3D Scanning

Sai-Kit Yeung

ISTD, SUTD

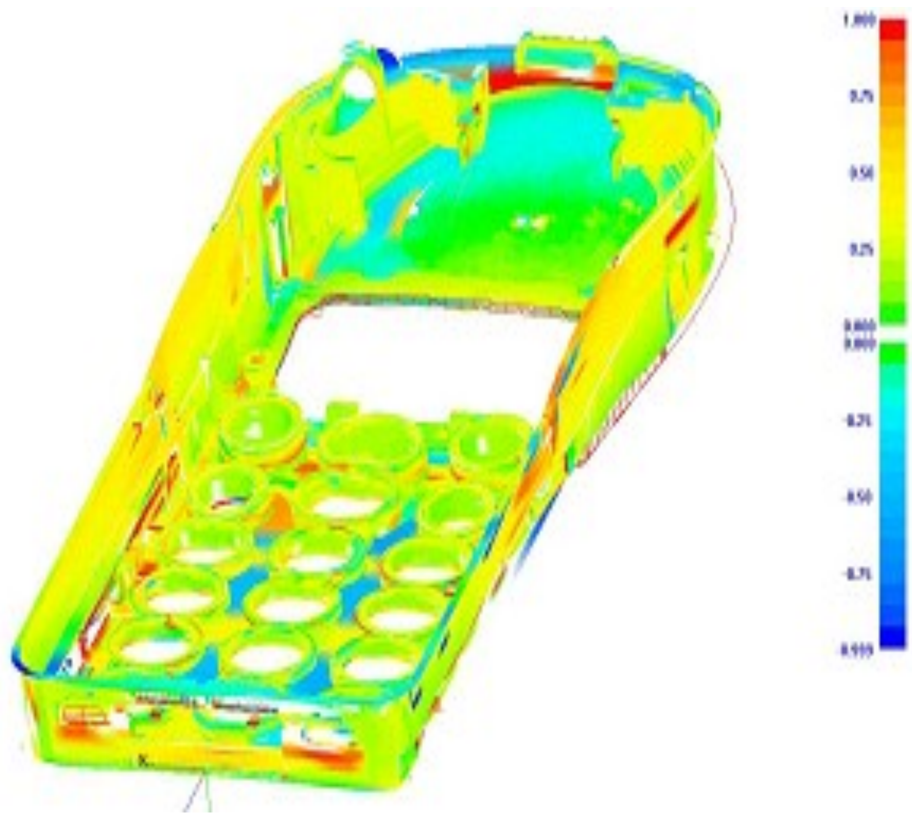
Notes courtesy by Wojciech Matusik, Some slides from Szymon Rusinkiewicz, Brian Curless, Srinivasa Narasimhan

