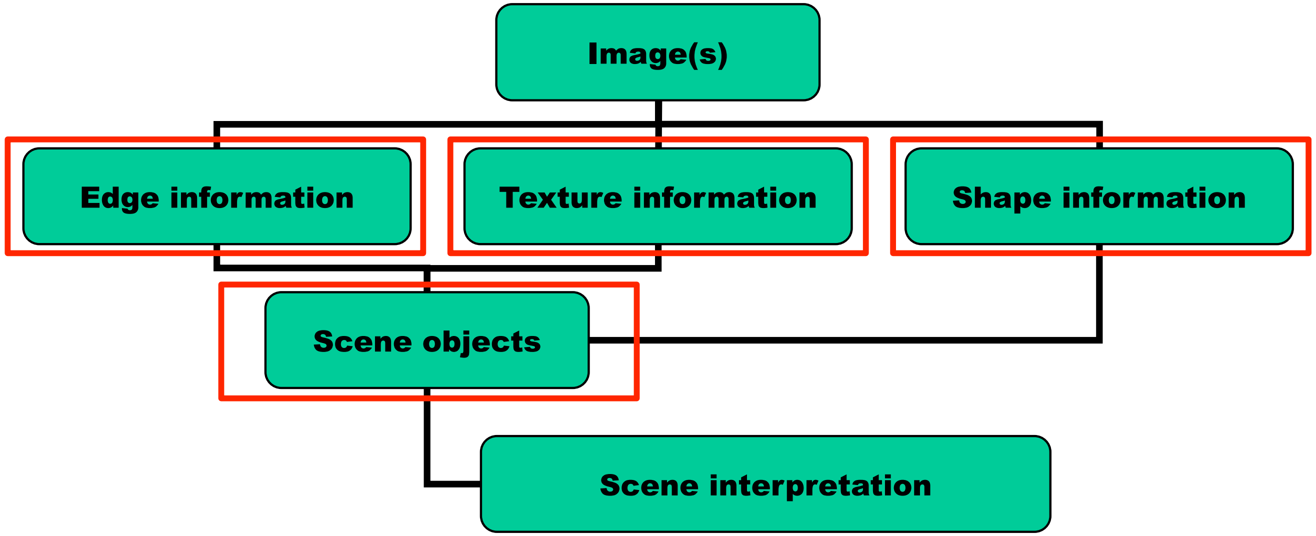
Reminder: A Teachable Scheme



Decomposition of the vision process into smaller manageable and implementable steps.

--> Paradigm followed in this course

--> May not be the one humans use

Shape From X

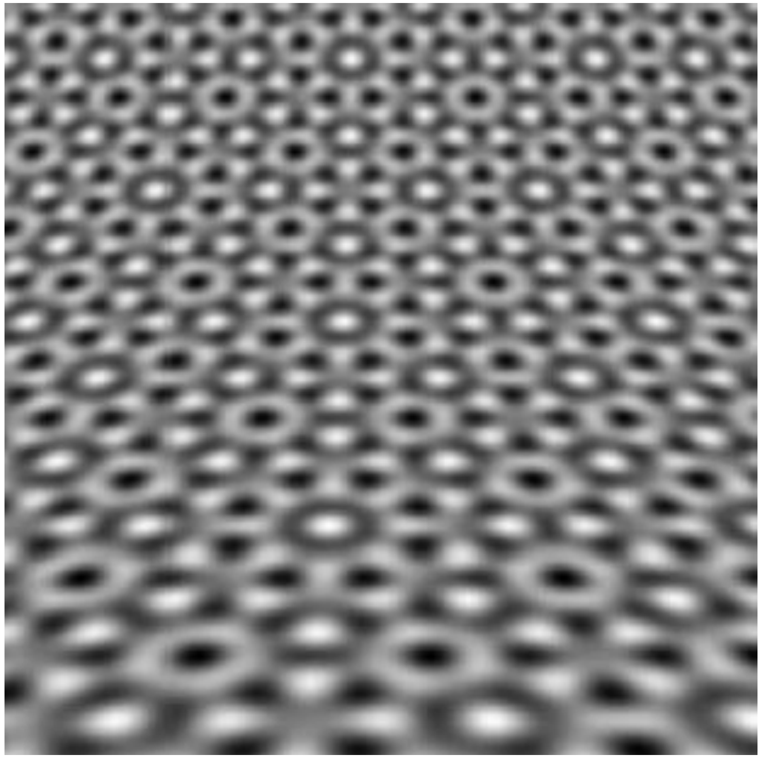
- One image:
  - **Shading**
  - Texture

- Two images or more:
  - Stereo
  - Contours
  - Motion

Shape From X

- One image:
    - Shading
    - **Texture**

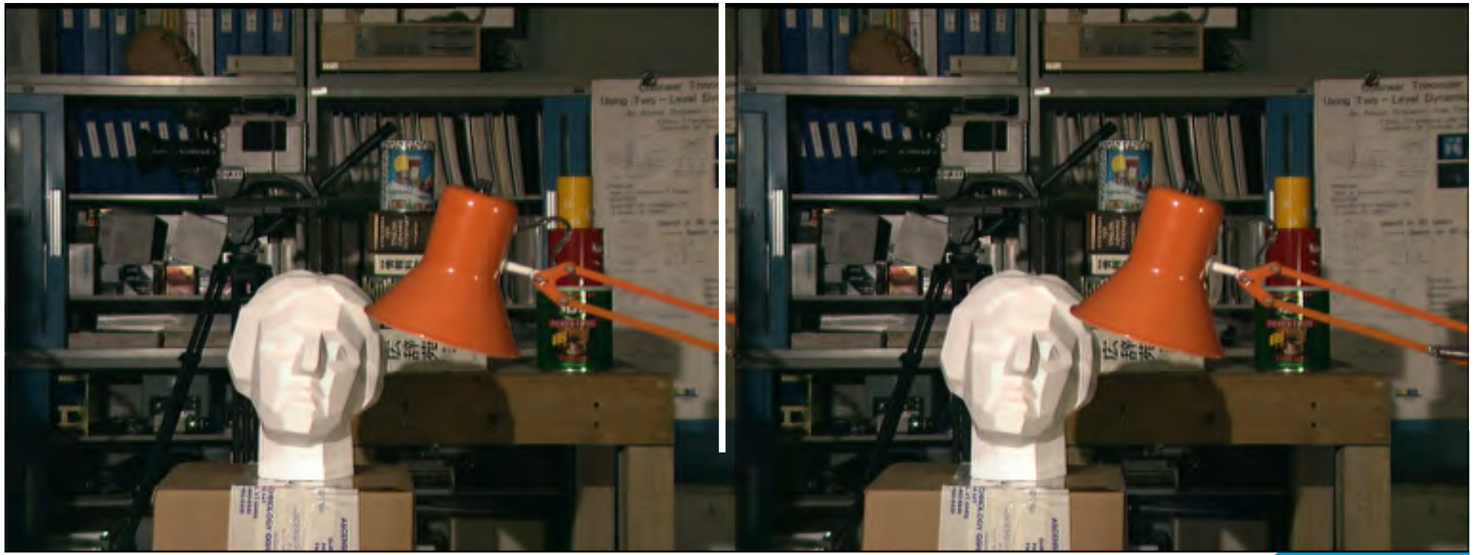
- Two images or more:
    - Stereo
    - Contours
    - Motion



Shape From X

- One image:
  - Shading
  - Texture

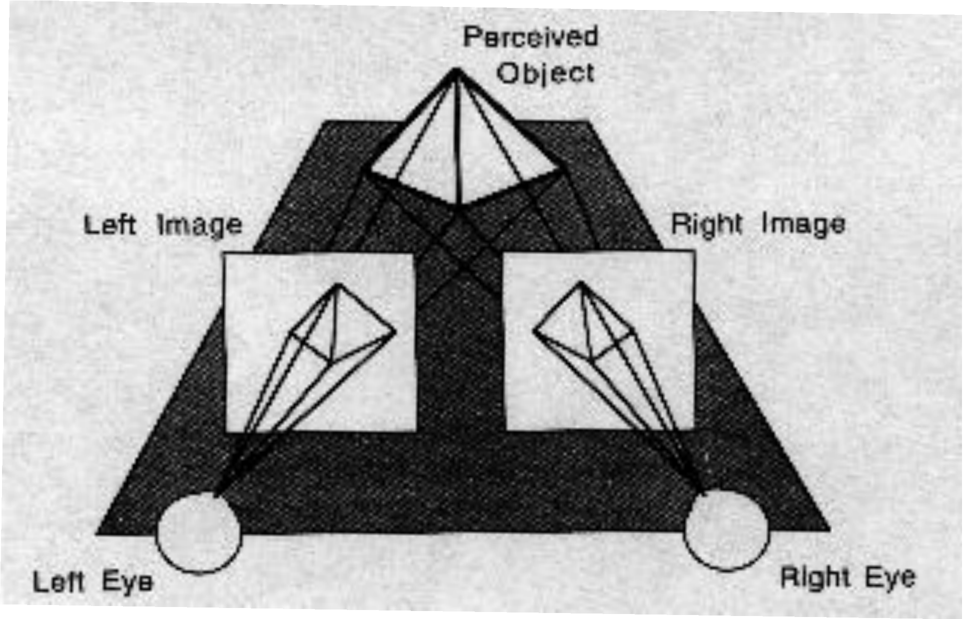
- Two images or more:
  - **Stereo**
  - Contours
  - Motion



Shape From X

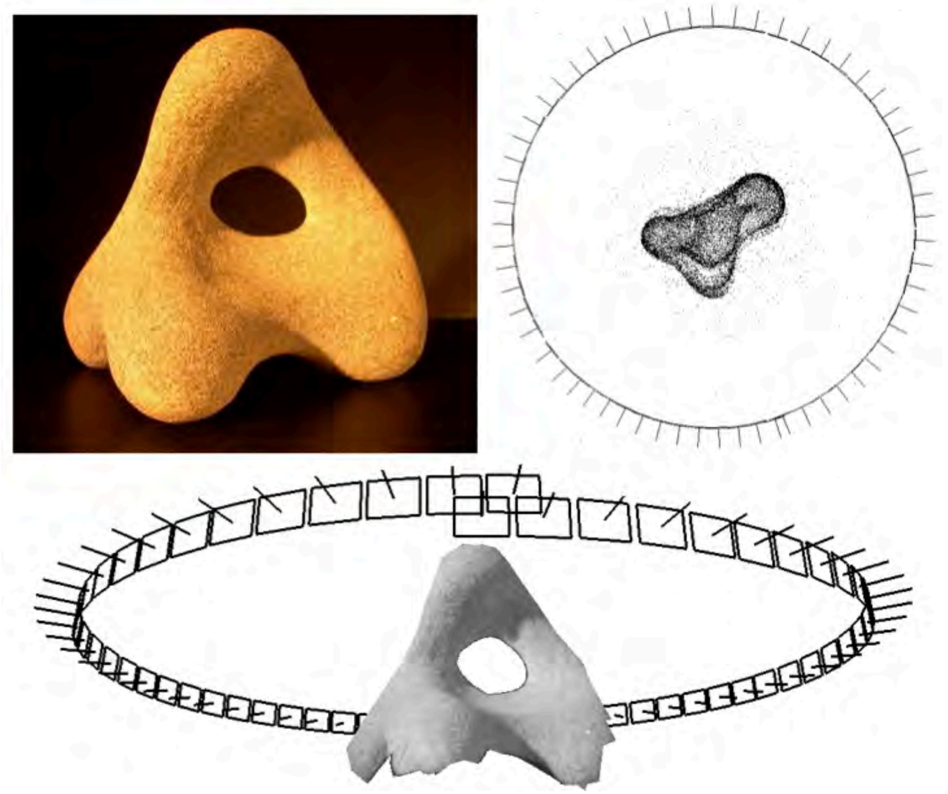
- One image:
  - Shading
  - Texture

- Two images or more:
  - Stereo
  - **Contours**
  - Motion



Shape From X

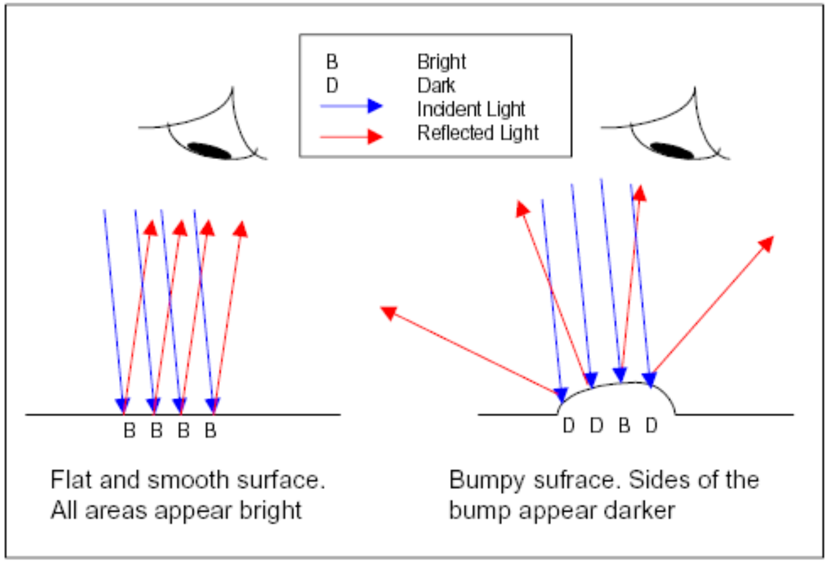
- One image:
  - Shading
  - Texture
- Two images or more:
  - Stereo
  - Contours
  - **Motion**

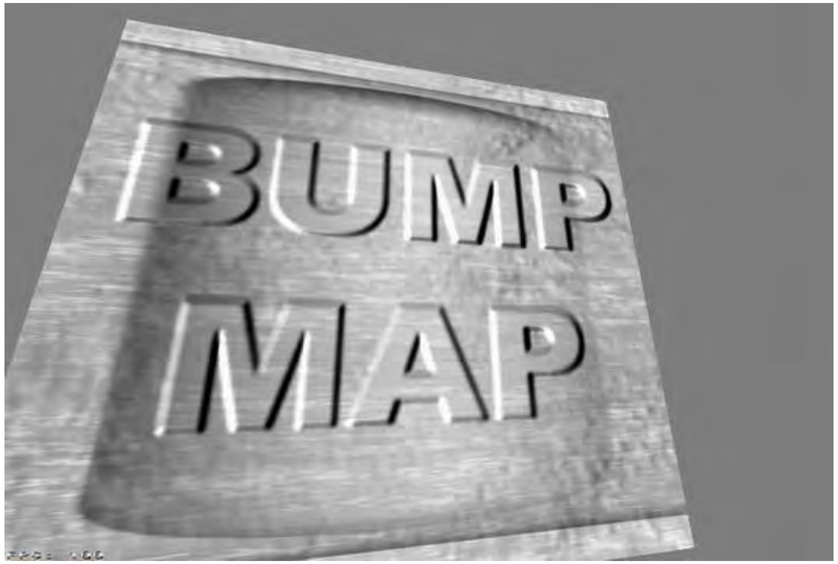


Shading

- Shading models
- Shape from shading
  - Variational Methods
  - Deep Learning Methods
  - Photometric Stereo

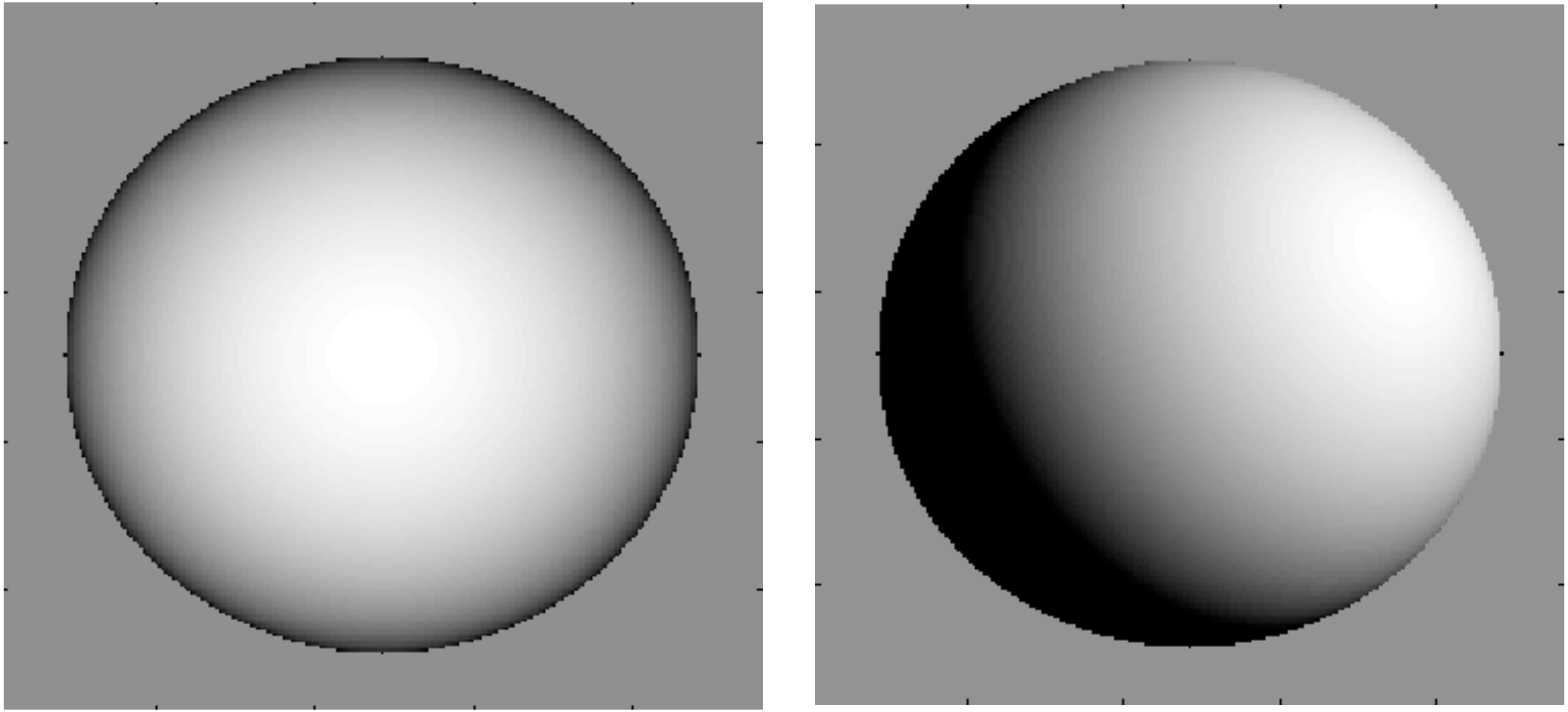
Bump Mapping





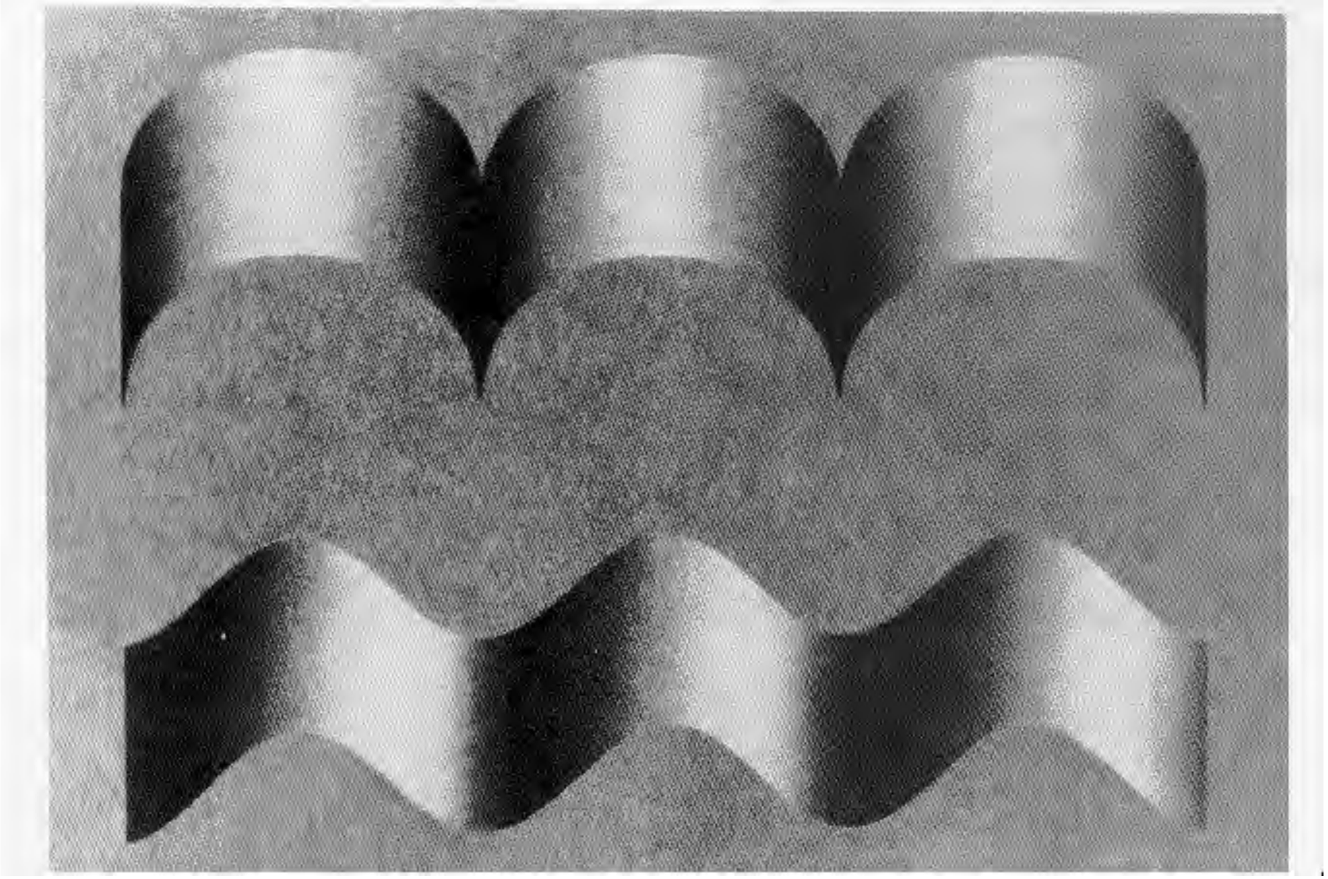
Simple mesh + 2D bump map = Complex looking object

Lambertian Half-Sphere



Gray level changes are interpreted as changes in the direction of the surface normal.

Solving an Inverse Problem



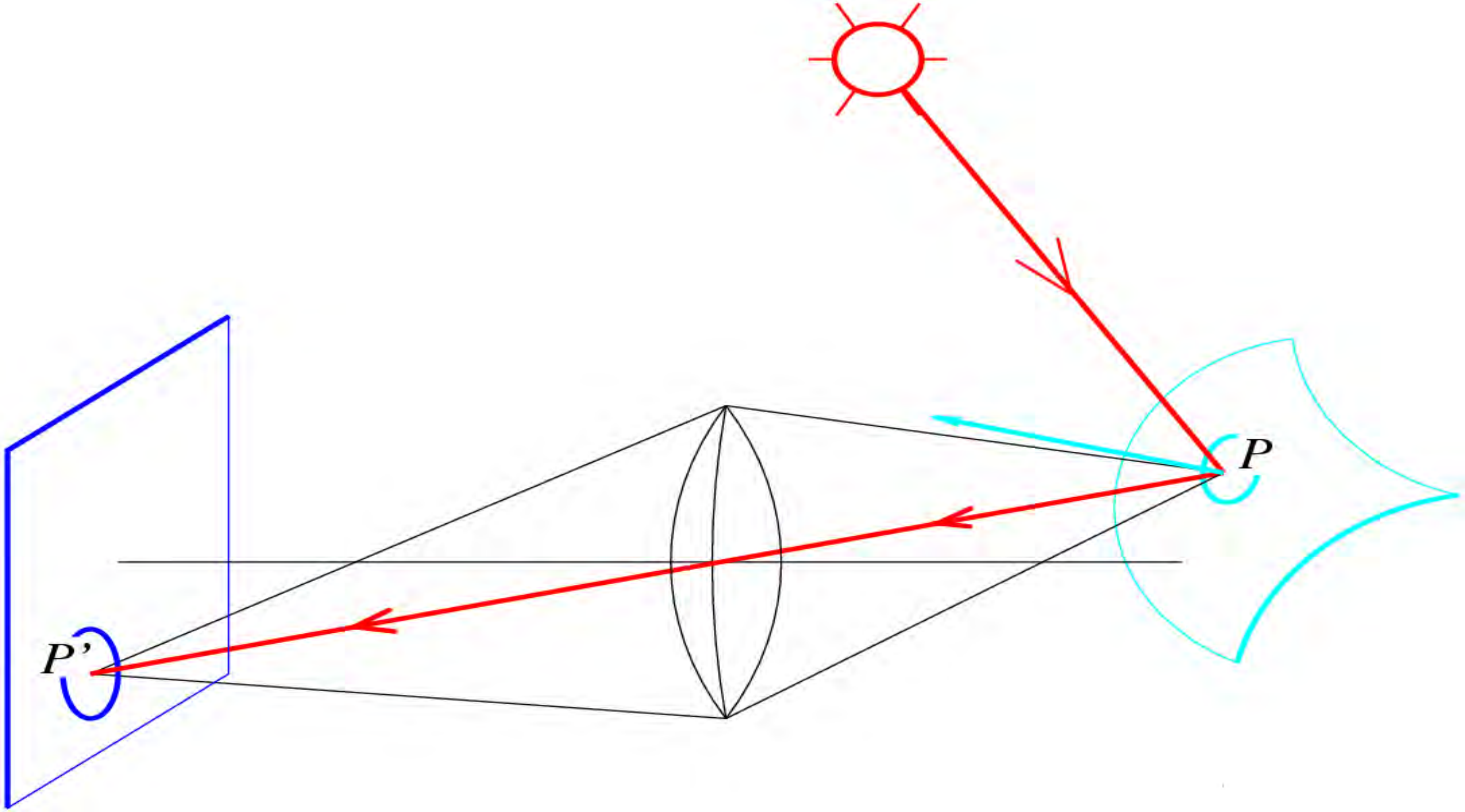
- Shading gives information about surface normals.
- Recovering the 3D surfaces amounts to solving a differential equation.

→Boundary conditions are required to do so.

Boundary Conditions

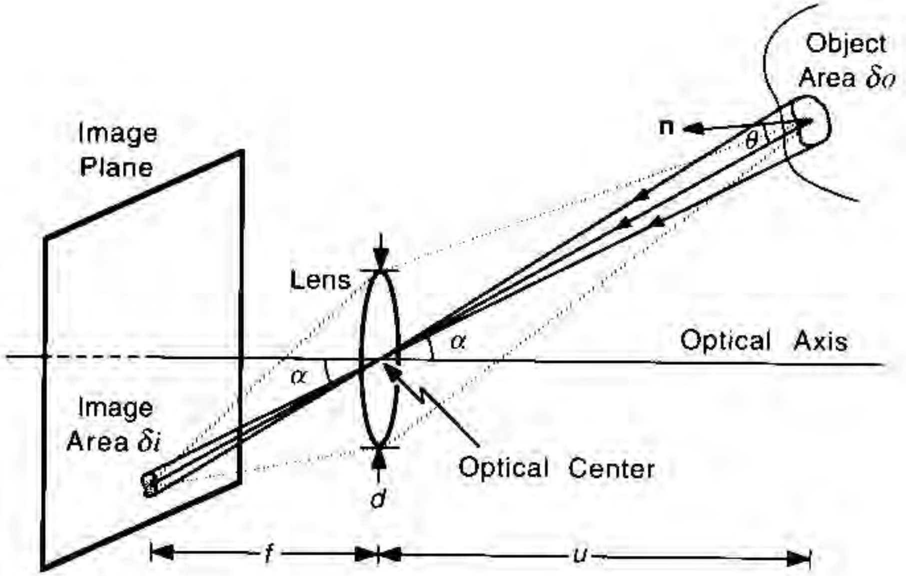
—> The carefully designed contour gives us an erroneous perception!

