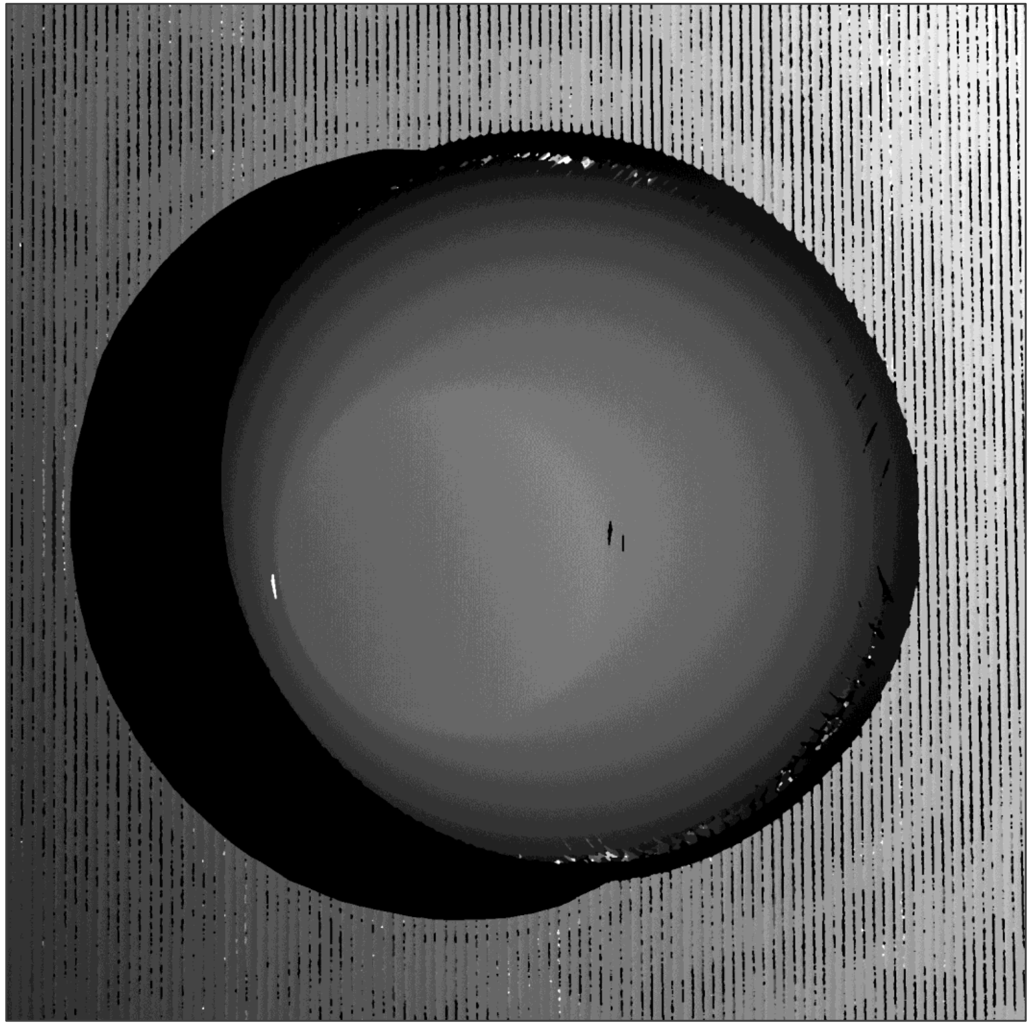
Stereo and structured light



15-463, 15-663, 15-862
Computational Photography
Fall 2018, Lecture 20

http://graphics.cs.cmu.edu/courses/15-463

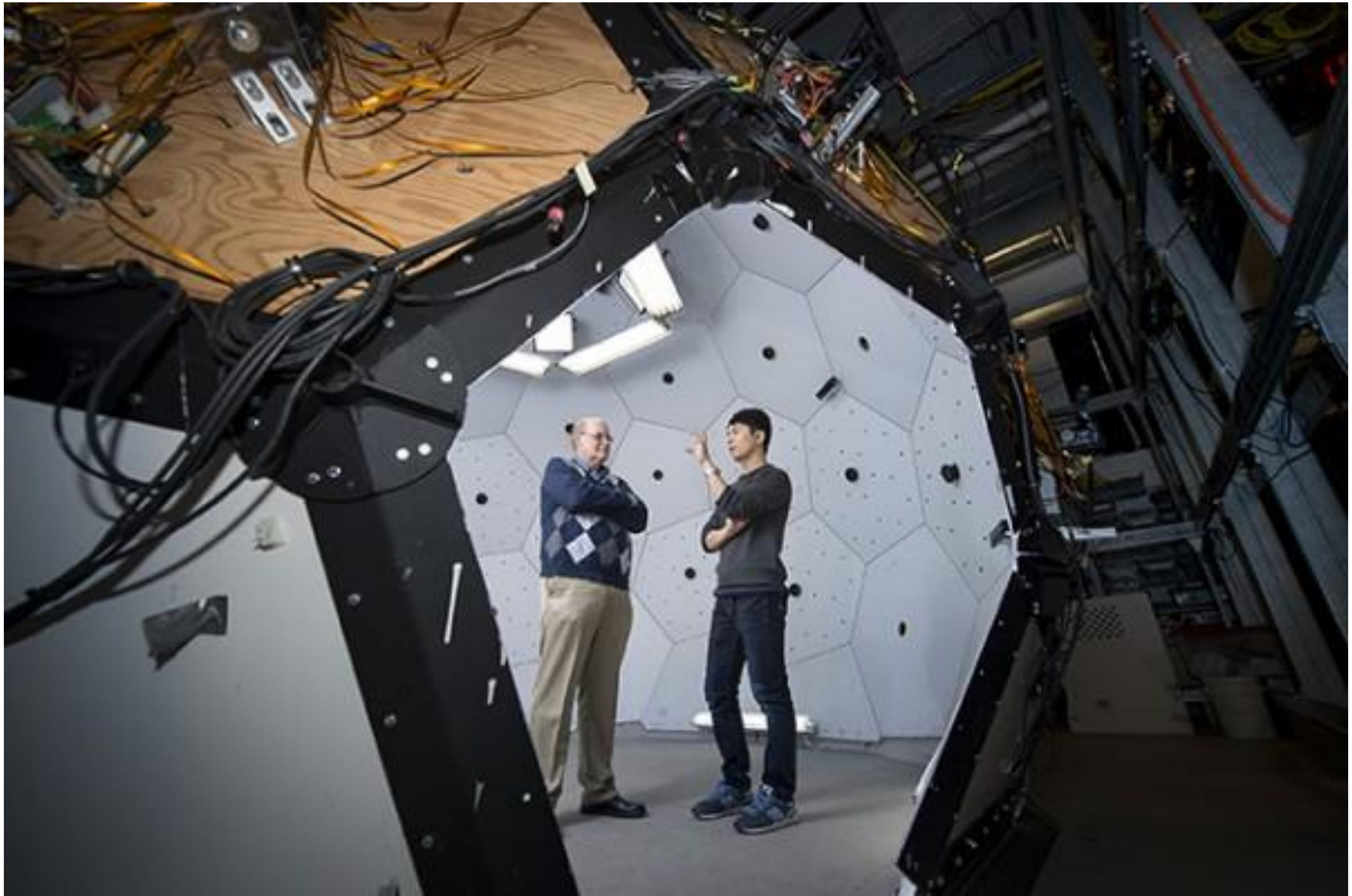
Course announcements

- Homework 5 is still ongoing.
  - **Make sure to download updated version!**
  - Any questions about homework 5?

- How many of you attended Eric Fossum's talk?

- Guest lecture on Wednesday: Han Joo will talk about the panoptic studio.

CMU’s Panoptic Studio



Overview of today’s lecture

- Revisiting triangulation.
- Disparity.
- Stereo rectification.
- Stereo matching.
- Structured light.
- Binary coding.
- Dealing with global illumination.
- Epipolar imaging.

Slide credits

Many of these slides were adapted directly from:

- Kris Kitani (16-385, Spring 2017).
- Srinivasa Narasimhan (16-820, Spring 2017).
- Mohit Gupta (Wisconsin).

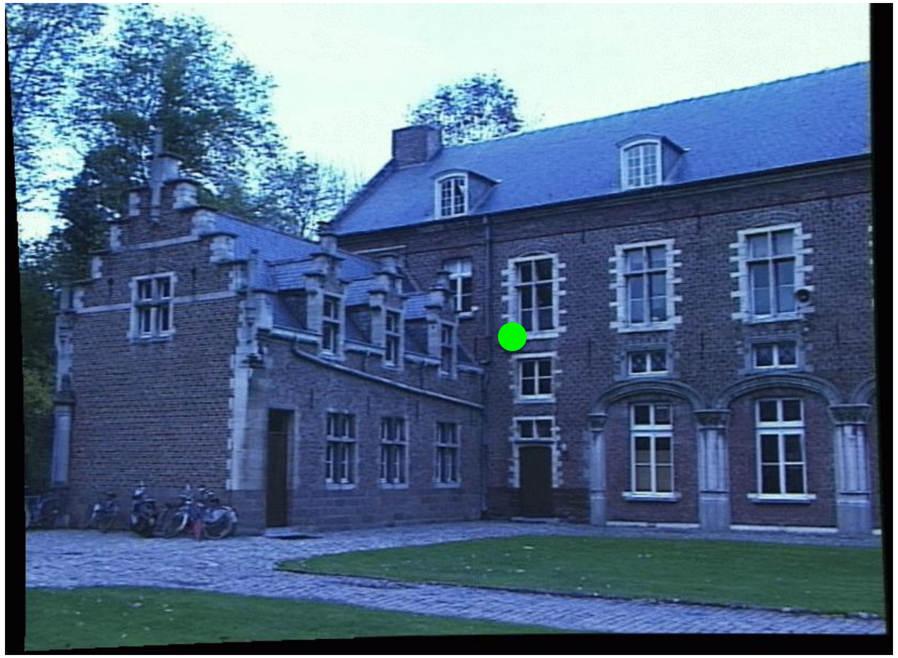
Revisiting triangulation

How would you reconstruct 3D points?

Left image

Right image

How would you reconstruct 3D points?

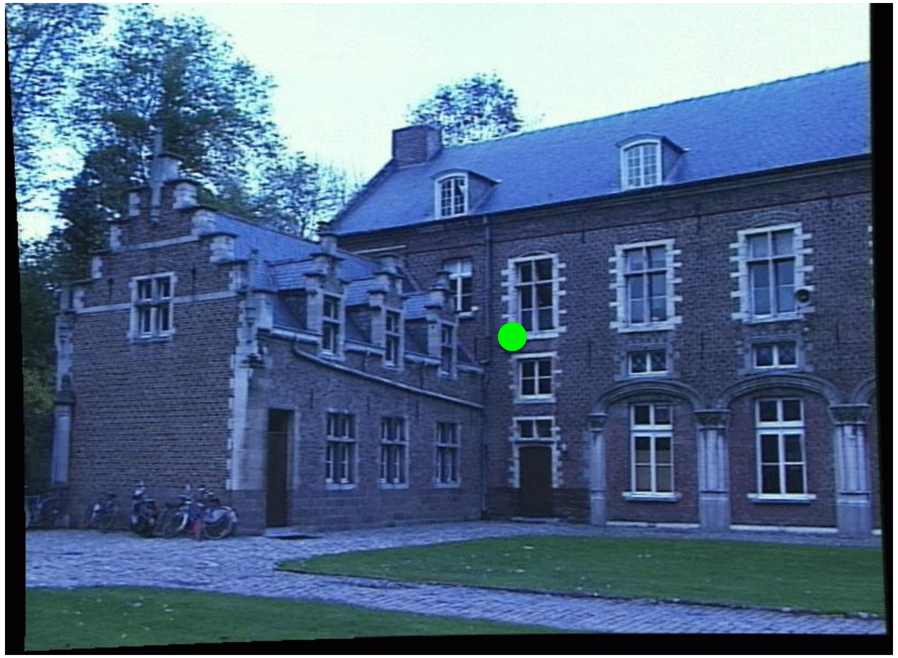


Left image

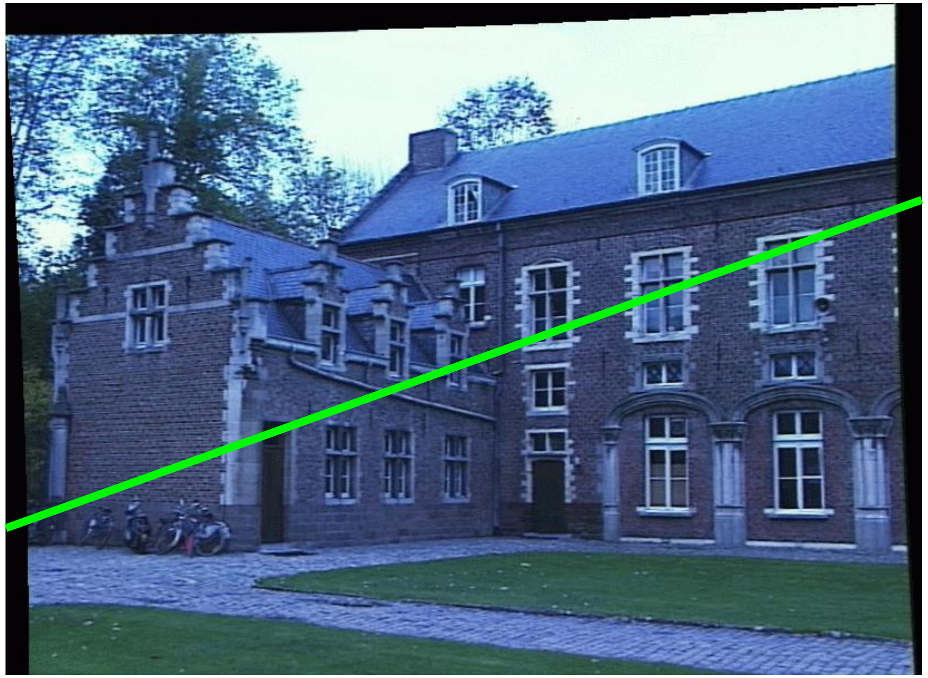
Right image

1. Select point in one image

How would you reconstruct 3D points?



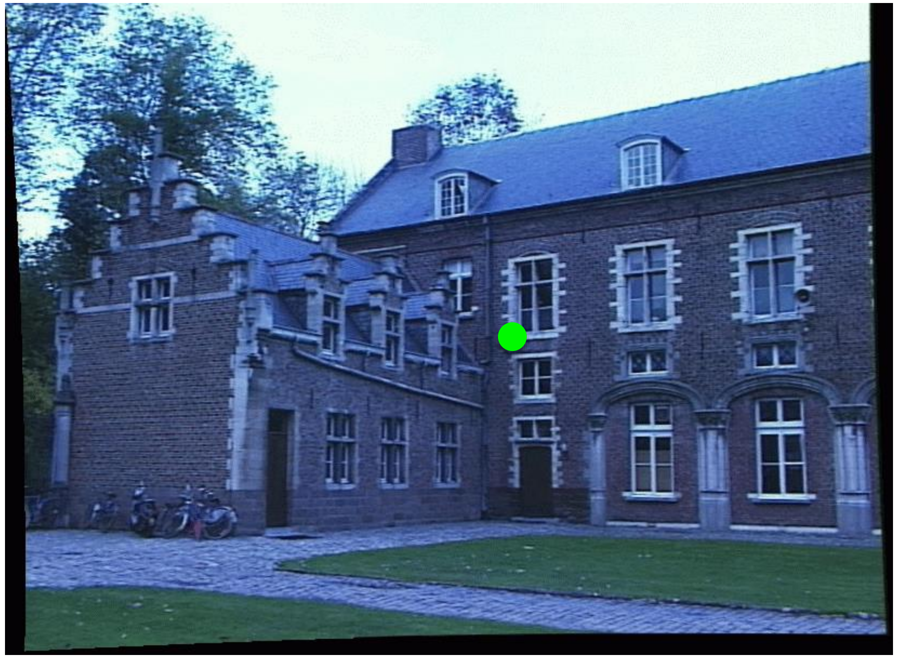
Left image



Right image

1. Select point in one image
2. Form epipolar line for that point in second image (how?)

How would you reconstruct 3D points?

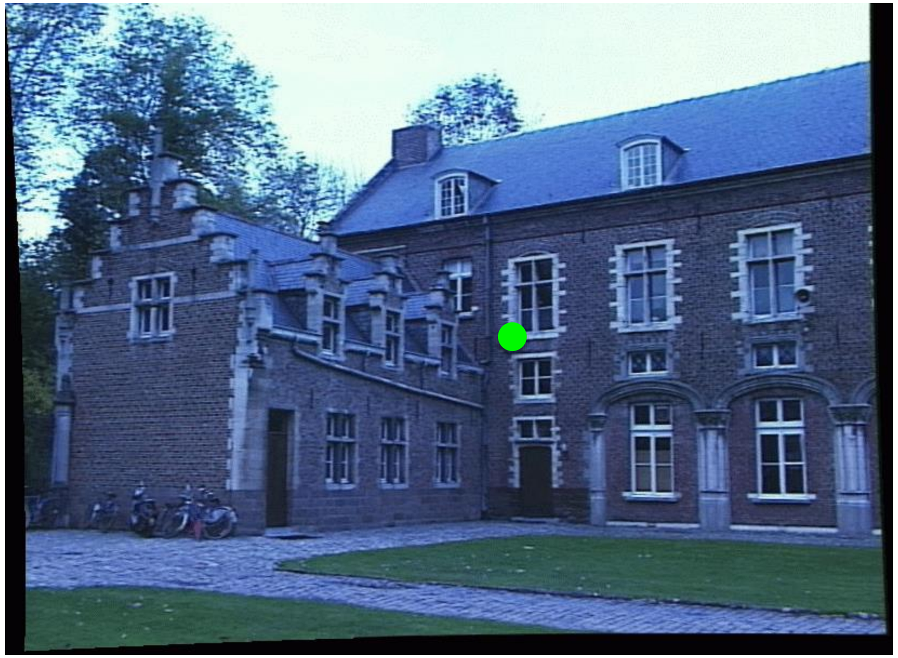


Left image

Right image

1. Select point in one image
2. Form epipolar line for that point in second image (how?)
3. Find matching point along line (how?)

How would you reconstruct 3D points?

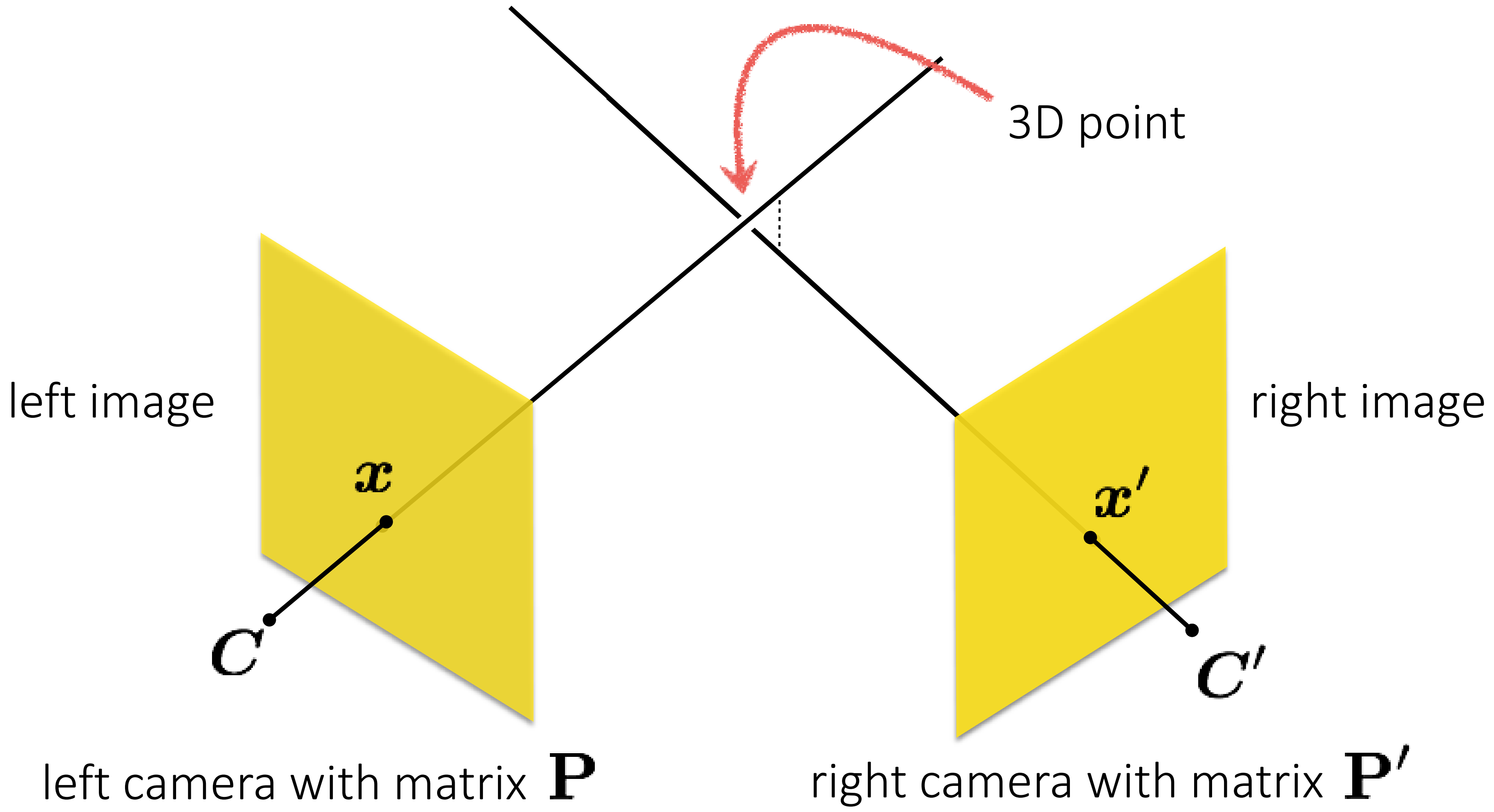


Left image

Right image

1. Select point in one image
2. Form epipolar line for that point in second image (how?)
3. Find matching point along line (how?)
4. Perform triangulation (how?)

Triangulation



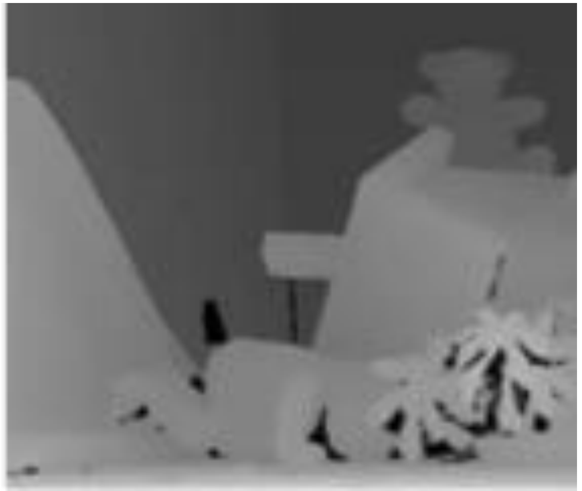
Stereo rectification

What's different between these two images?

Objects that are close move more or less?

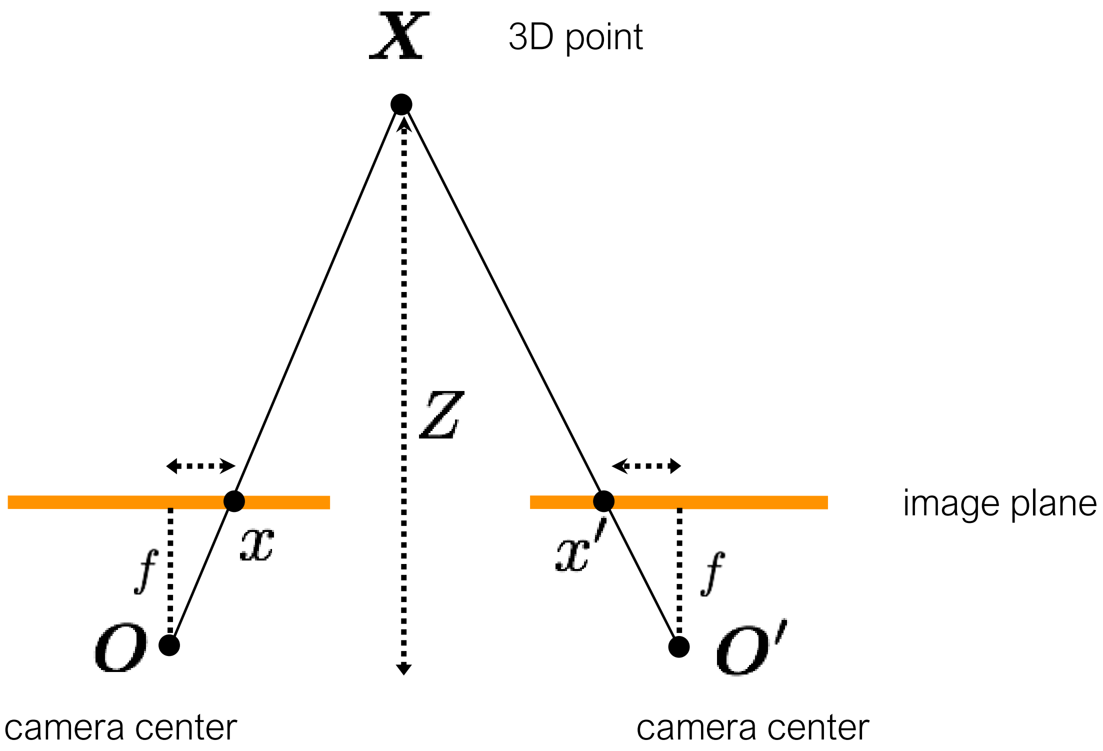
The amount of horizontal movement is inversely proportional to …

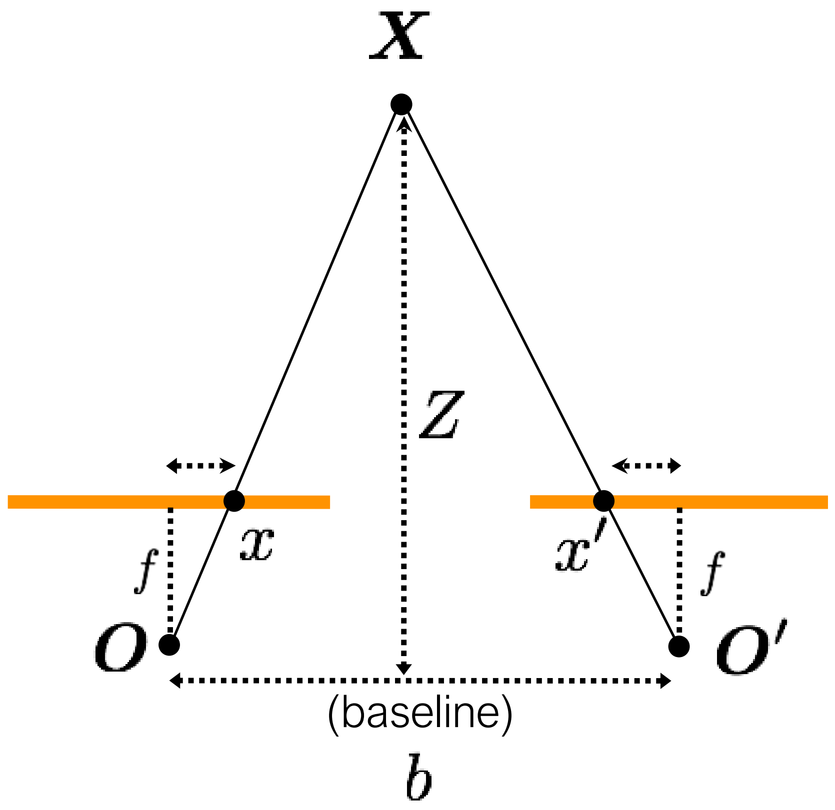


The amount of horizontal movement is inversely proportional to …

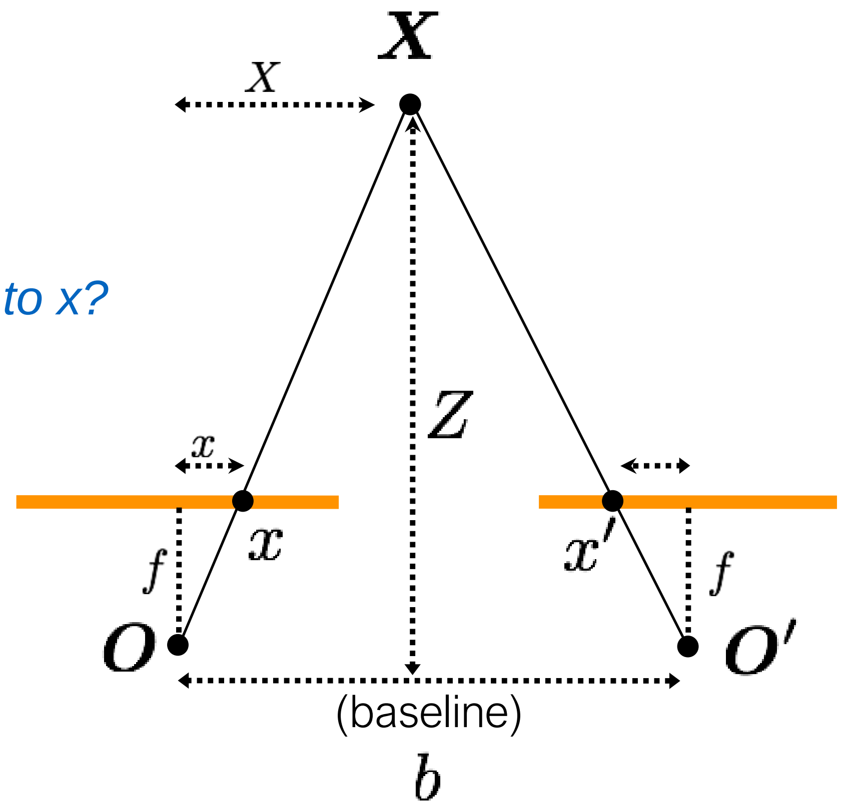
… the distance from the camera.

More formally…

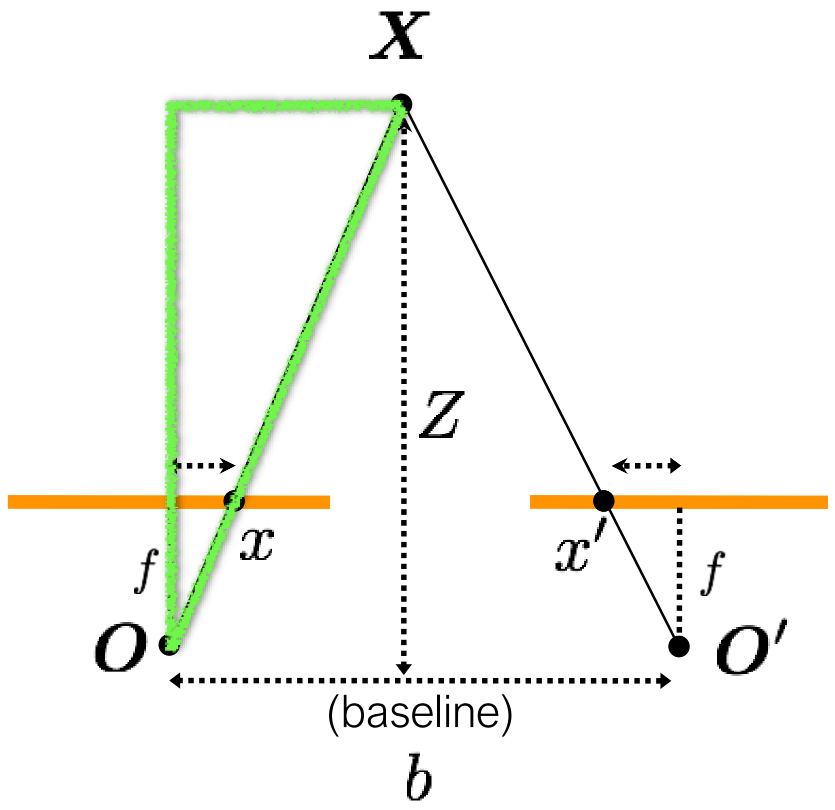

X
3D point
Z
x
x′
f
f
image plane
O
O′
camera center
camera center


X
Z
x
x′
f
f
O
O′
(baseline)
b

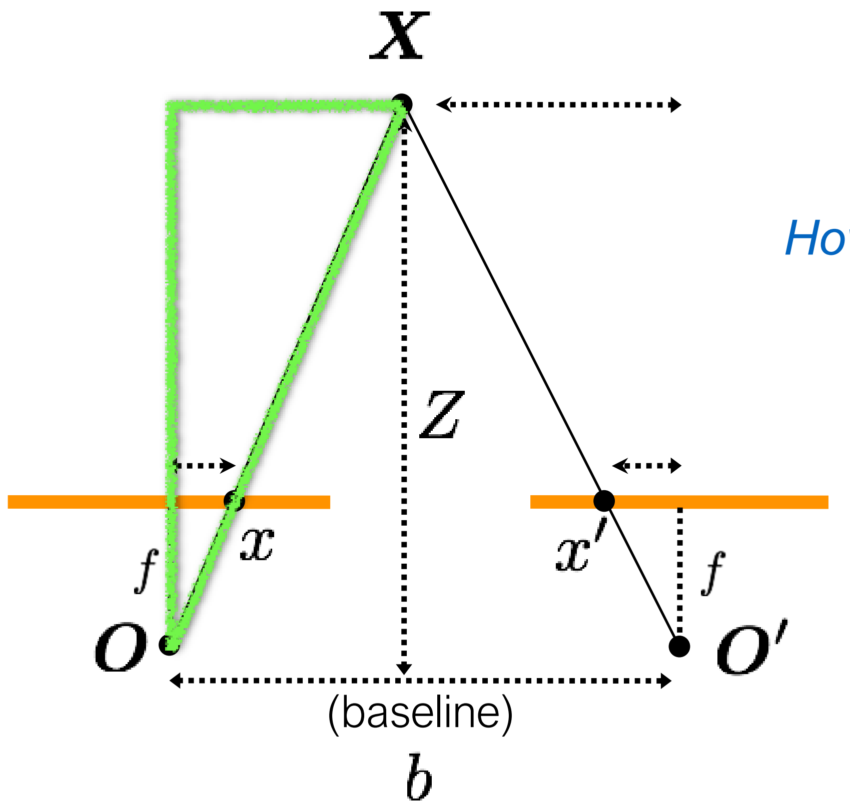
How is X related to x?



$$\frac{X}{Z} = \frac{x}{f}$$

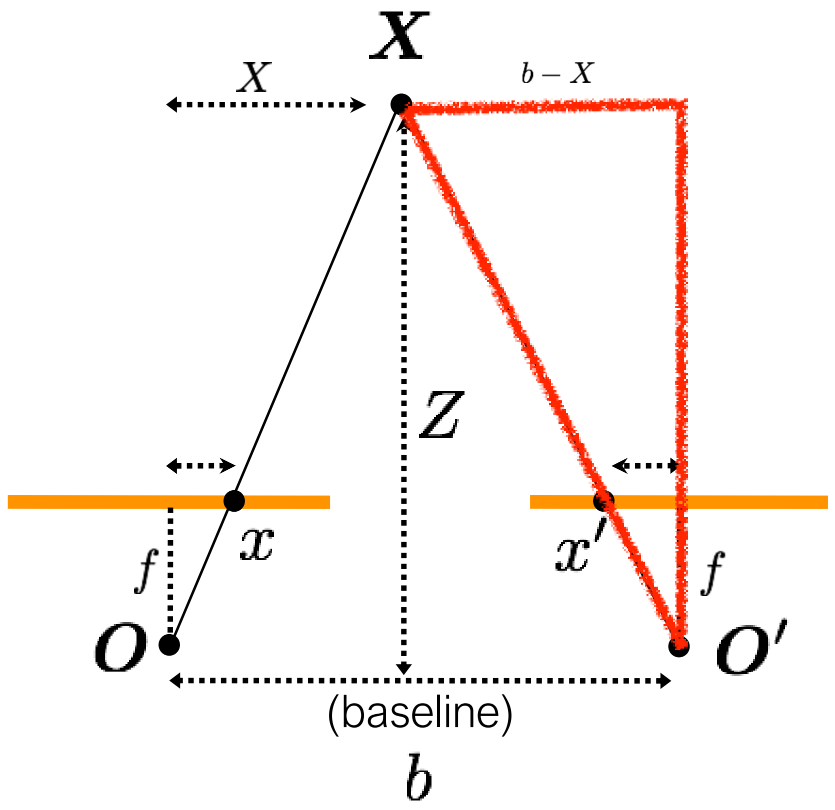


$$\frac{X}{Z} = \frac{x}{f}$$



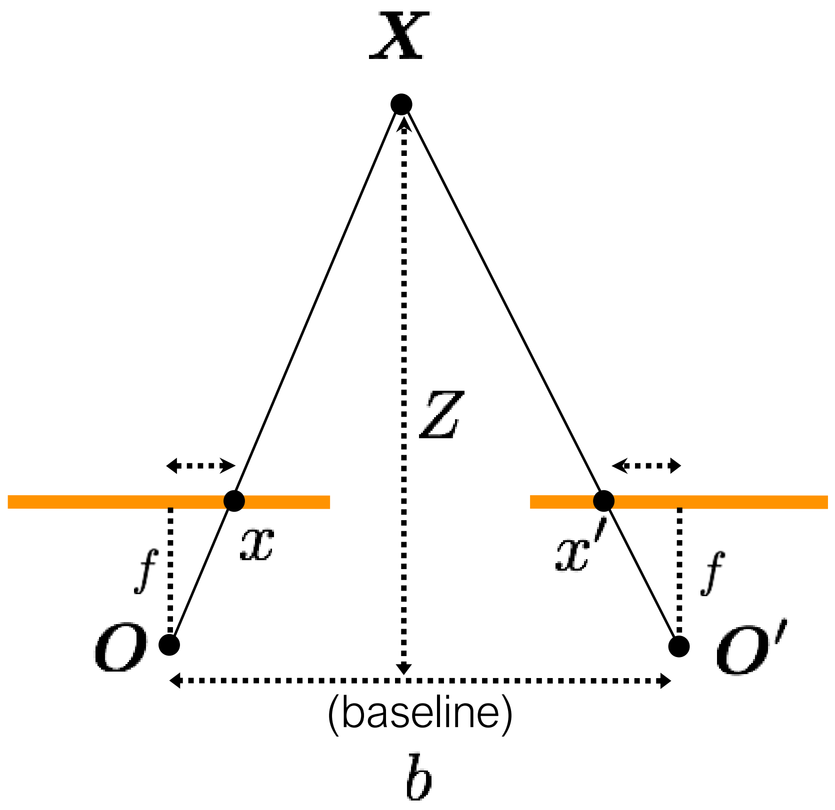
How is X related to x'?

$$\frac{X}{Z} = \frac{x}{f}$$



$$\frac{b - X}{Z} = \frac{x'}{f}$$

$$\frac{X}{Z} = \frac{x}{f}$$



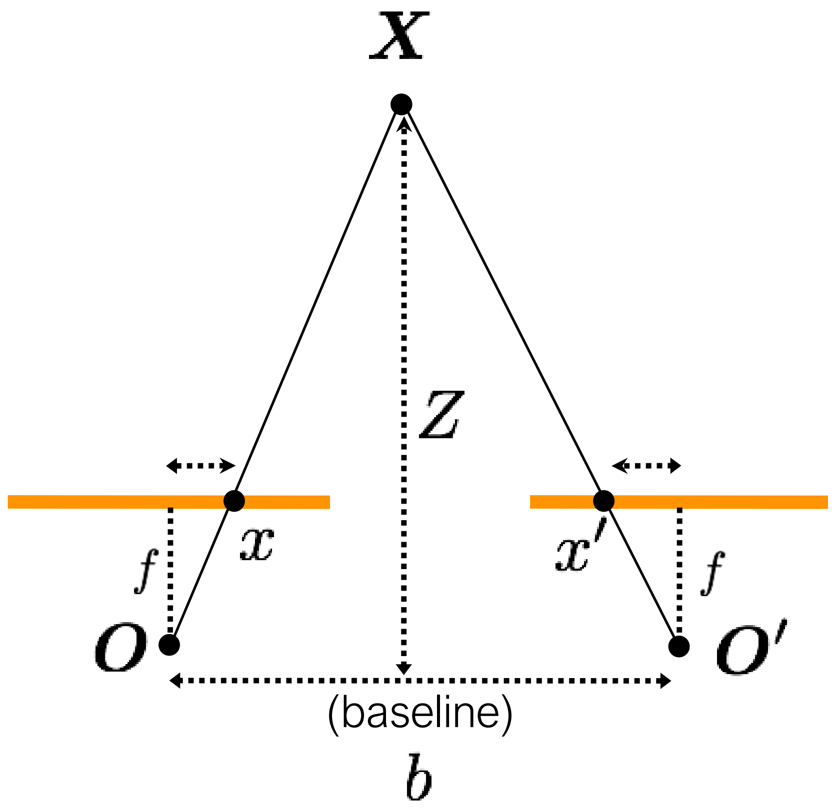
$$\frac{b - X}{Z} = \frac{x'}{f}$$

Disparity

$$d = x - x'$$ (wrt to camera origin of image plane)

$$= \frac{bf}{Z}$$

$$\frac{X}{Z} = \frac{x}{f}$$



$$\frac{b - X}{Z} = \frac{x'}{f}$$

Disparity

$$d = x - x'$$

$$= \frac{bf}{Z}$$

inversely proportional to depth

Real-time stereo sensing

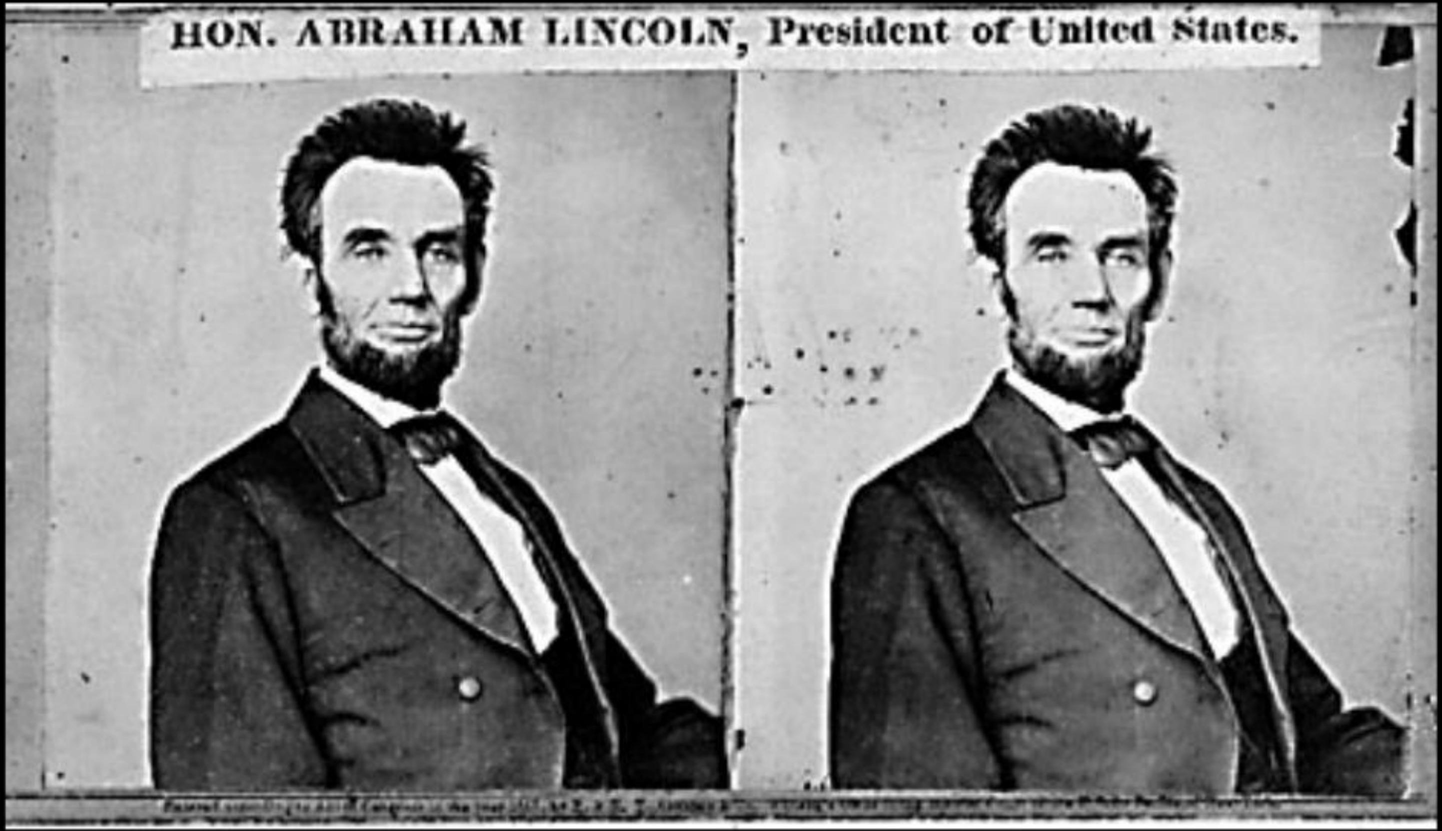
Nomad robot searches for meteorites in Antartica
http://www.frc.ri.cmu.edu/projects/meteorobot/index.html

Subaru
Eyesight system

Pre-collision
braking

What other vision system uses disparity for depth sensing?

