination of a Lambertian term and a specular term. If we want to be precise, we use a BRDF (Bidirectional Reflectance Distribution function) which is a 4D function corresponding to the ratio of outgoing radiance in a particular direction to the incoming irradiance in some other direction. This can be measured empirically.

Shape from Shading

A surface can be described as:

$$Z - f(x, y) = 0$$

For this surface, the surface normal can be expressed as:

$$\hat{n} = \frac{[-f_x; -f_y; 1]}{\sqrt{(1 + f_x^2 + f_y^2)}}$$

The light source is infinitely far away:

$$\hat{s} = [s_x; s_y; s_z]$$

Then radiance in image plane coordinate (x, y):

$$E(x, y) = \rho(\frac{-f_x s_x - f_y s_y - s_z}{\sqrt{(1 + f_x^2 + f_y^2)}})$$

SFS results in a partial differential equation

$$Z = f(x, y)$$

$$f_x = Z_x \qquad \frac{\partial Z}{\partial x} = Z_x(x, y)$$

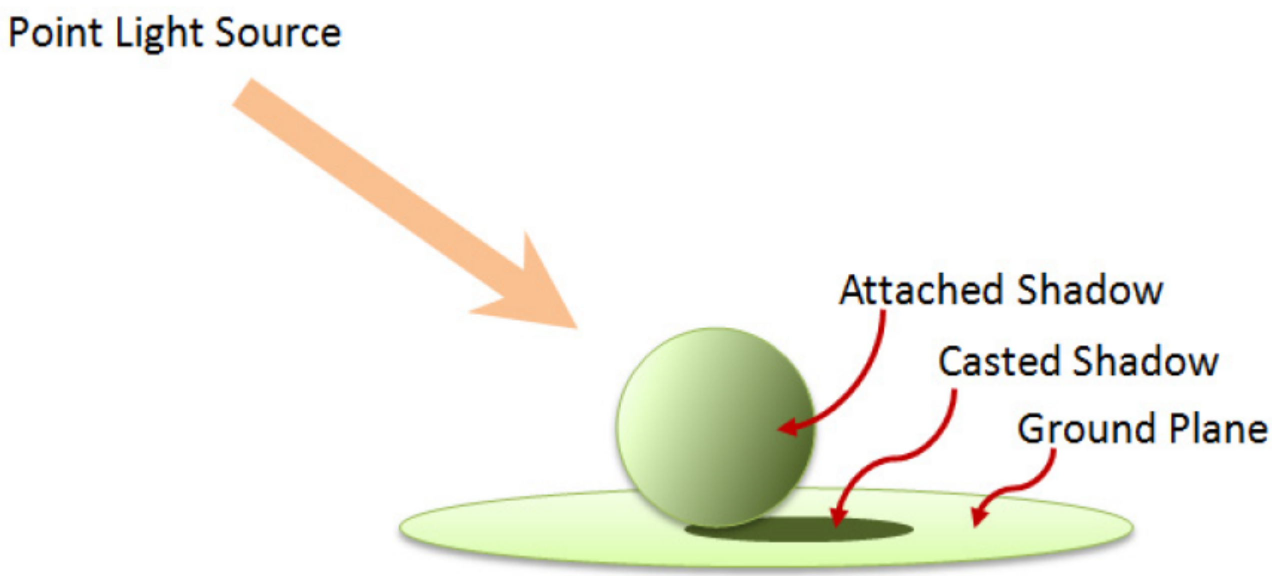
$$f_y = Z_y \qquad \frac{\partial Z}{\partial y} = Z_y(x, y)$$

$$E(x, y) = \rho(\frac{-s_x Z_x(x, y) - s_y Z_y(x, y) - s_z}{\sqrt{(1 + f_x^2 + f_y^2)}})$$

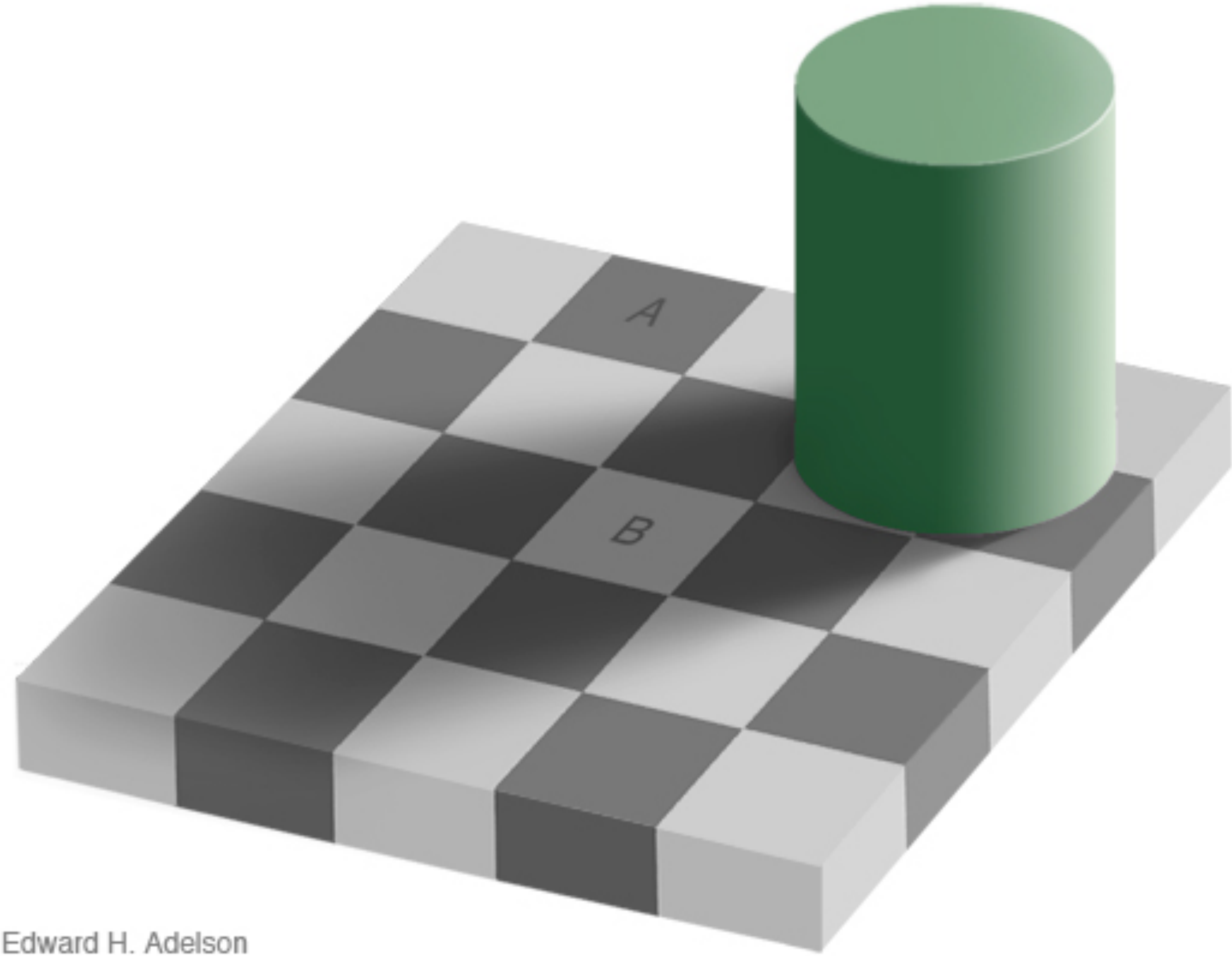
It was solved by Horn in 1970 using characteristic strips.

Real world scenes have additional complexity…

- Objects are illuminated not just by light sources, but also by reflected light from other surfaces. In computer graphics, ray tracing and radiosity are techniques that address this issue.
- Shadows

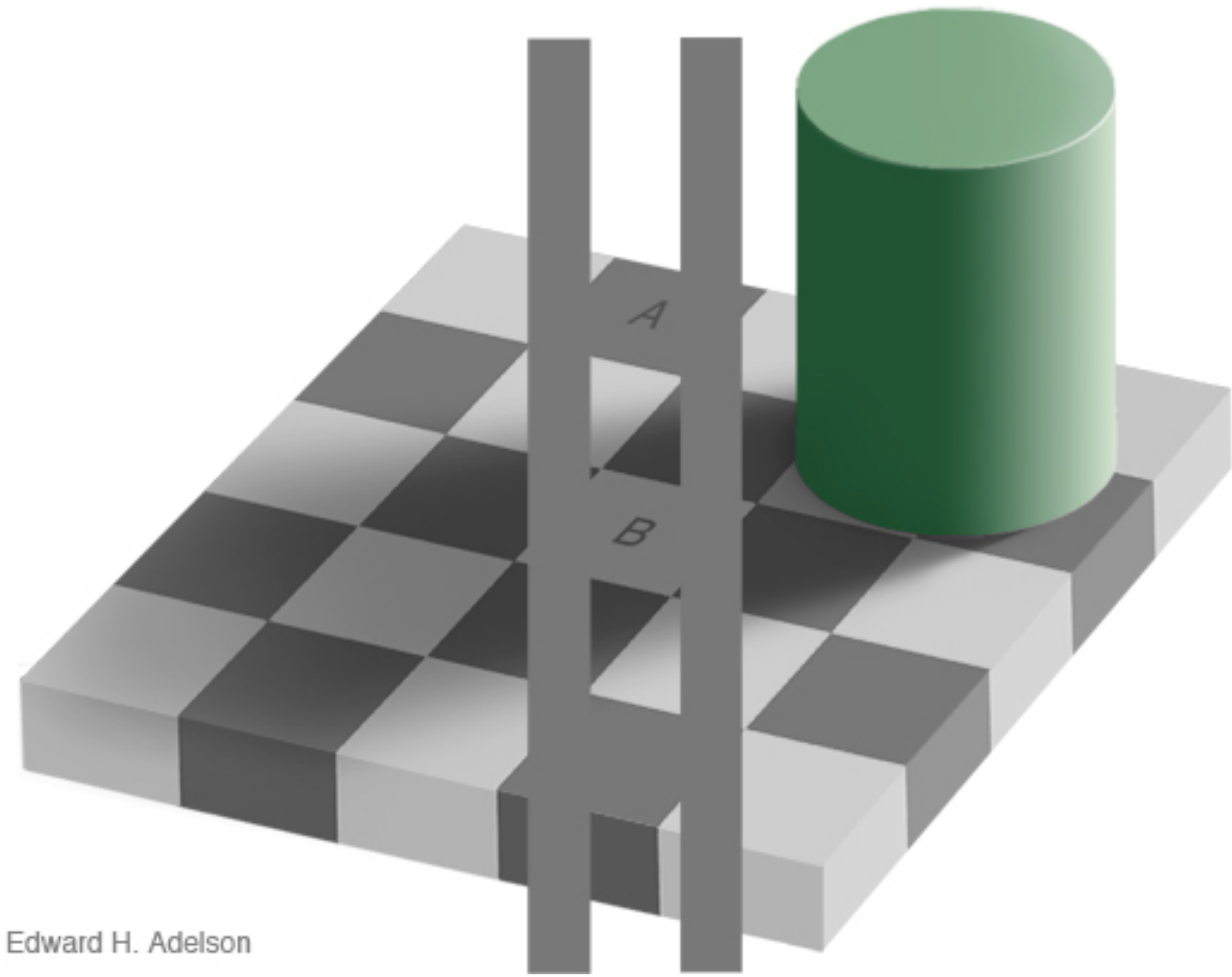


Adelson’s checkershadow

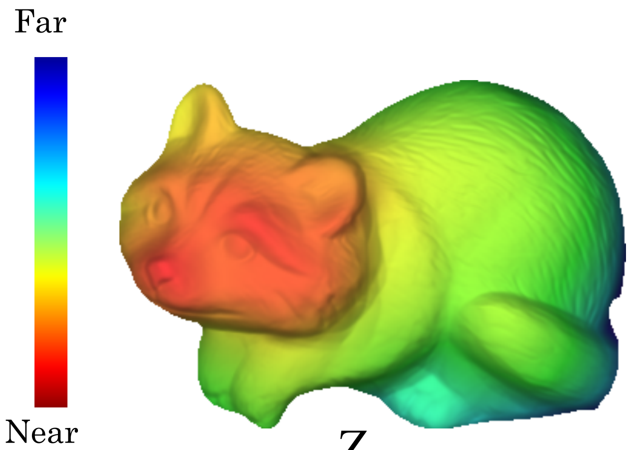


Edward H. Adelson

A and B have same luminance!

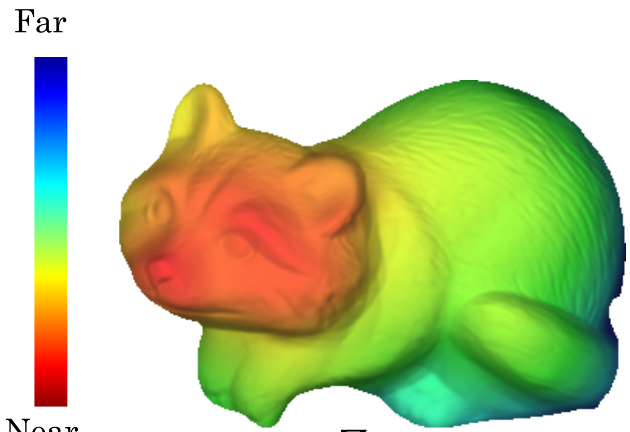


Forward Optics



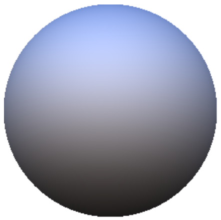
Z
shape / depth

Forward Optics



Z

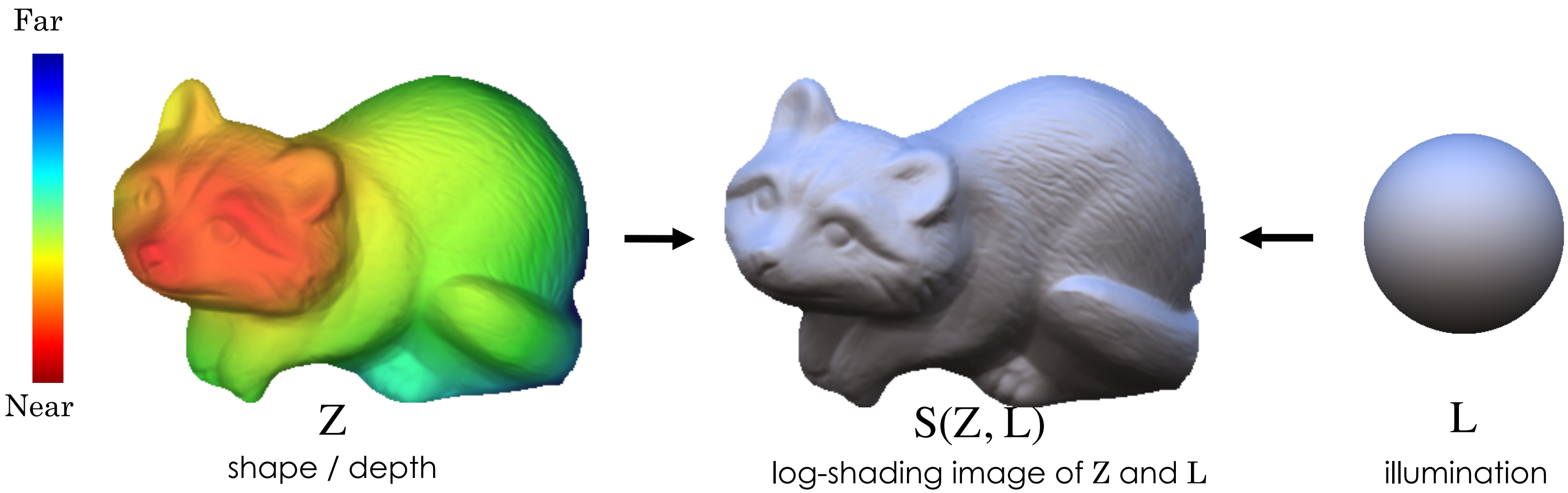
shape / depth



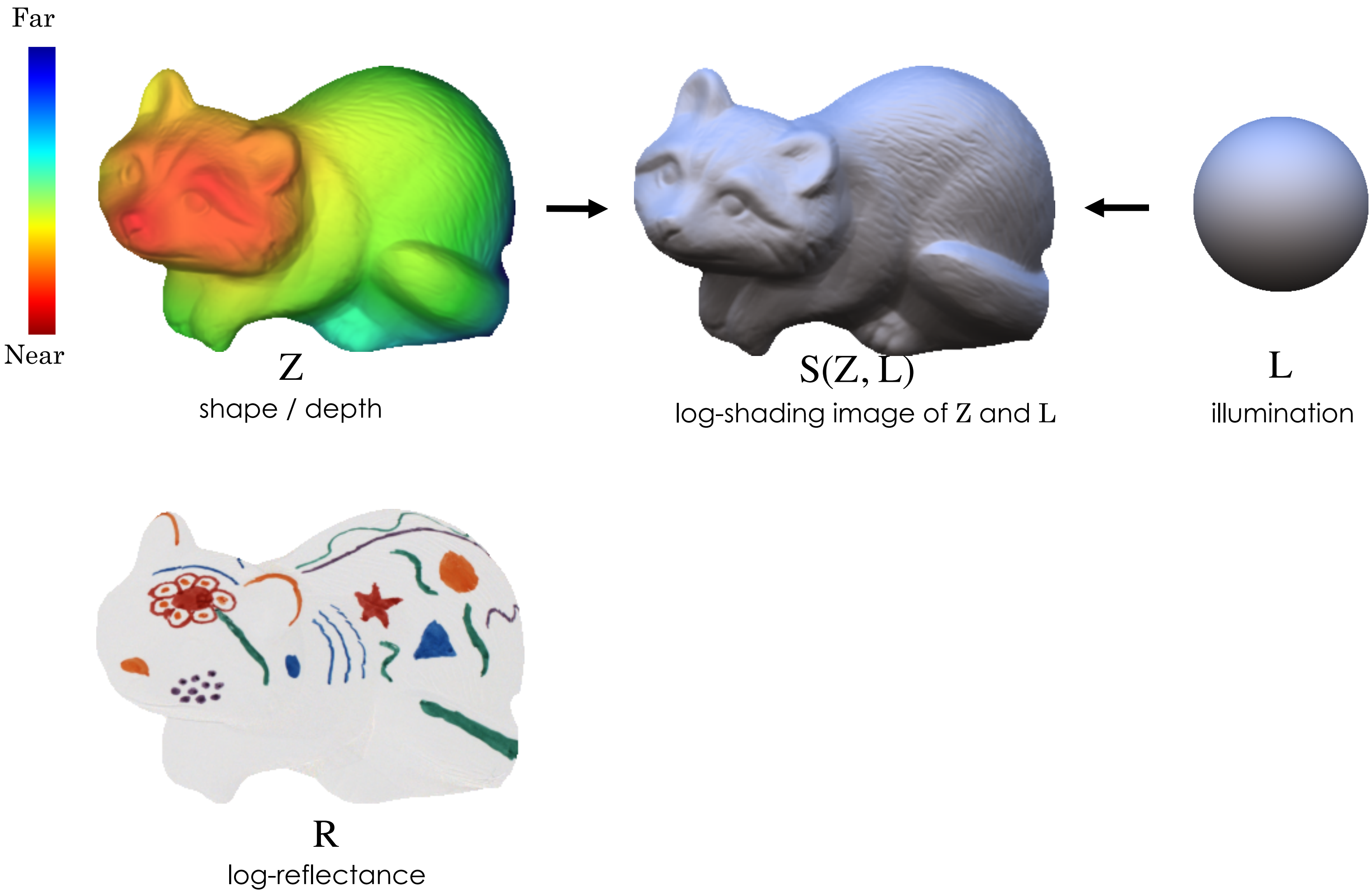
L

illumination

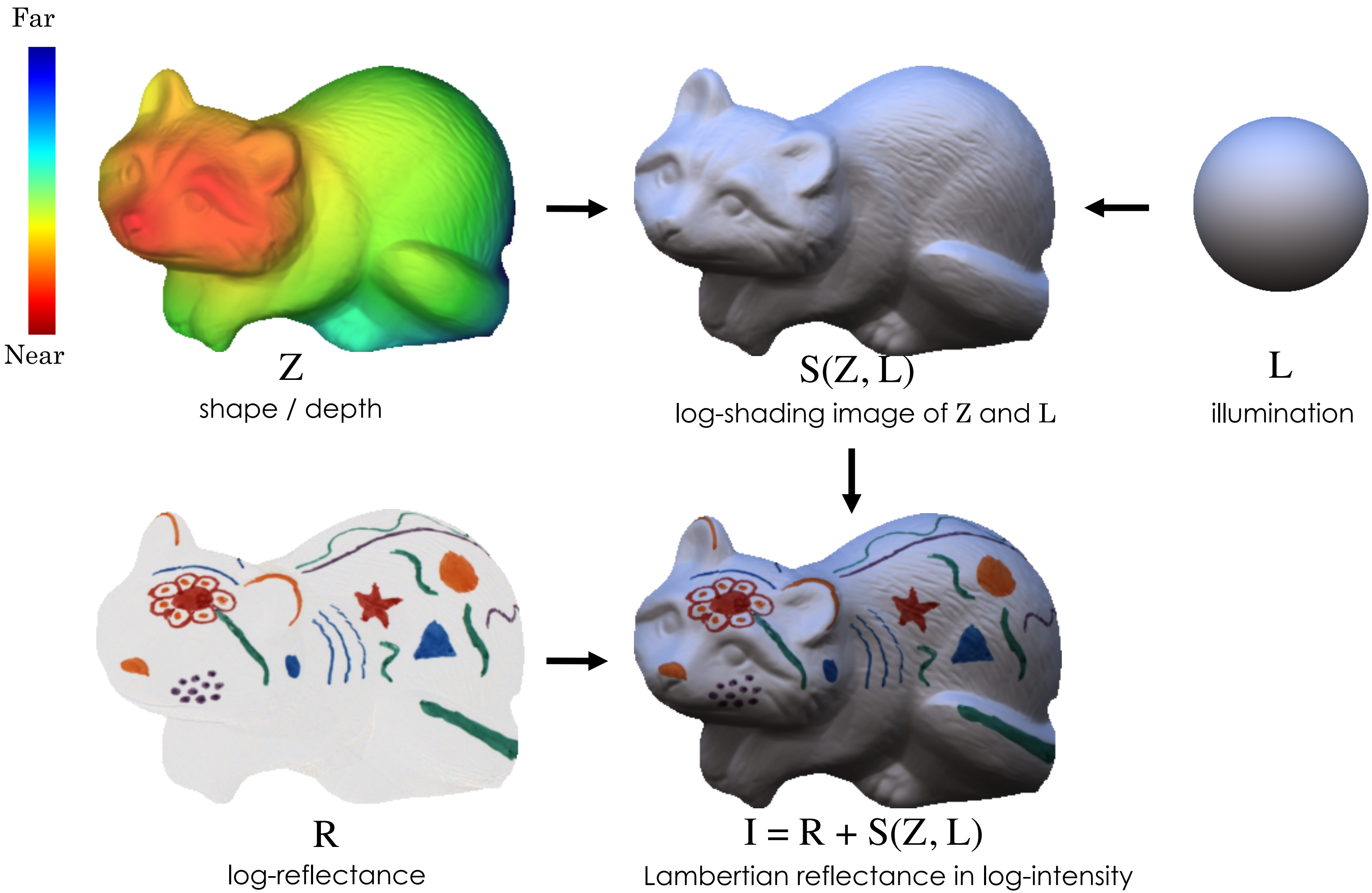
Forward Optics



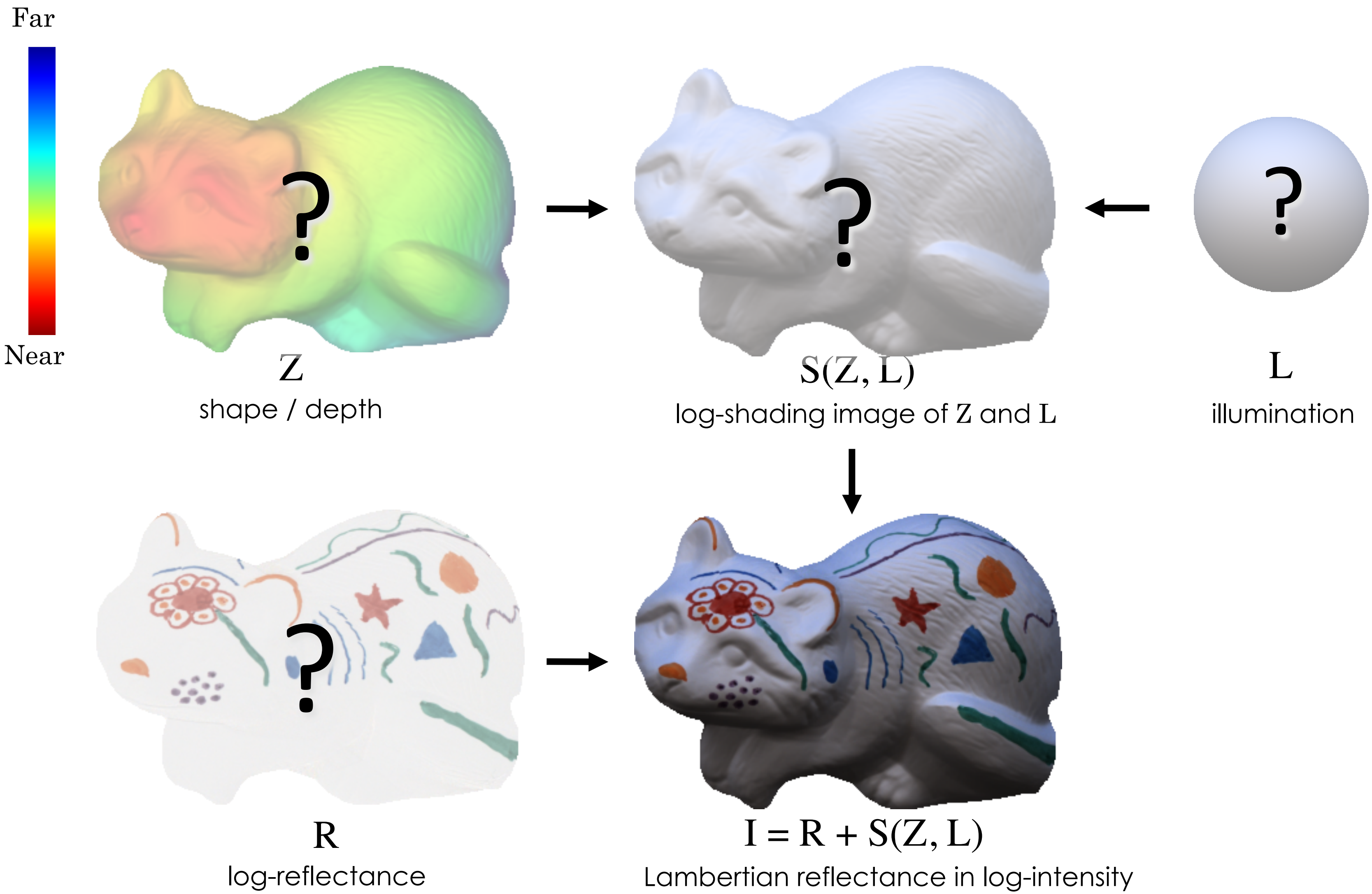
Forward Optics



Forward Optics

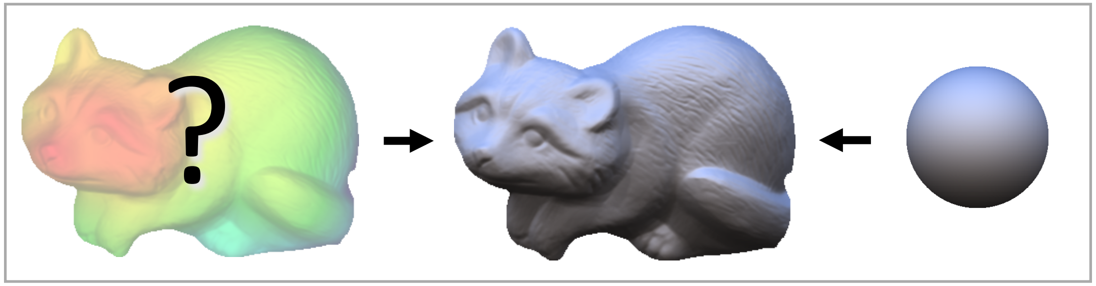


Our problem


Far
Near
?
Z
shape / depth
?
S(Z, L)
log-shading image of Z and L
?
L
illumination
?
R
log-reflectance
I = R + S(Z, L)
Lambertian reflectance in log-intensity

Past Work

Past Work: Shape from Shading



Basic Assumption: illumination and albedo are known.