Reminder: Image Formation



- The light source illuminates a 3D surface.
- The 3D surface reemits some the light.
- It goes through a lens and forms an image on the image plane.

Reminder: Fundamental Radiometric Equation



Scene Radiance (Rad): Amount of light radiation emitted from a surface point (Watt / m2 / Steradian)

Image Irradiance (Irr): Amount of light incident at the image of the surface point. (Watt / m2)

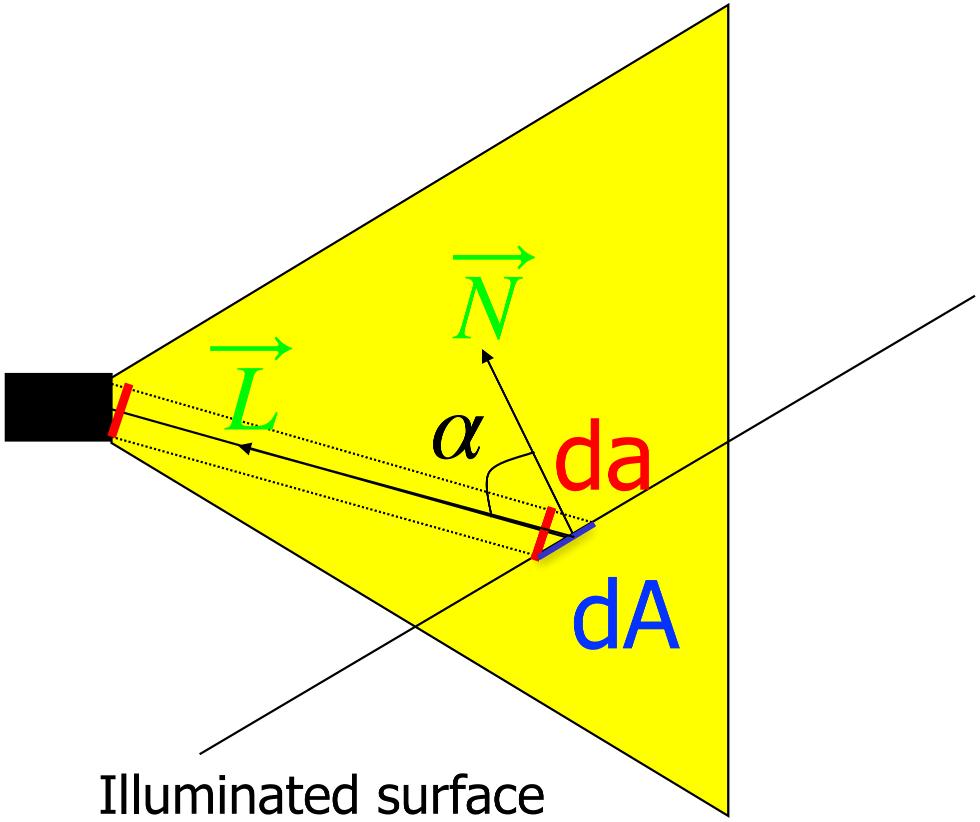
$$\text{Irr} = \frac{\pi}{4}\left(\frac{d}{f}\right)^2 \cos^4(\alpha)\,\text{Rad}\ ,$$

Image intensity

$$\Rightarrow I \propto \text{Rad}\ ,$$

when the camera is photometrically calibrated.

Lambertian Shading Model



- The amount of light radiation **P** in the cylinder of section da in direction $\vec{L}$ is spread over the surface area dA.

$$da = \cos(\alpha) dA$$

$$\cos(\alpha) = \vec{L} \cdot \vec{N}$$

- Some light is absorbed ($0 < \text{albedo} < 1$).

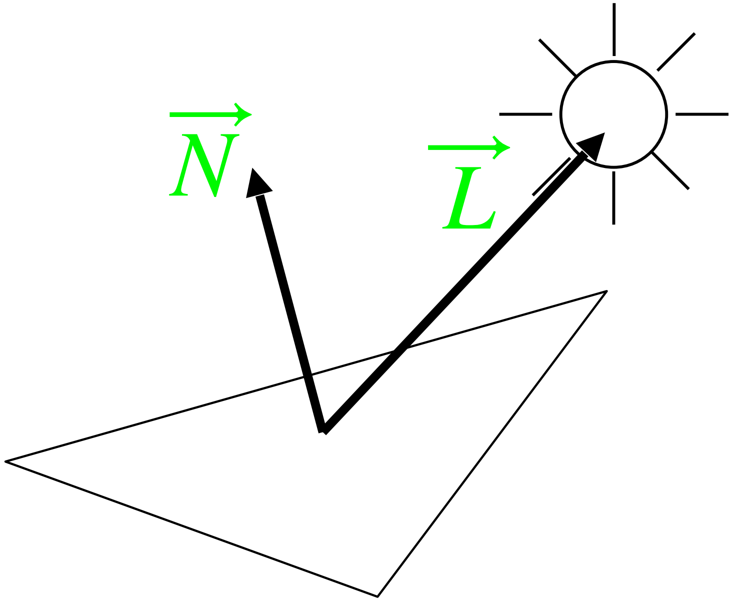
Radiance: $\text{Rad} \propto \text{albedo } \mathbf{P}/dA$

$$\mathbf{P} \propto da$$

$$\Rightarrow \mathbf{P} \propto albedo \frac{da}{dA}$$

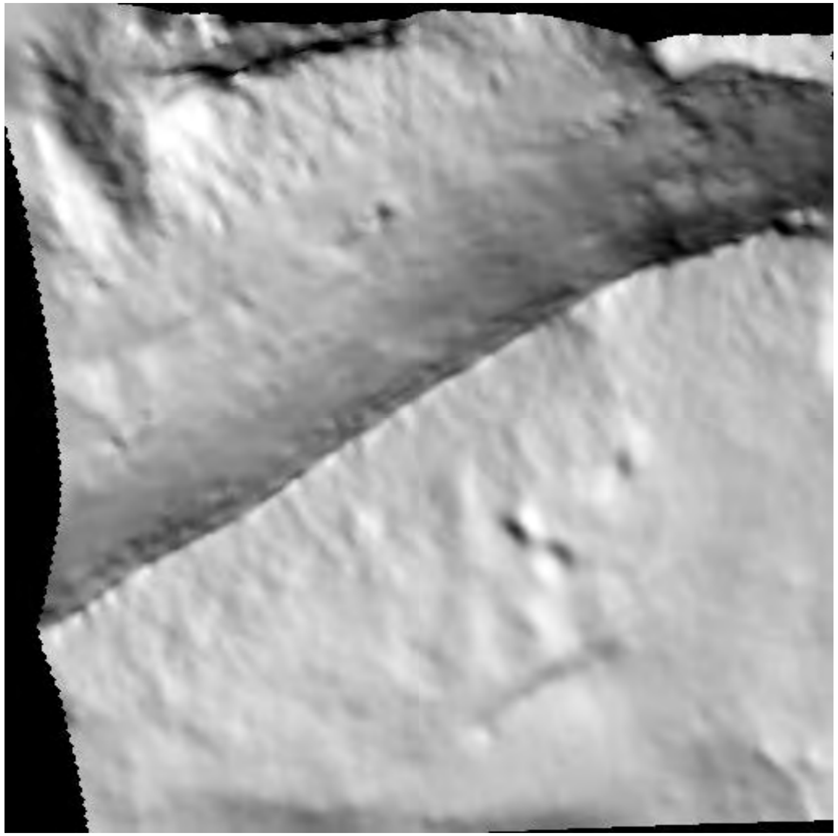
$$\boxed{\mathbf{P} \propto albedo \ \vec{L} \cdot \vec{N}}$$

Ideal Lambertian Surface



$$I = \max(0, \text{albedo}\ \overrightarrow{L} \cdot \overrightarrow{N})$$

No negative light!

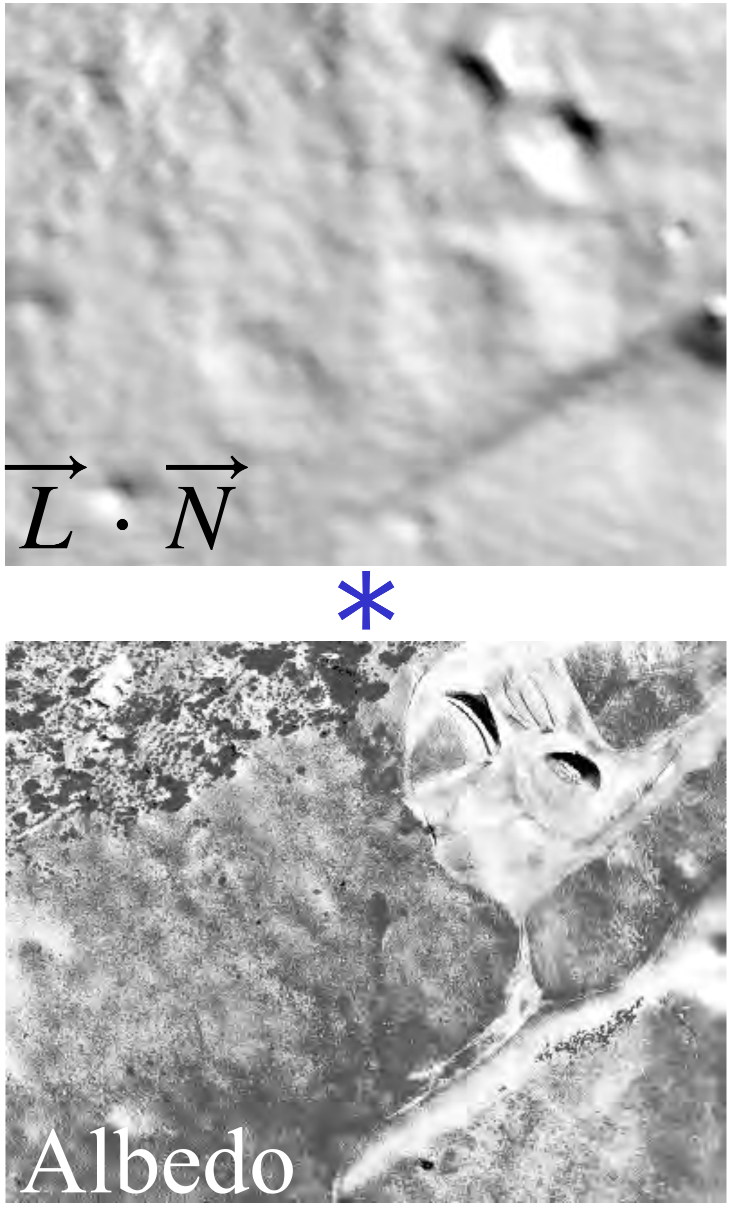


Perfectly matte surface: The radiance depends only on angle of incidence and not on viewing direction. This is known as **diffuse** reflection.

Estimated Albedo

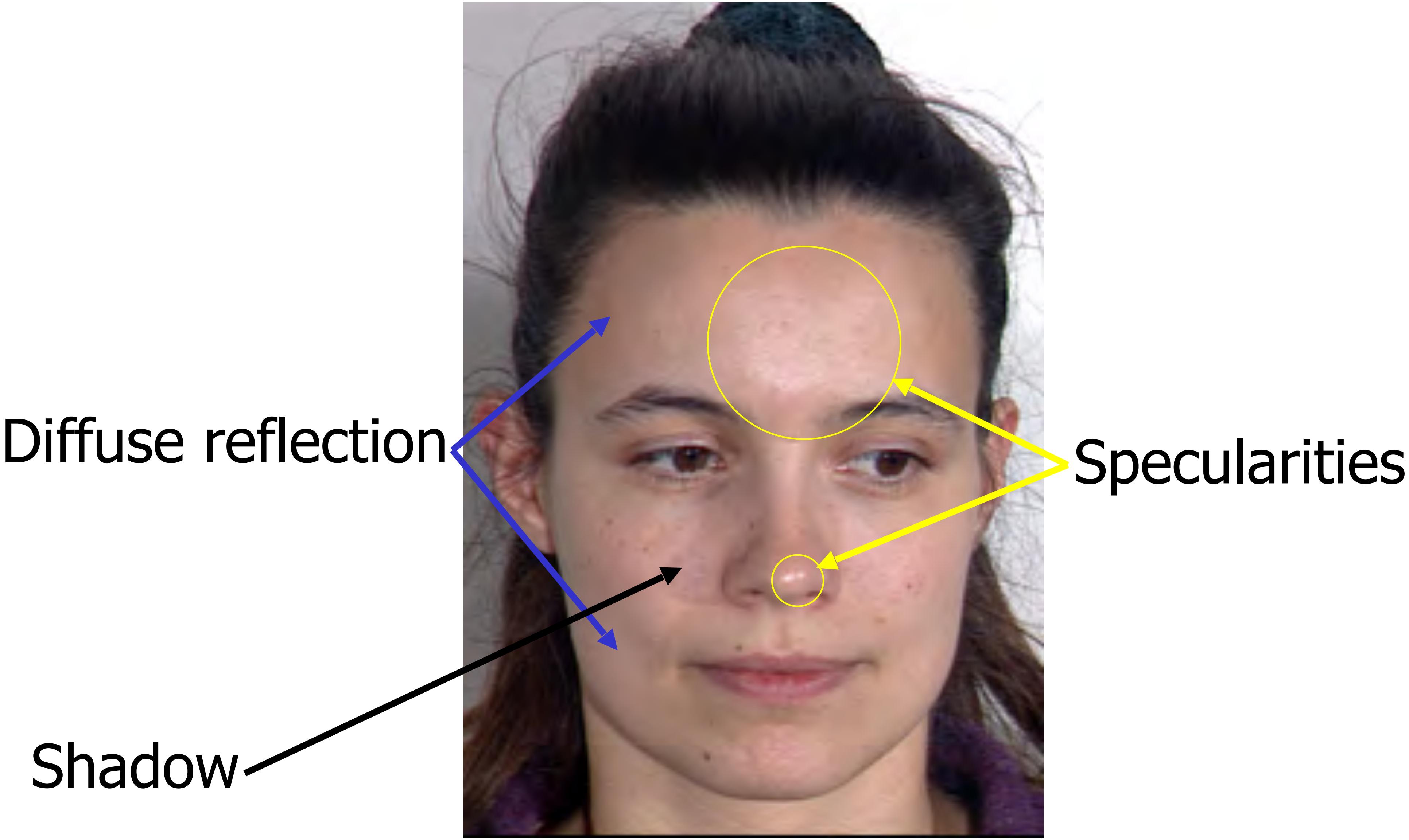
Original image.

=



—> The “albedo” image looks much flatter than the original one.

Diffuse vs Specular

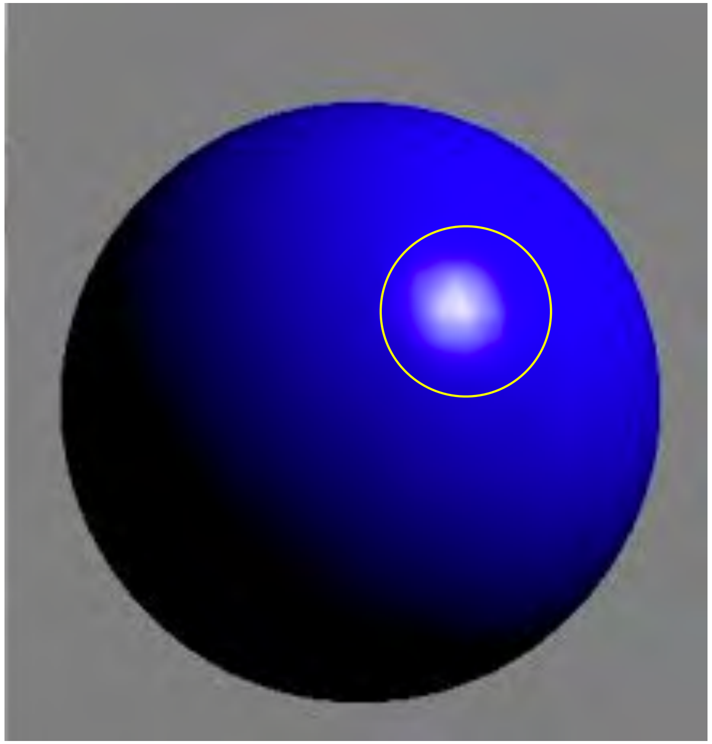

Diffuse reflection

Specularities

Shadow

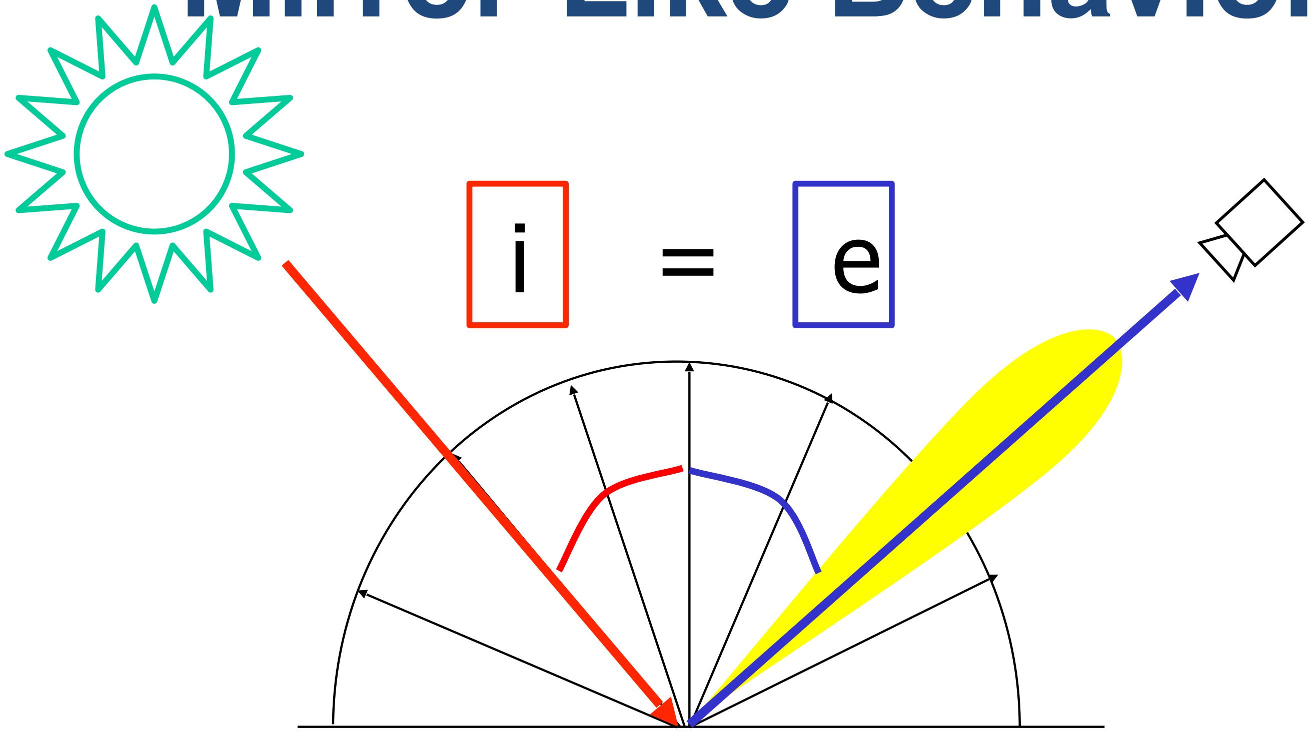
Real-world surfaces are not really Lambertian!

Specularities



- At specular points Lambertian assumptions are violated.
- The surface behaves like a mirror.
- This is known as specular reflection.
- Most surfaces are a combination of diffuse and specular reflectors.

Mirror-Like Behavior



- Specularities occur when the two directions are symmetric with respect to the normal.
- If the light source direction is known, they can be used to infer the normals.

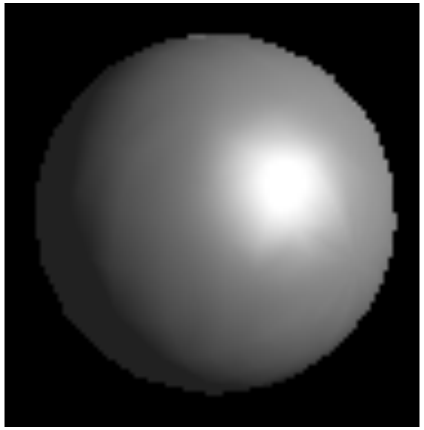
Radiance under Indirect Lighting

- The light source is not visible. Yet there still is light.
- The light enters through the windows and bounces of the walls.

Visualizing Secondary Illumination

Reflections produce indirect lighting.

Unique light source assumption doe not allow correct albedo recovery.

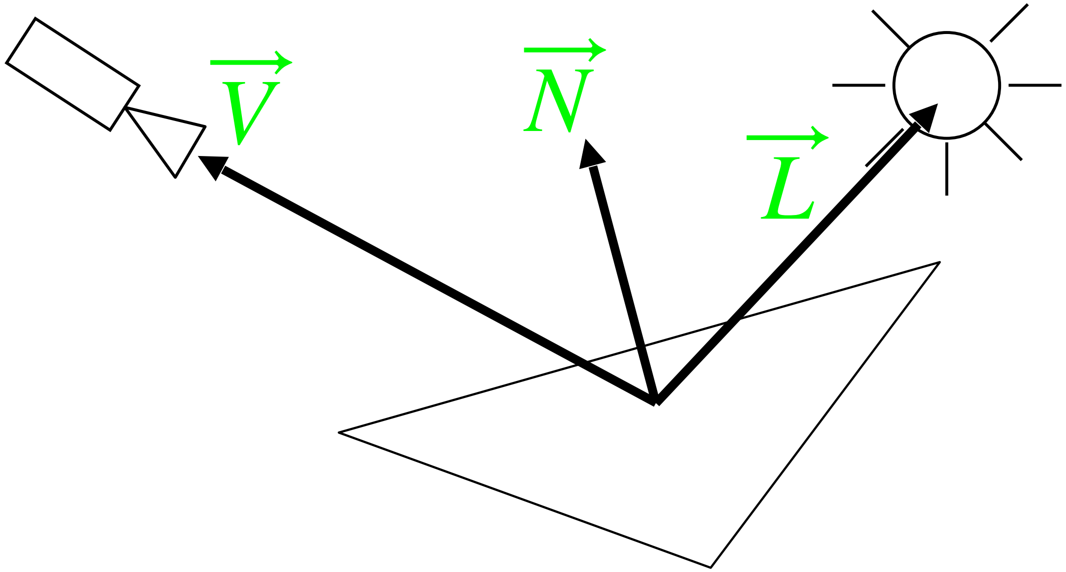


Lambertian shading

Lambertian albedoes

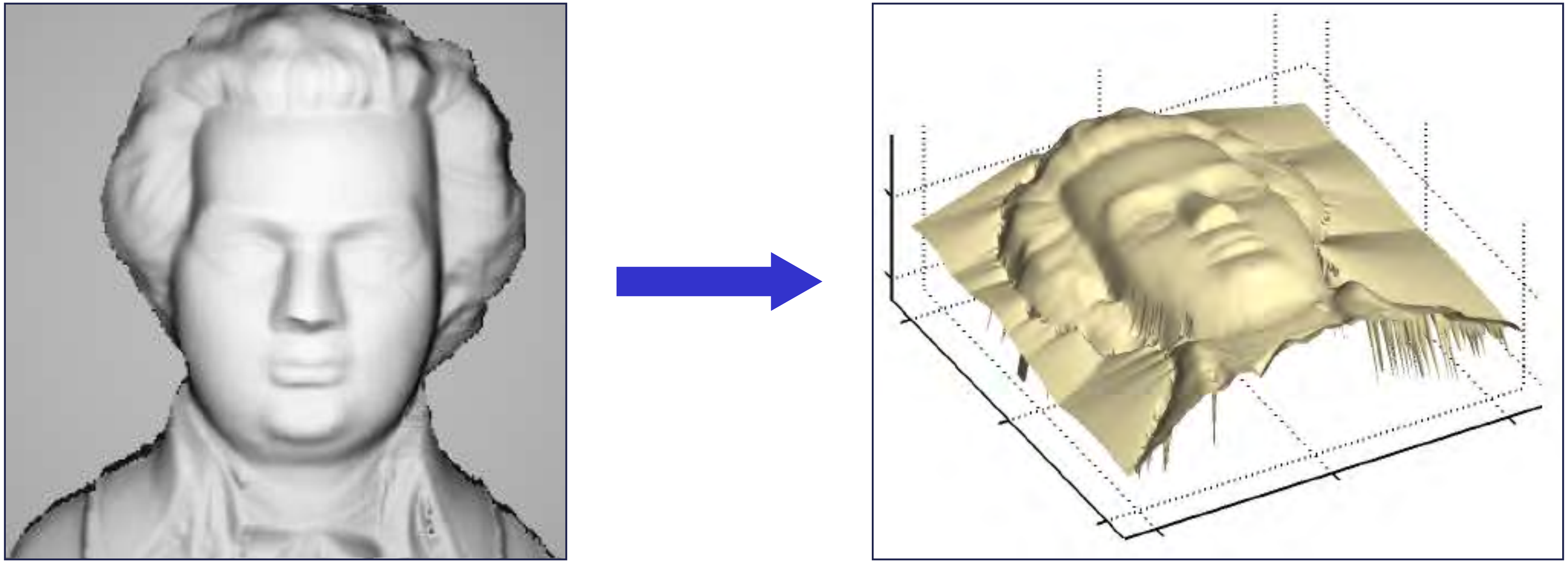
This is not right!

Simplifying Assumptions



- Accounting for secondary illumination in the computer vision context remains an open research problem.
- We will mostly ignore it in this class and make the following assumptions.
  - The illumination sources are distant from the imaged surfaces.
  - Secondary illumination is not significant.
  - There are no cast shadows.