Stereoscopes: A 19th Century Pastime

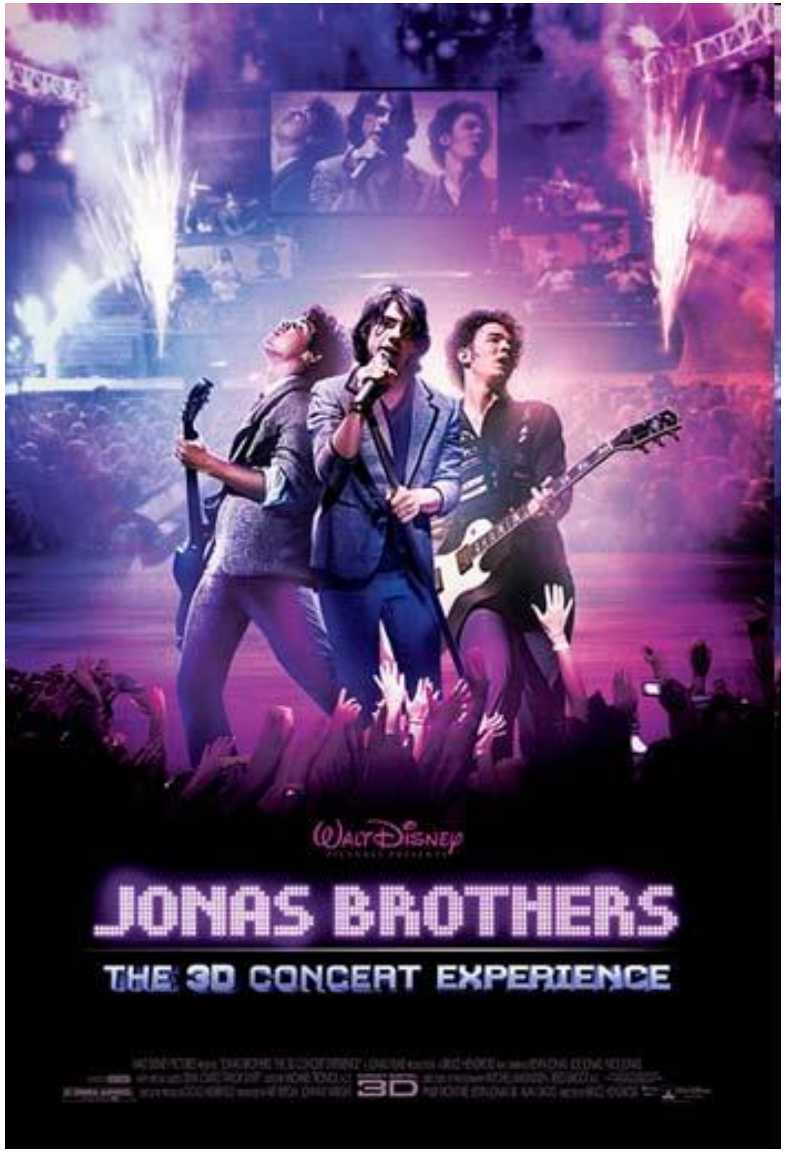

HON. ABRAHAM LINCOLN, President of United States.

Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923

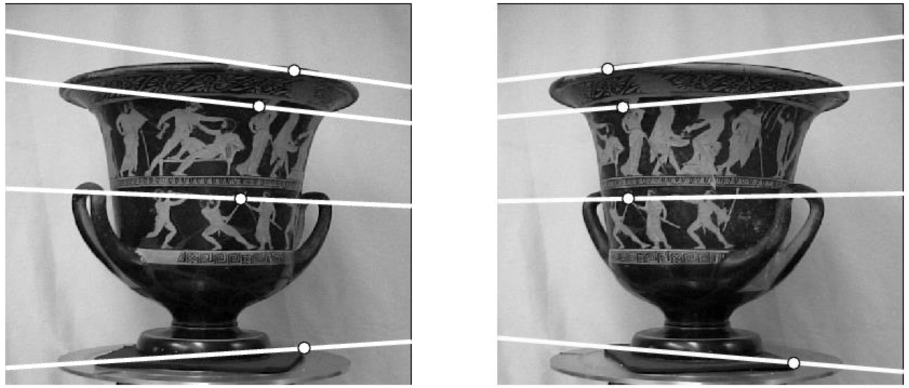
Teesta suspension bridge-Darjeeling, India

Mark Twain at Pool Table", no date, UCR Museum of Photography

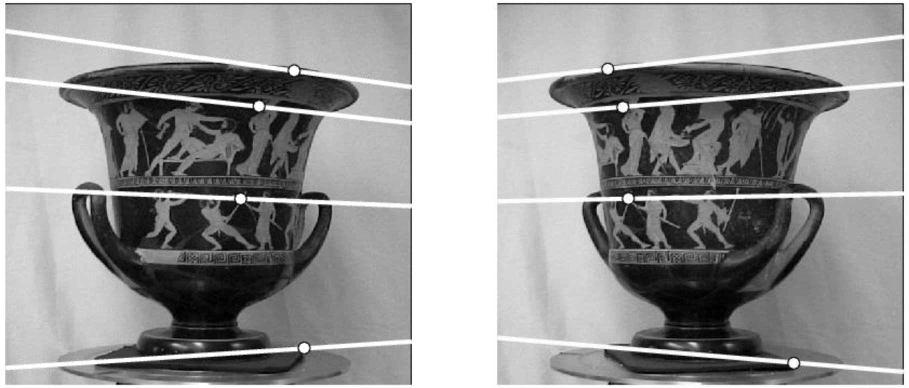
This is how 3D movies work



So can I compute depth using disparity from any two images of the same object?

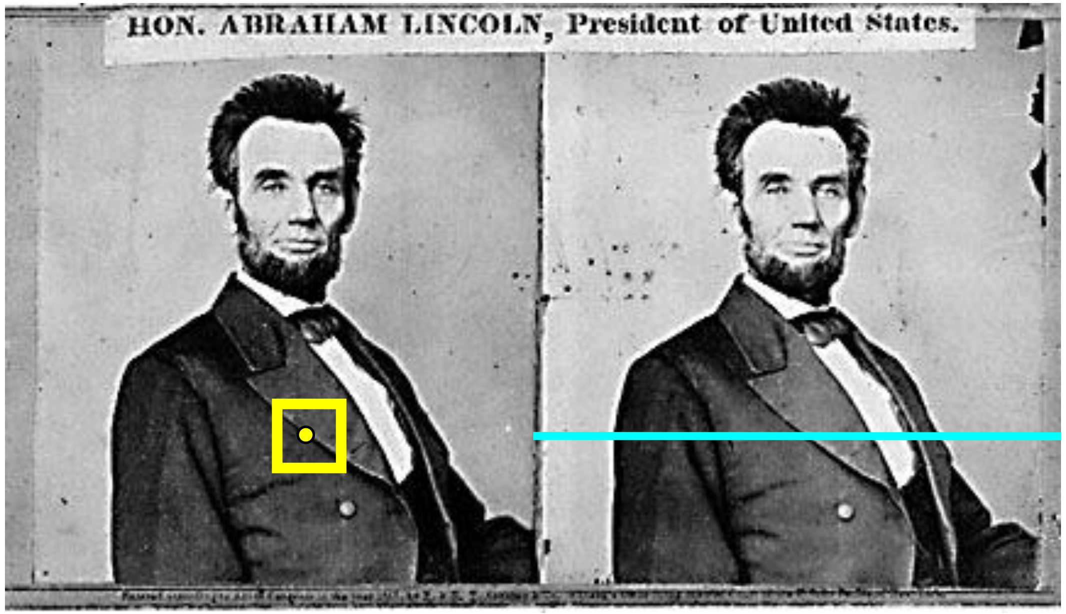


So can I compute depth using disparity from any two images of the same object?

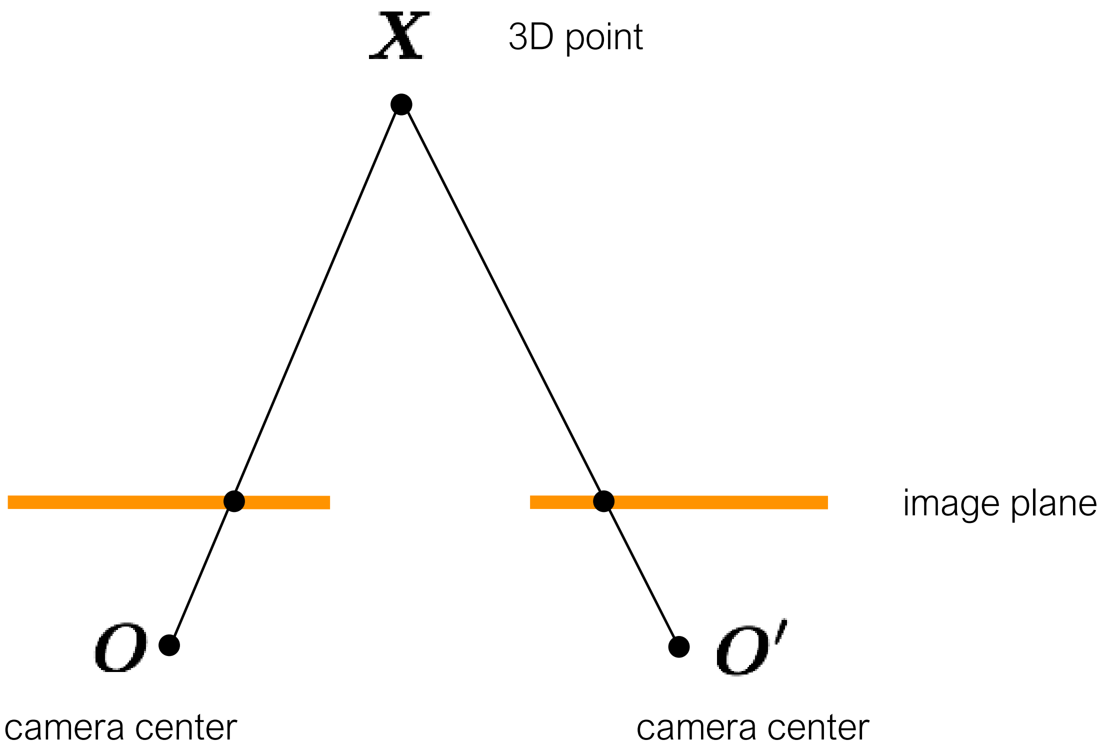


1. Need sufficient baseline

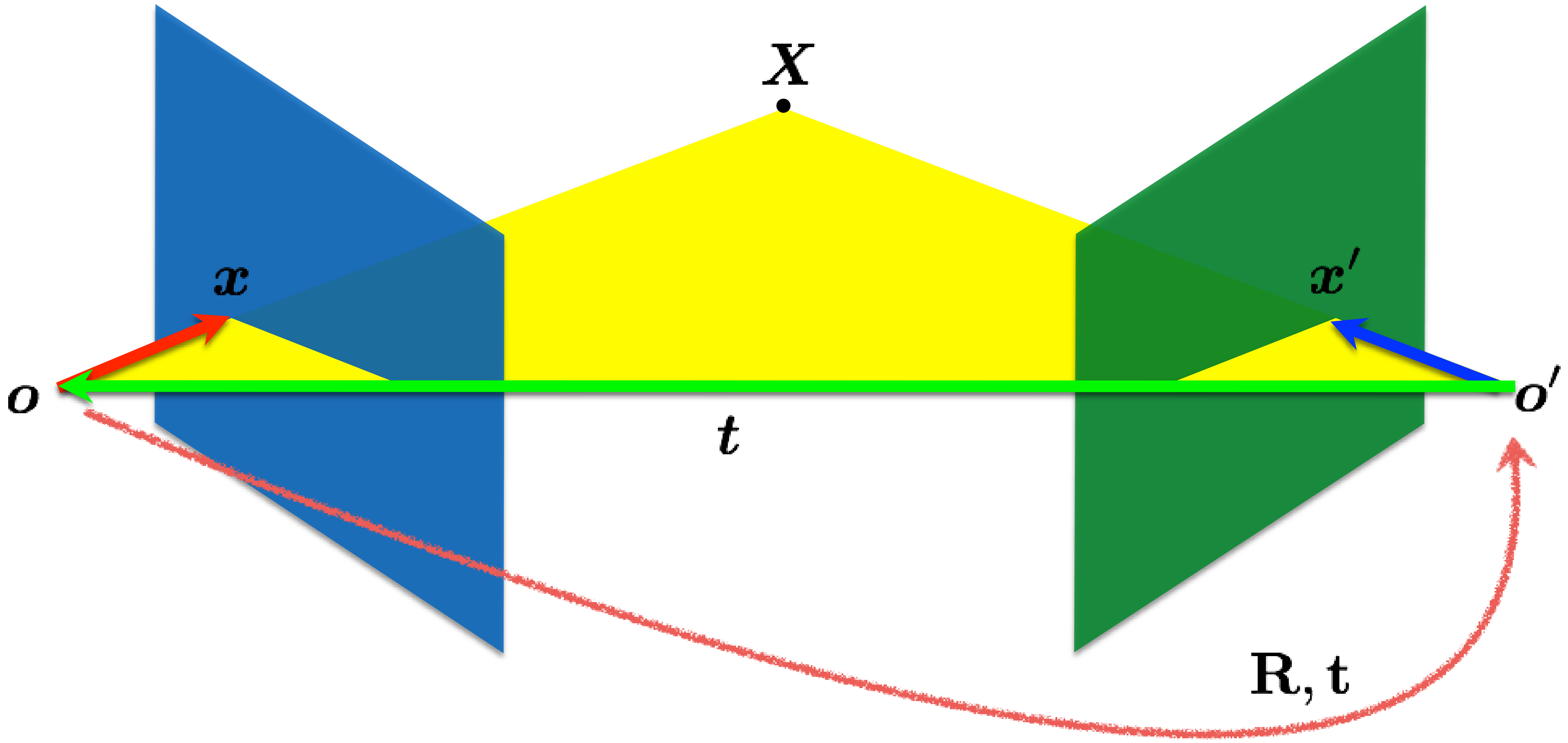
2. Images need to be ‘rectified’ first (make epipolar lines horizontal)



How can you make the epipolar lines horizontal?

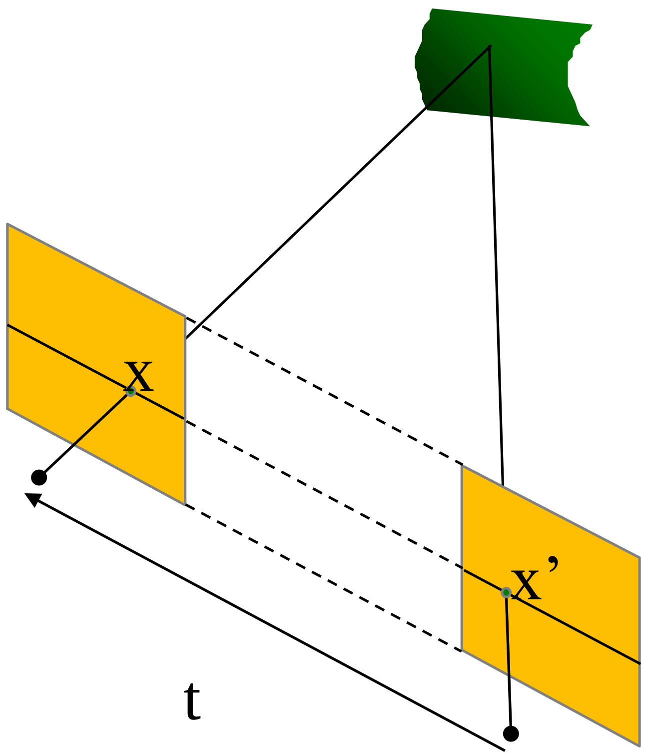


What's special about these two cameras?

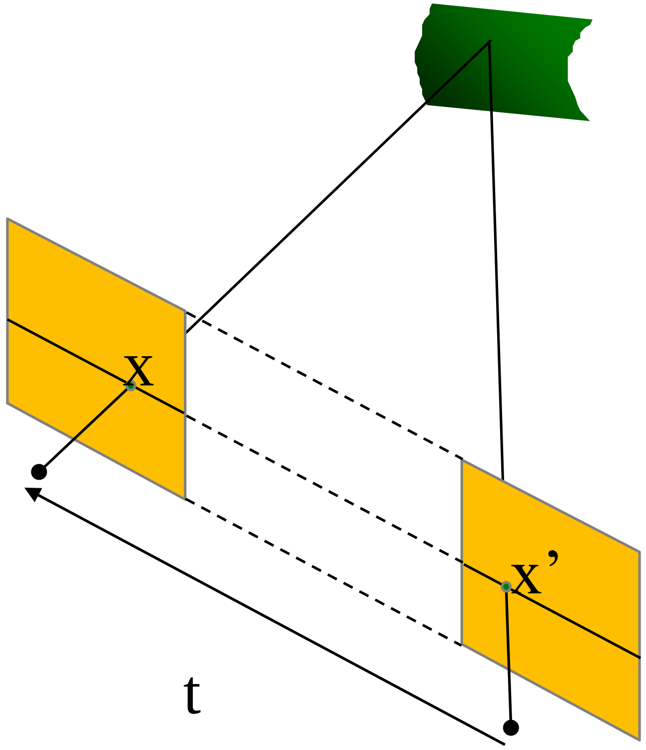


$$\boldsymbol{x}' = \mathbf{R}(\boldsymbol{x} - \boldsymbol{t})$$

When are epipolar lines horizontal?



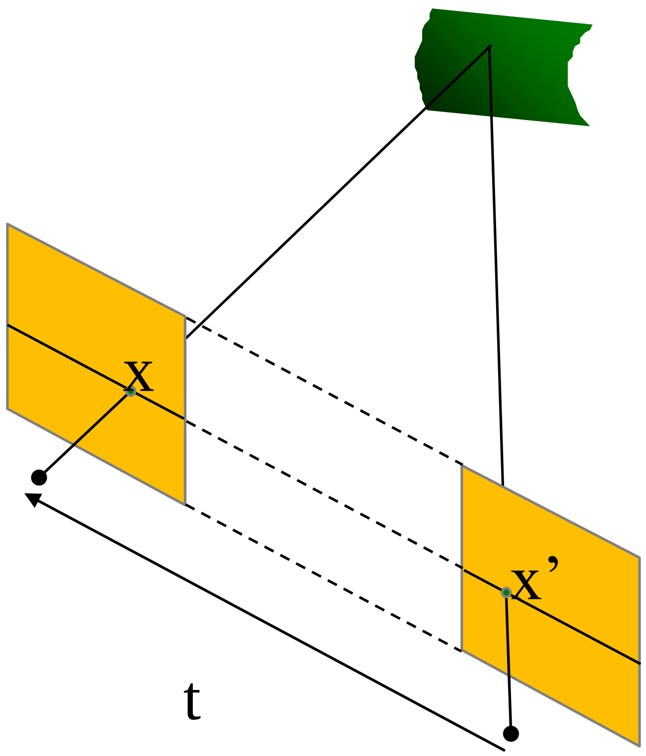
When are epipolar lines horizontal?



When this relationship holds:

$R = I \qquad t = (T, 0, 0)$

When are epipolar lines horizontal?



When this relationship holds:

$$R = I \qquad t = (T, 0, 0)$$

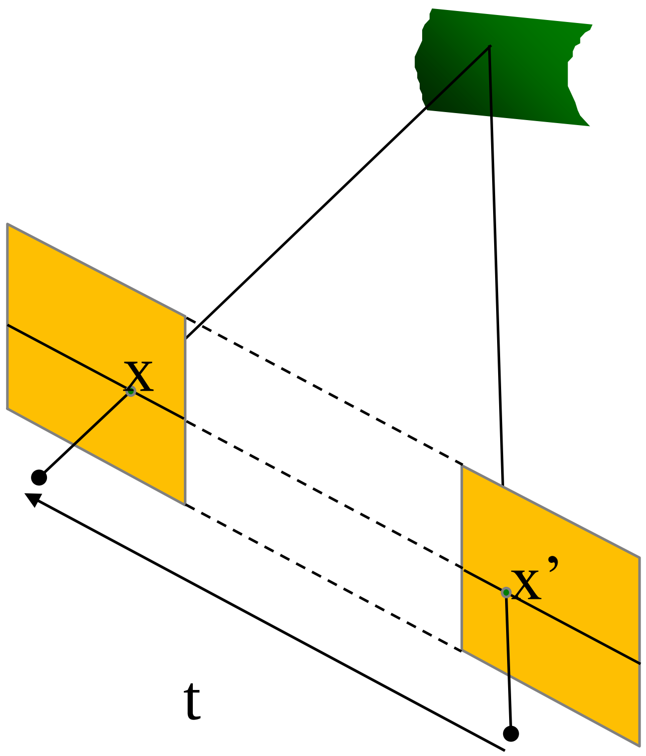
Let's try this out…

$$E = t \times R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$$

This always has to hold:

$$x^T E x' = 0$$

When are epipolar lines horizontal?



When this relationship holds:

$$R = I \qquad t = (T, 0, 0)$$

Let's try this out…

$$E = t \times R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$$

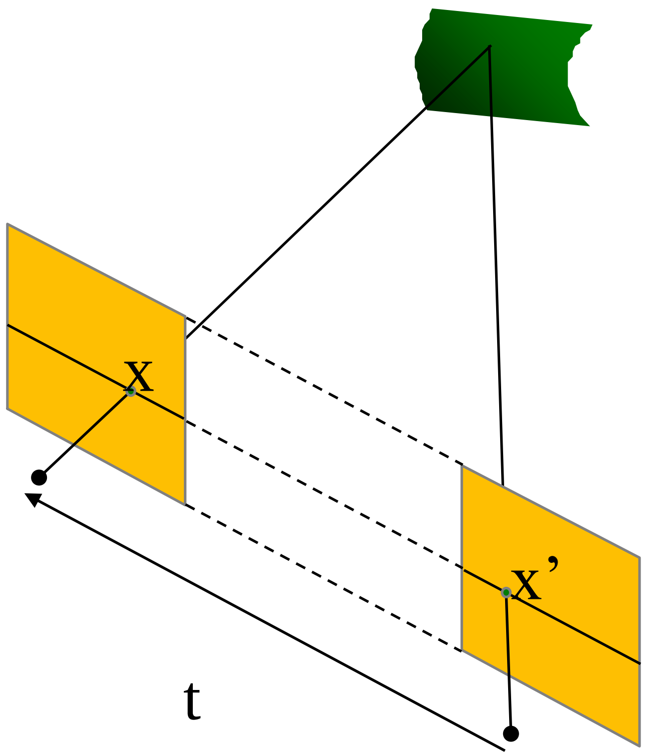
This always has to hold:

$$x^T E x' = 0$$

Write out the constraint

$$\begin{pmatrix} u & v & 1 \end{pmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \qquad \begin{pmatrix} u & v & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -T \\ Tv' \end{pmatrix} = 0$$

When are epipolar lines horizontal?



When this relationship holds:

$$R = I \qquad t = (T, 0, 0)$$

Let's try this out…

$$E = t \times R = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$$

This always has to hold:

$$x^T E x' = 0$$

Write out the constraint

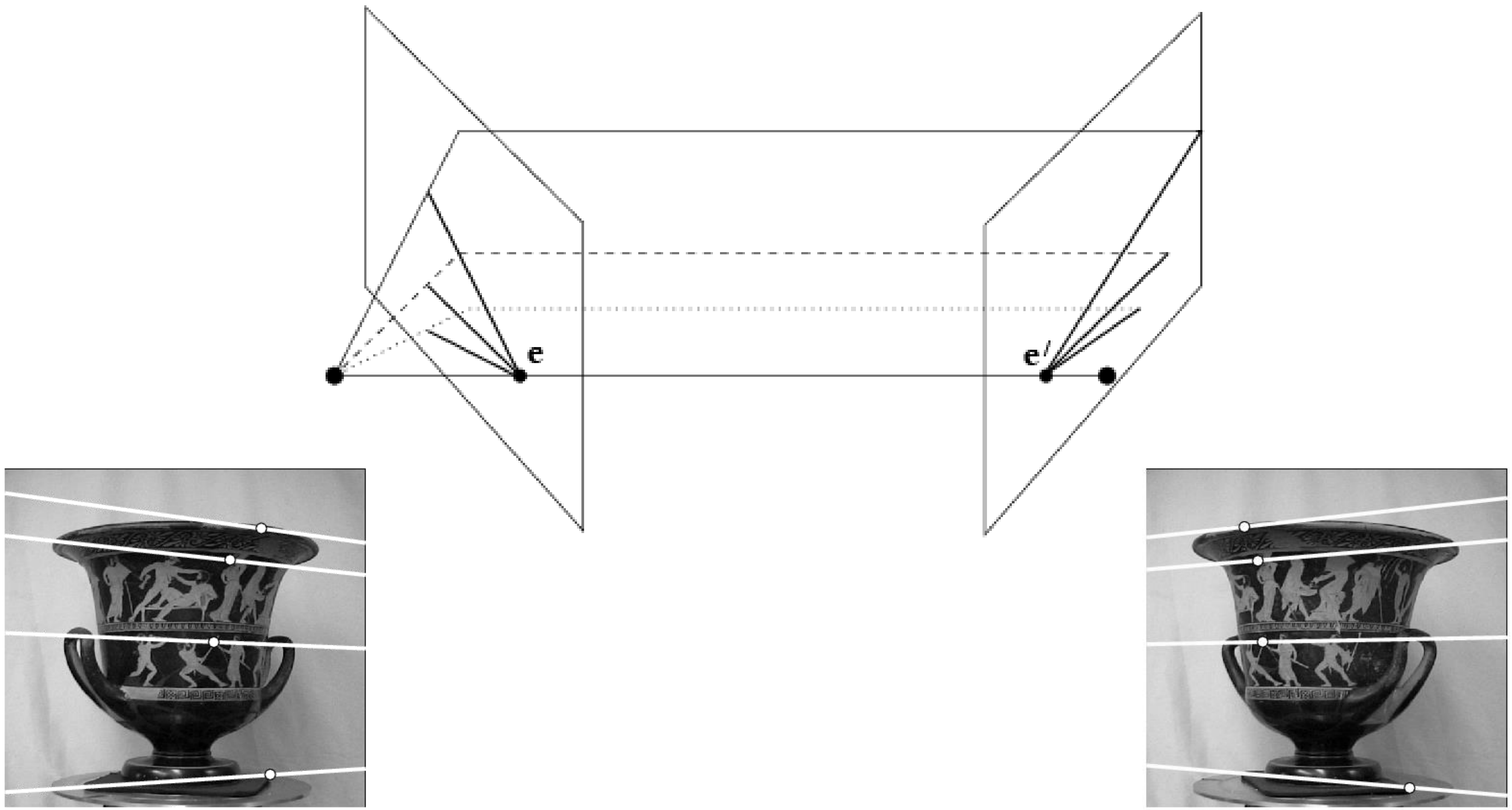
$$\begin{pmatrix} u & v & 1 \end{pmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0$$

$$\begin{pmatrix} u & v & 1 \end{pmatrix} \begin{pmatrix} 0 \\ -T \\ Tv' \end{pmatrix} = 0$$

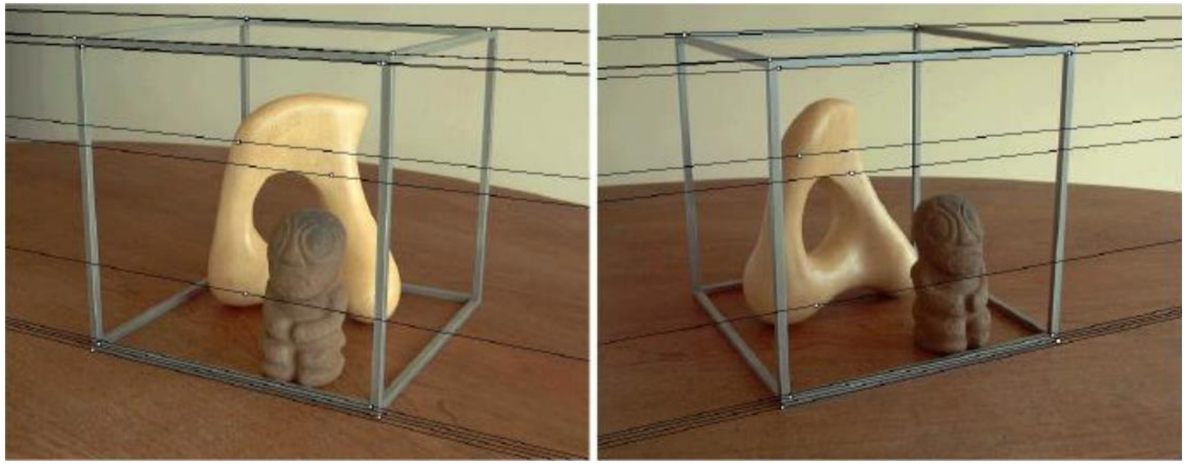
The image of a 3D point will always be on the same horizontal line

$$Tv = Tv'$$

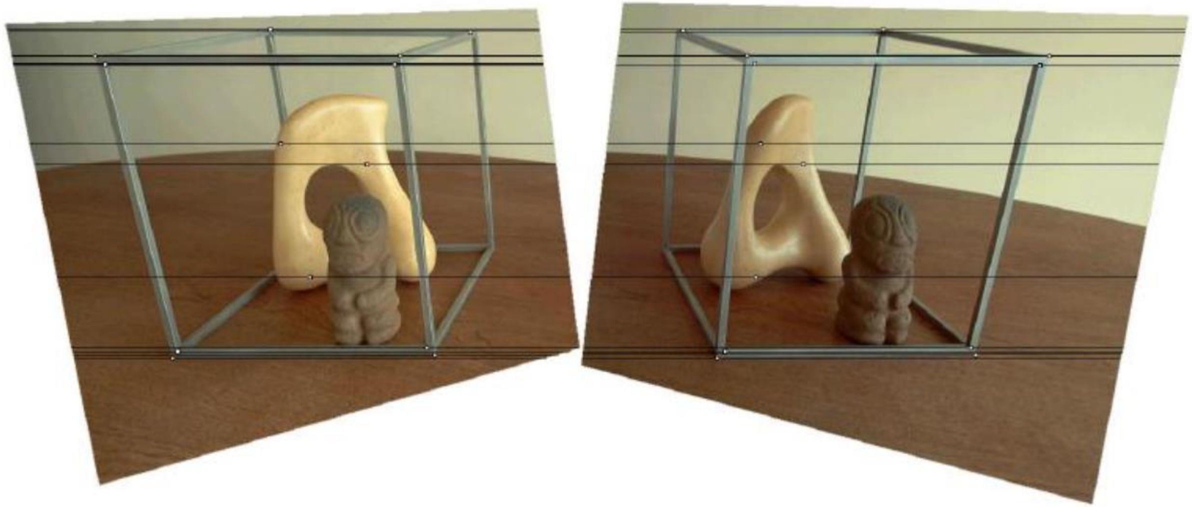
y coordinate is always the same!

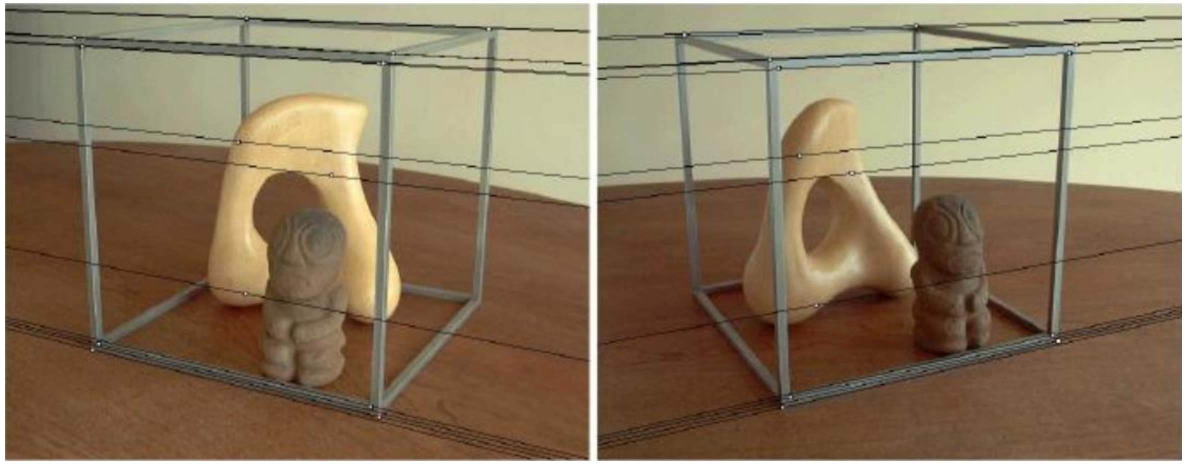


It’s hard to make the image planes exactly parallel

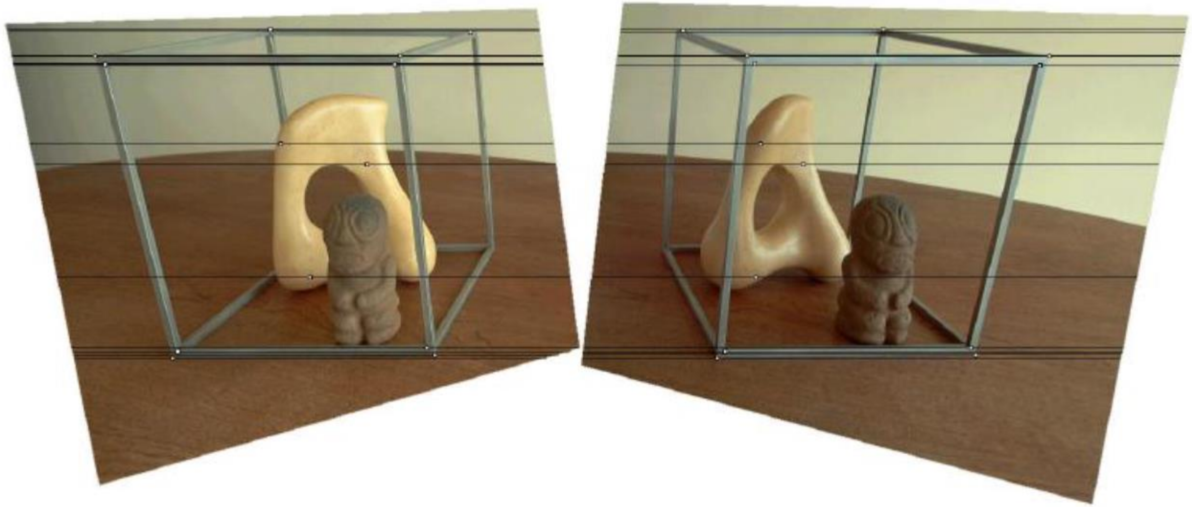


How can you make the epipolar lines horizontal?





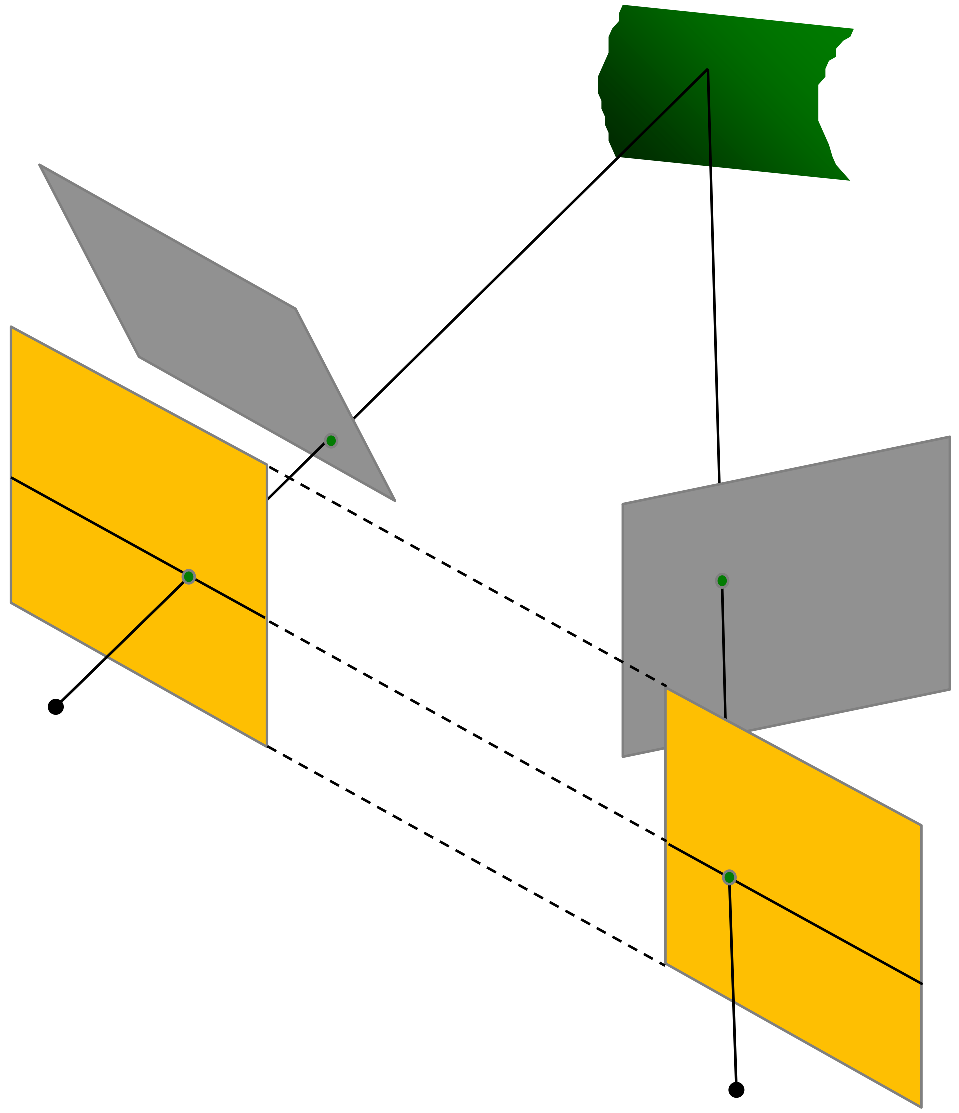
Use stereo rectification



What is stereo rectification?

Reproject image planes onto a common plane parallel to the line between camera centers

Need two homographies (3x3 transform), one for each input image reprojection



C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision.Computer Vision and Pattern Recognition, 1999.

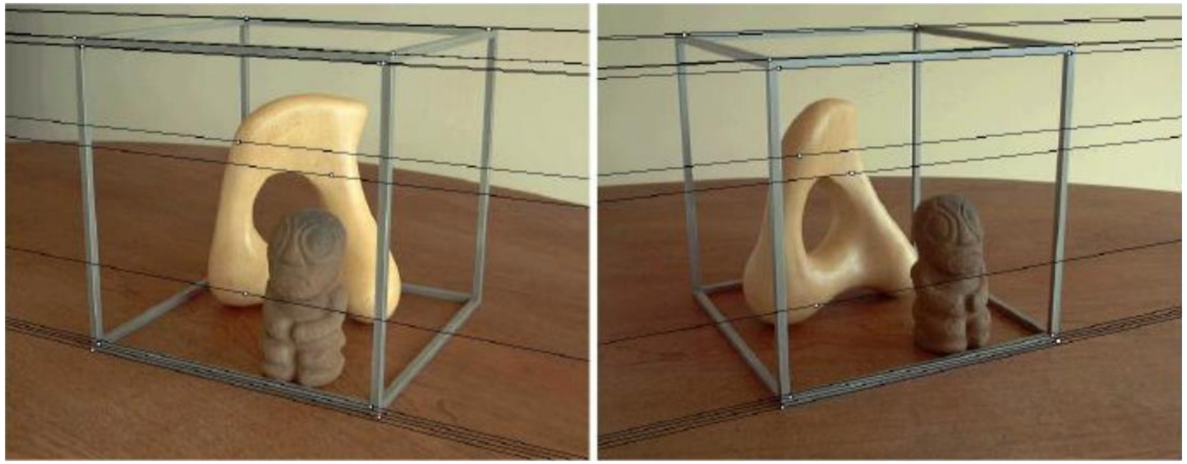
Stereo Rectification

1. **Rotate** the right camera by **R**
   (aligns camera coordinate system orientation only)

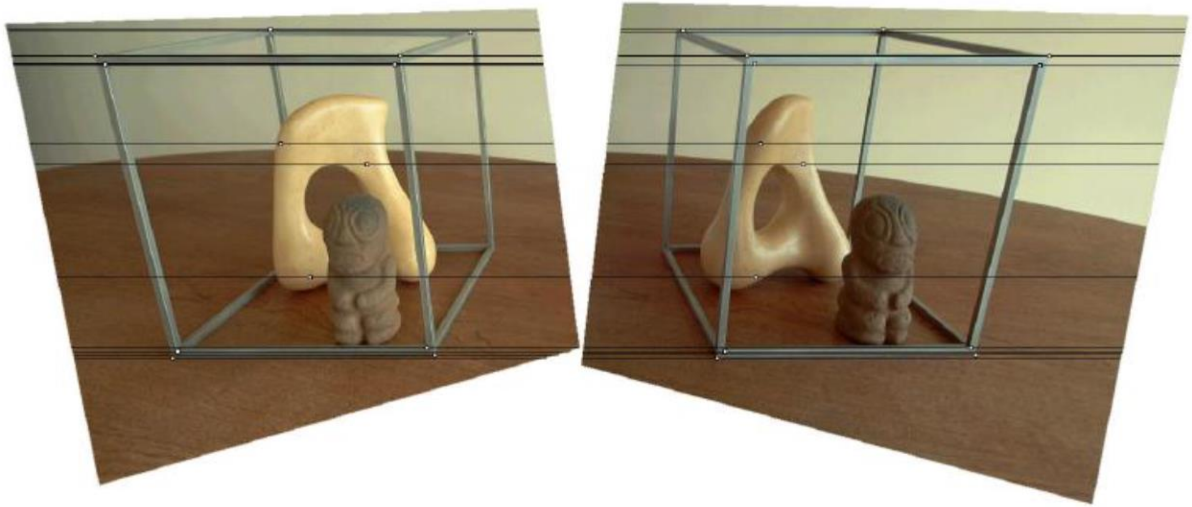
2. Rotate (**rectify**) the left camera so that the epipole is at infinity

3. Rotate (**rectify**) the right camera so that the epipole is at infinity

4. Adjust the **scale**

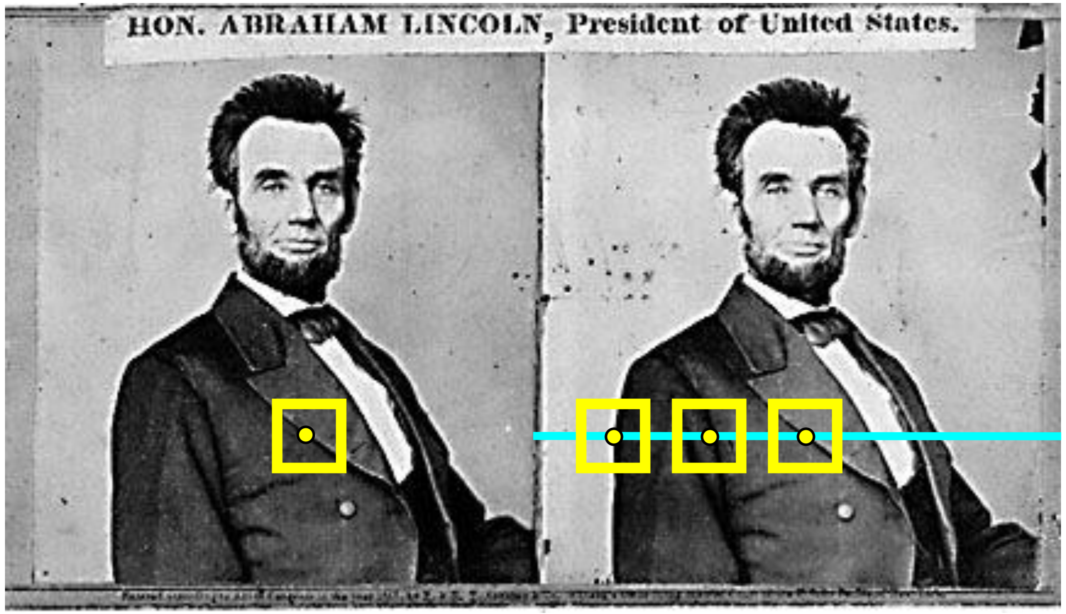


What do we do after rectifying the two image planes?



Stereo matching

Depth Estimation via Stereo Matching



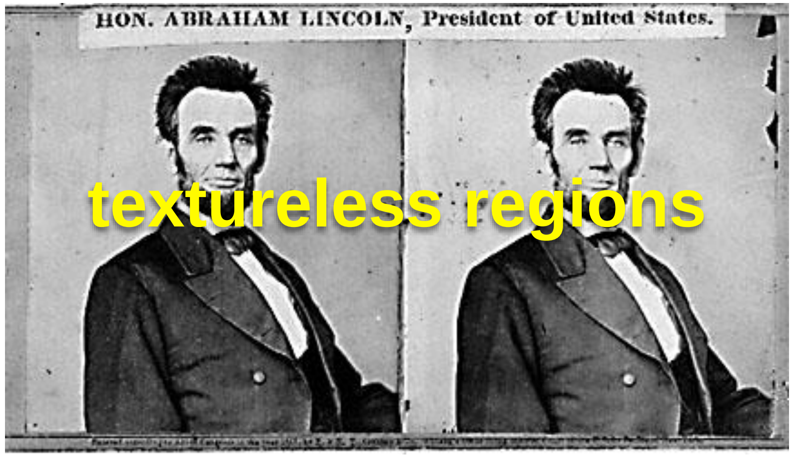
1. Rectify images
   (make epipolar lines horizontal)
2. For each pixel
   a. Find epipolar line
   b. Scan line for best match
   c. Compute depth from disparity

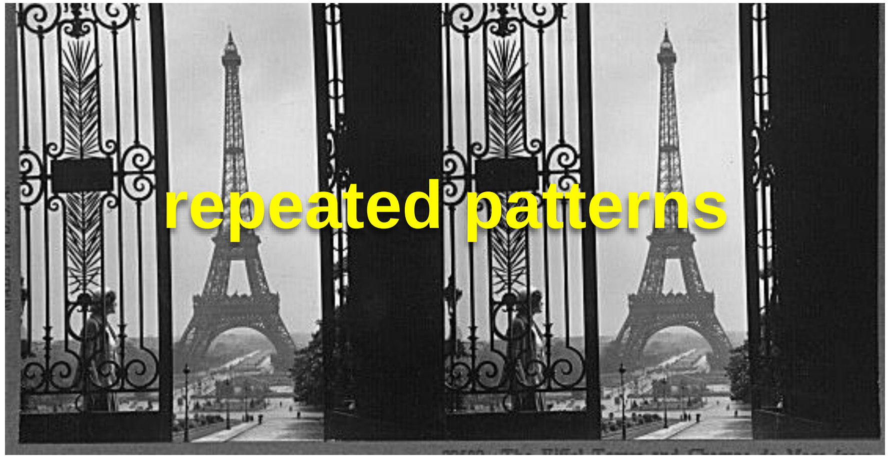
$$Z = \frac{bf}{d}$$

How would you do this?