3D Scanning

- Obtaining 3D shape (and appearance) of real-world objects

Applications: Industrial Inspection

- Determine whether manufactured parts are within tolerances

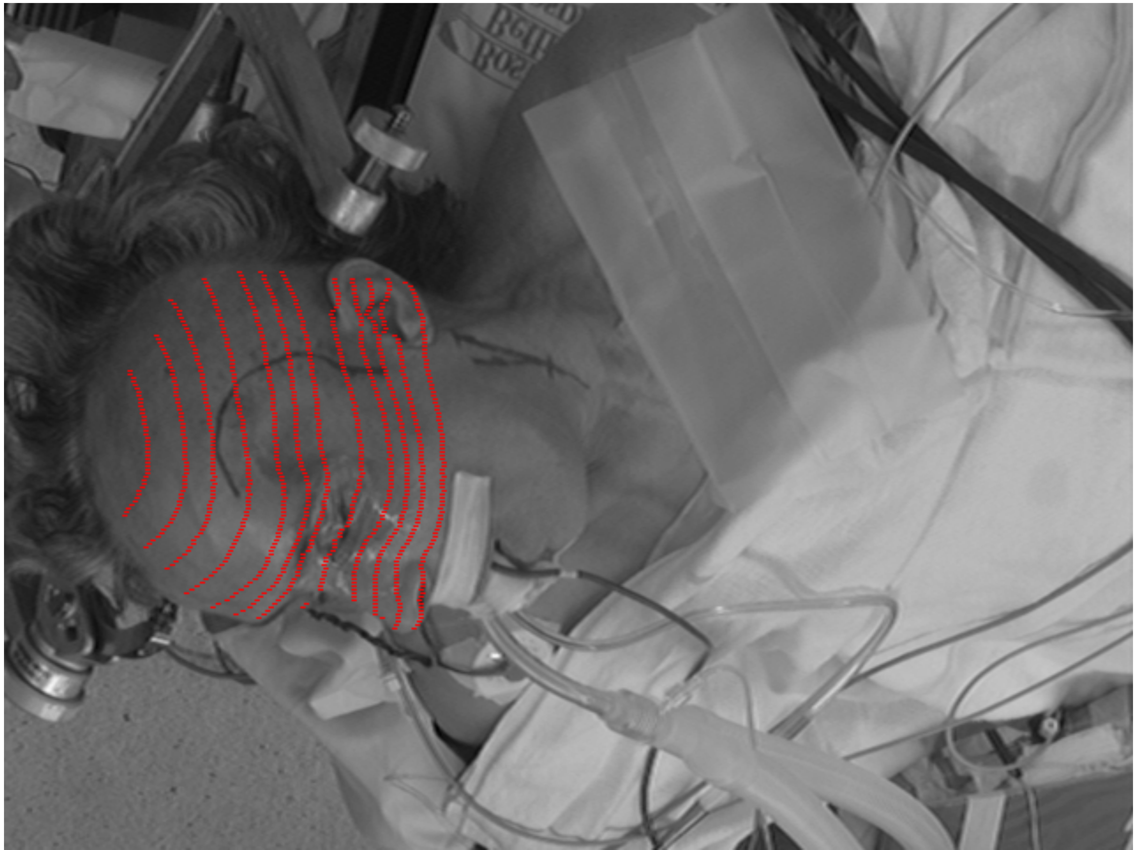


Applications: Medicine

- Plan surgery on computer model, visualize in real time

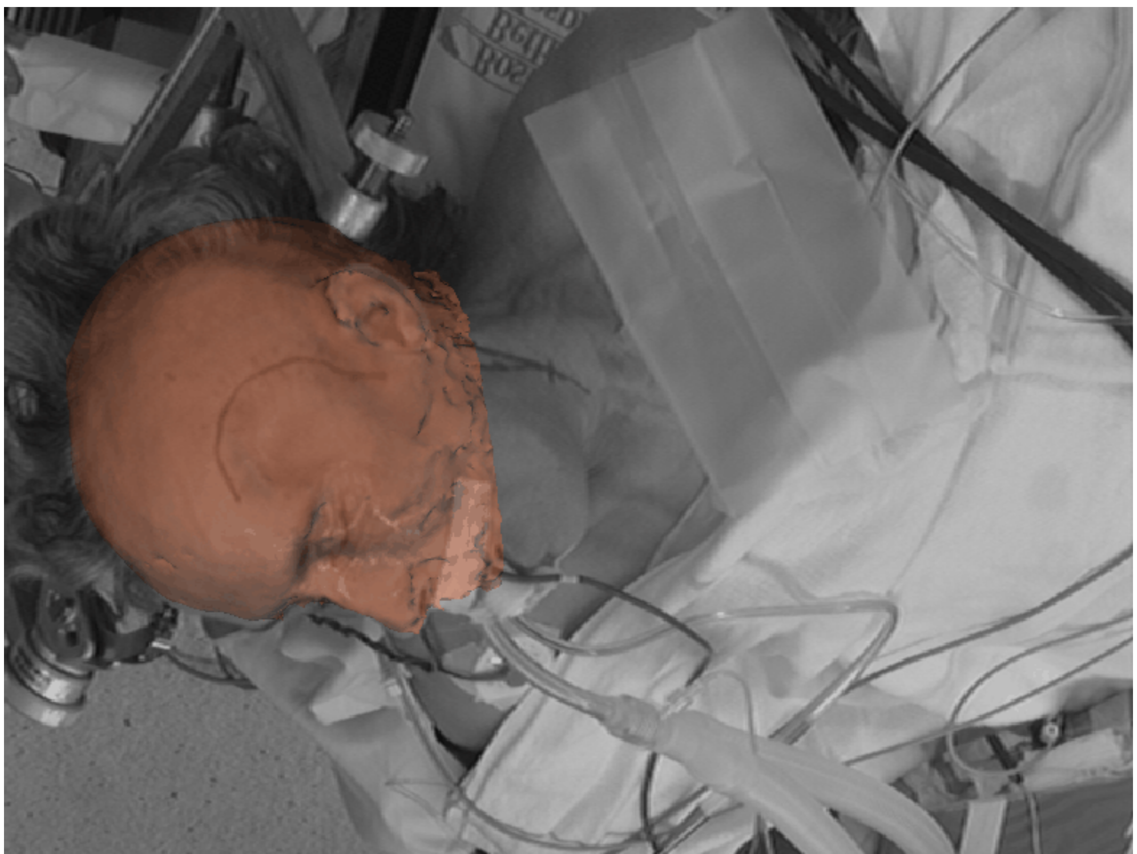
Applications: Medicine

- Plan surgery on computer model, visualize in real time



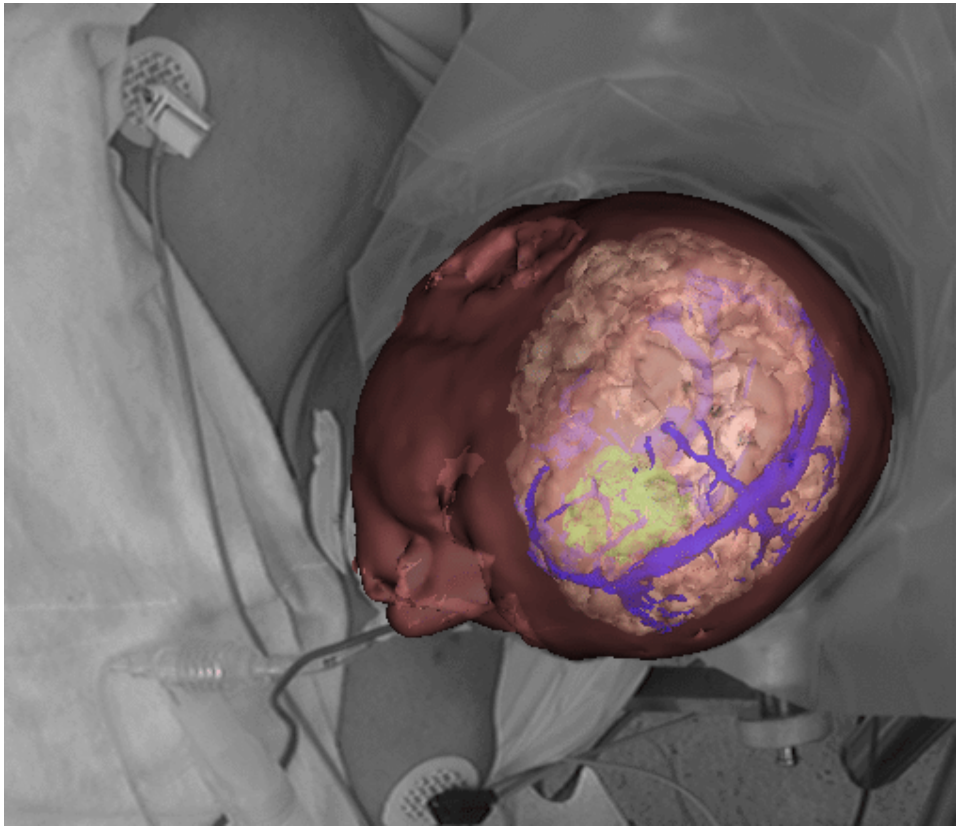