Ideal Synthetic Case



Goal:

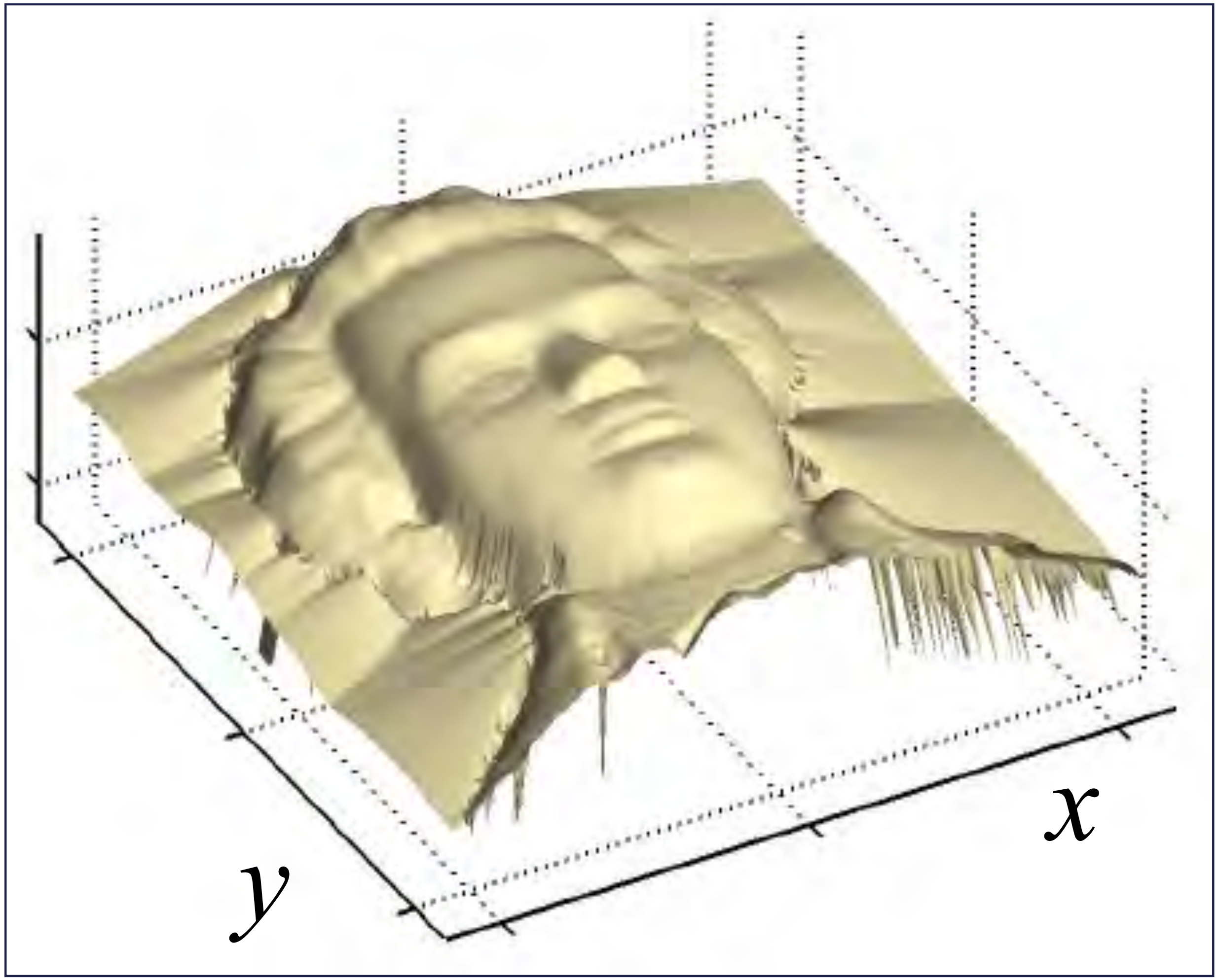
- Recover the 3D shape of the head from the 2D image.

Questions:

1. Given the surface normals, can we recover the surface?
2. Can we recover the normals?

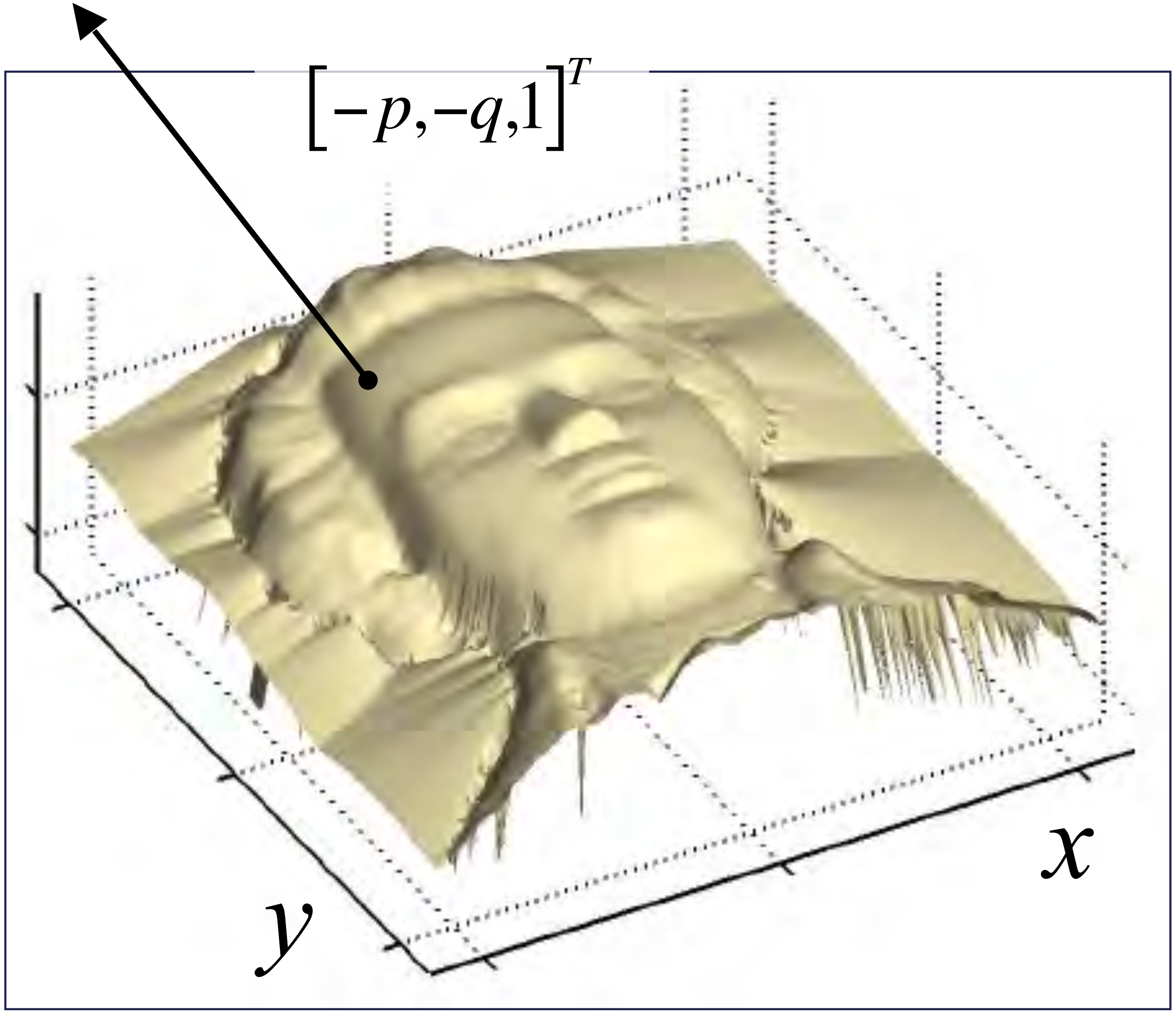
Monge Surface

- A Monge surface is defined by z = f(x,y).
- Not all surfaces can be represented this way.



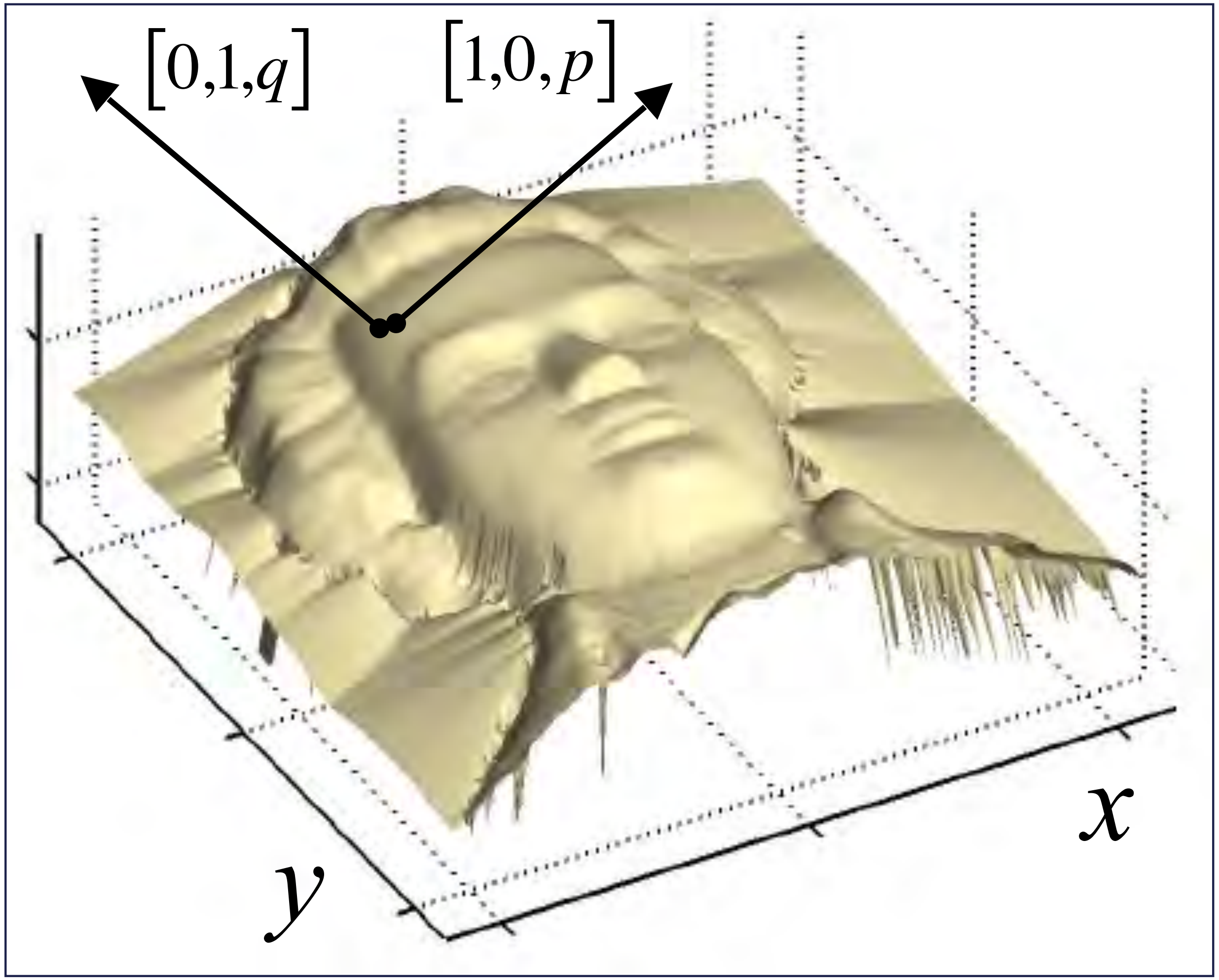
Surface Normals

$$
\begin{aligned}
z &= f(x, y) \\
p &= \frac{\delta z}{\delta x} \\
q &= \frac{\delta z}{\delta y}
\end{aligned}
$$



Tangent Vectors

$$
\begin{aligned}
z &= f(x, y) \\
p &= \frac{\delta z}{\delta x} \\
q &= \frac{\delta z}{\delta y}
\end{aligned}
$$



Shape from Normals

Normals

3D shape

Elevation and normal:

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

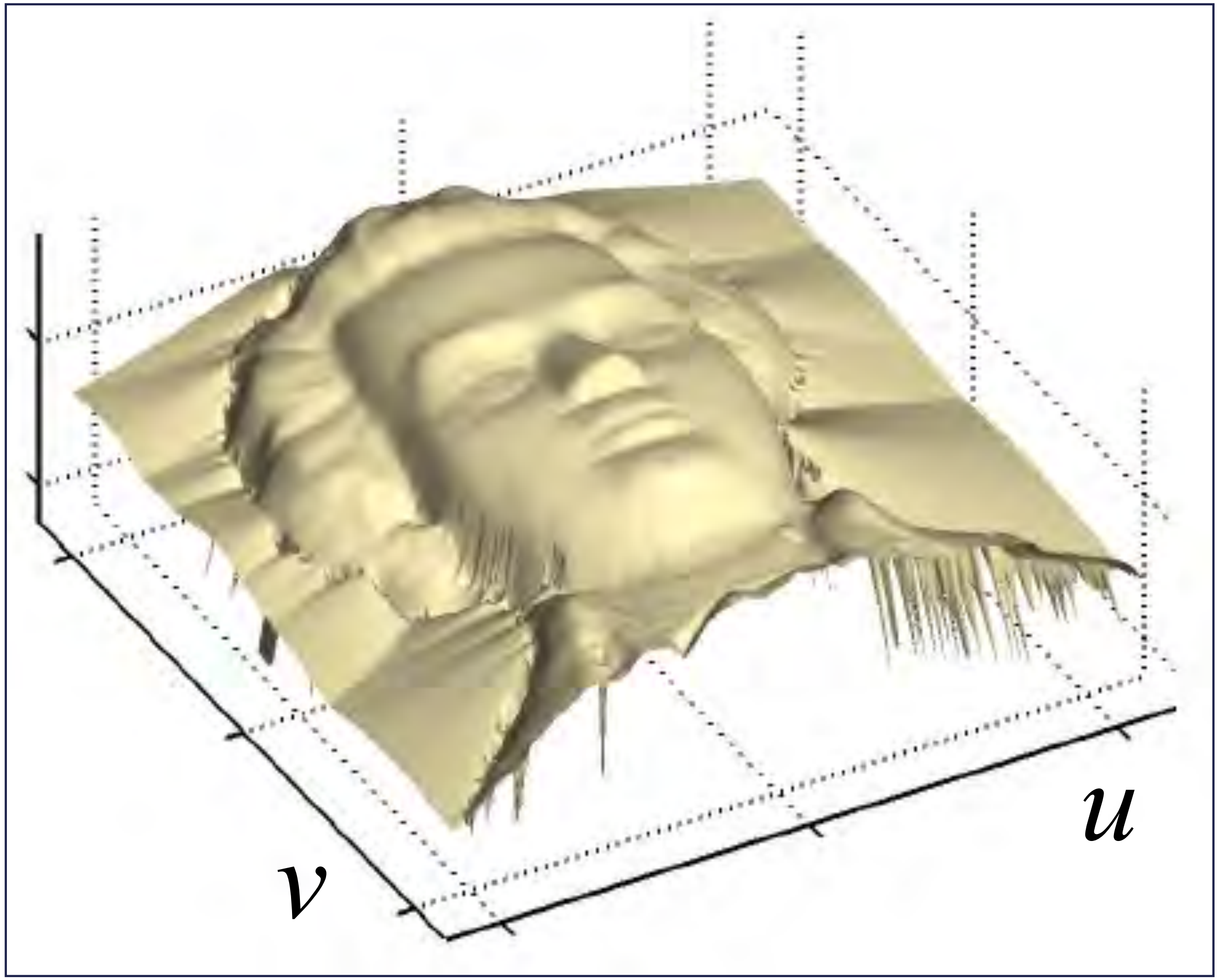
$$\mathbf{N} = \frac{1}{\sqrt{1 + p^2 + q^2}} \begin{bmatrix} -p \\ -q \\ 1 \end{bmatrix}$$

Orthographic projection:

$$u = sx$$
$$v = sy$$

Re-Parametrization

Perform a change of variables z = f(u,v) where u and v are image coordinates.



Shape From Normals (1)

Since u = sx and v = sy, the normal vector **N** is

$$\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \frac{1}{\sqrt{1 + \frac{\delta z}{\delta x}^2 + \frac{\delta z}{\delta y}^2}} \begin{bmatrix} -\frac{\delta z}{\delta x} \\ -\frac{\delta z}{\delta y} \\ 1 \end{bmatrix} ,$$

$$\Rightarrow \frac{\delta z}{\delta x} = -\frac{n_x}{n_z} \text{ and } \frac{\delta z}{\delta y} = -\frac{n_y}{n_z} ,$$

$$\Rightarrow \frac{\delta z}{\delta u} = -\frac{1}{s}\frac{n_x}{n_z} \text{ and } \frac{\delta z}{\delta v} = -\frac{1}{s}\frac{n_y}{n_z} ,$$

$$\Rightarrow \frac{\delta \bar{z}}{\delta u} = -\frac{n_x}{n_z} = n_1 \text{ and } \frac{\delta \bar{z}}{\delta v} = -\frac{n_y}{n_z} = n_2 ,$$

where $\bar{z} = sz$ is the scaled distance.

Shape From Normals (2)

- Let us assume we are given the normal at each pixel (u,v).
- Let $n_1(u,v) = -n_x(u,v)/n_z(u,v)$ and $n_2(u,v) = -n_y(u,v)/n_z(u,v)$.
- From the previous slide, we have

$$\forall u, v \quad \begin{cases} n_1(u,v) = \dfrac{\delta \bar{z}}{\delta u} \approx \bar{z}(u+1, v) - \bar{z}(u,v) \\ n_2(u,v) = \dfrac{\delta \bar{z}}{\delta v} \approx \bar{z}(u, v+1) - \bar{z}(u,v) \end{cases}$$

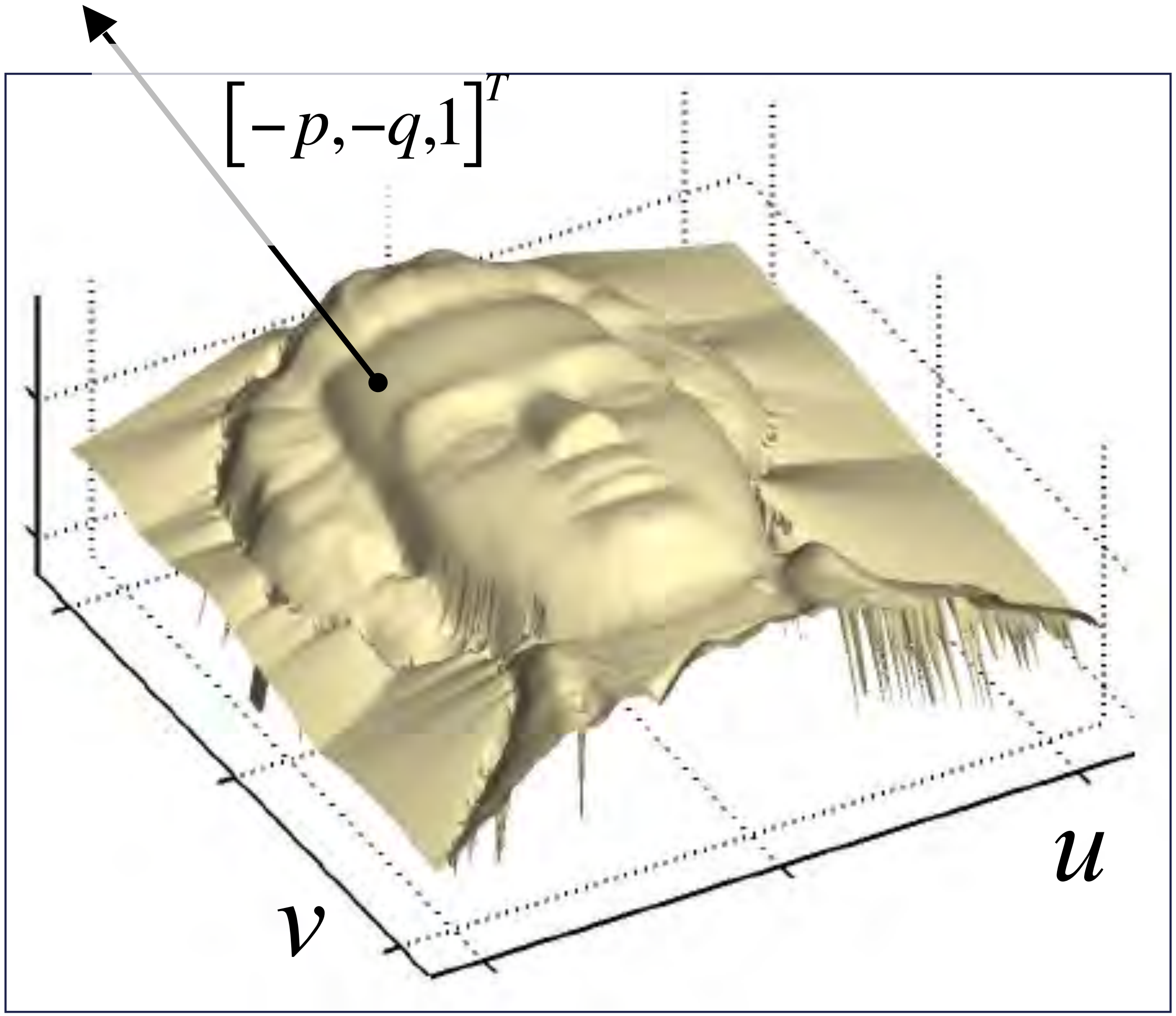
- We therefore have roughly twice as many equations as we have unknowns, the scaled distances $\bar{z}(u,v)$.
- This can be solved in the least squares sense.

—> Given the normals at every pixel we can recover the distances up to a scale factor.

But how do we get the normals?

Back to Estimating the Normals

$$
\begin{aligned}
z &= f(u, v) \\
p &= \frac{\delta z}{\delta u} \\
q &= \frac{\delta z}{\delta v}
\end{aligned}
$$



What does the image tell us about them?

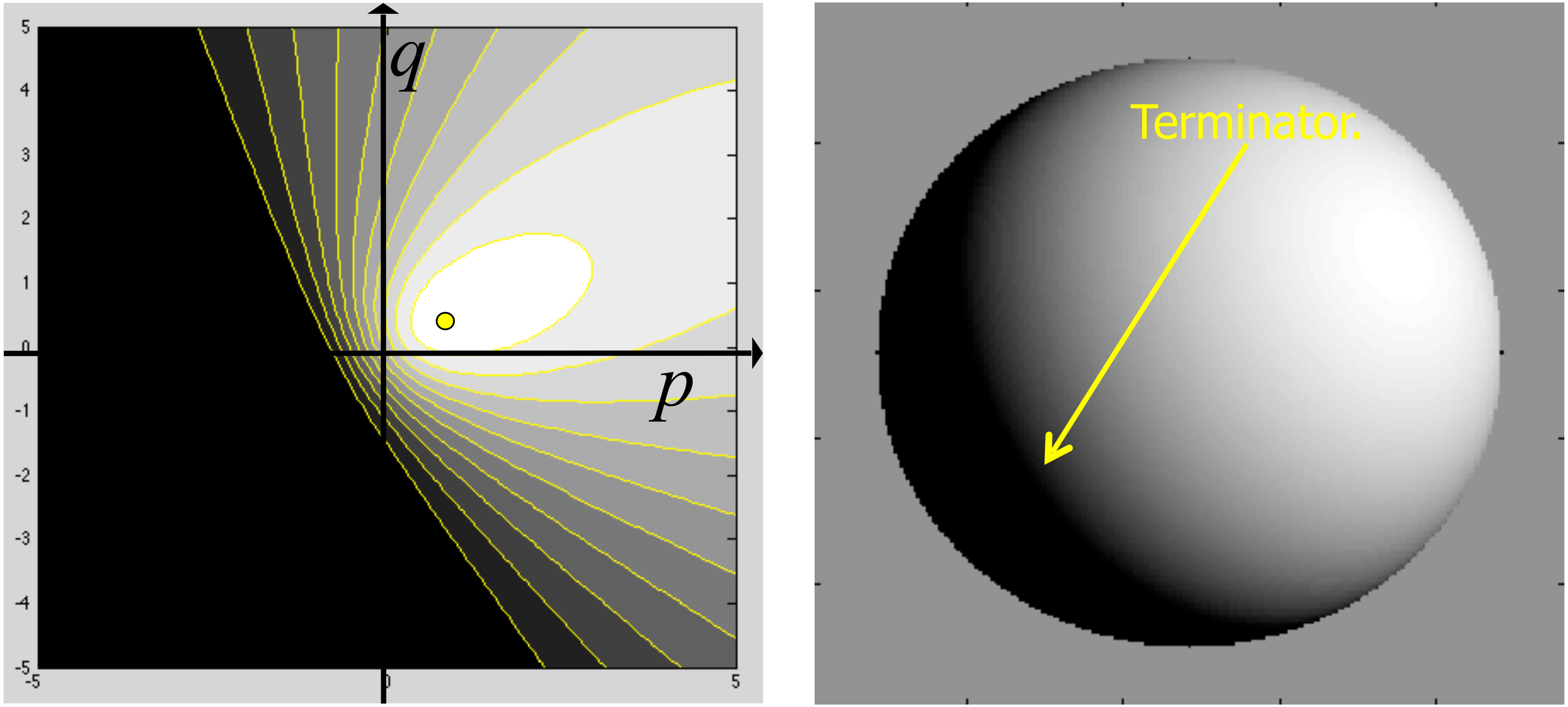
Reflectance Map

In the Lambertian case and for a constant albedo:

$$
\begin{aligned}
I(u,v) &\propto \mathbf{L} \cdot \overrightarrow{\mathbf{N}} \\
&\propto \mathbf{L} \cdot \left[-p(u,v), -q(u,v), 1\right]^T \\
&\propto Ref(p(u,v), q(u,v))
\end{aligned}
$$

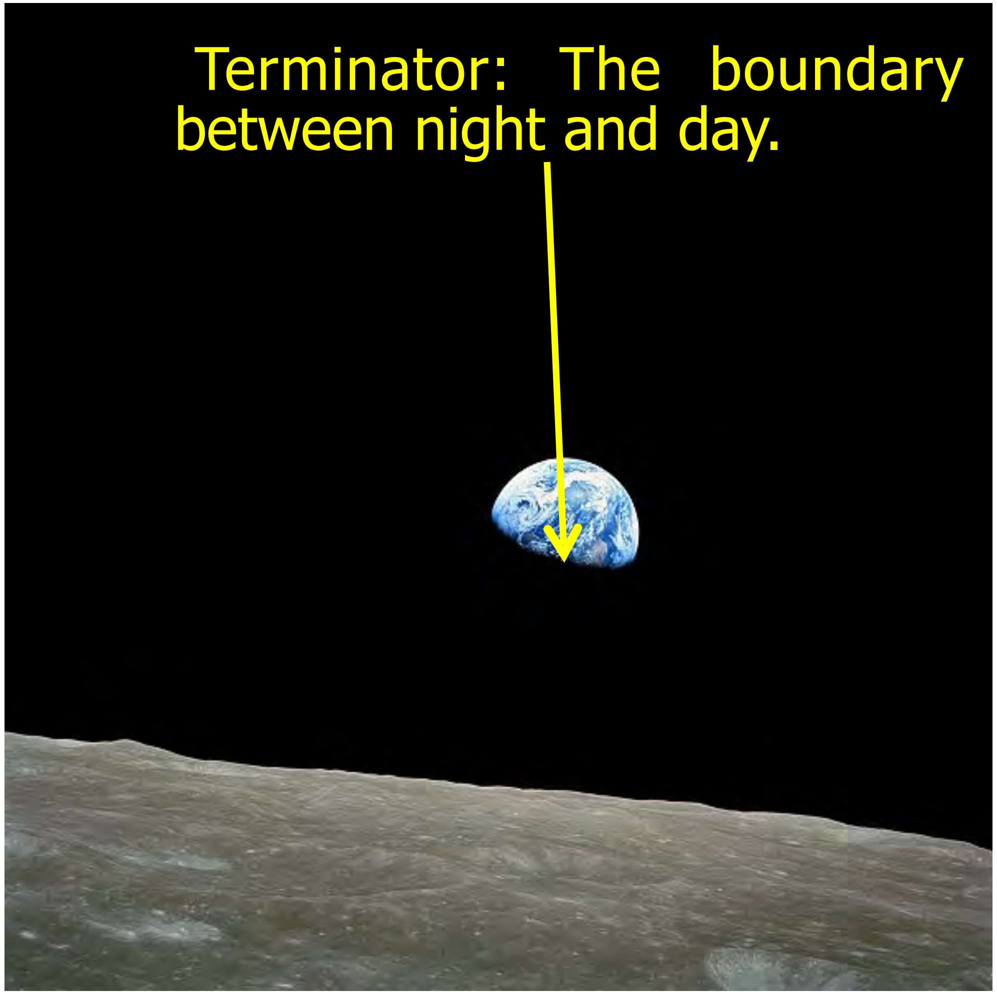
- The function Ref is known as the reflectance map.
- For non-Lambertian surfaces it can be more complex.