When are correspondences difficult?

When are correspondences difficult?

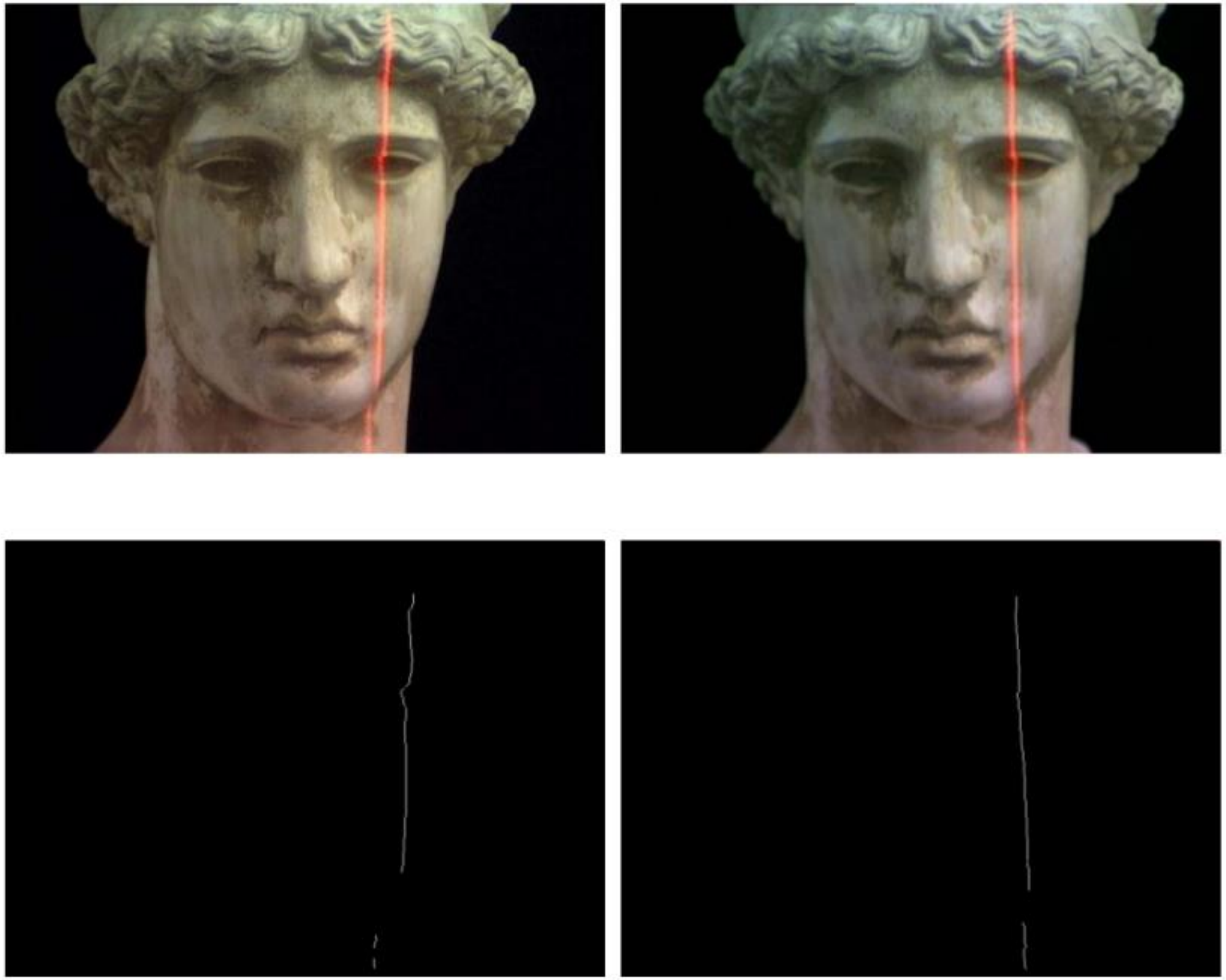




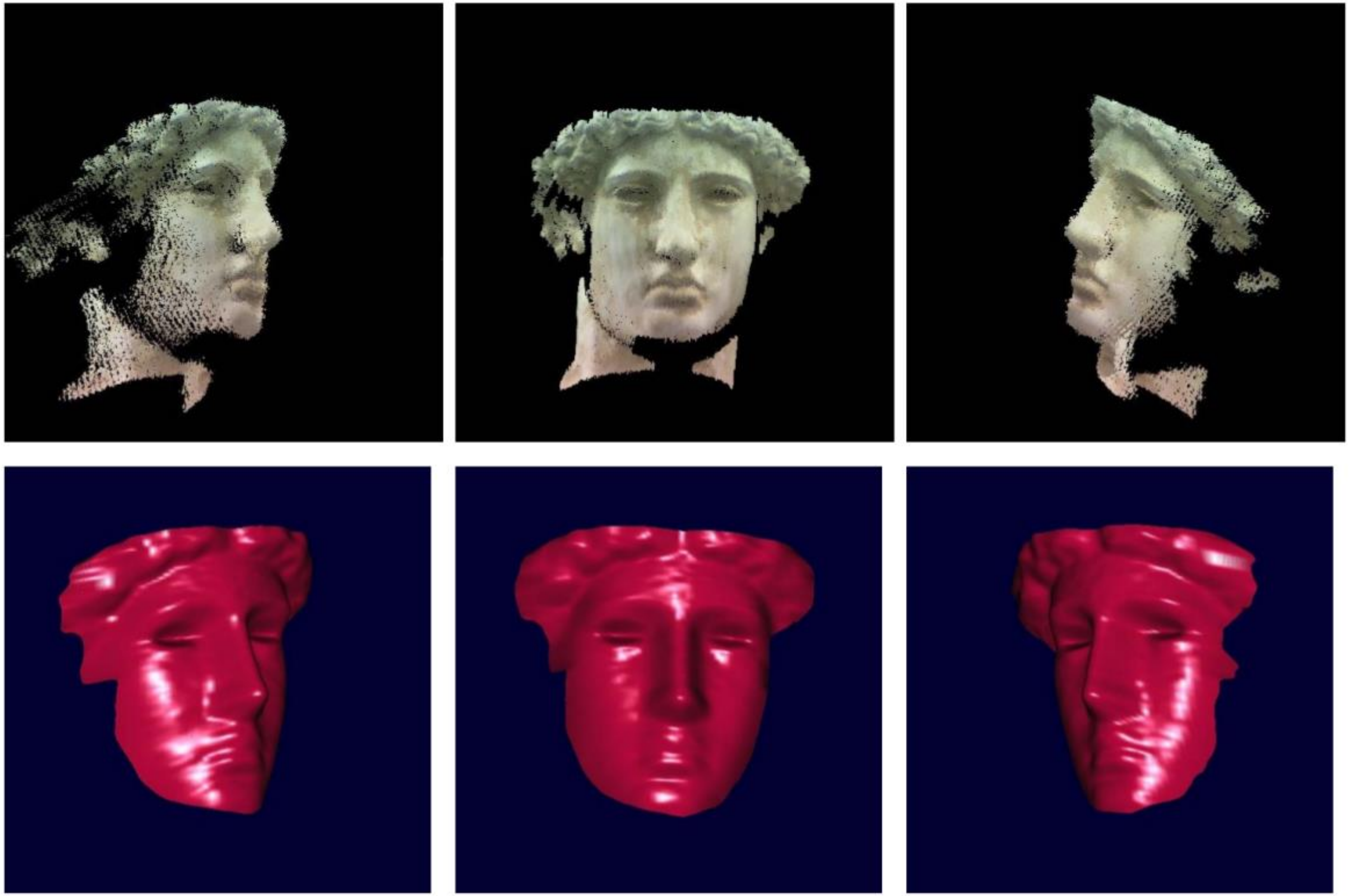
Structured light

Use controlled ("structured") light to make correspondences easier

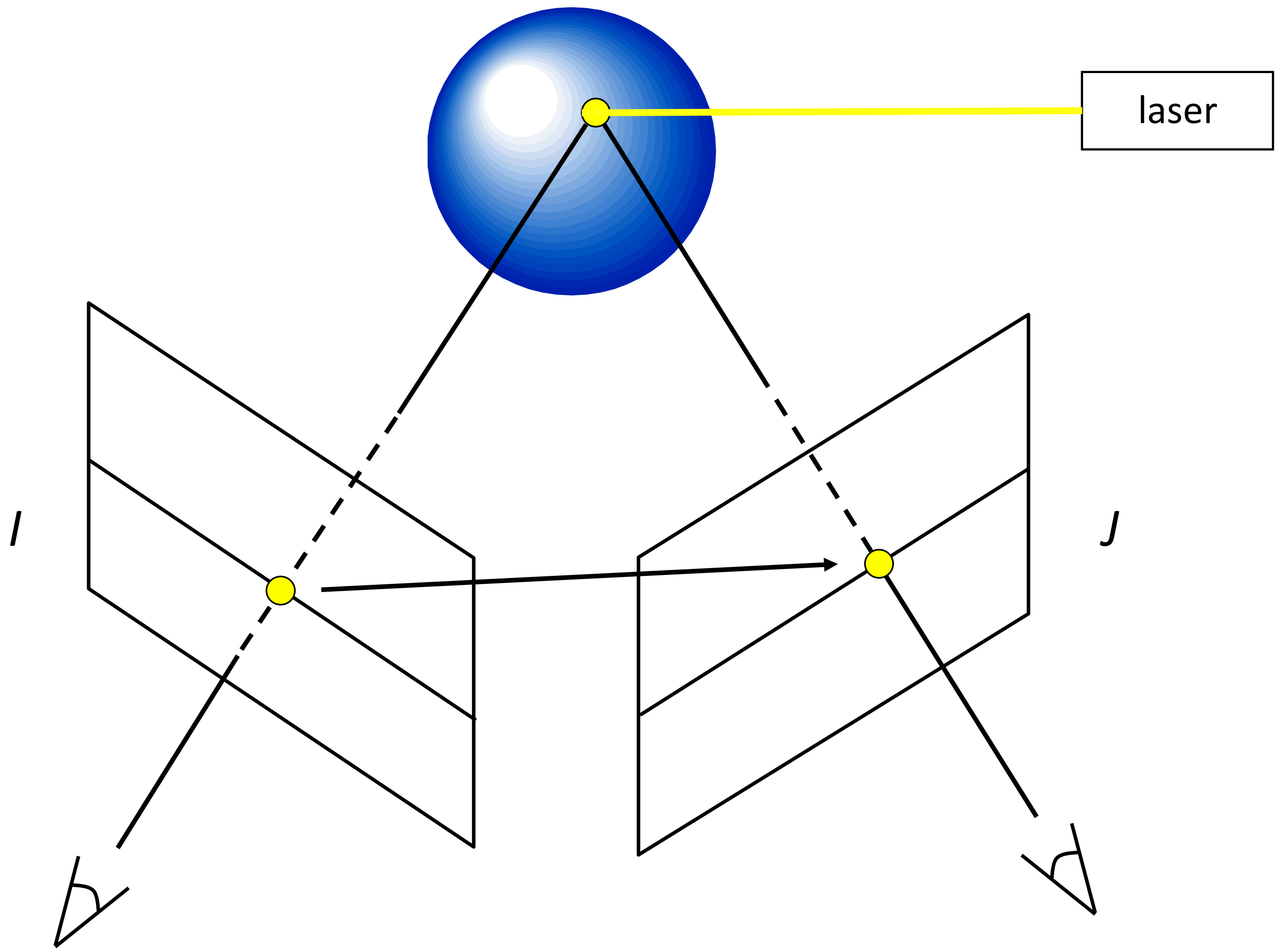
Disparity between laser points on the same scanline in the images determines the 3-D coordinates of the laser point on object



Use controlled ("structured") light to make correspondences easier

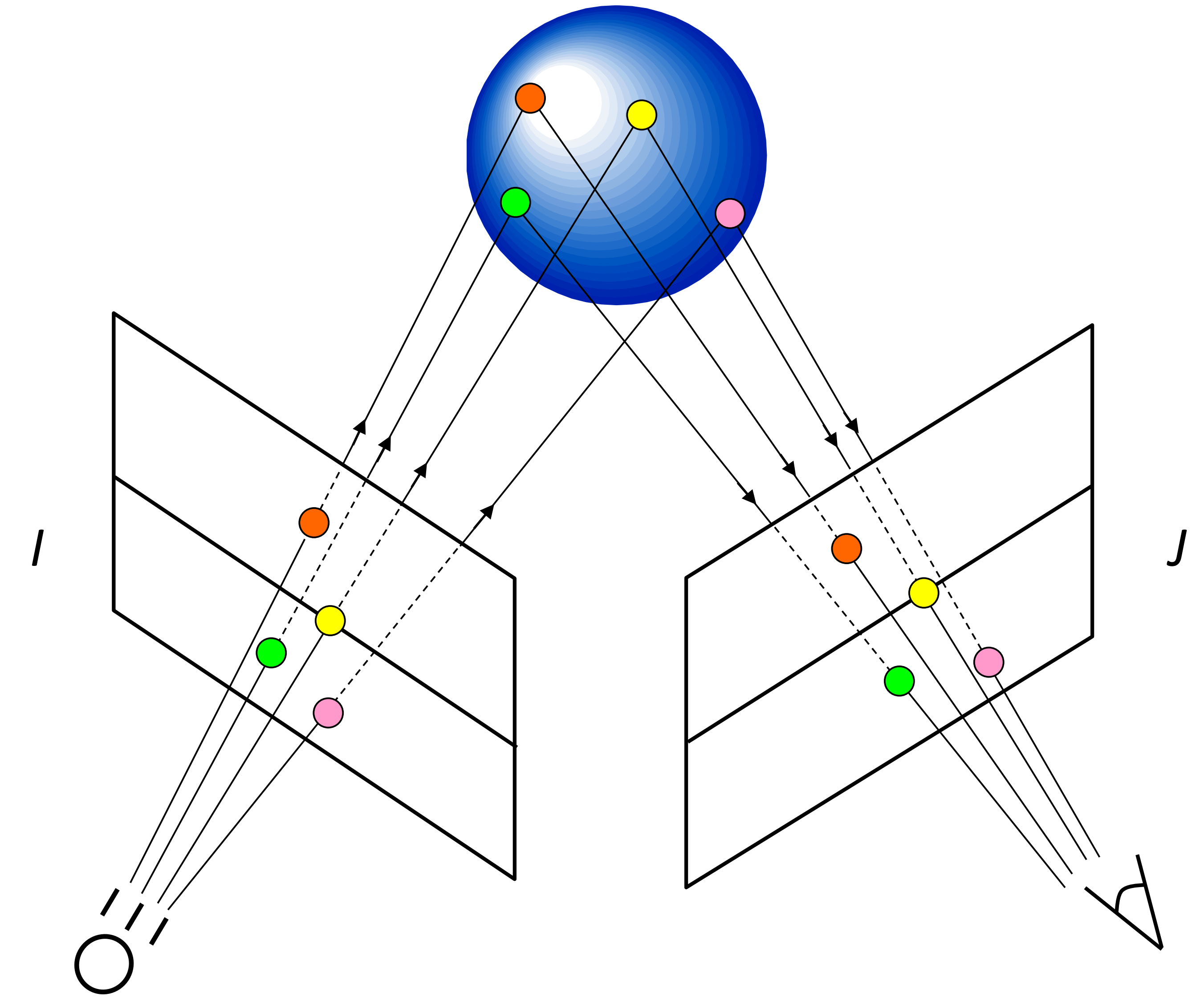


Structured light and two cameras



Structured light and one camera

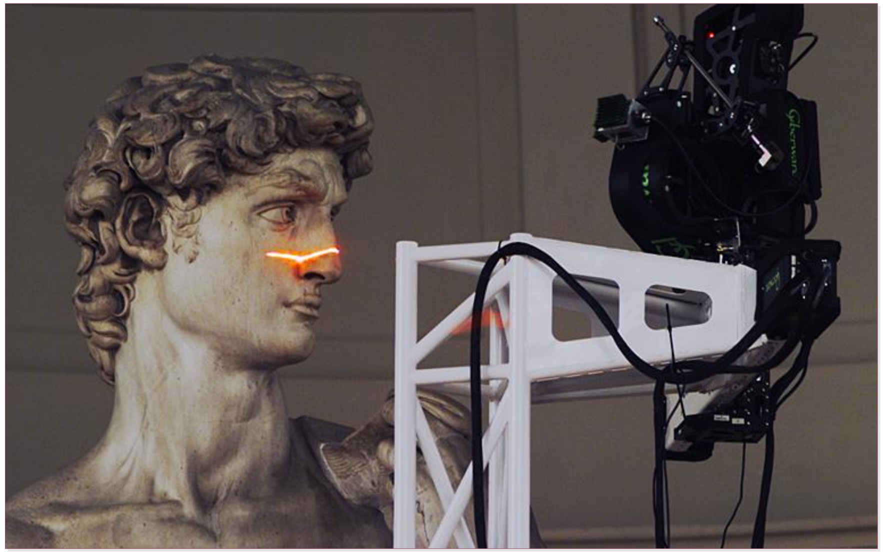
Projector acts like "reverse" camera



Structured Light

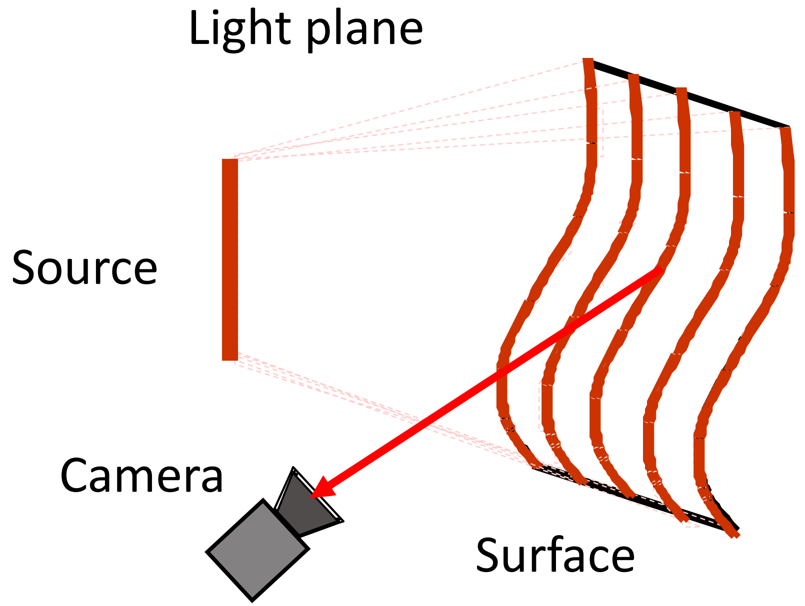
- Any spatio-temporal pattern of light projected on a surface (or volume).
- Cleverly illuminate the scene to extract scene properties (eg., 3D).
- Avoids problems of 3D estimation in scenes with complex texture/BRDFs.
- Very popular in vision and successful in industrial applications (parts assembly, inspection, etc).

3D Scanning using structured light



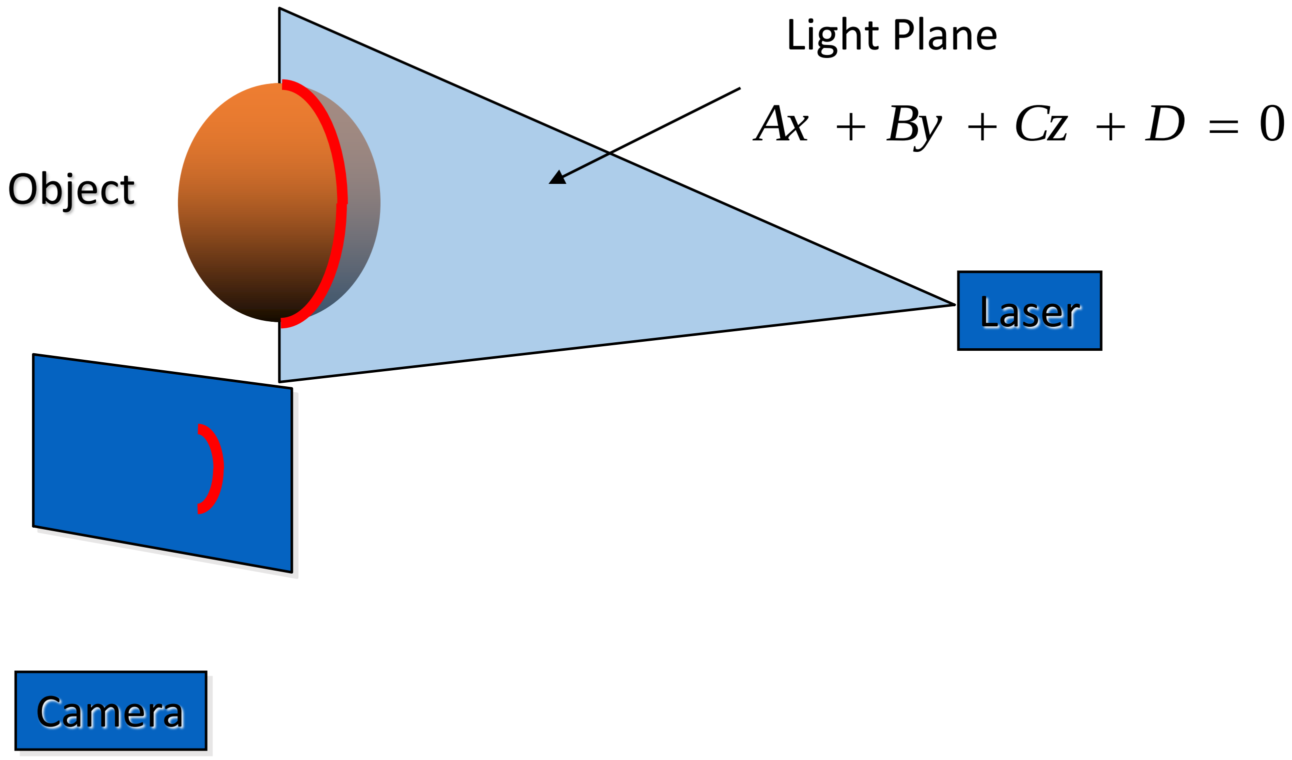
Do we need to illuminate the scene point by point?

Light Stripe Scanning – Single Stripe



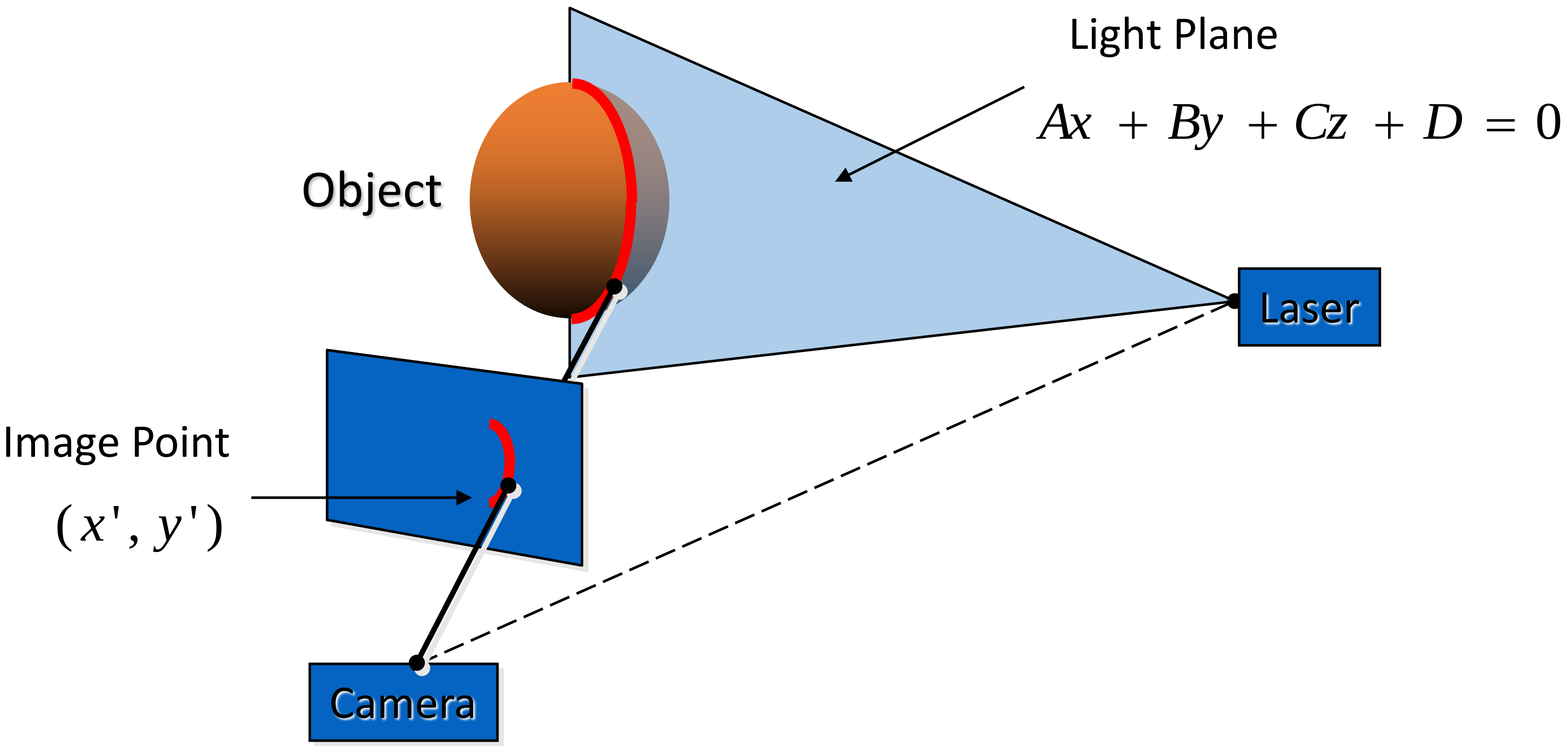
- Faster optical triangulation:
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning
  - Good for high resolution 3D, but still needs many images and takes time

Triangulation



- Project laser stripe onto object

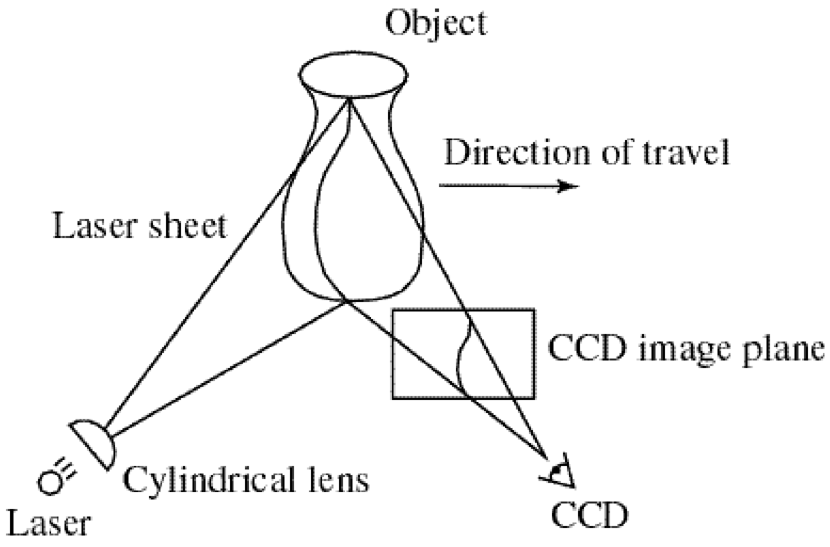
Triangulation

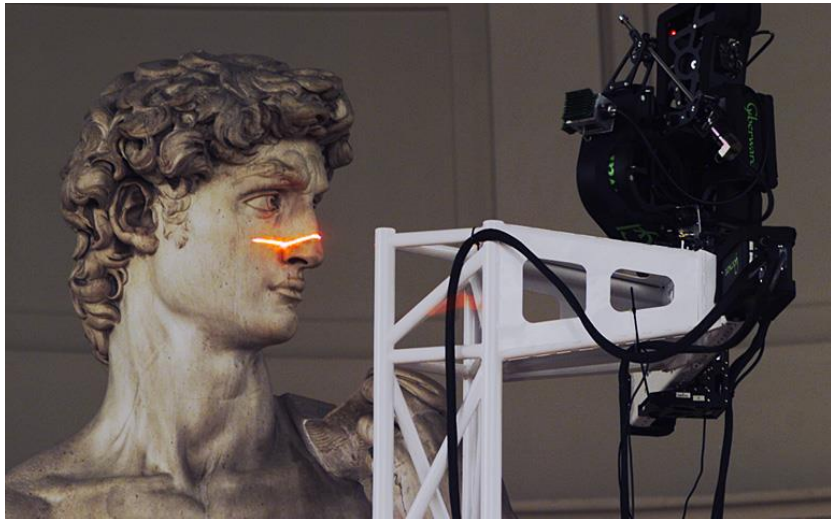


- Depth from ray-plane triangulation:
  - Intersect camera ray with light plane

$$x = x' z / f \qquad z = \frac{-Df}{Ax' + By' + Cf}$$
$$y = y' z / f$$

Example: Laser scanner

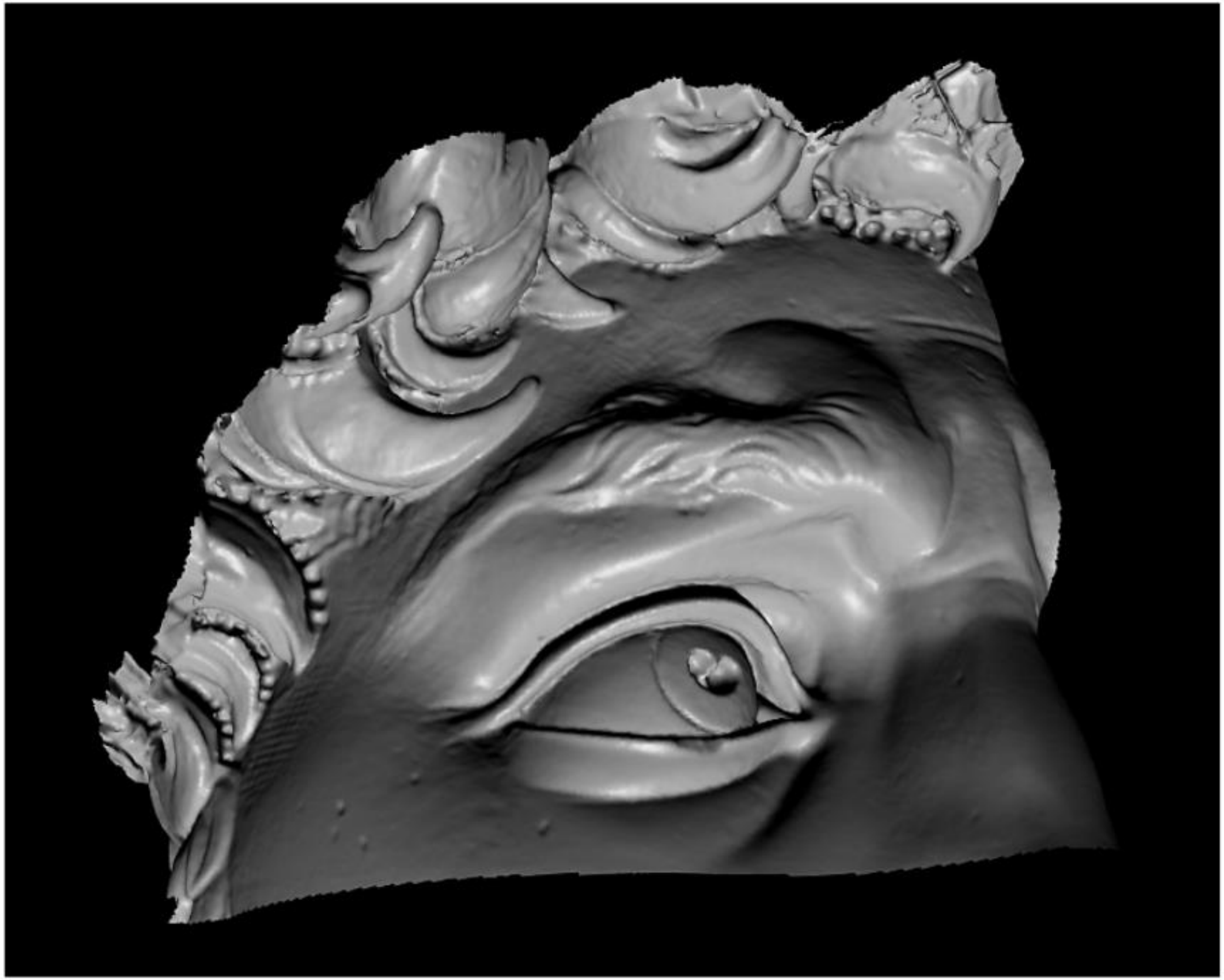




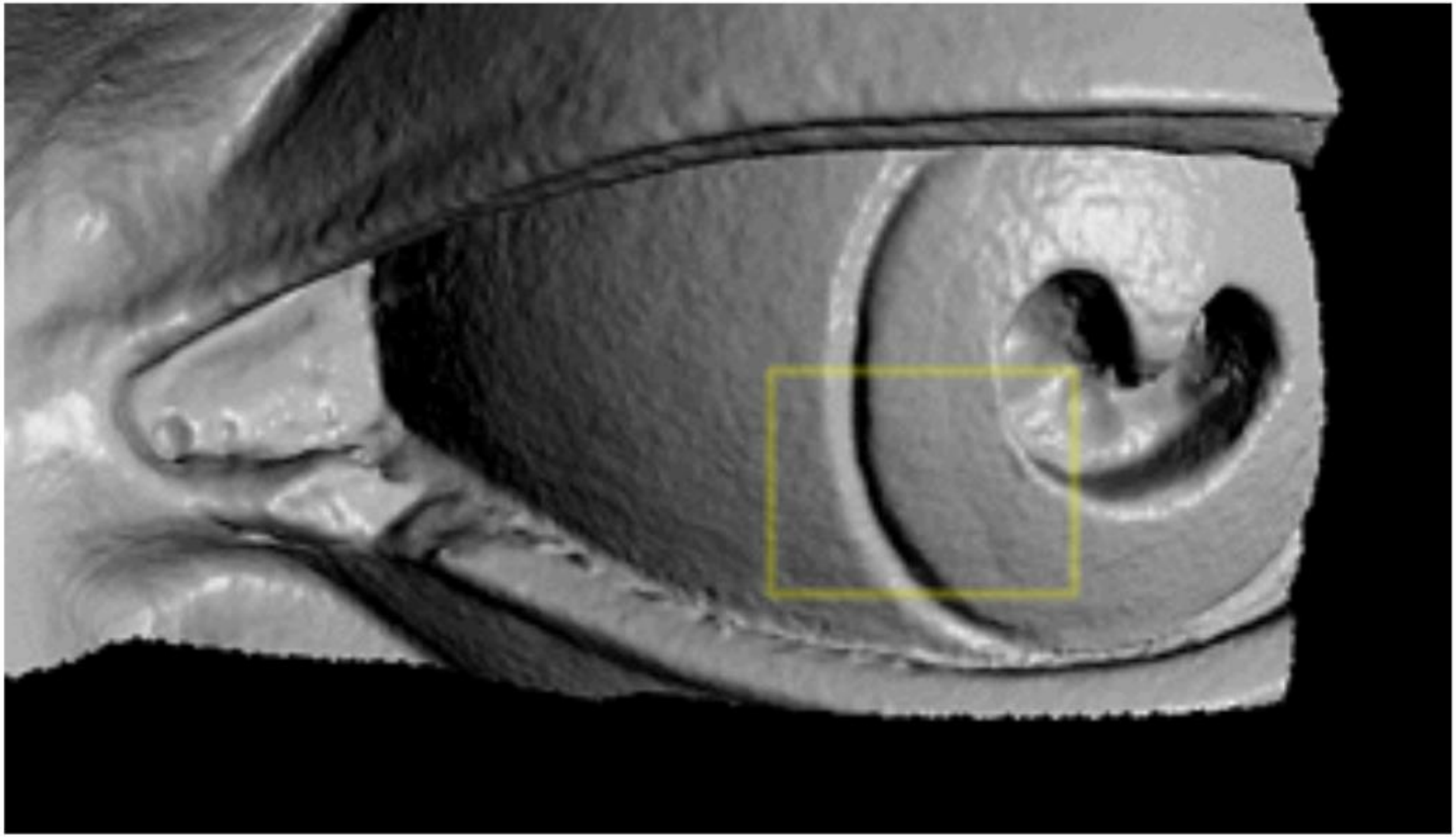
Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

The Digital Michelangelo Project, Levoy et al.

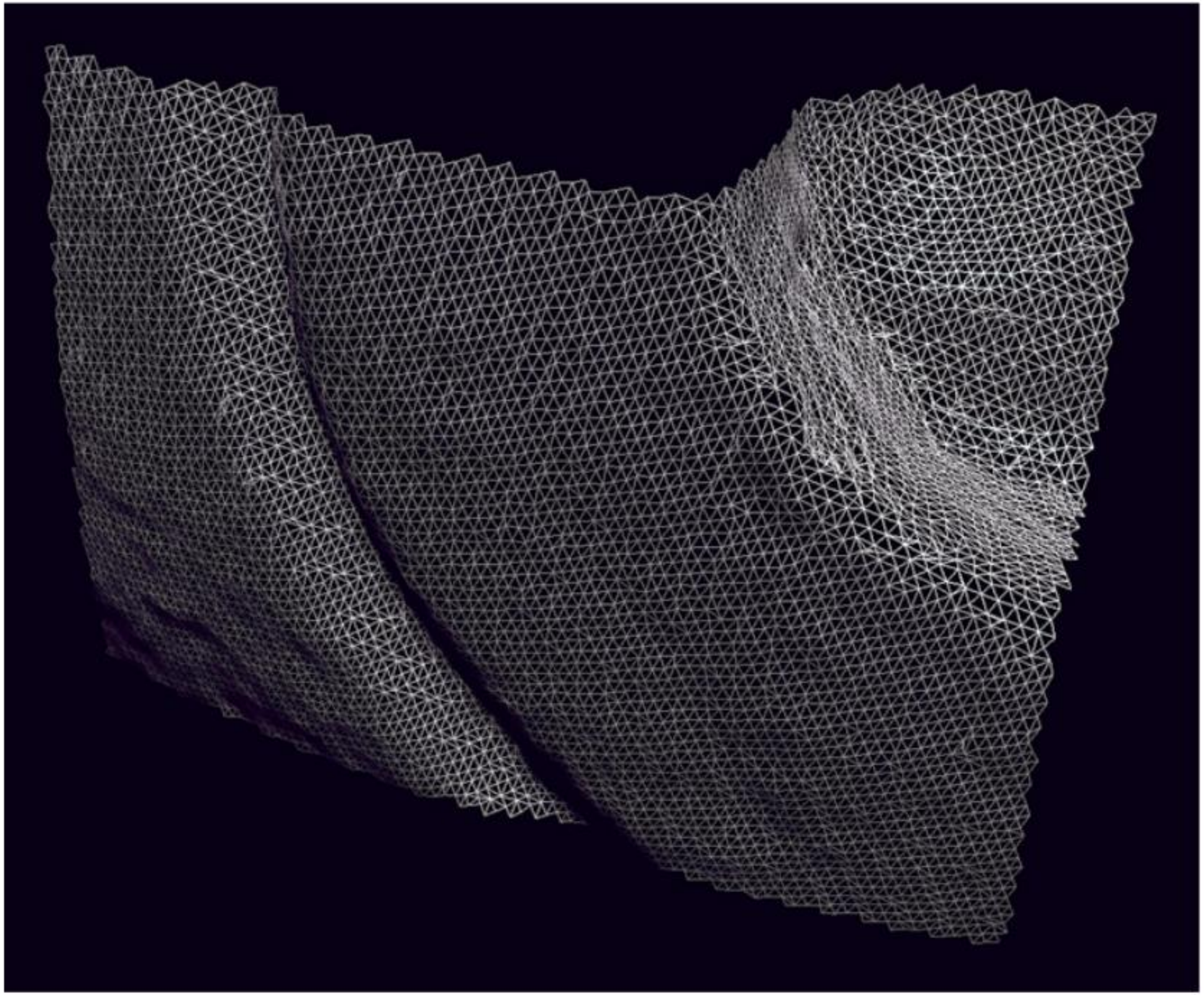
The Digital Michelangelo Project, Levoy et al.



The Digital Michelangelo Project, Levoy et al.



The Digital Michelangelo Project, Levoy et al.



The Digital Michelangelo Project, Levoy et al.

Portable 3D laser scanner (this one by Minolta)

Faster Acquisition?

Binary coding

Faster Acquisition?

- Project multiple stripes simultaneously
- What is the problem with this?

Faster Acquisition?

- Project multiple stripes simultaneously
- Correspondence problem: which stripe is which?

- Common types of patterns:
    - Binary coded light striping
    - Gray/color coded light striping

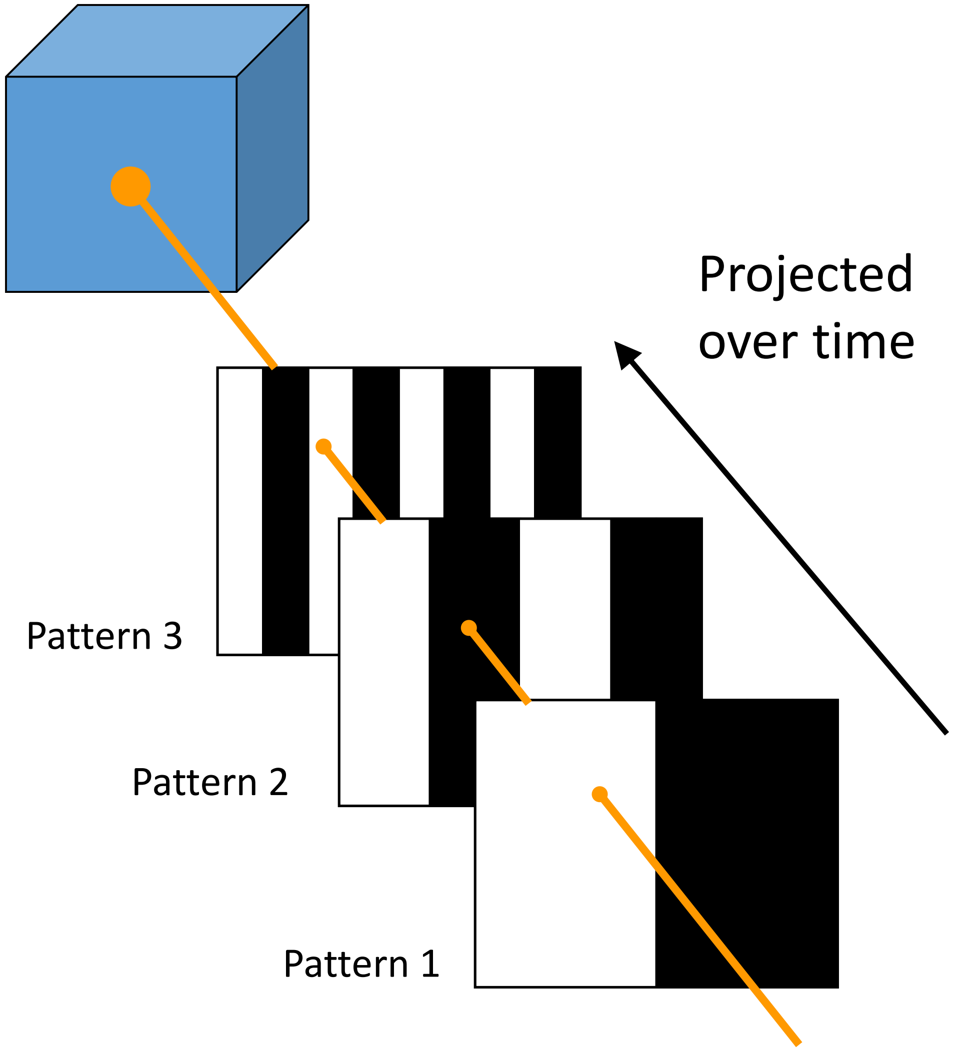
Binary Coding

Faster:

$2^n - 1$ stripes in n images.