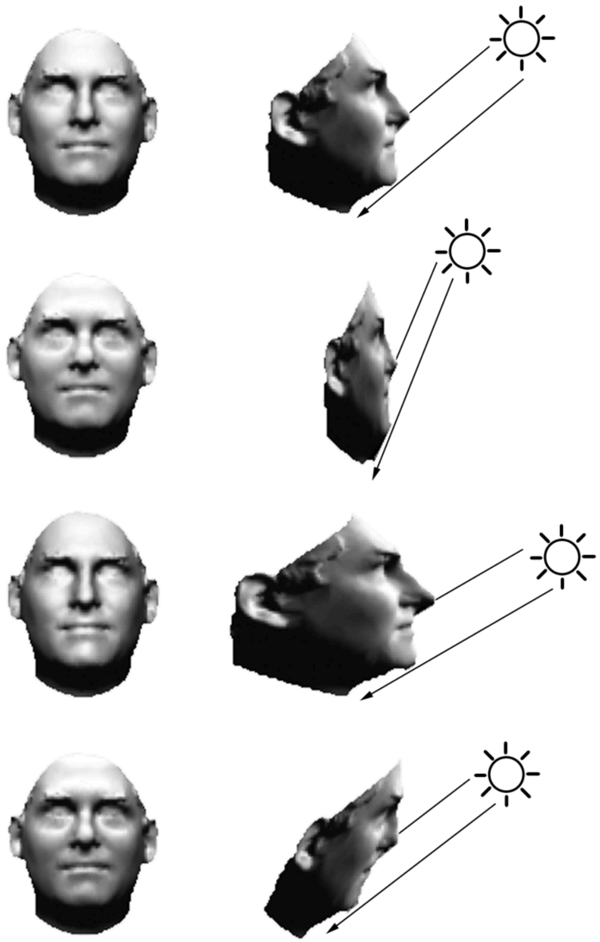
If the reflectivity function is $\phi(I,E,G)$, the normalized incident light intensity at the point $\underset{\sim}{r} = (x,y,z)$ is $A(\underset{\sim}{r})$ and the intensity at the corresponding image point $\underset{\sim}{r}' = (x',y',f)$ is $b(\underset{\sim}{r}')$, then:

$$A(\underset{\sim}{r})\ \phi(I,E,G) = b(\underset{\sim}{r}')$$

This image illumination equation is the main equation studied here. When finding a solution we assume $A(\underset{\sim}{r})$ and $\phi(I,E,G)$ are known and $b(\underset{\sim}{r}')$ is obtained from the image. We want to

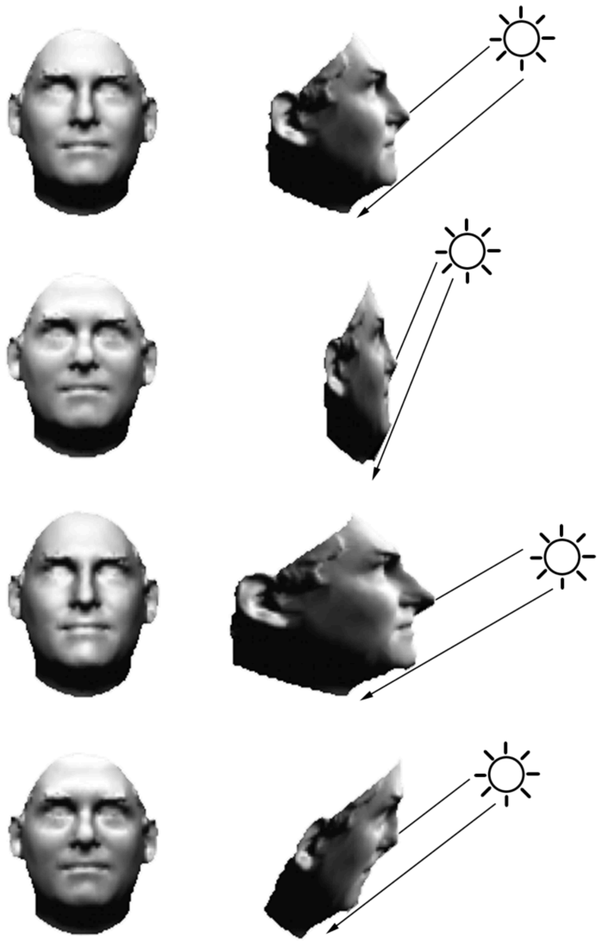
B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, MIT, 1970.

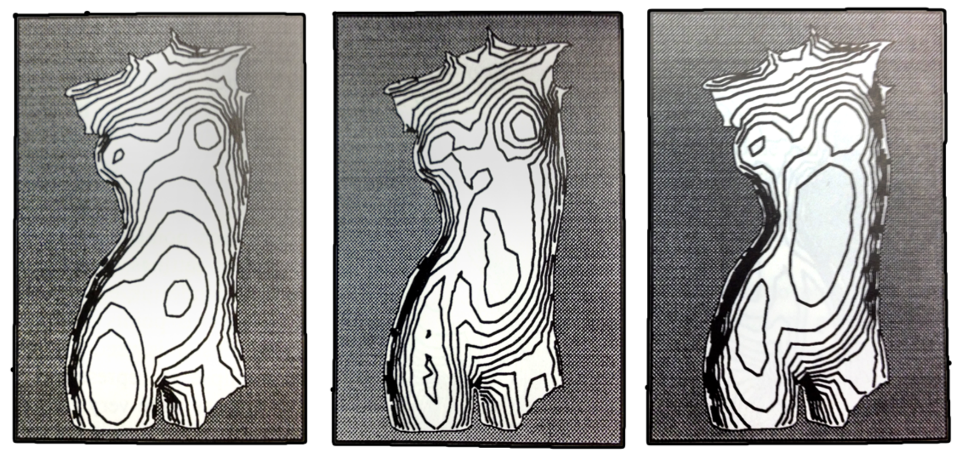
Past Work: Shape from Shading



P. Belhumeur, D. Kriegman, and A. Yuille.
The Bas-Relief Ambiguity. *IJCV*, 1999.

Past Work: Shape from Shading





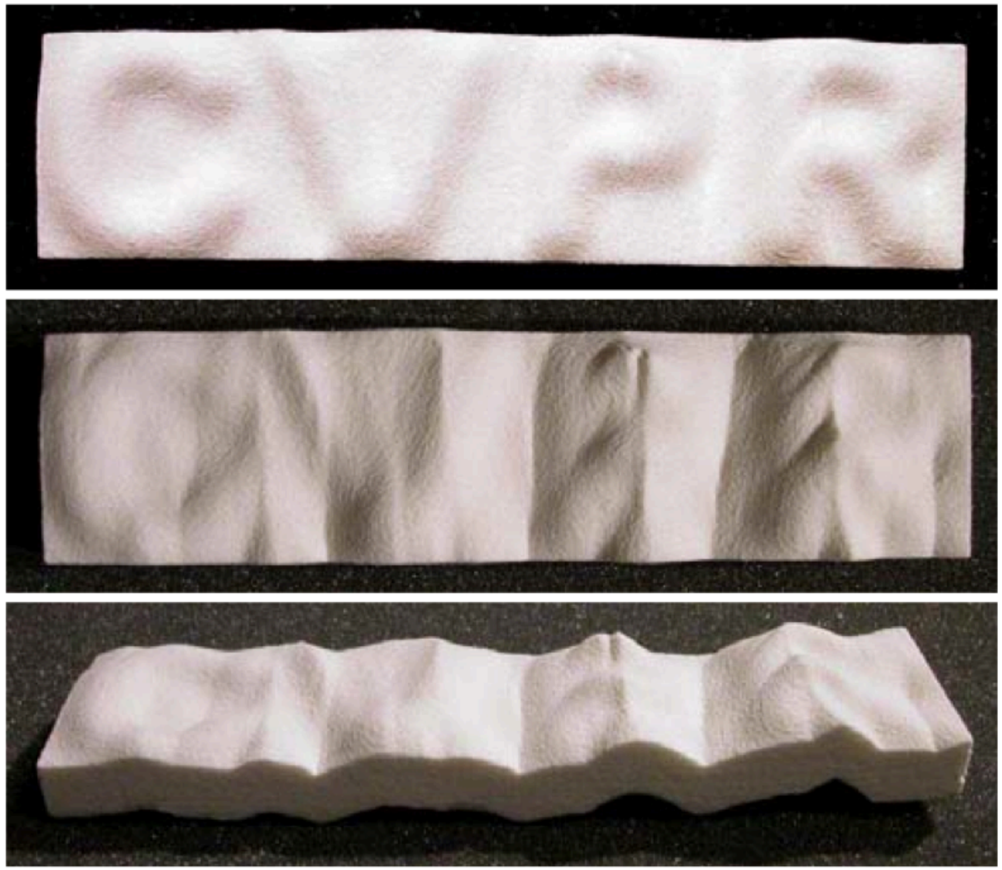
P. Belhumeur, D. Kriegman, and A. Yuille. The Bas-Relief Ambiguity. *IJCV*, 1999.

J. Koenderink, A. van Doorn, C. Christou, and J. Lappin. Shape constancy in pictorial relief. *Perception*, 1996.

Past Work: Shape from Shading

Ecker & Jepson, Polynomial Shape from Shading, *CVPR* 2010

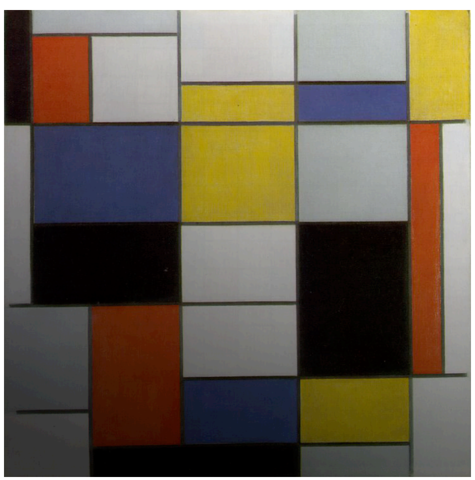
Past Work: Shape from Shading



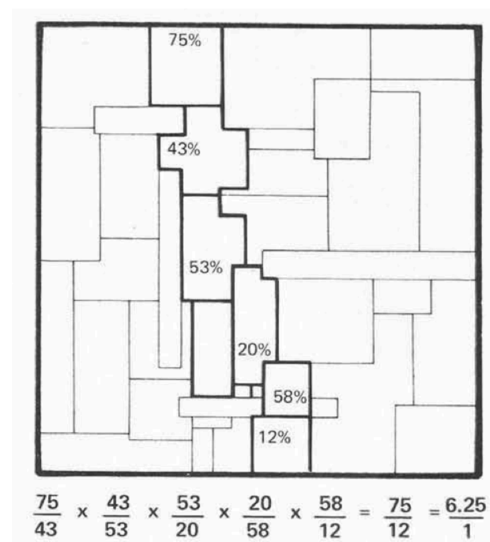
Ecker & Jepson, Polynomial Shape from Shading, *CVPR* 2010

Past Work: Lightness Recovery

Basic assumption: Shape is ignored, illumination varies slowly, therefore all edges are reflectance edges.

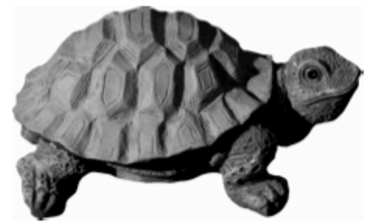


Piet Mondrian, Composition A. Oil on Canvas, 1920.



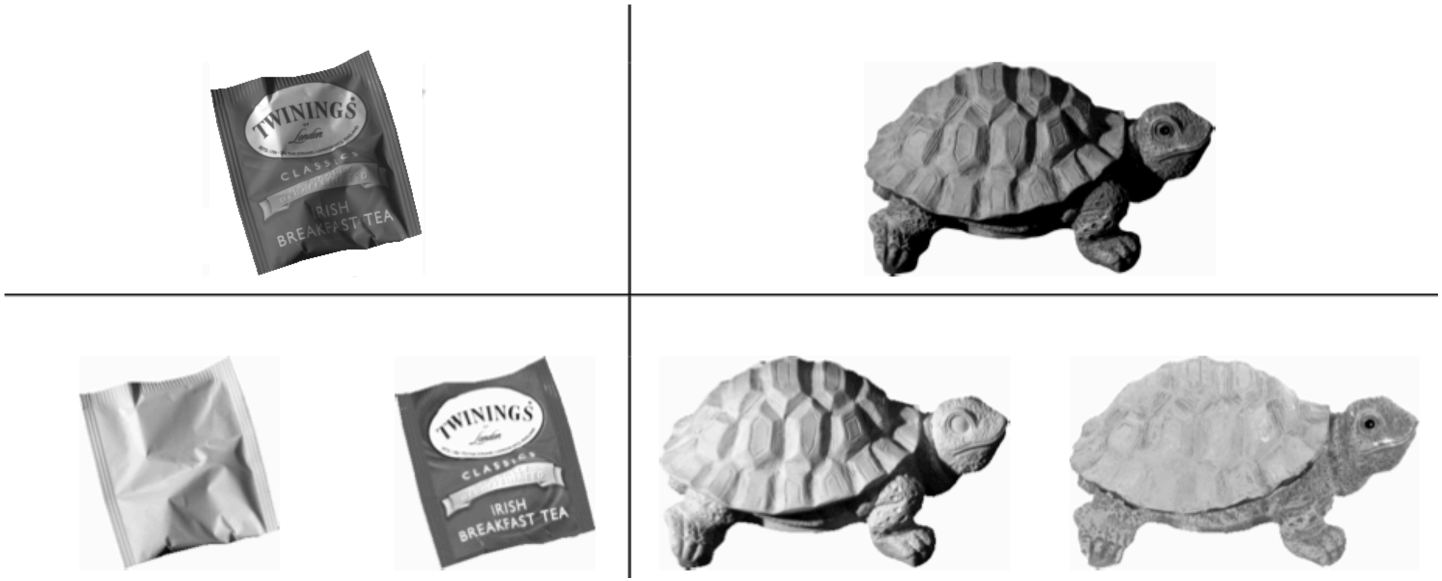
E. H. Land and J. J. McCann. Lightness and retinex theory. *JOSA*, 1971.

Past Work: Lightness Recovery



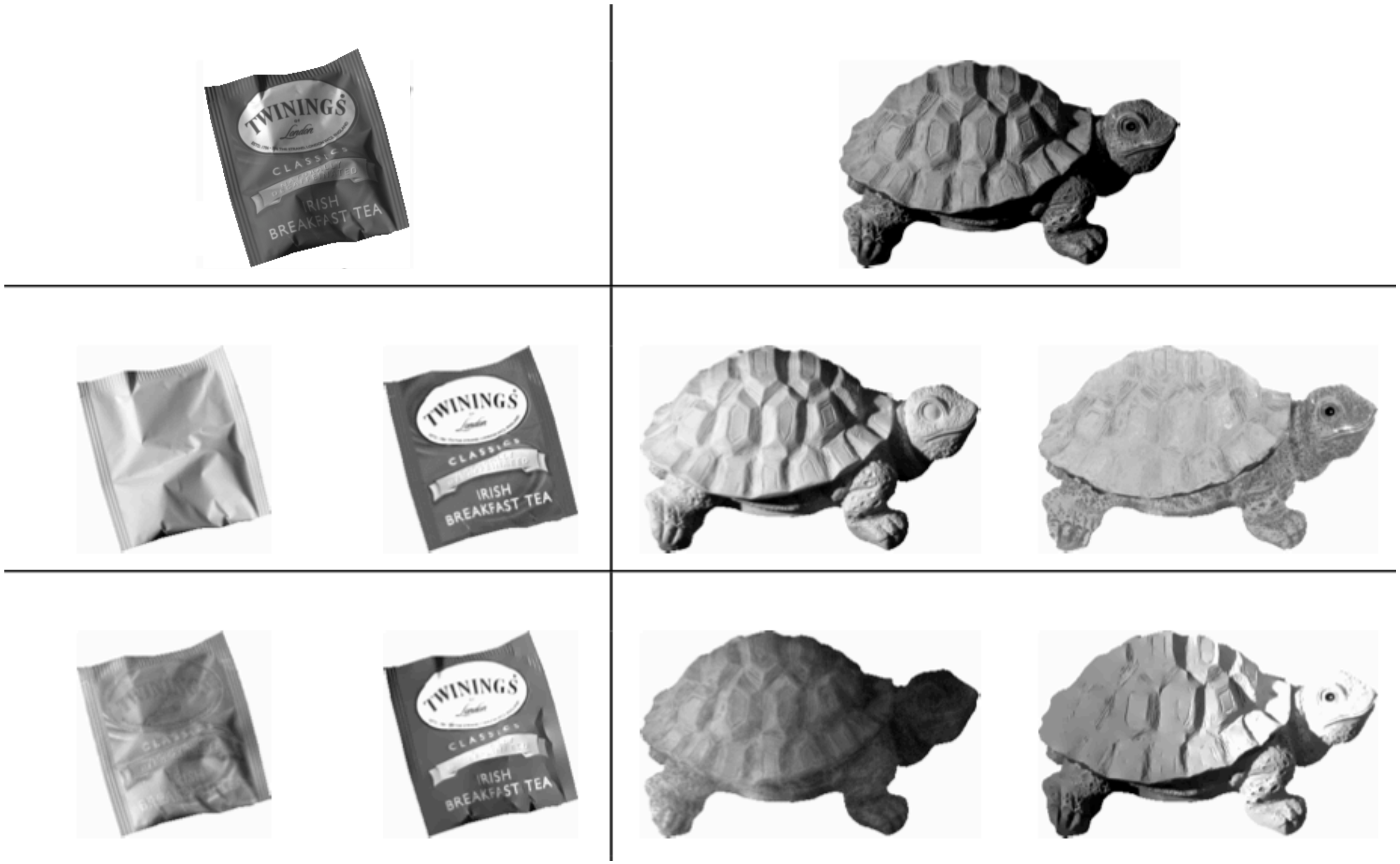
Horn. Determining lightness from an image. CGIP, 1974
Grosse et el., Ground-truth dataset and baseline evaluations for intrinsic image algorithms, ICCV, 2009

Past Work: Lightness Recovery



Horn. Determining lightness from an image. CGIP, 1974
Grosse et el., Ground-truth dataset and baseline evaluations for intrinsic image algorithms, ICCV, 2009

Past Work: Lightness Recovery



Horn. Determining lightness from an image. CGIP, 1974
Grosse et el., Ground-truth dataset and baseline evaluations for intrinsic image algorithms, ICCV, 2009

Past Work: Color Constancy

Past Work: Natural Image Statistics

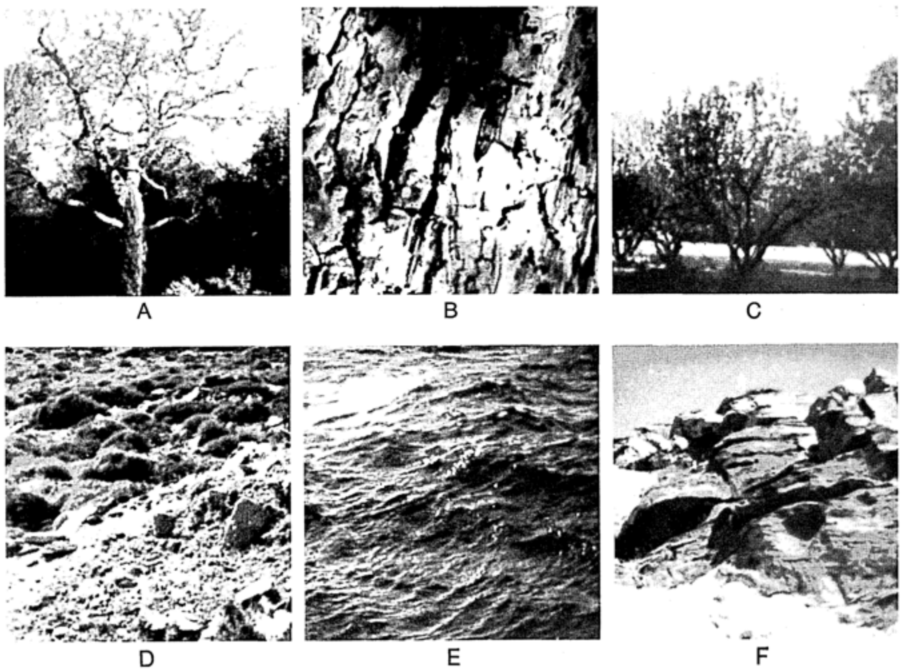


Fig. 6. Examples of the six images (A–F) in this study. Each image consists of 256 × 256 pixels with 256 gray levels (8 bits). However, only the central region was directly analyzed (160 × 160). See the text or details.

D. Field. Relations between the statistics of natural images and the response properties of cortical cells. JOSA A, 1987.

Past Work: Natural Image Statistics

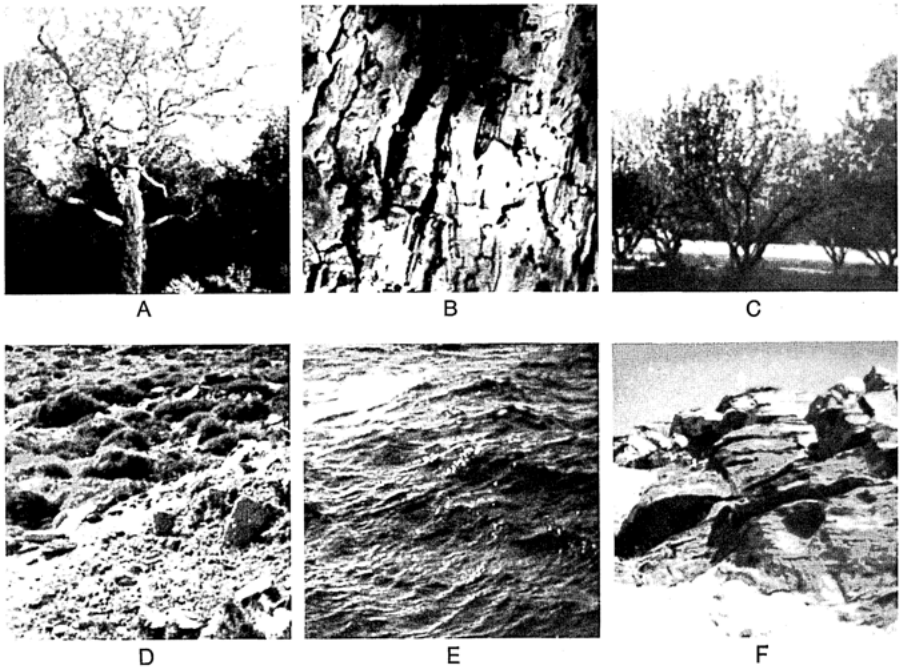


Fig. 6. Examples of the six images (A–F) in this study. Each image consists of 256 × 256 pixels with 256 gray levels (8 bits). However, only the central region was directly analyzed (160 × 160). See the text or details.

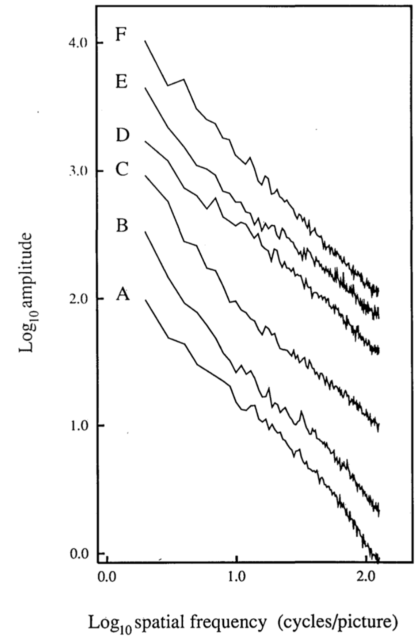


Fig. 8. Amplitude spectra for the six images A–F, averaged across all orientations. The spectra have been shifted up for clarity. On these log–log coordinates the spectra fall off by a factor of roughly $1/f$ (a slope of −1). Therefore the power spectra fall off as $1/f^2$.

D. Field. Relations between the statistics of natural images and the response properties of cortical cells. JOSA A, 1987.

Past Work: Natural Image Statistics

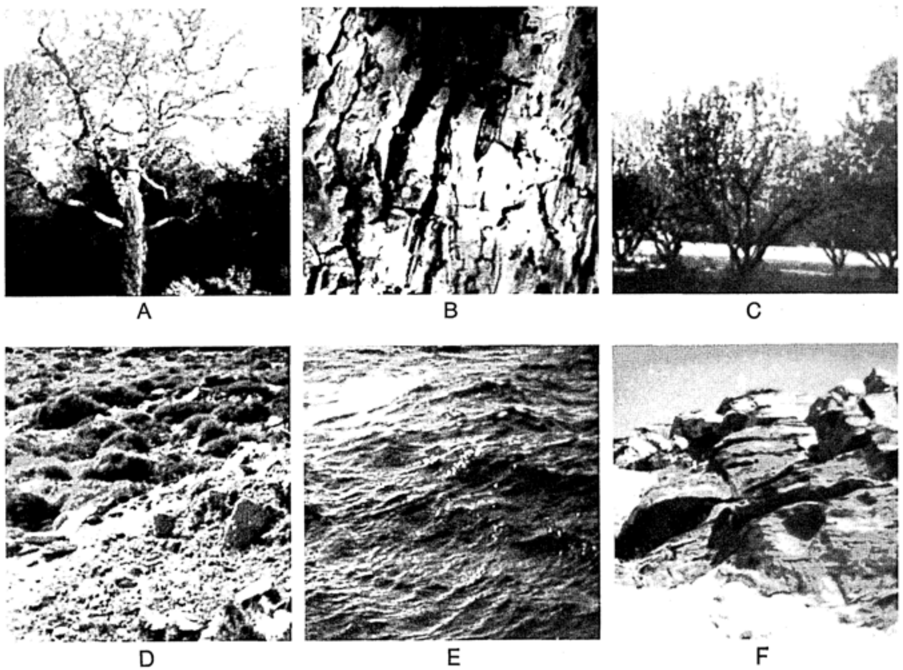


Fig. 6. Examples of the six images (A–F) in this study. Each image consists of 256 × 256 pixels with 256 gray levels (8 bits). However, only the central region was directly analyzed (160 × 160). See the text or details.

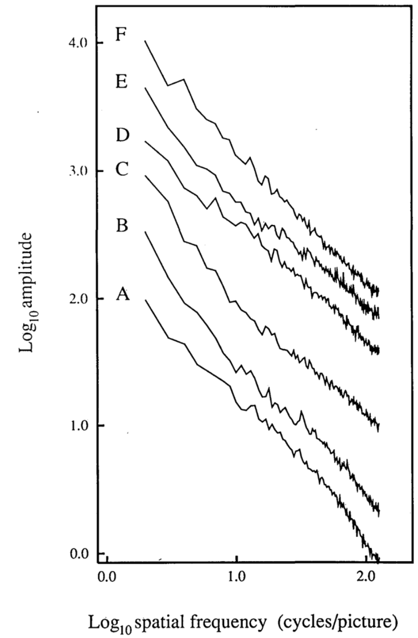


Fig. 8. Amplitude spectra for the six images A–F, averaged across all orientations. The spectra have been shifted up for clarity. On these log–log coordinates the spectra fall off by a factor of roughly $1/f$ (a slope of –1). Therefore the power spectra fall off as $1/f^2$.

D. Field. Relations between the statistics of natural images and the response properties of cortical cells. JOSA A, 1987.

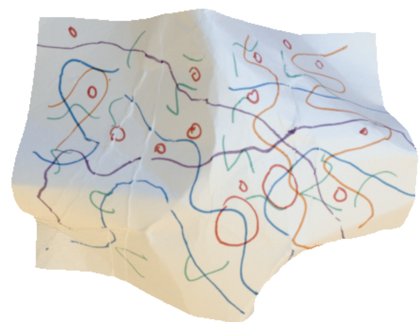
Statistical regularities arise in natural images (mostly) because of statistical regularities in natural environments!

Our Work

Shape, Illumination and Reflectance from Shading

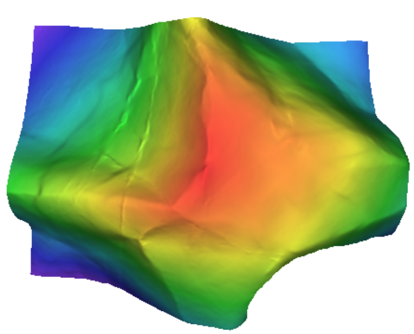
Barron & Malik, CVPR 2011, CVPR 2012, ECCV 2012

Input:

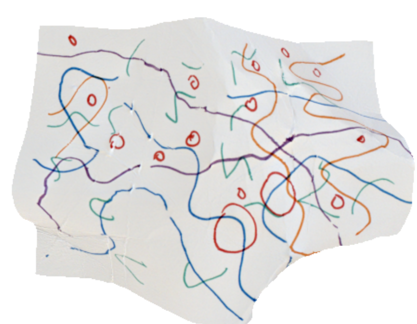


Image

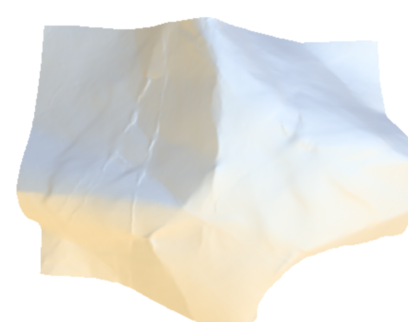
Output:



Shape



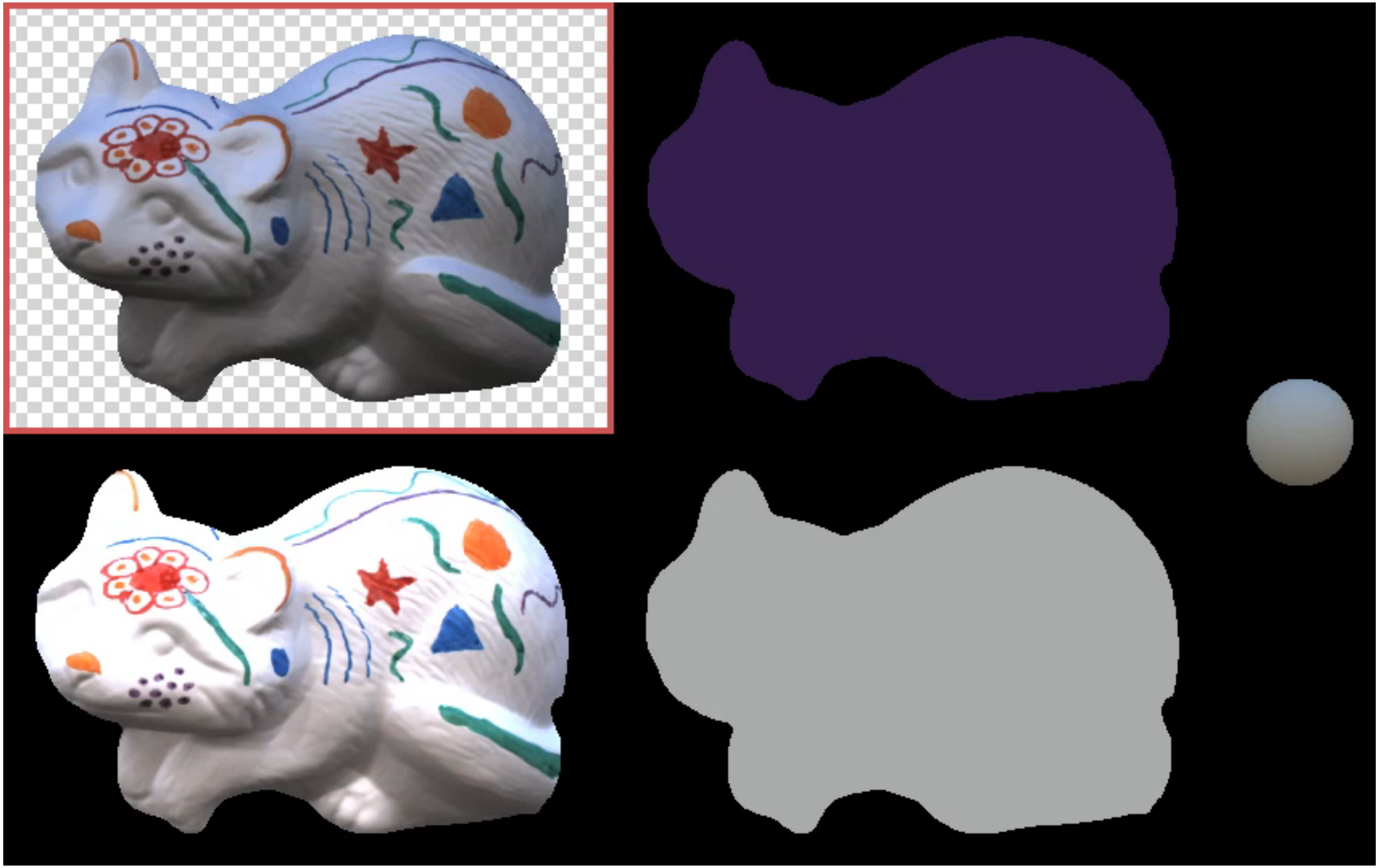
Albedo



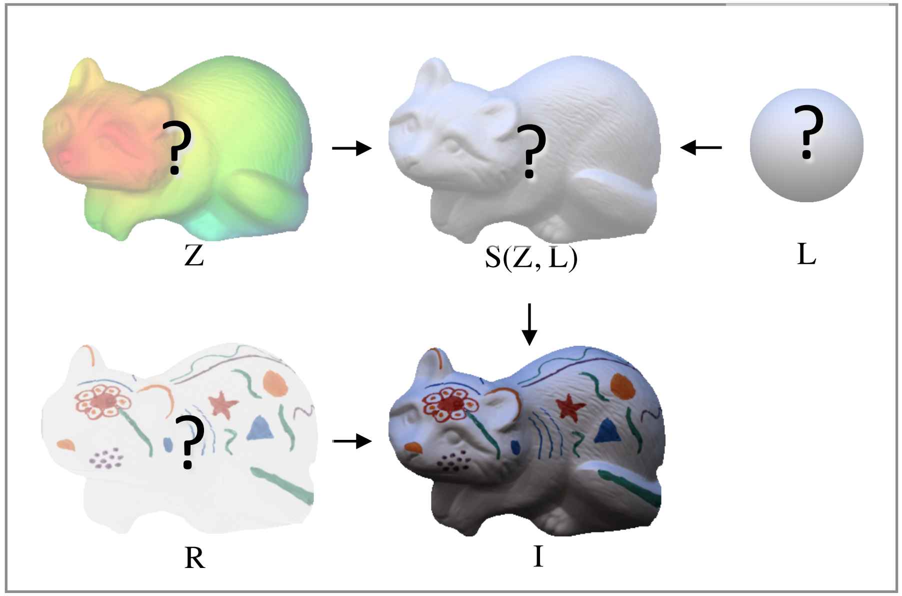
Shading

Illumination

Demo!