Lambertian Reflectance Map



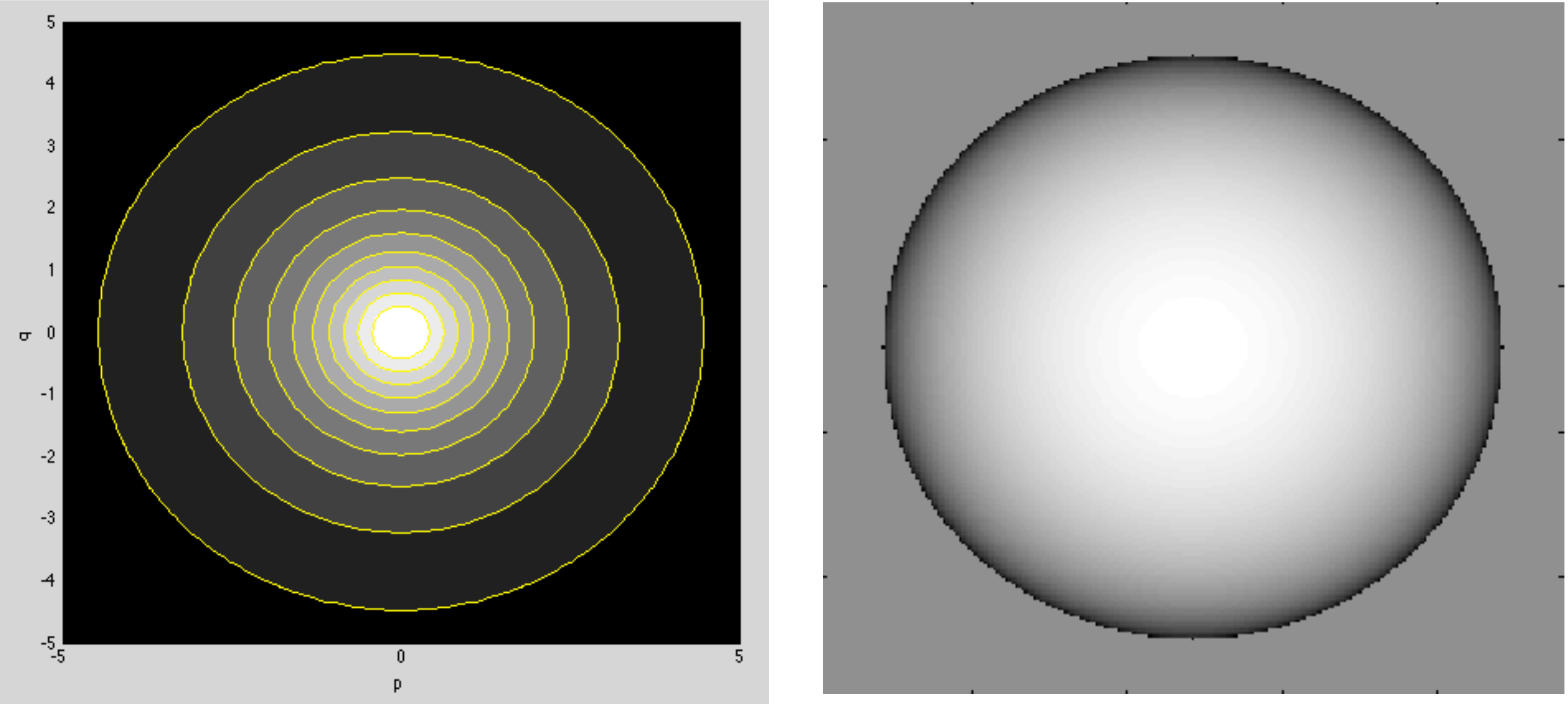
Reflectance map and shaded surface for Lambertian surface illuminated in the direction [-1 -0.5 -1].

Earth Seem from the Moon


Terminator: The boundary between night and day.

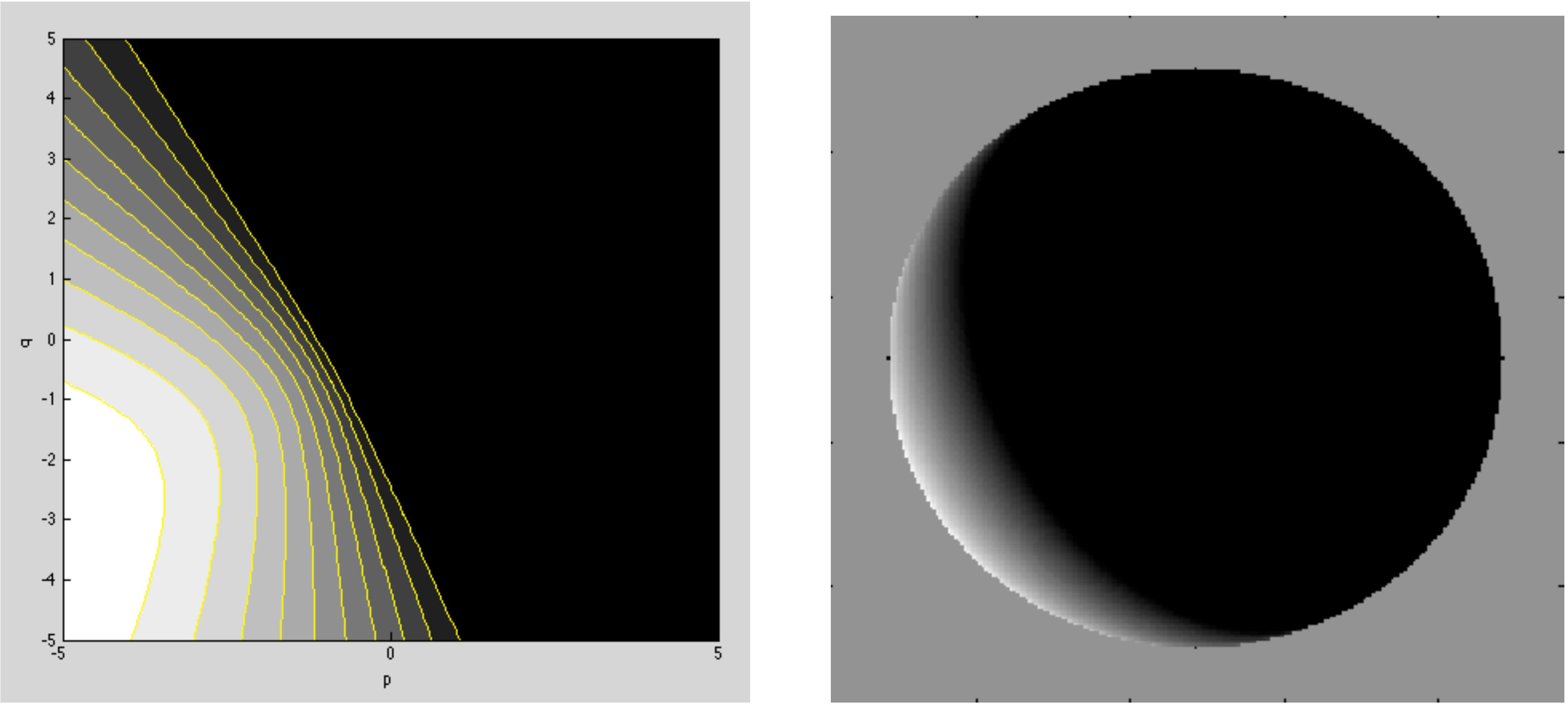
Apollo 8, 1968.

Lambertian Reflectance Map



Reflectance map and shaded surface for Lambertian surface illuminated in the direction [0 0 -1].

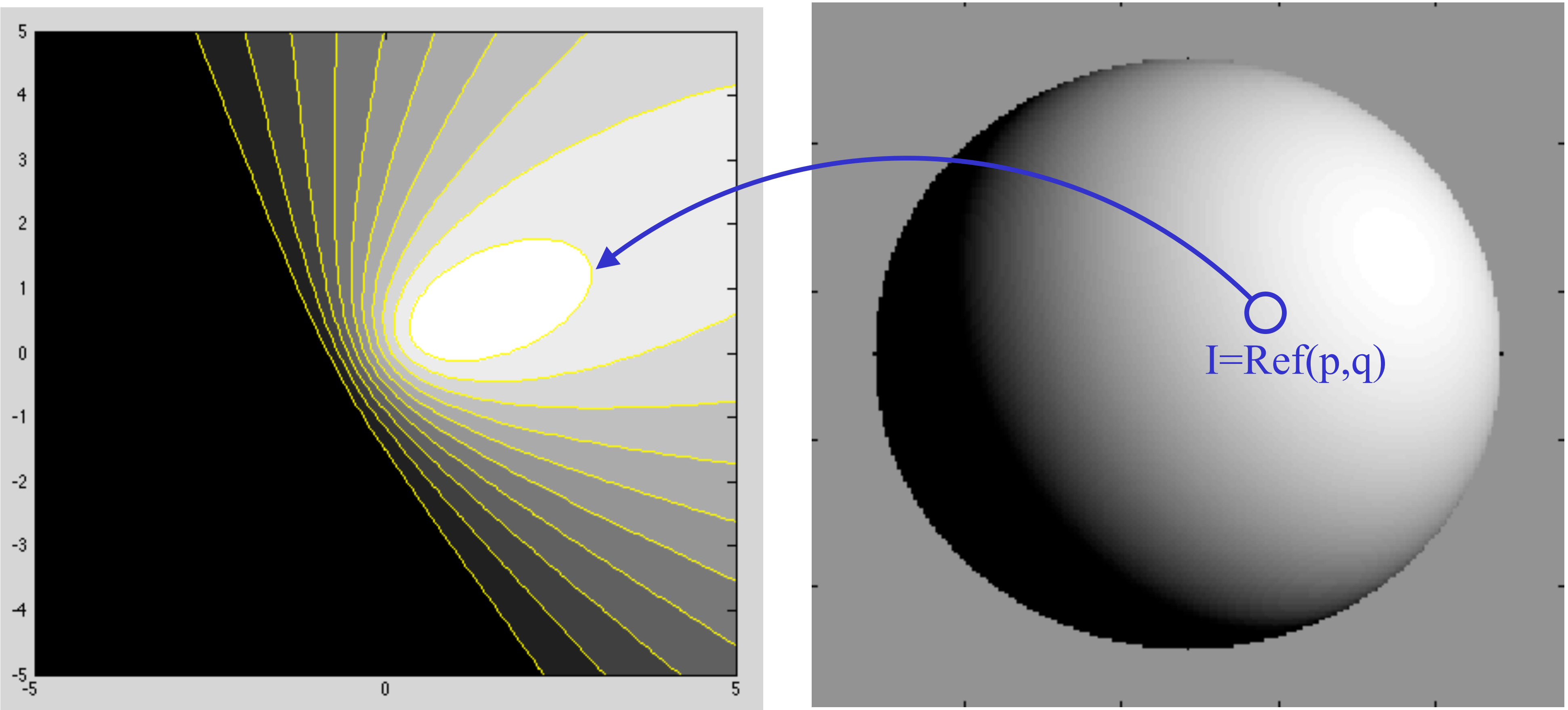
Lambertian Reflectance Map



Reflectance map and shaded surface for Lambertian surface illuminated in the direction [1 0.5 -1].

Inverse Problem

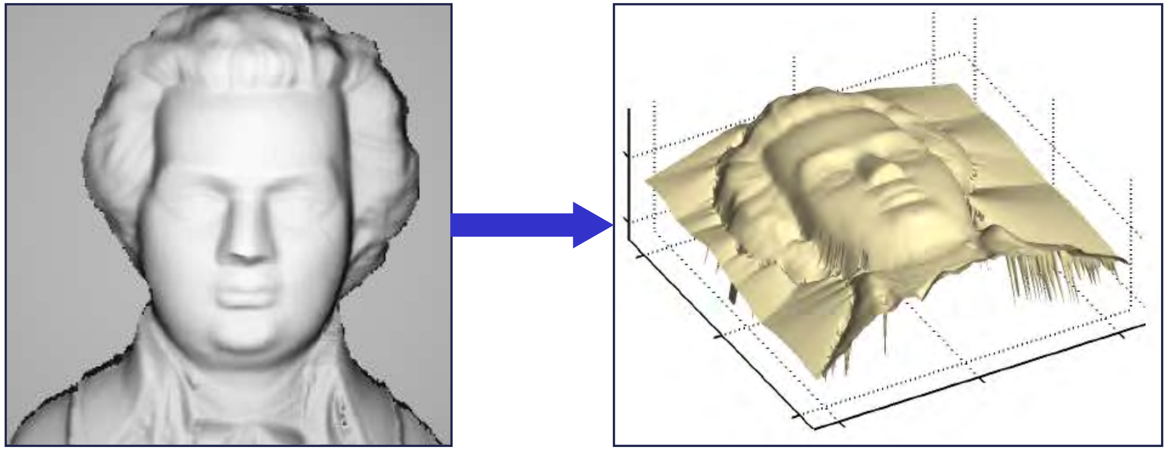
Can we determine (p,q) uniquely for each image point independently?



No because many p and q yield the same Ref(p,q).

—> Global optimization required.

Variational Method (1)



Minimize:

$$\int\int\left([I(u,v)-Ref(p,q)]^2+\lambda\left[\left(\frac{\delta p}{\delta u}\right)^2+\left(\frac{\delta p}{\delta v}\right)^2+\left(\frac{\delta q}{\delta u}\right)^2+\left(\frac{\delta q}{\delta v}\right)^2\right]+\mu\left[\frac{\delta p}{\delta v}-\frac{\delta q}{\delta u}\right]^2\right)dudv$$

Data term

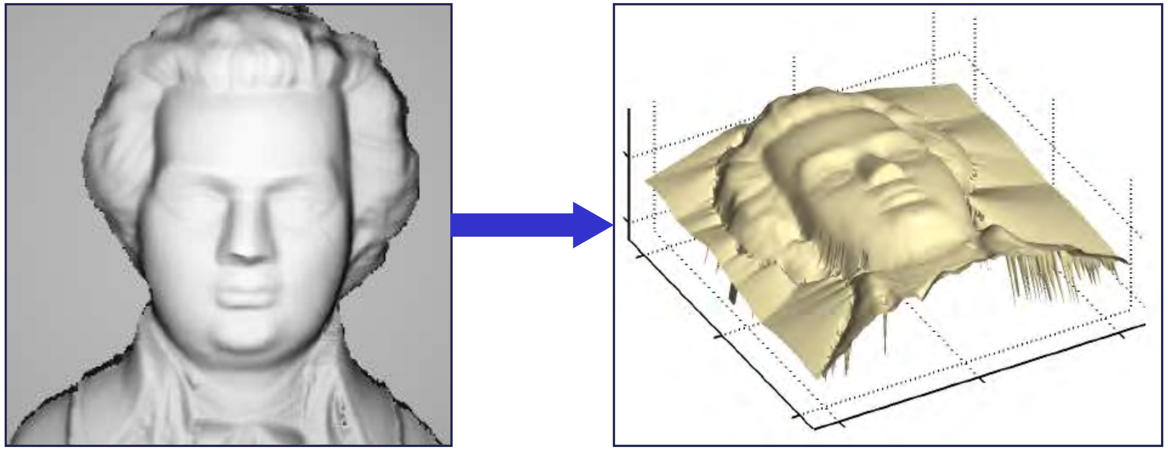
Smoothness term

Integrability term

$$\frac{\frac{\delta z}{\delta u}}{\delta v}=\frac{\frac{\delta z}{\delta v}}{\delta u}=\frac{\delta^2 z}{\delta u\delta v}$$

1. Recover the normals and make them integrable.
2. Integrate to recover the 3D surface.

Variational Method (2)



Minimize:

$$\int\int\left(\left[I(u,v) - Ref(\frac{\delta z}{\delta u},\frac{\delta z}{\delta v})\right]^2 + \lambda\left[\left(\frac{\delta^2 z}{\delta u^2}\right)^2 + \left(\frac{\delta^2 z}{\delta u\delta v}\right)^2 + \left(\frac{\delta^2 z}{\delta v^2}\right)^2\right]\right) dudv$$

Data term

Smoothness term

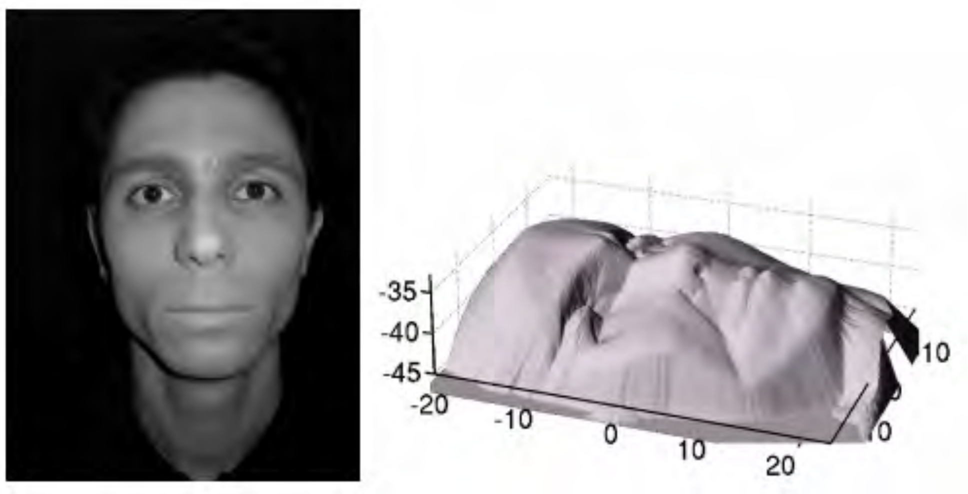
- Recover the surface directly, which means solving a second order differential equation instead of a first order one.
- Both approaches are valid and require boundary conditions.

Moonscape

Moon image

Depth map

Faces from Shading



- Generic Monge surface
- Low resolution images

Prados and Faugeras, CVPR’05.

- Deep face model
- High resolution images

Bagautdinov et al. , CVPR’18.

T-Shirt

Works because the albedo is constant!

Barron & Malik, PAMI'15

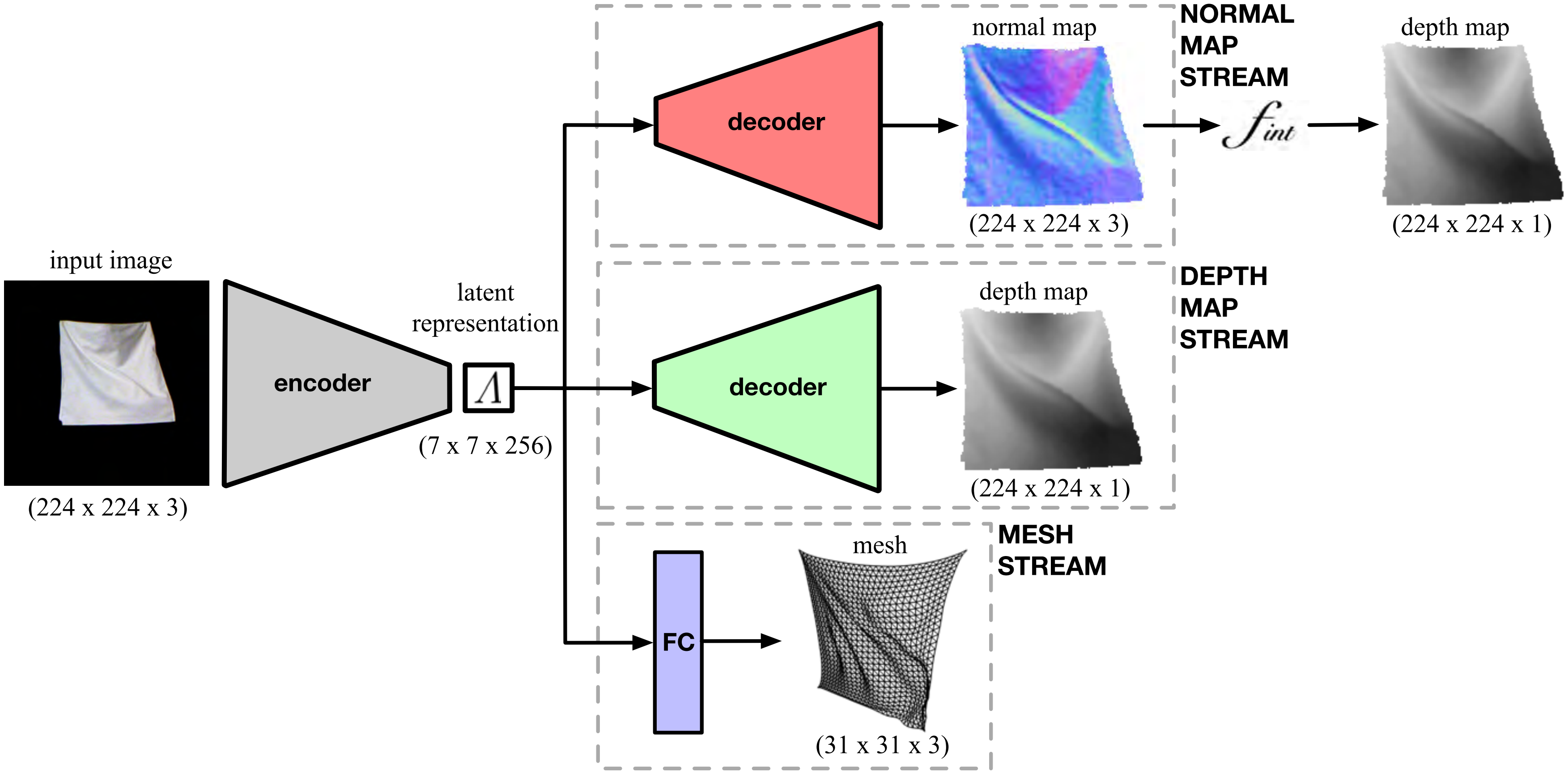
Combining Stereo and Shading

Stereo Only Stereo + Flow

- Shape-from-shading can be used in conjunction with other modalities to refine the shape and provide high-frequency details.
- We will come back to this when we talk about stereo and motion.

Valgaerts et al. SIGGRAPH Asia’12

From Variational to Deep



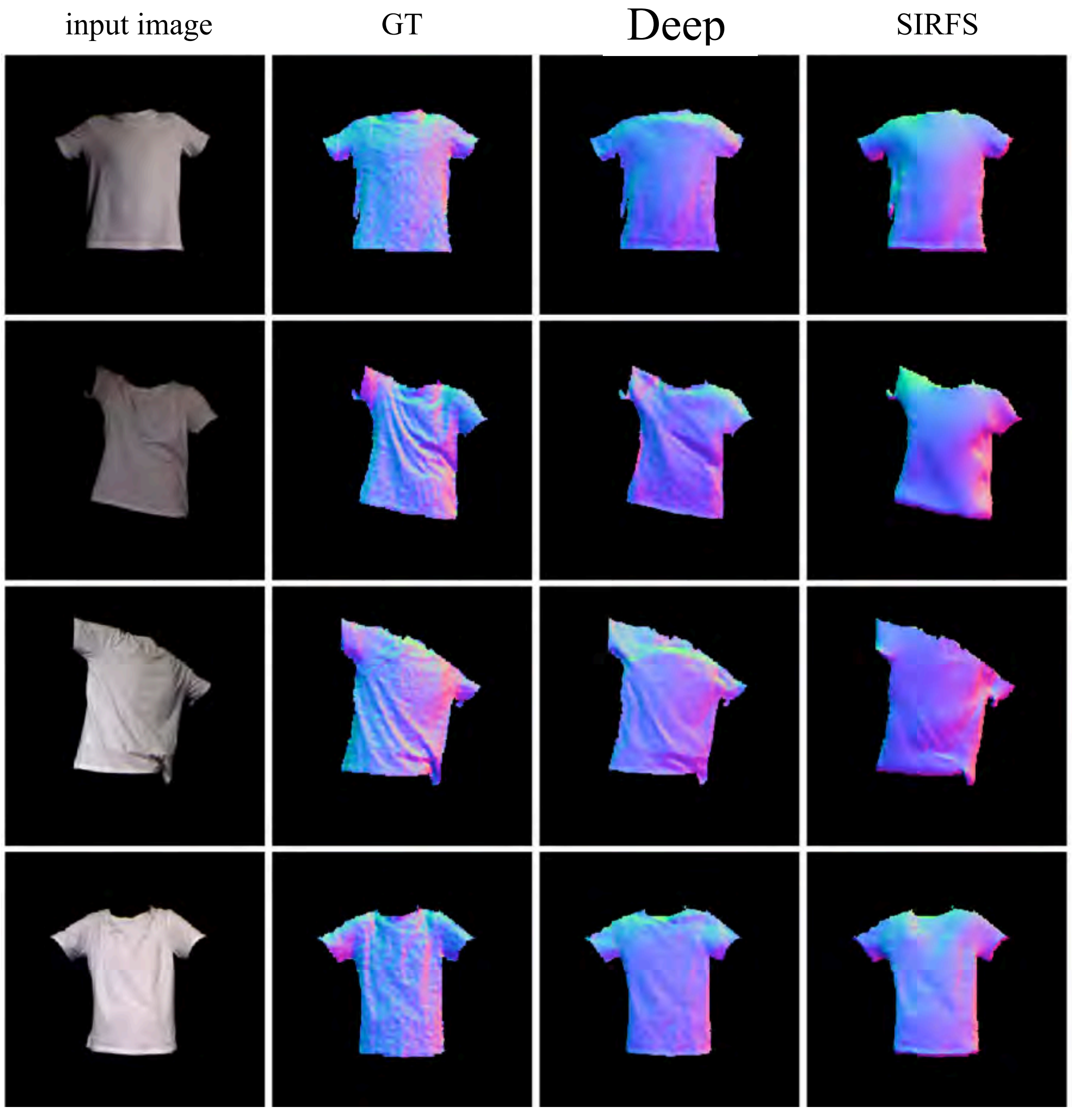
The deep net is trained to:

- produce both a depth map and a map of normals,
- ensure they are consistent with each other.

—> Can be understood as another way to solve the variational problem.

Bednarik et al. , ECCV'18

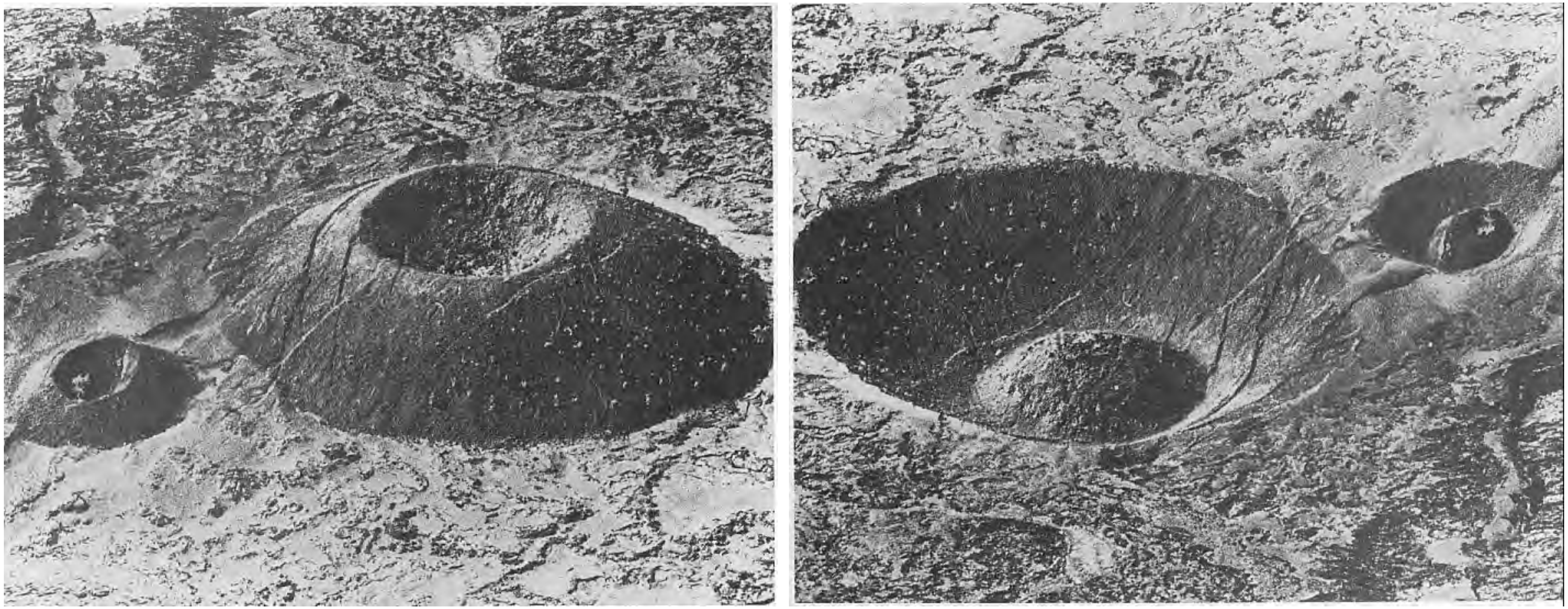
Improved Results



The deep net recovers more details ……

…… but requires training data.