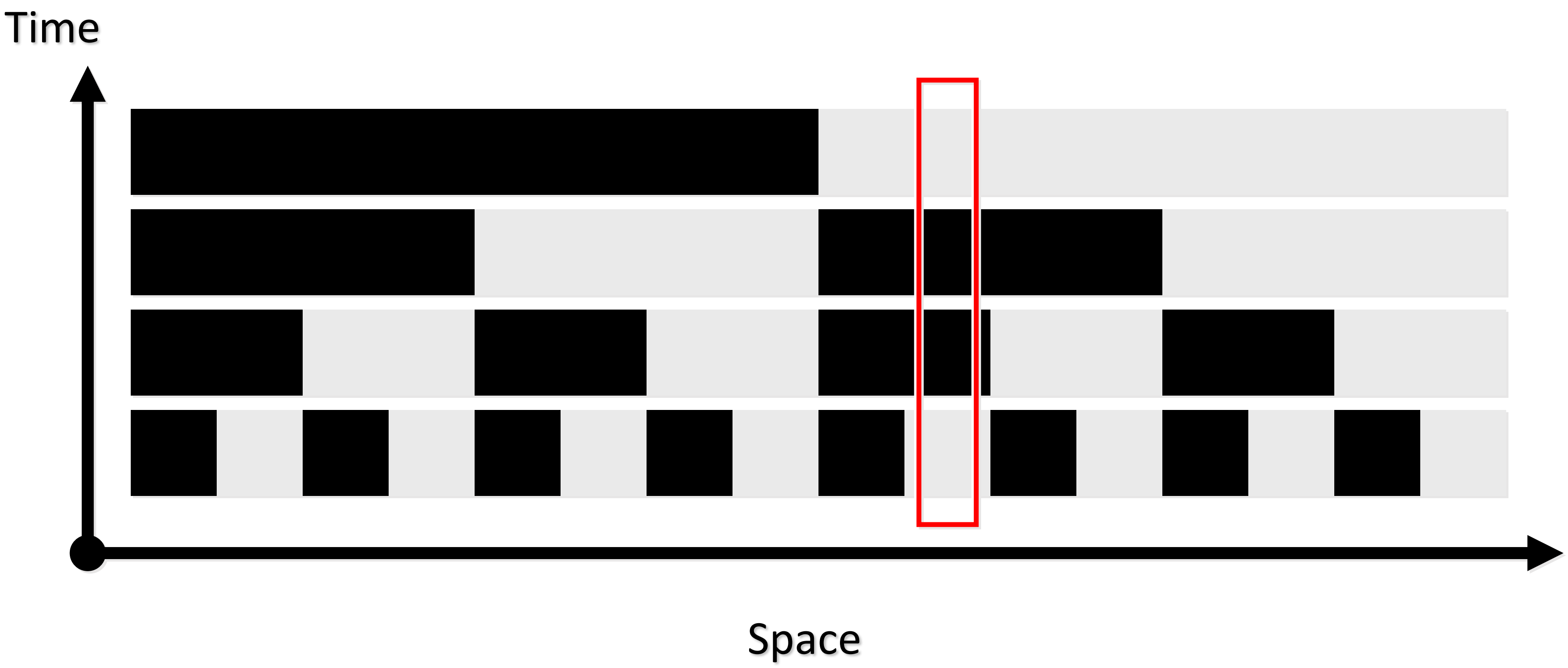
Example:

3 binary-encoded patterns which allows the measuring surface to be divided in 8 sub-regions

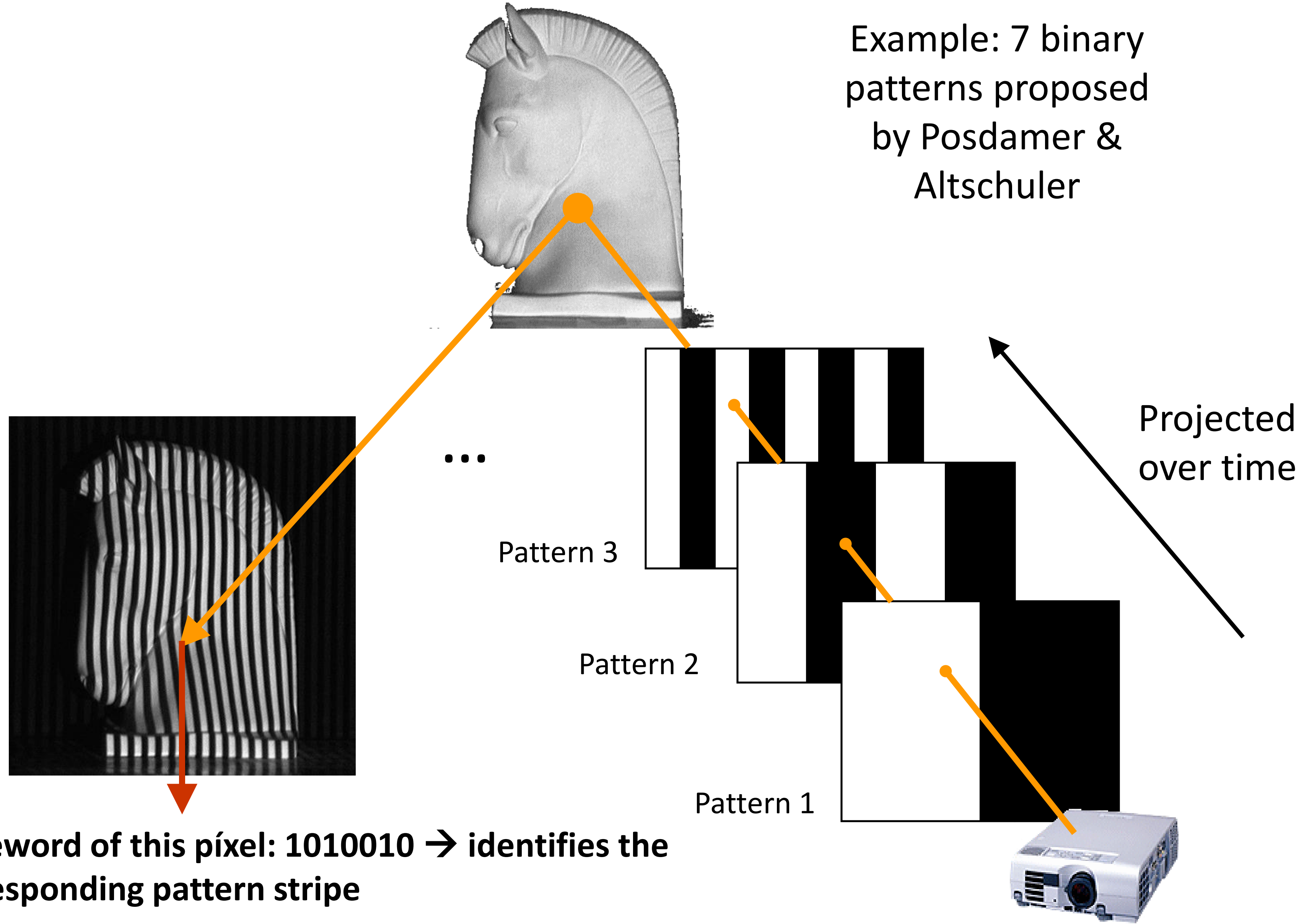


Binary Coding

- Assign each stripe a unique illumination code over time [Posdamer 82]

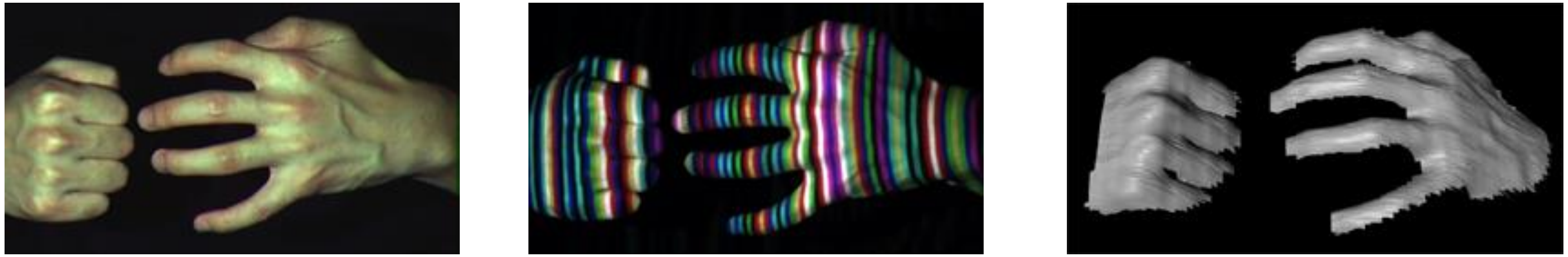


Binary Coding

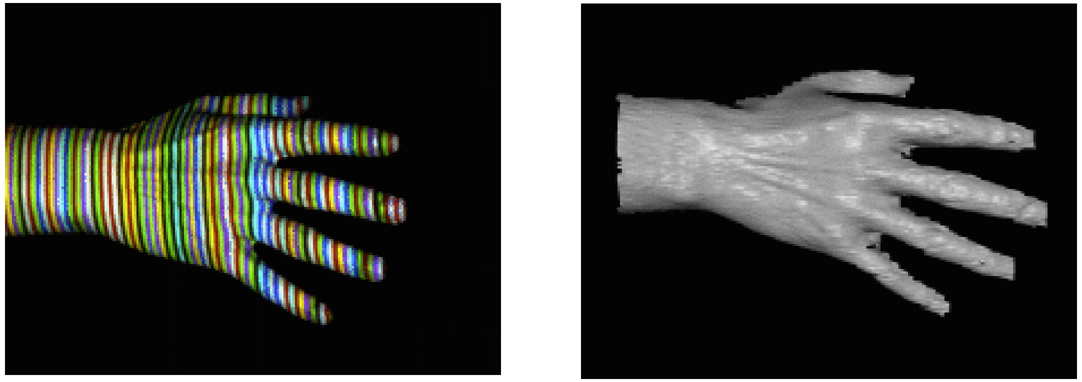


Codeword of this píxel: 1010010 → identifies the corresponding pattern stripe

More complex patterns



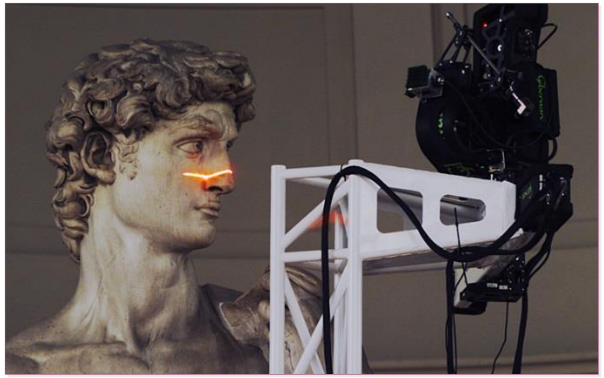
Works despite complex appearances



Works in real-time and on dynamic scenes

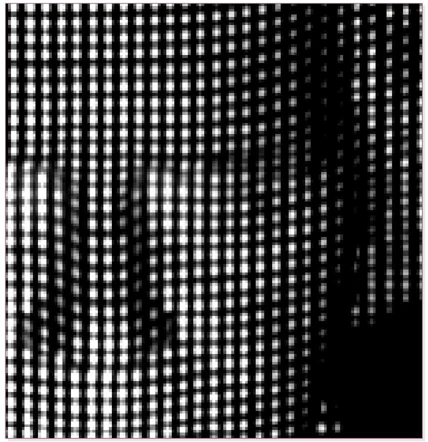
- Need very few images (one or two).
- But needs a more complex correspondence algorithm