Problem Formulation

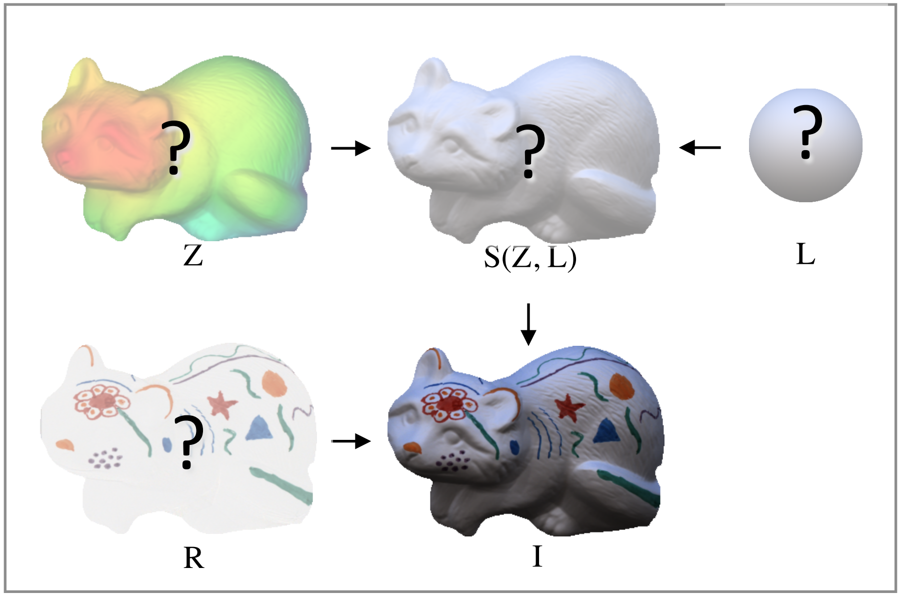


$$\underset{Z,R,L}{\text{maximize}} \quad P(R)P(Z)P(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the most likely explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Problem Formulation



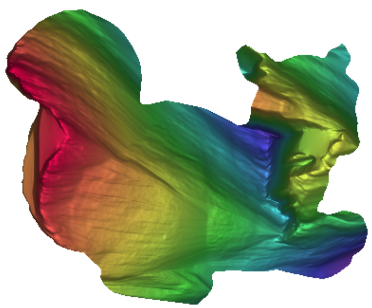
$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

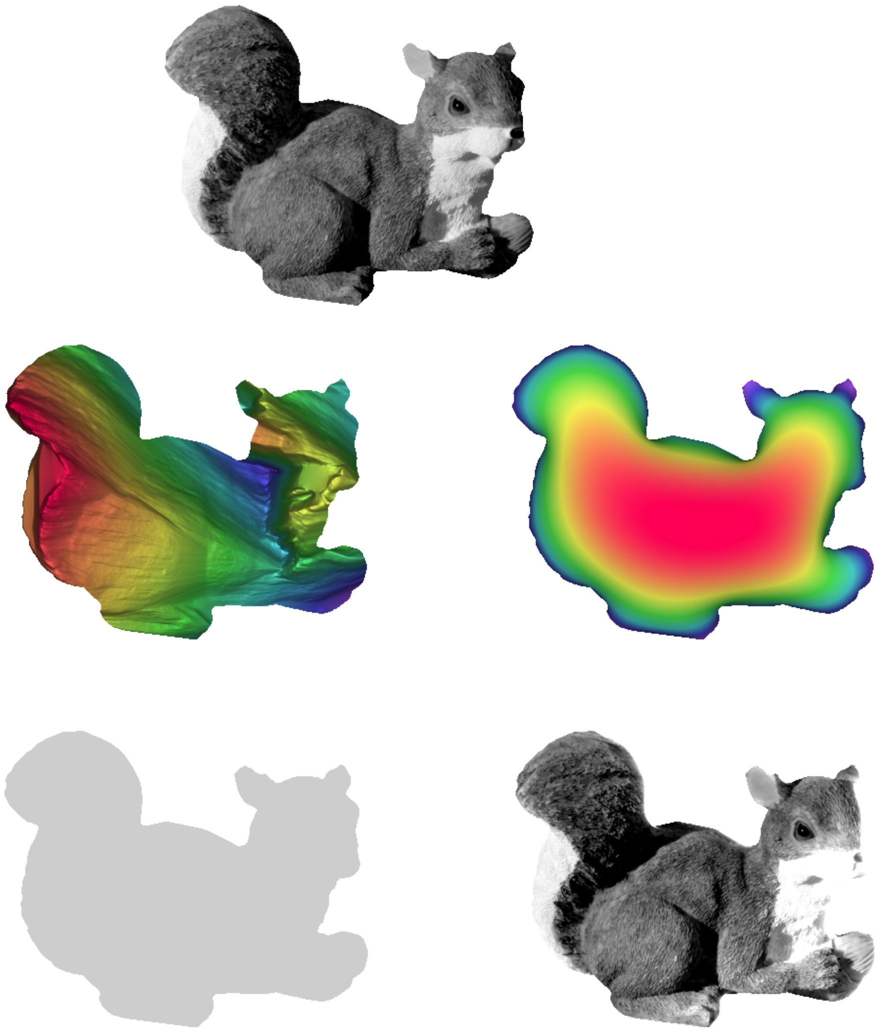
"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Some Explanations

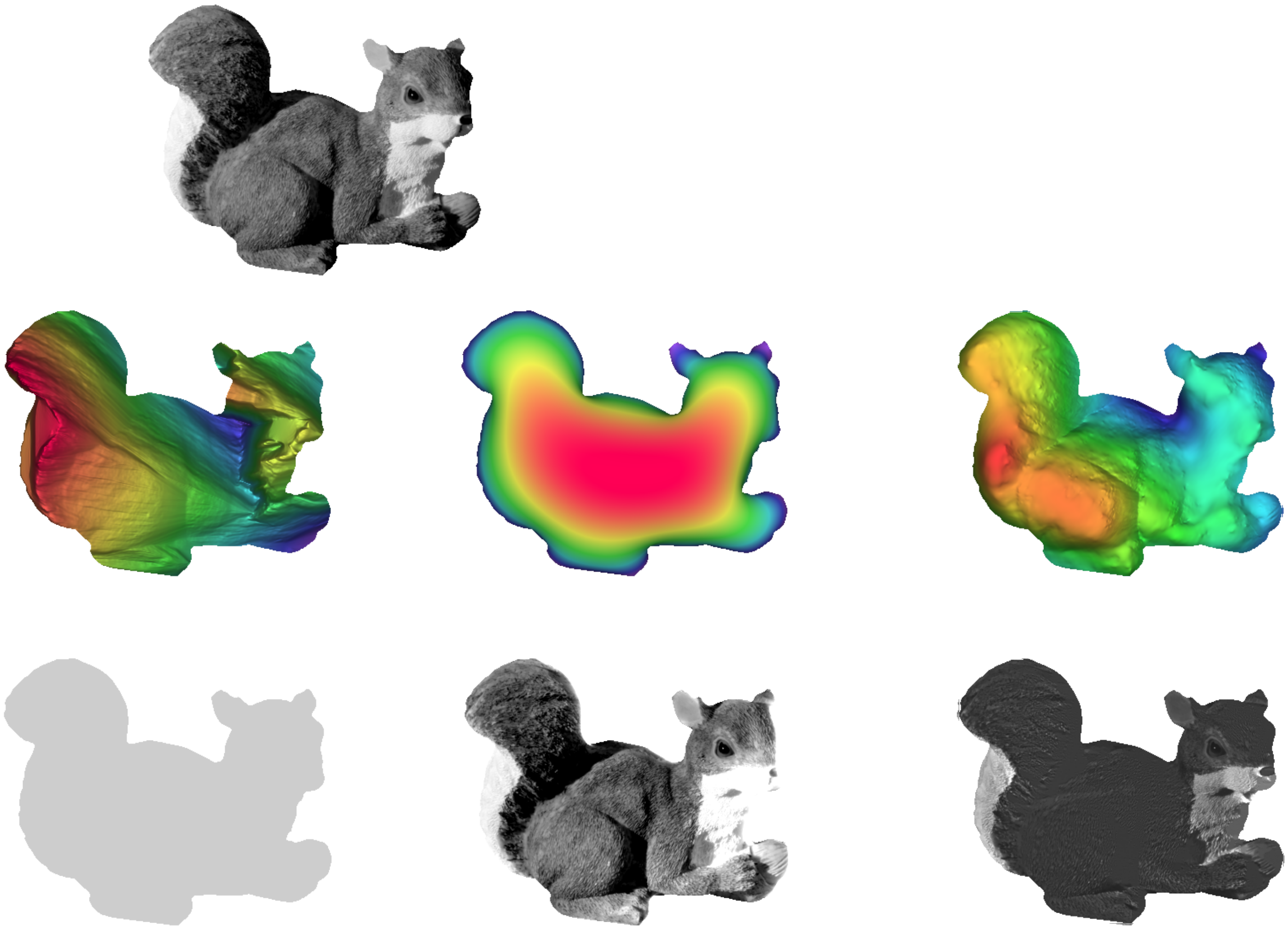
Some Explanations



Some Explanations

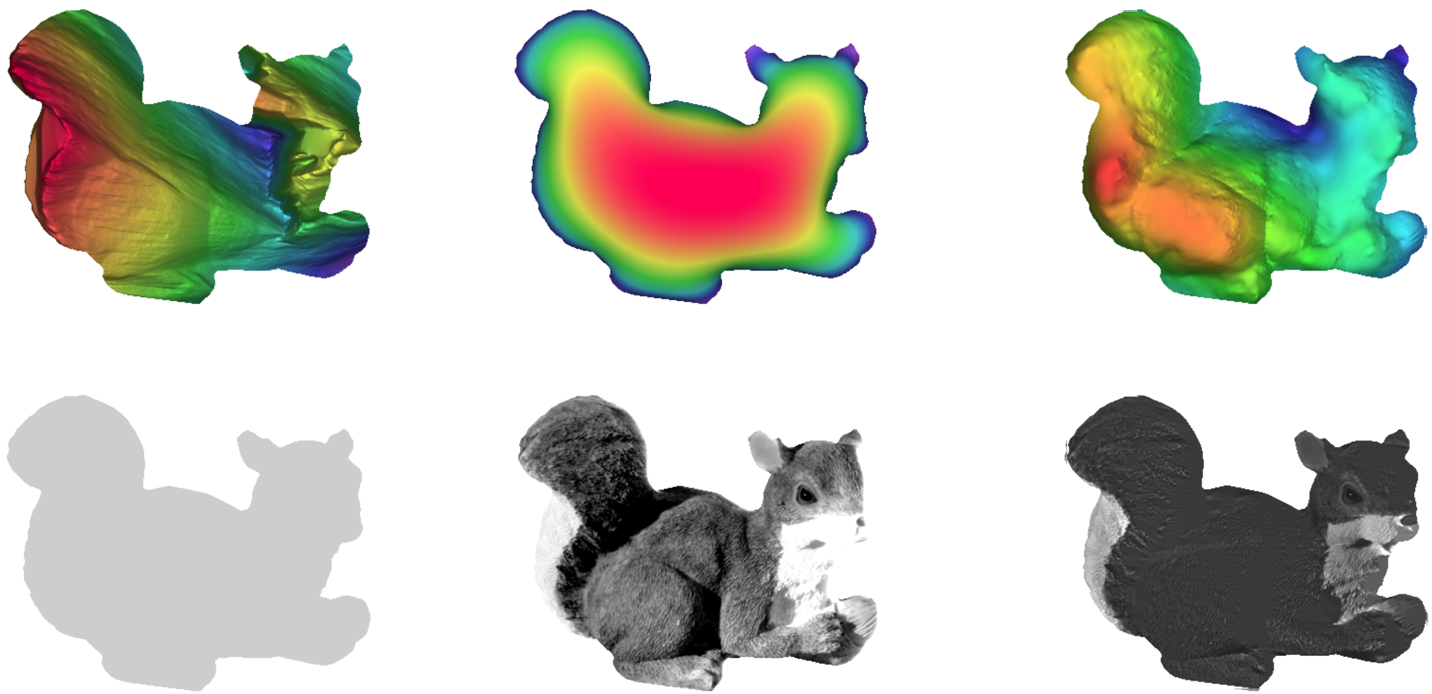


Some Explanations



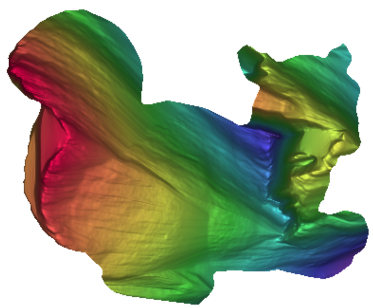
Some Explanations

$$P(R) \quad \propto \quad e^{-g(R)}$$
$$P(Z) \quad \propto \quad e^{-g(Z)}$$

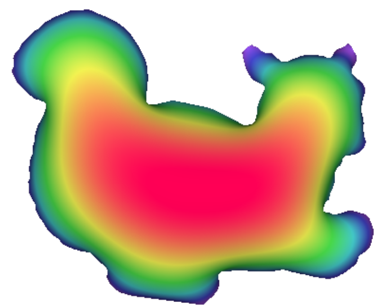


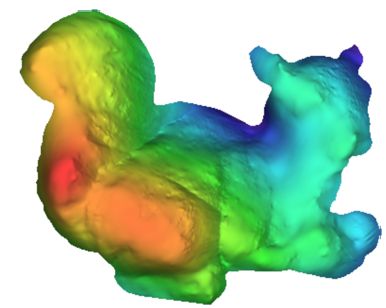
Some Explanations

$$P(R) \propto e^{-g(R)}$$
$$P(Z) \propto e^{-g(Z)}$$



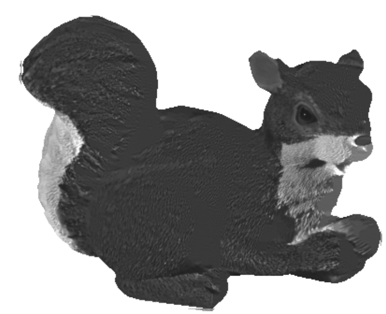
$f(Z) = 18.87$





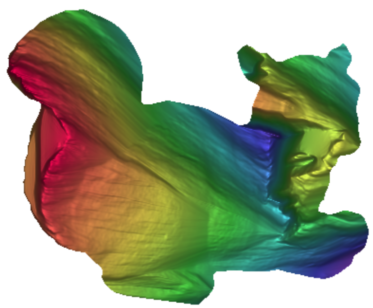
$g(R) = 0.00$

$f(Z) + g(R) = 18.87$

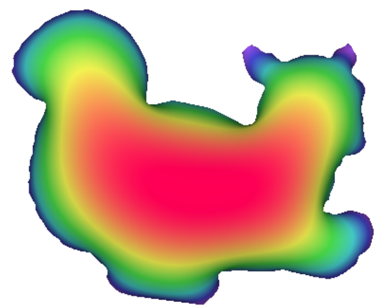


Some Explanations

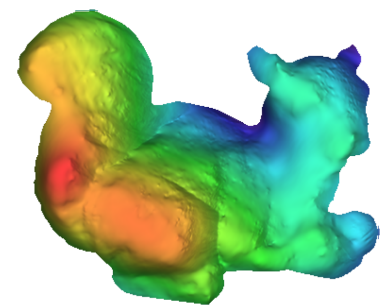
$$P(R) \quad \propto \quad e^{-g(R)}$$
$$P(Z) \quad \propto \quad e^{-g(Z)}$$



$f(Z) = 18.87$

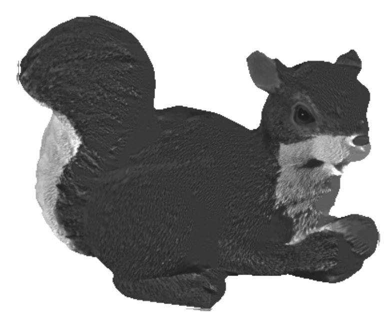


$f(Z) = 0.78$



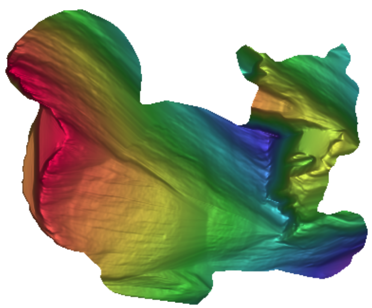
$g(R) = 0.00$
$f(Z) + g(R) = 18.87$

$g(R) = 19.12$
$f(Z) + g(R) = 19.91$

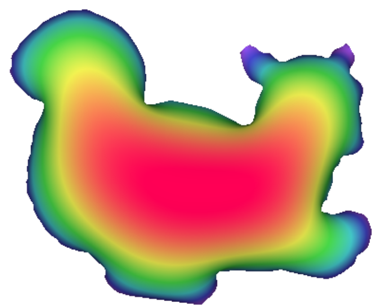


Some Explanations

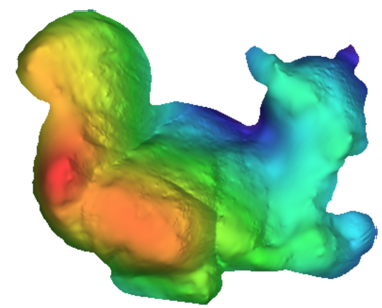
$$P(R) \quad \propto \quad e^{-g(R)}$$
$$P(Z) \quad \propto \quad e^{-g(Z)}$$



$f(Z) = 18.87$



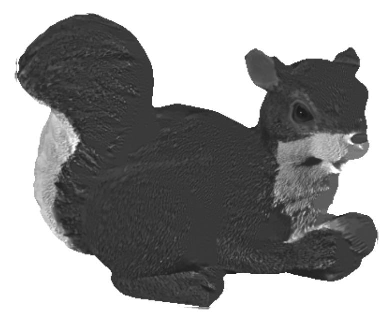
$f(Z) = 0.78$



$f(Z) = 2.27$

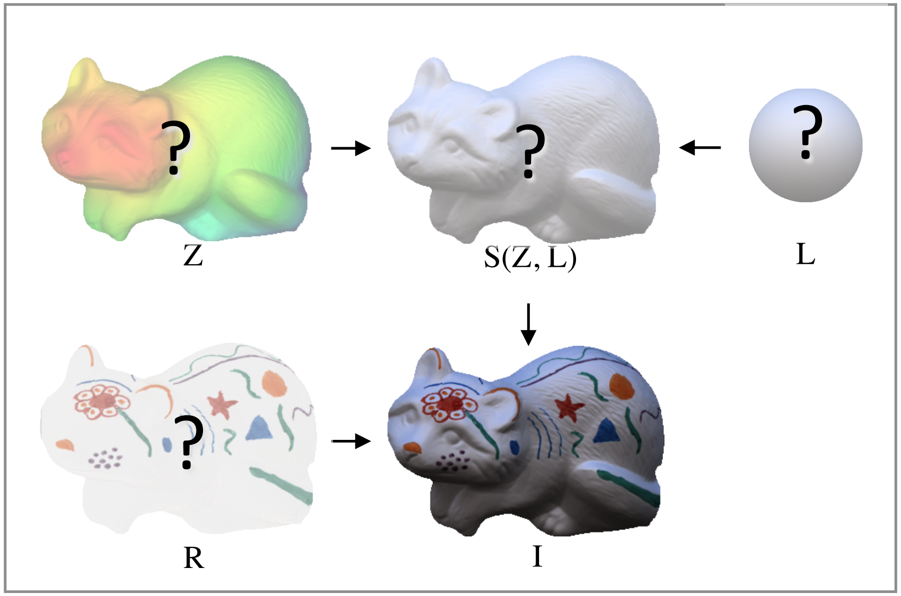
$g(R) = 0.00$
$f(Z) + g(R) = 18.87$

$g(R) = 19.12$
$f(Z) + g(R) = 19.91$



$g(R) = 9.66$
$f(Z) + g(R) = 11.93$

Problem Formulation

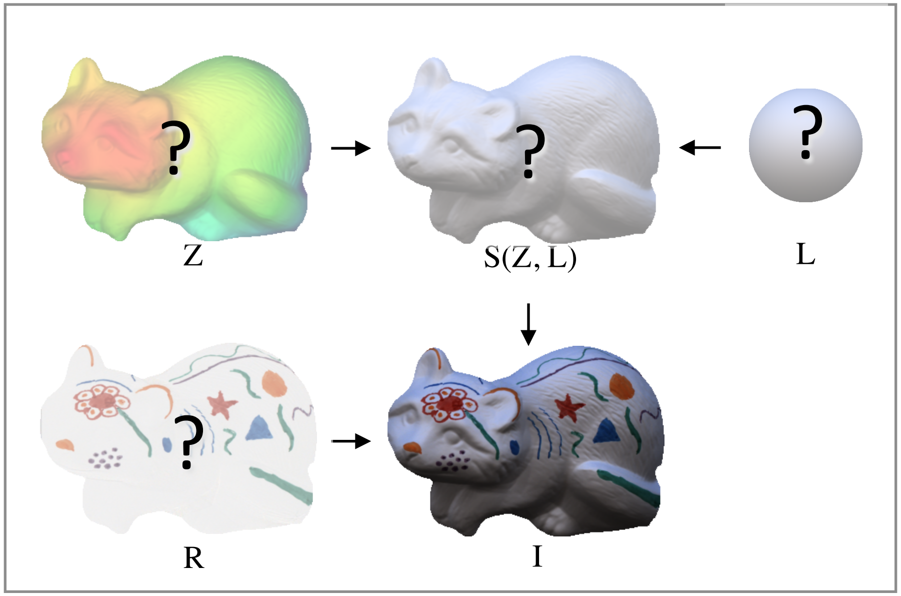


$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Problem Formulation



$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

What do we know about **reflectance**?

What do we know about **reflectance**?

1) Piecewise smooth
(variation is small and sparse)

$$g(R) = \lambda_s \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(R_i - R_j \,; \mathbf{0}, \boldsymbol{\sigma}_k\right) \right)$$

What do we know about **reflectance**?

1) Piecewise smooth
   (variation is small and sparse)

2) Palette is small
   (distribution is low-entropy)

$$g(R) = \lambda_s \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(R_i - R_j \,; \mathbf{0}, \boldsymbol{\sigma}_k\right) \right) - \lambda_e \log \left(\sum_i \sum_j \exp \left(-\frac{\left(R_i - R_j\right)^2}{4\sigma_e^2} \right) \right)$$

What do we know about **reflectance**?

1) Piecewise smooth
(variation is small and sparse)

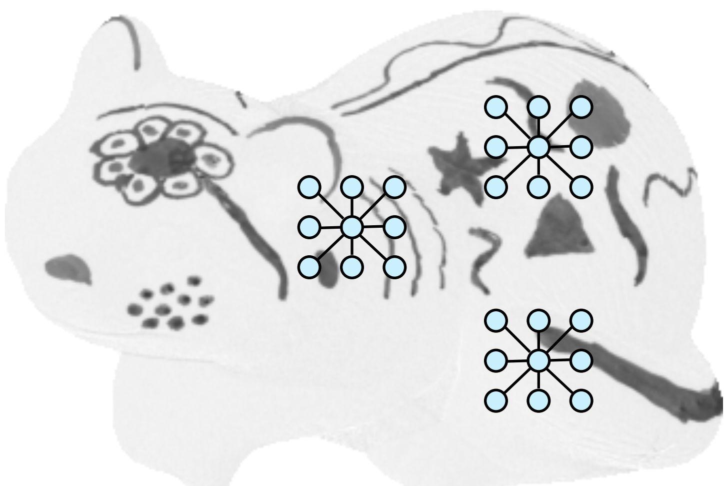
2) Palette is small
(distribution is low-entropy)

3) Some colors are common
(maximize likelihood under density model)

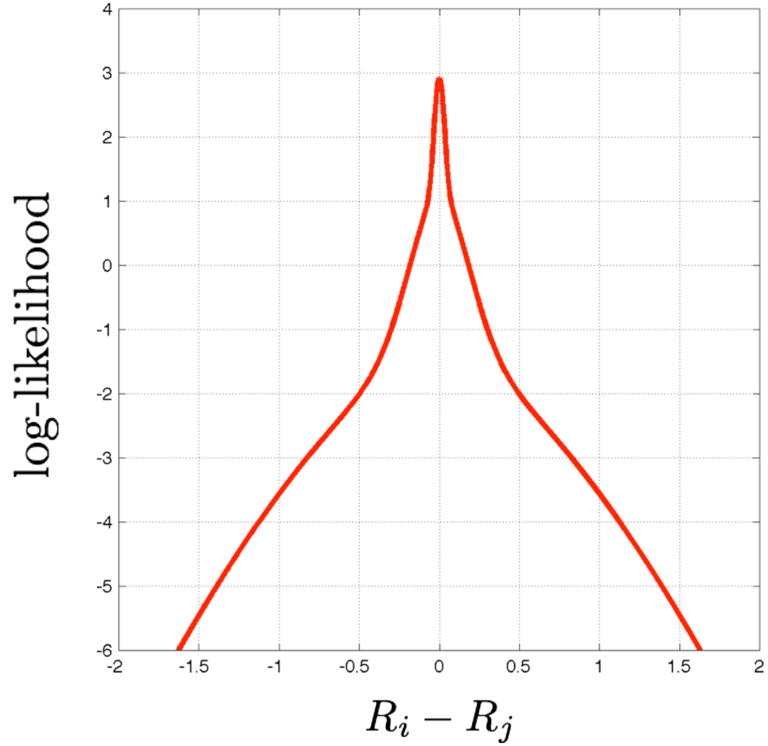
$$g(R) = \lambda_s \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(R_i - R_j\,; \mathbf{0}, \boldsymbol{\sigma}_k\right) \right) - \lambda_e \log \left(\sum_i \sum_j \exp \left(- \frac{(R_i - R_j)^2}{4\sigma_e^2} \right) \right) + \lambda_a \sum_i \mathrm{F}(R_i)$$