Reminder: Ambiguities



- Back where we started.
- Let us look at them more closely.

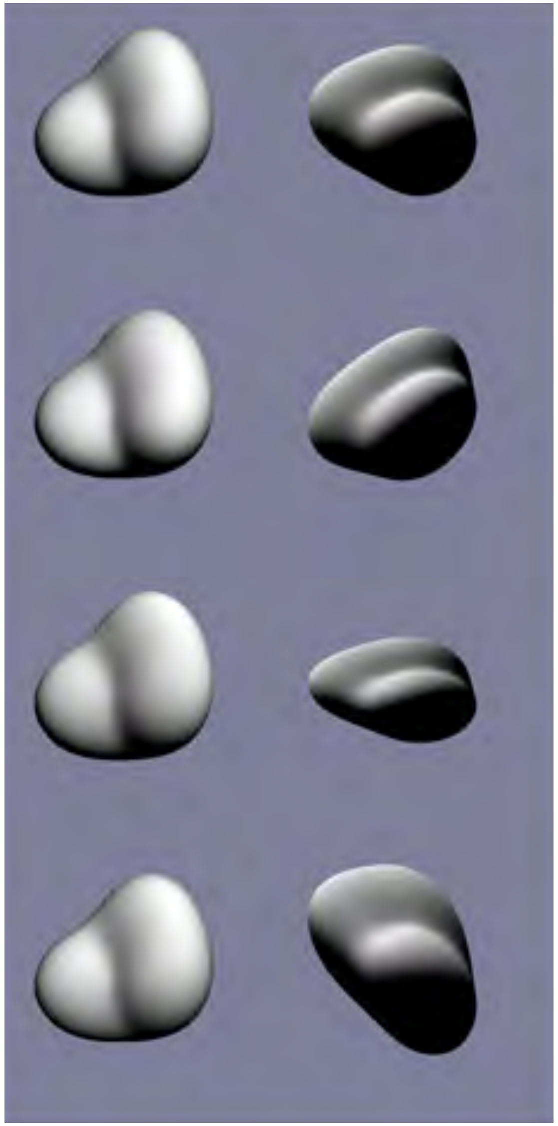
Bas-Relief Ambiguity

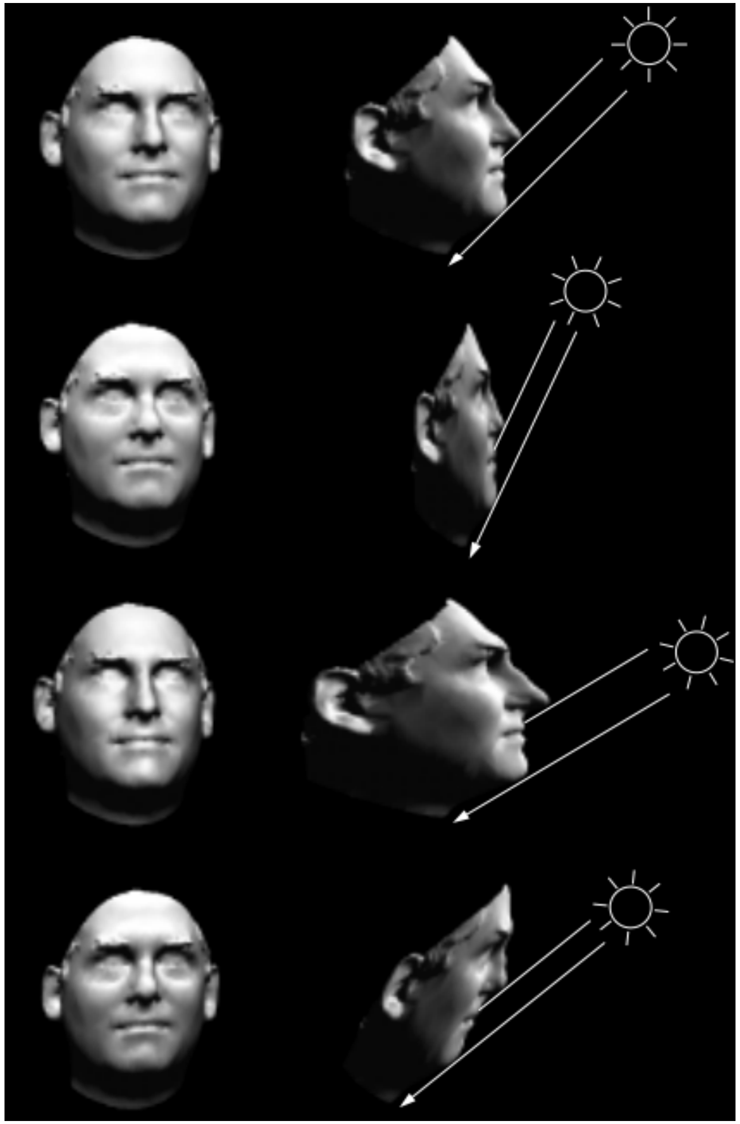
Looks like a normal human head …

… but not when seen from the side.

Why is that?

Bas-Relief Ambiguity

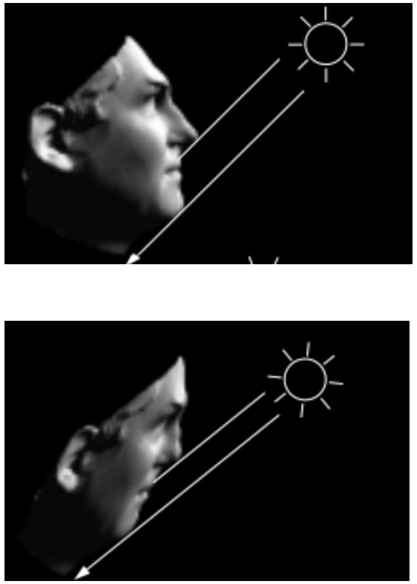




By moving the light source, it is possible to produce the same image for different 3D shapes.

Bas-Relief Ambiguity Explained (1)

$$Ref = \mathbf{N} \cdot \mathbf{L}$$



For any invertible 3x3 linear transformation A:

$$\begin{aligned}(\mathbf{AN}) \cdot (\mathbf{A}^{-T}\mathbf{L}) &= (\mathbf{AN})^T(\mathbf{A}^{-T}\mathbf{L}) \\ &= \mathbf{N}^T\mathbf{A}^T\mathbf{A}^{-T}\mathbf{L} \\ &= \mathbf{N}^T\mathbf{L} \\ &= \mathbf{N} \cdot \mathbf{L}\end{aligned}$$

- In theory, applying **A** to the normals and $\mathbf{A}^{-1}$ to the light source would not change the image.
- However, the normals must remain integrable, which means that not all transformations of the normals are valid.
- In particular, for a Monge surface z = f(u,v), we must have

$$\frac{\frac{\delta z}{\delta u}}{\delta v} = \frac{\frac{\delta z}{\delta v}}{\delta u} = \frac{\delta^2 z}{\delta u \delta v}$$

Bas-Relief Ambiguity Explained (2)

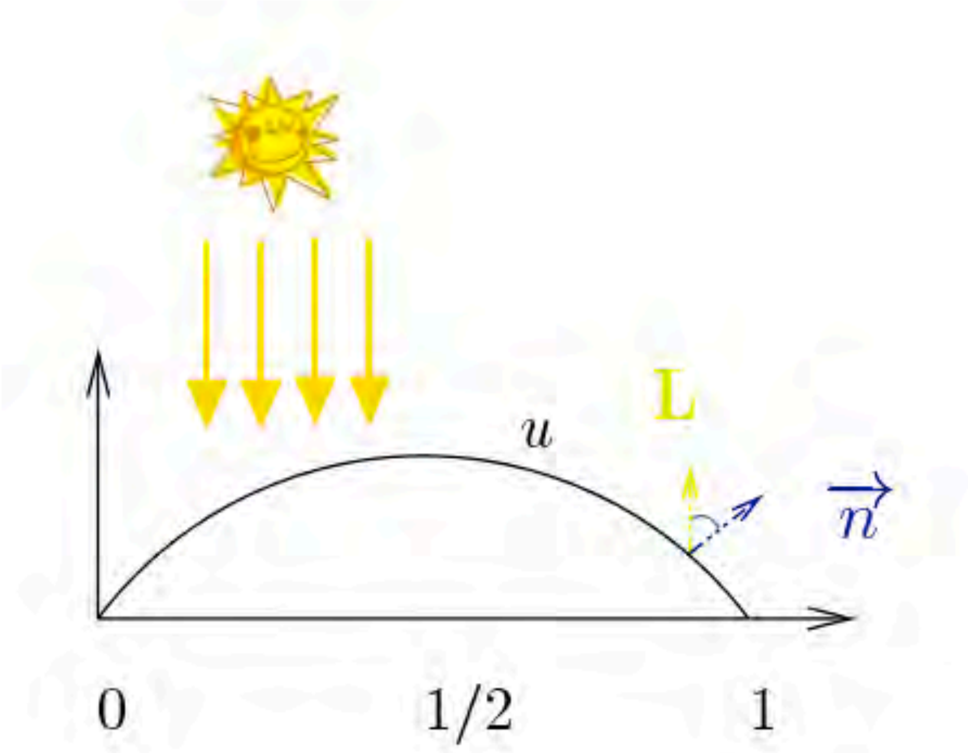
Let us write the integrability constraint in our specific case:

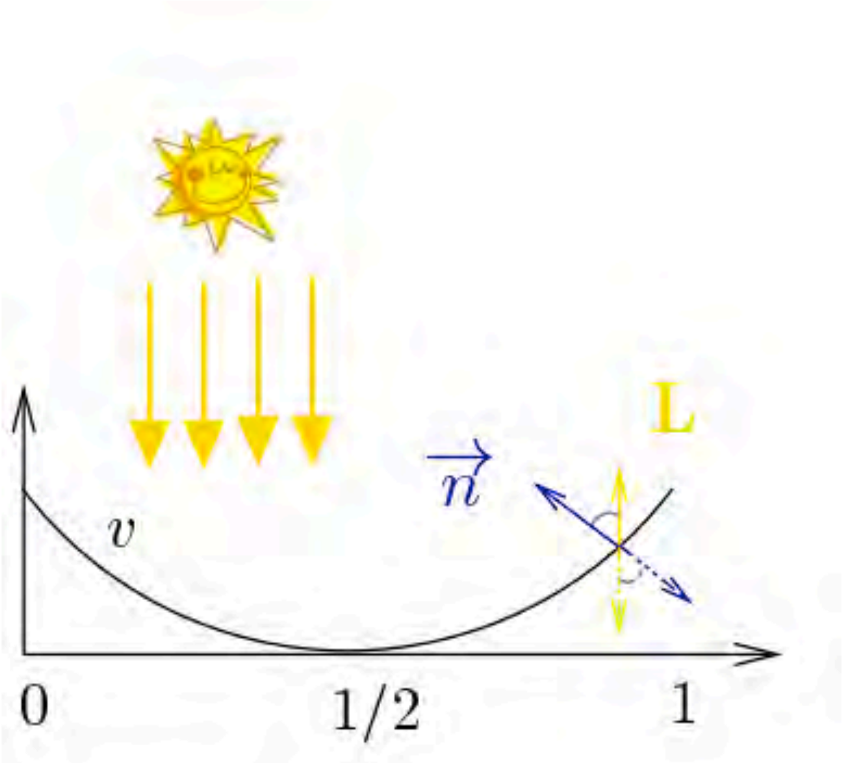
$$\left.\begin{aligned}\frac{\delta z}{\delta u} &= -\frac{n_x^*}{n_z^*}\\ \frac{\delta z}{\delta v} &= -\frac{n_y^*}{n_z^*}\end{aligned}\right\} \Rightarrow \frac{\delta \frac{n_x^*}{n_z^*}}{\delta v} = \frac{\delta \frac{n_y^*}{n_z^*}}{\delta u} \text{ with } \begin{bmatrix} n_x^* \\ n_y^* \\ n_z^* \end{bmatrix} = \mathbf{A}^{-T}\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}$$

$$\Rightarrow \mathbf{A} \text{ restricted to } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}$$

—> The surface f(u, v) can be changed into λf(u,v) + μu + νv and still produce the same image.

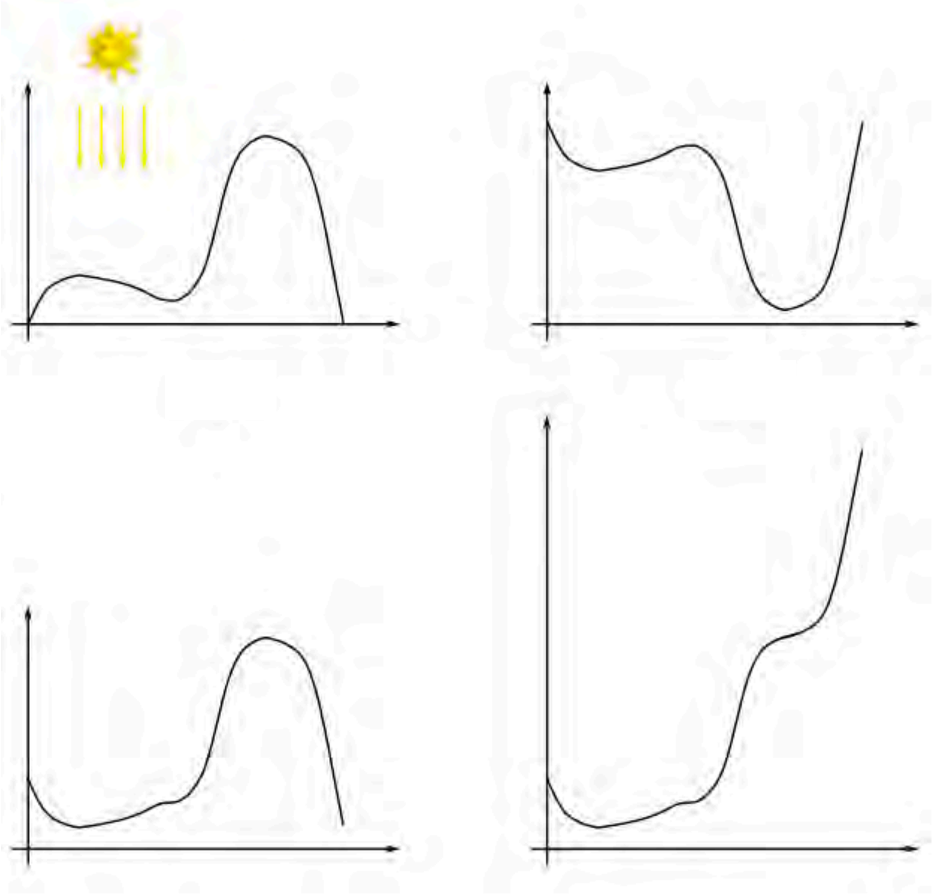
Convex/Concave Ambiguity





It can happen even when the light source is known!

Convex/Concave Ambiguity

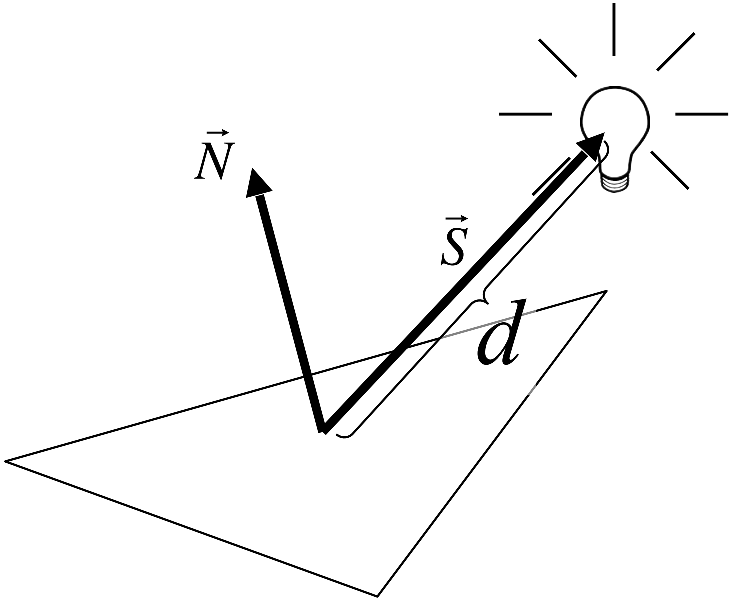


- All four profiles will produce the same image under the illumination shown here.
- The SfS problem under orthographic projection with a distant light-source is ill-posed.

Making the SfS Problem Well-Posed

- Use perspective projection model.
- Radiance depends on distance to light source:

$$I = \frac{Albedo \cdot (\mathbf{N} \bullet \mathbf{S})}{d^2}$$



instead of

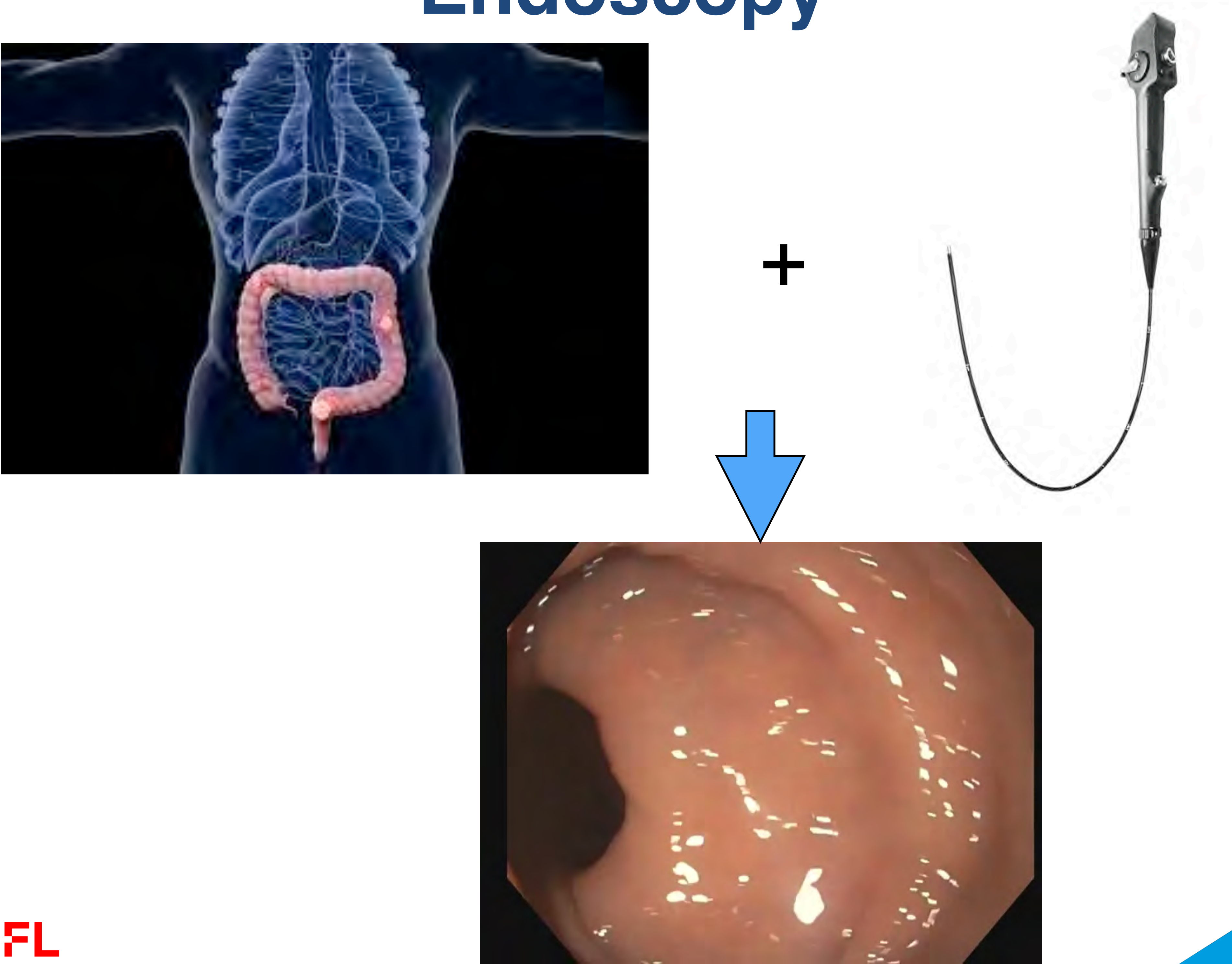
$$I = Albedo \cdot (\mathbf{N} \bullet \mathbf{S})$$

- Light source located at the optical center.

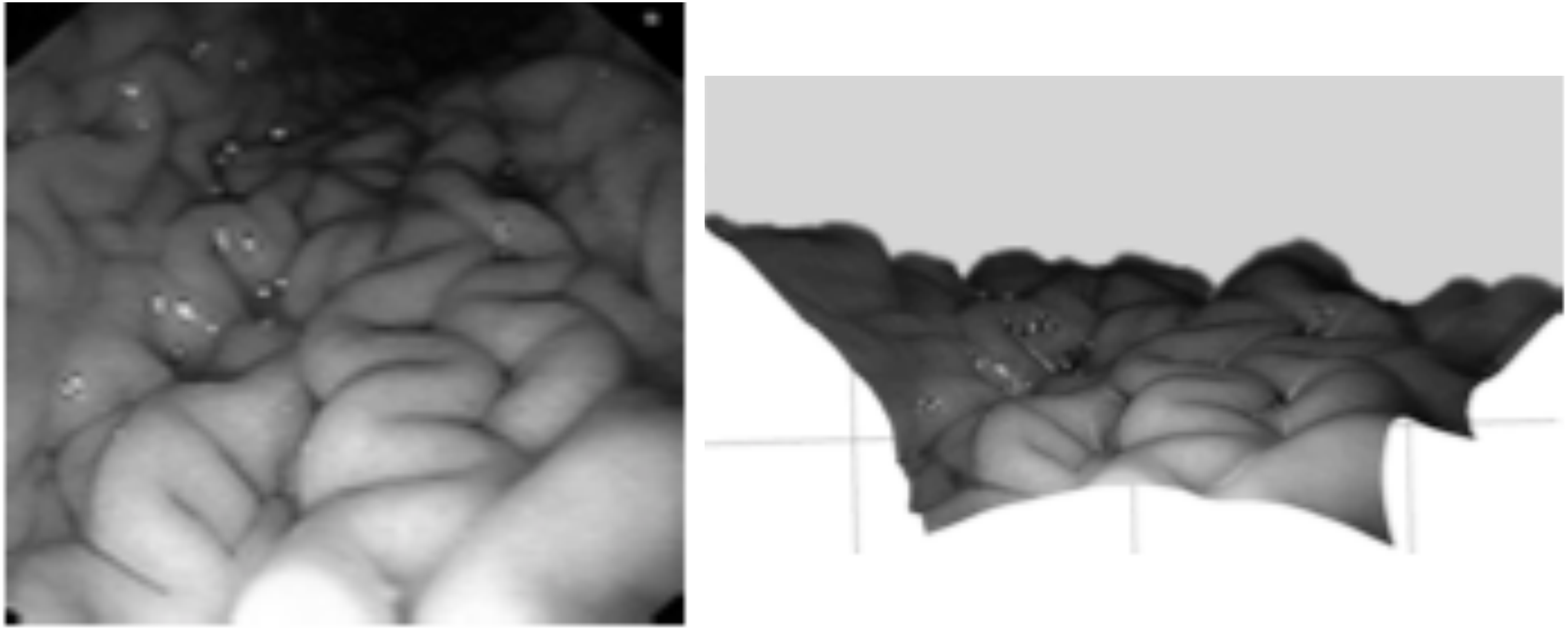
-> Unique solution but more complex computations.

Prados and Faugeras, CVPR'05.

Endoscopy



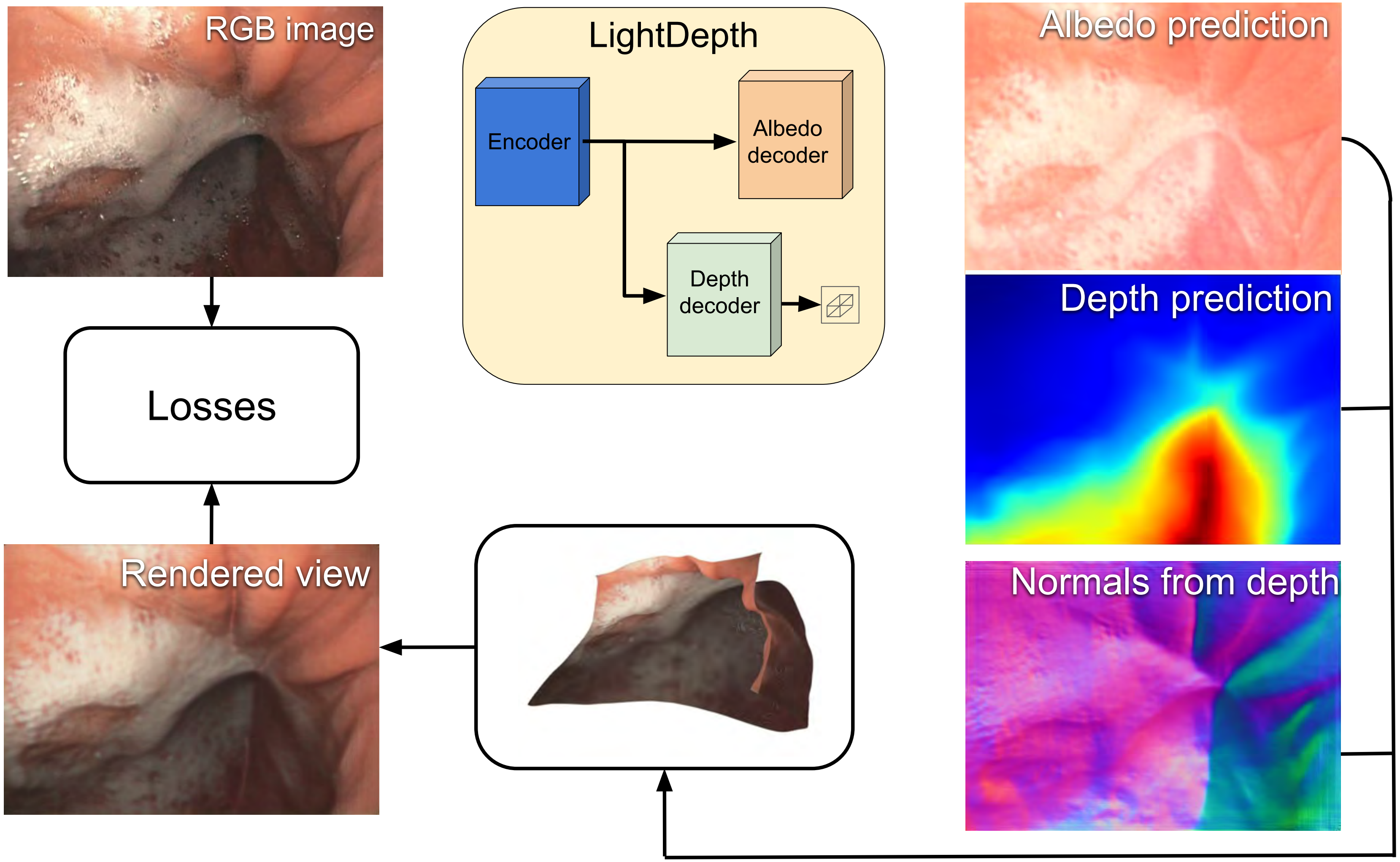
Endoscopy 2005



- The problem becomes well-posed but the variational approach assumes constant albedo, which is limiting.
- Can we take advantage of deep learning to overcome this problem?