Applications: Medicine

- Plan surgery on computer model, visualize in real time



Applications: Medicine

- Plan surgery on computer model, visualize in real time

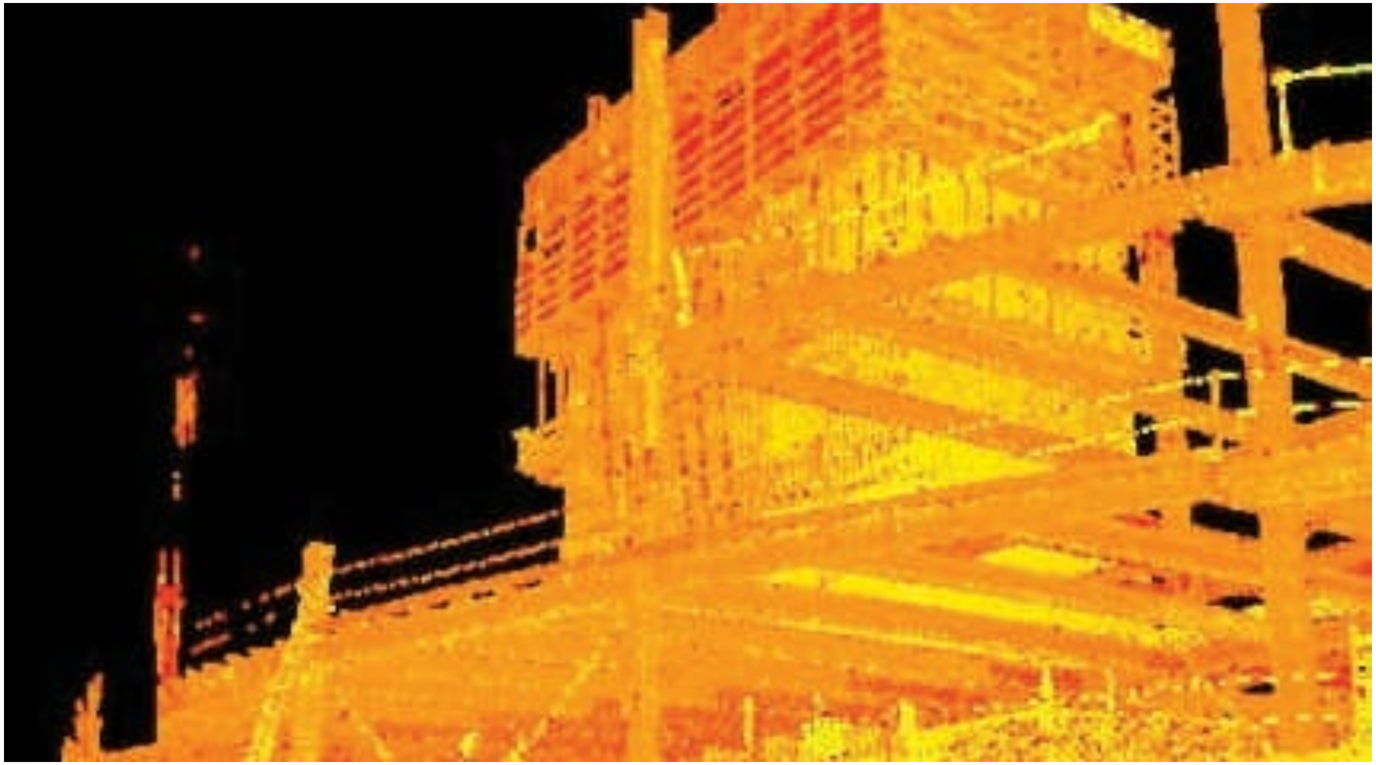


Applications: Scanning Buildings

- Quality control during building
- As-build models

Applications: Scanning Buildings

- Quality control during building
- As-build models



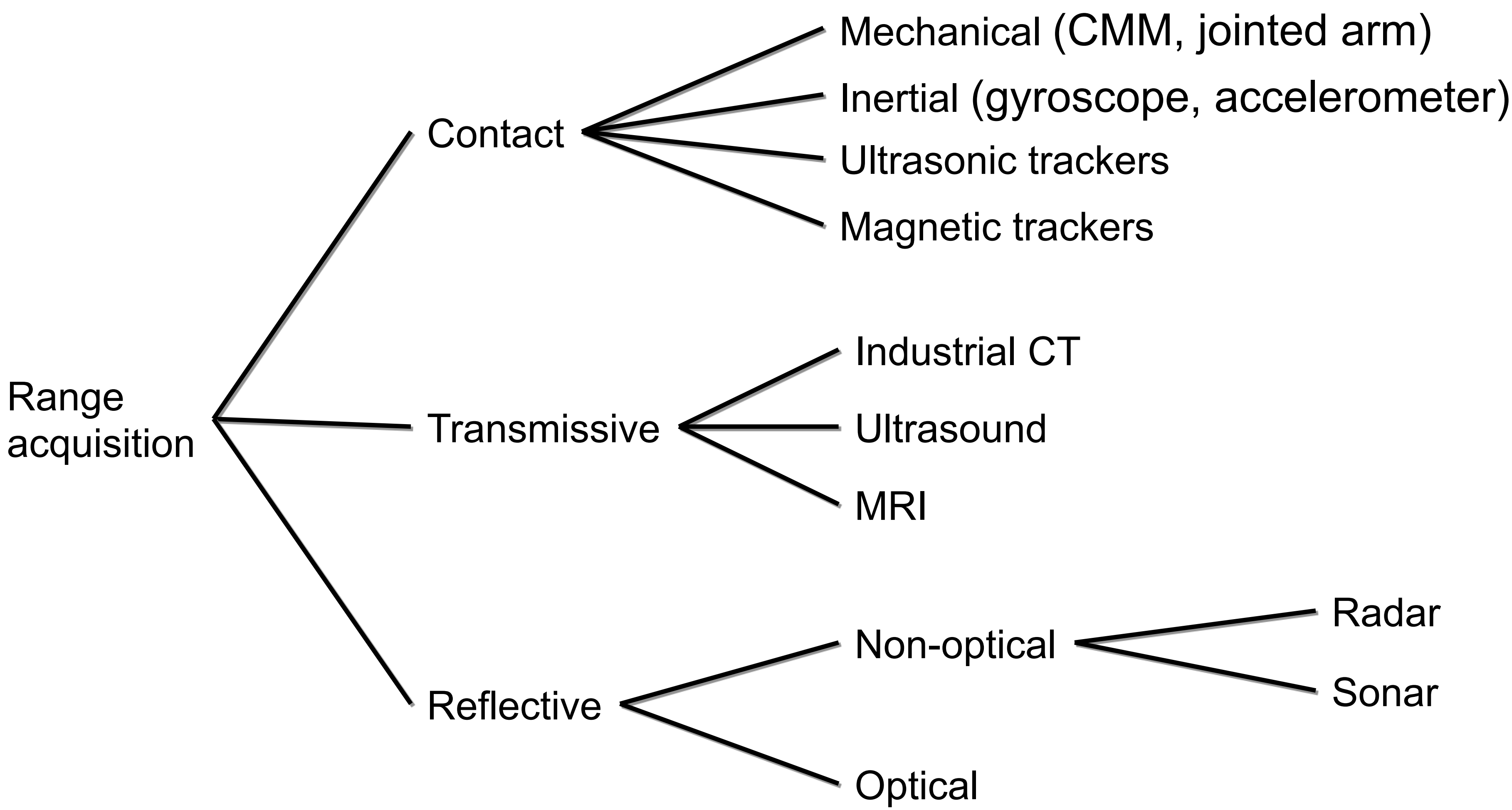
Applications: Virtual Clothing