Reflectance:Smoothness

Reflectance: Smoothness

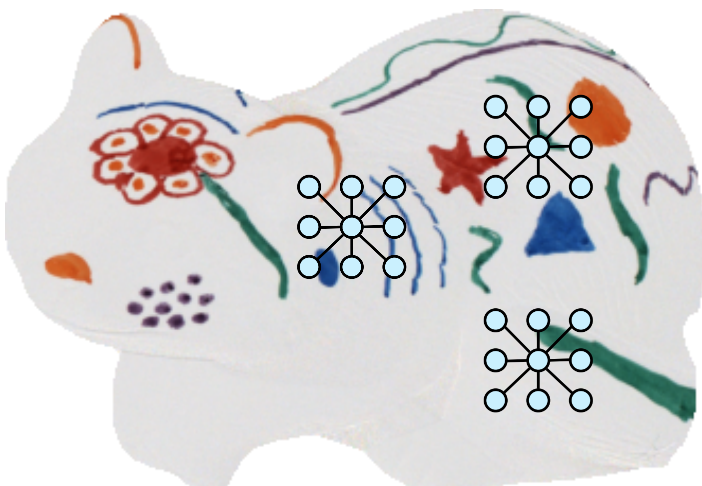


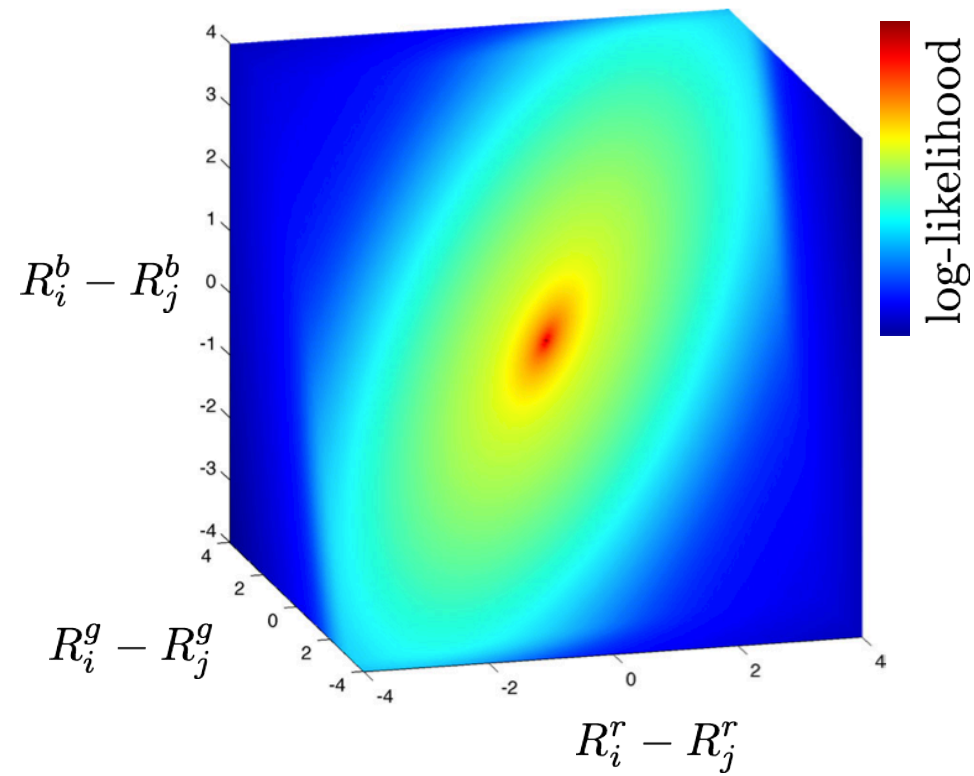
Reflectance:Smoothness



$$\sum_{i} \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \, \mathcal{N} \left(R_i - R_j \, ; \mathbf{0}, \boldsymbol{\sigma}_k \right) \right)$$

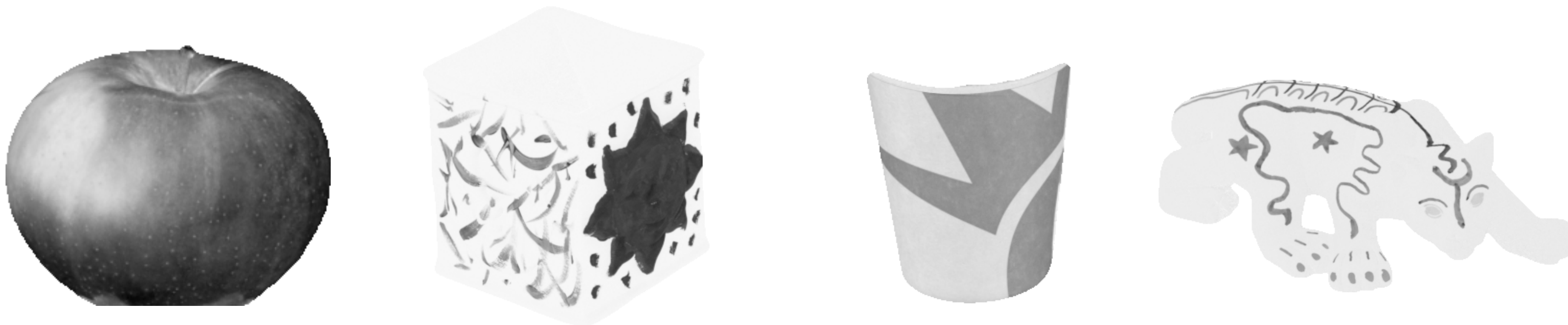
Reflectance:Smoothness



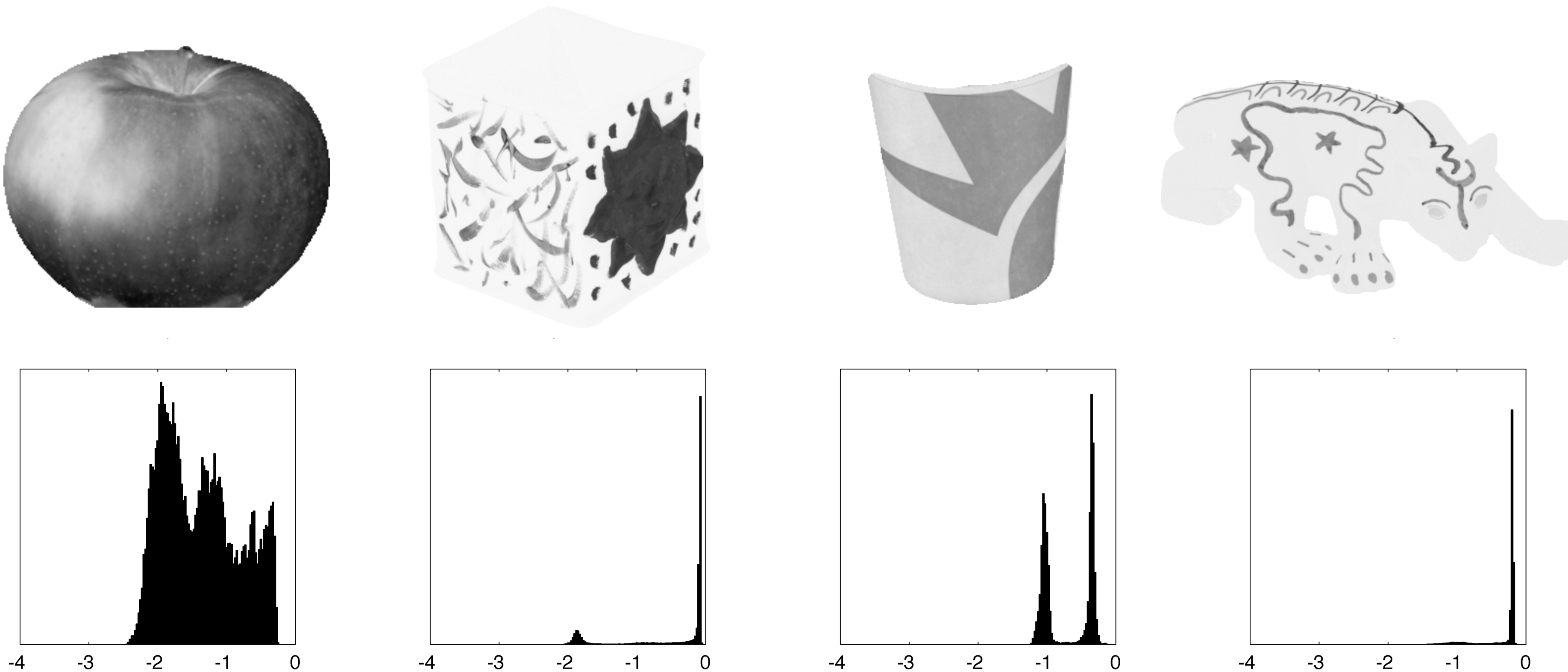


$$\sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(R_i - R_j \,; \mathbf{0}, \boldsymbol{\sigma}_k\right) \right)$$

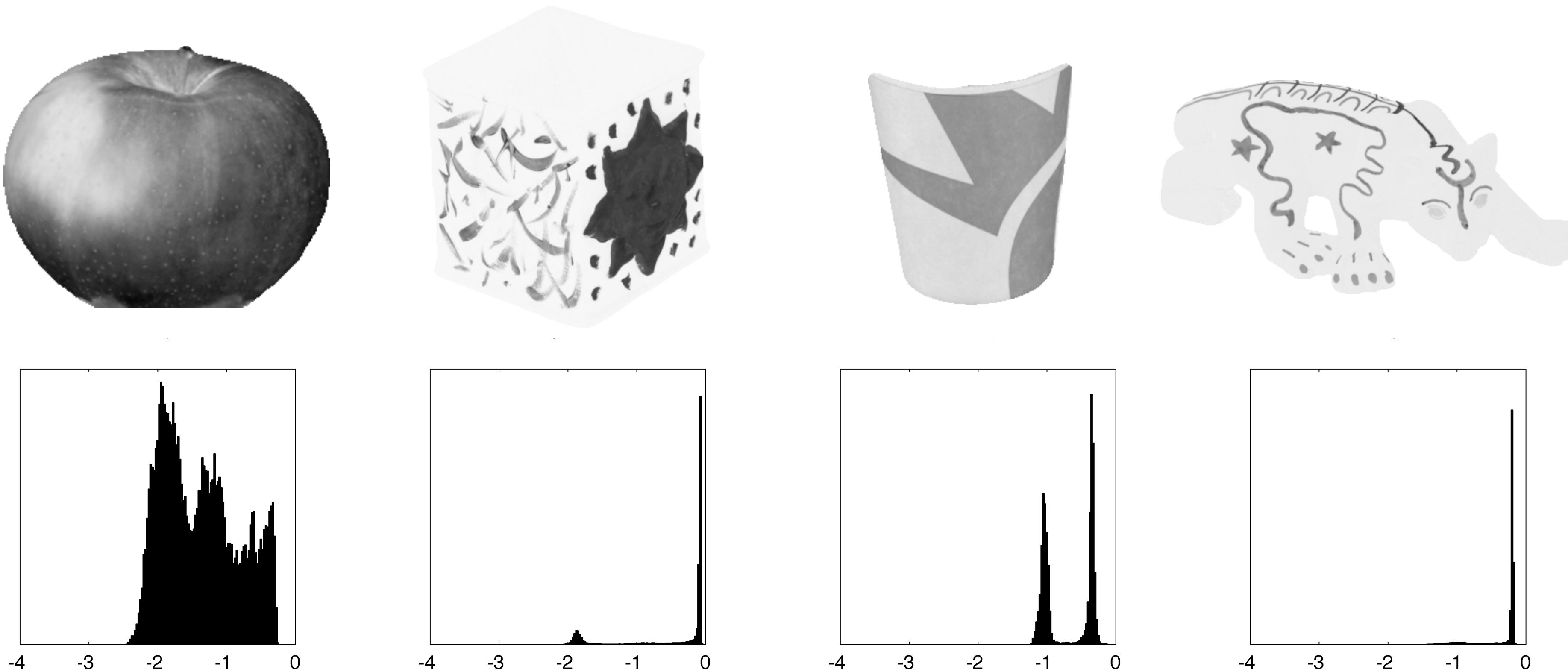
Reflectance: Minimal Entropy



Reflectance: Minimal Entropy



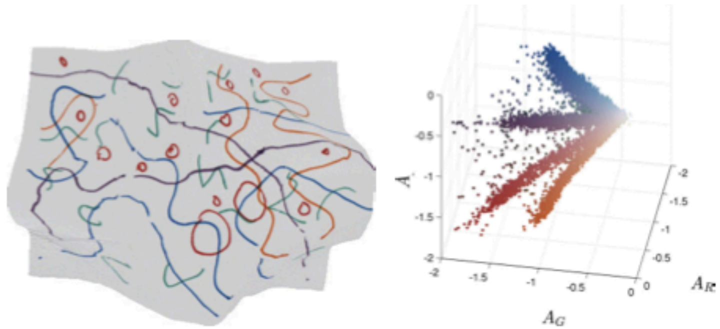
Reflectance: Minimal Entropy



1.094 $\quad -0.694$ $\quad -0.254$ $\quad -1.078$

$$-\log\left(\sum_i \sum_j \exp\left(-\frac{(R_i - R_j)^2}{4\sigma_e^2}\right)\right)$$

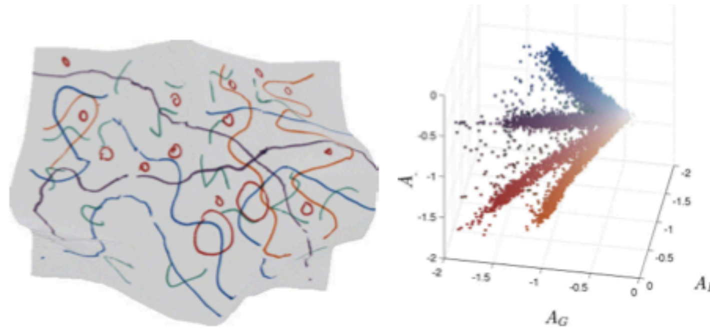
Reflectance: Minimal Entropy



(a) Correct Everything

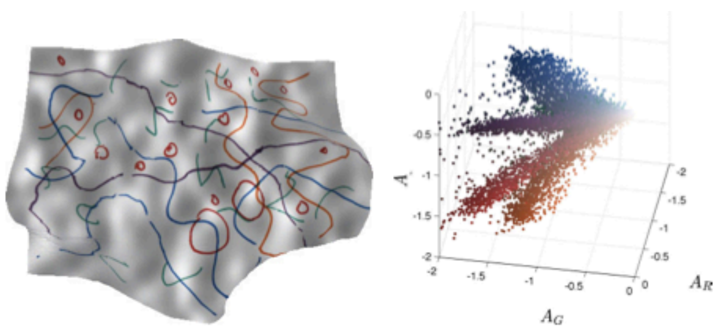
$$g_e(R) = 0.913$$

Reflectance: Minimal Entropy



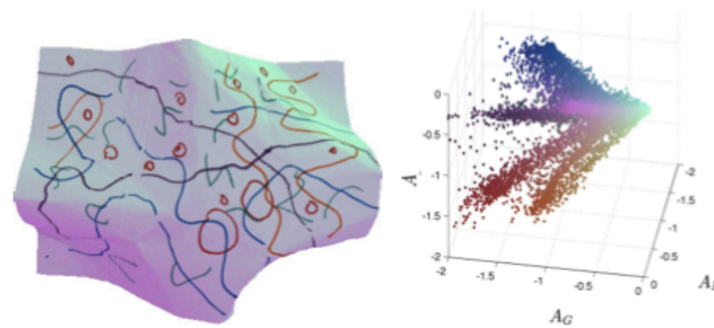
(a) Correct Everything

$g_e(R) = 0.913$



(b) Wrong Shape

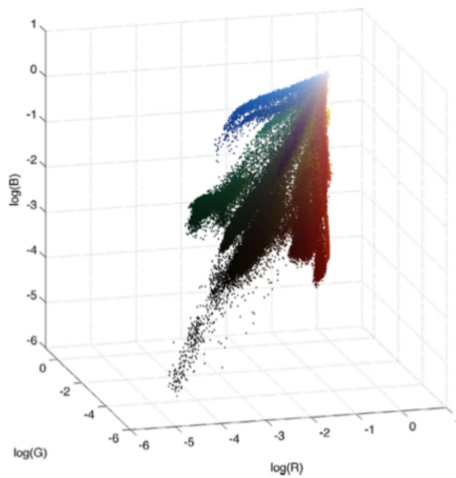
$g_e(R) = 1.325$



(c) Wrong Light

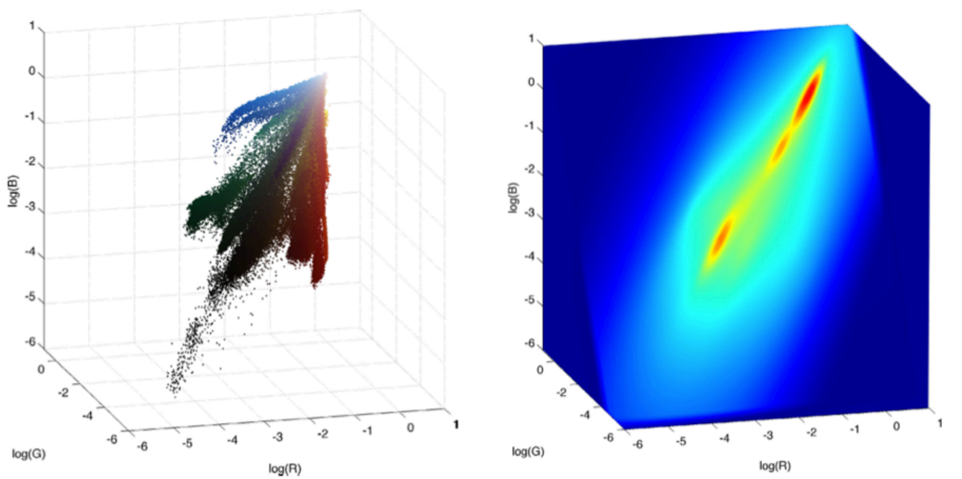
$g_e(R) = 2.366$

Reflectance: Absolute Color



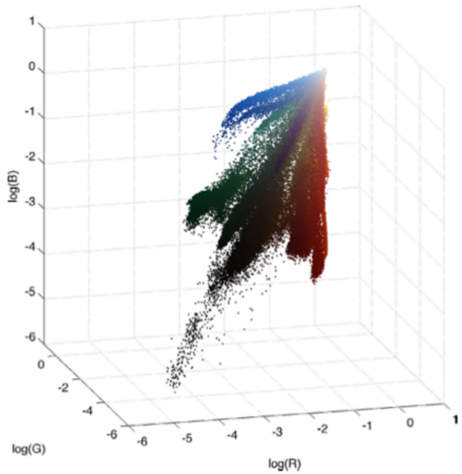
(a) Training reflectances

Reflectance: Absolute Color

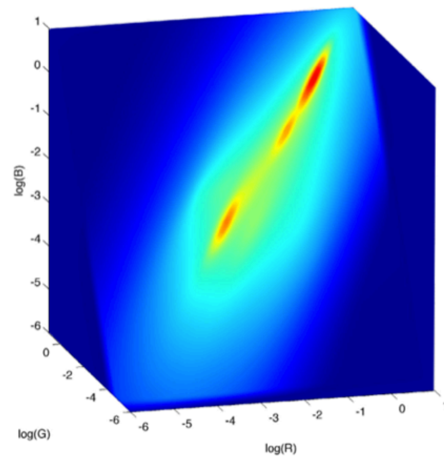


(a) Training reflectances (b) Our PDF of reflectance

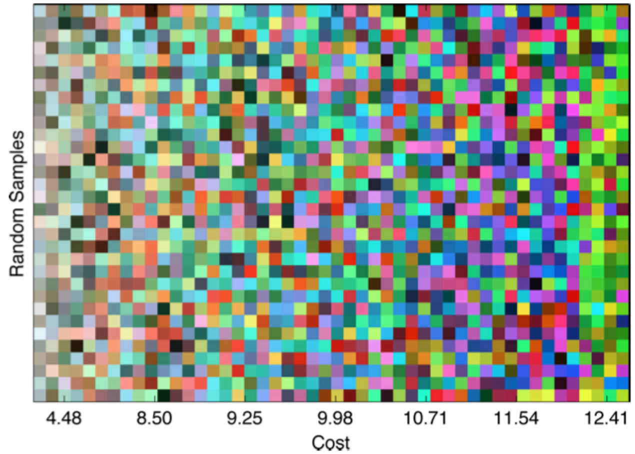
Reflectance: Absolute Color



(a) Training reflectances

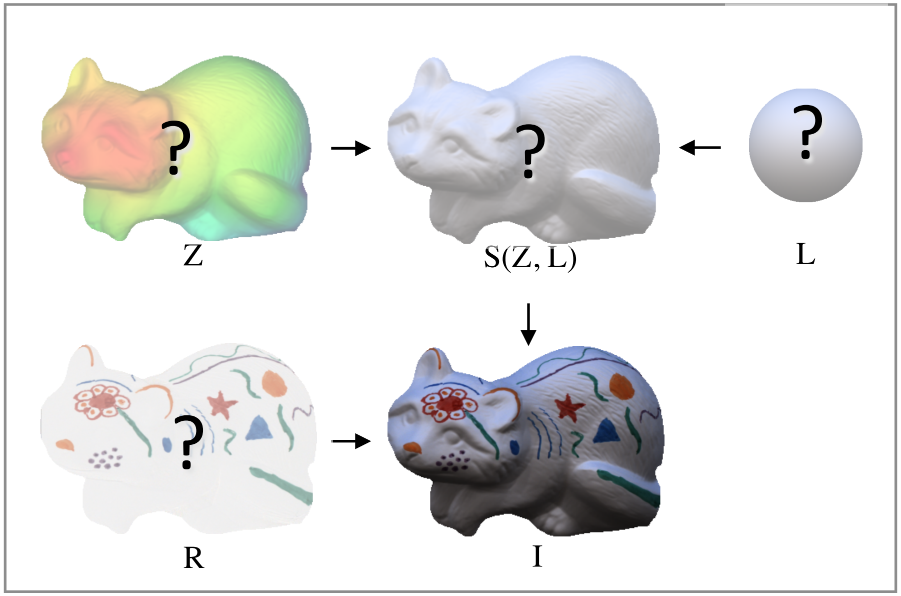


(b) Our PDF of reflectance



(c) Reflectances sorted by cost

Problem Formulation

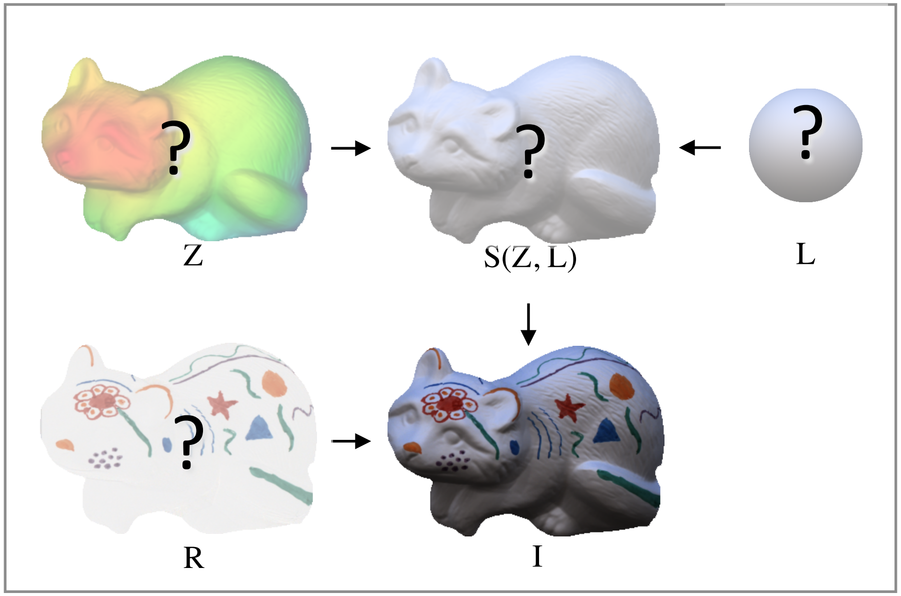


$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Problem Formulation



$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

What do we know about **shapes**?

What do we know about **shapes**?

1) Piecewise smooth
(variation in mean curvature is small and sparse)

$$f(Z) = \lambda_k \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N} \left(H(Z)_i - H(Z)_j \, ; 0, \boldsymbol{\sigma}_k \right) \right) .$$

What do we know about **shapes**?

1) Piecewise smooth
   (variation in mean curvature is small and sparse)

2) Face outward at the occluding contour

$$f(Z) = \lambda_k \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(H(Z)_i - H(Z)_j \,; 0, \boldsymbol{\sigma}_k\right) \right) + \lambda_c \sum_{i \in C} \sqrt{\left(N_i^x(Z) - n_i^x\right)^2 + \left(N_i^y(Z) - n_i^y\right)^2} \,.$$

What do we know about **shapes**?

1) Piecewise smooth
(variation in mean curvature is small and sparse)

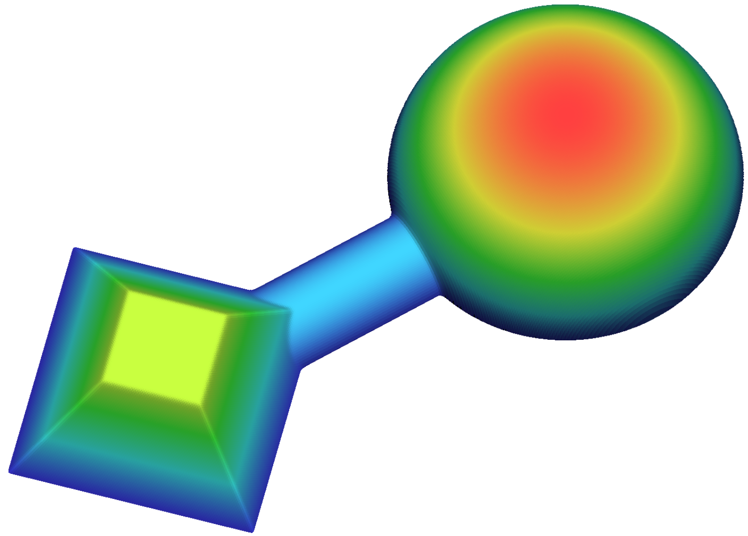
2) Face outward at the occluding contour

3) Tend to be fronto-parallel
(slant tends to be small)

$$f(Z) = \lambda_k \sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(H(Z)_i - H(Z)_j \,; 0, \boldsymbol{\sigma}_k\right) \right) + \lambda_c \sum_{i \in C} \sqrt{\left(N_i^x(Z) - n_i^x\right)^2 + \left(N_i^y(Z) - n_i^y\right)^2} - \lambda_f \sum_{x,y} \log\left(2 N_{x,y}^z(Z)\right)$$

Shapes:Smoothness

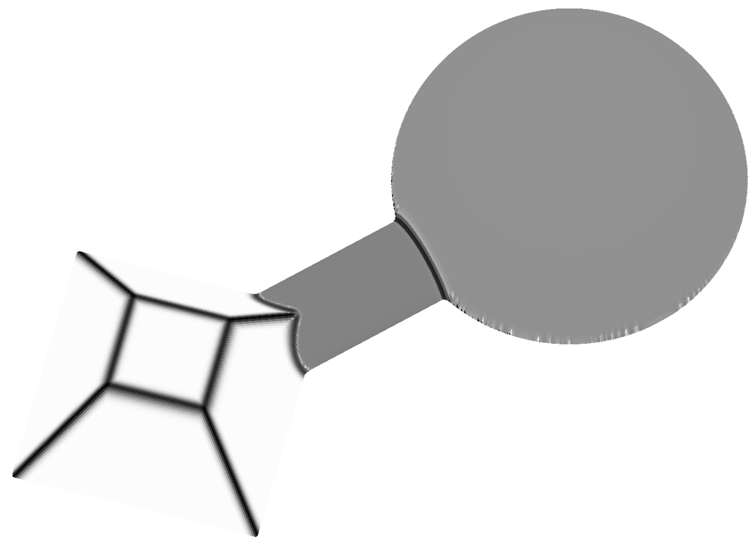
Shapes:Smoothness



$$Z$$

What's a good representation of shape for imposing priors?