Continuum of Triangulation Methods



Single-stripe

Multi-stripe
Multi-frame



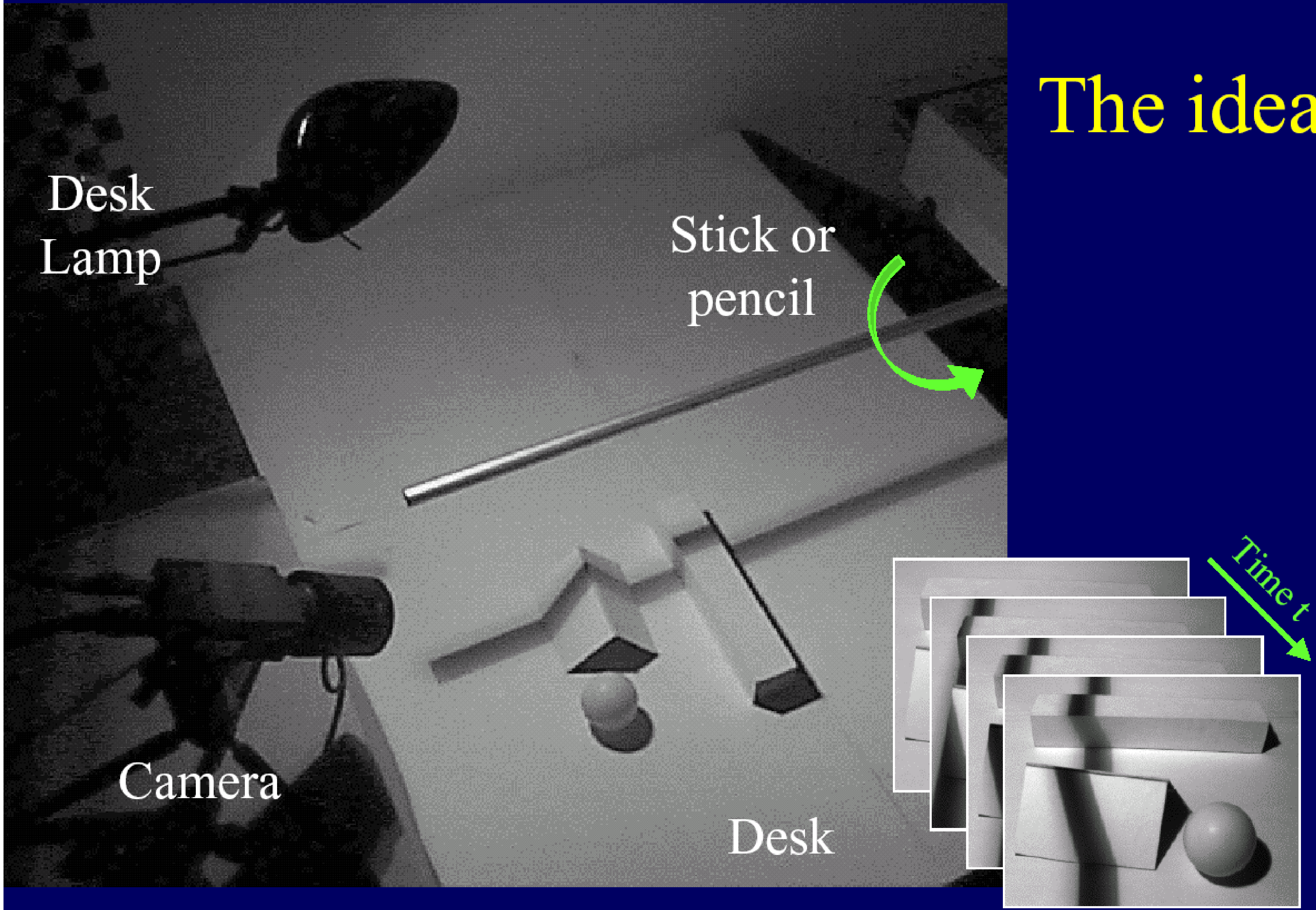
Single-frame

Slow, robust

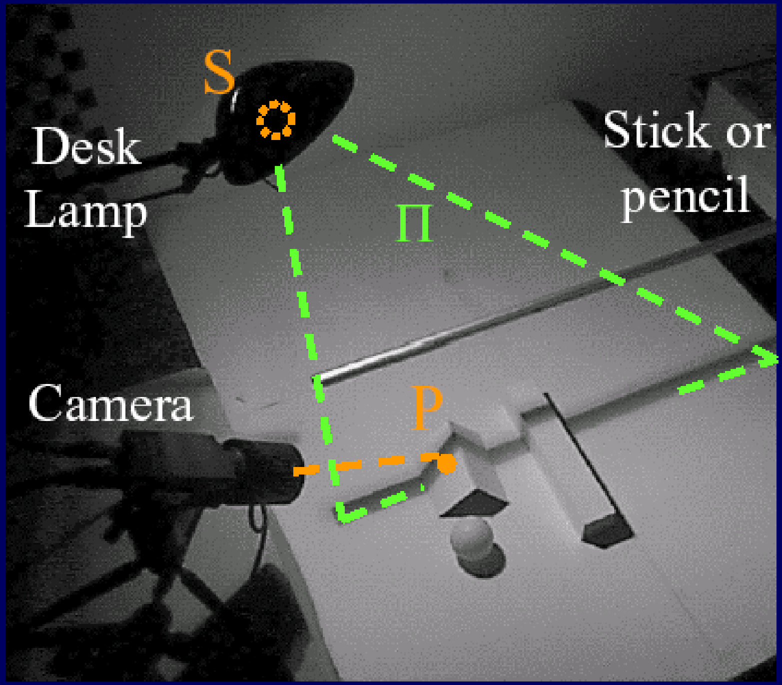
Fast, fragile

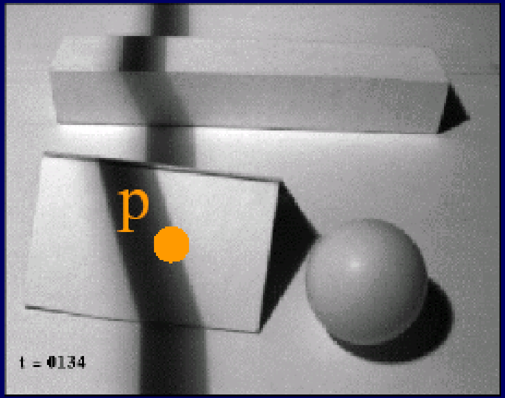
Using shadows

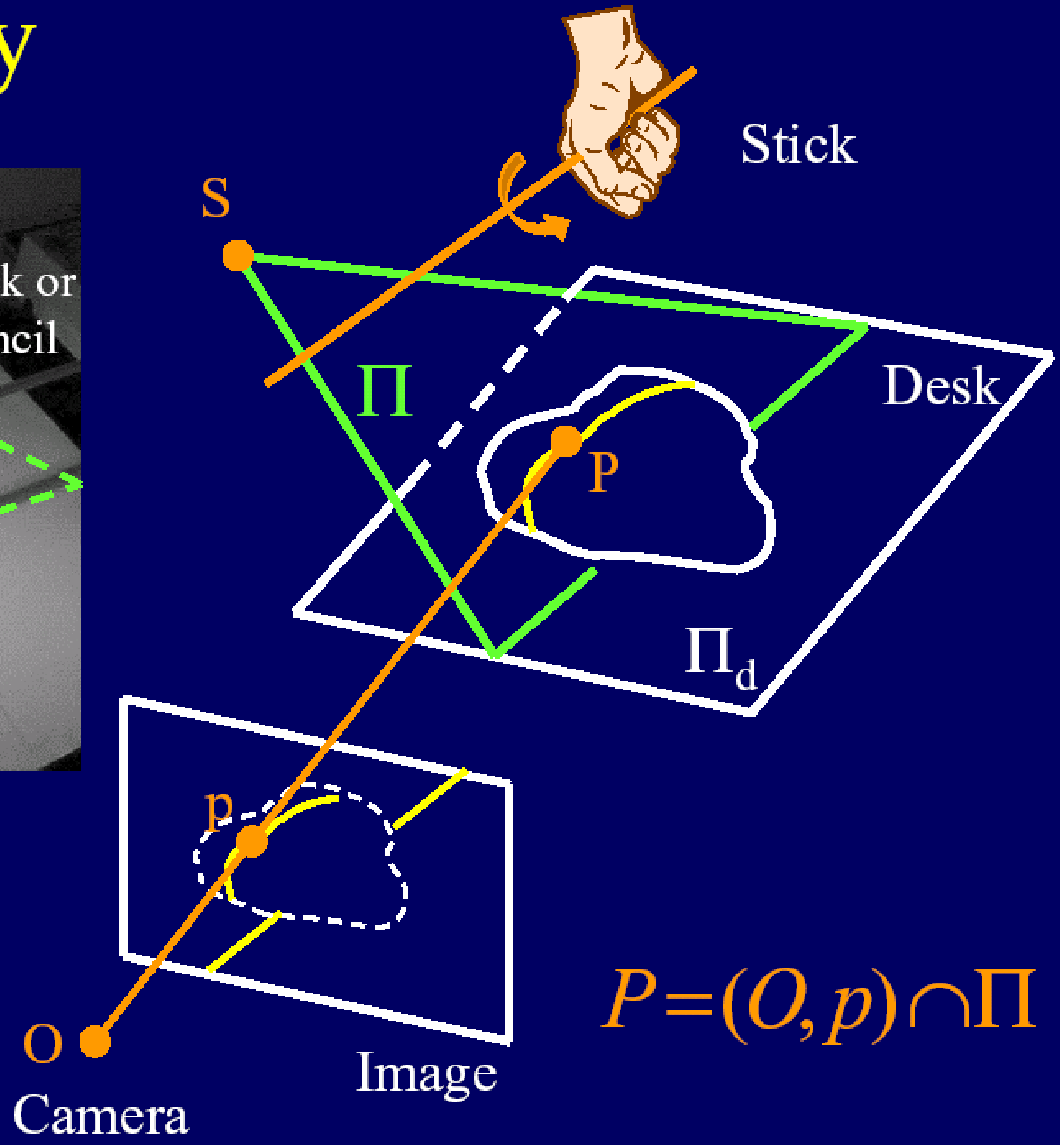
The idea



The geometry

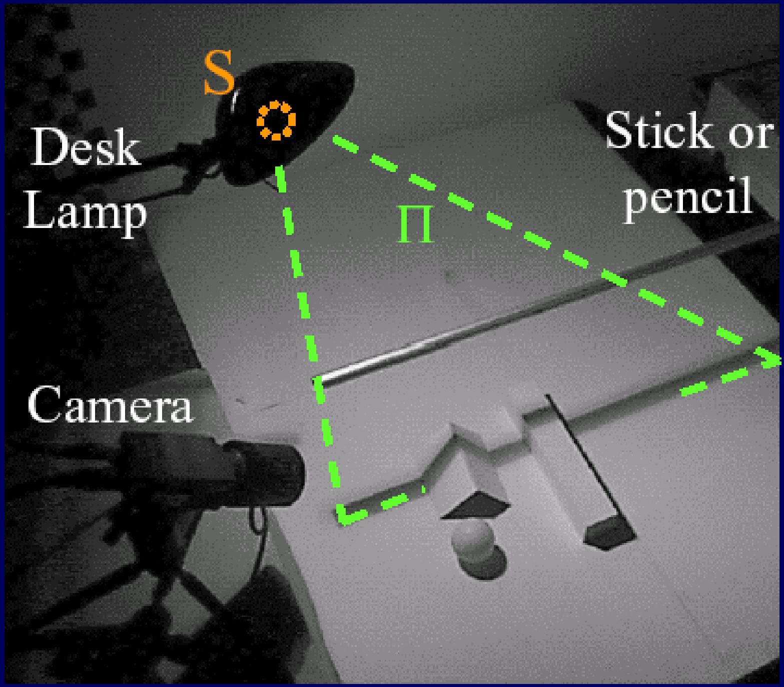


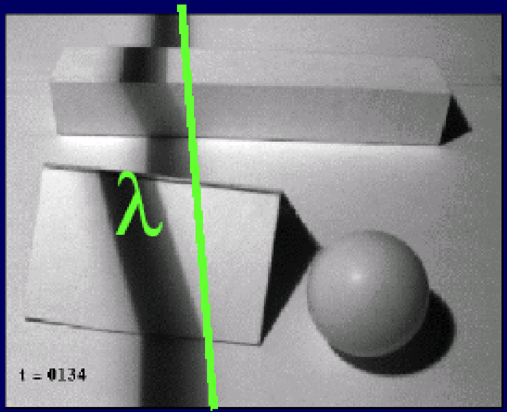


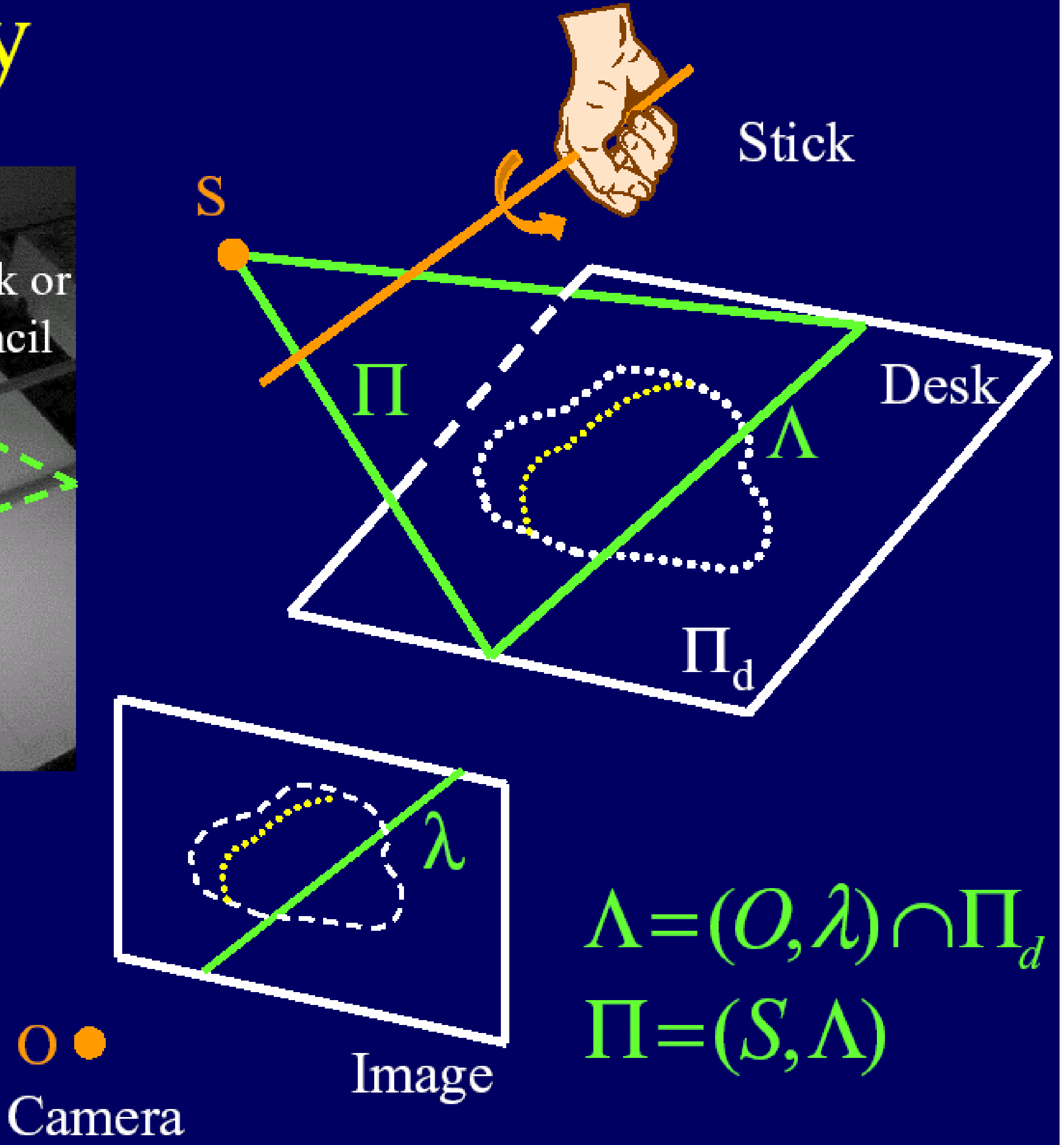


$$P = (O, p) \cap \Pi$$

The geometry



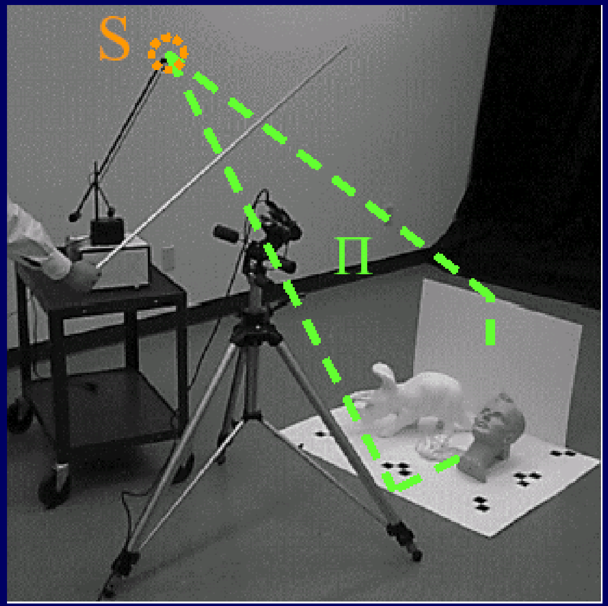


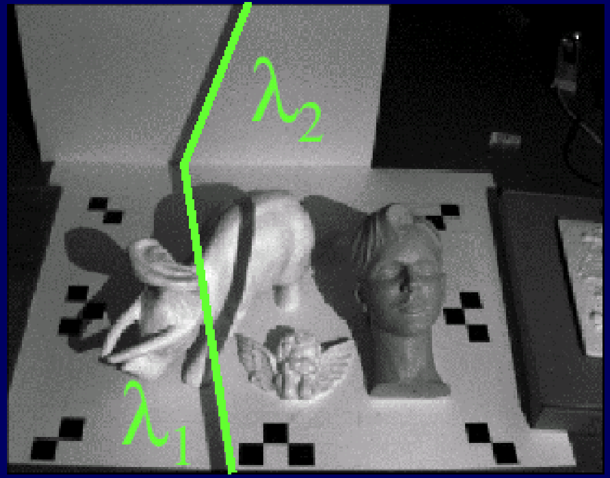


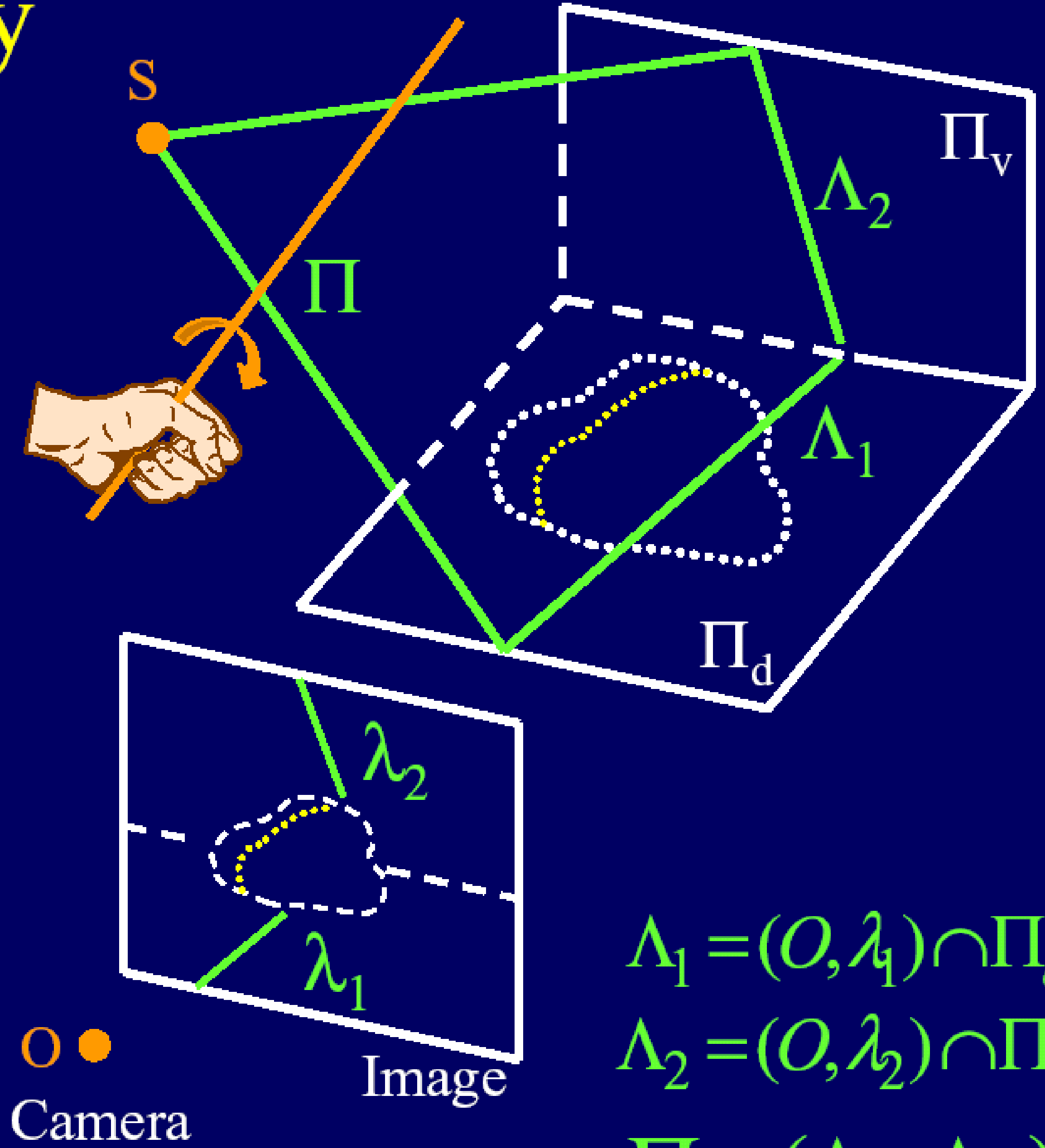
$$\Lambda = (O, \lambda) \cap \Pi_d$$

$$\Pi = (S, \Lambda)$$

The geometry





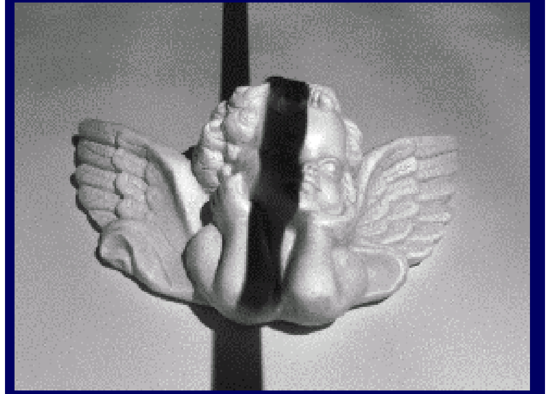


$$\Lambda_1 = (O, \lambda_1) \cap \Pi_d$$

$$\Lambda_2 = (O, \lambda_2) \cap \Pi_v$$

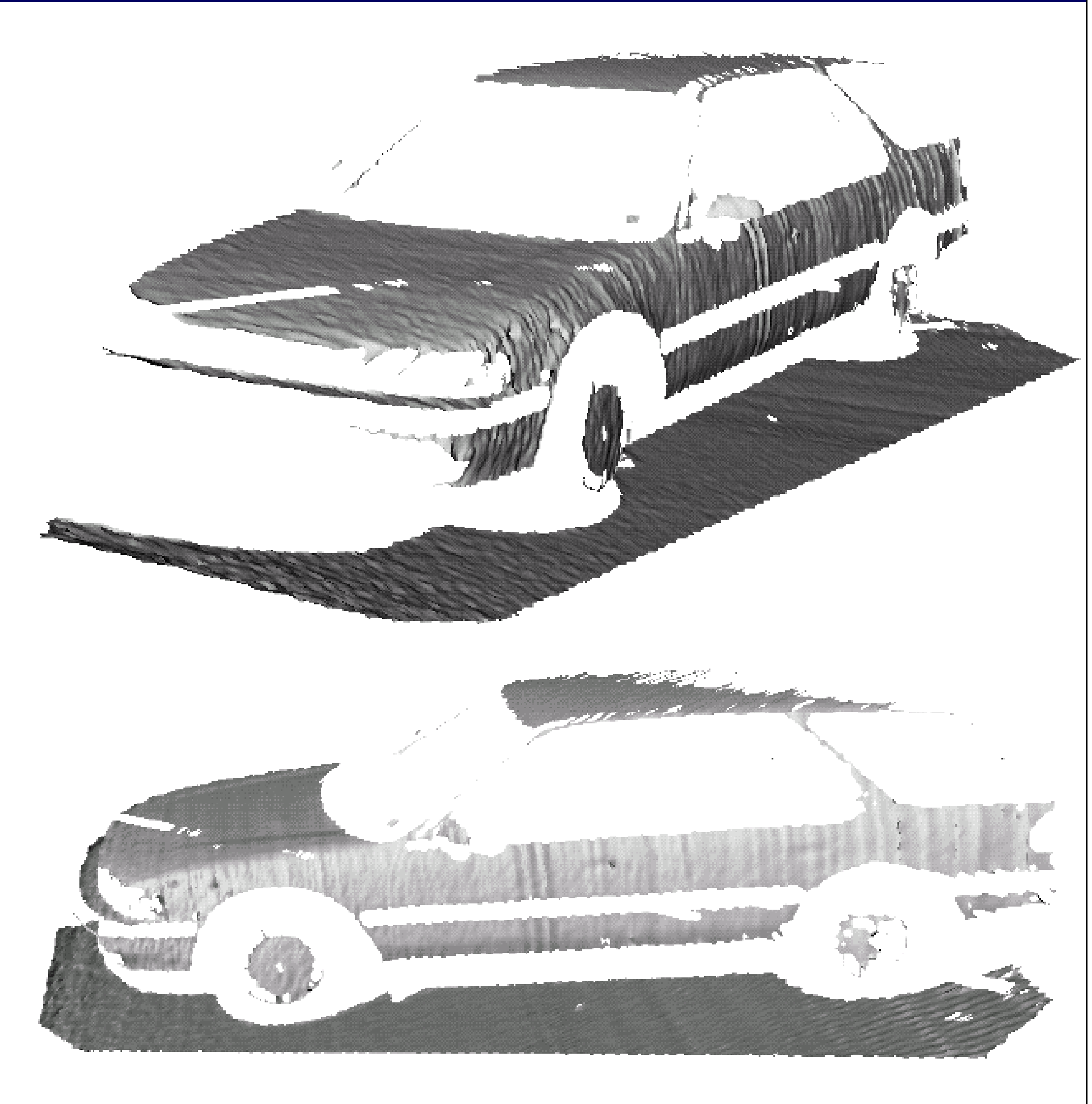
$$\Pi = (\Lambda_1, \Lambda_2)$$

Angel experiment



Accuracy: 0.1mm over 10cm → ~ 0.1% error

Scanning with the sun



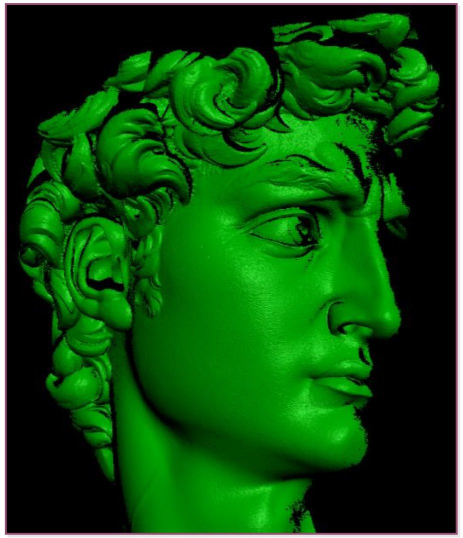
Accuracy: 1cm over 2m

→ ~ 0.5% error

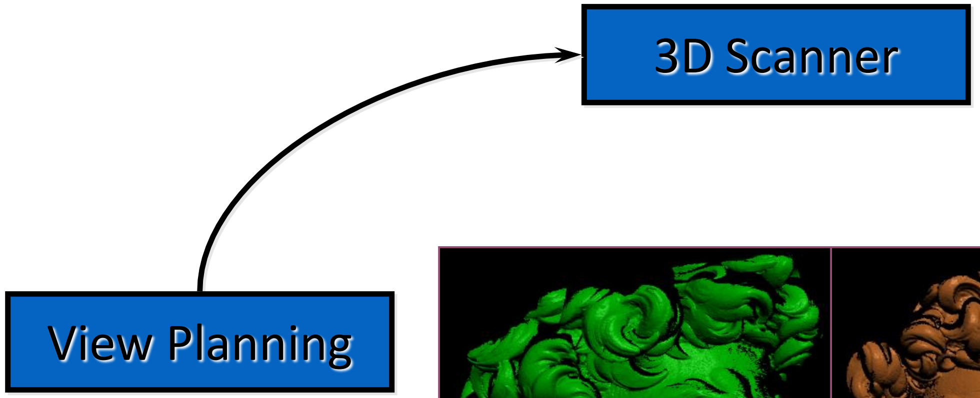
The 3D scanning pipeline

3D Model Acquisition Pipeline

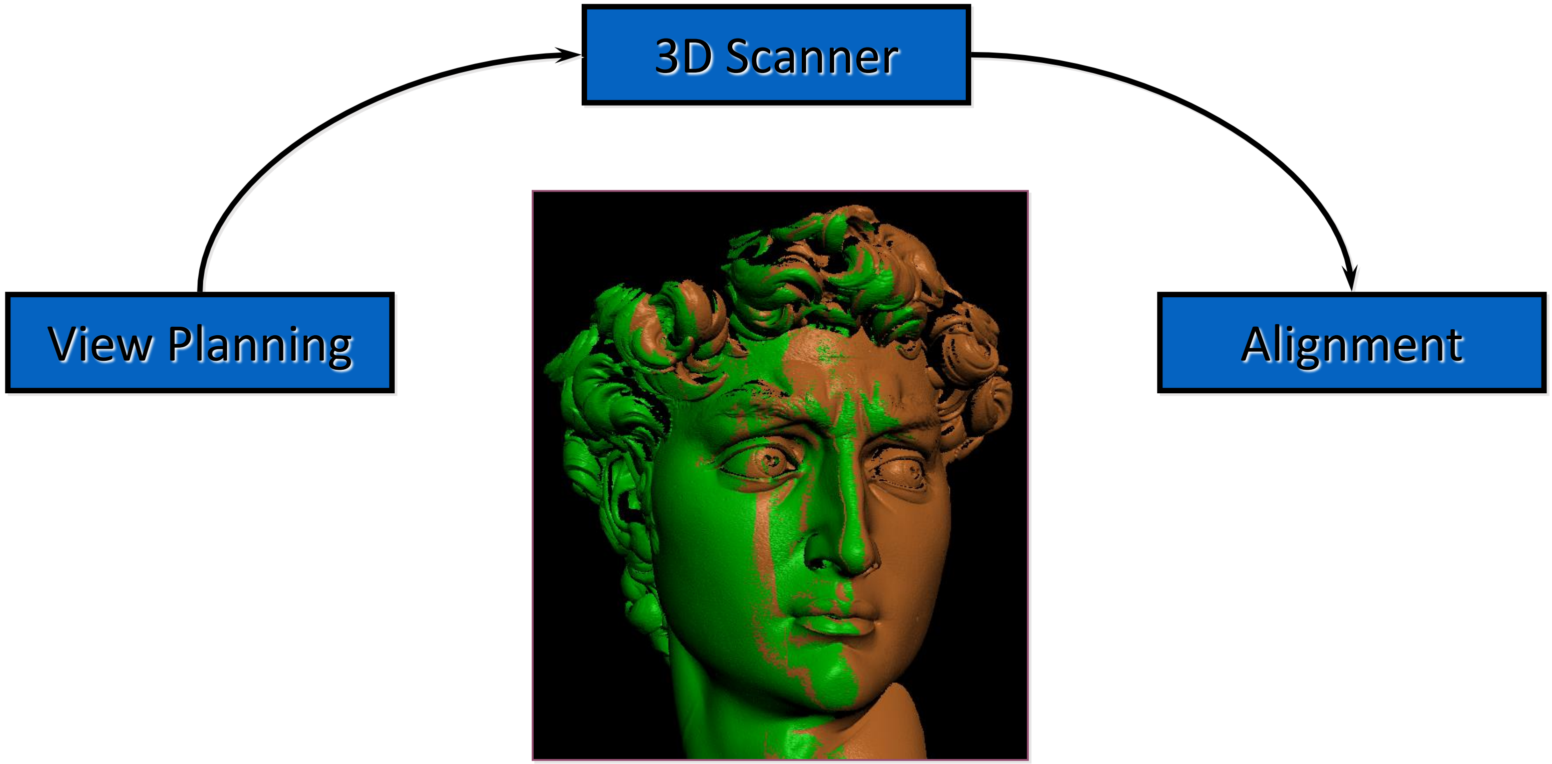
3D Scanner



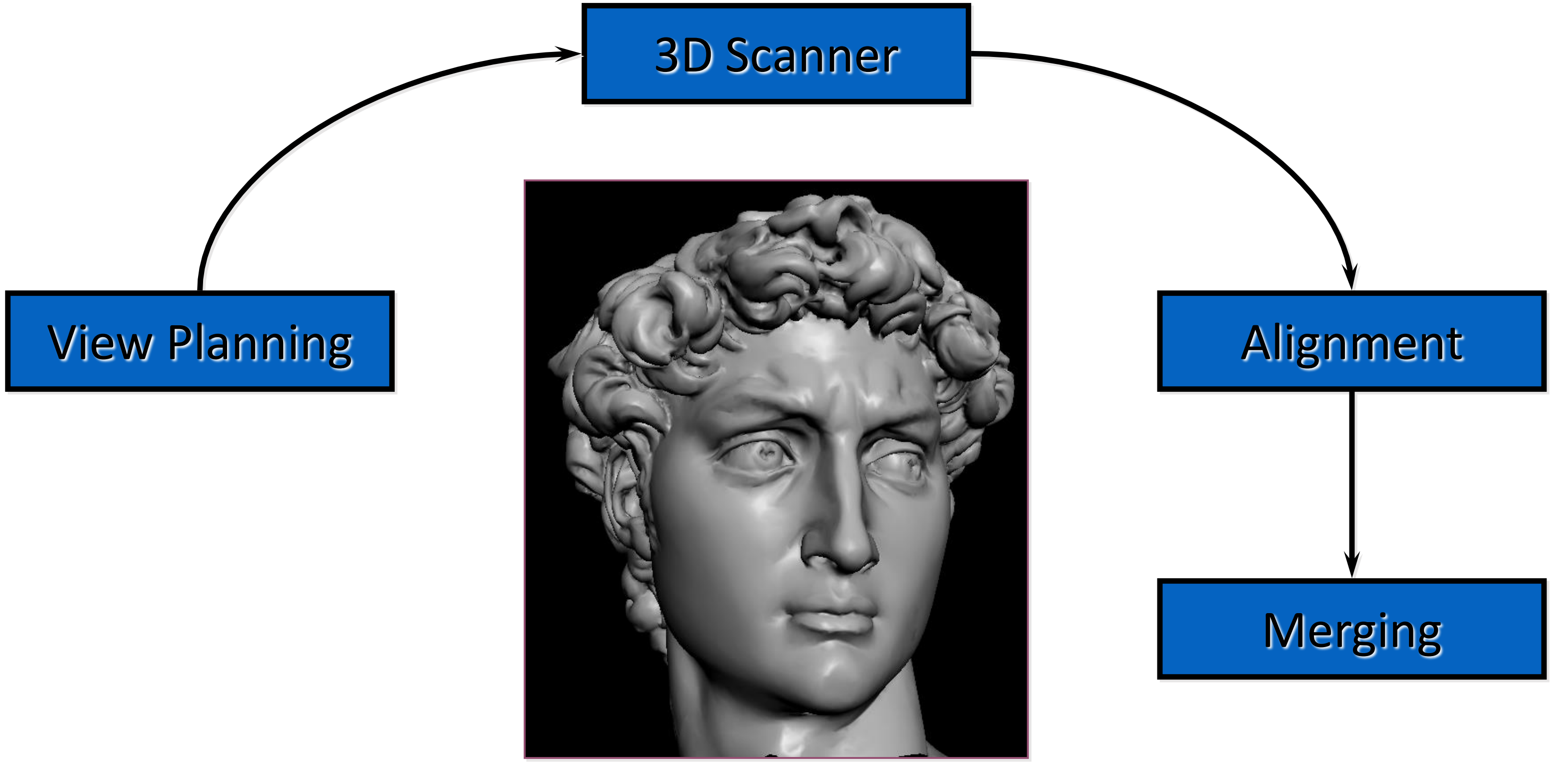
3D Model Acquisition Pipeline



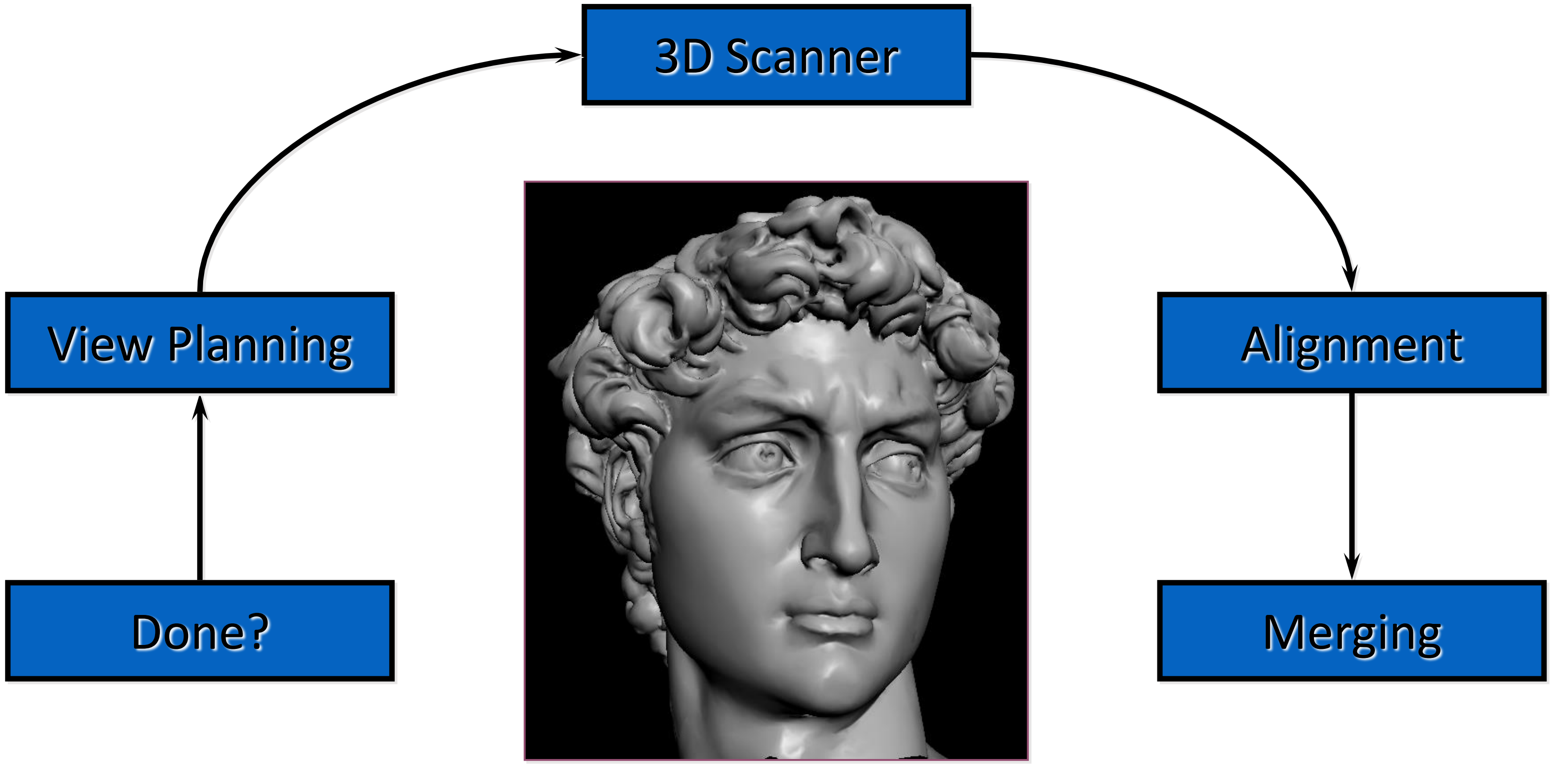
3D Model Acquisition Pipeline



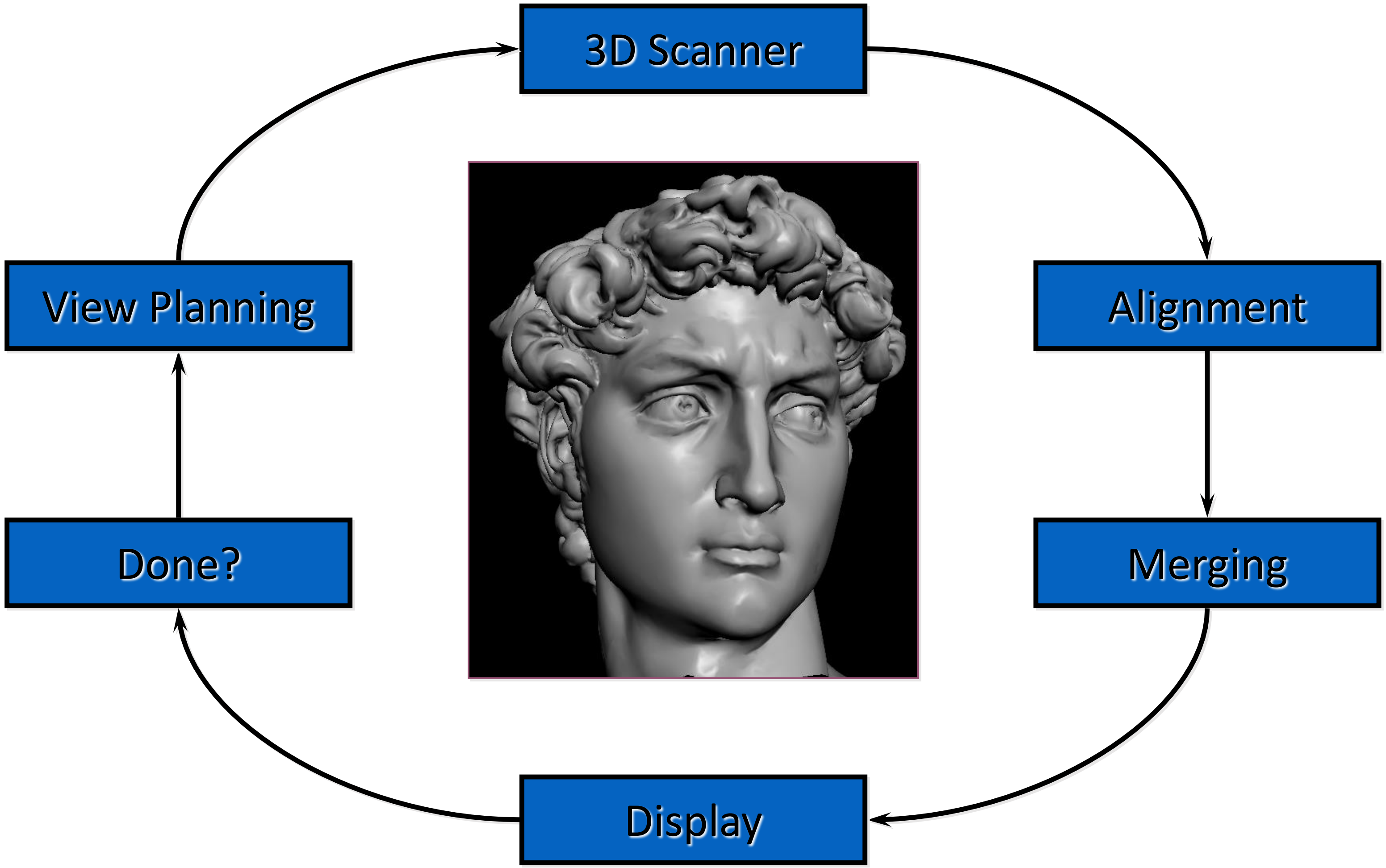
3D Model Acquisition Pipeline



3D Model Acquisition Pipeline

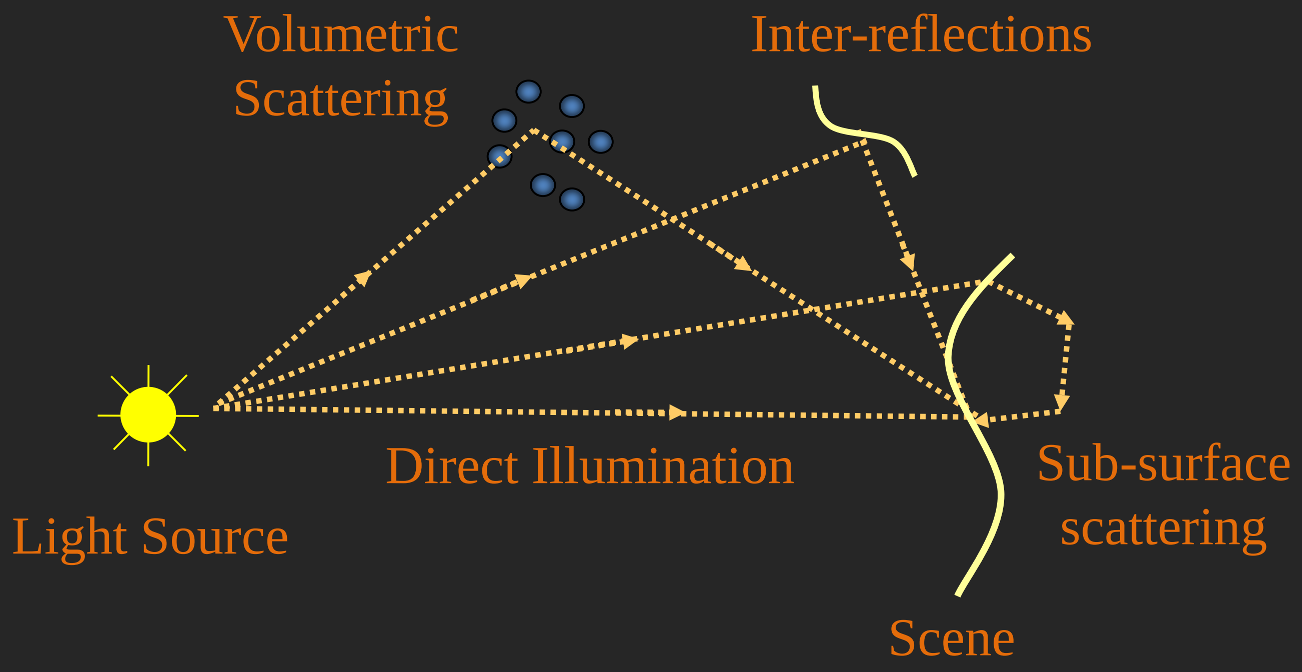


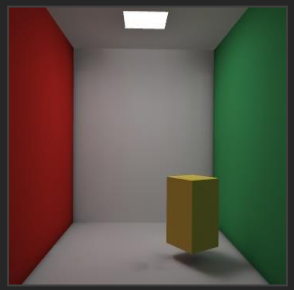
3D Model Acquisition Pipeline



Dealing with global illumination

Light Transport

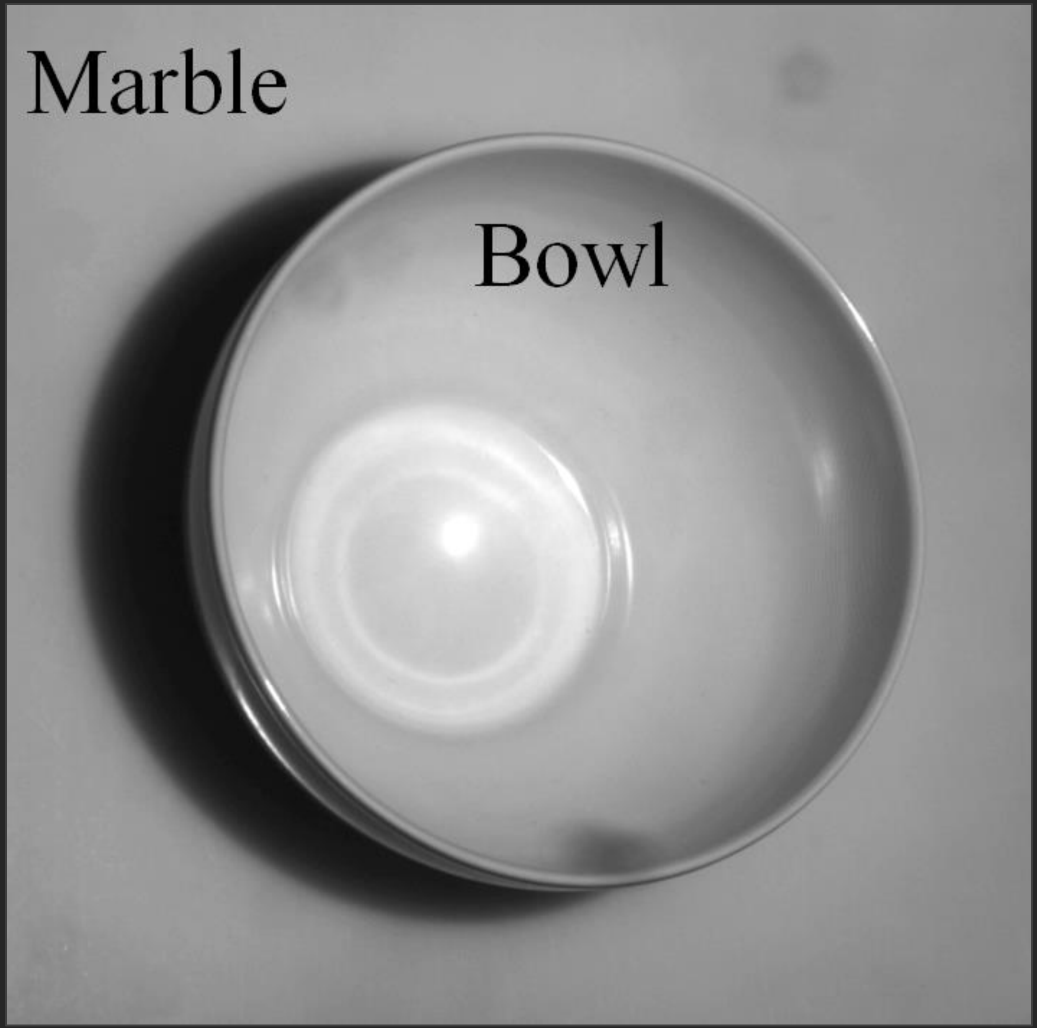




Slide adapted from Nayar et al

Why is global illumination a problem?

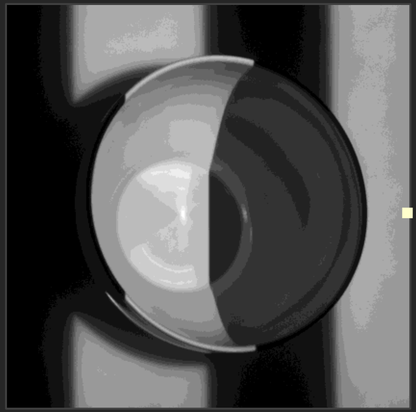
Bowl on a Marble Slab



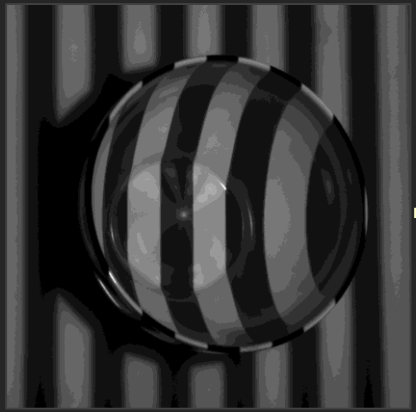
Captured images under conventional Gray codes

Lowest Frequency Illumination → Highest Frequency Illumination

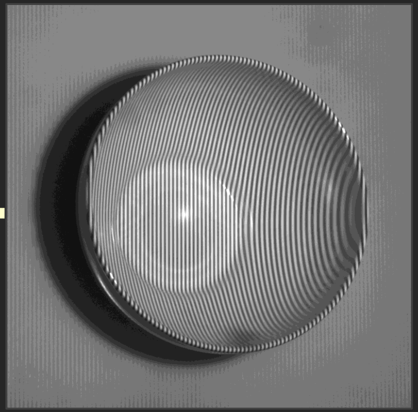
Pattern 1



Pattern 4

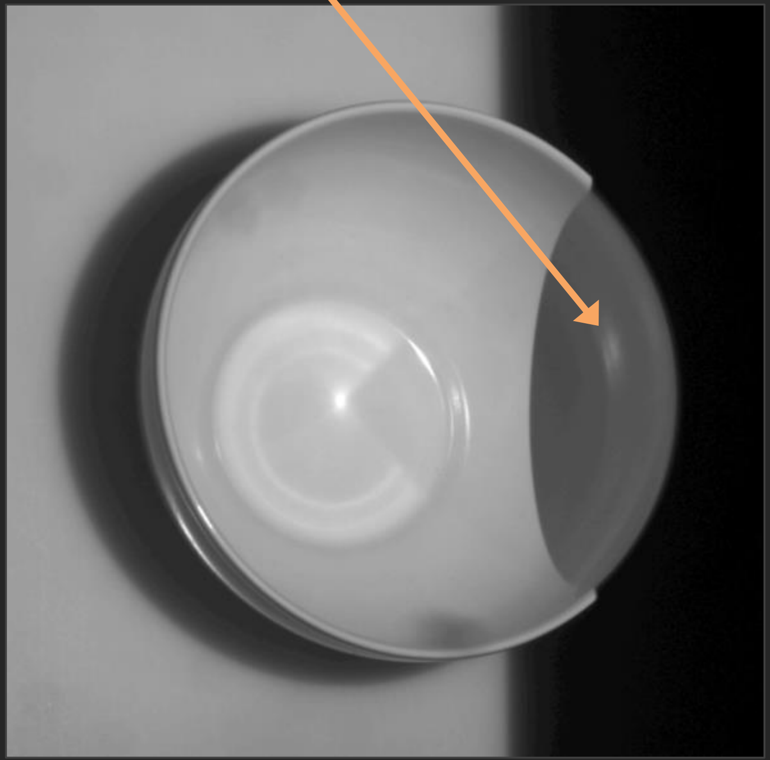


Pattern 7

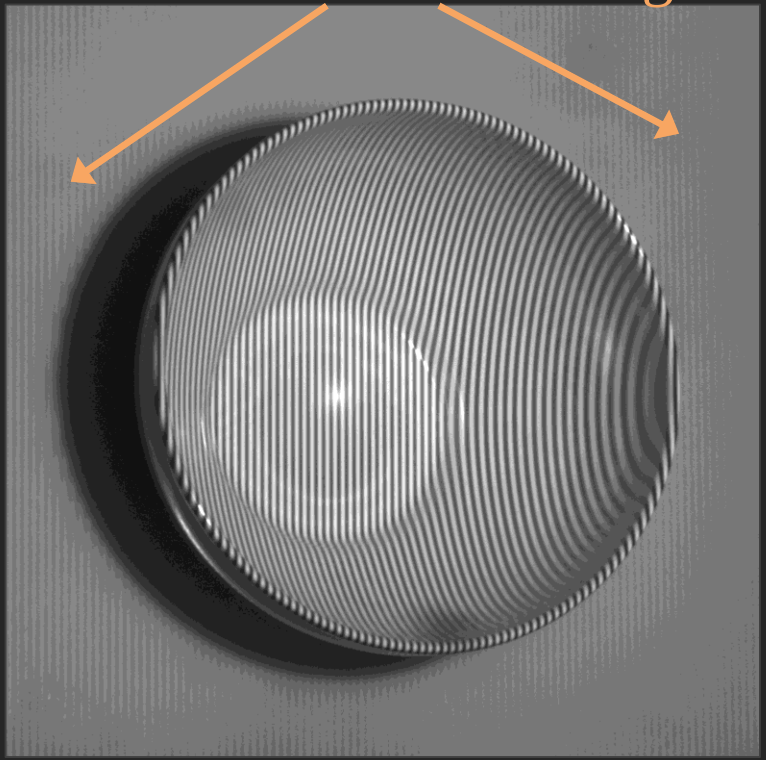


Pattern 10

Issues due to global illumination effects

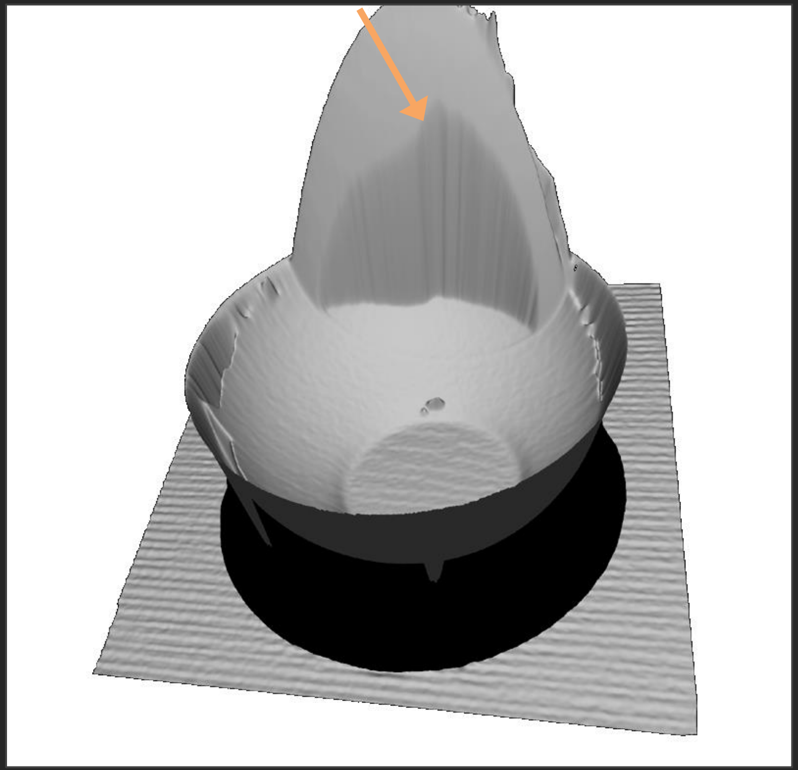


Low-frequency pattern

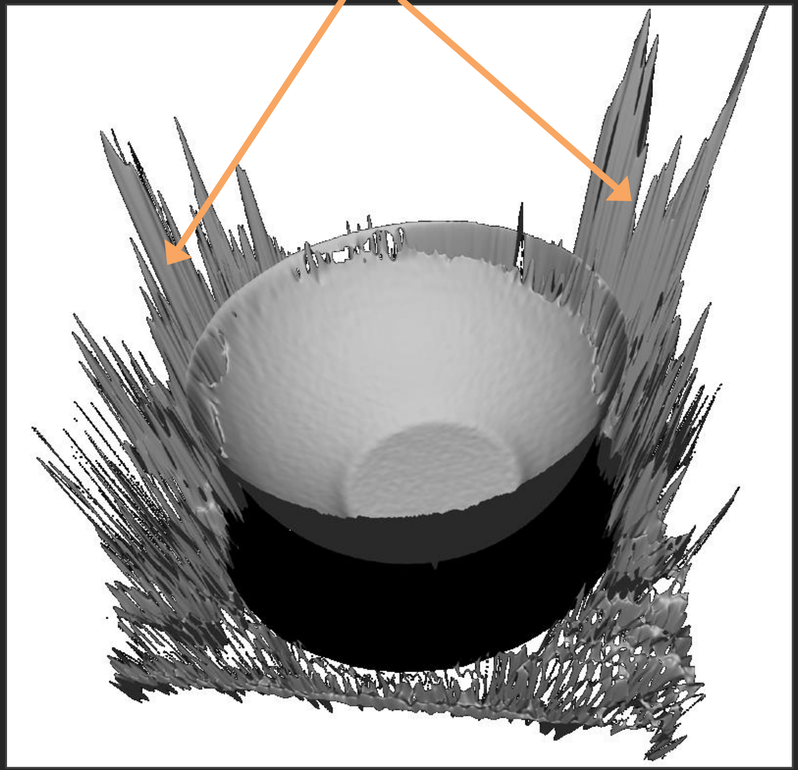


High-frequency pattern

3D Visualizations: State of the Art



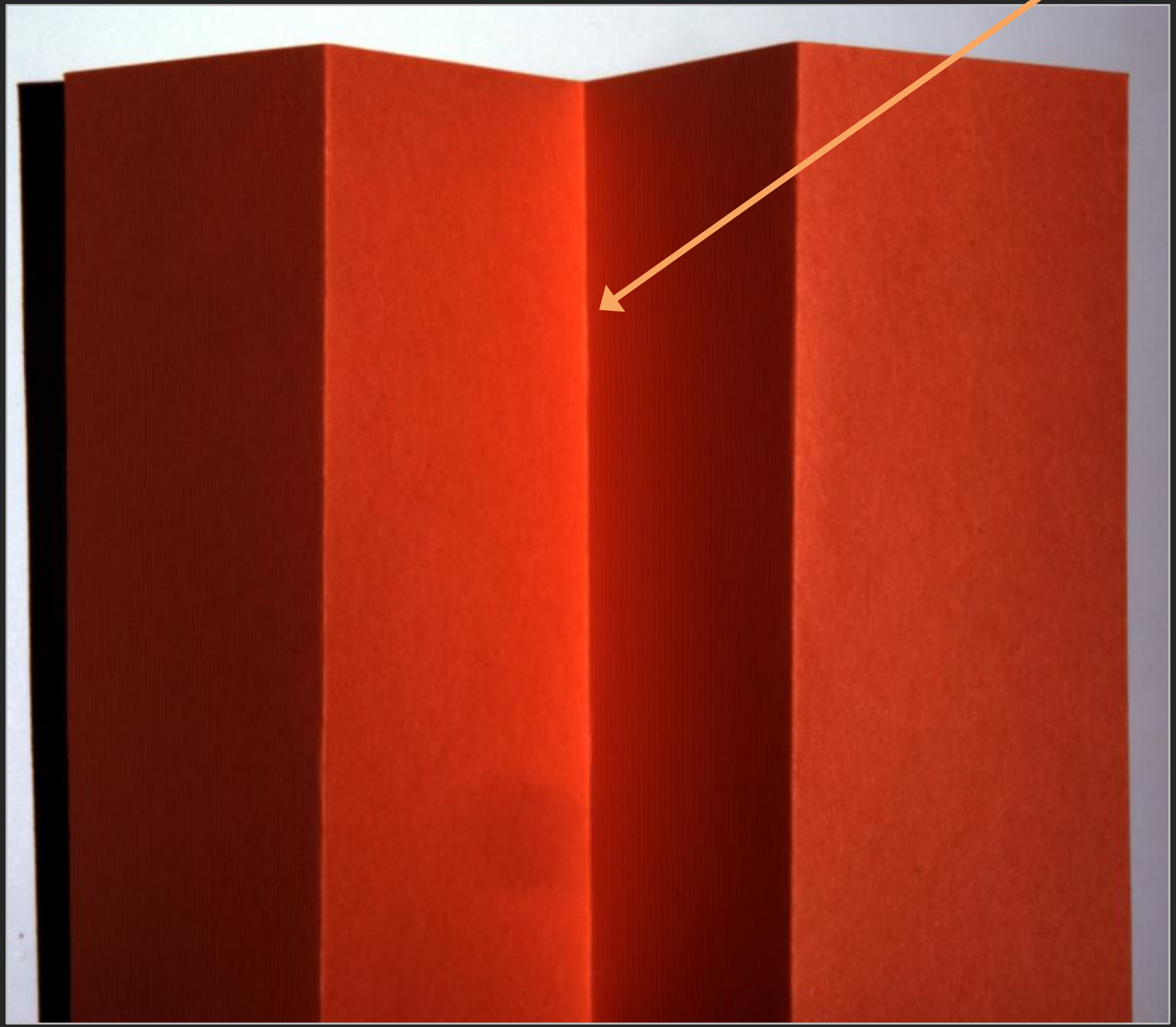
Conventional Gray
(11 images)



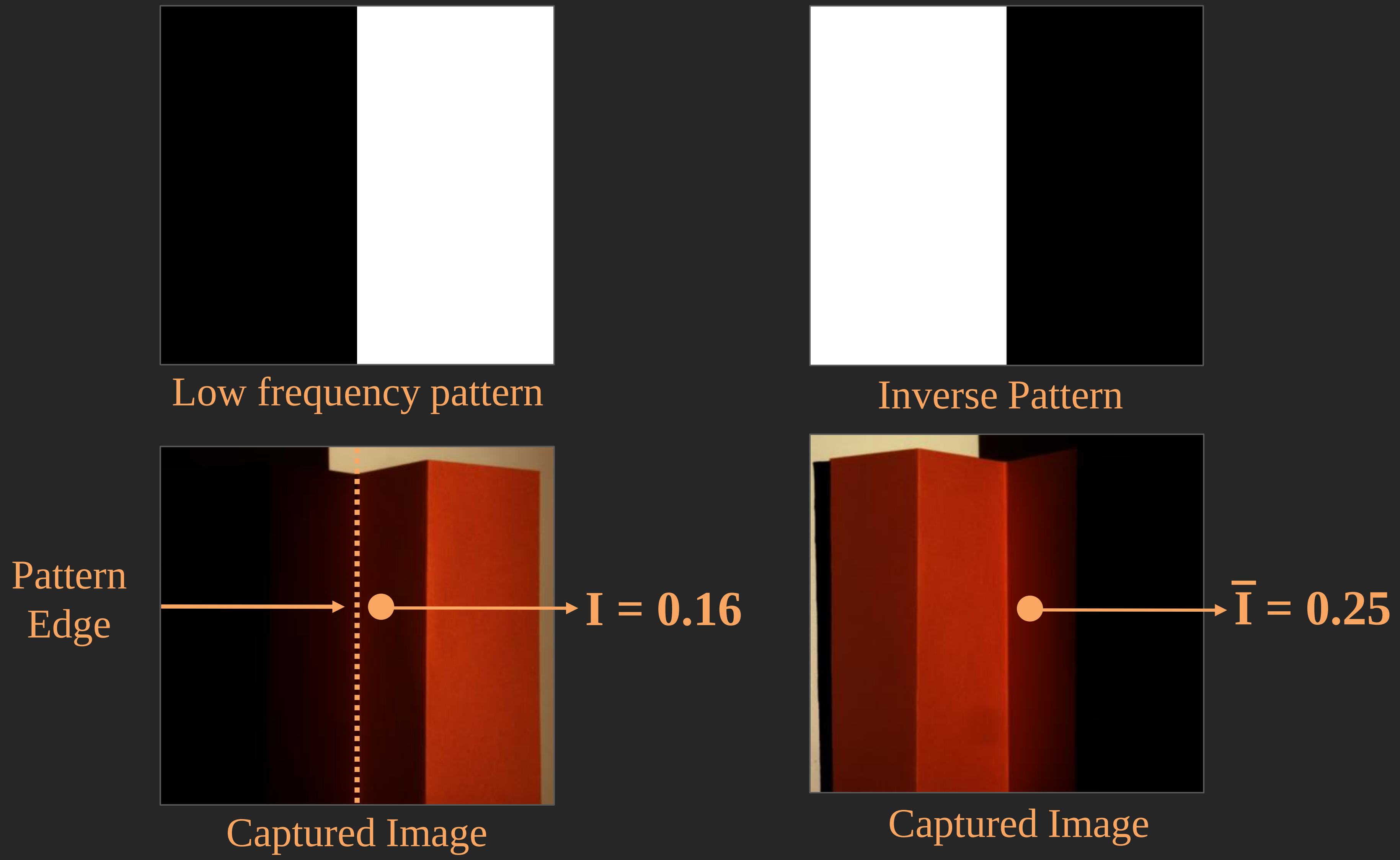
Modulated Phase-Shifting
(162 images)

V-Groove Scene

Inter-reflections

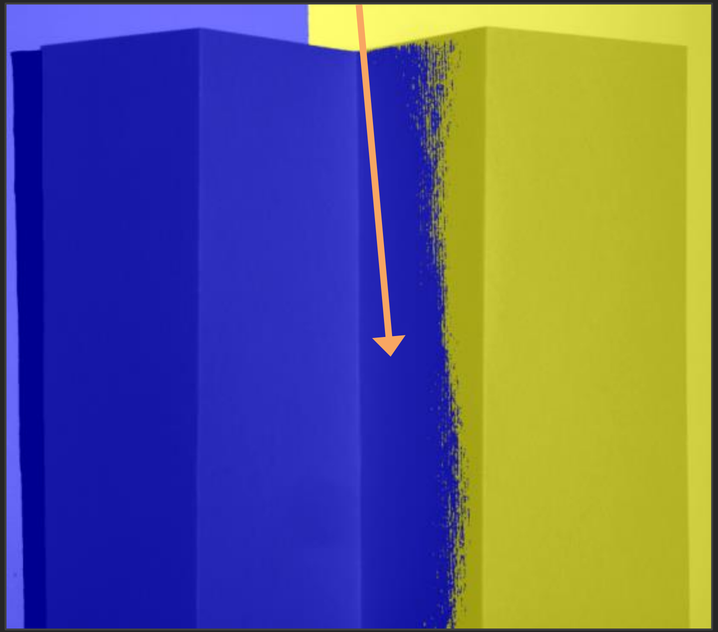


Conventional Gray codes

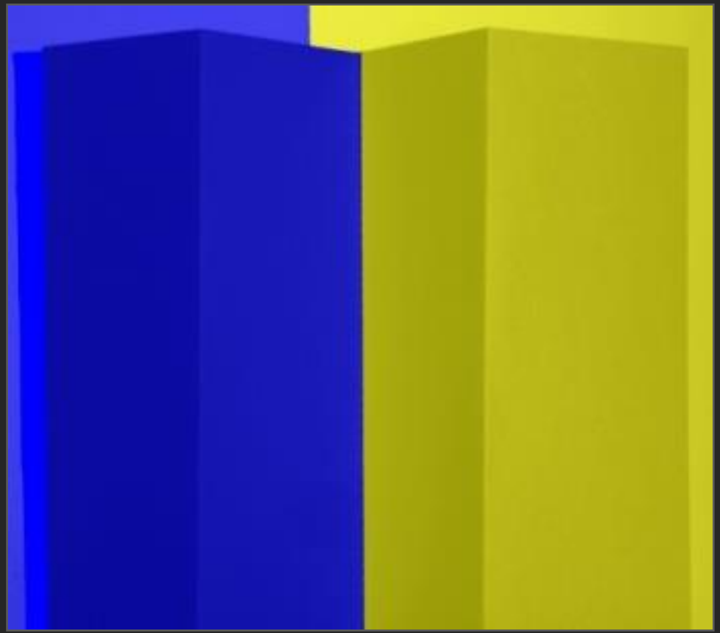


Binarization error

Errors due to inter-reflections



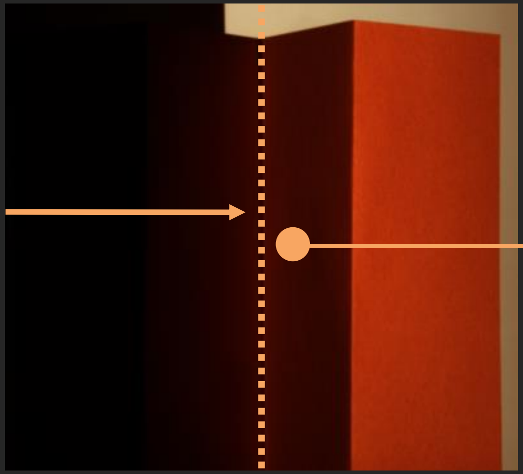
Incorrect Binarization



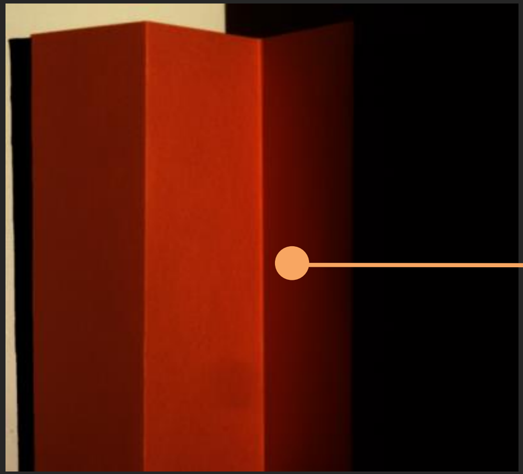
Ground-truth Binarization

One (illuminated) Zero (not-illuminated)

Why is the Decoding Incorrect for Low-frequencies?



Captured Image



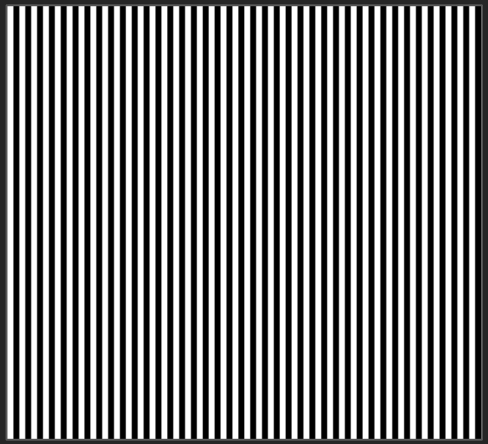
Captured Image

$$I = \text{Direct} + \alpha \,.\, \text{Global}$$

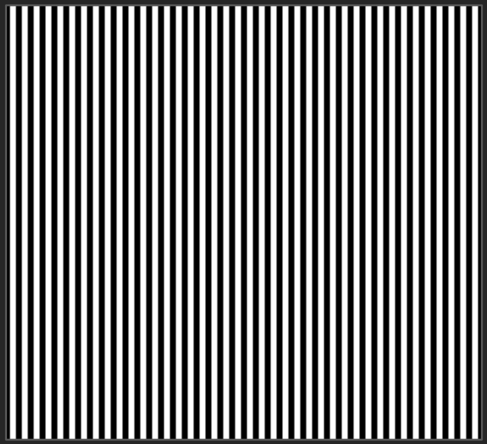
$$\bar{I} = (1 - \alpha) \,.\, \text{Global}$$

$$\alpha \sim= 0, \;\; \text{Direct} < \text{Global} \quad => \quad I < \bar{I}$$

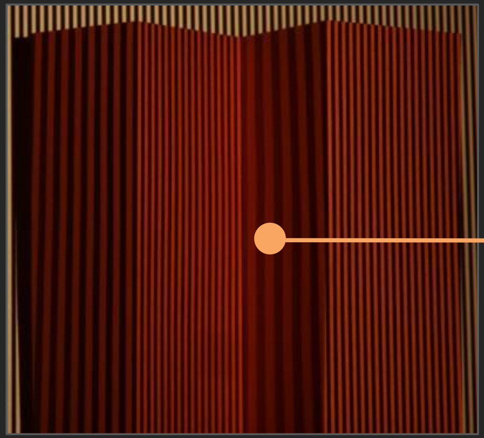
Binarization for high-frequency pattern



Pattern

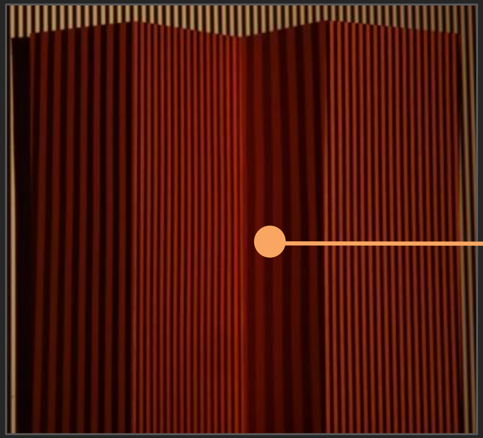


Inverse Pattern



$I = 0.25$

Captured Image

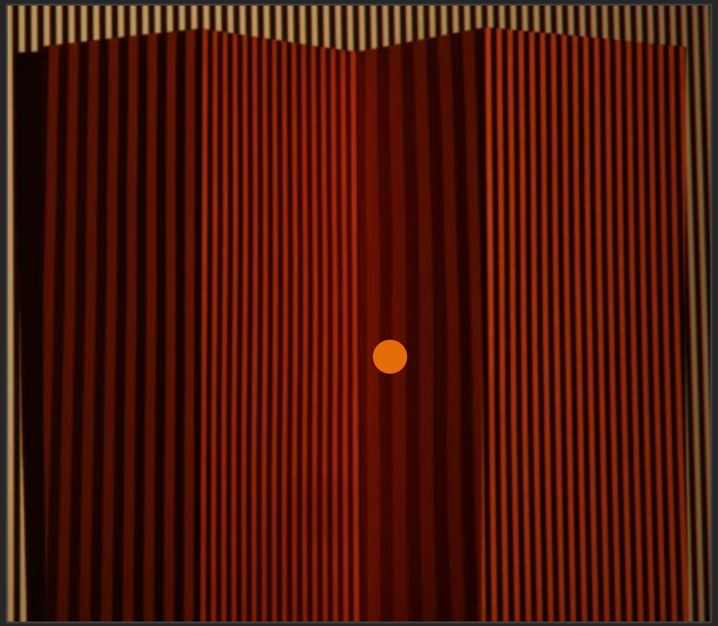


$\bar{I} = 0.16$

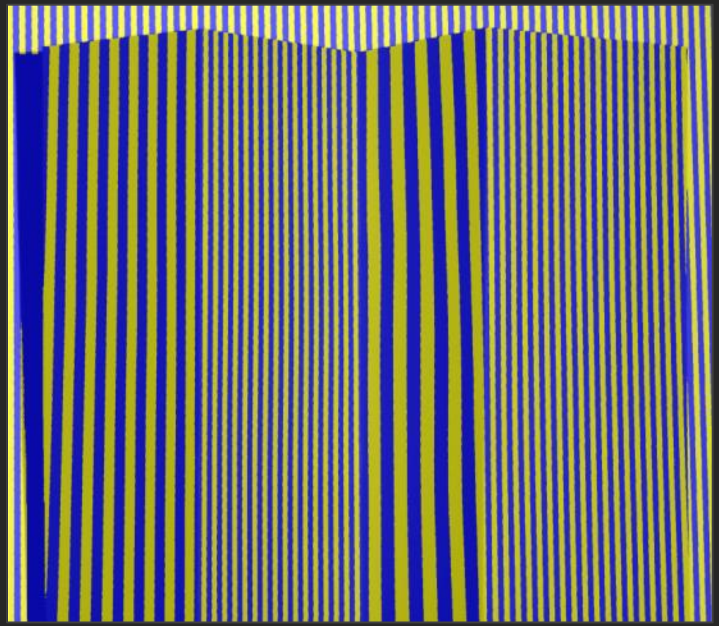
Captured Image

$$I = \text{Direct} + \mathbf{0.5}\ \text{Global} \quad > \quad \bar{I} = 0.5\ \text{Global}$$