- Scan a person, custom-fit clothing
  - U.S. Army; booths in malls

Applications: Cultural Heritage

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

Range Acquisition Taxonomy



Touch Probes

- Jointed arms with angular encoders
- Return position, orientation of tip

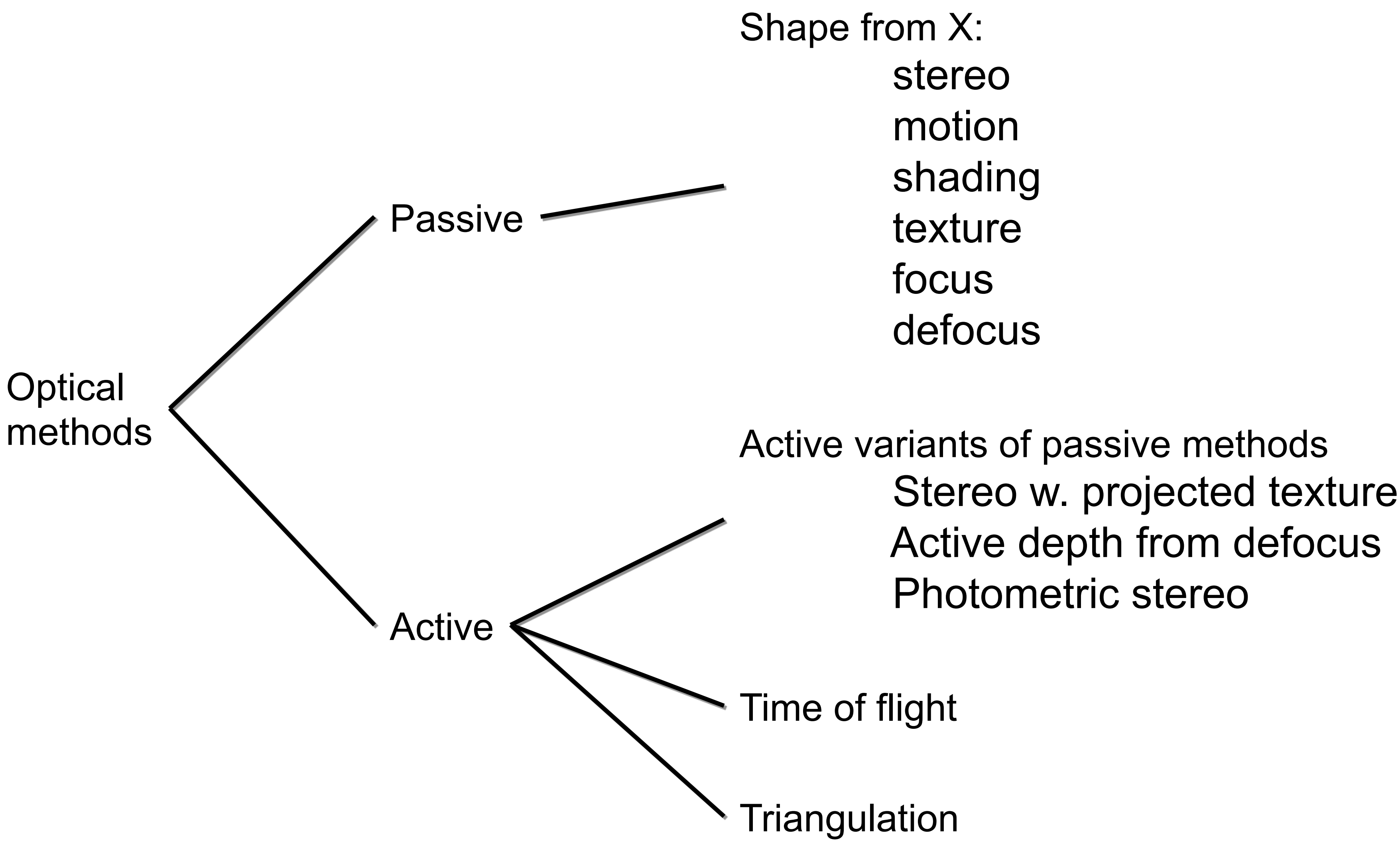
Faro Arm – Faro Technologies, Inc.

Touch Probes

- Jointed arms with angular encoders
- Return position, orientation of tip

Faro Arm – Faro Technologies, Inc.