Prados and Faugeras, CVPR’05.

Endoscopy 2023

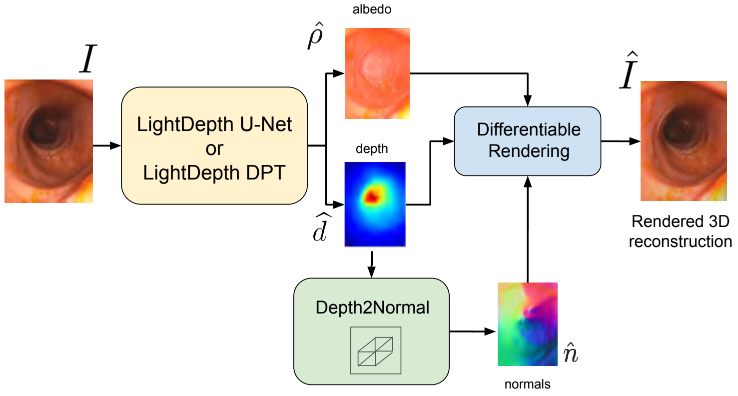


$1/d^2$ term is incorporated in the renderer and makes the problem well-posed by relating depths, albedos, and pixel intensities.

—> The training is fully self-supervised.

Puigvert et al. ICCV'23.

Network Architecture


I

LightDepth U-Net or LightDepth DPT

$\hat{\rho}$ albedo

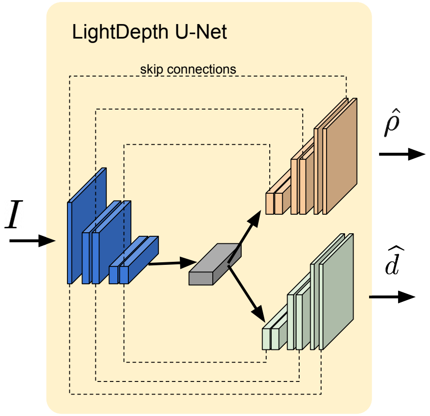
depth $\widehat{d}$

Depth2Normal

normals $\hat{n}$

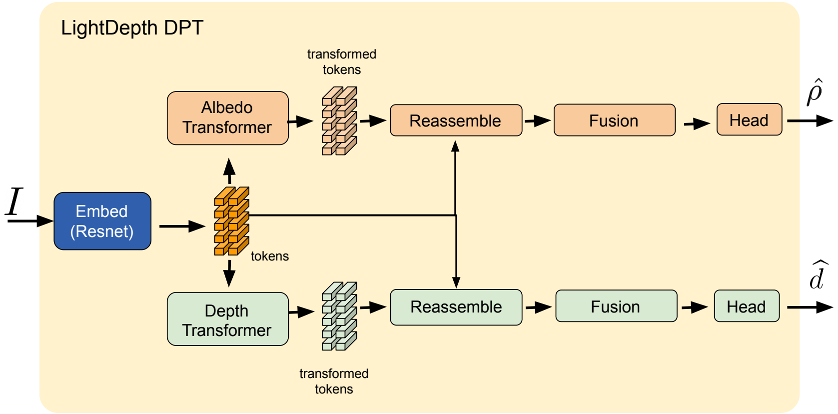
Differentiable Rendering

$\hat{I}$ Rendered 3D reconstruction


LightDepth U-Net

skip connections

I $\hat{\rho}$ $\widehat{d}$


LightDepth DPT

I Embed (Resnet) → tokens

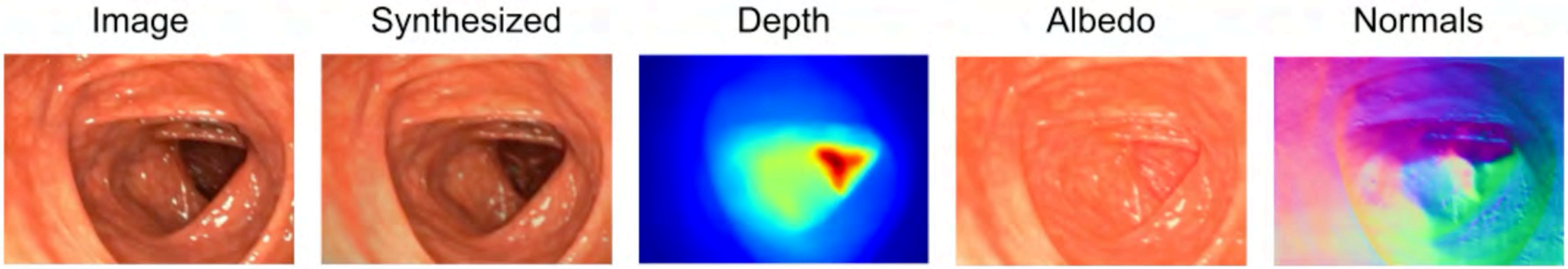
Albedo Transformer → transformed tokens → Reassemble → Fusion → Head → $\hat{\rho}$

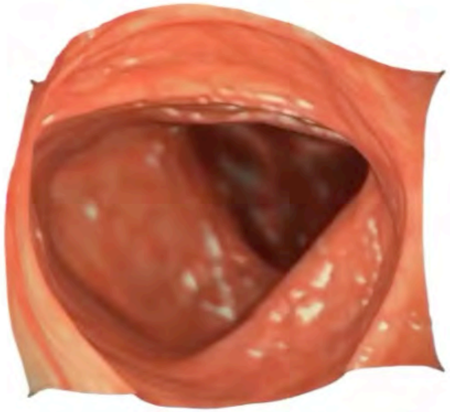
Depth Transformer → transformed tokens → Reassemble → Fusion → Head → $\widehat{d}$

Convolutional

Transformer

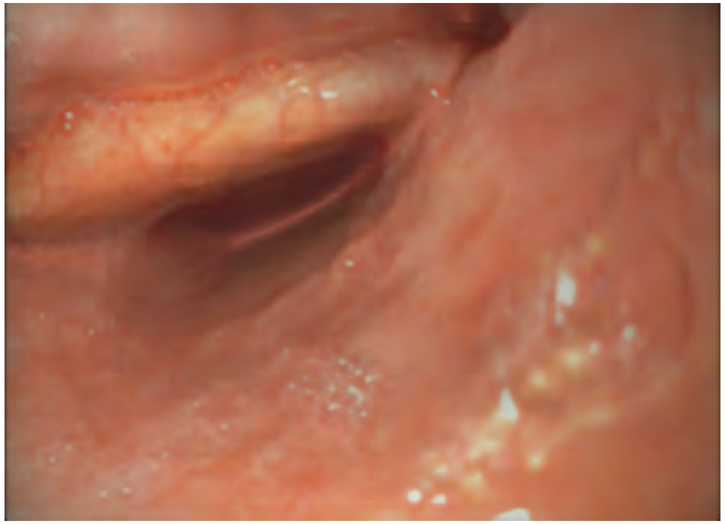
Colonoscopy Images

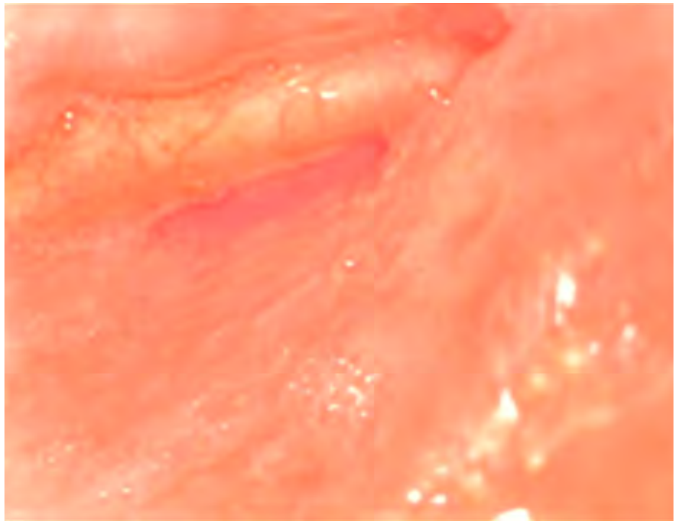


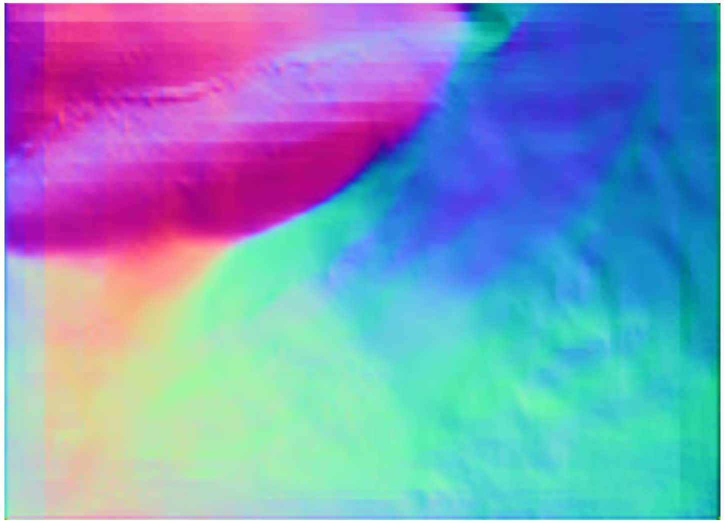


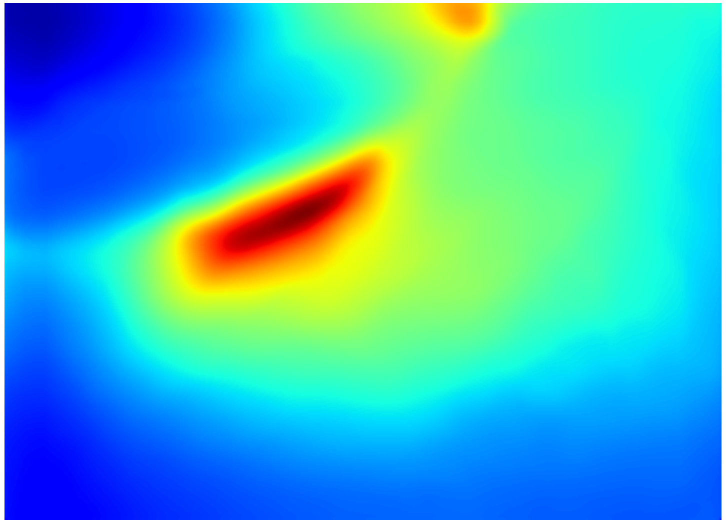
Gastroscopy Images


Image

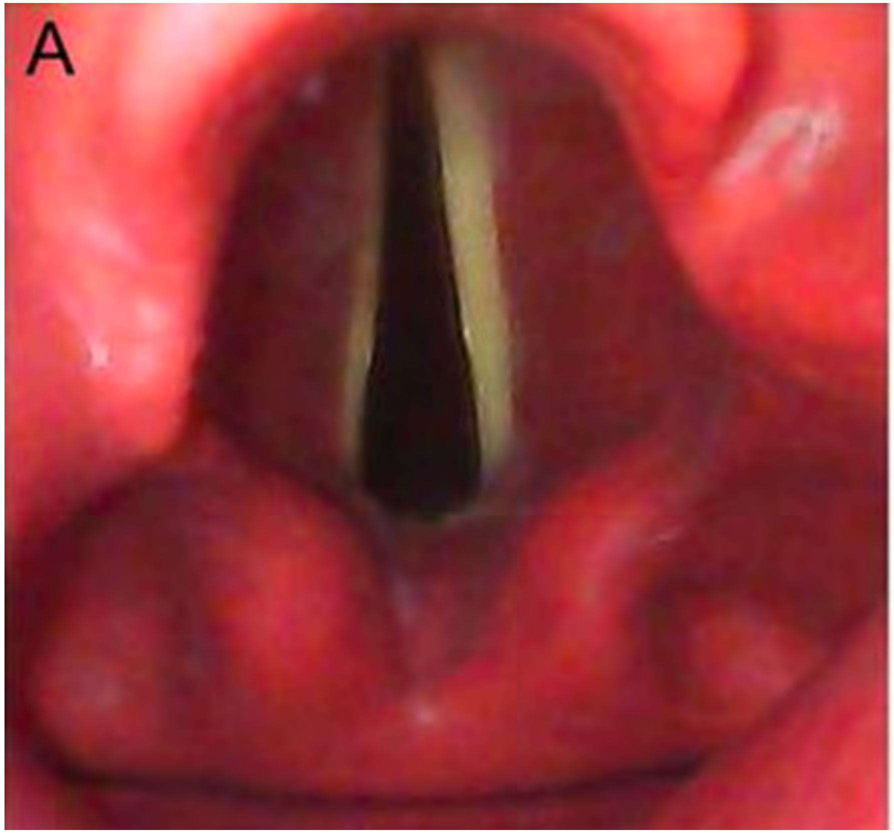

Synthesized


Albedo


Normals


Depth

Automating Intubation



Goal:

- Provide real-time automated guidance.
- Enable a broader range of personnel to perform one in an emergency.

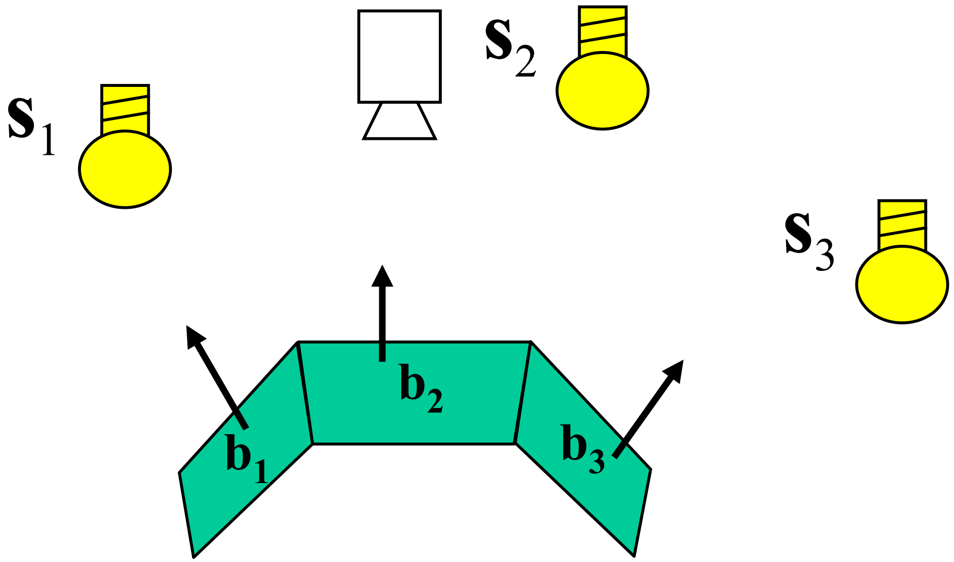
Phantom Reconstructions

Real Image

Synthetic Image

Depth

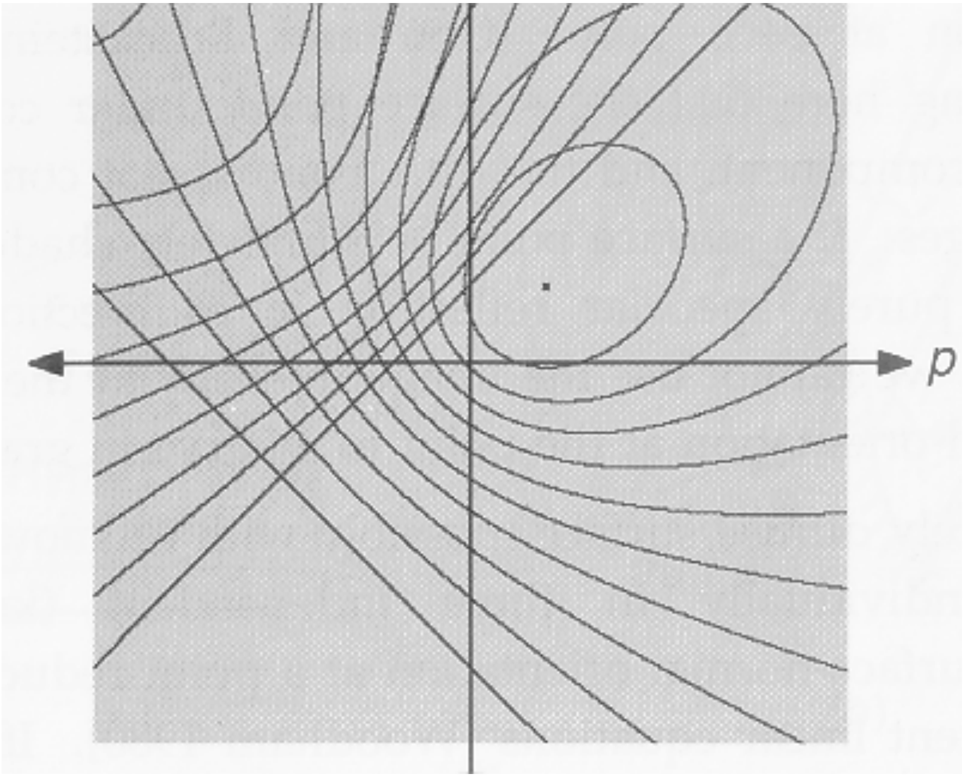
Photometric Stereo

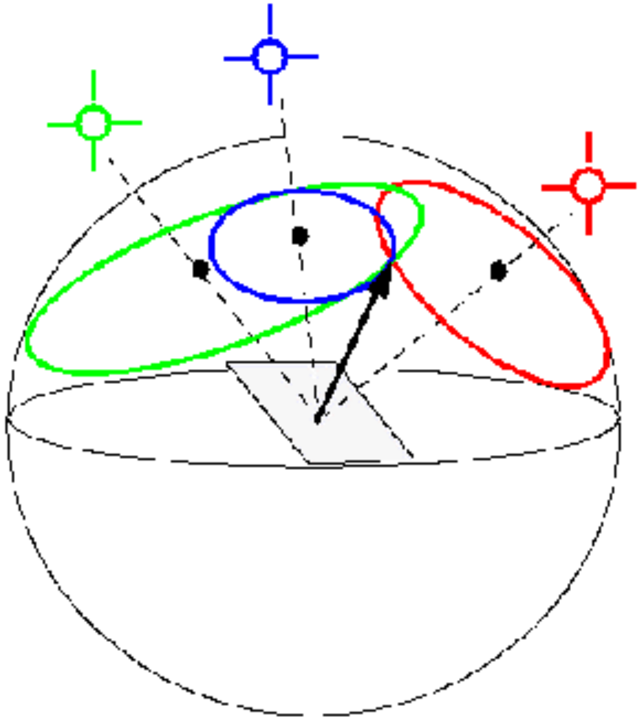


Given multiple images of the same surface under different known lighting conditions, can we recover the surface shape?

- **Yes!** (Woodham, 1978).
- This gave rise to an early and still practical application of computer vision.

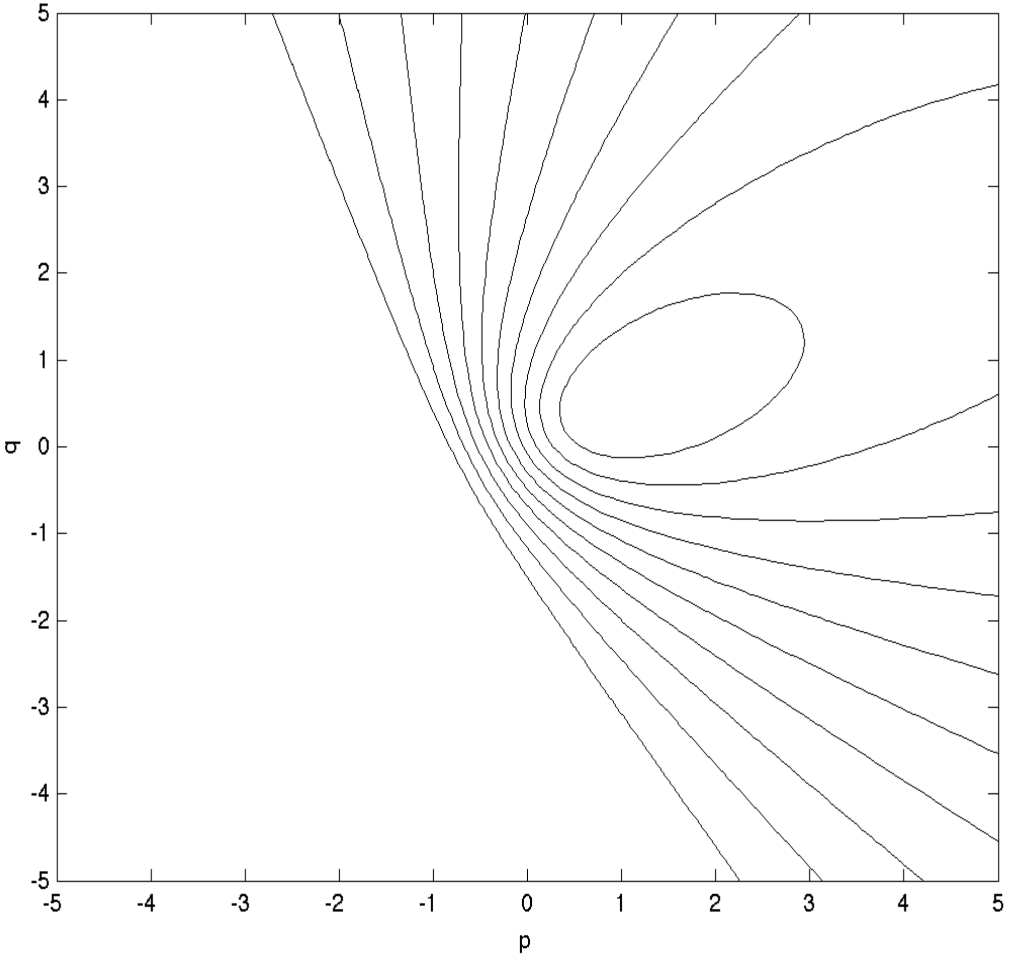
Basic Idea

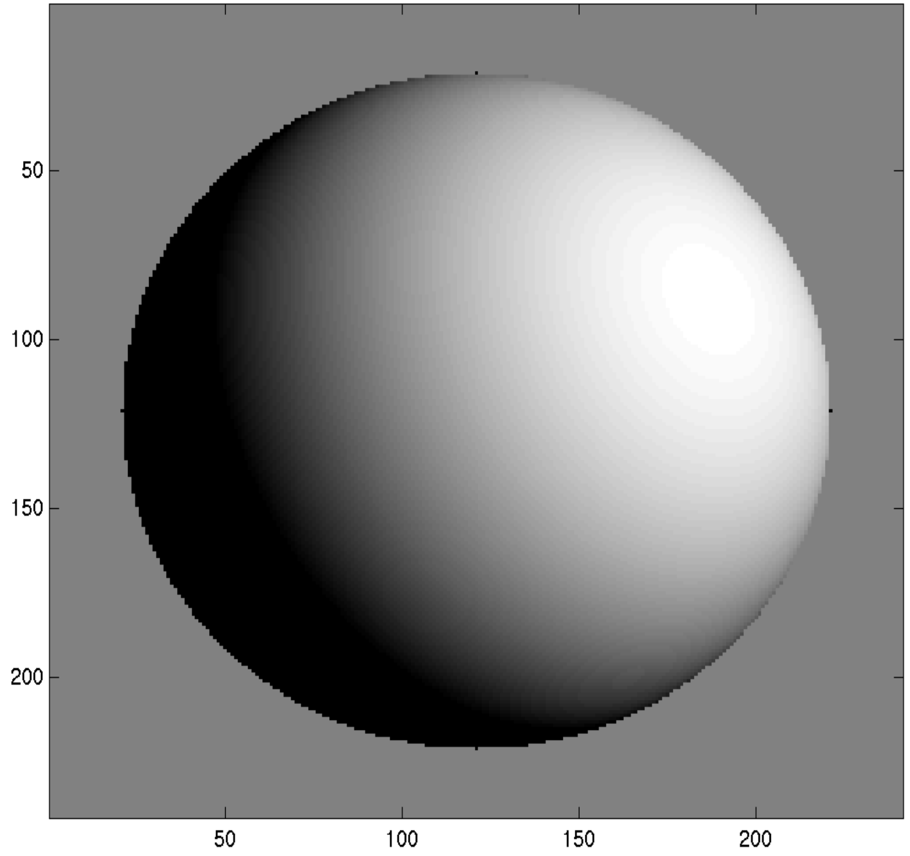




- Take several images under different lighting conditions.
- Infer the normals from the changes in illumination.
- Given at least three different lights, there are no more ambiguities.