High-frequency Patterns are Decoded Correctly



Captured Image

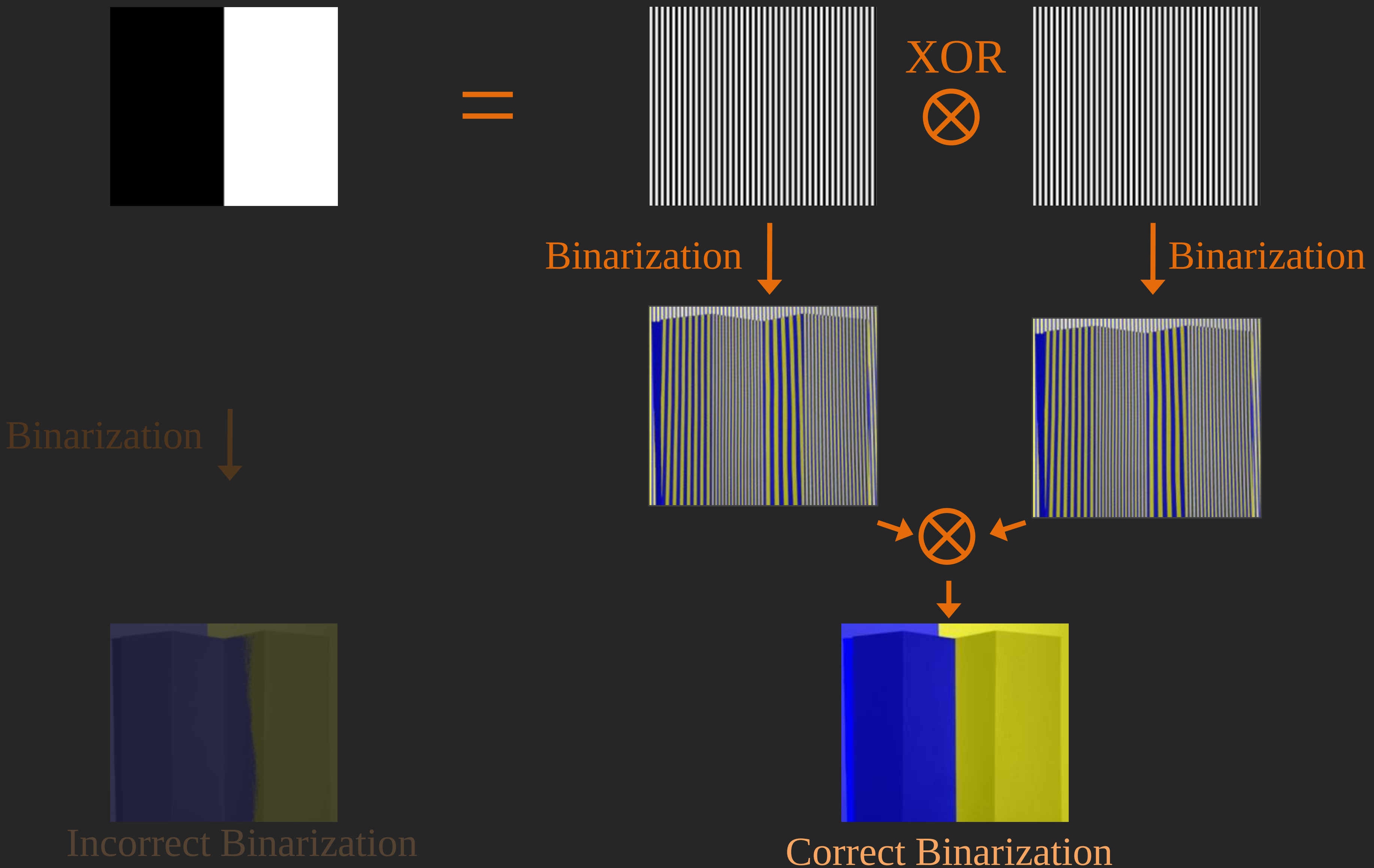


Binary Decoding

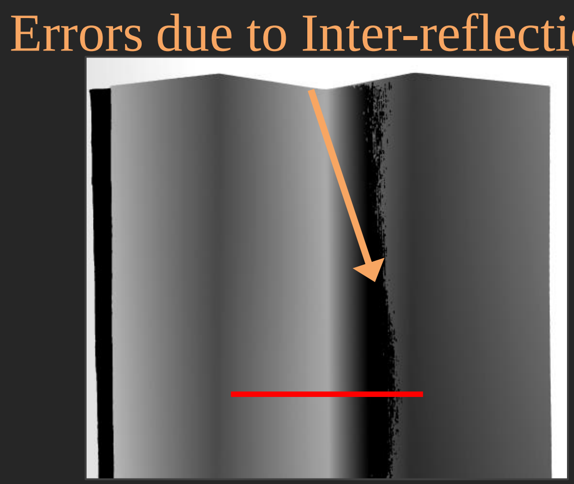
Preventing errors due to long-range effects

Logical Coding and Decoding

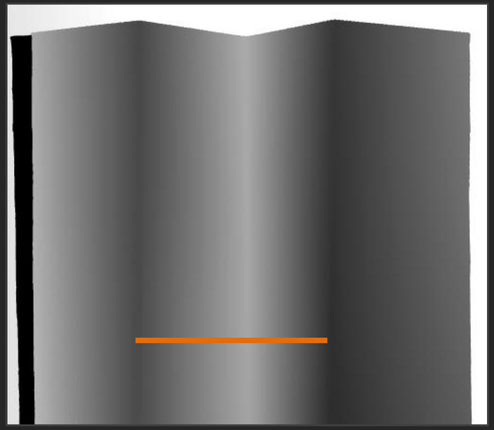
Logical Coding and Decoding



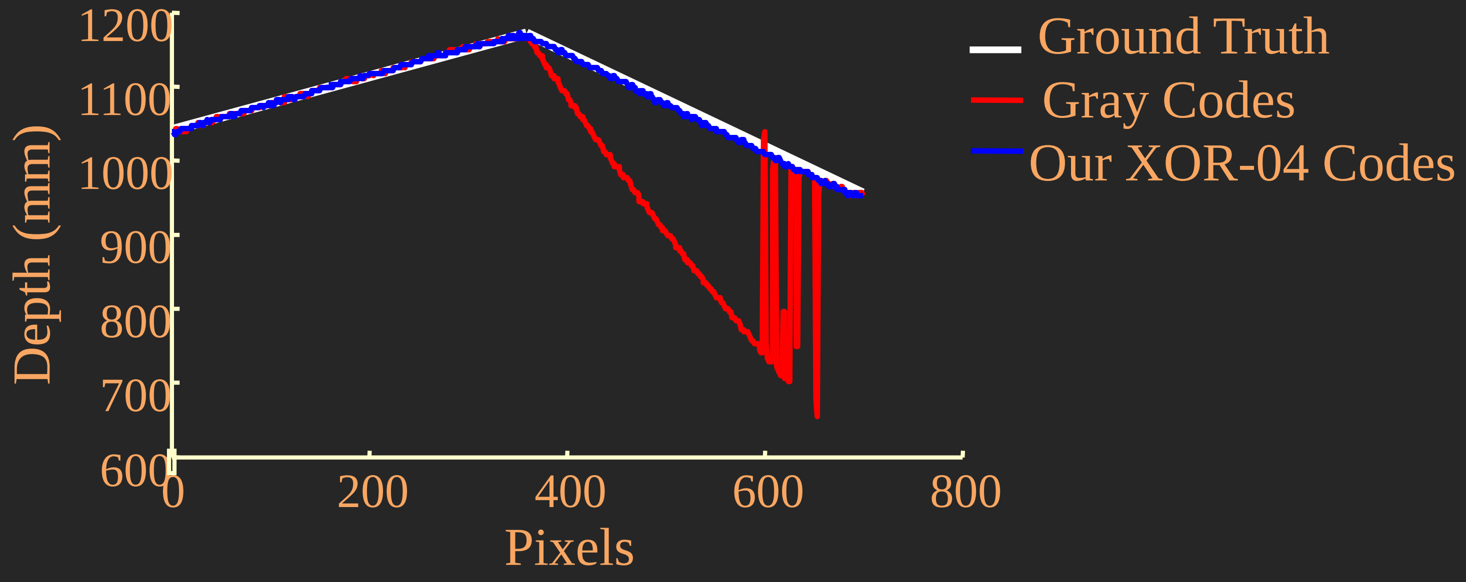
Depth Map Comparison



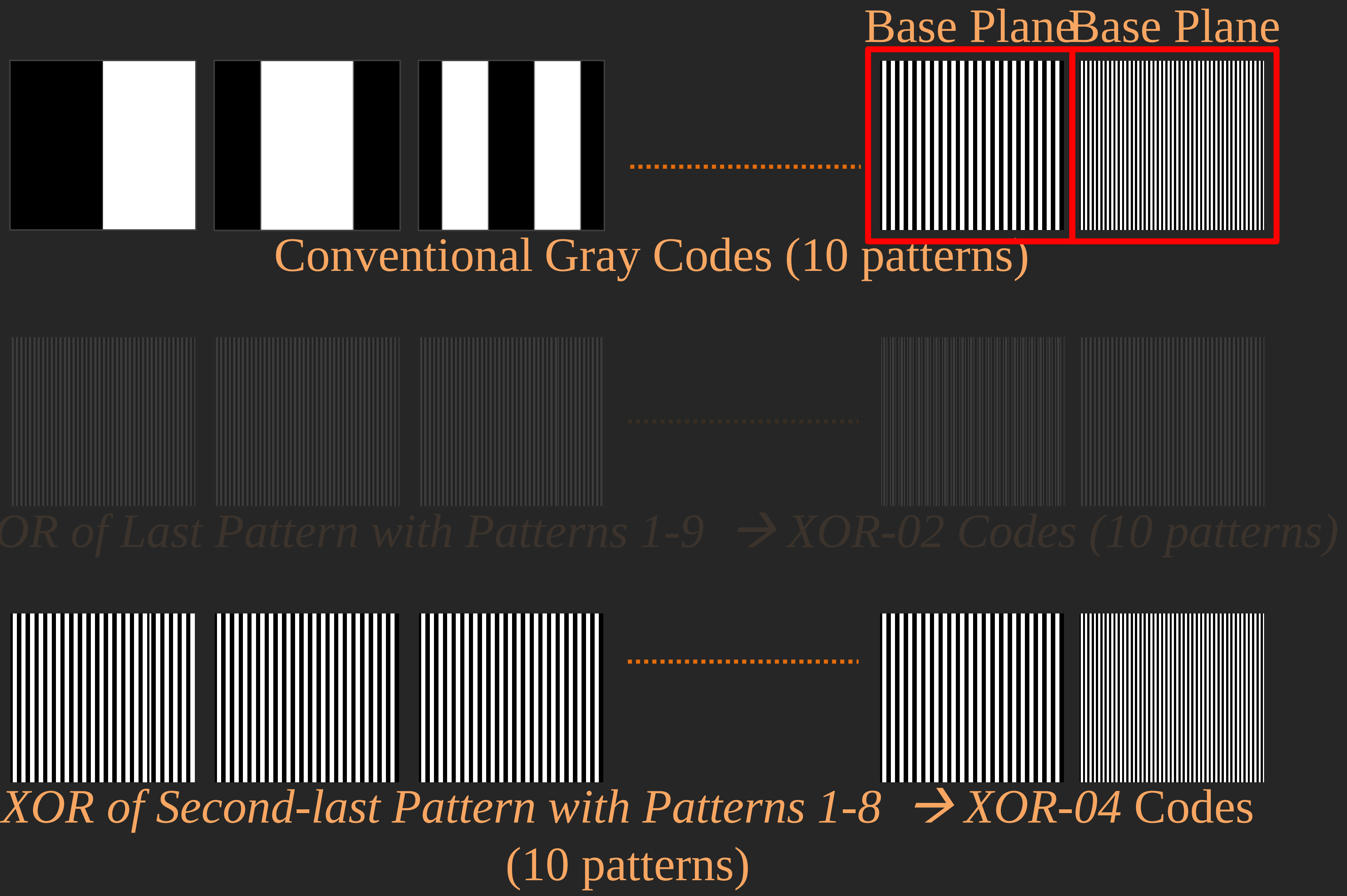
Conventional Gray Codes (11 images)



Our XOR-04 Codes (11 images)



Making the Logical XOR Codes

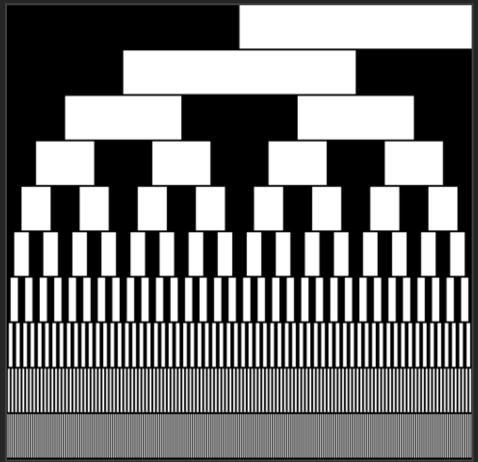


Gray Codes with Low Spatial Frequencies

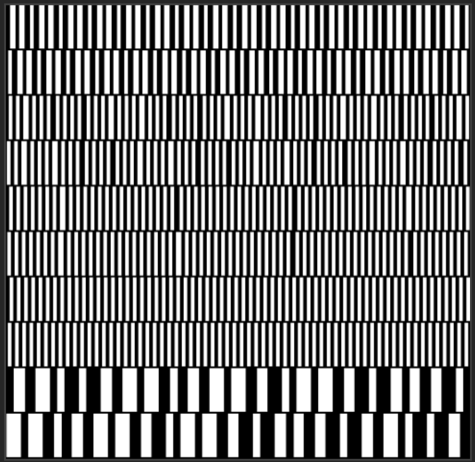
Max min-stripe-width Gray Codes

Conventional Gray Codes

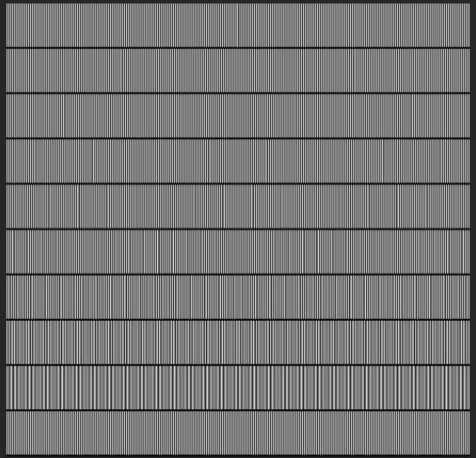
Ensemble of Codes for General Scenes



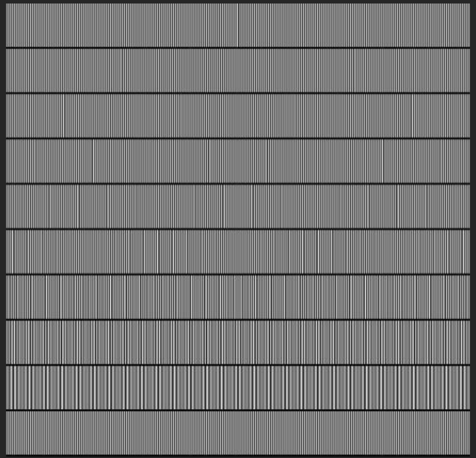
Conventional Gray (10 images)



Max min-SW Gray (10 images)

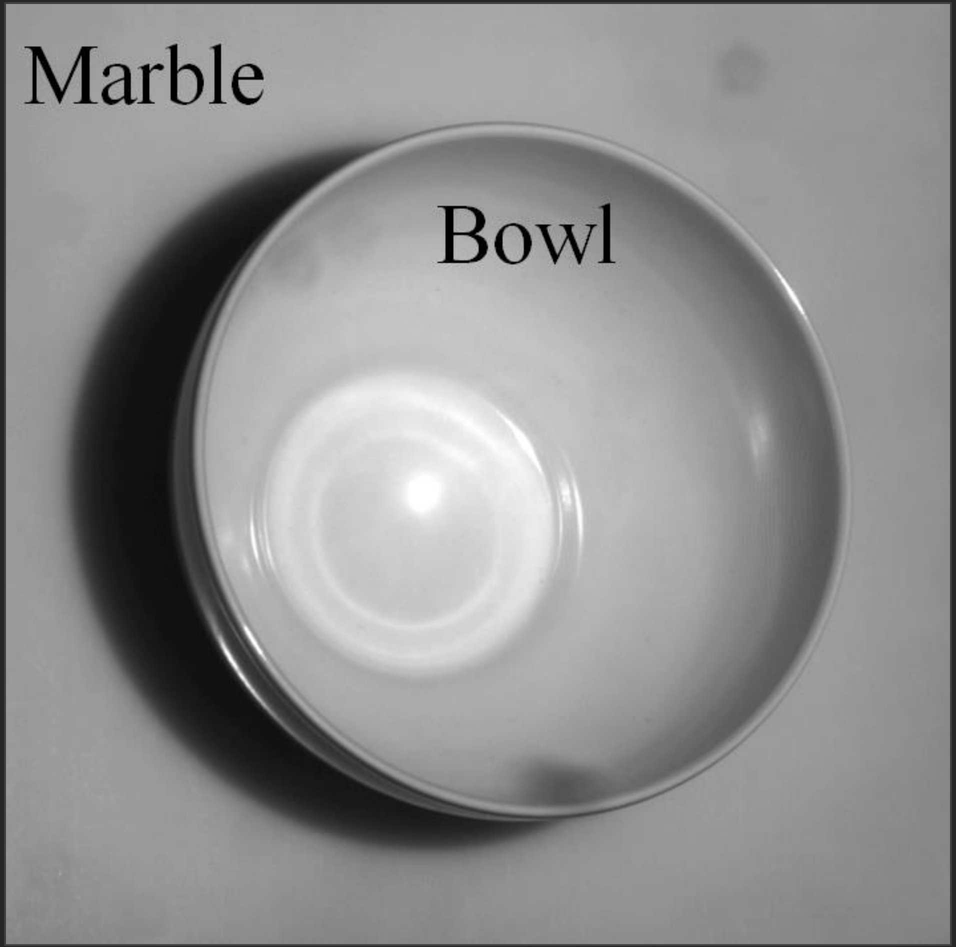


XOR-04 (10 images)

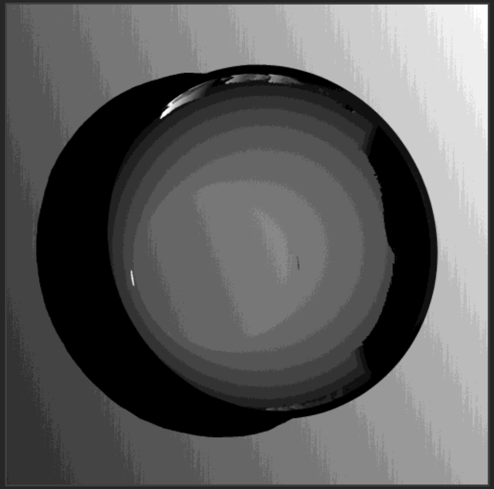


XOR-02 (10 images)

Reconstructing General Scenes

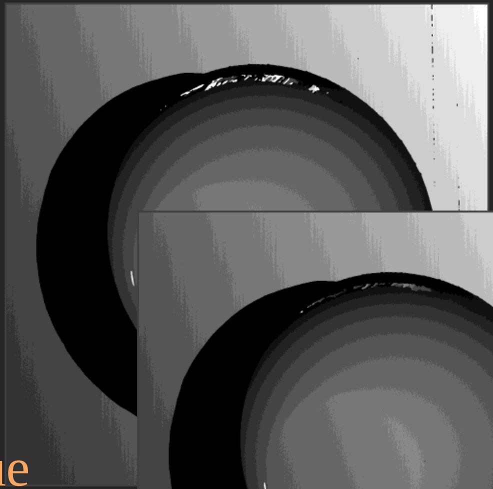


Ensemble of Codes for General Scenes

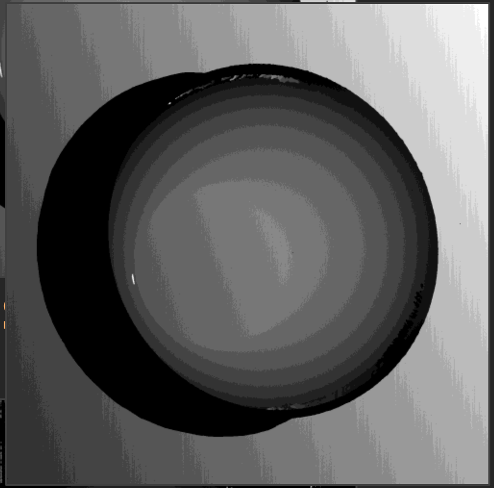


Conventional Gray (10 images)

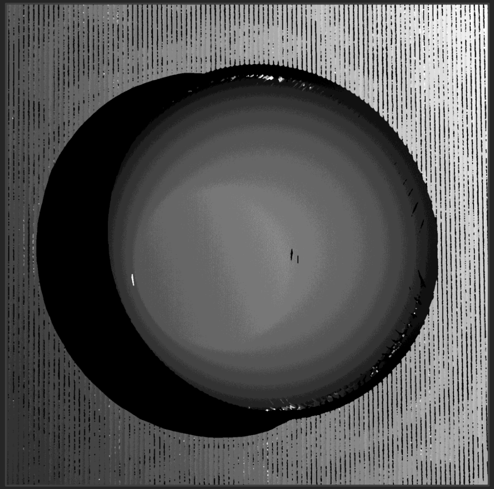
Return the consistent value



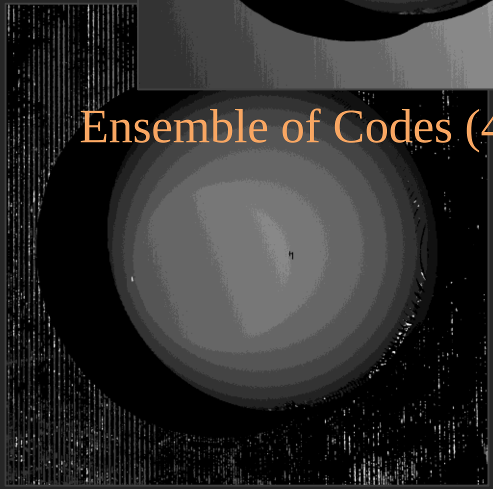
Max min-



Ensemble of Codes (41 images)

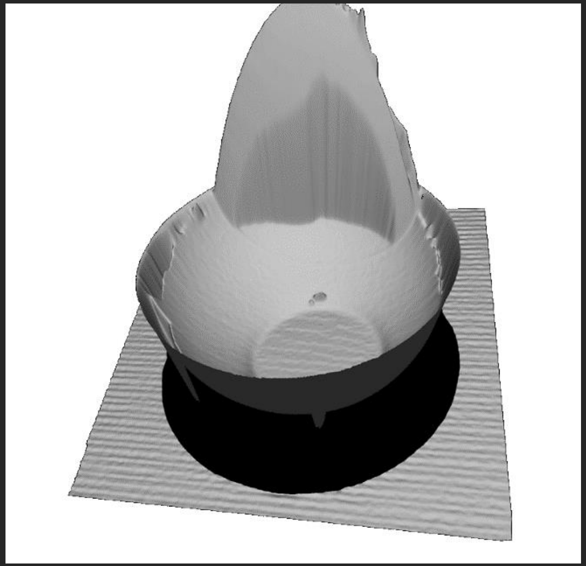


XOR-04 (10 images)

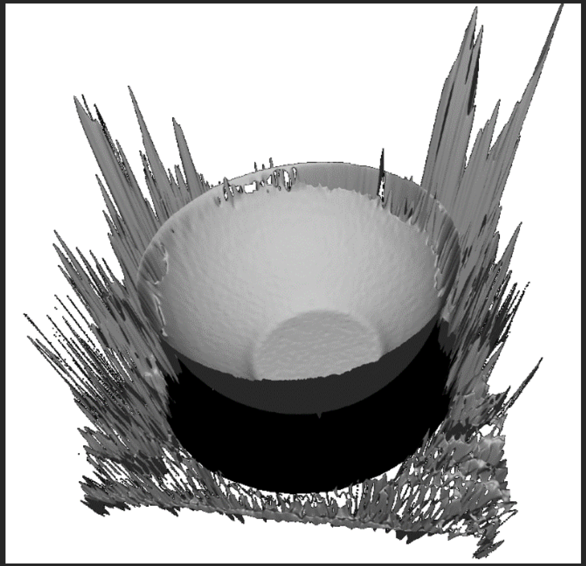


XOR-02 (10 images)

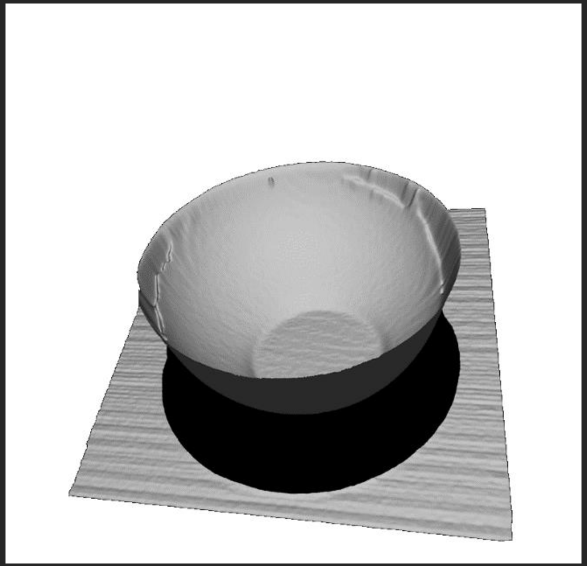
Shape Comparison



Conventional Gray
(11 images)

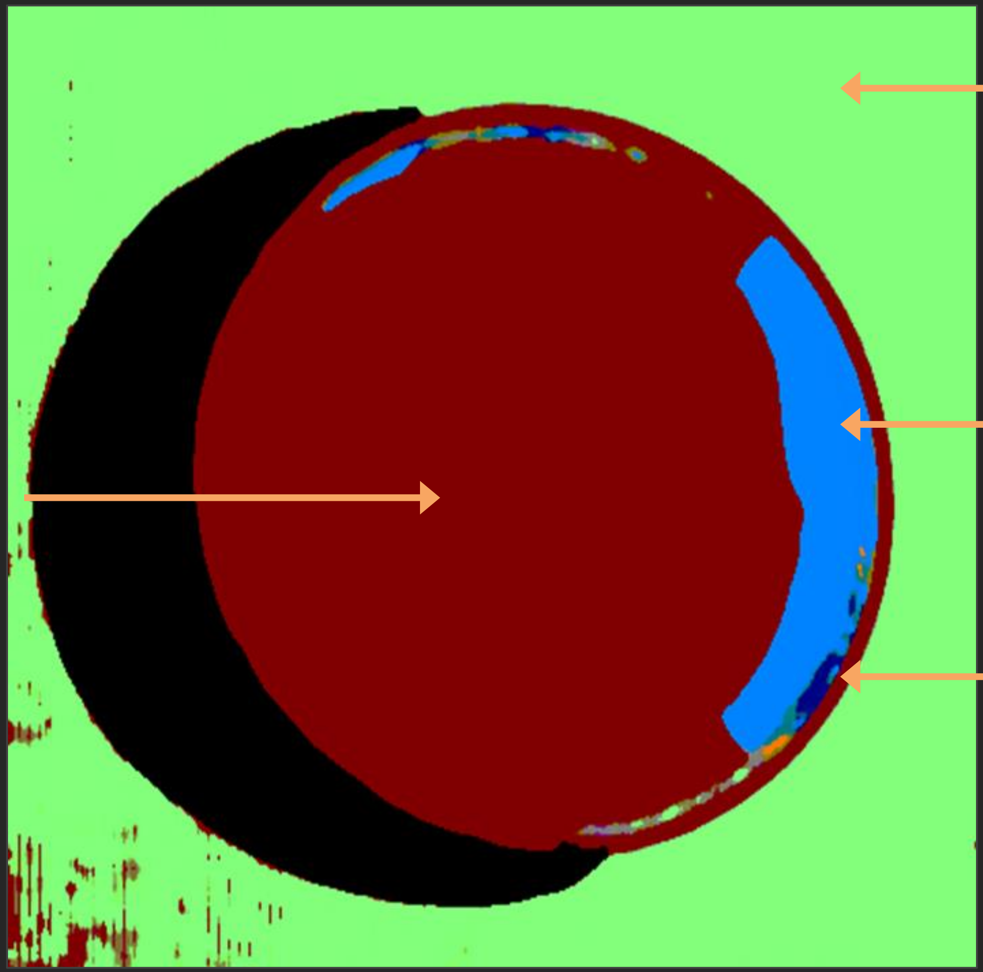


Modulated Phase-Shifting
(162 images)

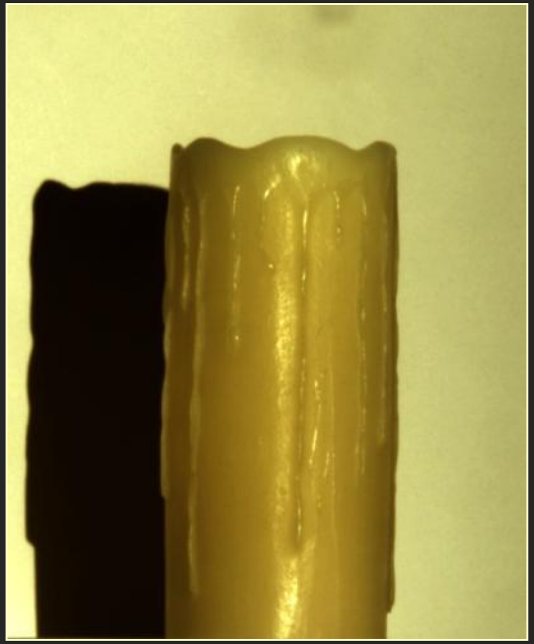


Our Technique
(41 images)

Qualitative Light Transport Analysis

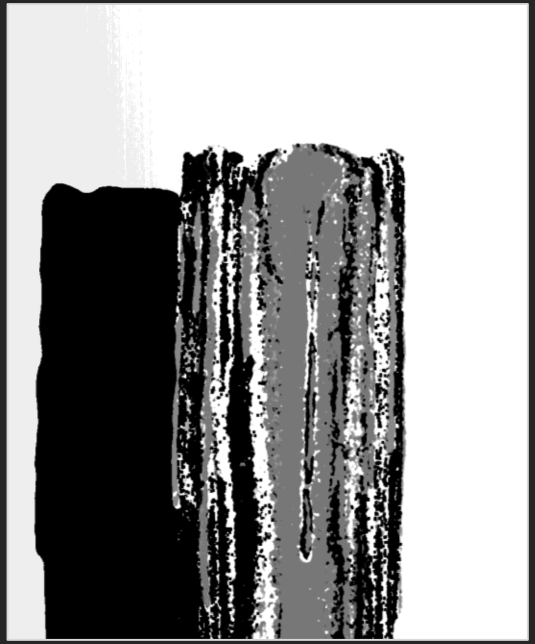


Translucent Wax Candle

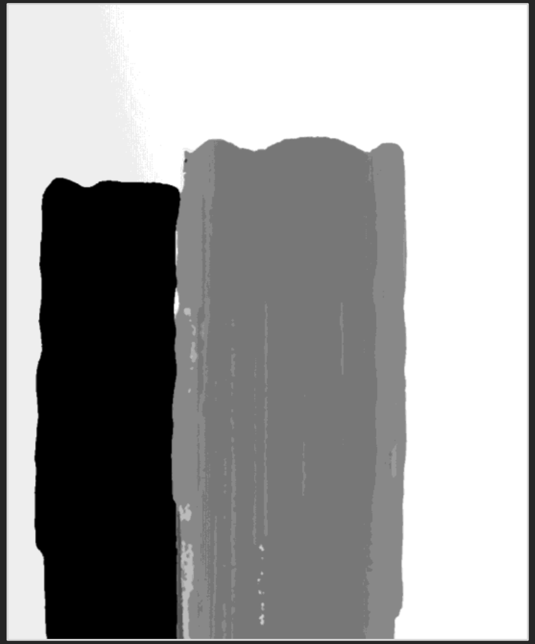


Scene

Errors due to strong
sub-surface scattering



Modulated Phase-
Shifting (162 images)

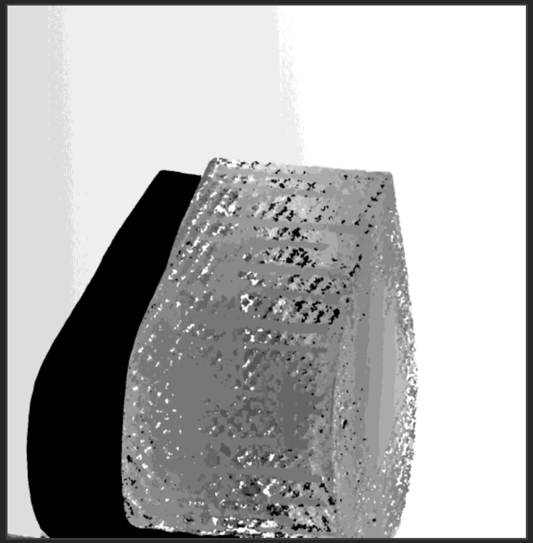


Our Ensemble Codes
(41 images)

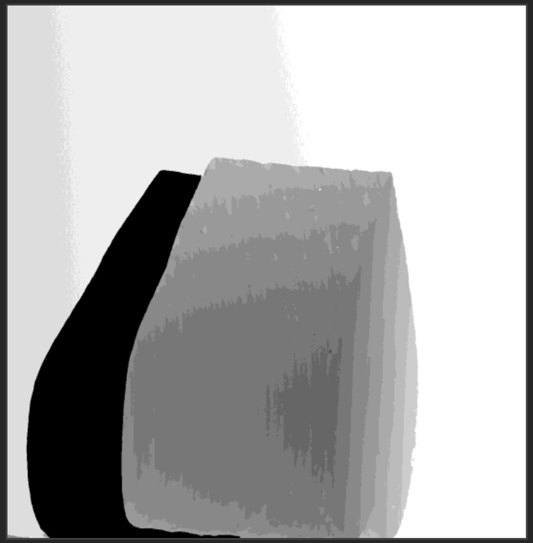
Translucent Wax Object

Errors due to strong sub-surface scattering

Scene



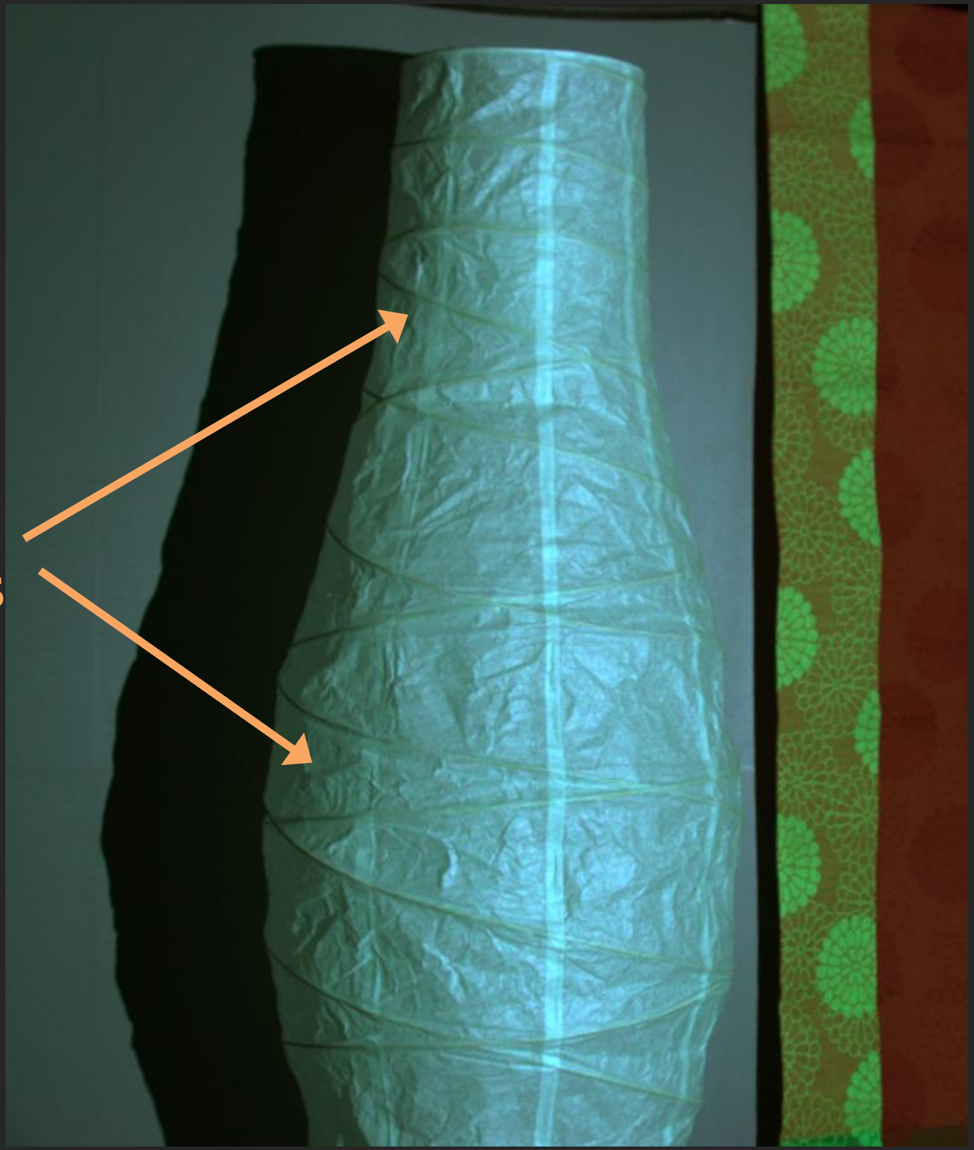
Modulated Phase-Shifting (162 images)



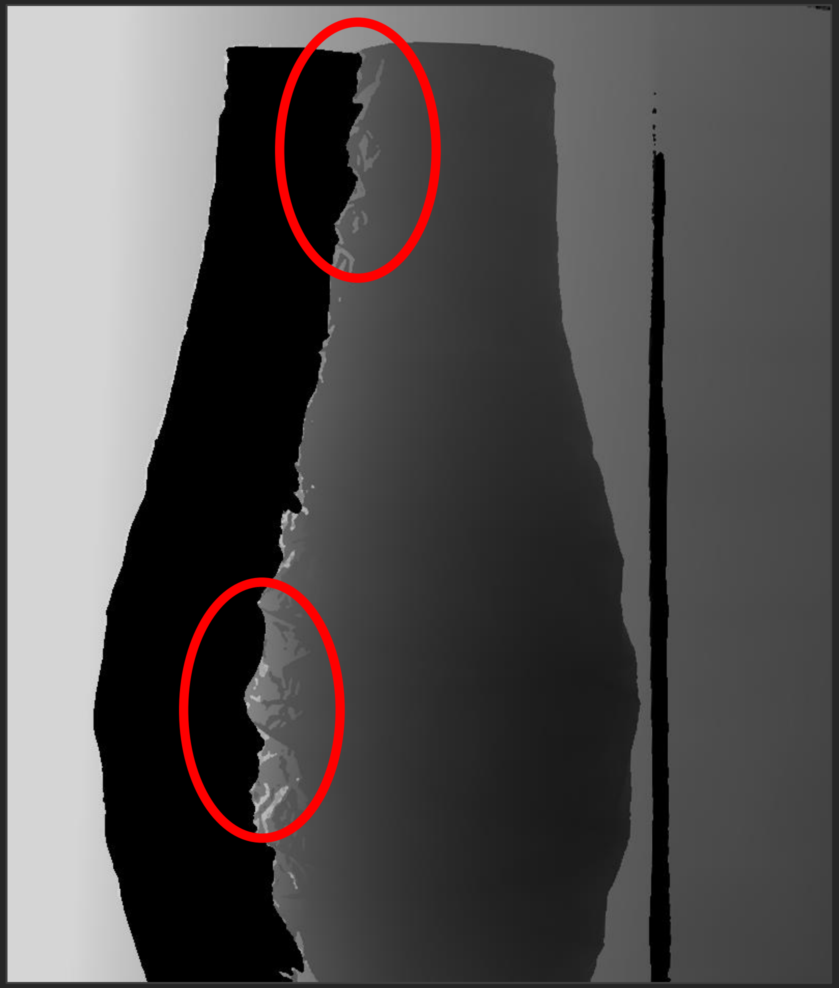
Our Ensemble Codes (41 images)

Ikea Lamp

Diffusion +
Inter-reflections



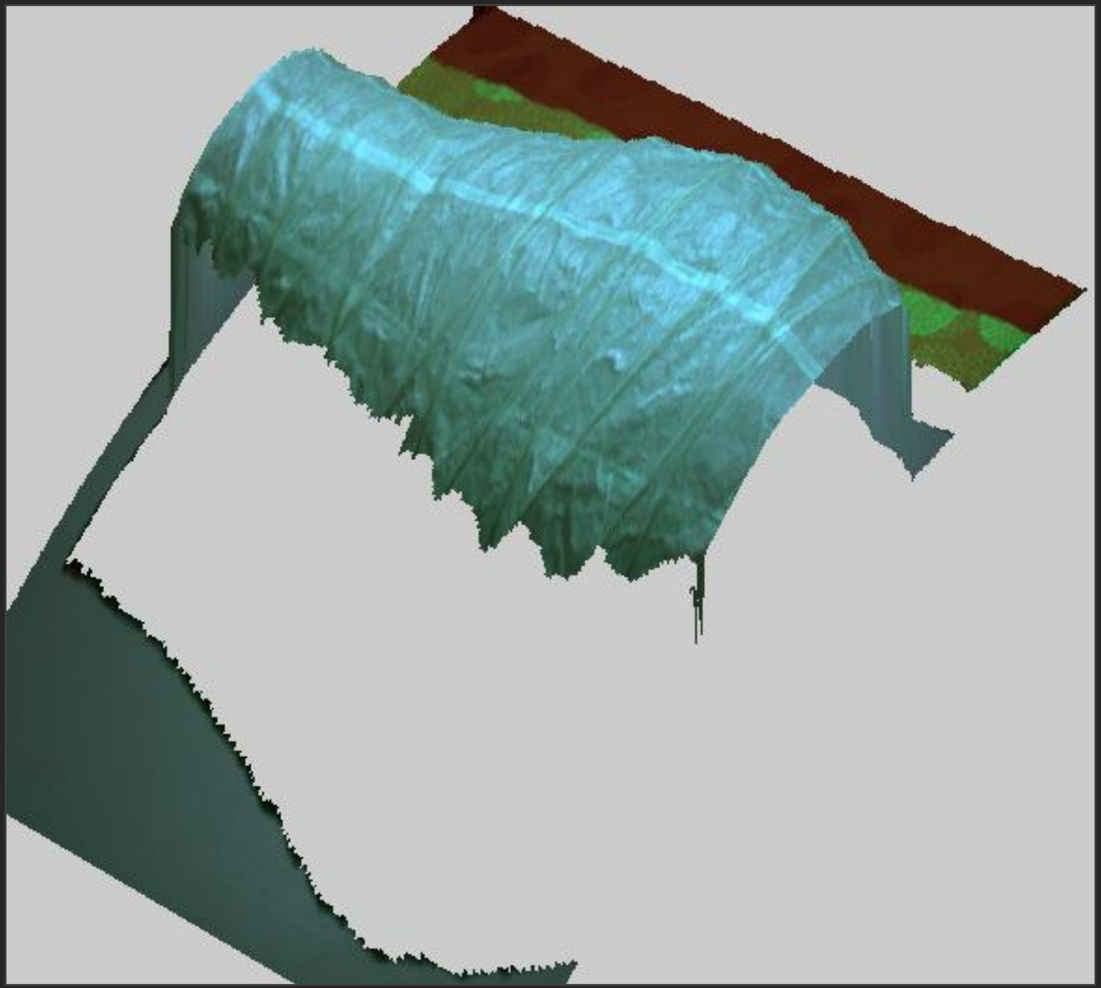
Depth-Map Comparison



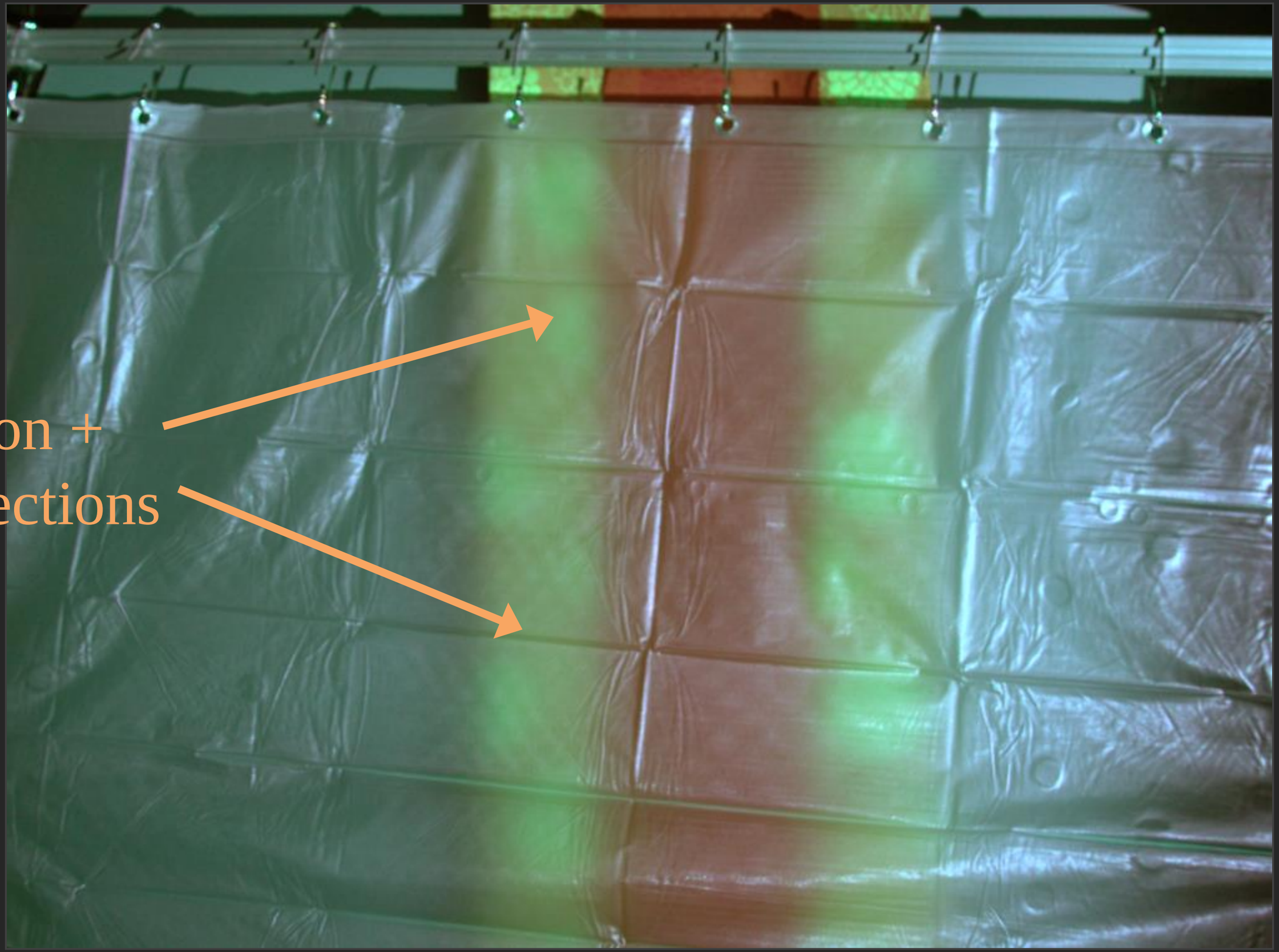
Regular Gray Codes (11 images)

Our Ensemble Codes (41 images)

3D Visualization using our ensemble codes



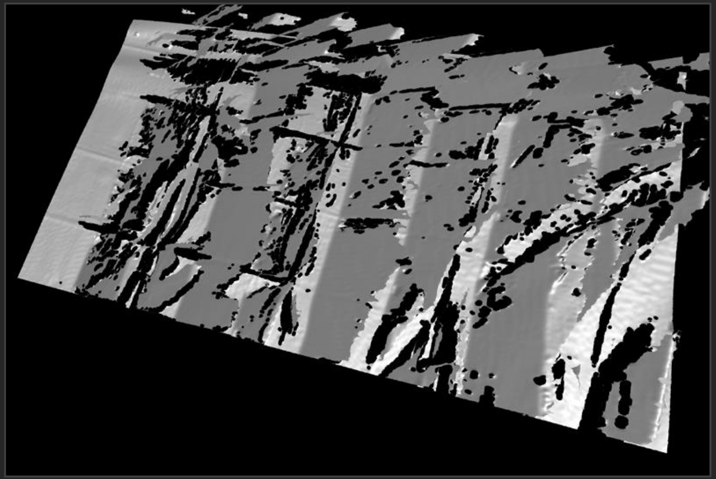
Shower Curtain



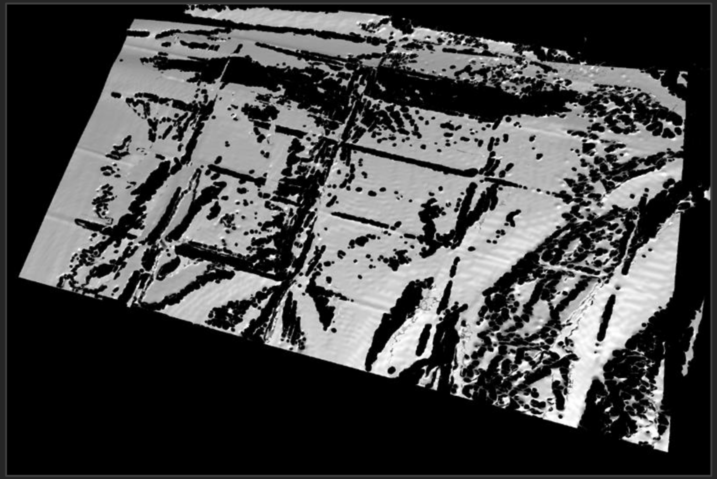
Diffusion + Inter-reflections

Goal is to reconstruct the shape of the shower-curtain. Shape of the curtain is planar because it was taped to the rod to avoid movement while capture.