Range Acquisition Taxonomy



Optical Range Acquisition Methods

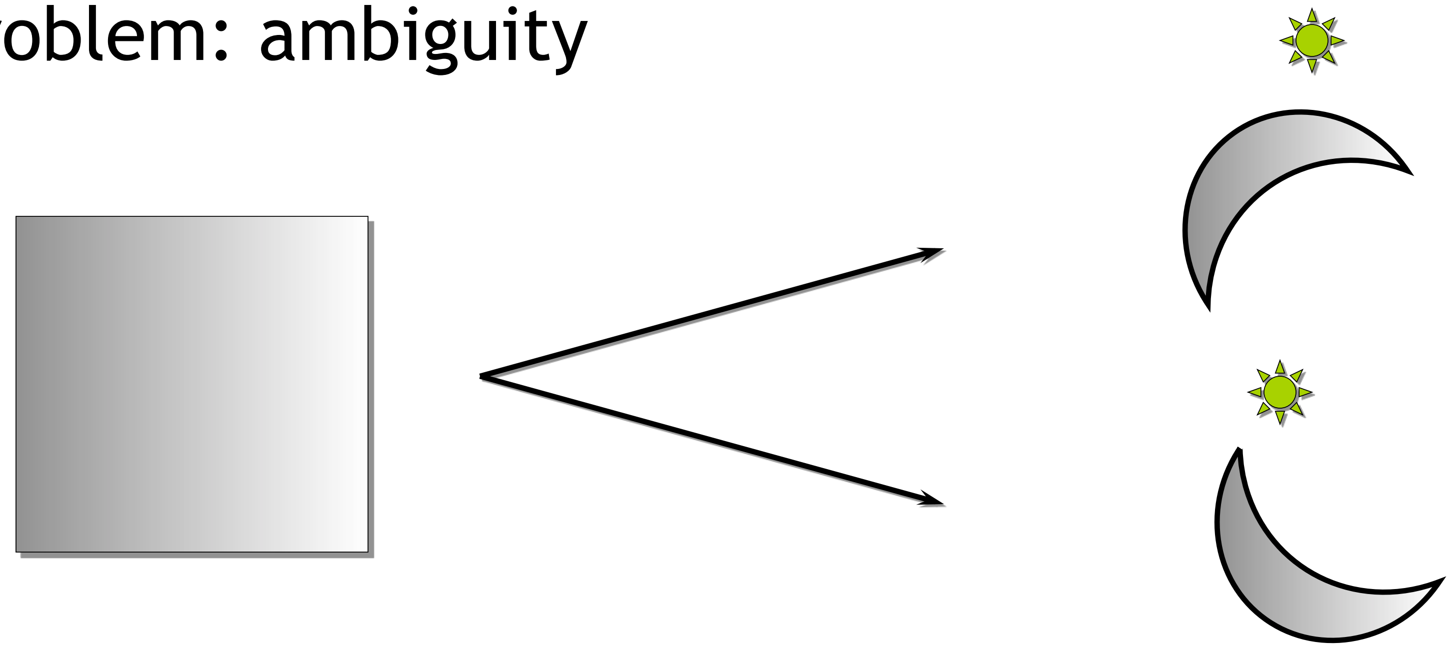
- Advantages:
  - Non-contact
  - Safe
  - Usually inexpensive
  - Usually fast
- Disadvantages:
  - Sensitive to transparency
  - Confused by specularity and interreflection
  - Texture (helps some methods, hurts others)

Range Acquisition Taxonomy

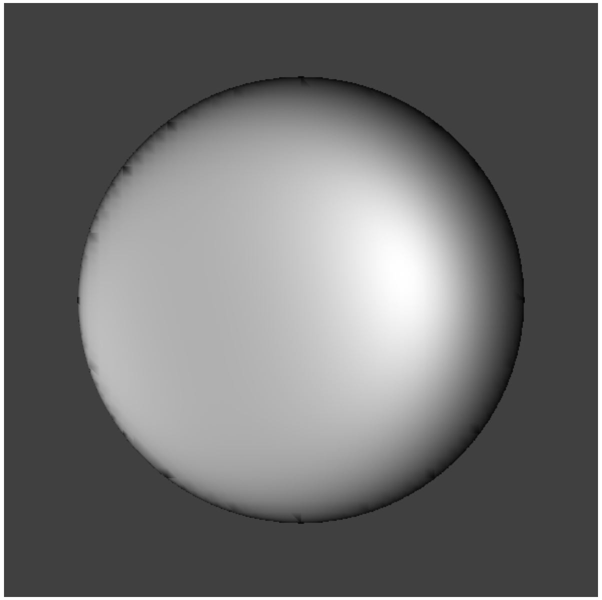
- Optical methods
  - Passive
    - Shape from X:
      - stereo
      - motion
      - shading
      - texture
      - focus
      - defocus
  - Active
    - Active variants of passive methods
      - Stereo w. projected texture
      - Active depth from defocus
      - Photometric stereo
    - Time of flight
    - Triangulation

Shape from Shading

- Single image
- Given: image of surface with known, constant reflectance under known point light
- Estimate normals, integrate to find surface
- Problem: ambiguity



Shape from Shading



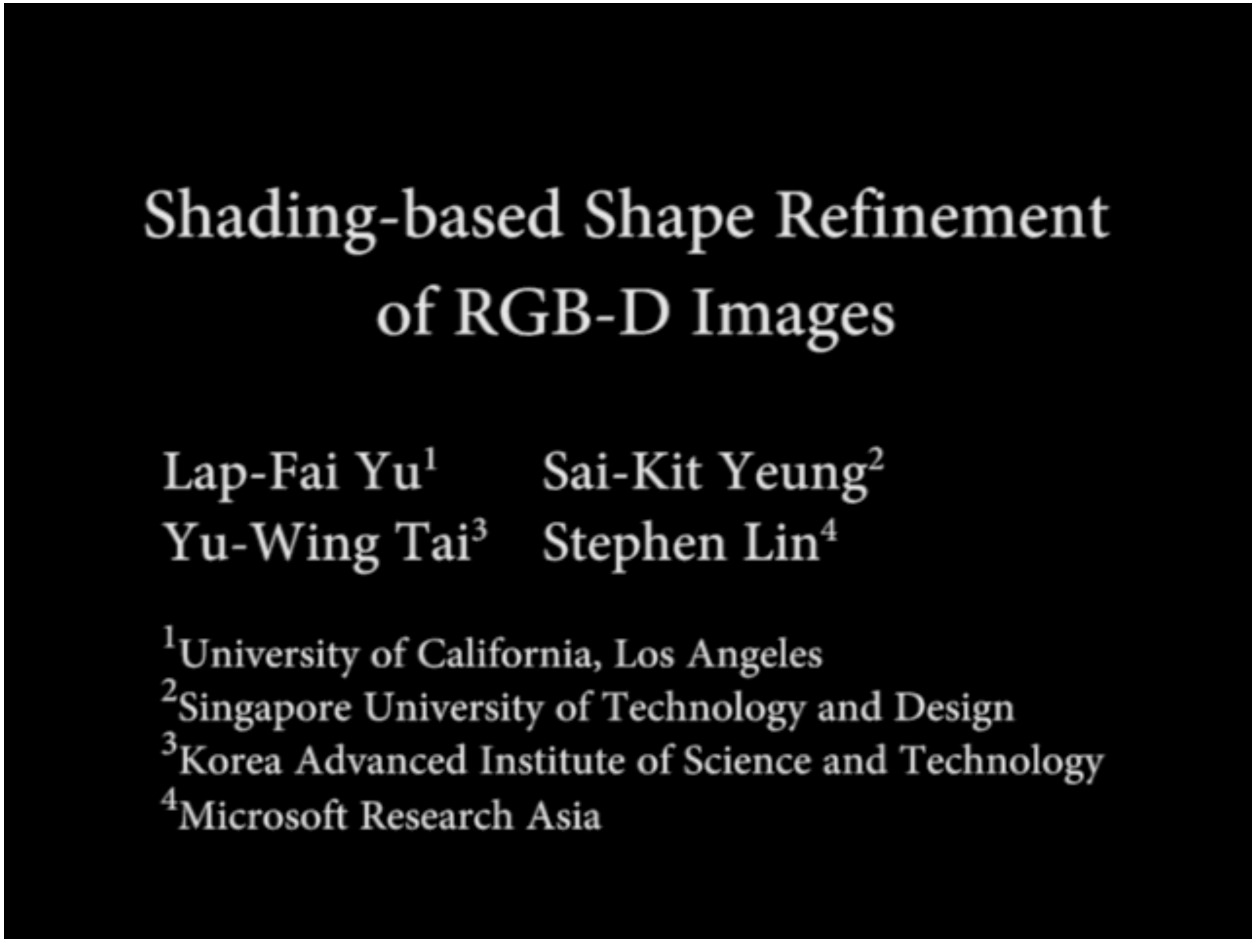
- Advantages:
  - Single image
  - No correspondences
  - Analogue in human vision
- Disadvantages:
  - Mathematically unstable
  - Can't have texture
- "Photometric stereo" (active method) more practical than passive version
- Or with Shape prior (not strictly passive anymore!)