Shapes:Smoothness



$$H(Z)$$

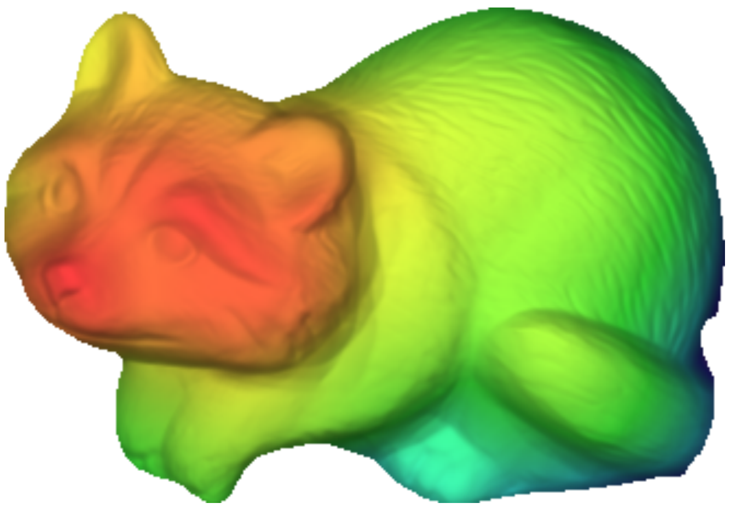
Mean Curvature of Z
(zero on planes, constant on cylinders and spheres)

Shapes:Smoothness

$$\nabla H(Z)$$

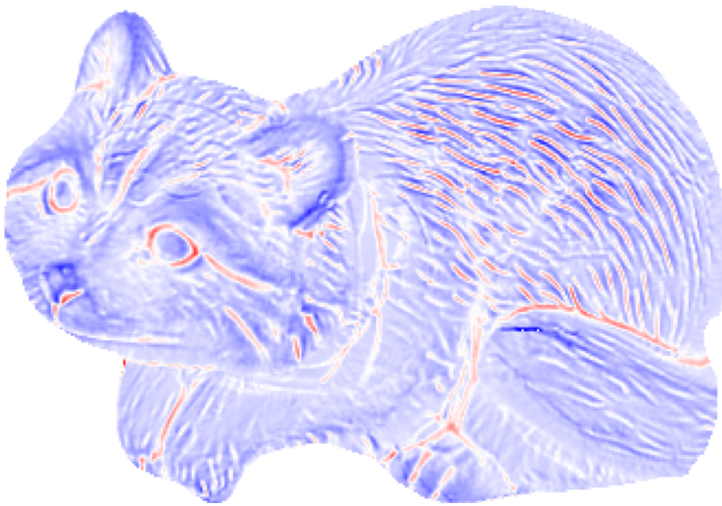
Variation of Mean Curvature of Z
"bending"

Shapes:Smoothness



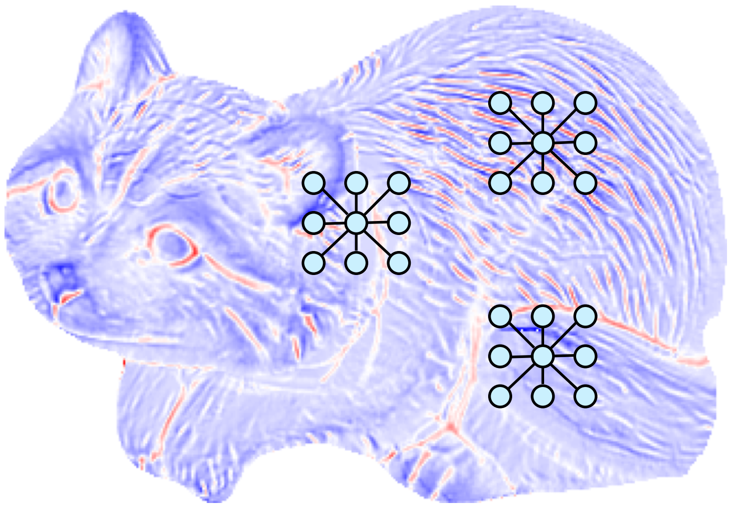
Z

Shapes:Smoothness



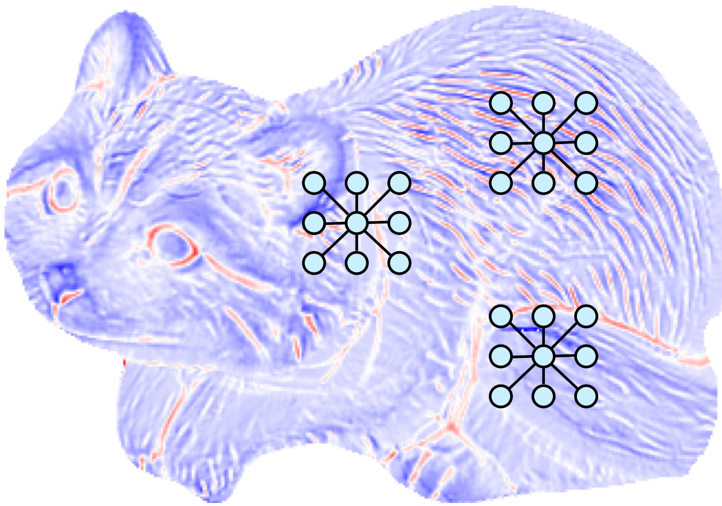
$H(Z)$

Shapes:Smoothness

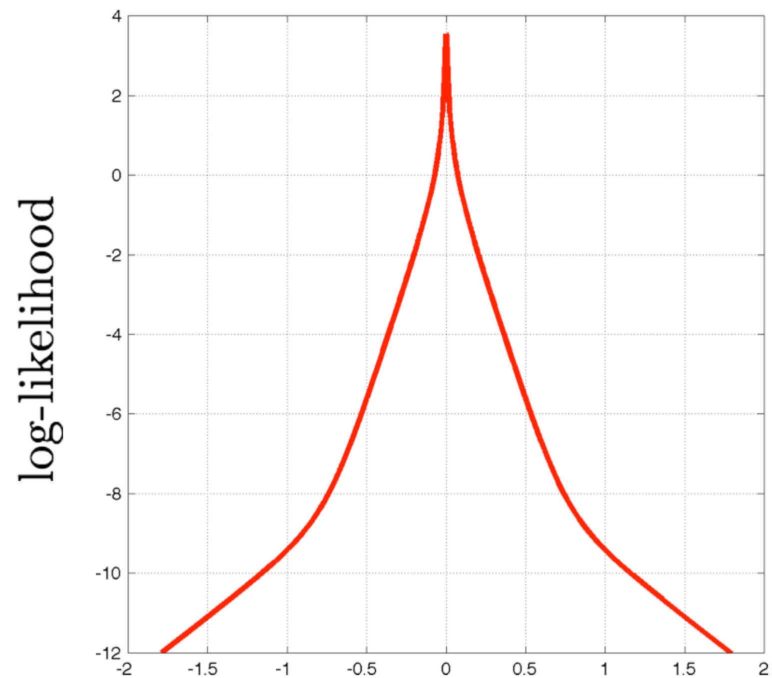


$H(Z)$

Shapes:Smoothness



$H(Z)$

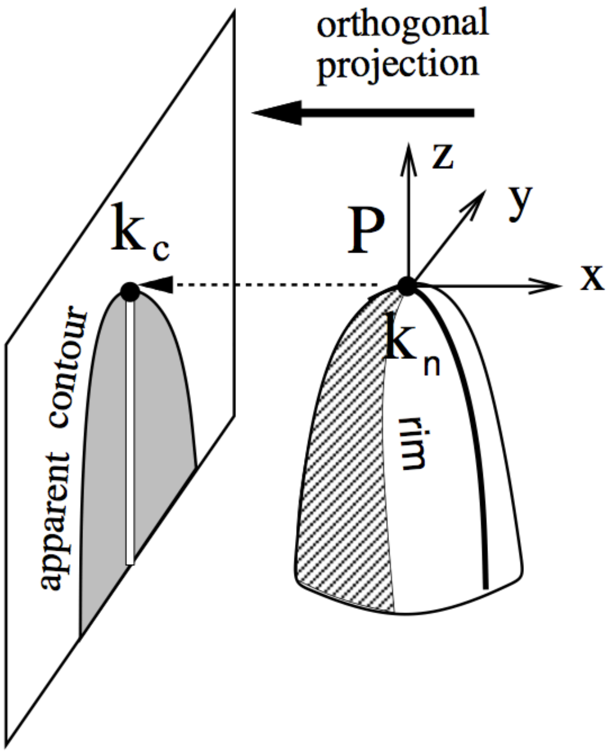


$H(Z)_i - H(Z)_j$

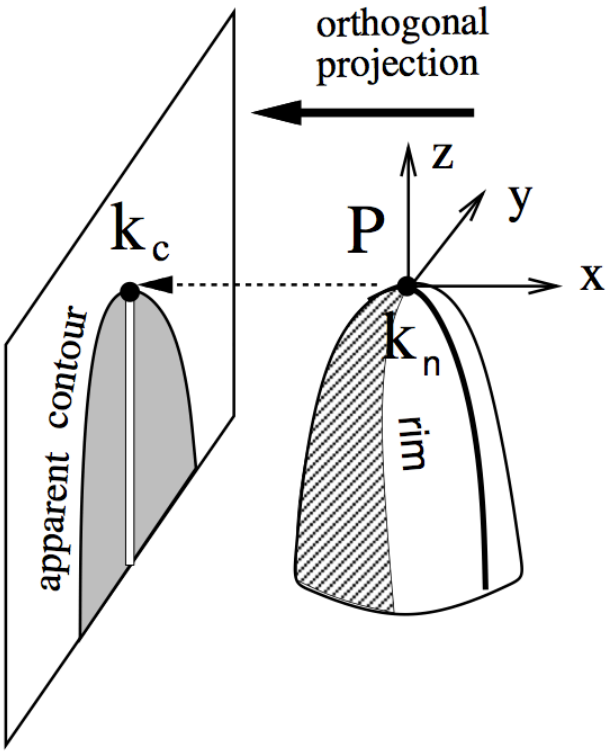
$$\sum_i \sum_{j \in N(i)} \log \left(\sum_{k=1}^{K} \boldsymbol{\alpha}_k \mathcal{N}\left(H(Z)_i - H(Z)_j \,; 0, \boldsymbol{\sigma}_k \right) \right)$$

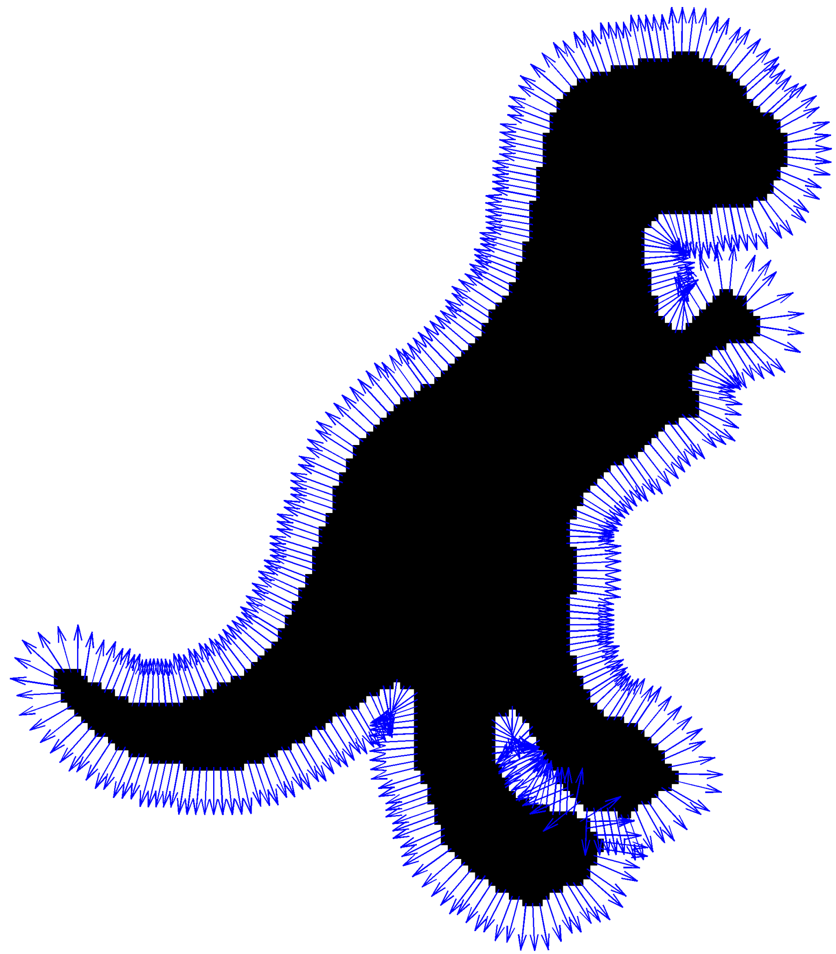
Shapes: Occluding Contours

Shapes: Occluding Contours

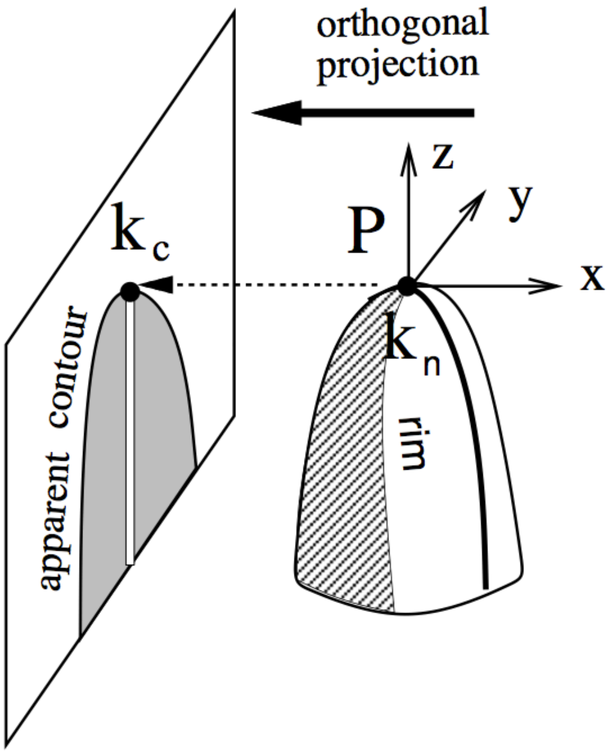


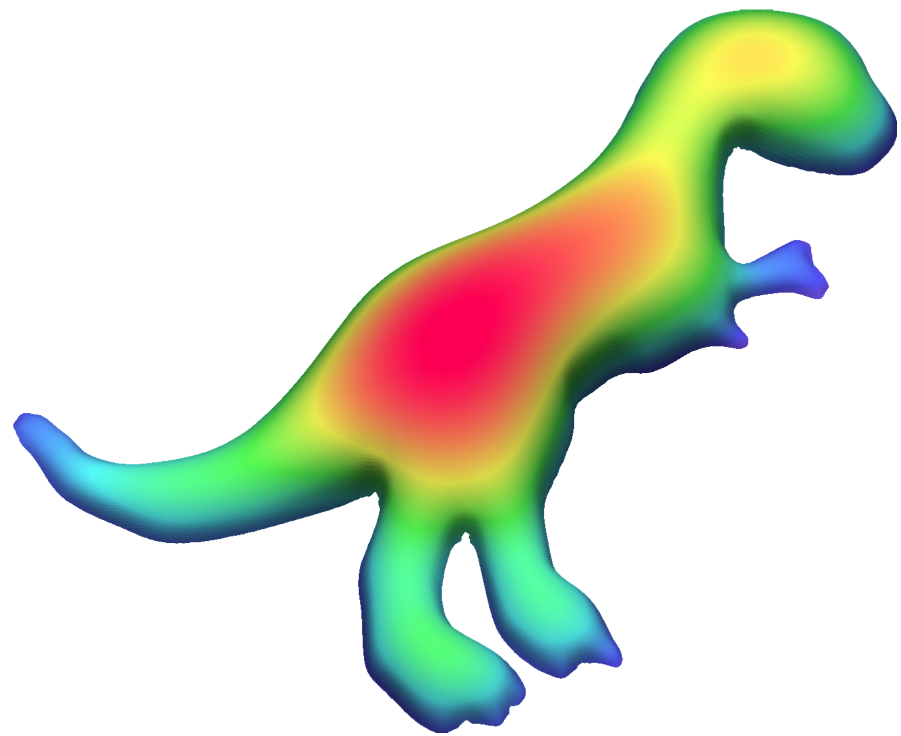
Shapes: Occluding Contours





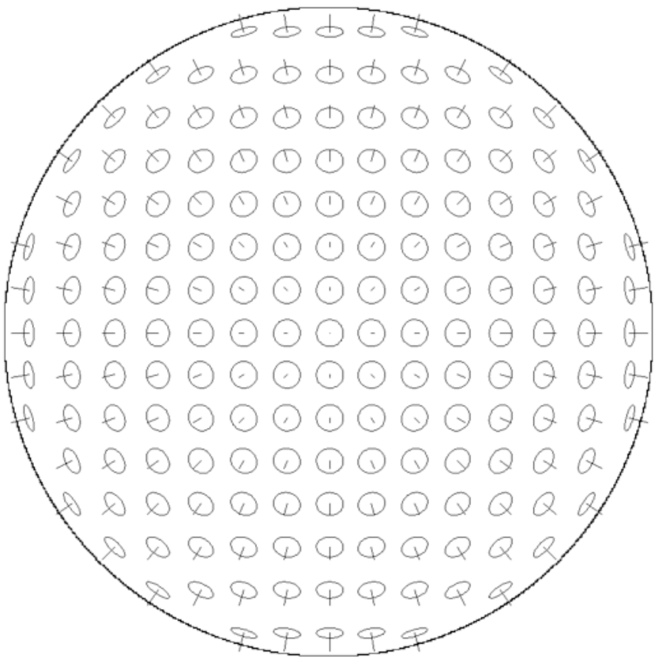
Shapes: Occluding Contours





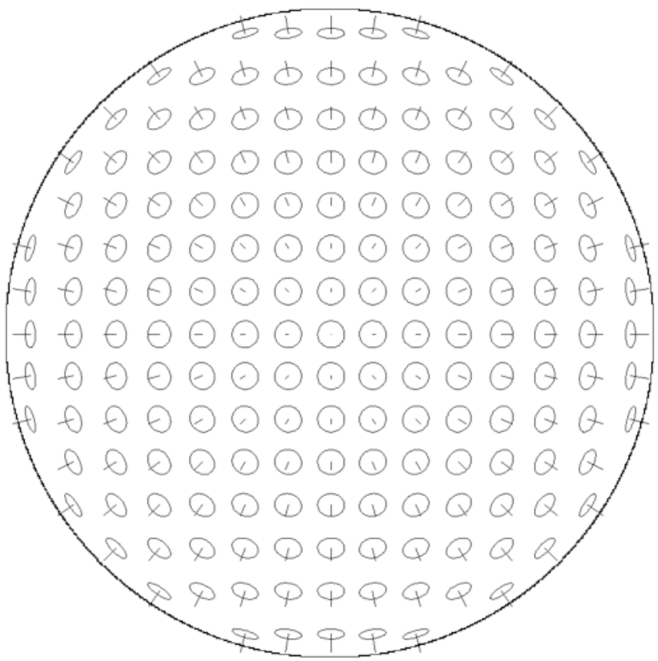
Shapes: Fronto-Parallel

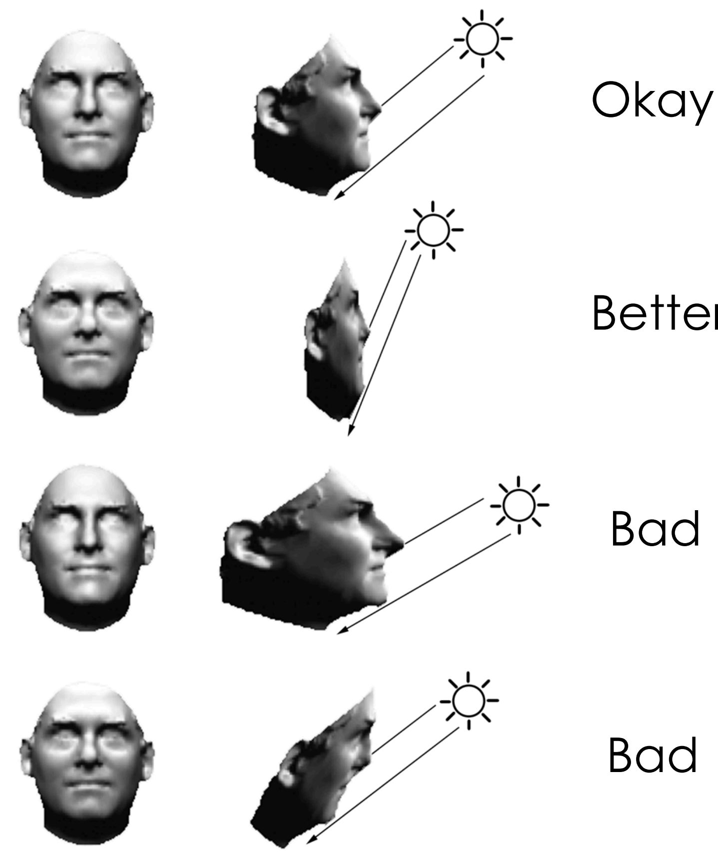
Shapes: Fronto-Parallel



If we observe a surface,
it is more likely that it faces us $(N^Z \approx 1)$
than that it is perpendicular to us $(N^Z \approx 0)$

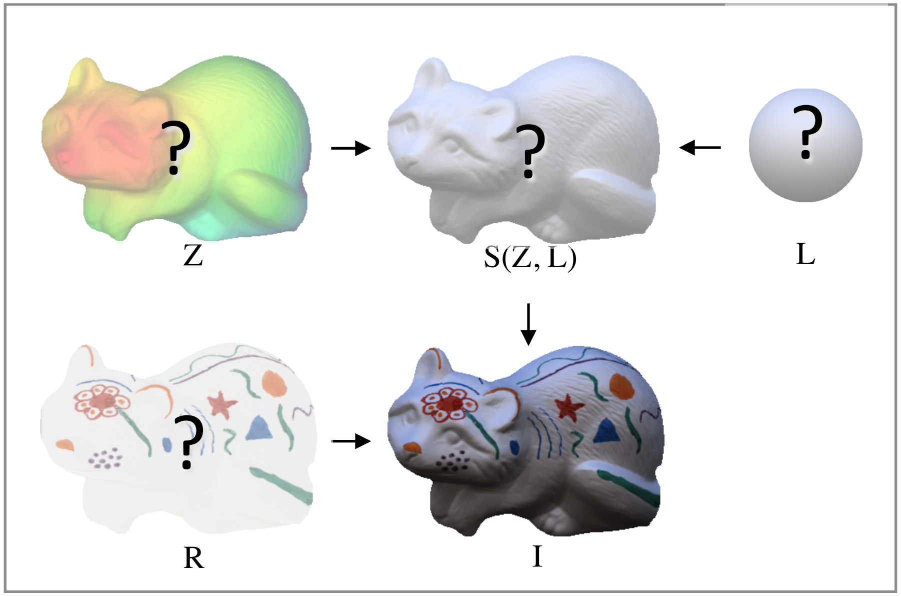
Shapes: Fronto-Parallel





If we observe a surface,
it is more likely that it faces us ($N^Z \approx 1$)
than that it is perpendicular to us ($N^Z \approx 0$)

Problem Formulation

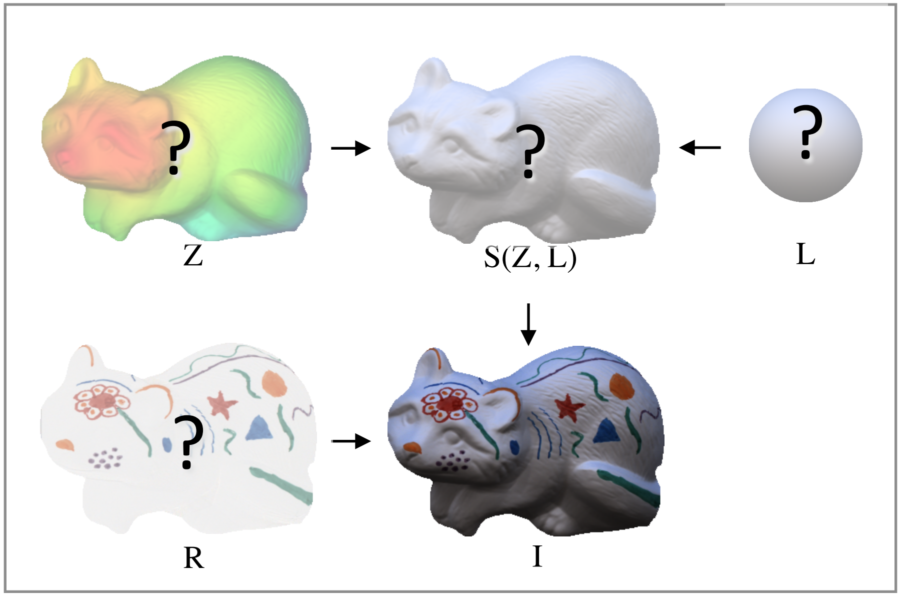


$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation (shape Z, log-reflectance R and illumination L) that together exactly reconstructs log-image I"

Problem Formulation



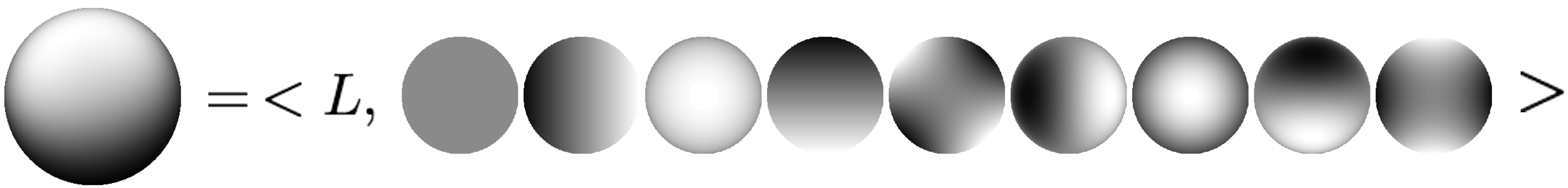
$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation (shape Z, log-reflectance R and illumination L) that together exactly reconstructs log-image I"

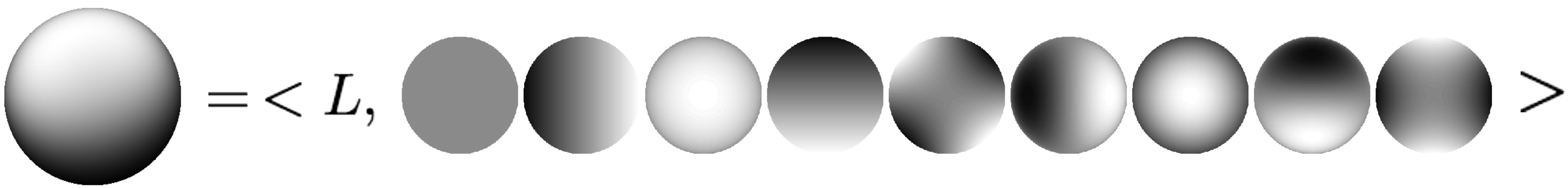
What do we know about **light**?

1) Global illumination is well modeled with spherical harmonics:



What do we know about **light**?

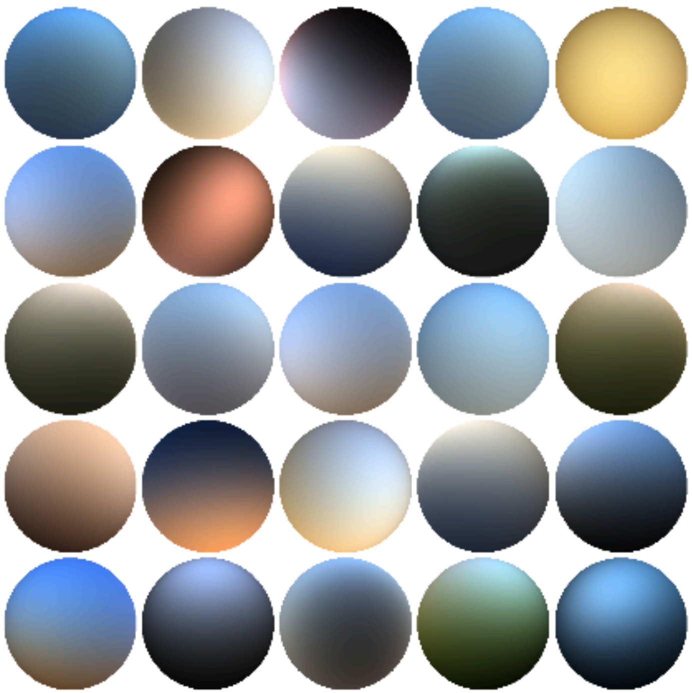
1) Global illumination is well modeled with spherical harmonics:



2) Spherical harmonic coefficients are well-modeled with a Gaussian

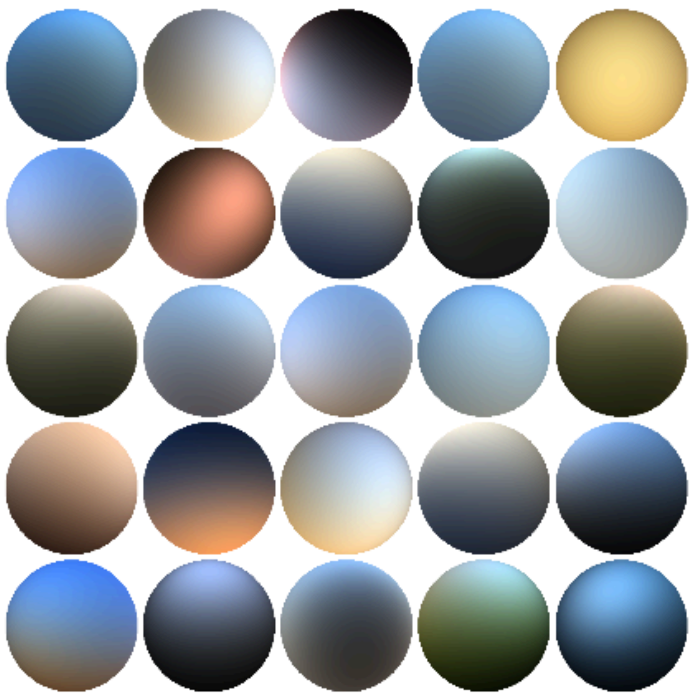
$$h(L) = \lambda_L (L - \boldsymbol{\mu}_L)^{\mathrm{T}} \Sigma_L^{-1} (L - \boldsymbol{\mu}_L)$$

What do we know about **light**?

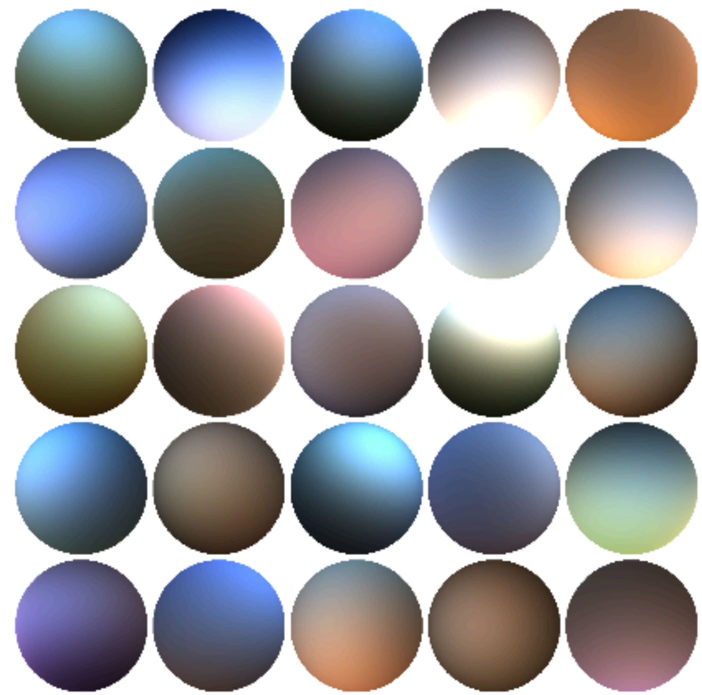


Natural Illuminations from our dataset

What do we know about **light**?