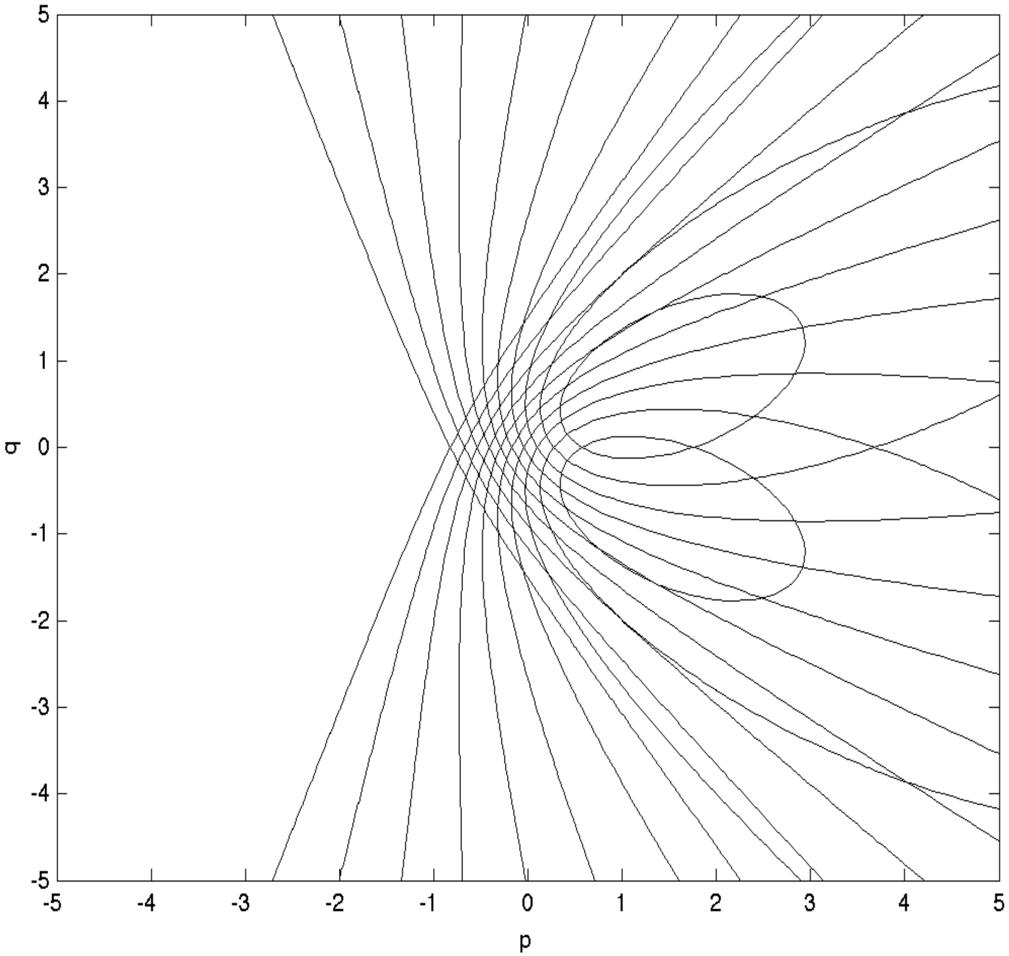
One Single Light Source

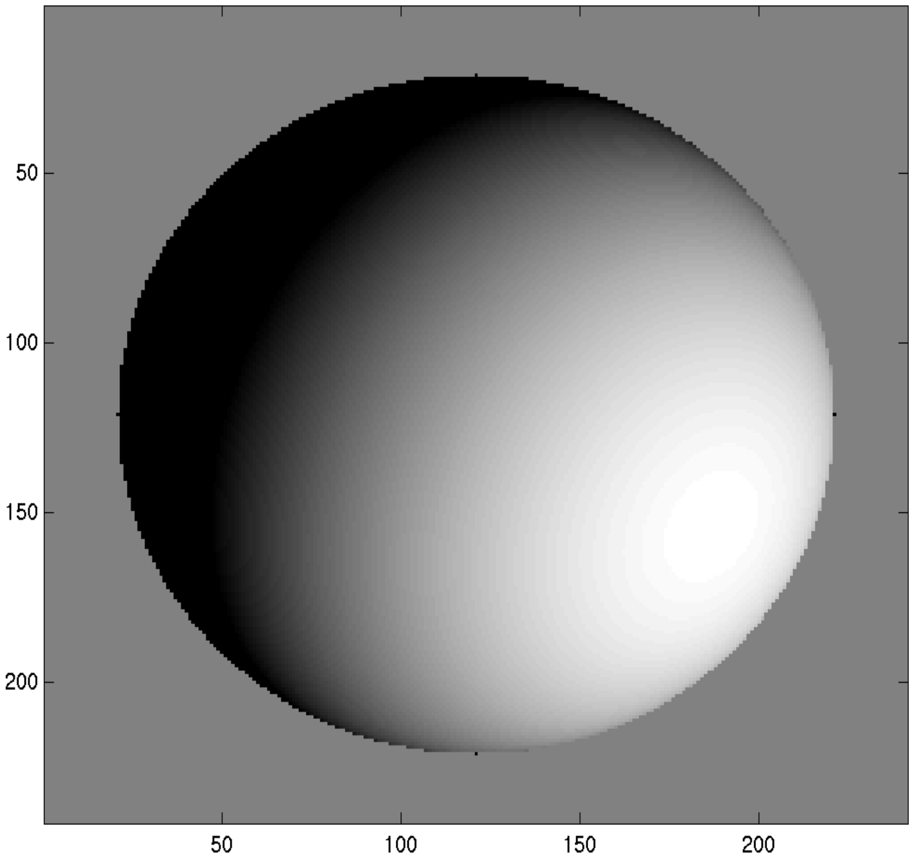




Many potential normals for each image point.

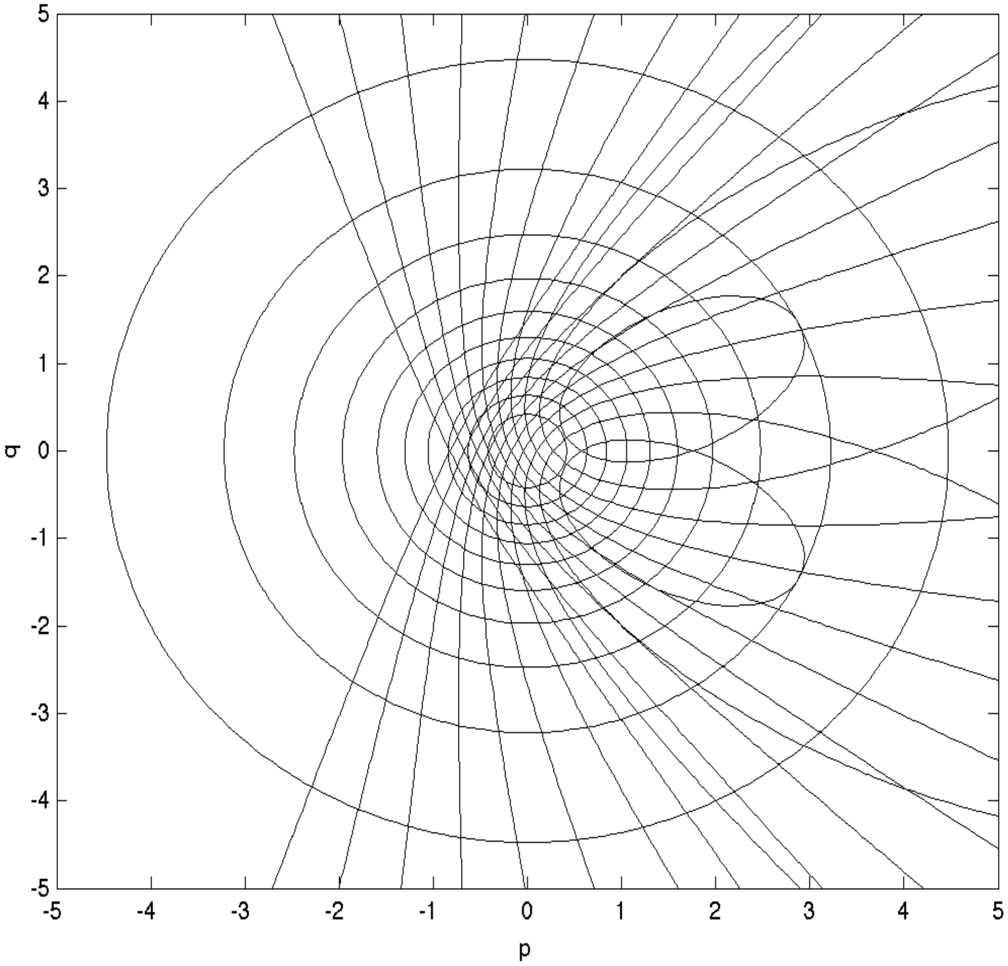
Two Light Sources

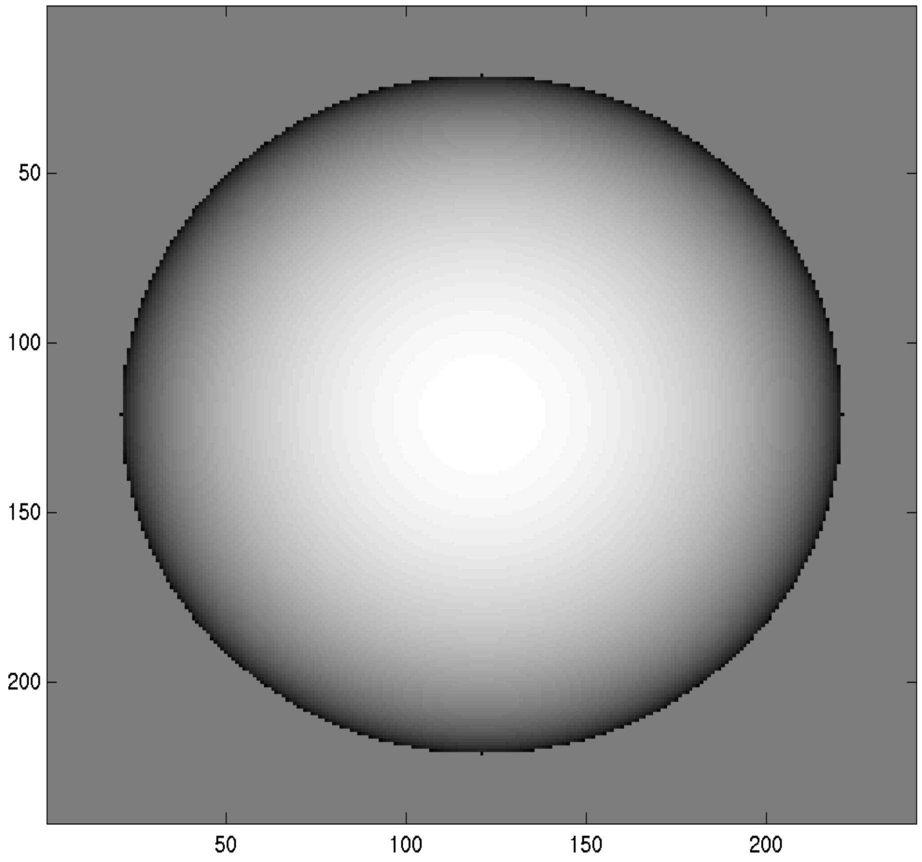




Still some ambiguities.

Three Light Sources





No more ambiguities even if the albedo is unknown.

Algebraic Formulation

Lambertian model:

$$I \quad = \quad \alpha(\mathbf{L} \cdot \mathbf{N}) = (\mathbf{L} \cdot \mathbf{M})$$

Three light sources:

Unknown 3 vector that can be estimated by solving a 3x3 linear system.

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \\ \mathbf{L}_3 \end{bmatrix} \mathbf{M}$$

$$\mathbf{N} = \frac{\mathbf{M}}{||\mathbf{M}||}$$

$$\alpha = ||\mathbf{M}||$$

N and a can then be inferred from **M**.

Using More Lights

One can use as many lights as one wants:

$$\mathbf{I} = \mathbf{L}\mathbf{M}, \text{ with } \mathbf{I} = \begin{bmatrix} I_1 \\ : \\ I_n \end{bmatrix} \text{ and } \mathbf{L} = \begin{bmatrix} \mathbf{L}_1 \\ : \\ \mathbf{L}_n \end{bmatrix}$$

$$\Rightarrow \mathbf{L}^t\mathbf{L}\mathbf{M} = \mathbf{L}^t\mathbf{I} \text{ (Least - squares solution)}$$

—> This is known as an over-constraint problem and, the more camera, the more robust to noise the solution is.

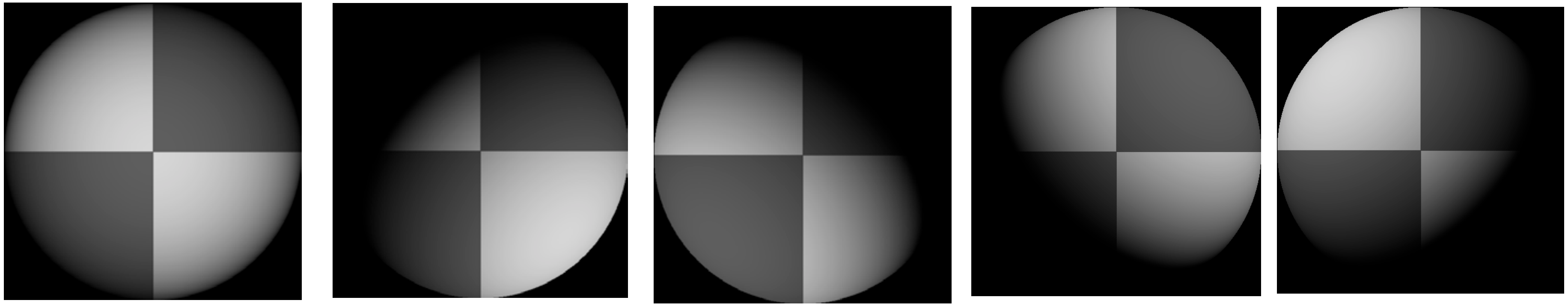
Discounting Shadows

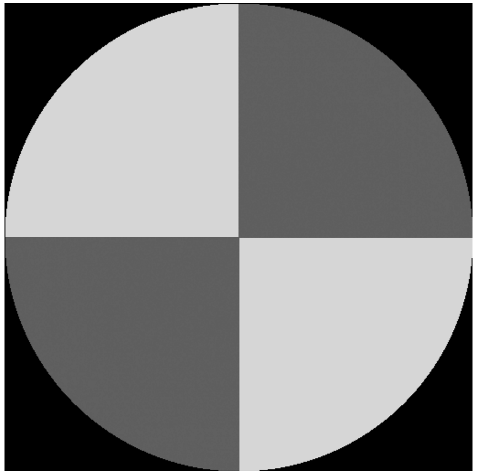
- Shadowed pixels for a given light source position do not conform to the model.
- Premultiplying by the intensities reduces their contributions because their intensities tend be lower.

$$\begin{bmatrix} I_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & I_n \end{bmatrix} \boldsymbol{I} = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & I_n \end{bmatrix} \boldsymbol{LM}$$

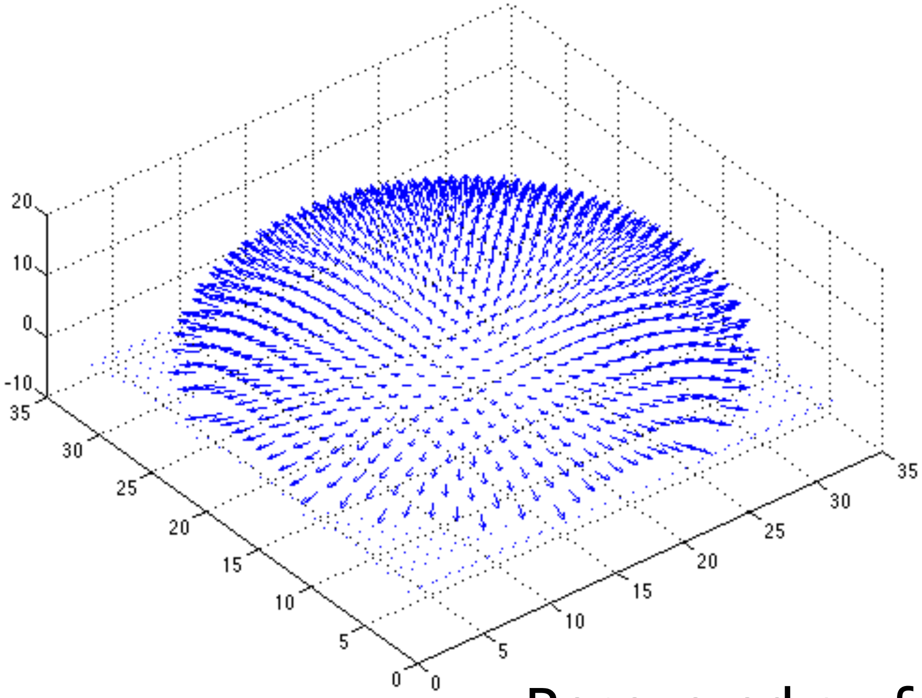
Synthetic Sphere Images

Five different light sources:



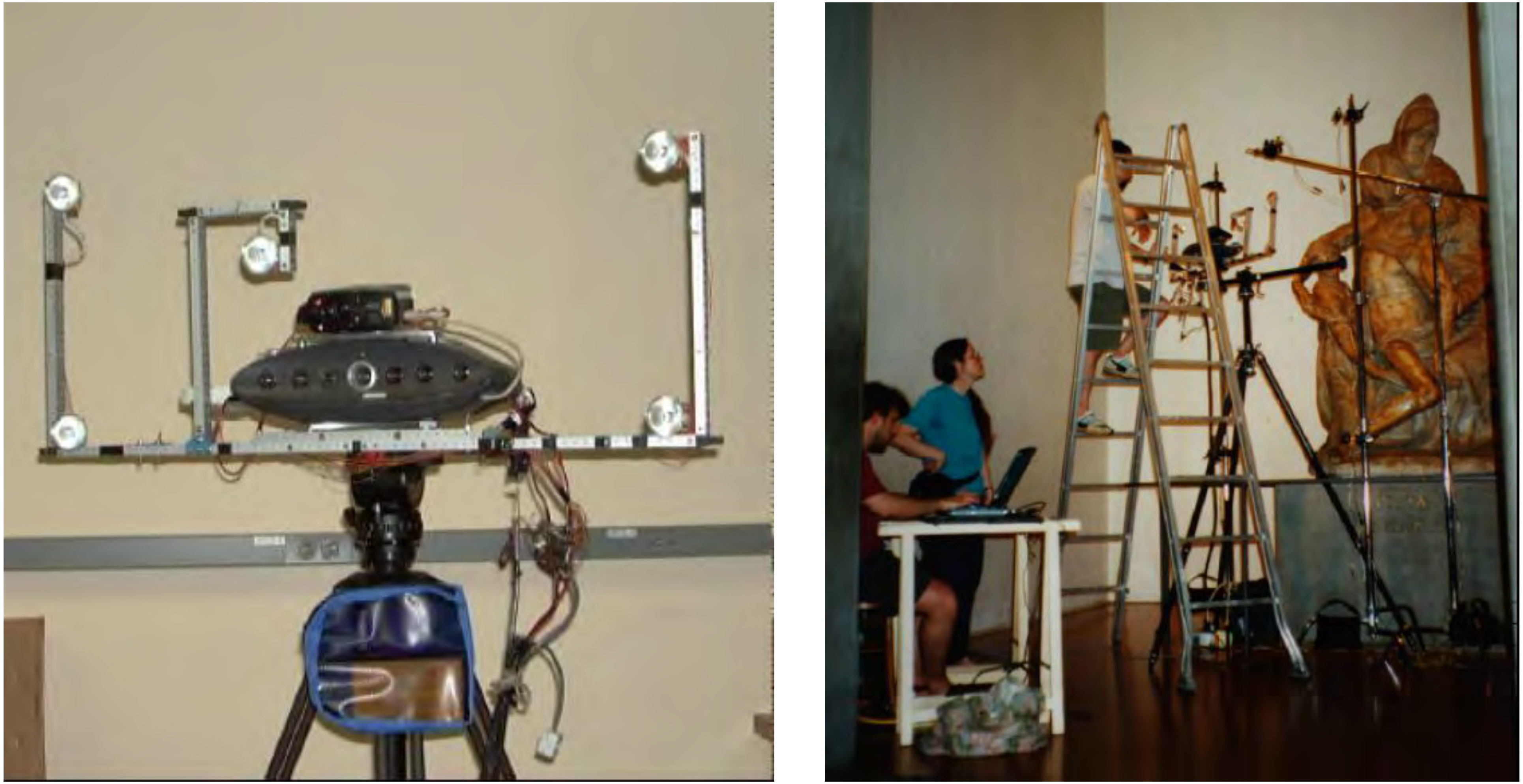


Recovered albedo



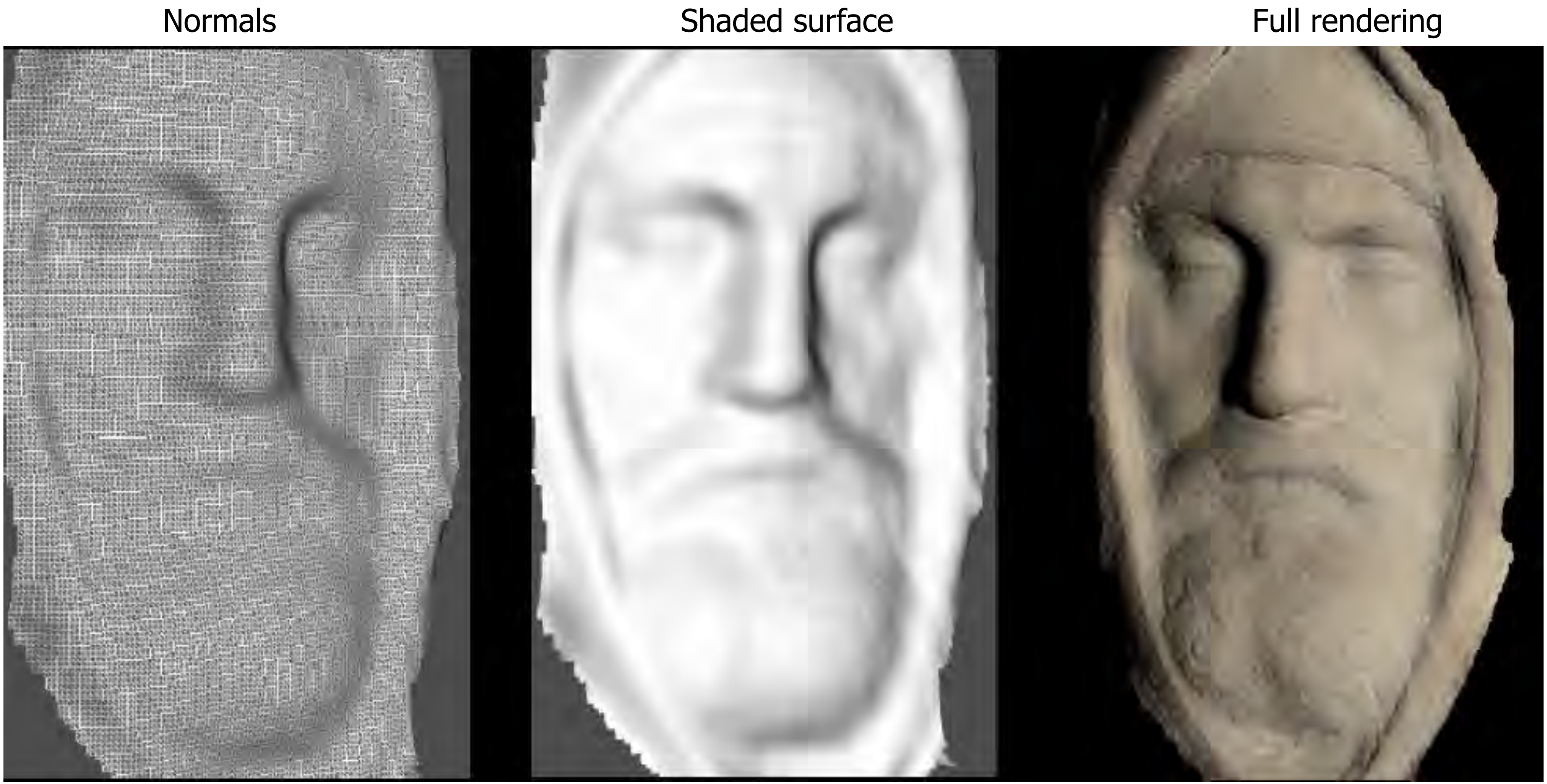
Recovered surface normals

Scanning Michel Angelo’s Pieta



- One camera and five light sources.
- The positions of the light sources w.r.t. the camera are exactly known.

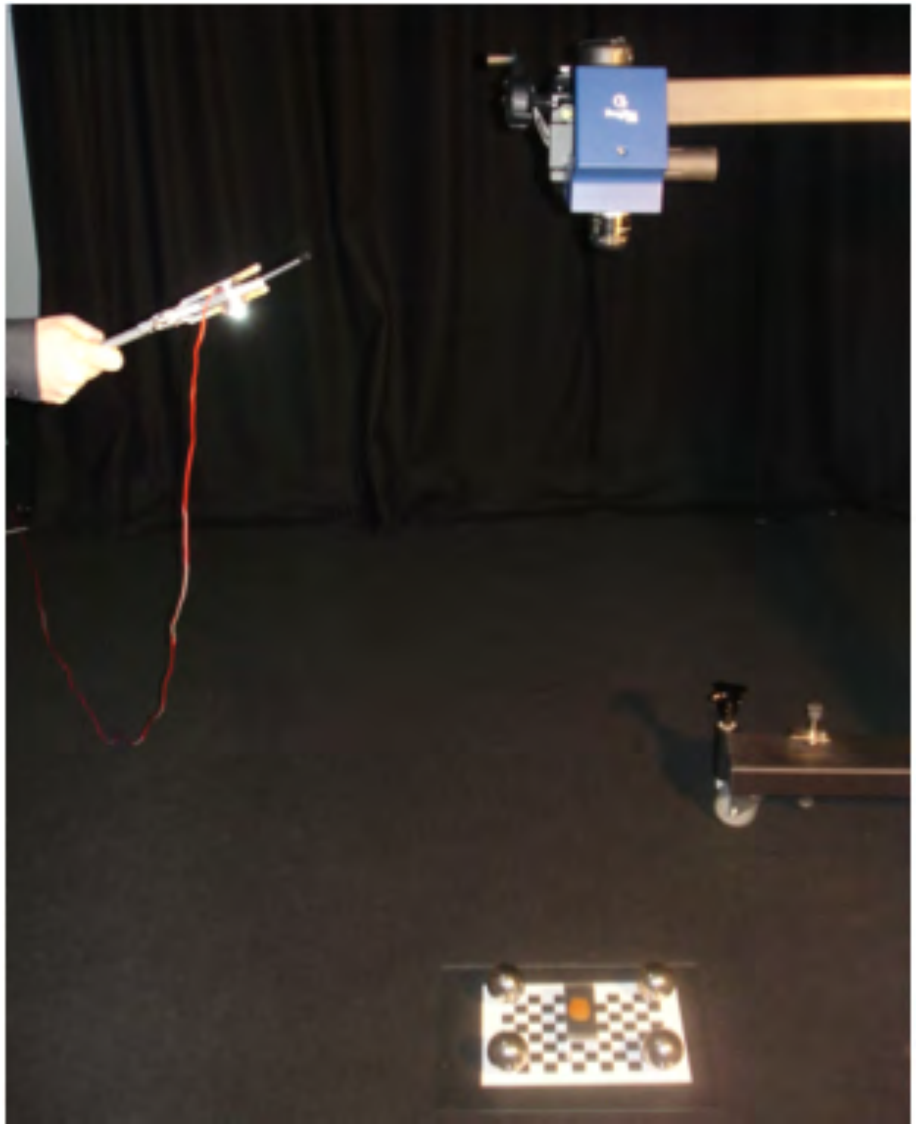
Full 3D Model

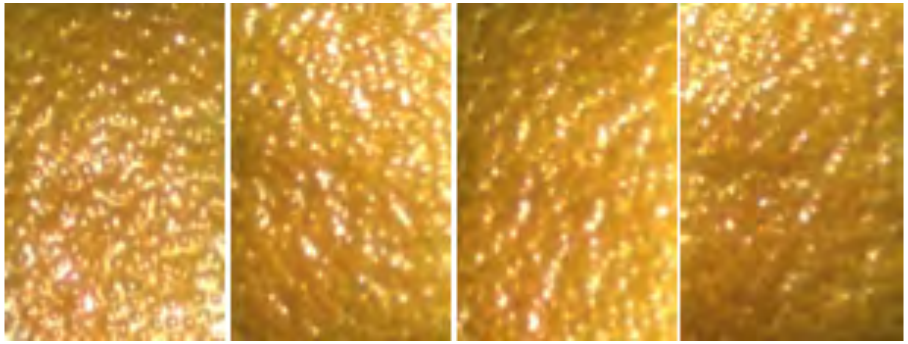


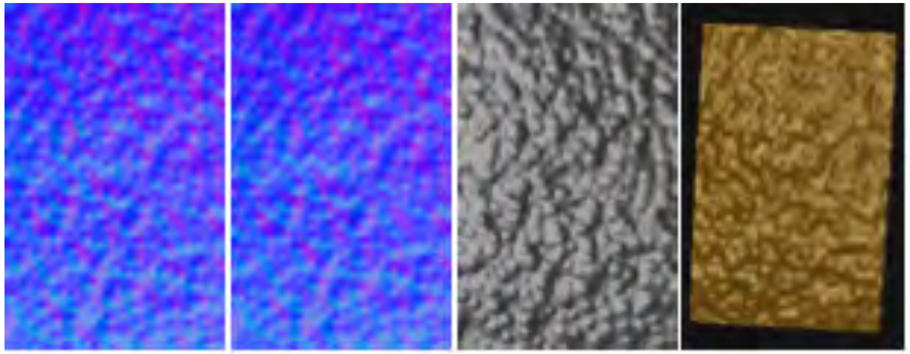
- This was done for the whole statue.
- All the fragments were then “glued” together.

—> A full 3D model that can be visualized from any viewpoint and under any illumination conditions.