Shape from Shading with Shape Prior

- Shape prior from Kinect:
  - *Lap-Fai Yu*, **Sai-Kit Yeung**, Yu-Wing Tai, Stephen Lin, Shading-based Shape Refinement of RGB-D Images. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013.
  - Jonathan T. Barron, Jitendra Malik, Intrinsic Scene Properties from a Single RGB-D Image. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013
  - Gyeongmin Choe, Jaesik Park, Yu-Wing Tai, In So Kweon, Exploiting Shading Cues in Kinect IR Images for Geometry Refinement. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014
  - Chenglei Wu, Michael Zollhöfer, Matthias Nießner, Marc Stamminger, Shahram Izadi, Christian Theobalt, Real-time Shading-based Refinement for Consumer Depth Cameras, ACM SIGGRAPHAsia 2014
  - Michael Zollhöfer, Angela Dai, Matthias Innmann, Chenglei Wu, Marc Stamminger, Christian Theobalt, Matthias Nießner, Shading-based Refinement on Volumetric Signed Distance Function, ACM Transactions on Graphics (TOG) 2015

Shape from Shading with Shape Prior

- *Lap-Fai Yu*, **Sai-Kit Yeung**, Yu-Wing Tai, Stephen Lin, Shading-based Shape Refinement of RGB-D Images. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2013.

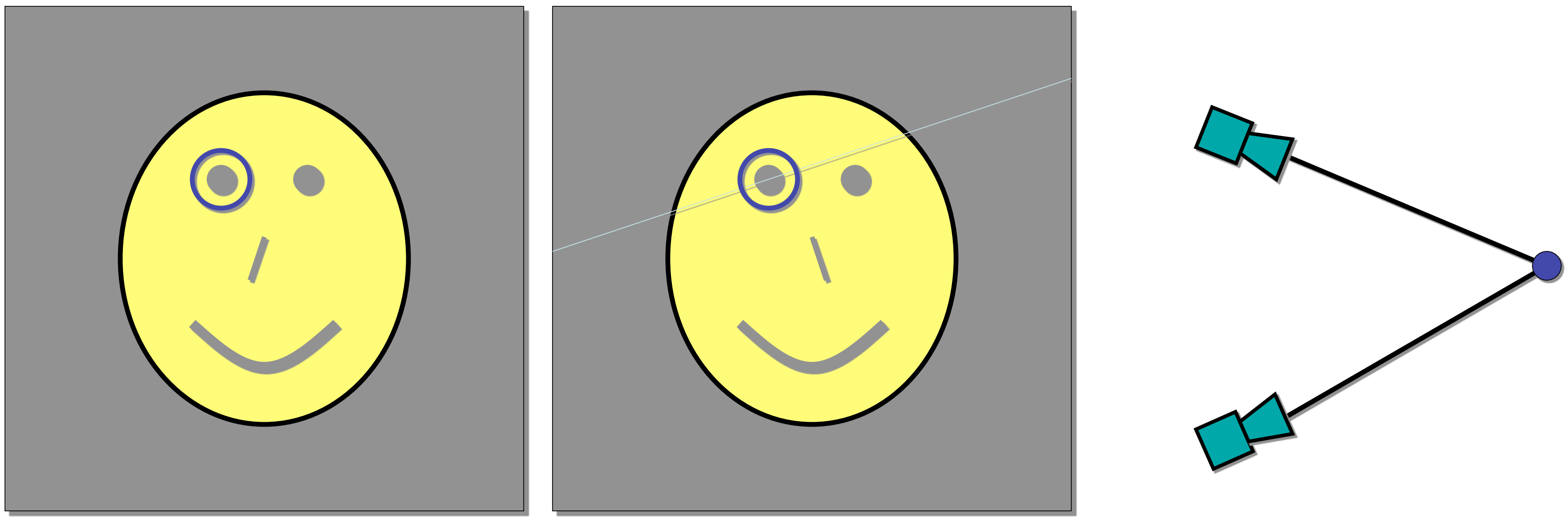


Shape from Shading with Shape Prior

- Chenglei Wu, Michael Zollhöfer, Matthias Nießner, Marc Stamminger, Shahram Izadi, Christian Theobalt,
  Real-time Shading-based Refinement for Consumer Depth Cameras, ACM SIGGRAPHAsia 2014