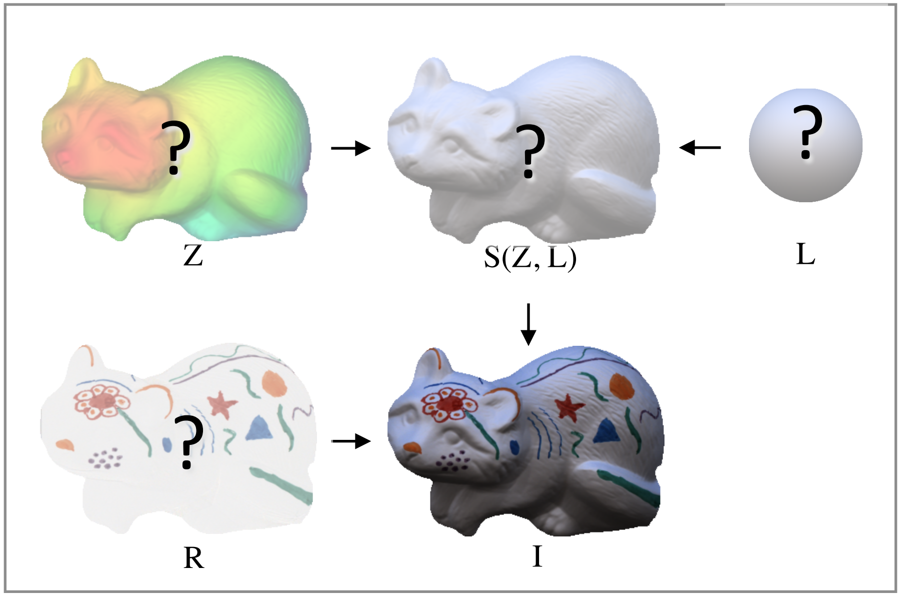


Natural Illuminations from our dataset



Samples from a Gaussian fit to the training set

Problem Formulation

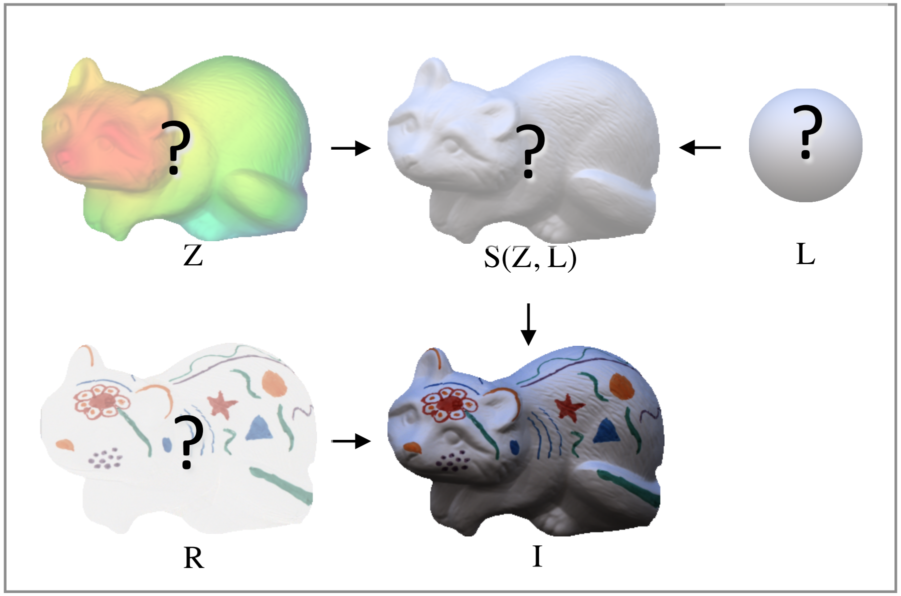


$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Problem Formulation

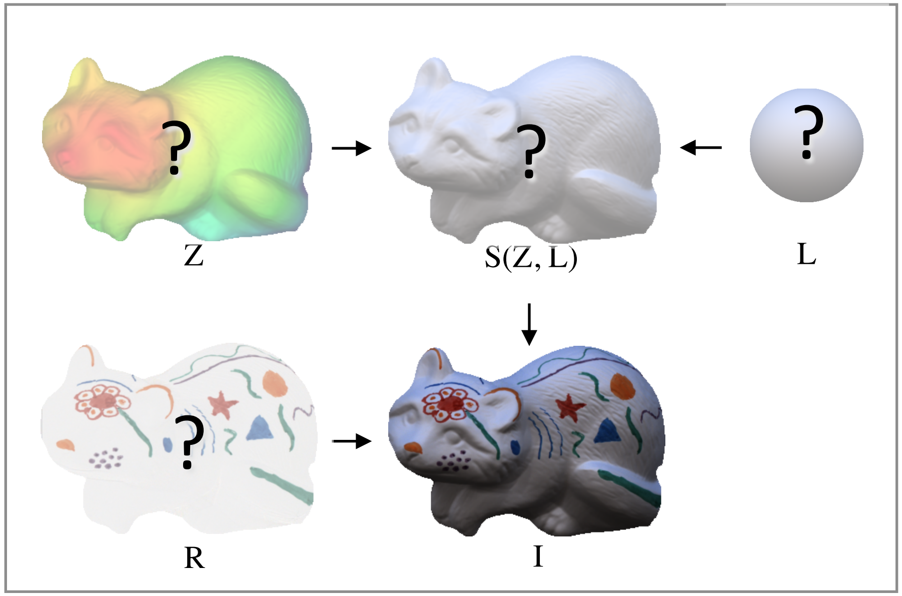


$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$

$$\text{subject to} \quad I = R + S(Z, L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Problem Formulation

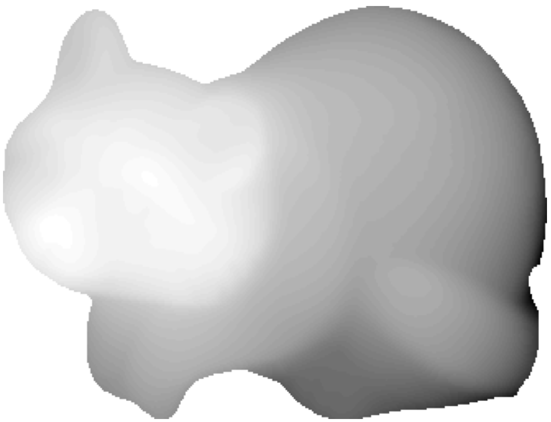


$$\underset{Z,L}{\text{minimize}} \quad g(I - S(Z, L)) + f(Z) + h(L)$$

"Search for the least costly explanation
(shape Z, log-reflectance R and illumination L)
that together exactly reconstructs log-image I"

Optimization

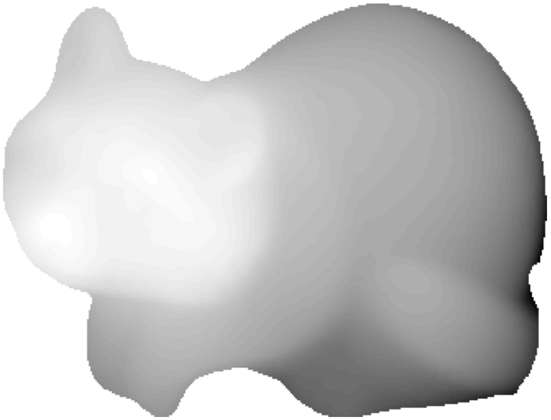
Straightforward L-BFGS with respect to Z fails!



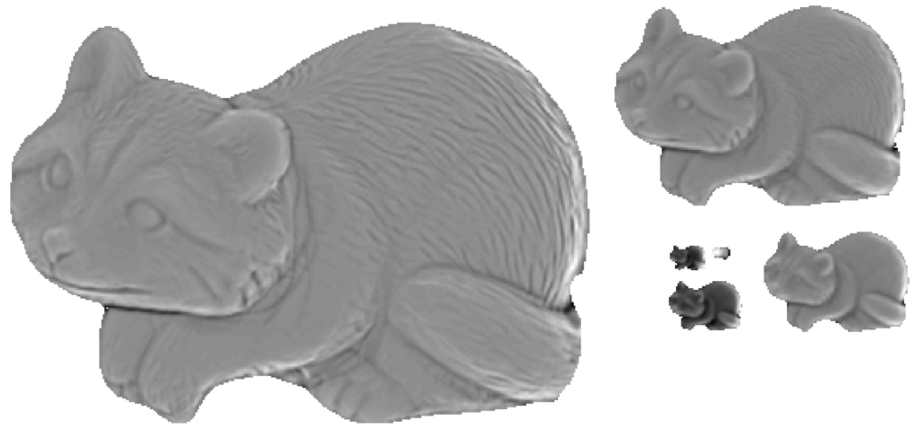
Z

Optimization

Straightforward L-BFGS with respect to Z fails!



Z

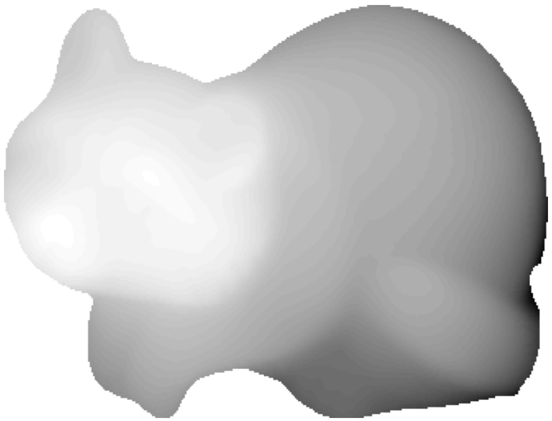


$\mathcal{L}(Z)$

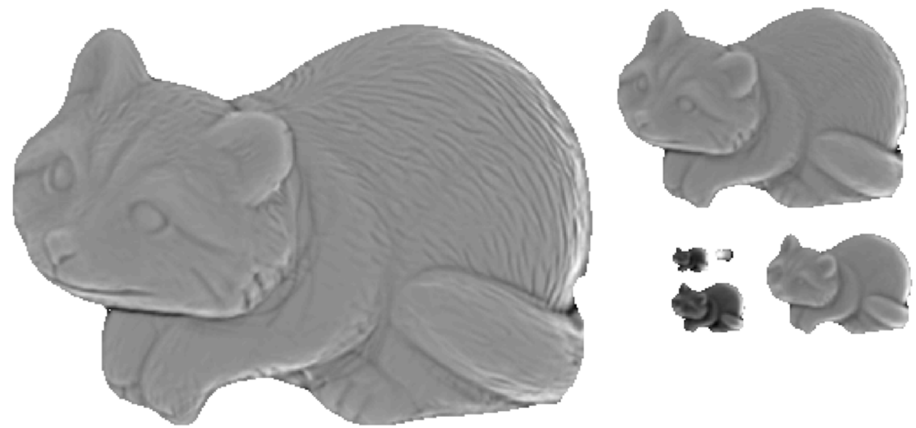
Instead, optimize over $\mathcal{L}(Z)$, a Laplacian pyramid of Z

Optimization

Straightforward L-BFGS with respect to Z fails!



Z



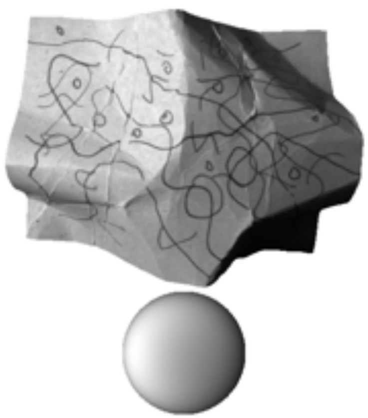
$\mathcal{L}(Z)$

Instead, optimize over $\mathcal{L}(Z)$, a Laplacian pyramid of Z

Psuedocode:

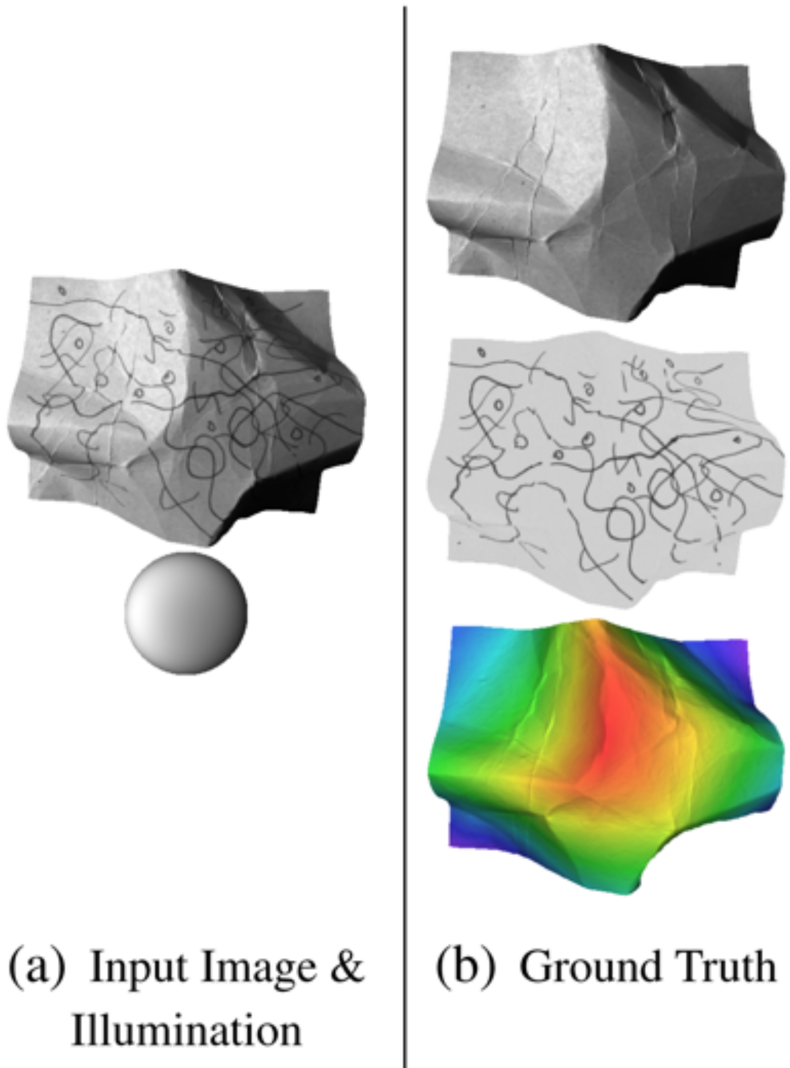
$$
\begin{aligned}
&[\ell, \nabla_Y \ell] = f'(Y): \\
&\quad Z \leftarrow \mathcal{L}^{-1}(Y) \\
&\quad [\ell, \nabla_Z \ell] \leftarrow f(Z) \\
&\quad \nabla_Y \ell \leftarrow \mathcal{G}(\nabla_Z \ell)
\end{aligned}
$$

Evaluation: Known Lighting



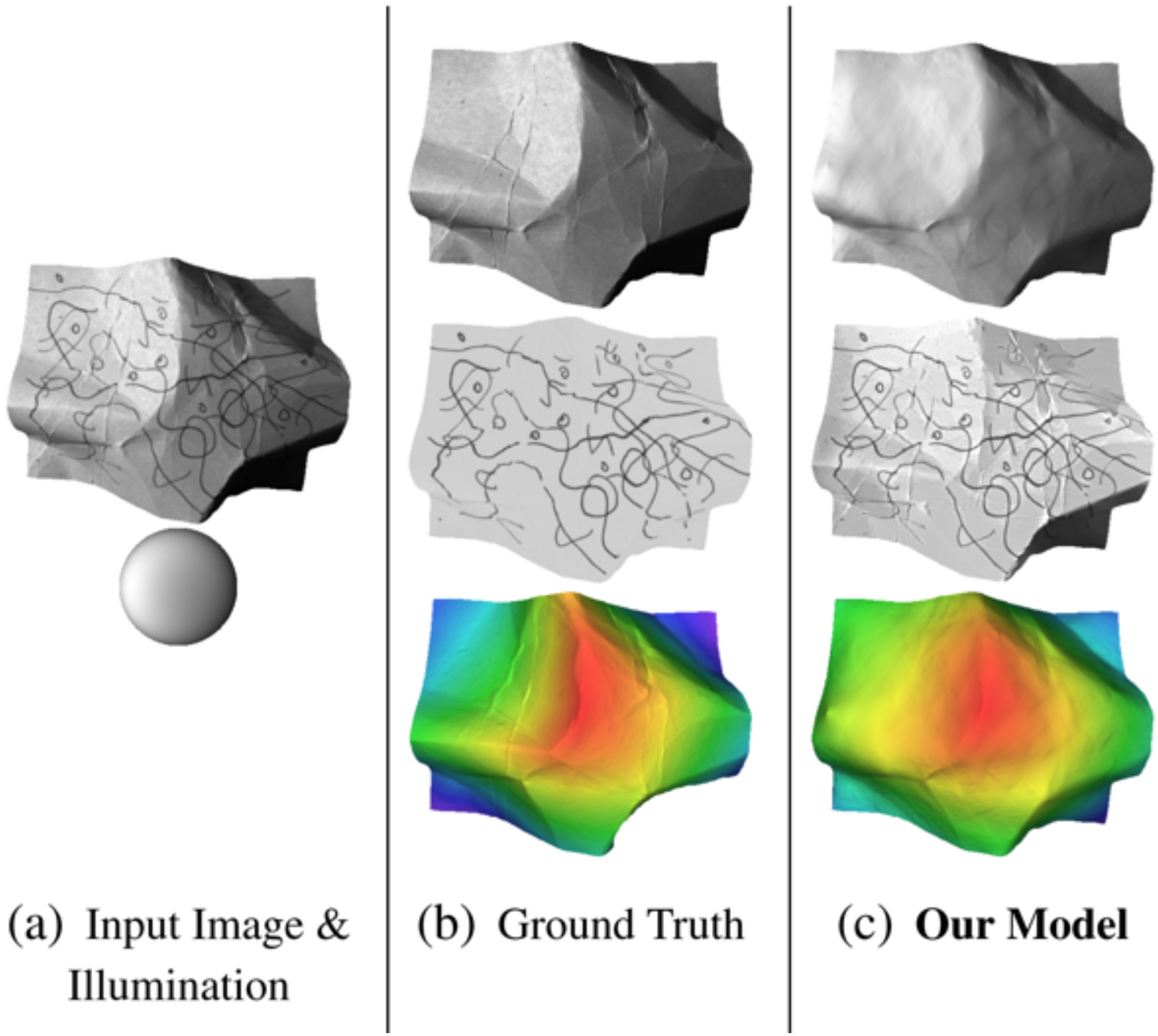
(a) Input Image & Illumination

Evaluation: Known Lighting



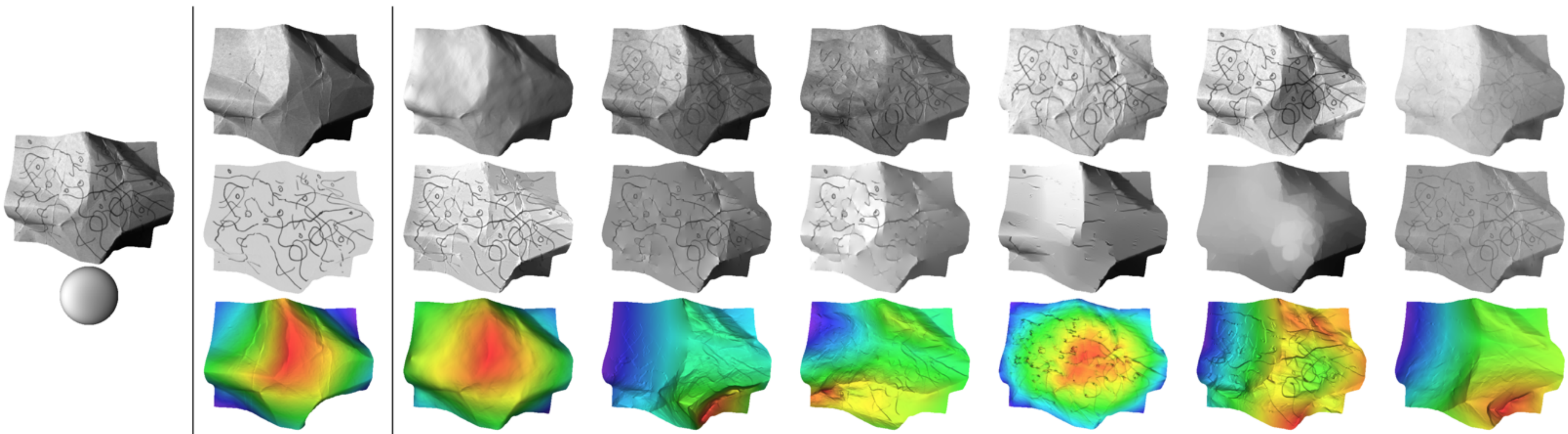
(a) Input Image & Illumination | (b) Ground Truth

Evaluation: Known Lighting



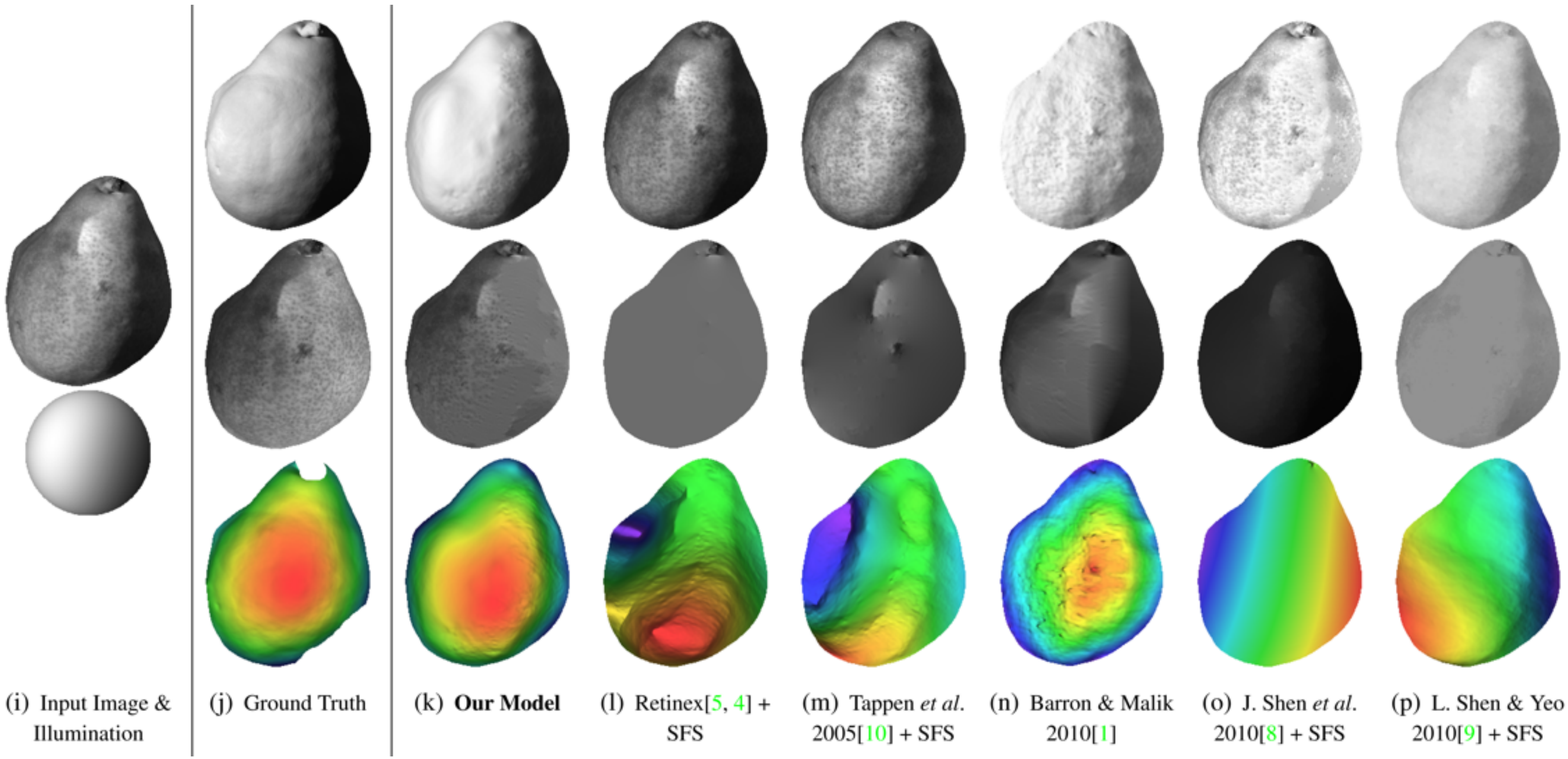
(a) Input Image & Illumination | (b) Ground Truth | (c) **Our Model**

Evaluation: Known Lighting



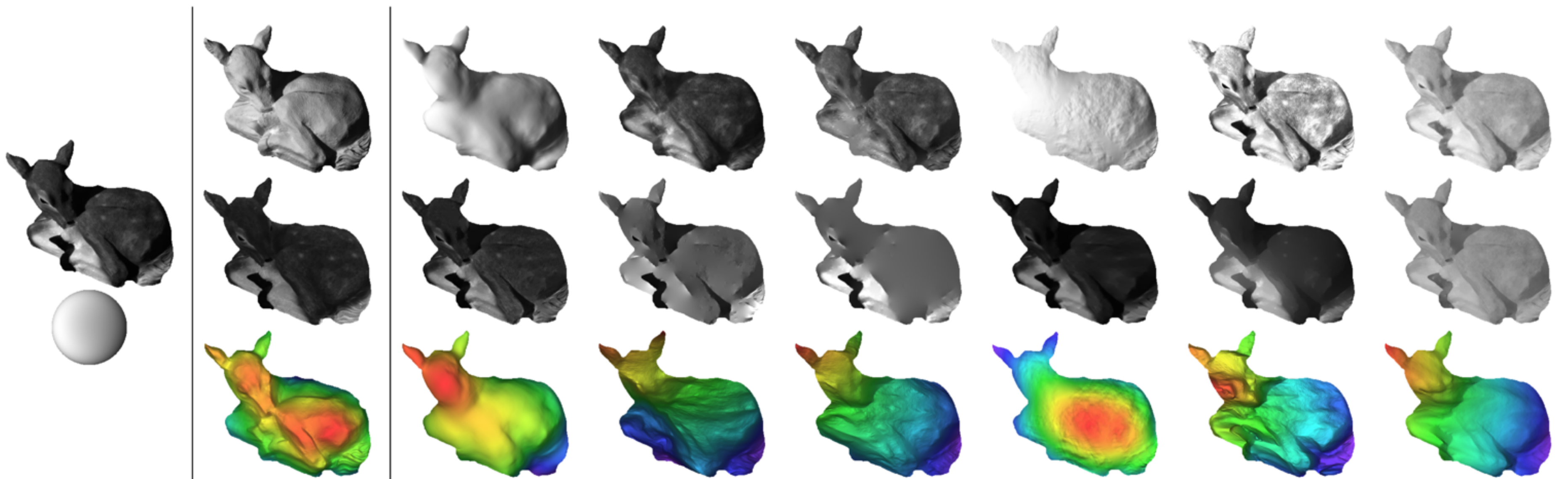
(a) Input Image & Illumination (b) Ground Truth (c) **Our Model** (d) Retinex[5, 4] + SFS (e) Tappen *et al.* 2005[10] + SFS (f) Barron & Malik 2010[1] (g) J. Shen *et al.* 2010[8] + SFS (h) L. Shen & Yeo 2010[9] + SFS

Evaluation: Known Lighting



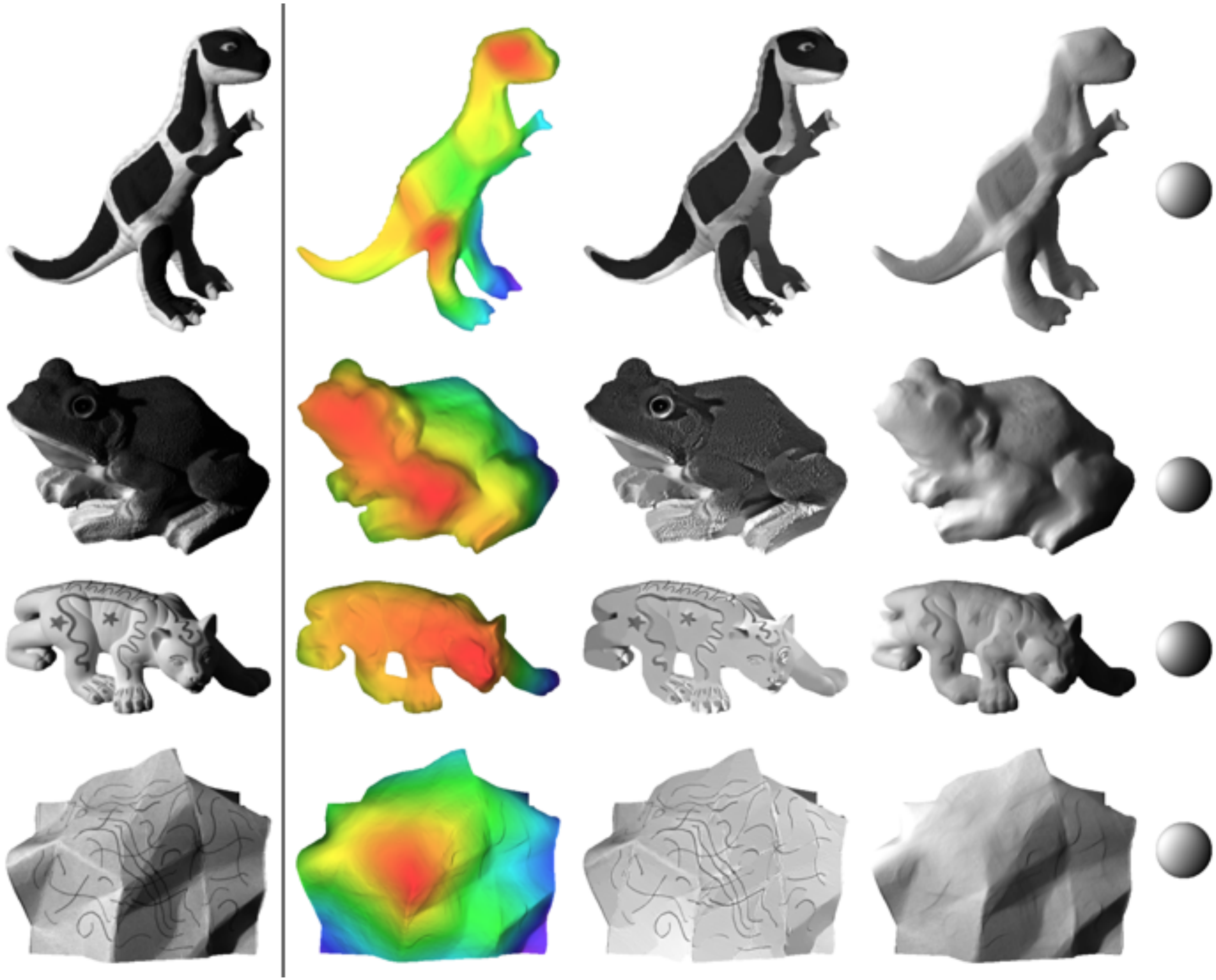
(i) Input Image & Illumination | (j) Ground Truth | (k) **Our Model** | (l) Retinex[5, 4] + SFS | (m) Tappen *et al.* 2005[10] + SFS | (n) Barron & Malik 2010[1] | (o) J. Shen *et al.* 2010[8] + SFS | (p) L. Shen & Yeo 2010[9] + SFS