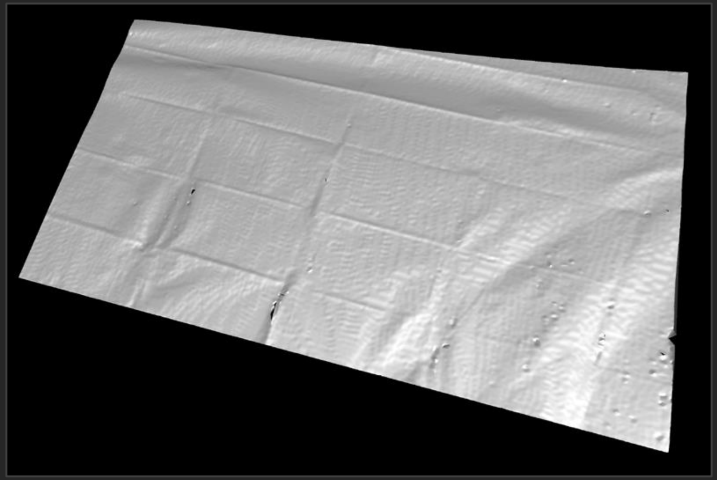
Shape Comparisons



Regular Gray Codes (11 images)

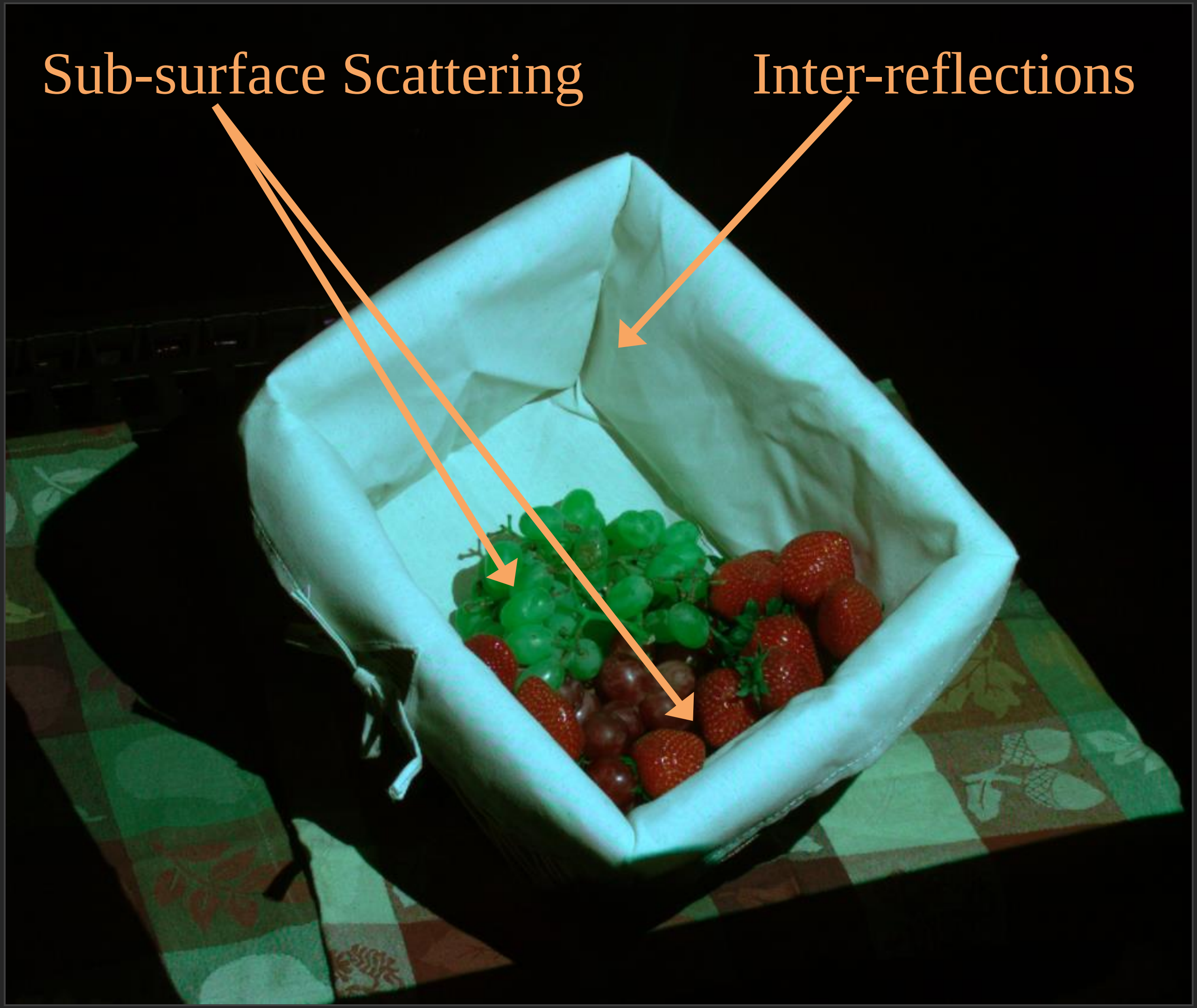


Phase-Shifting (18 images)

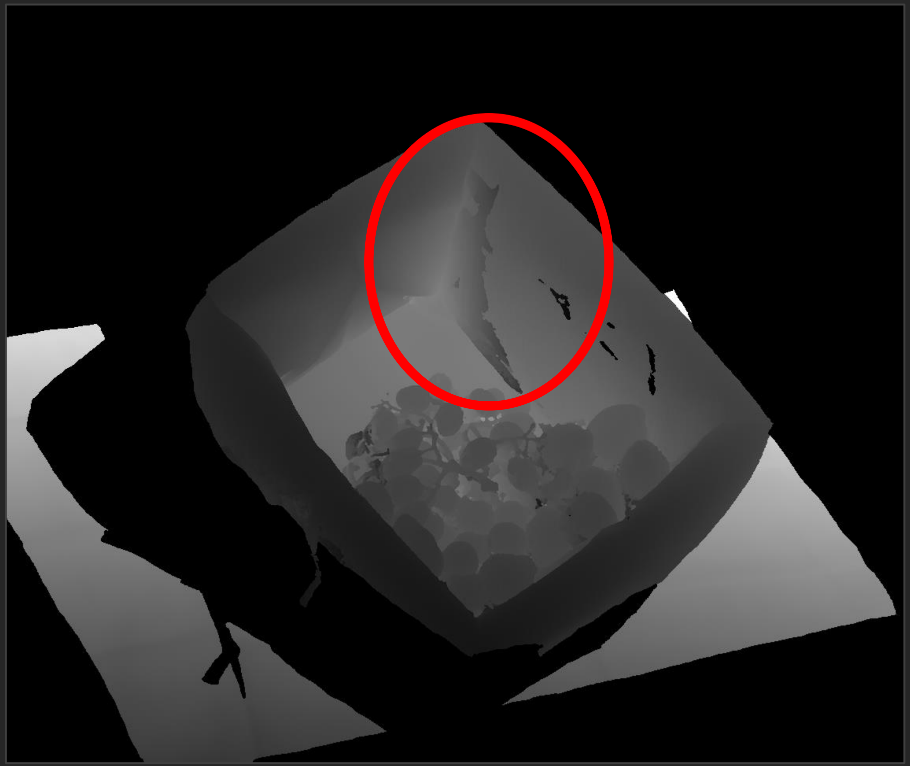


Our XOR Codes (11 images)

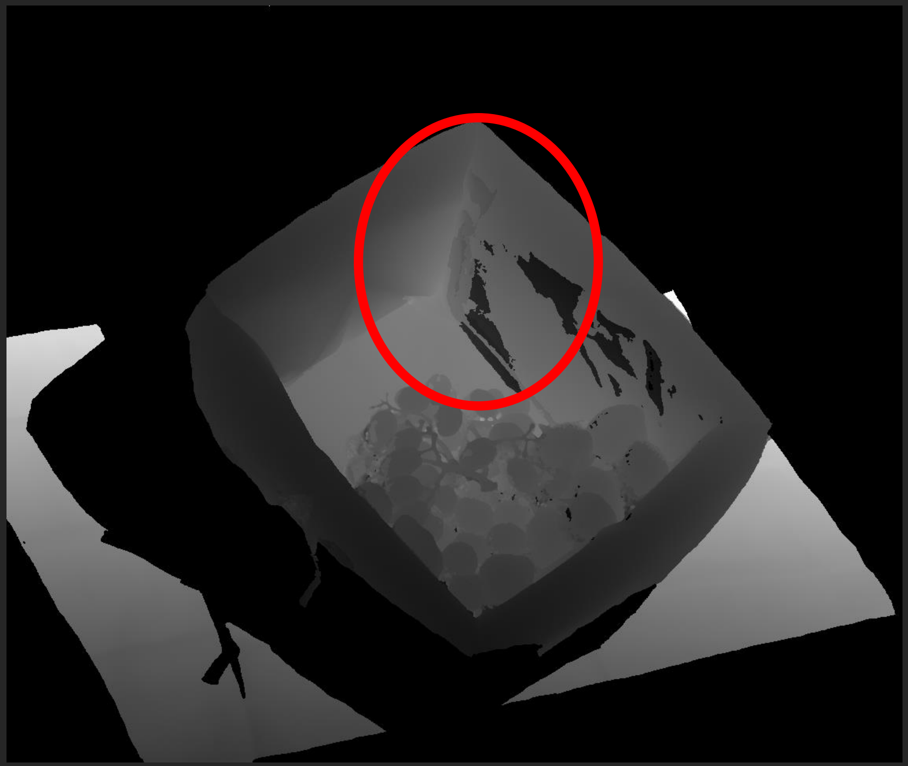
Fruit Basket: Multiple Effects



Depth-maps with previous state of the art

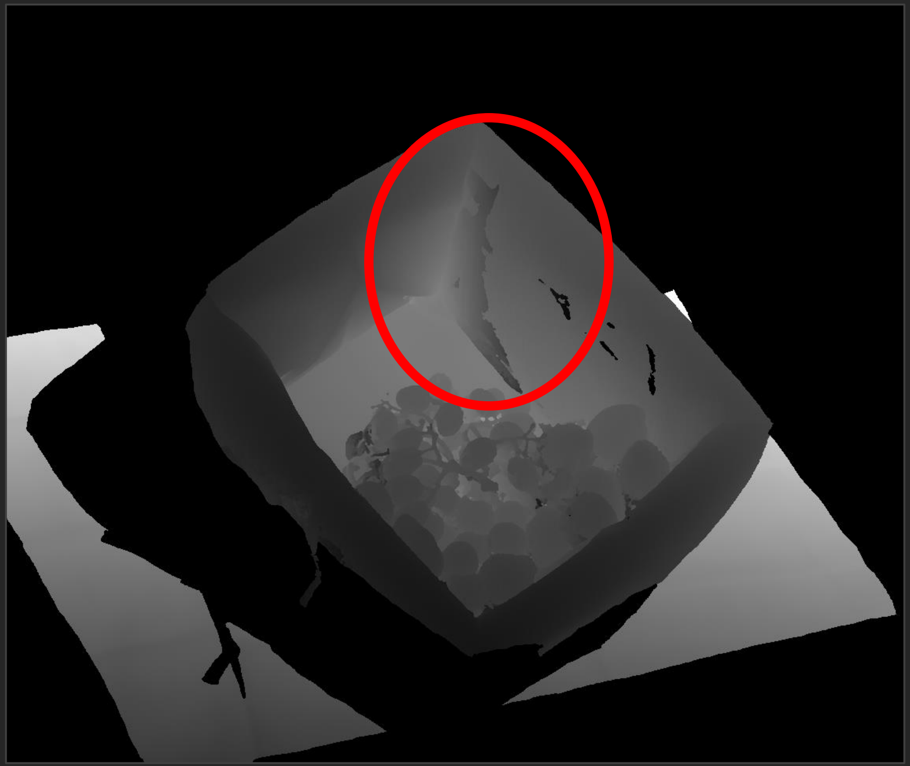


Regular Gray (11 images)

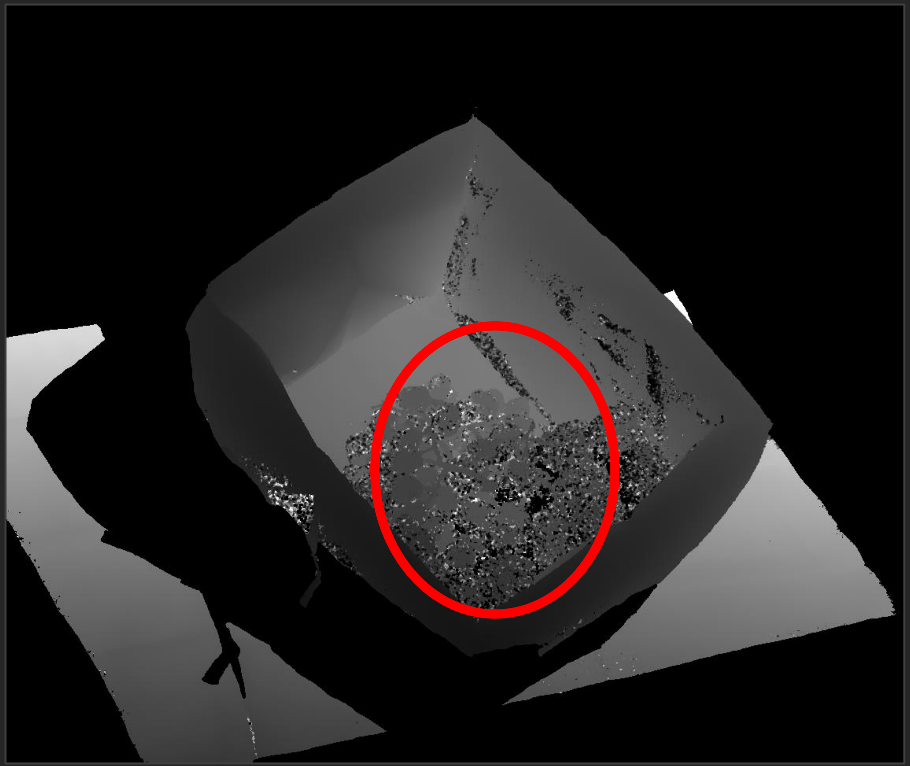


Phase-Shifting (18 images)

Depth-maps with previous state of the art

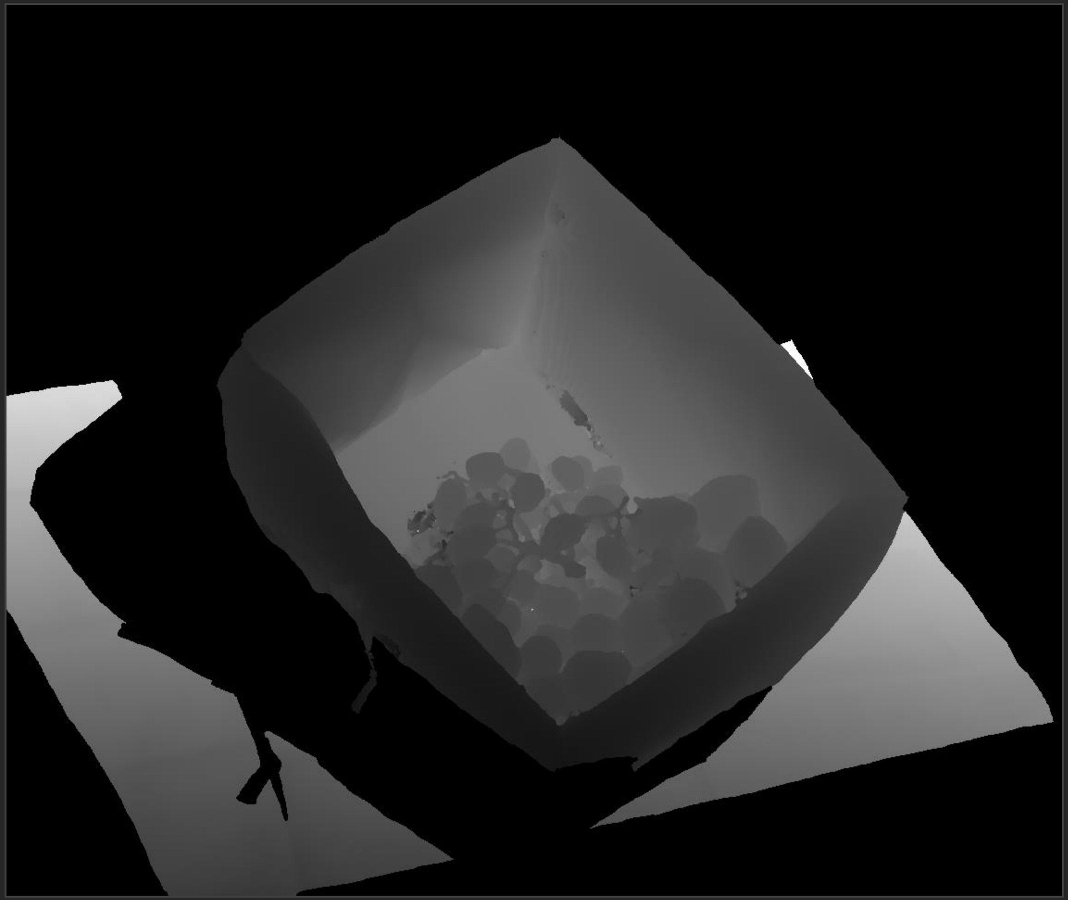


Regular Gray (11 images)



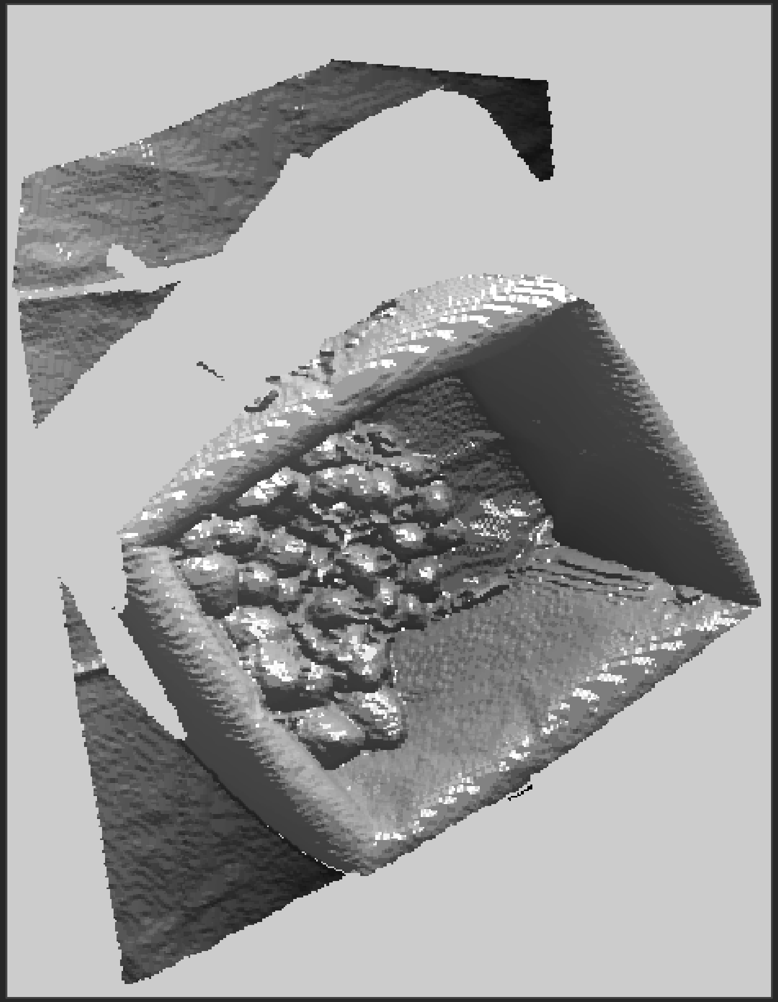
Modulated Phase-Shifting (162 images)

Depth-maps with our Ensemble Codes



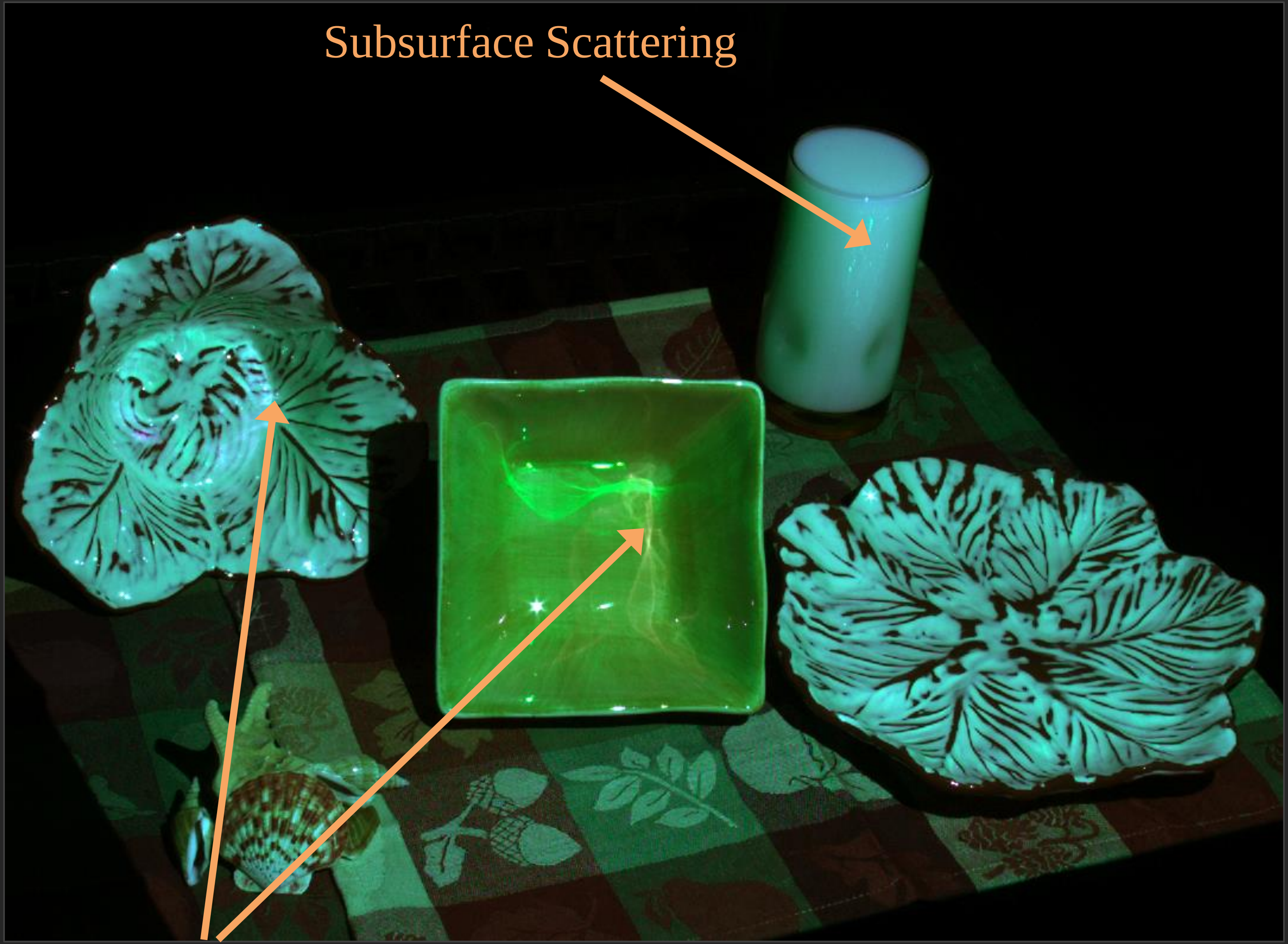
Our Ensemble Codes (41 images)

3D Visualizations with our ensemble codes

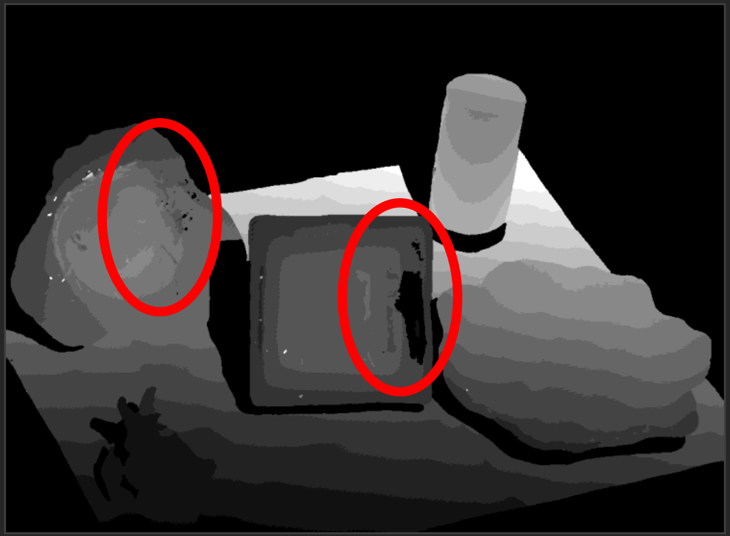


3D Visualization with our ensemble codes

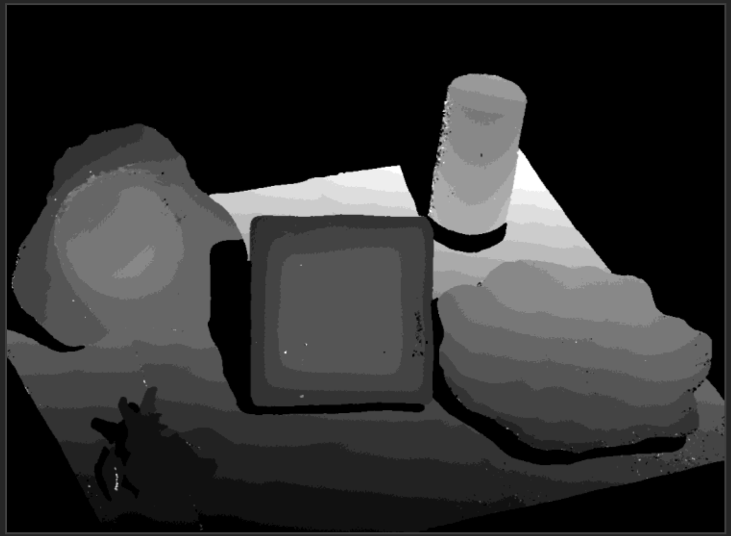
Bowls and Milk: Multiple Effects



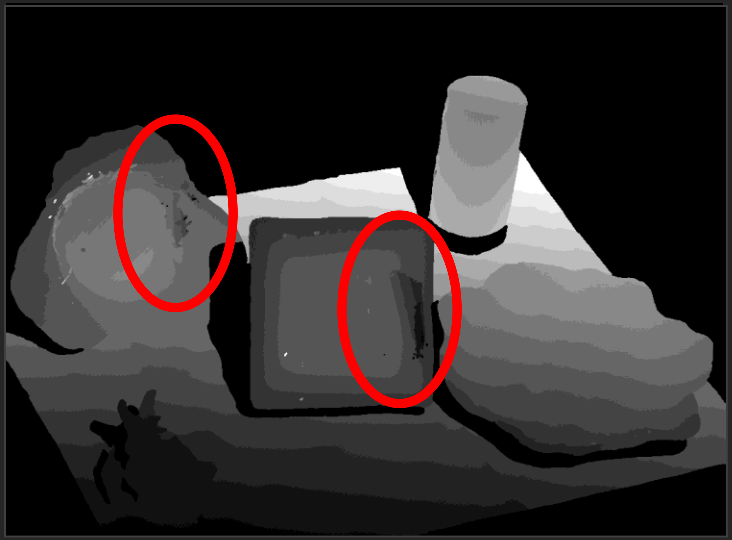
Interreflections



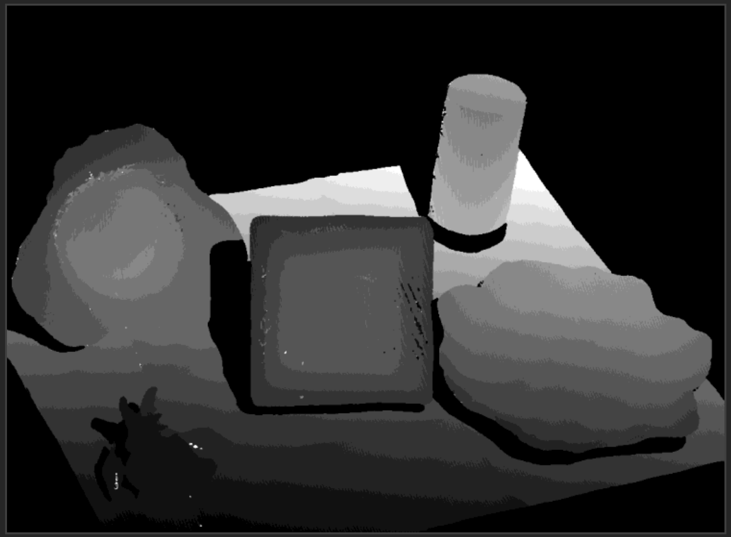
Phase-Shifting (18 images)



Modulated Phase-Shifting (162 images)

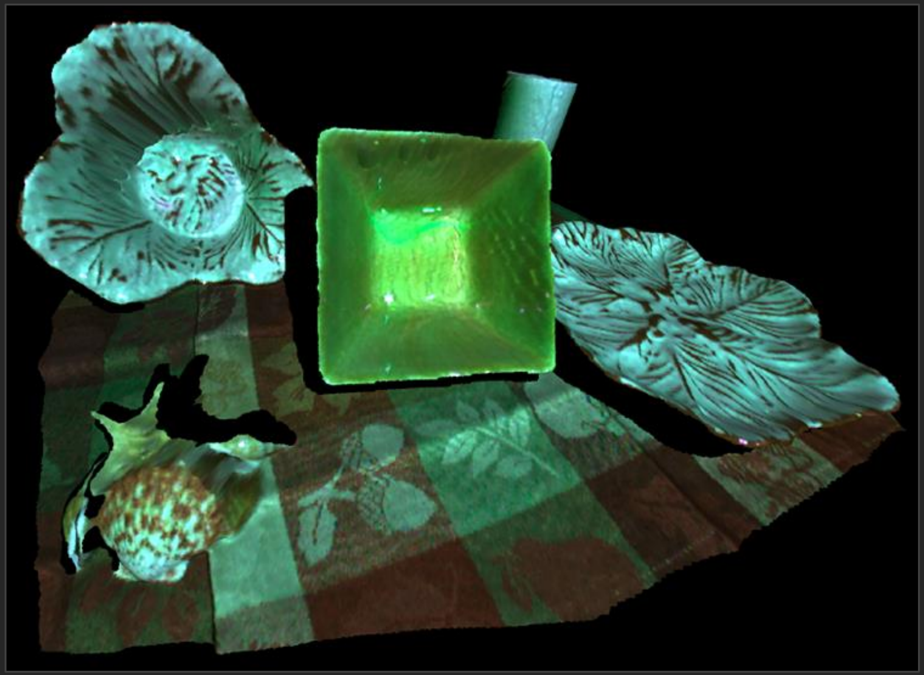


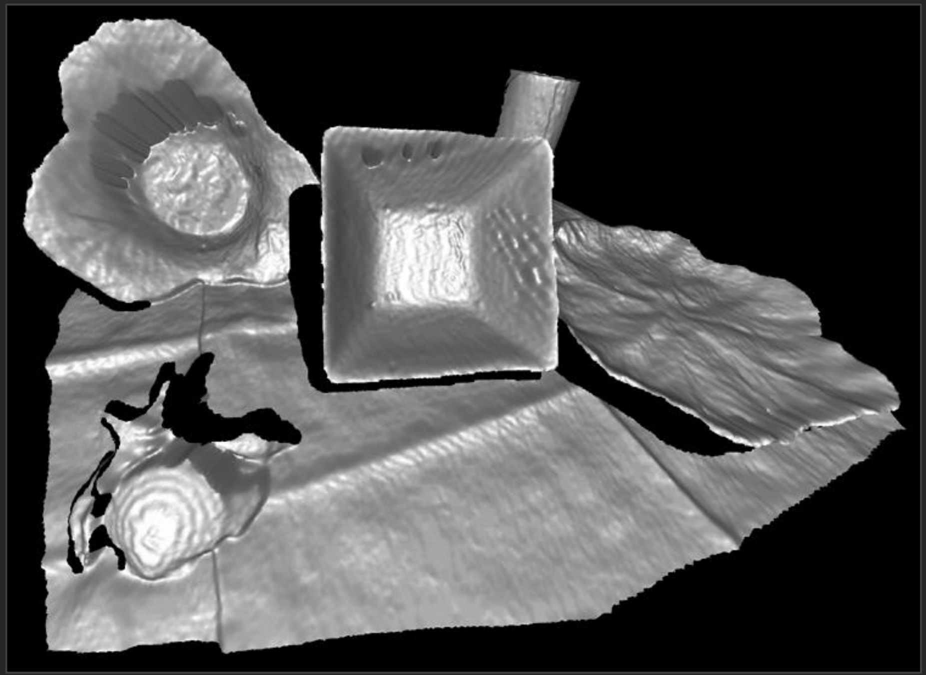
Regular Gray Codes (11 images)



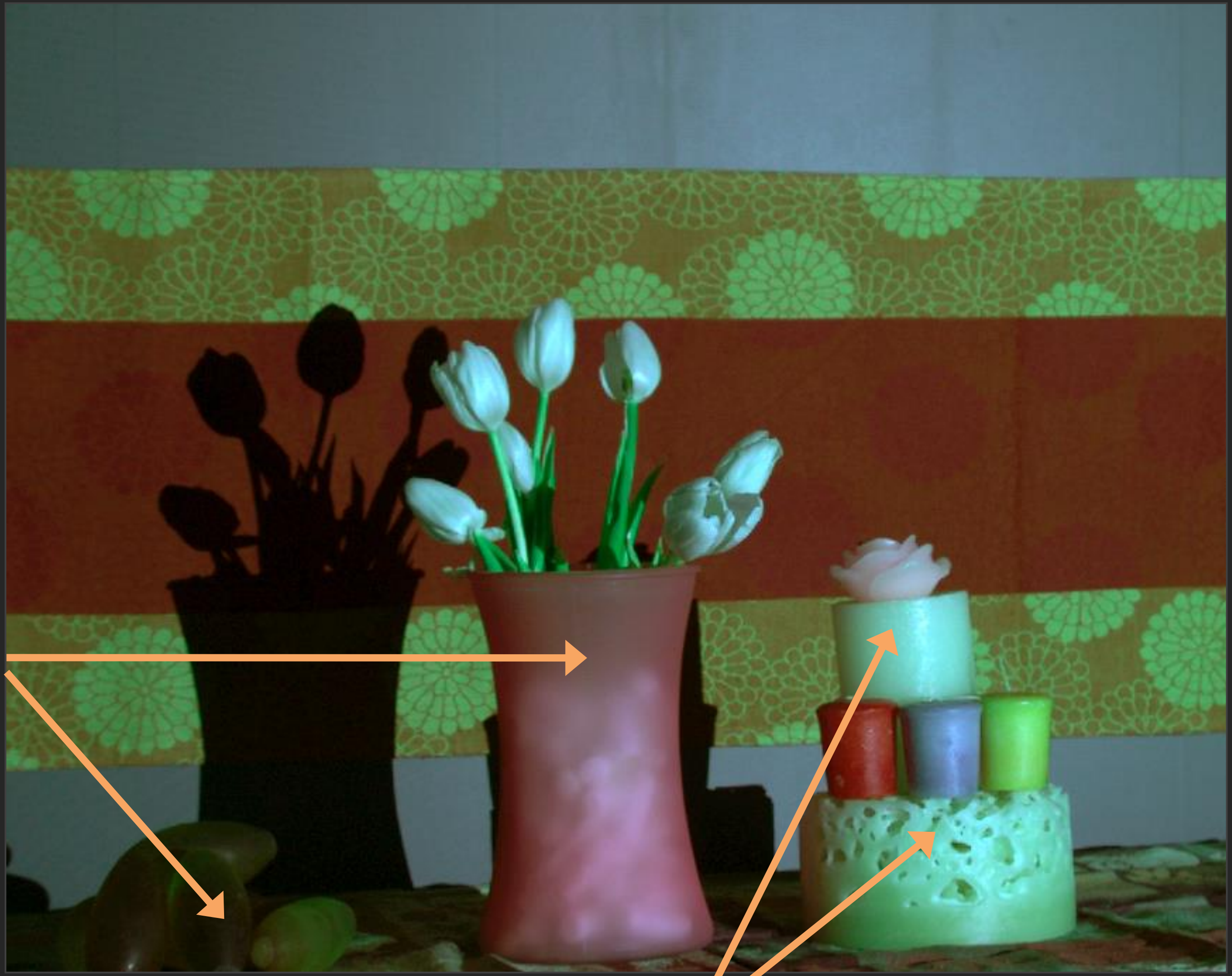
Our XOR Codes (11 images)

3D Visualizations with our ensemble codes





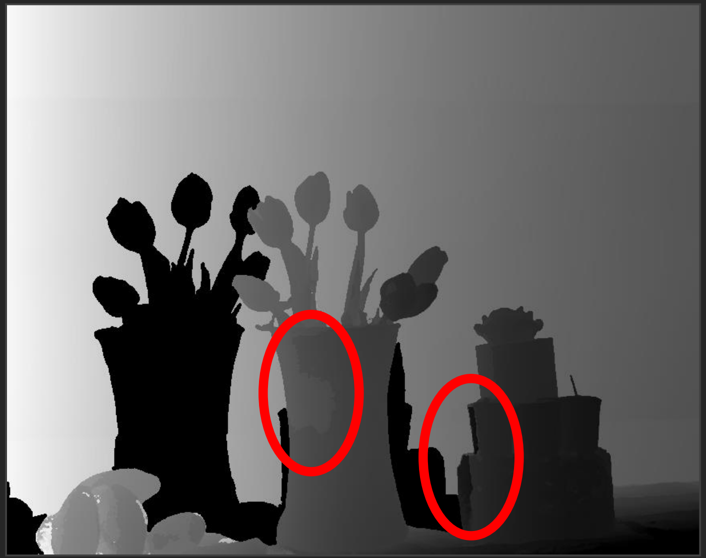
Flower-Vase



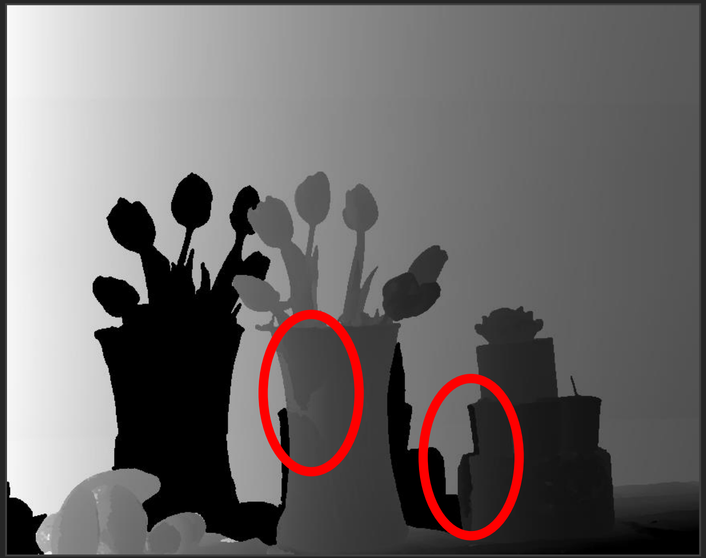
Diffusion

Sub-surface Scattering

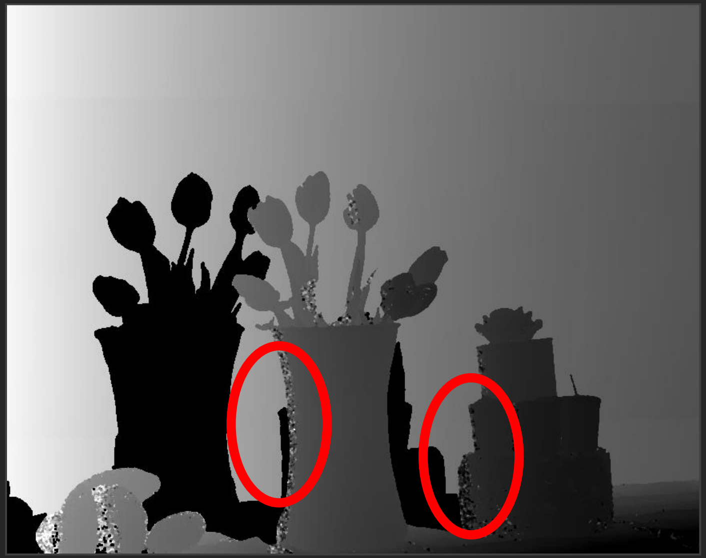
Comparison



Phase-Shifting (18 images)



Regular Gray Code (11 images)



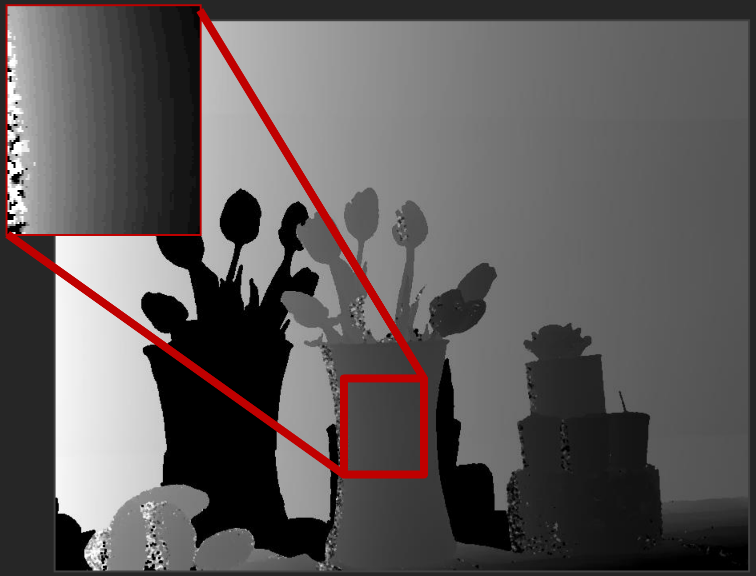
Modulated Phase-Shifting (162 images)

Our Ensemble Codes (41 images)

Comparison

Phase-Shifting (18 images)

Regular Gray Code (11 images)