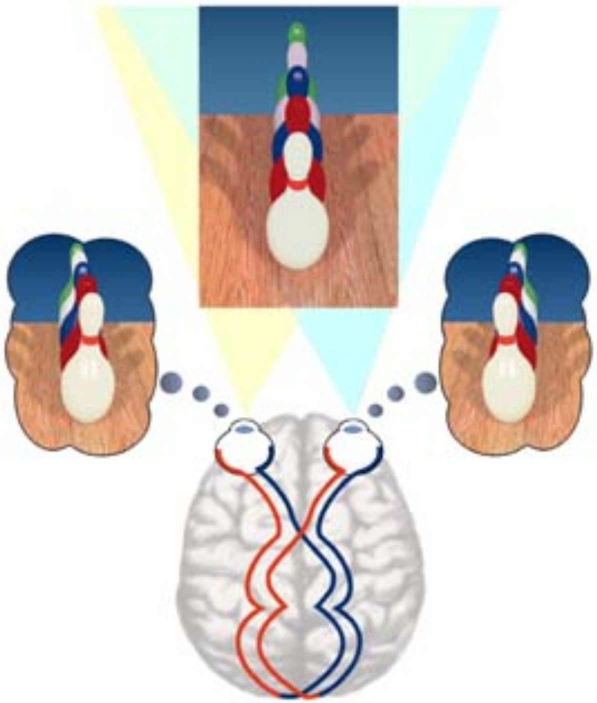
Stereo

- Two or multiple images
- Find feature in one image, search along epipolar line in other image for correspondence



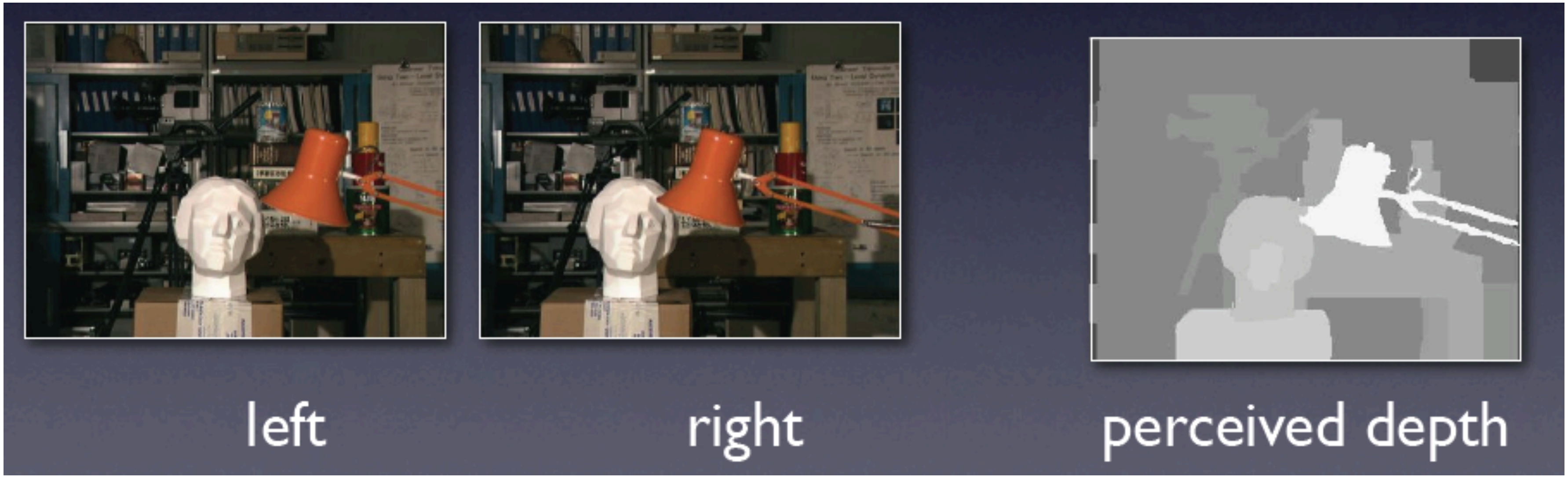
Stereo

- Disparity:
  - Informally: difference between 2 images
  - Allows us to gain a sense of depth
- Stereopsis:
  - Ability to perceive depth from disparity
- Goal:
  - Design algorithms that mimic stereopsis



Stereo

- Binocular fusion of features observed from the eyes
- Reconstruction of their 3D preimage

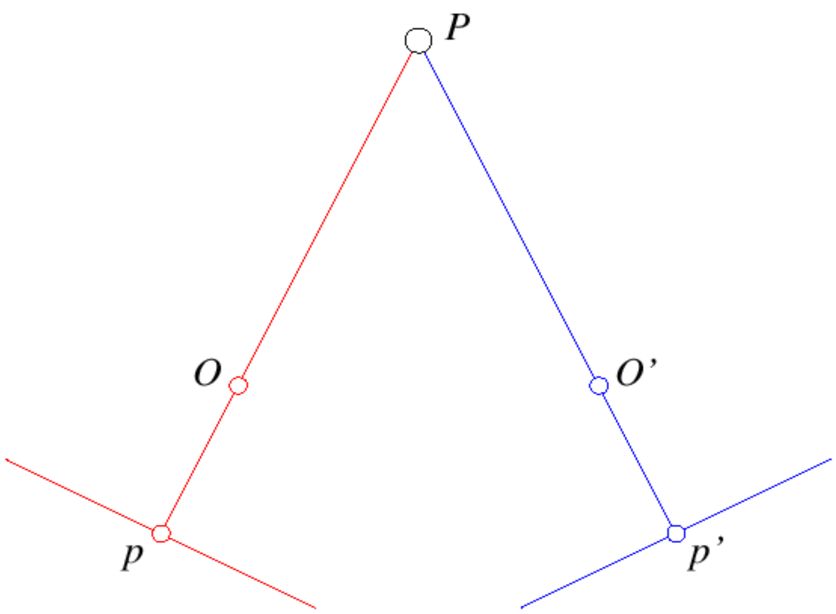


Stereo

- Advantages:
  - Passive
  - Cheap hardware (2 cameras)
  - Easy to accommodate motion
  - Intuitive analogue to human vision