Evaluation: Known Lighting

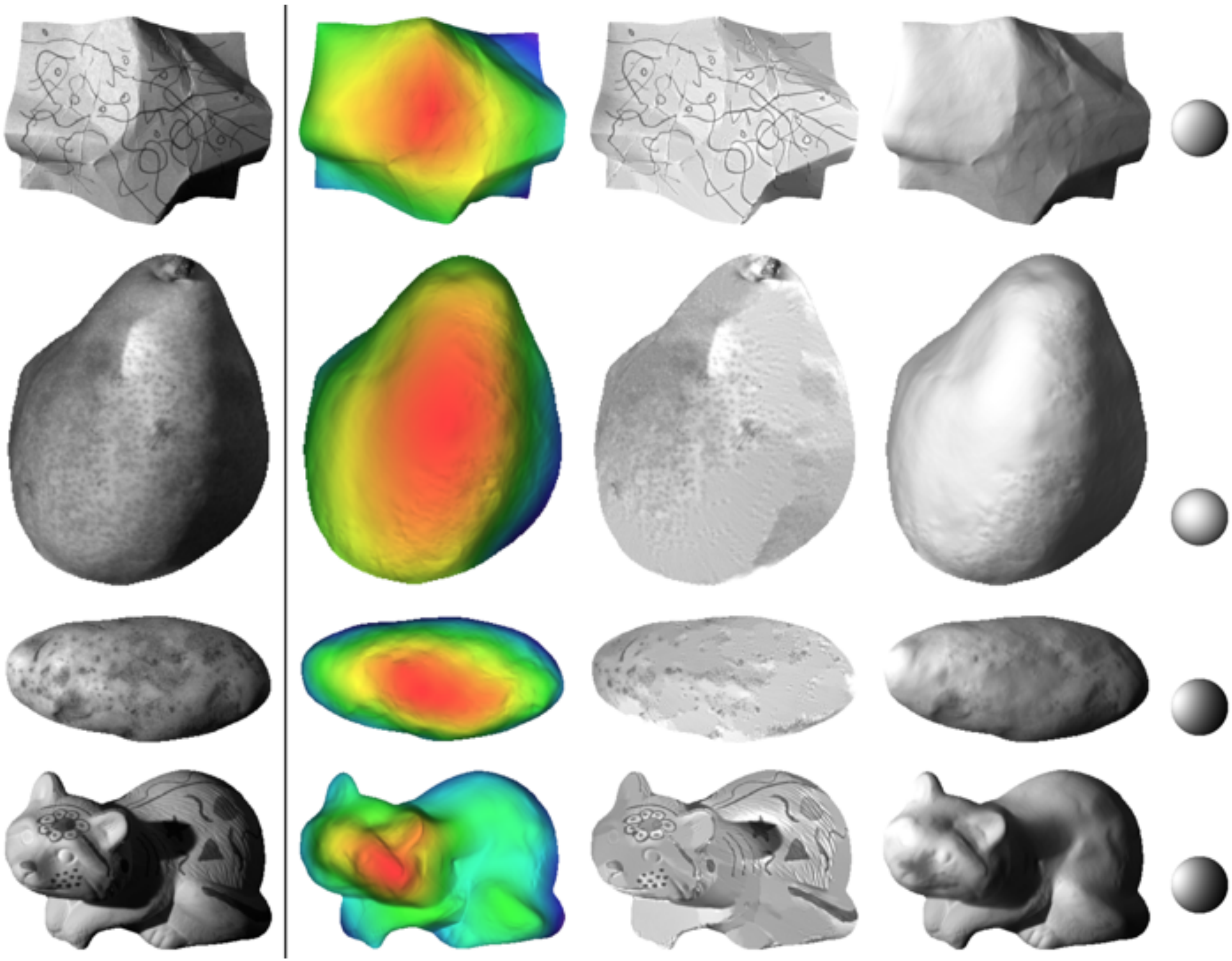


(q) Input Image & Illumination | (r) Ground Truth | (s) **Our Model** | (t) Retinex[5, 4] + SFS | (u) Tappen *et al.* 2005[10] + SFS | (v) Barron & Malik 2010[1] | (w) J. Shen *et al.* 2010[8] + SFS | (x) L. Shen & Yeo 2010[9] + SFS

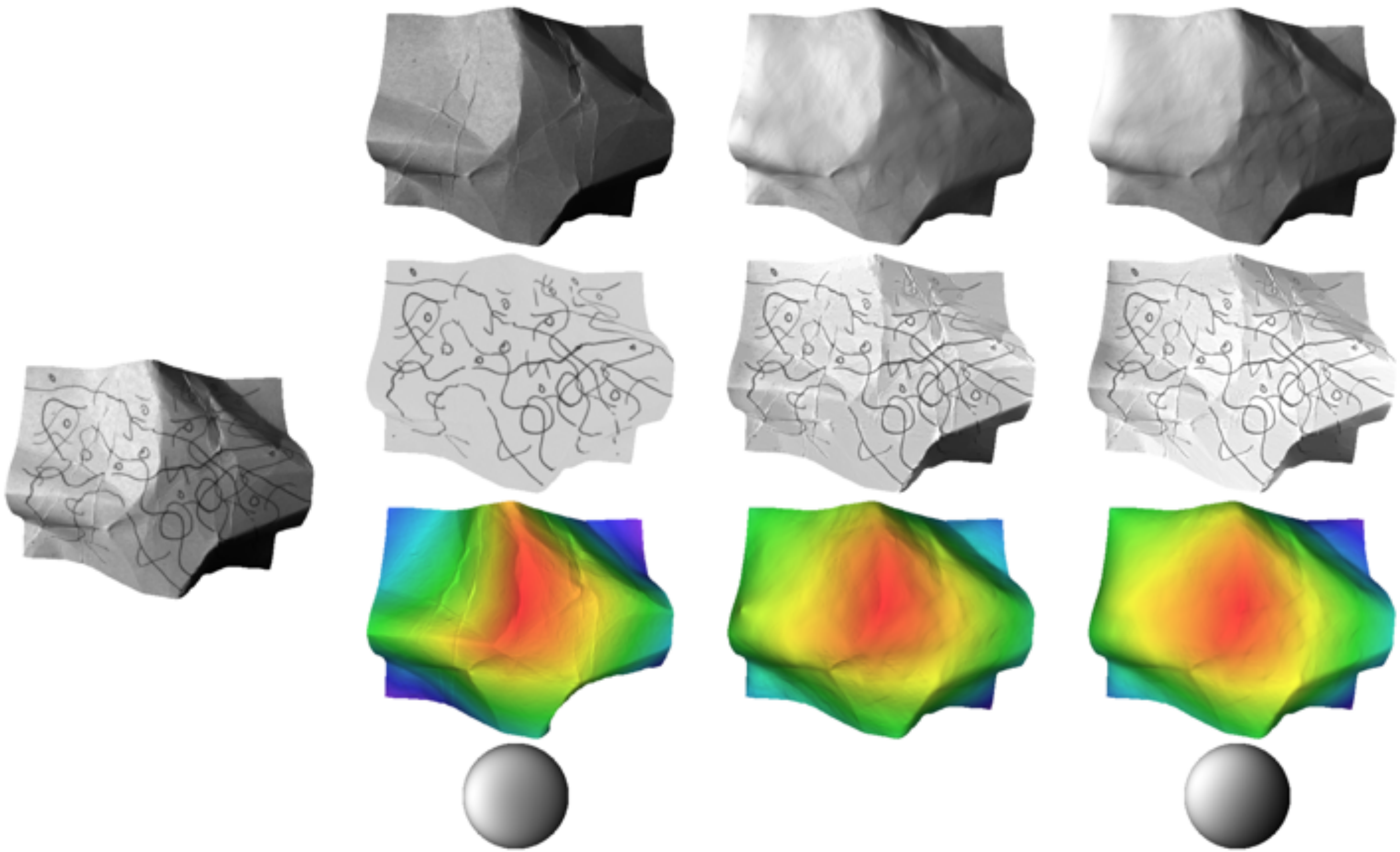
Evaluation: Unknown Lighting



Evaluation: Unknown Lighting

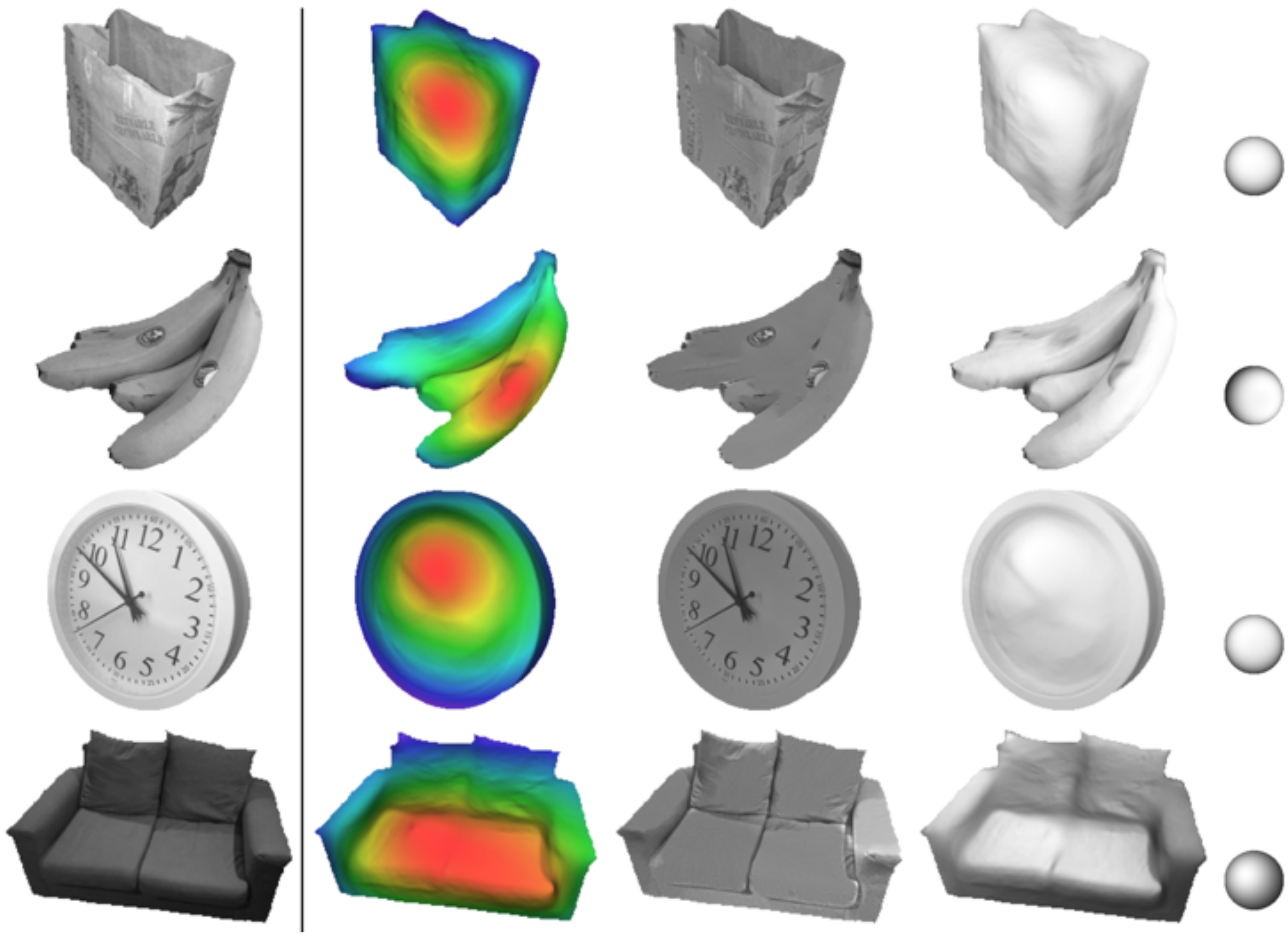


Evaluation: Known vs Unknown

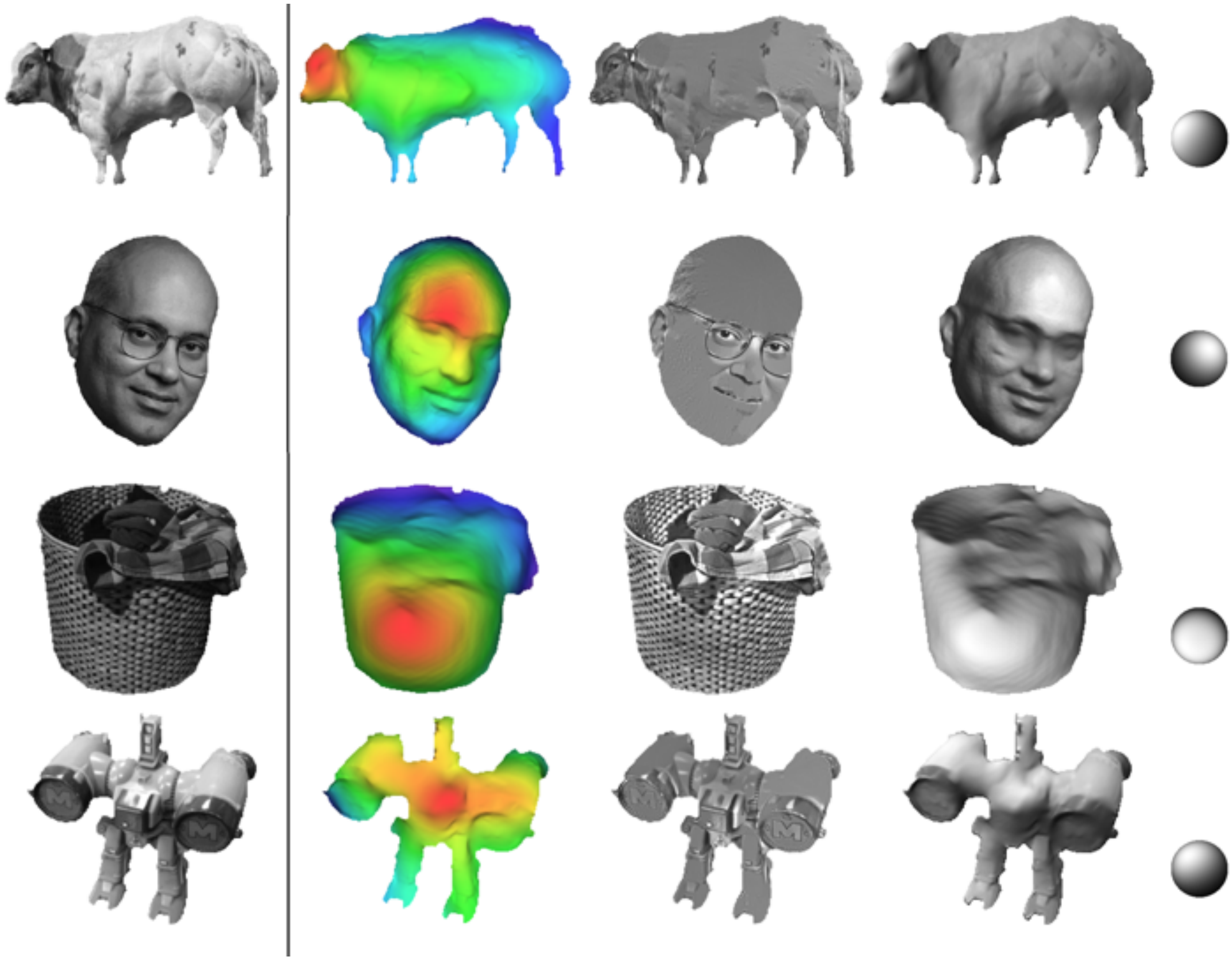


(a) Input Image (b) Ground Truth (c) Known Illumination (d) Unknown Illumination

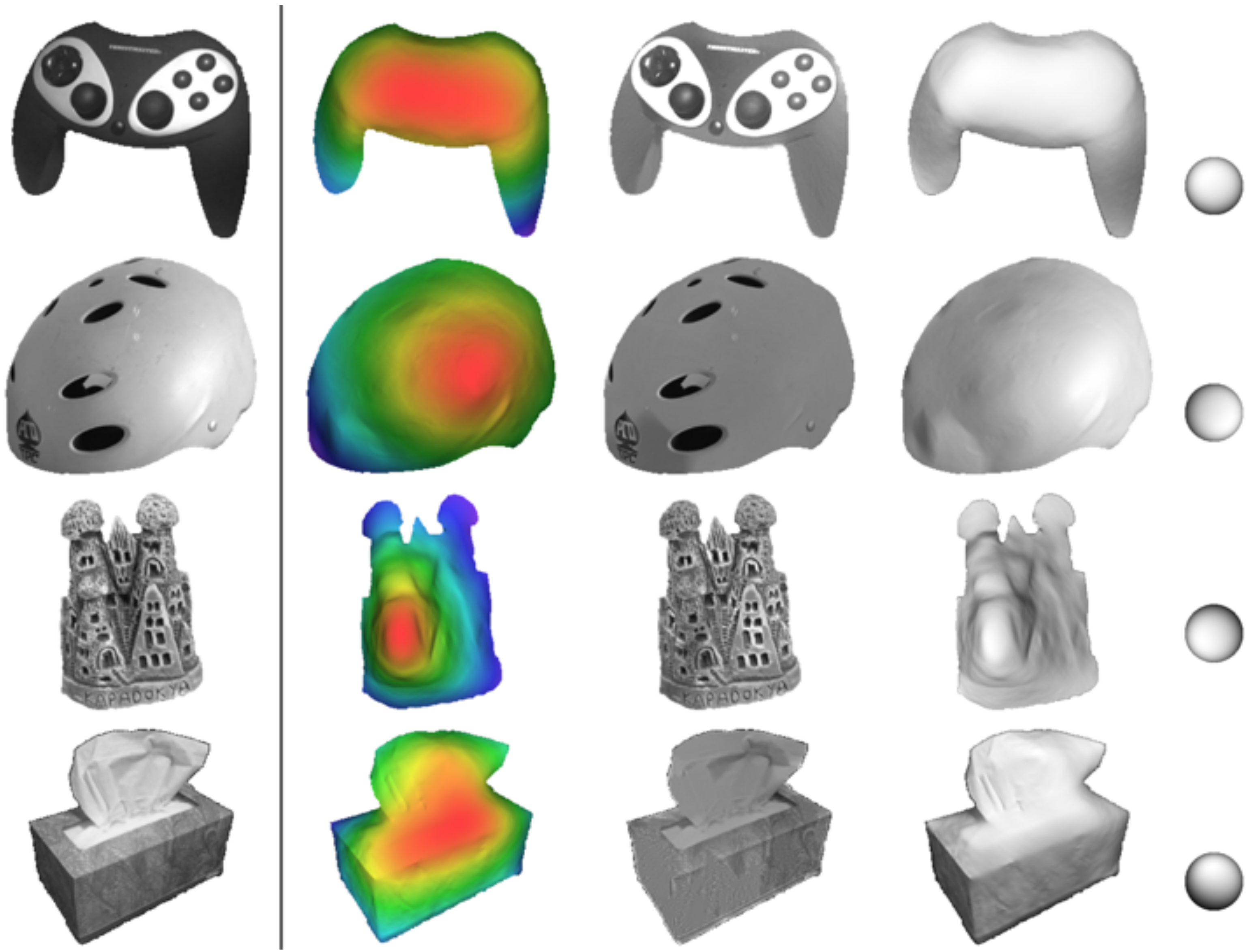
Evaluation: Real World Images



Evaluation: Real World Images



Evaluation: Real World Images



Evaluation: The Numbers

Known Illumination

Algorithm	Avg.
Flat Baseline	0.2004
Retinex + SFS	0.2009
Tappen *et al*. 2005 + SFS	0.1761
Barron & Malik 2011	0.1682
J. Shen *et al*. 2011 + SFS	0.2376
Our Model (All Priors)	0.0856

Evaluation: The Numbers

Known Illumination

Algorithm	Avg.
Flat Baseline	0.2004
Retinex + SFS	0.2009
Tappen *et al*. 2005 + SFS	0.1761
Barron & Malik 2011	0.1682
J. Shen *et al*. 2011 + SFS	0.2376
Our Shape from Contour	0.1394
Our Model (No $\|\|\nabla A\|\|$)	0.1070
Our Model (No $\|\|\nabla H(Z)\|\|$)	0.1244
Our Model (No Flatness)	0.1002
Our Model (No Contour)	0.1082
Our Model (No Albedo Entropy)	0.0865
Our Model (All Priors)	0.0856

Evaluation: The Numbers

Known Illumination

Algorithm	Avg.
Flat Baseline	0.2004
Retinex + SFS	0.2009
Tappen *et al*. 2005 + SFS	0.1761
Barron & Malik 2011	0.1682
J. Shen *et al*. 2011 + SFS	0.2376
Our Shape from Contour	0.1394
Our Model (No $\|\|\nabla A\|\|$)	0.1070
Our Model (No $\|\|\nabla H(Z)\|\|$)	0.1244
Our Model (No Flatness)	0.1002
Our Model (No Contour)	0.1082
Our Model (No Albedo Entropy)	0.0865
Our Model (All Priors)	0.0856

Evaluation: The Numbers

Known Illumination

Algorithm	Z-MAE	I-MSE	LMSE	S-MSE	ρ-MSE	Avg.
Flat Baseline	25.56	0.1369	0.0385	0.0563	0.0427	0.2004
Retinex + SFS	82.06	0.1795	0.0289	0.0291	0.0264	0.2009
Tappen *et al.* 2005 + SFS	43.30	0.1522	0.0292	0.0343	0.0256	0.1761
Barron & Malik 2011	21.10	0.0829	0.0584	0.0282	0.0468	0.1682
J. Shen *et al.* 2011 + SFS	48.51	0.1629	0.0445	0.0478	0.0450	0.2376
Our Shape from Contour	21.42	0.0805	0.0350	0.0280	0.0311	0.1394
Our Model (No $\|\|\nabla A\|\|$)	17.50	0.0620	0.0289	0.0188	0.0238	0.1070
Our Model (No $\|\|\nabla H(Z)\|\|$)	21.81	0.1011	0.0341	0.0205	0.0194	0.1244
Our Model (No Flatness)	35.11	0.0651	0.0190	0.0148	0.0157	0.1002
Our Model (No Contour)	28.45	0.0811	0.0204	0.0167	0.0189	0.1082
Our Model (No Albedo Entropy)	21.23	0.0523	0.0196	0.0138	0.0162	0.0865
Our Model (All Priors)	21.86	0.0521	0.0191	0.0136	0.0156	0.0856

Evaluation: The Numbers

Known Illumination

Algorithm	Z-MAE	I-MSE	LMSE	S-MSE	ρ-MSE	Avg.
Flat Baseline	25.56	0.1369	0.0385	0.0563	0.0427	0.2004
Retinex + SFS	82.06	0.1795	0.0289	0.0291	0.0264	0.2009
Tappen *et al.* 2005 + SFS	43.30	0.1522	0.0292	0.0343	0.0256	0.1761
Barron & Malik 2011	21.10	0.0829	0.0584	0.0282	0.0468	0.1682
J. Shen *et al.* 2011 + SFS	48.51	0.1629	0.0445	0.0478	0.0450	0.2376
Our Shape from Contour	21.42	0.0805	0.0350	0.0280	0.0311	0.1394
Our Model (No $\|\|\nabla A\|\|$)	17.50	0.0620	0.0289	0.0188	0.0238	0.1070
Our Model (No $\|\|\nabla H(Z)\|\|$)	21.81	0.1011	0.0341	0.0205	0.0194	0.1244
Our Model (No Flatness)	35.11	0.0651	0.0190	0.0148	0.0157	0.1002
Our Model (No Contour)	28.45	0.0811	0.0204	0.0167	0.0189	0.1082
Our Model (No Albedo Entropy)	21.23	0.0523	0.0196	0.0138	0.0162	0.0865
Our Model (All Priors)	21.86	0.0521	0.0191	0.0136	0.0156	0.0856

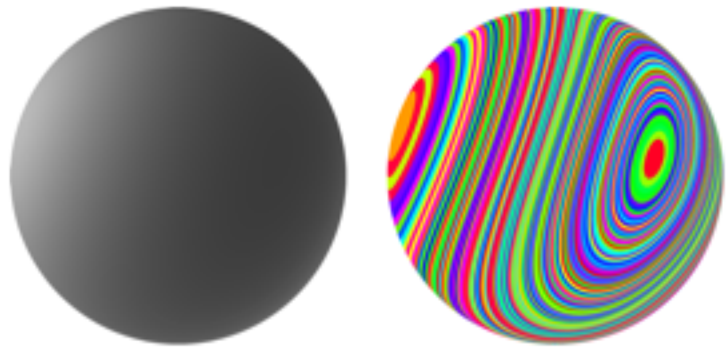
Unknown Illumination

Our Model (All Priors)	19.41	0.0577	0.0197	0.0178	0.0193	0.0946

Color!

Color: The Good

Color light tells you a lot about shape

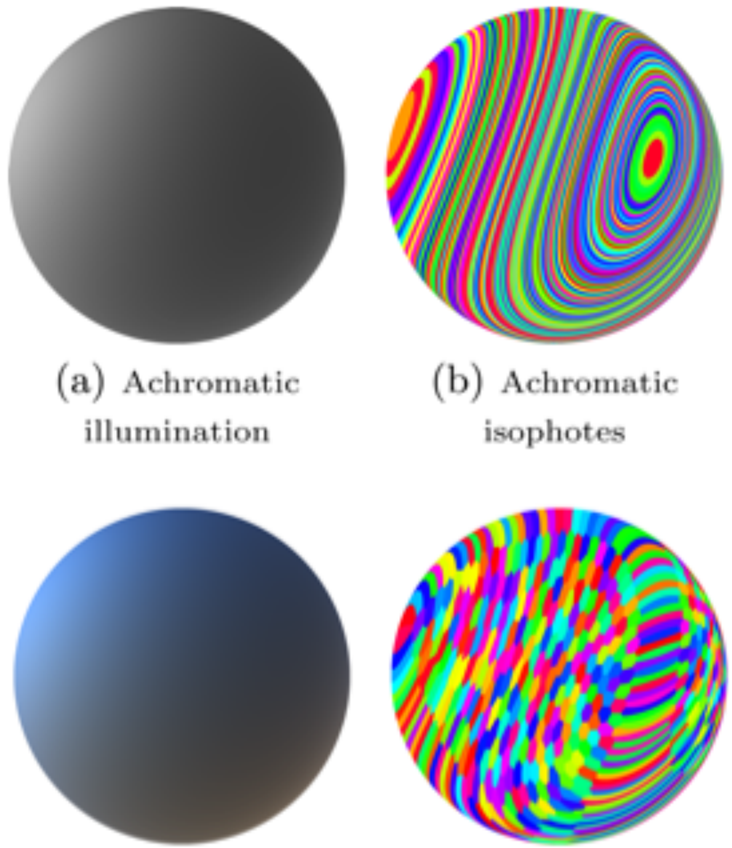


(a) Achromatic illumination

(b) Achromatic isophotes

Color: The Good

Color light tells you a lot about shape



(a) Achromatic illumination

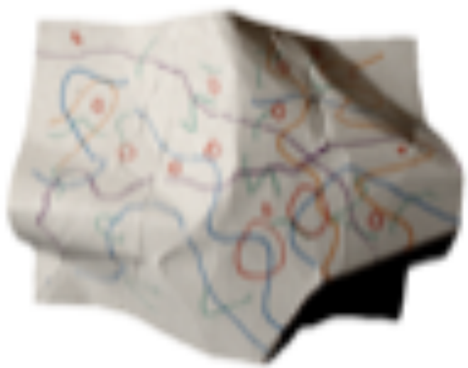
(b) Achromatic isophotes

(c) Chromatic illumination

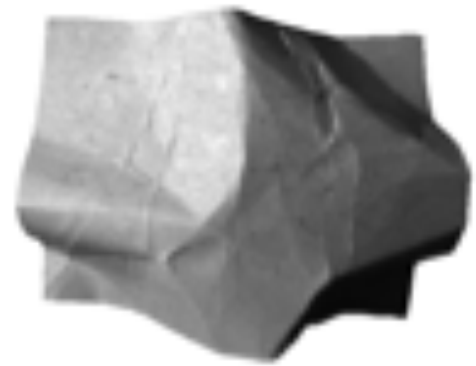
(d) Chromatic isophotes

Color: The Good

Color images help distinguish between albedo and shading...