Optional: Shape from Specularities



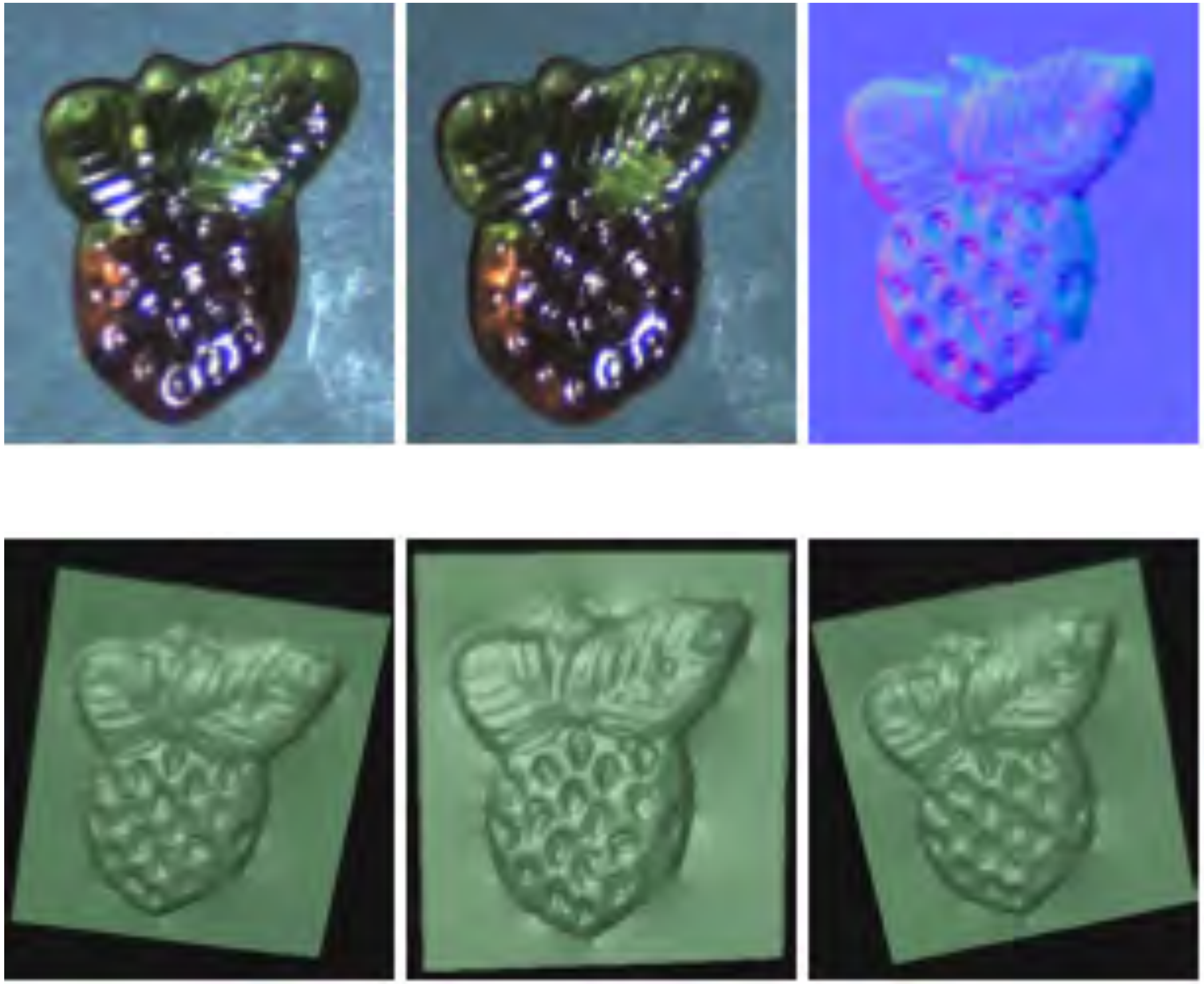




- Move the light around and compute its position each time.
- Find the bright spots and the image and assume they are secularities.
- Infer the normals at those points.

Chen et al., CVPR 2006

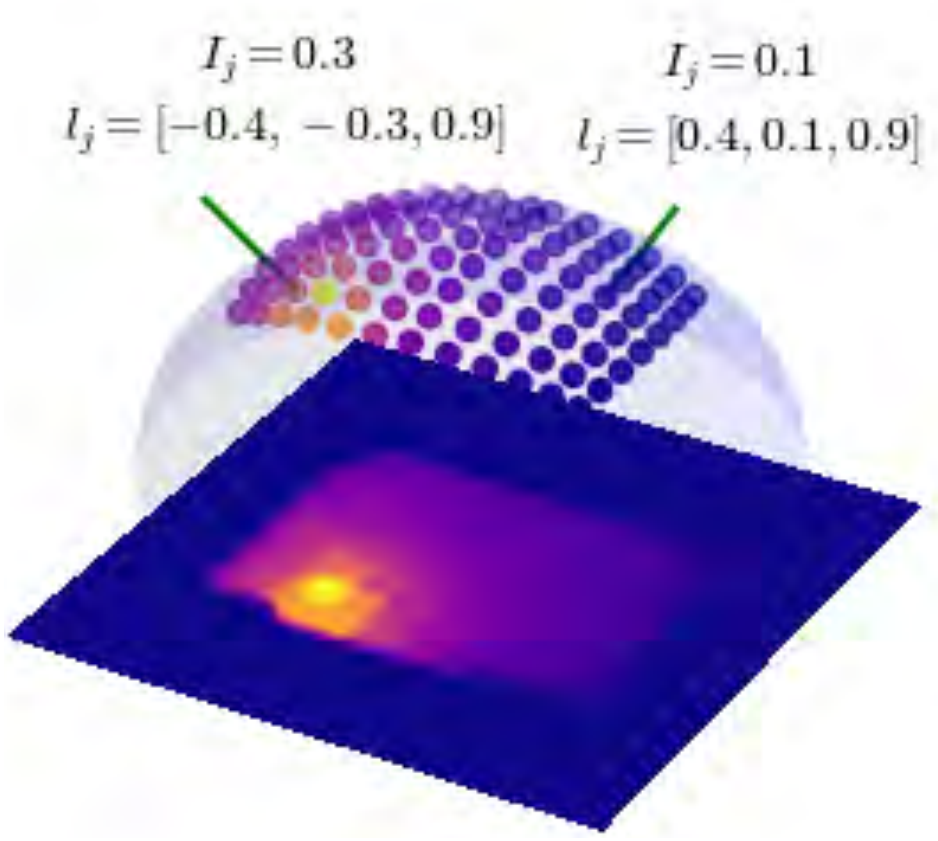
Optional: Shape from Specularities



- Excellent precision can be achieved because the secularities are very sensitive to the exact normal direction.
- However, this only works well for shiny, that is, highly specular, objects.

Chen et al., CVPR 2006

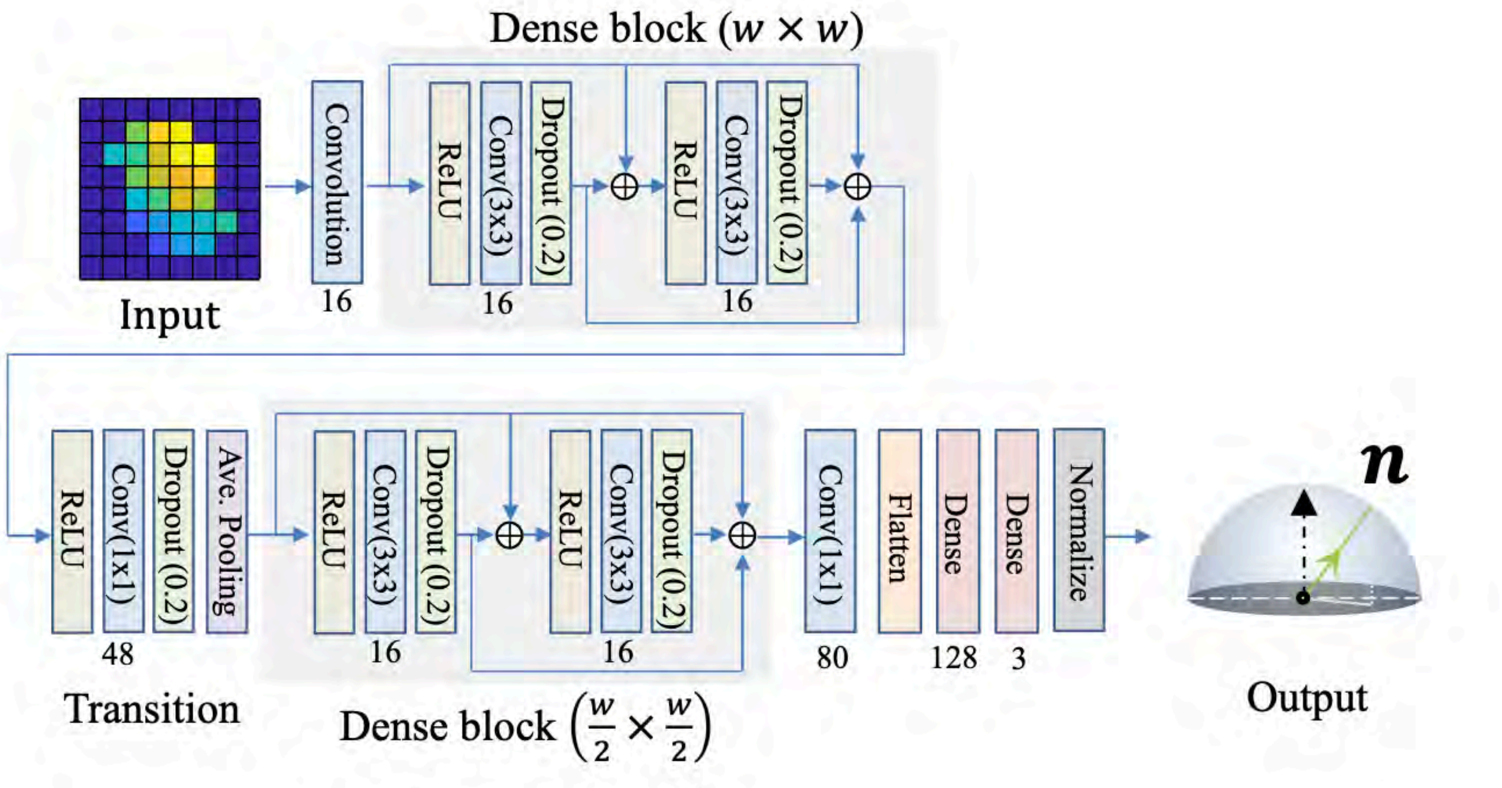
Deep Photogrammetric Stereo



- Build an observation map at each pixel.
- Each map pixel represents an observation under an illumination direction defined on a unit-hemisphere.

Ikehata, ECCV’18

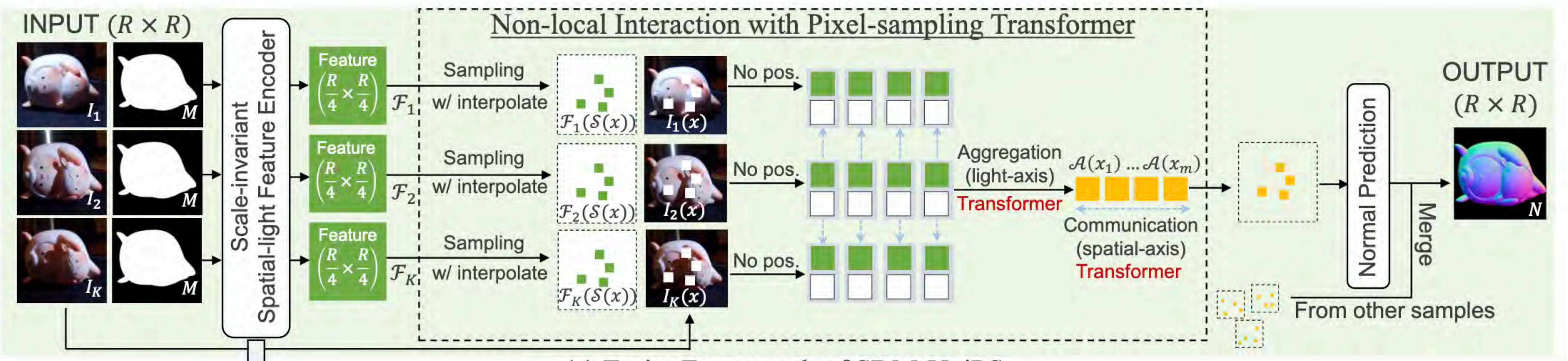
Deep Photogrammetric Stereo



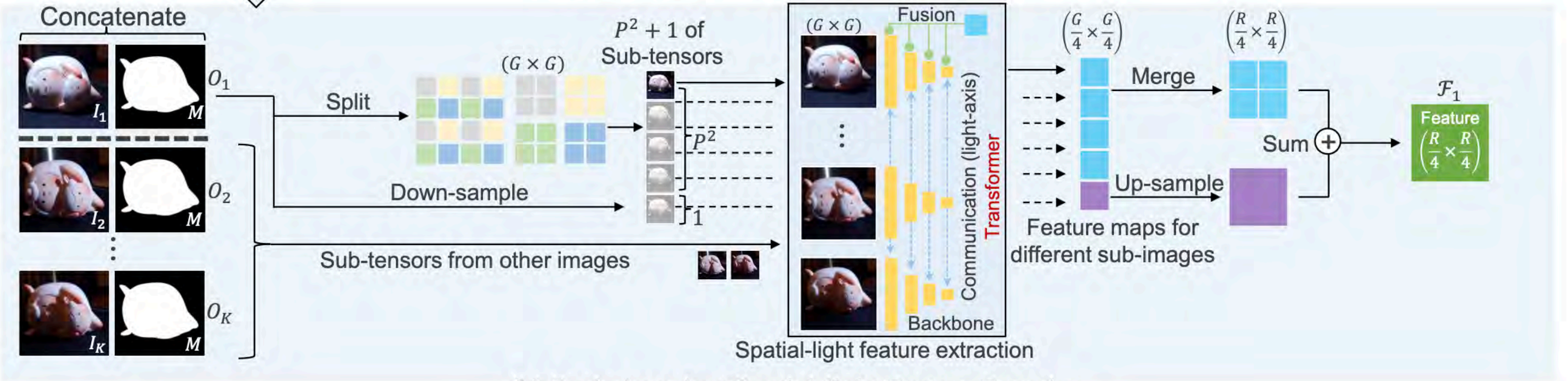
- Observation maps as input to a CNN.
- Take the relationship between neighboring pixels into account.

Ikehata, ECCV'18

Optional: Bring in the Transformers



(a) Entire Framework of SDM-UniPS

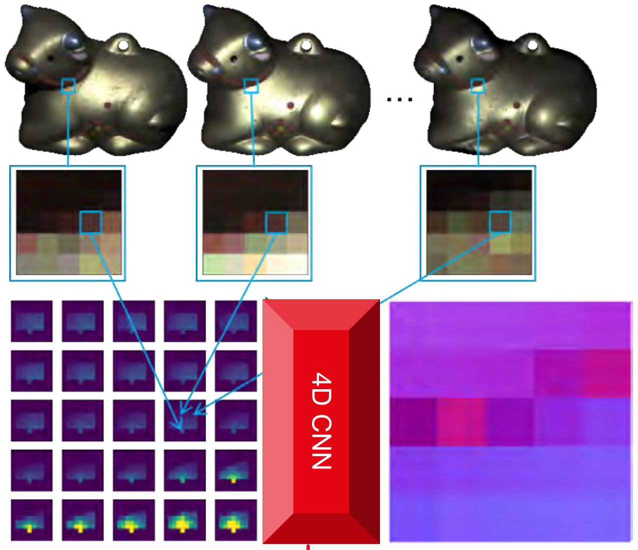


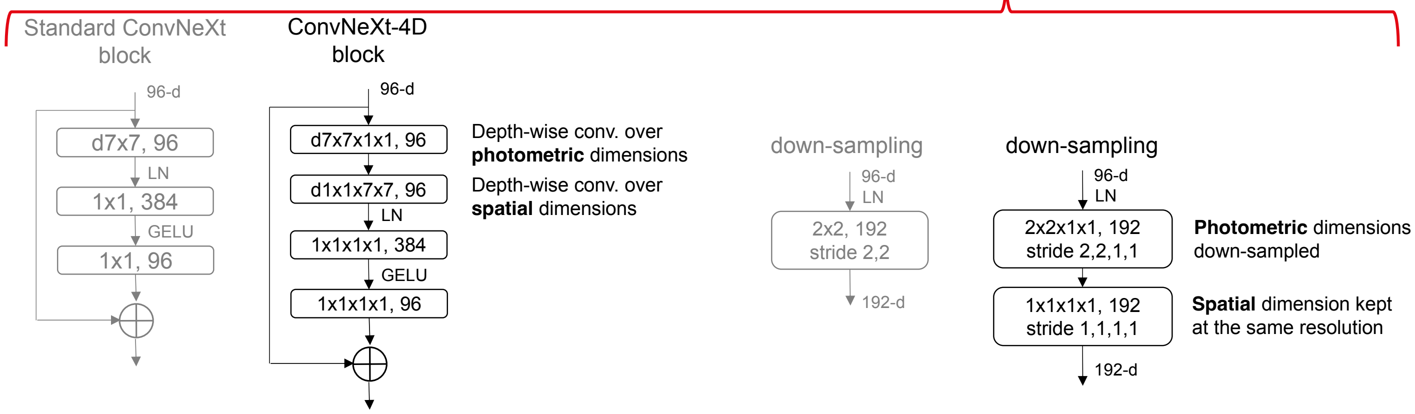
(b) Scale-invariant Spatial-light Feature Encoder

Ikehata, ICCV'23

Optional: ConvNeXt-4D

- Leverages both spatial and photometric context **in every layer**
- Input: **Arbitrarily large** patches of observation maps
- Output: **Patches** of normals





Standard ConvNeXt block
96-d
d7x7, 96
LN
1x1, 384
GELU
1x1, 96

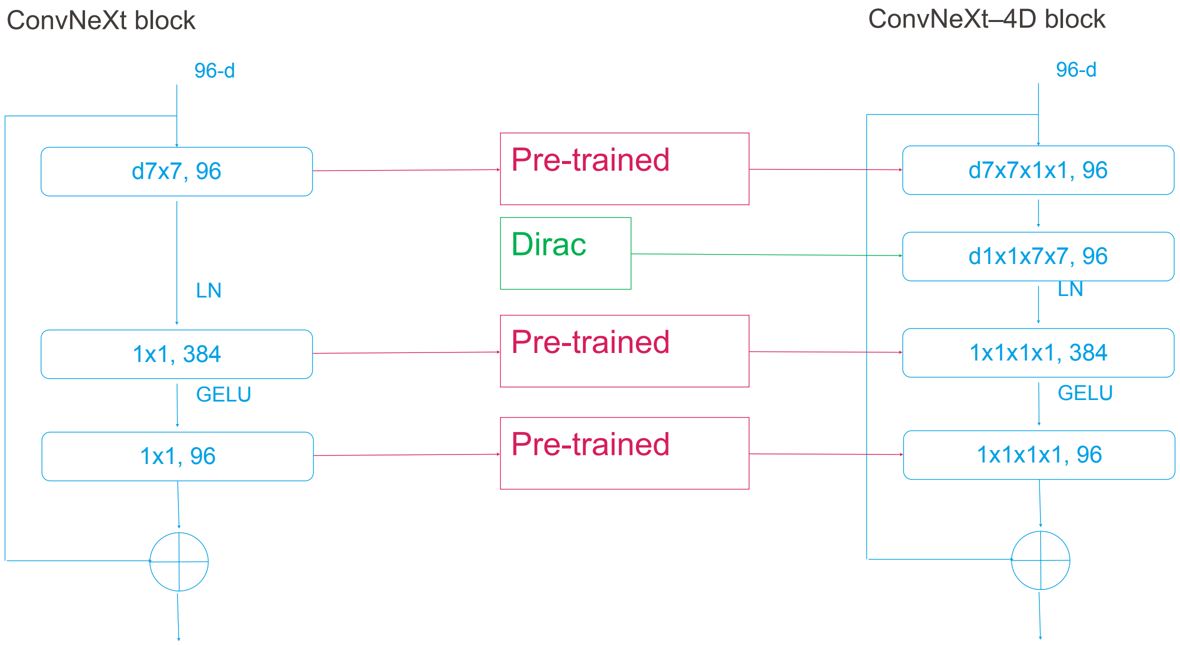
ConvNeXt-4D block
96-d
d7x7x1x1, 96 — Depth-wise conv. over **photometric** dimensions
d1x1x7x7, 96 — Depth-wise conv. over **spatial** dimensions
LN
1x1x1x1, 384
GELU
1x1x1x1, 96

down-sampling
96-d
LN
2x2, stride 2,2
192-d

down-sampling
96-d
LN
2x2x1x1, 192 stride 2,2,1,1 — **Photometric** dimensions down-sampled
1x1x1x1, 192 stride 1,1,1,1 — **Spatial** dimension kept at the same resolution
192-d

Honzátko, arXiv'24

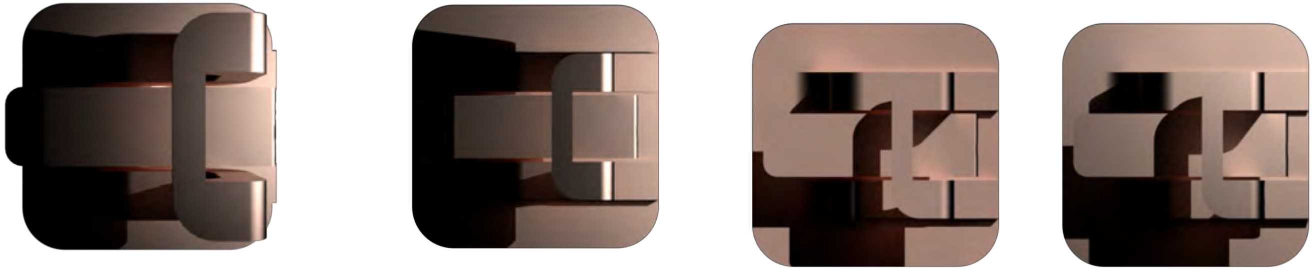
Optional: Two-Stage Training



- Pre-train ConvNeXt as a per-pixel method.
- Fine-tune ConvNeXt-4.

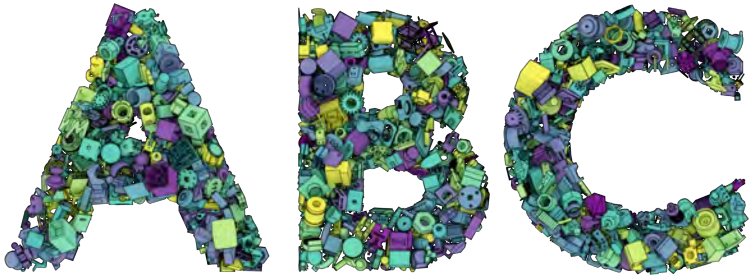
Optional: Online Rendering

using Mitsuba3 ray-tracer

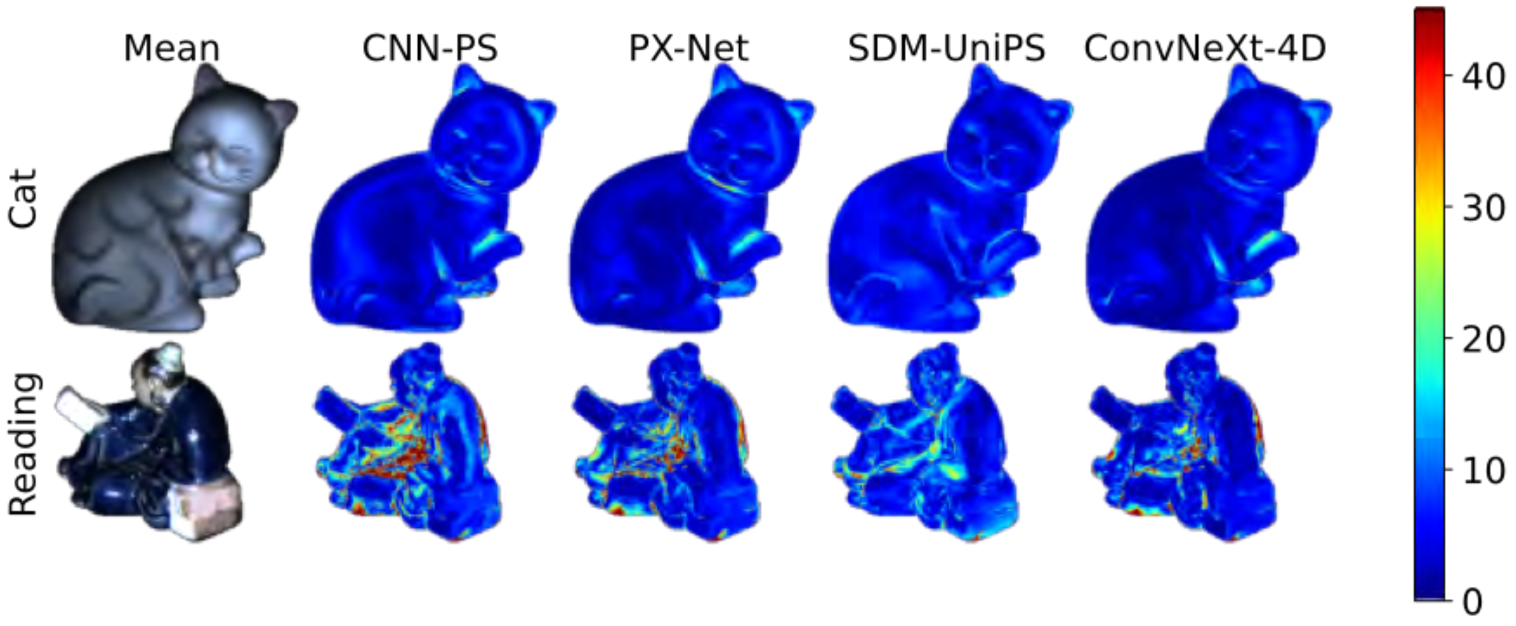


- ABC dataset with 10e6 objects
- Possibility to use custom dataset of objects relevant for particular application
- Infinite number of BRDFs (e.g. Disney, MERL, RGL, NBRDF, Plastic, Metallic, …)
- No spatial limitation

➡ Designed a PS path tracer that renders 8 illums. at once



Optional: BenchMarking



Method	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	Average
Lambertian	4.1	8.4	14.9	8.4	25.6	18.5	30.6	8.9	16.7	19.8	15.4
PS-FCN	2.8	7.6	7.9	6.2	7.3	8.6	15.9	7.1	7.3	13.3	8.4
Attention-PSN	2.9	4.9	7.8	6.1	8.9	8.4	15.4	6.9	7.0	12.9	7.9
CNN-PS K=10 (EECV 2018)	2.2	4.1	7.9	4.6	8.0	7.3	14.0	5.4	6.0	12.6	7.2
PX-NET K=10 *(ICCV 2021)*	*2.0*	***3.5***	*7.6*	*4.3*	*4.7*	*6.7*	*13.3*	*4.9*	*5.0*	*9.8*	*6.2*
U-NET 4D K=12 (3DV 2021)	2.0	**3.5**	6.9	4.4	4.8	6.7	12.6	4.8	4.6	12.0	6.2
SDM-UniPS ICCV 2023	**1.5**	3.6	7.5	5.4	**4.5**	8.5	**10.2**	**4.7**	**4.1**	8.2	**5.8**
ConvNeXt-4D	1.8	3.6	**6.2**	**3.9**	5.1	**7.1**	12.8	4.8	4.6	**8.0**	**5.8**

Shape-from-Shading in Short

Traditional Shape-from-Shading requires strong assumptions:

- Constant or piece-wise constant albedo
- No interreflections
- No shadows
- No specularities

➡ In a single image context, it is most useful in conjunction with other information sources.

➡ These assumptions can be relaxed when the light-source is nearby or when using multiple images.