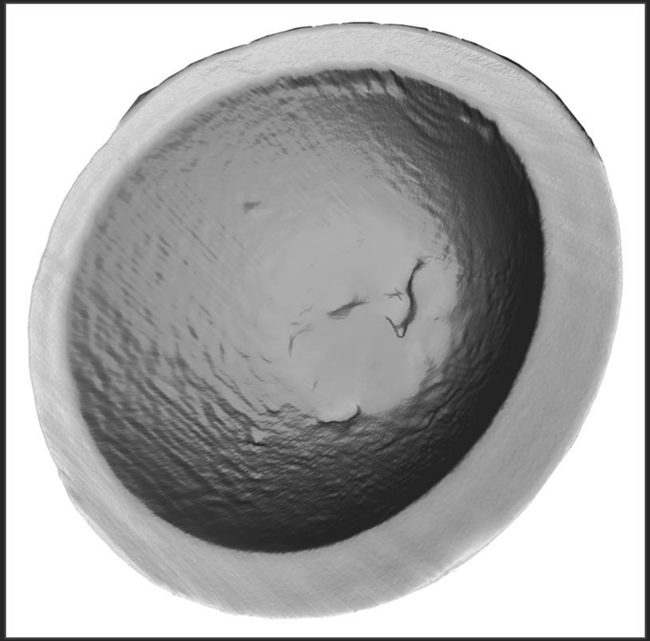


Modulated Phase-Shifting (162 images)

Our Ensemble Codes (41 images)

Multiple Global Illumination Effects

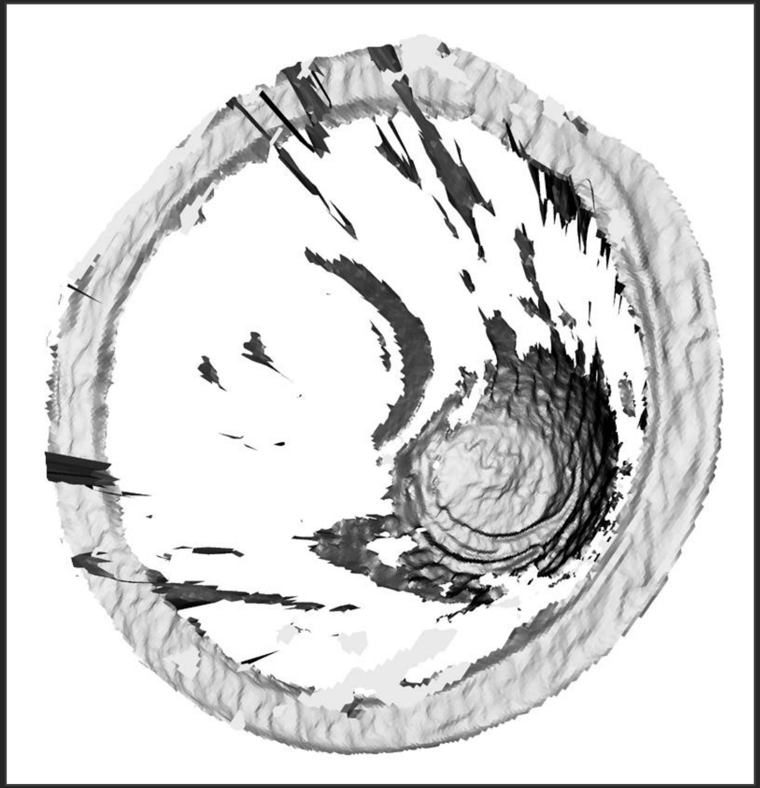
Wax Bowl



Shape Using Ensemble Codes

Multiple Global Illumination Effects

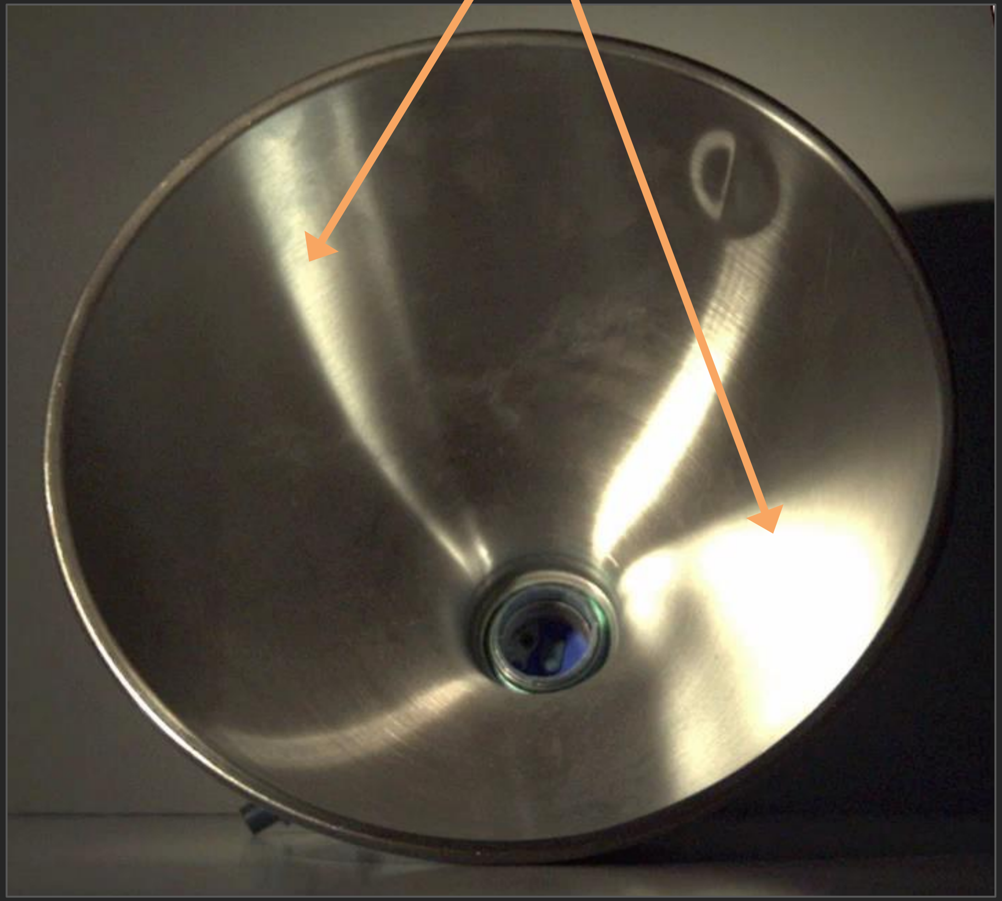
Deep Wax Container



Shape Using Ensemble Codes

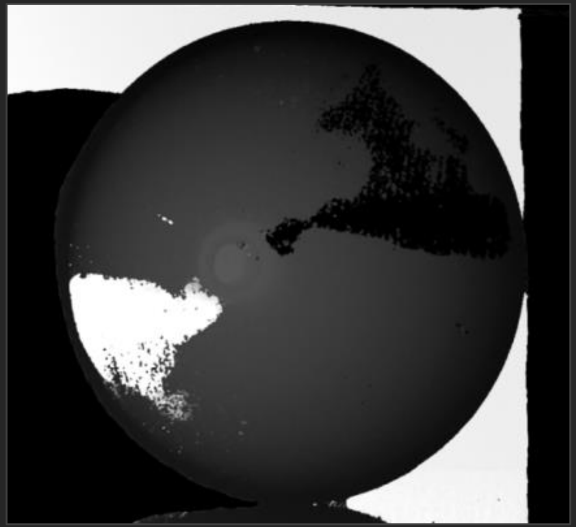
Lamp made of shiny brushed metal

Strong and high-frequency inter-reflections



Depth Map Comparison

Regular Gray (11 images)

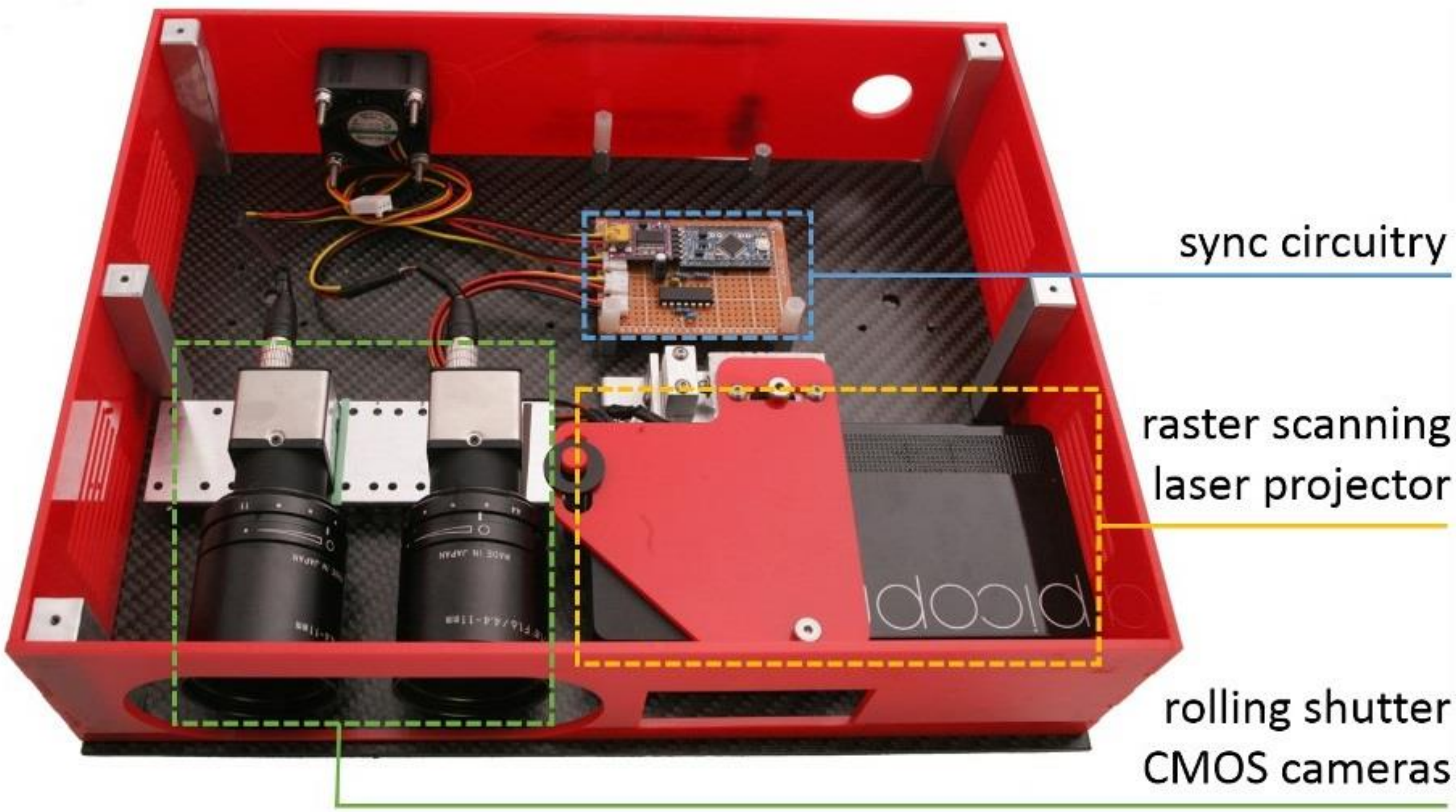


Our Ensemble Codes (41 images)

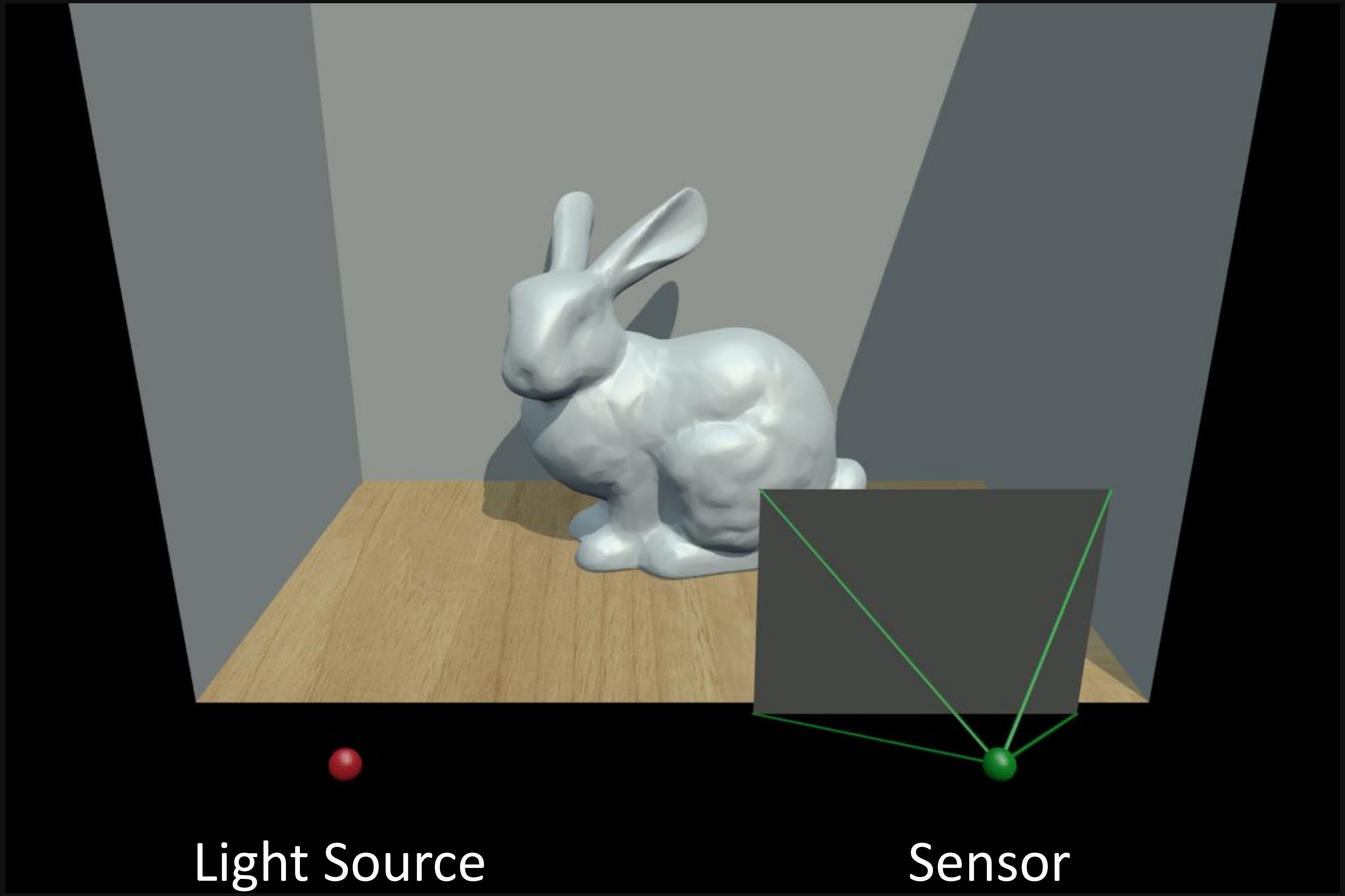
Another look at epipolar imaging

Epipolar imaging camera

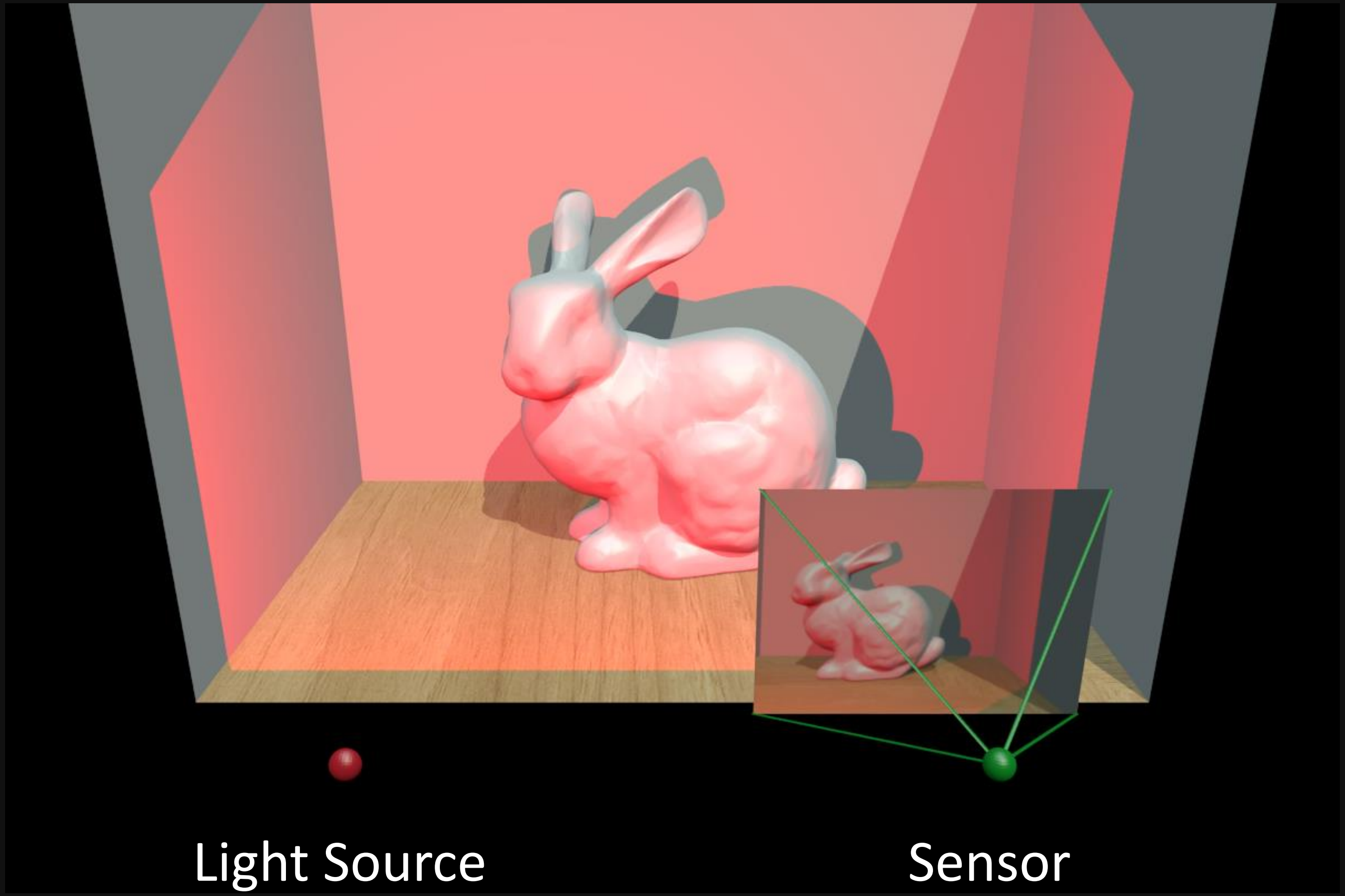
Rectified camera-projector pair.



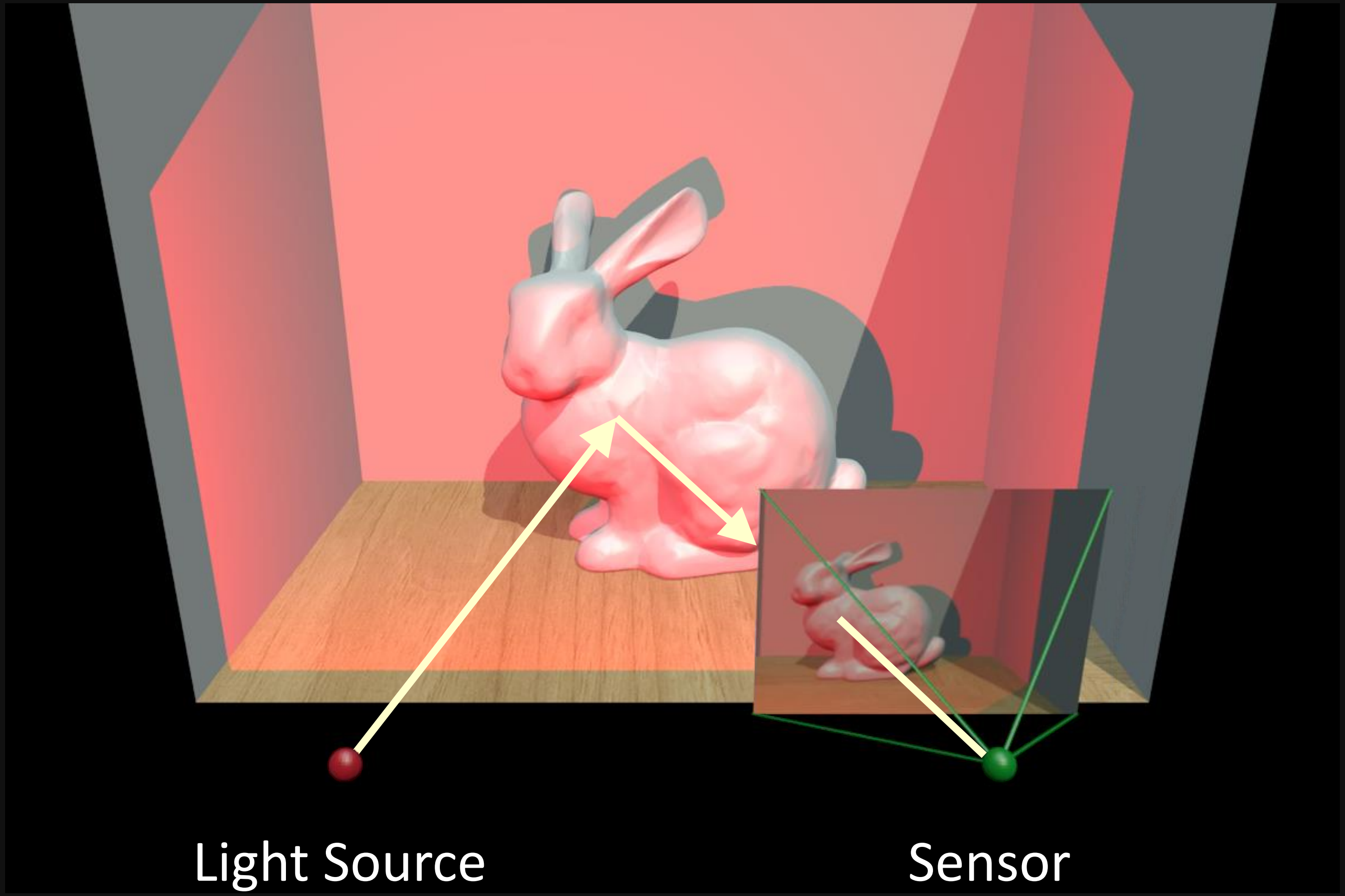
Regular Imaging



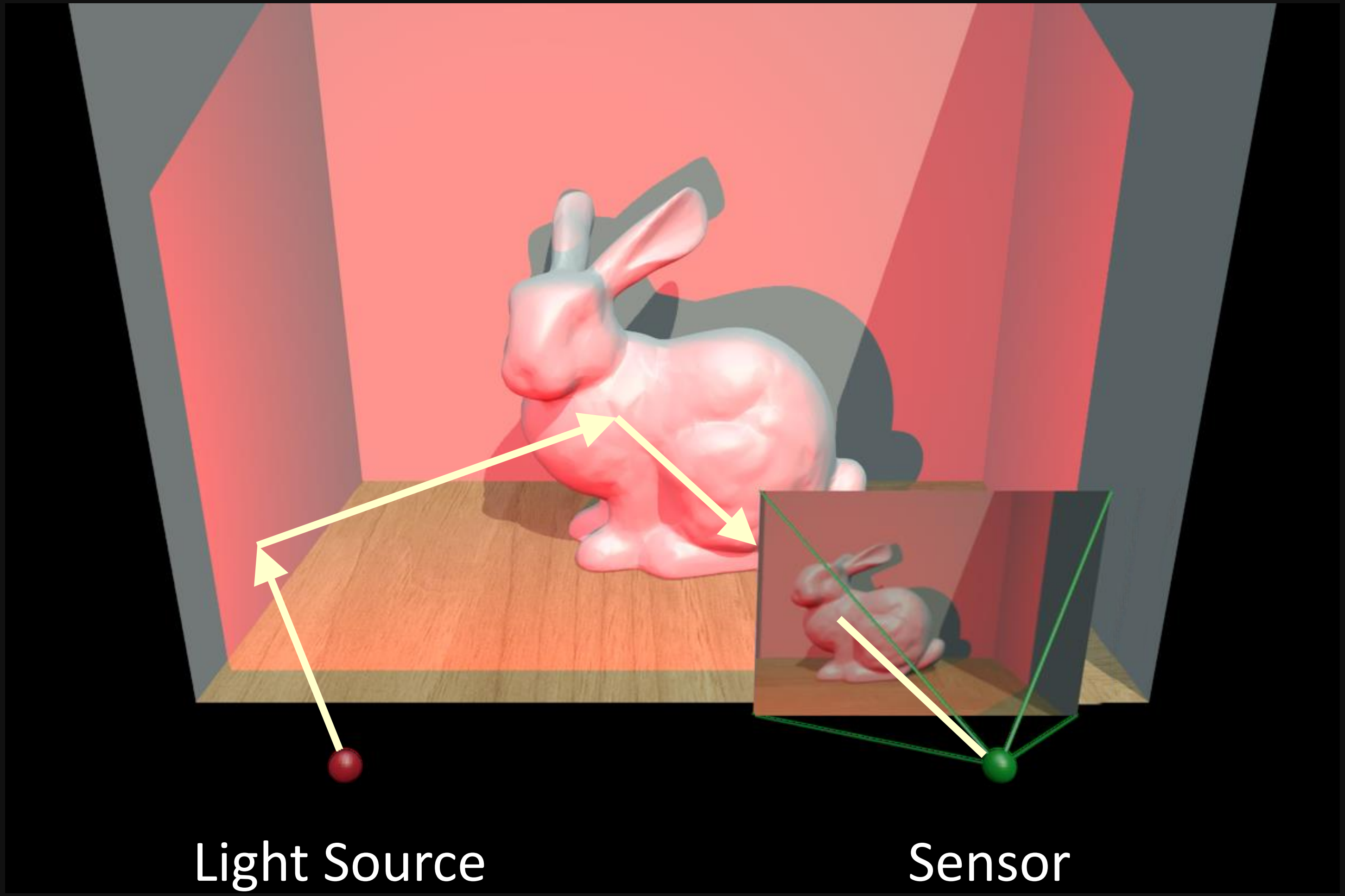
Regular Imaging



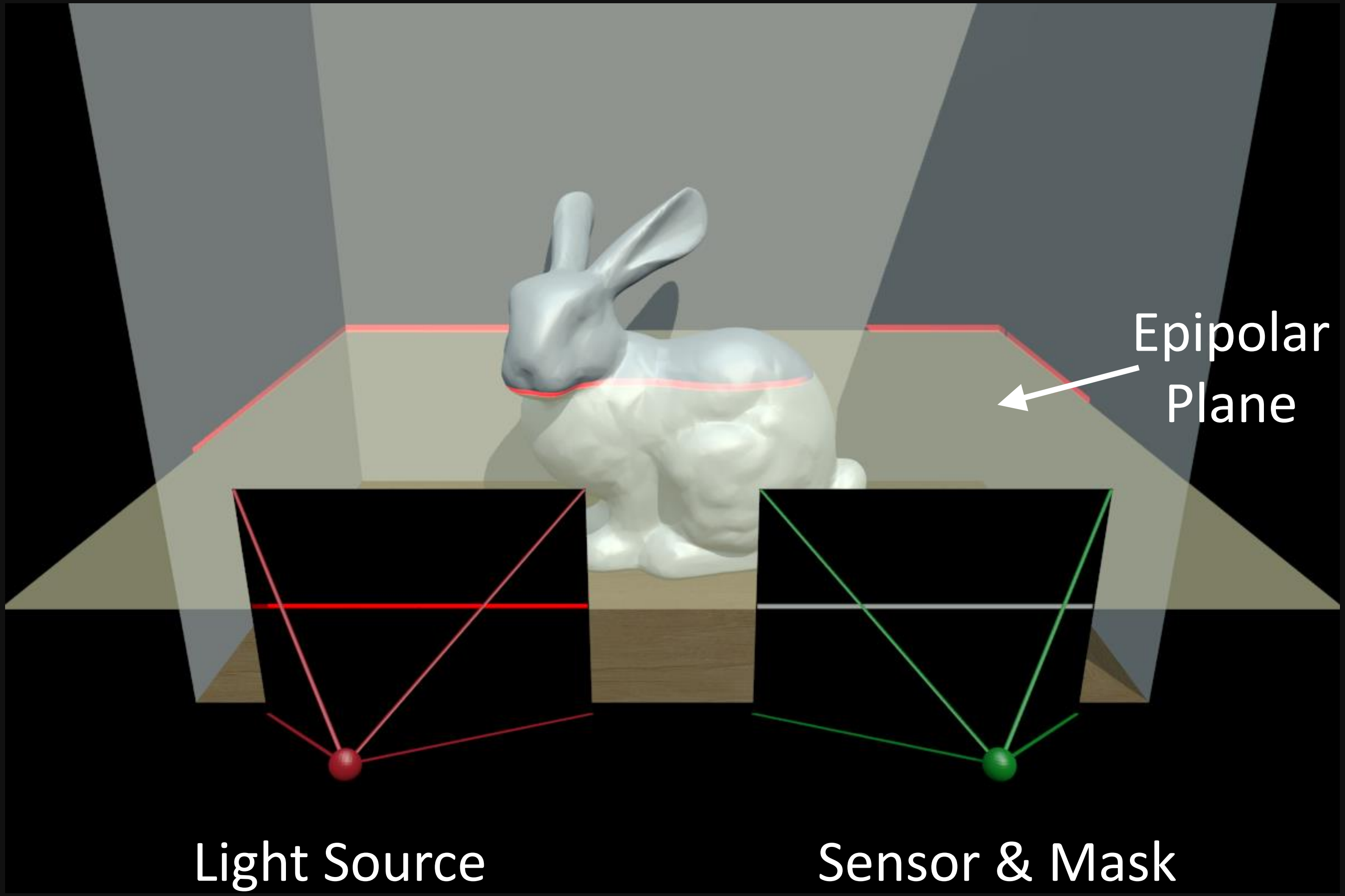
Regular Imaging



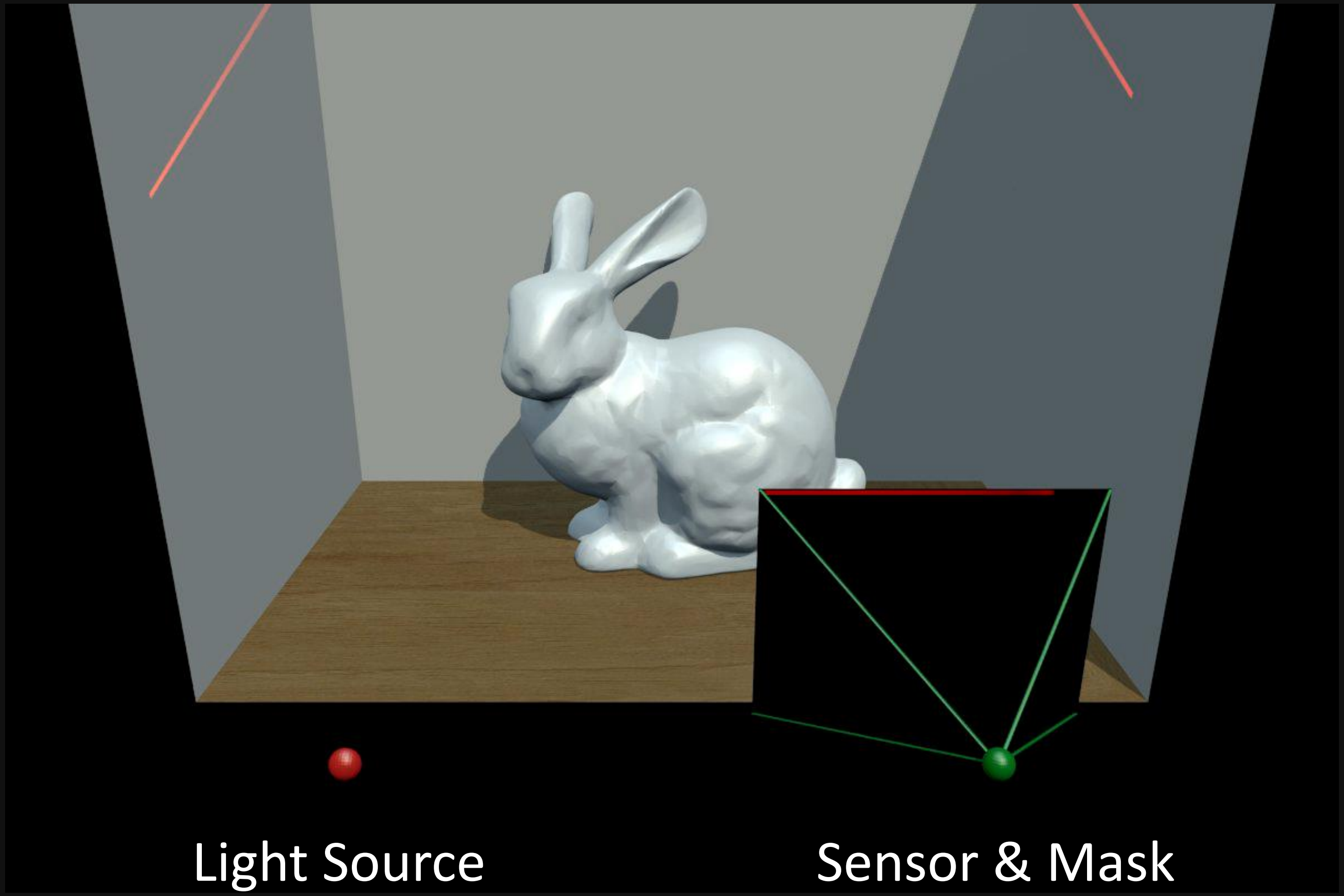
Regular Imaging



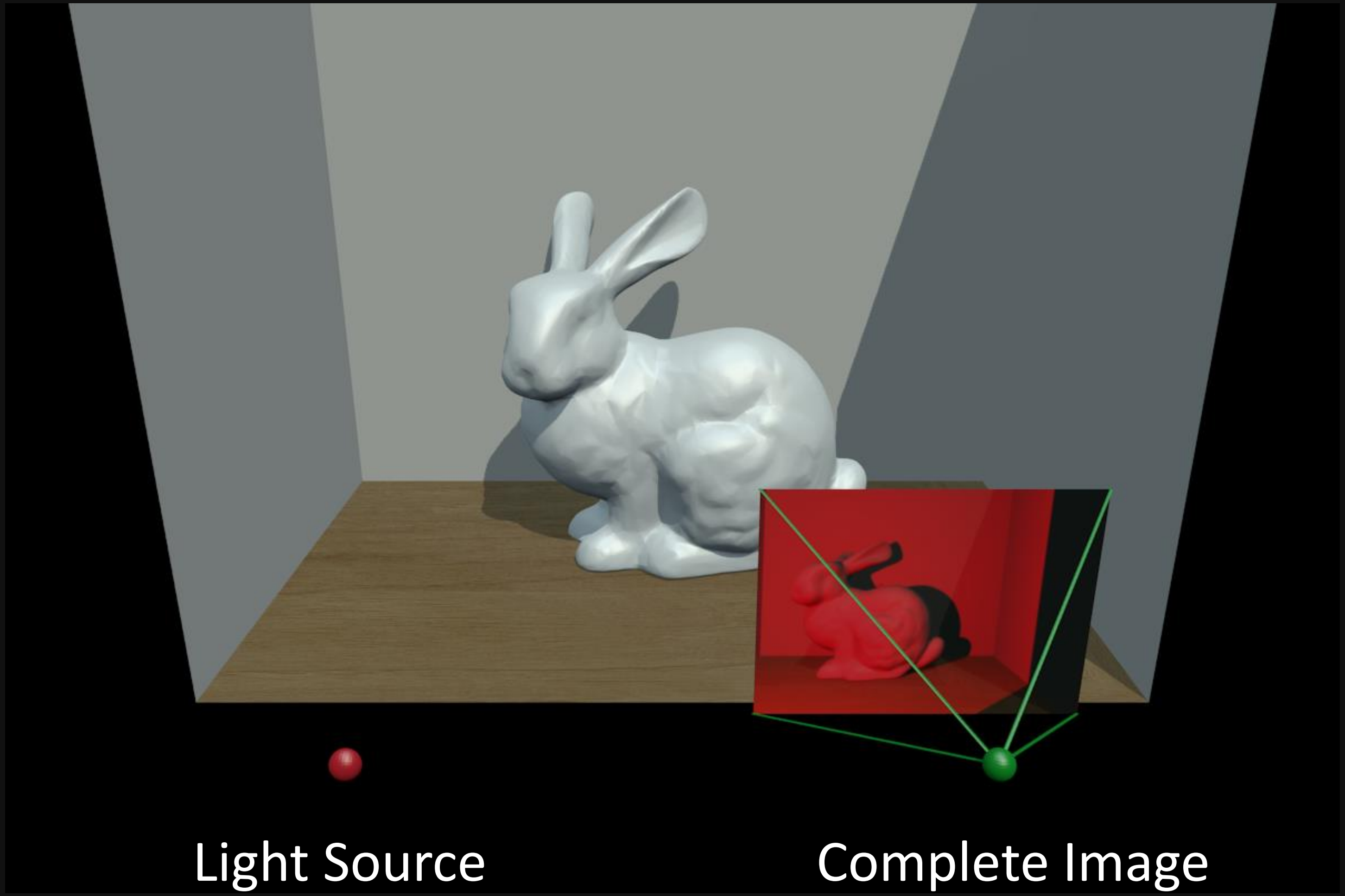
Epipolar Imaging



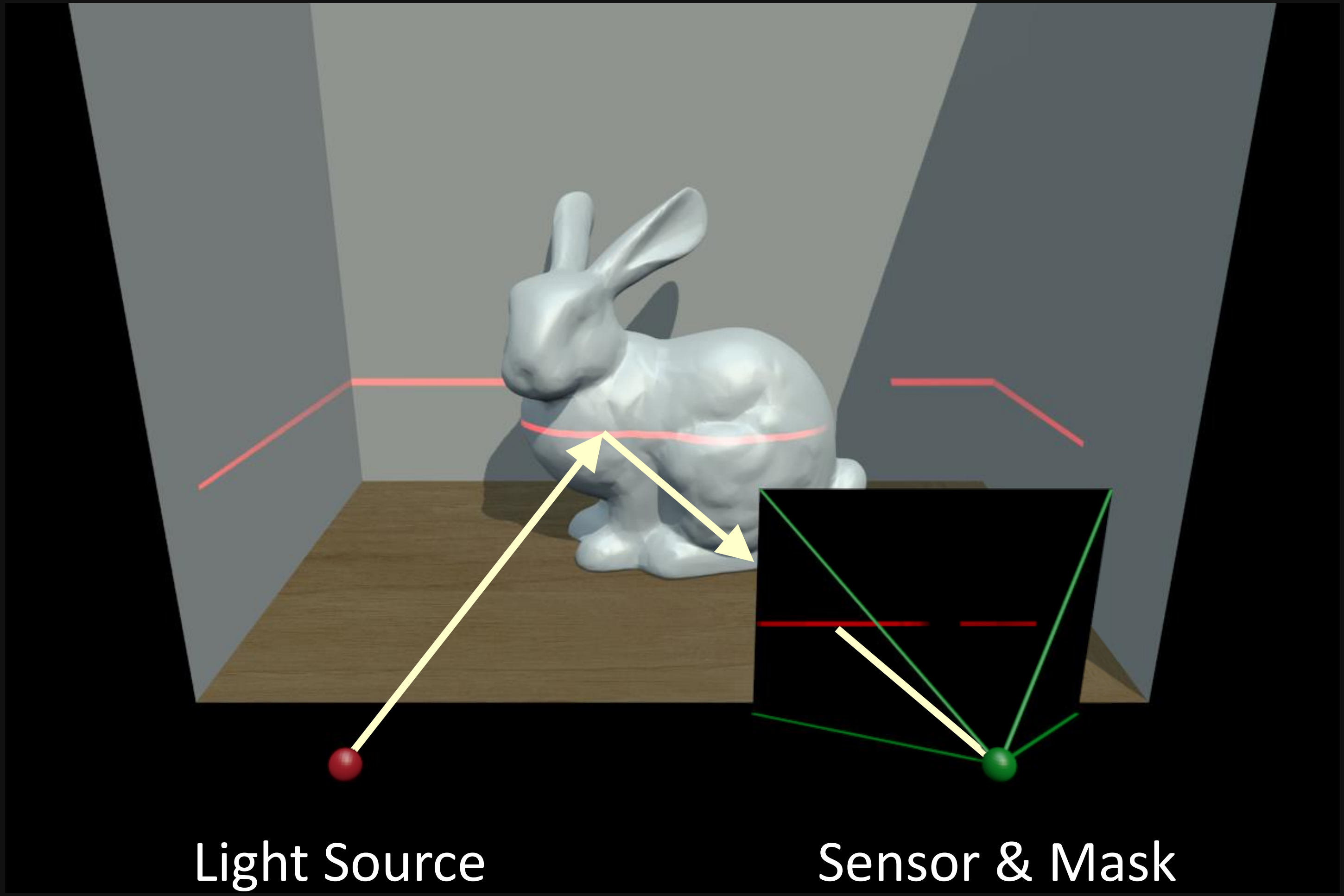
Epipolar Imaging



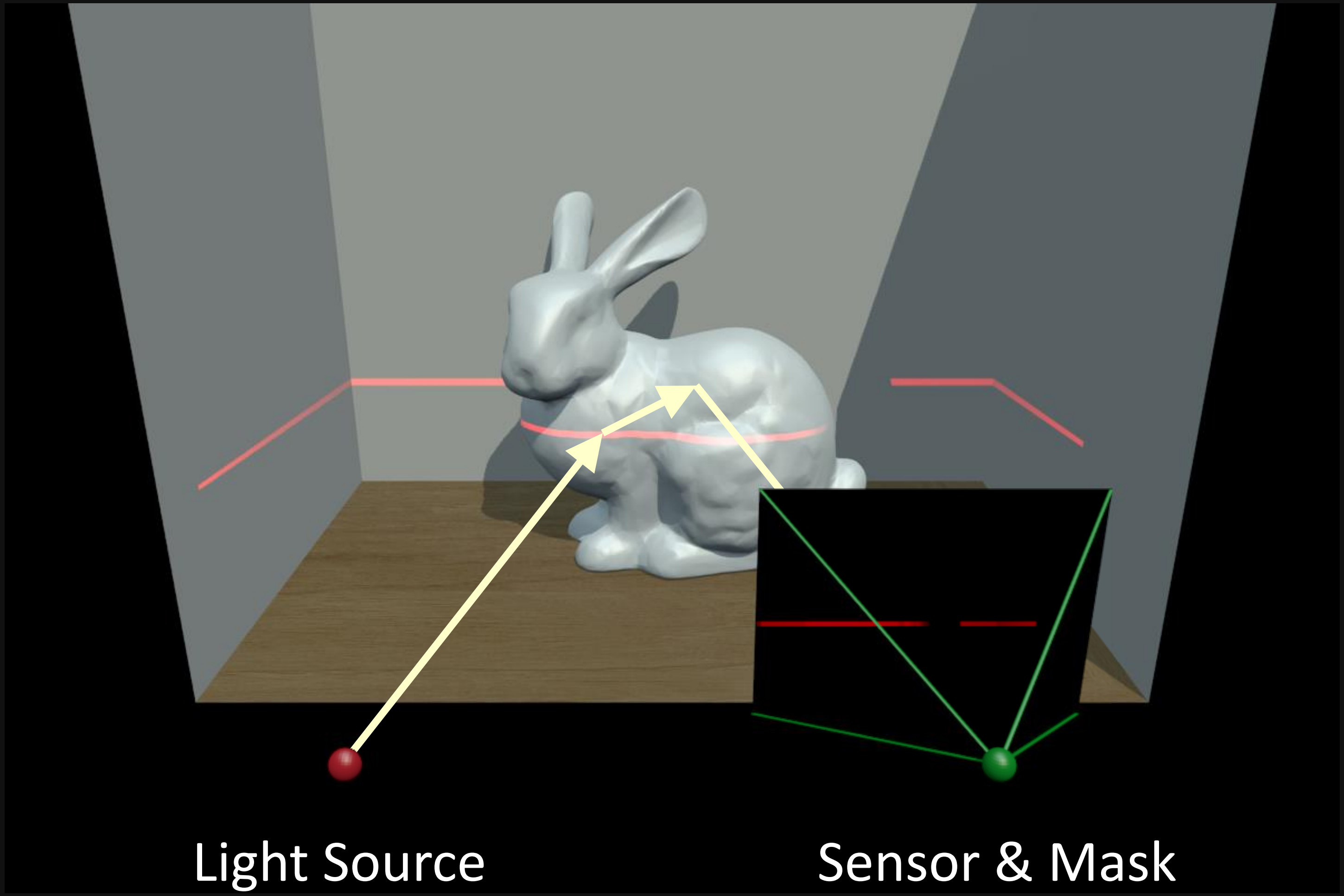
Epipolar Imaging



Epipolar Imaging



Epipolar Imaging



References

Basic reading:

- Szeliski textbook, Sections 7.1, 11.1, 12.1.
- Lanman and Taubin, “Build Your Own 3D Scanner: Optical Triangulation for Beginners,” SIGGRAPH course 2009.

  this very comprehensive course has everything you need to know about 3D scanning using structured light, including details on how to build your own.
- Gupta et al., “A Practical Approach to 3D Scanning in the Presence of Interreflections, Subsurface Scattering and Defocus,” IJCV 2013.

  this paper has a very detailed treatment of problems to structured-light-based 3D scanning caused because of global illumination, and proposes the robust XOR patterns we discussed.

Additional reading:

- Gupta et al., “A Combined Theory of Defocused Illumination and Global Light Transport,” IJCV 2012.

  an earlier paper discussing global illumination and 3D scanning with structured light.
- O’Toole et al., “Homogeneous codes for energy-efficient illumination and imaging,” SIGGRAPH 2015.

  the epipolar imaging paper we covered in a previous class also includes a discussion of how epipolar imaging helps when performing stereo-based 3D scanning in the presence of global illumination.

References

Basic reading:

- Szeliski textbook, Section 8.1 (not 8.1.1-8.1.3), Chapter 11, Section 12.2.
- Hartley and Zisserman, Section 11.12.