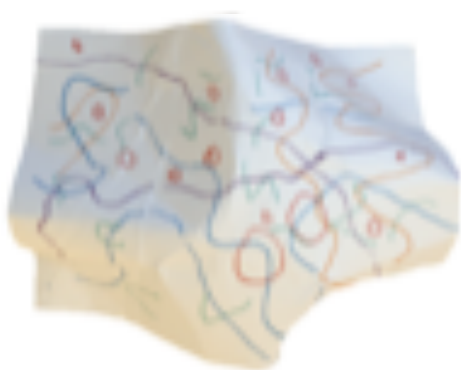
Input Image



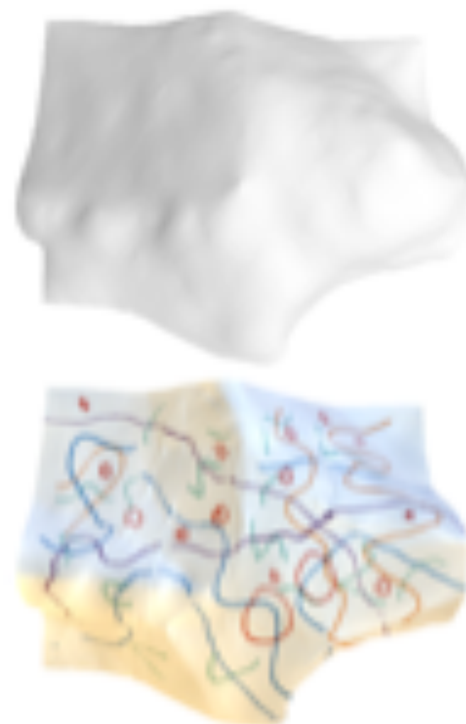
Gehler et al.
(the best-performing
intrinsic image algorithm)

Color: The Bad

...but things get tricky if illumination isn't white
(and illumination is almost never white)

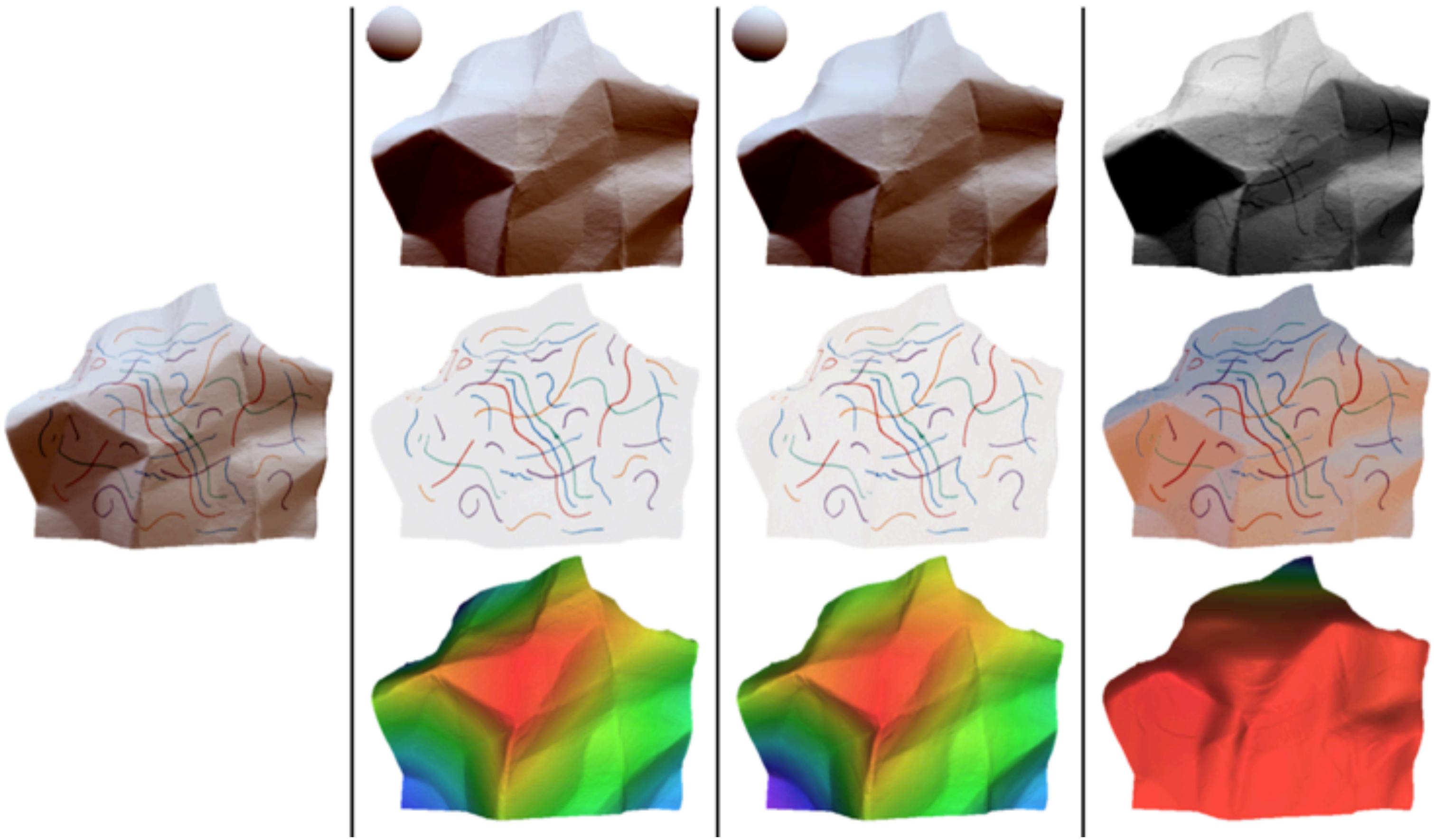


Input Image



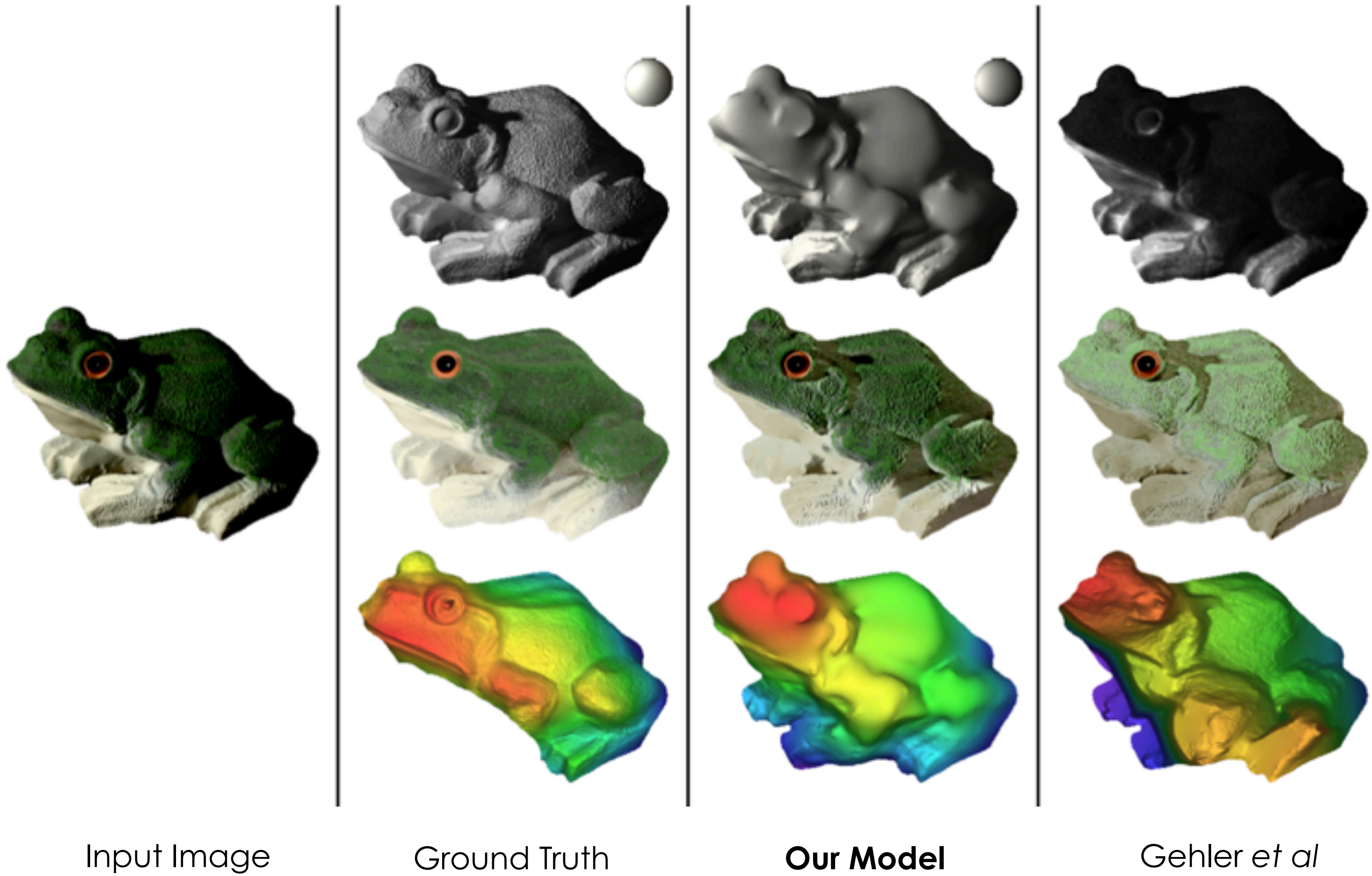
Gehler et al.
(the best-performing
intrinsic image algorithm)

Results

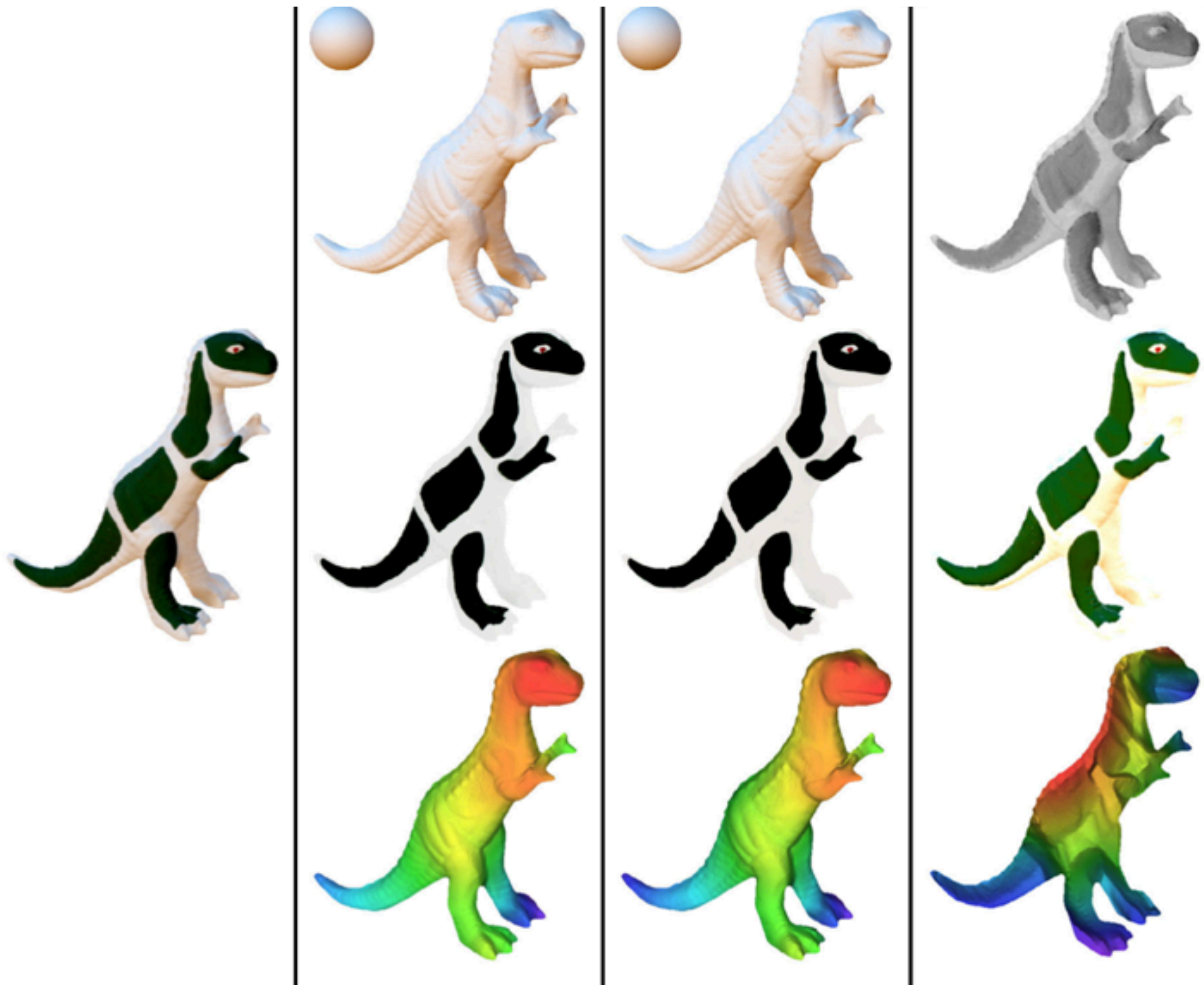


Input Image | Ground Truth | **Our Model** | Gehler *et al*

Results

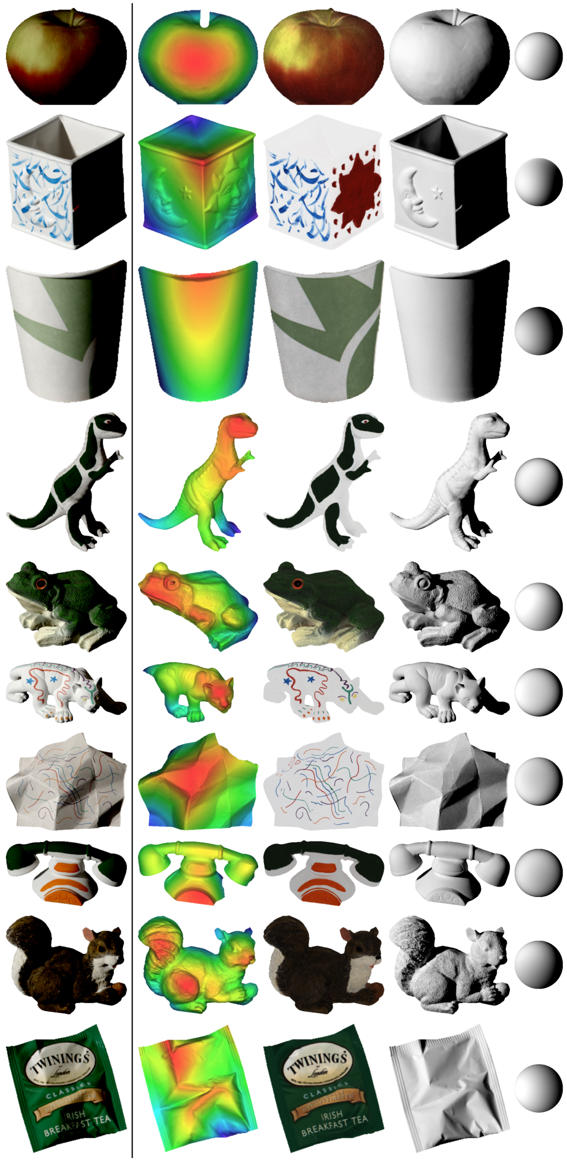


Results



Input Image | Ground Truth | **Our Model** | Gehler *et al*

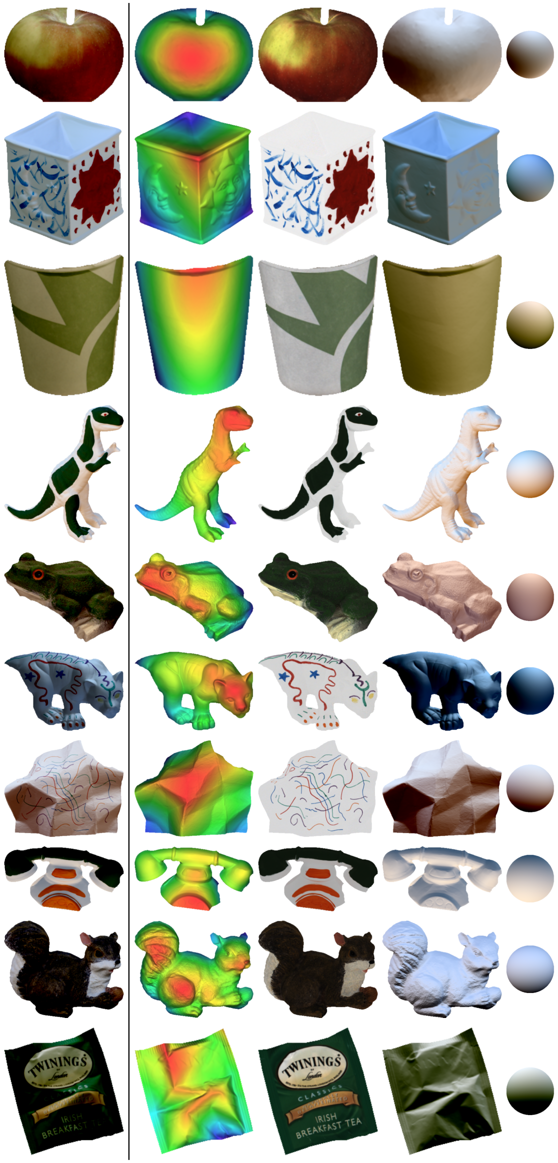
Results: Laboratory Illumination



Known Illumination

Algorithm	N-MSE	s-MSE	r-MSE	rs-MSE	L-MSE	Avg.
Flat Baseline	0.6141	0.0572	0.0452	0.0354	-	0.0866
Retinex [2, 5] + SFS [1]	0.8412	0.0204	0.0186	0.0163	-	0.0477
Tappen *et al.* 2005 [14] + SFS [1]	0.7052	0.0361	0.0379	0.0347	-	0.0760
Shen *et al.* 2011 [15] + SFS [1]	0.9232	0.0528	0.0458	0.0398	-	0.0971
Gehler *et al.* 2011 [12] + SFS [1]	0.6342	0.0106	0.0101	0.0131	-	0.0307
Barron & Malik 2012A [1]	0.2032	0.0142	0.0160	0.0181	-	0.0302
Shape from Contour [1]	0.2464	0.0296	0.0412	0.0309	-	0.0552
Our Model (Complete)	0.2151	0.0066	0.0115	0.0133	-	0.0215
Unknown Illumination						
Barron & Malik 2012A [1]	0.1975	0.0194	0.0224	0.0190	0.0247	0.0332
Our Model (Complete)	0.2793	0.0075	0.0112	0.0136	0.0085	0.0188

Results: Natural Illumination



Known Illumination

Algorithm	N-MSE	s-MSE	r-MSE	rs-MSE	L-MSE	Avg.
Flat Baseline	0.6141	0.0246	0.0243	0.0125	-	0.0463
Retinex [2, 5] + SFS [1]	0.4258	0.0174	0.0174	0.0083	-	0.0322
Tappen *et al.* 2005 [14] + SFS [1]	0.6707	0.0255	0.0280	0.0268	-	0.0599
Gehler *et al.* 2011 [12] + SFS [1]	0.5549	0.0162	0.0150	0.0105	-	0.0346
Gehler *et al.* 2011 [12] + [11] + SFS [1]	0.6282	0.0163	0.0164	0.0106	-	0.0365
Barron & Malik 2012A [1]	0.2044	0.0092	0.0094	0.0081	-	0.0195
Shape from Contour [1]	0.2502	0.0126	0.0163	0.0106	-	0.0271
Our Model (Complete)	0.0867	0.0022	0.0017	0.0026	-	0.0054
Unknown Illumination						
Barron & Malik 2012A [1]	0.2172	0.0193	0.0188	0.0094	0.0206	0.0273
Our Model (Complete)	0.2348	0.0060	0.0049	0.0042	0.0084	0.0119

Evaluation: Graphics!

Conclusions

- Unification shape-from-shading, intrinsic images, and color constancy

Conclusions

- Unification shape-from-shading, intrinsic images, and color constancy
- Solving the unified problem > Solving any sub-problem

Conclusions

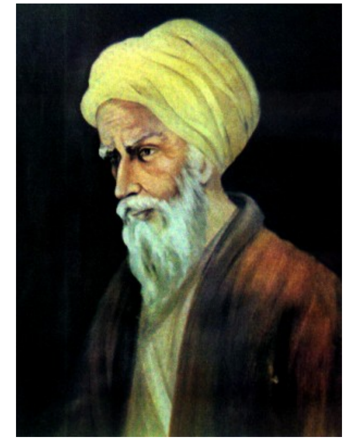
- Unification shape-from-shading, intrinsic images, and color constancy
- Solving the unified problem > Solving any sub-problem
- Not a toy

Conclusions

- Unification shape-from-shading, intrinsic images, and color constancy
- Solving the unified problem > Solving any sub-problem
- Not (and can never be?) metrically accurate

Closing thoughts…

"Nothing of what is visible, apart from light and color, can be perceived by pure sensation, but only by discernment, inference, and recognition, in addition to sensation."



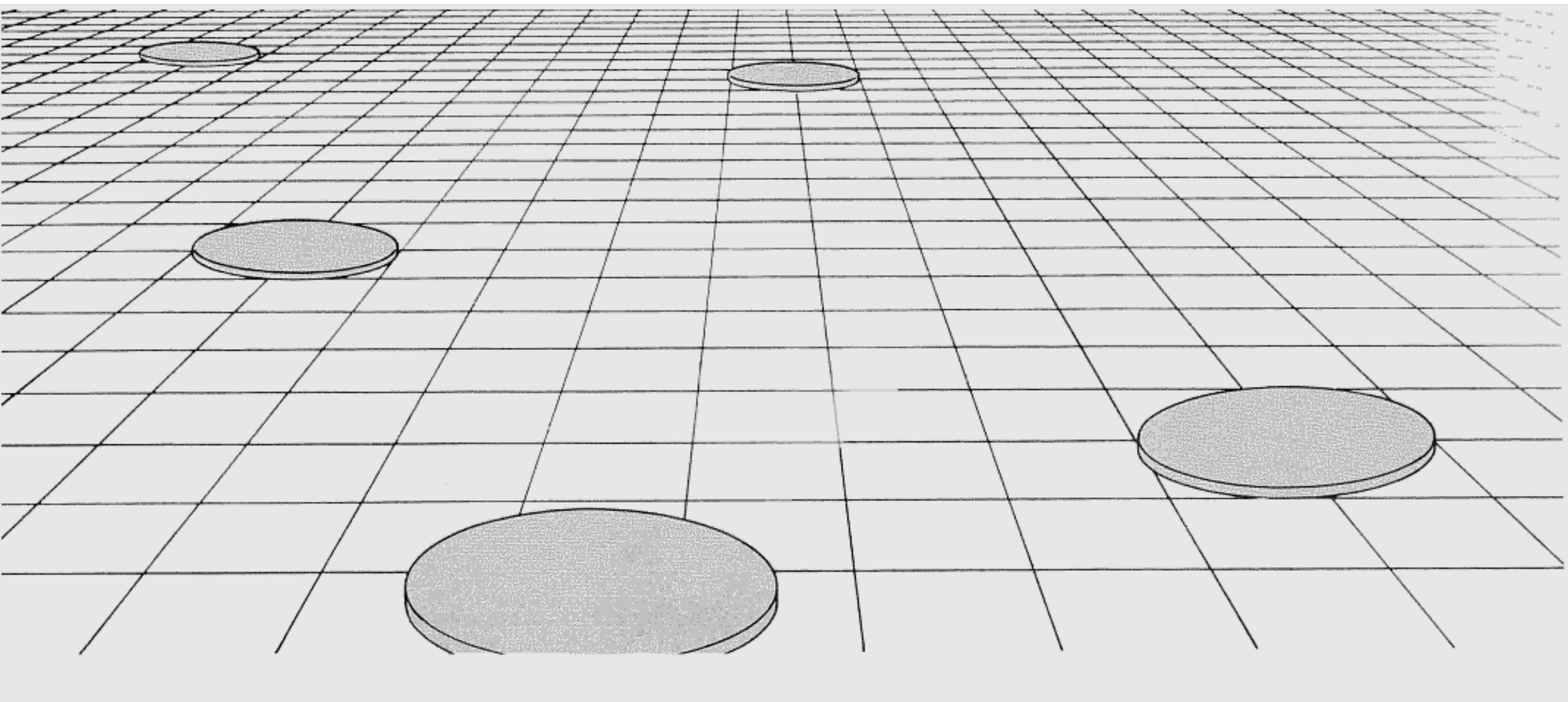
$$F_*(\mathbf{p}) = \begin{bmatrix} r/\cos\sigma & 0 \\ 0 & r \end{bmatrix}$$

Alhazen
965-1040

"Vision can only be the result of some form of unconscious inferences: a matter of making assumptions and conclusions from incomplete data, based on previous experiences."

Hermann von Helmholtz
1821-1894

Texture gradient cues



Journal of Mathematical Imaging and Vision 2, 327–350 (1992).

Shape from Texture for Smooth Curved Surfaces in Perspective Projection

JONAS GÅRDING
Computational Vision and Active Perception Laboratory (CVAP), Department of Numerical Analysis and Computing Science, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Abstract. Projective distortion of surface texture observed in a perspective image can provide direct information about the shape of the underlying surface. Previous theories have generally concerned planar surfaces; this paper presents a systematic analysis of first- and second-order texture distortion cues for the case of a smooth, curved surface. In particular, several kinds of texture gradients are analyzed and are related to surface orientation and surface curvature. The local estimates obtained from these cues can be integrated to obtain a global surface shape, and it is shown that the two surfaces resulting from the well-known tilt ambiguity in the local foreshortening cue typically have qualitatively different shapes. As an example of a practical application of the analysis, a shape-from-texture algorithm based on local orientation-selective filtering is described, and some experimental results are shown.

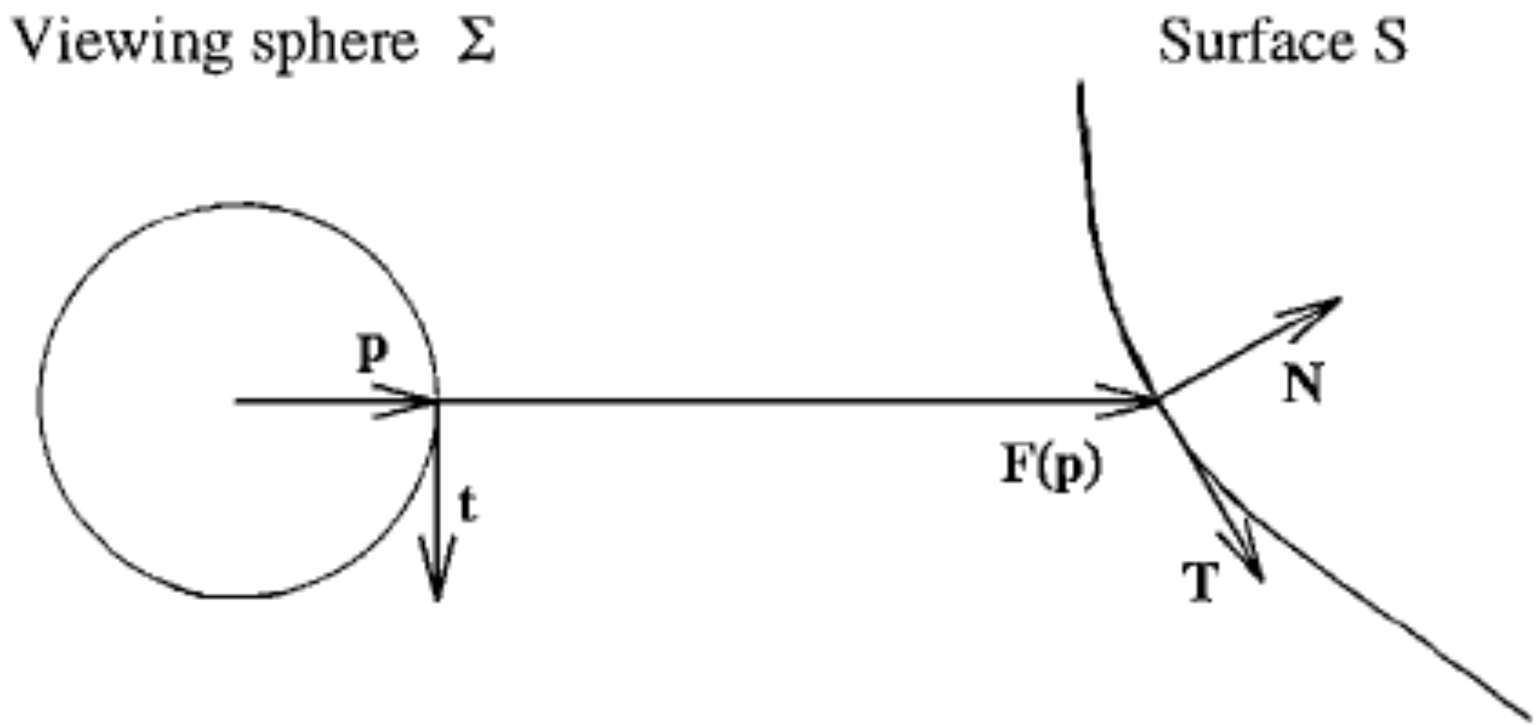
Computing Local Surface Orientation and Shape from Texture for Curved Surfaces

JITENDRA MALIK AND RUTH ROSENHOLTZ*
Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA 94720
malik@cs.berkeley.edu
rruth@parc.xerox.com

Received April 5, 1994; Accepted September 27, 1994

Abstract. Shape from texture is best analyzed in two stages, analogous to stereopsis and structure from motion: (a) Computing the 'texture distortion' from the image, and (b) Interpreting the 'texture distortion' to infer the orientation and shape of the surface in the scene. We model the texture distortion for a given point and direction on the image plane as an affine transformation and derive the relationship between the parameters of this transformation and the shape parameters. We have developed a technique for estimating affine transforms between nearby image patches which is based on solving a system of linear constraints derived from a differential analysis. One need not explicitly identify texels or make restrictive assumptions about the nature of the texture such as isotropy. We use non-linear minimization of a least squares error criterion to recover the surface orientation (slant and tilt) and shape (principal curvatures and directions) based on the estimated affine transforms in a number of different directions. A simple linear algorithm based on singular value decomposition of the linear parts of the affine transforms provides the initial guess for the minimization procedure. Experimental results on both planar and curved surfaces under perspective projection demonstrate good estimates for both orientation and shape. A sensitivity analysis yields predictions for both computer vision algorithms and human perception of shape from texture.

Projection mapping



$$F_*(\mathbf{p}) = \begin{bmatrix} r/\cos\sigma & 0 \\ 0 & r \end{bmatrix}$$

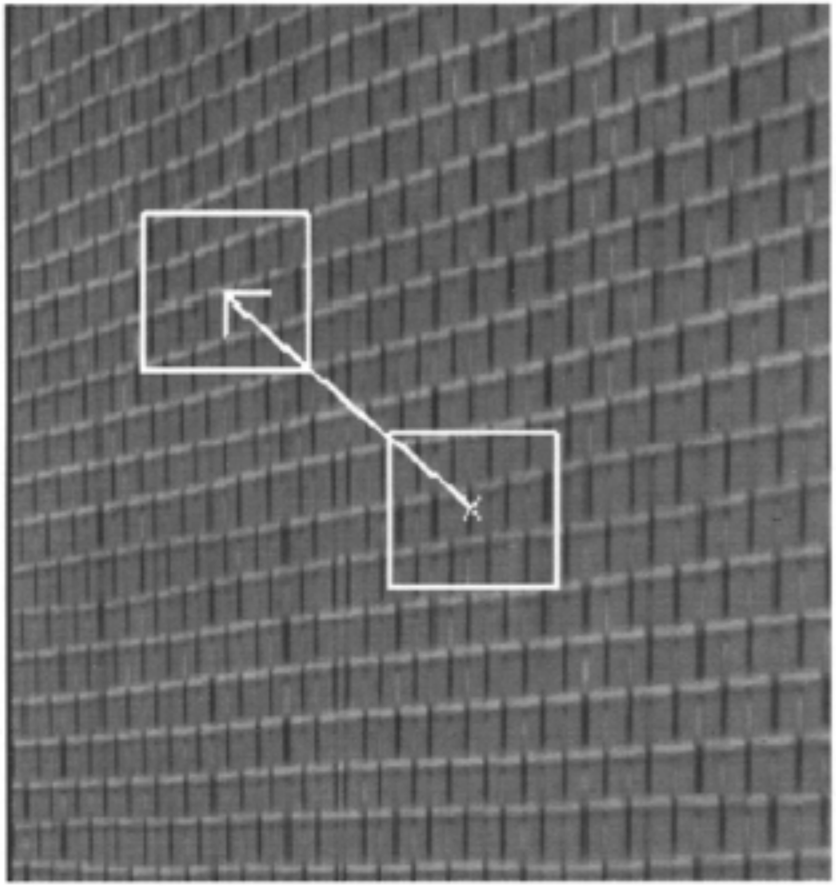


FIGURE 2. The texture distortion between two image patches can be modeled as an affine transformation: $[x', y']^T = A[x, y]^T + [\Delta x, \Delta y]^T$, where A is a 2×2 matrix. A depends on the local surface shape and orientation. A computational model of how this affine texture distortion can be used to recover the local surface geometry has been presented in (Malik & Rosenholtz, 1994, 1996).

Thanks!