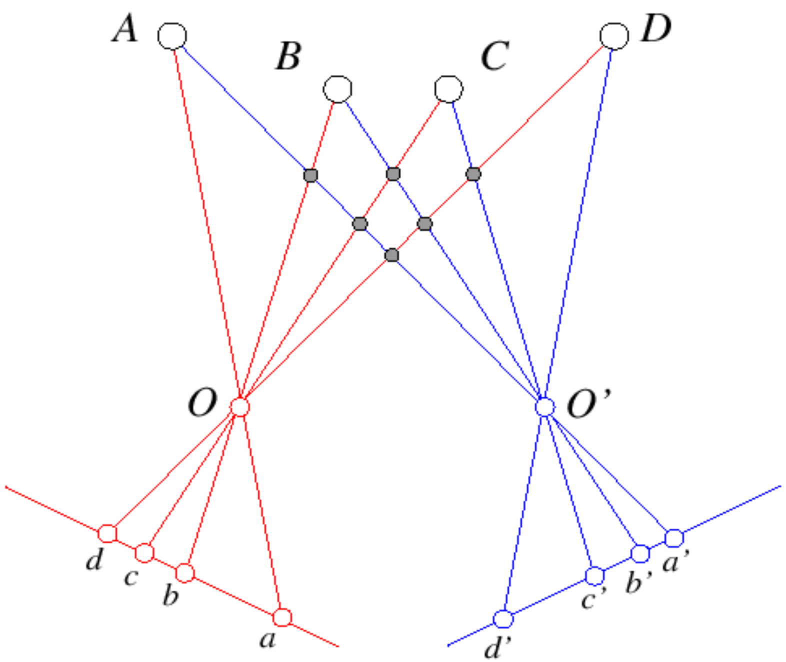
Stereo Vision - Easy Case

- A single point being observed
  - The preimage can be found at the intersection of the rays from the focal points to the image points



Stereo Vision - Hard Case

- Many points being observed
  - Need some method to establish correspondences

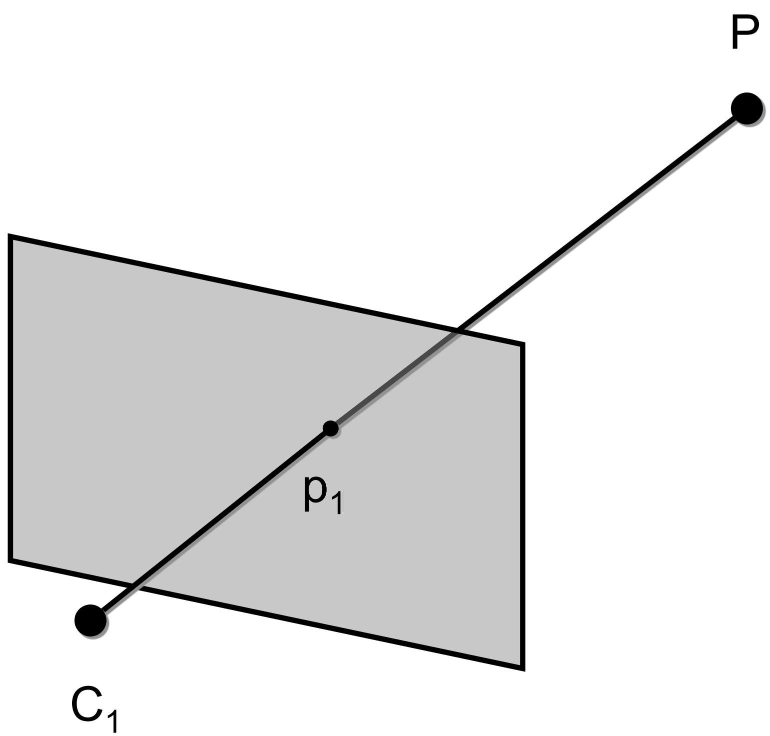


Components of Stereo Vision Systems

- Camera calibration: geometric properties of the cameras
- Image rectification: simplifies the search for correspondences
- Correspondence: which item in the left image corresponds to which item in the right image
- Reconstruction: recovers 3-D information from the 2-D correspondences

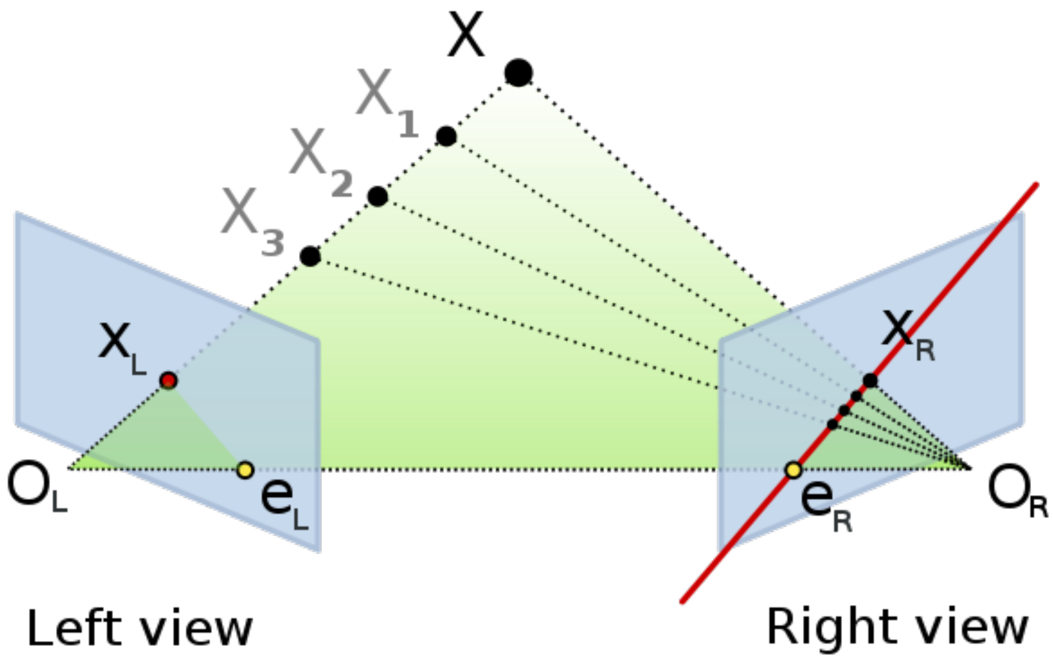
Camera Calibration

- From pixels to rays in 3D
  - given 2D pixel coordinates, what is the corresponding ray in 3D (a point and a vector) ?
- From 3D points to 2D pixels
  - given a 3D point, what are the 2D coordinates of its projection on the image plane?



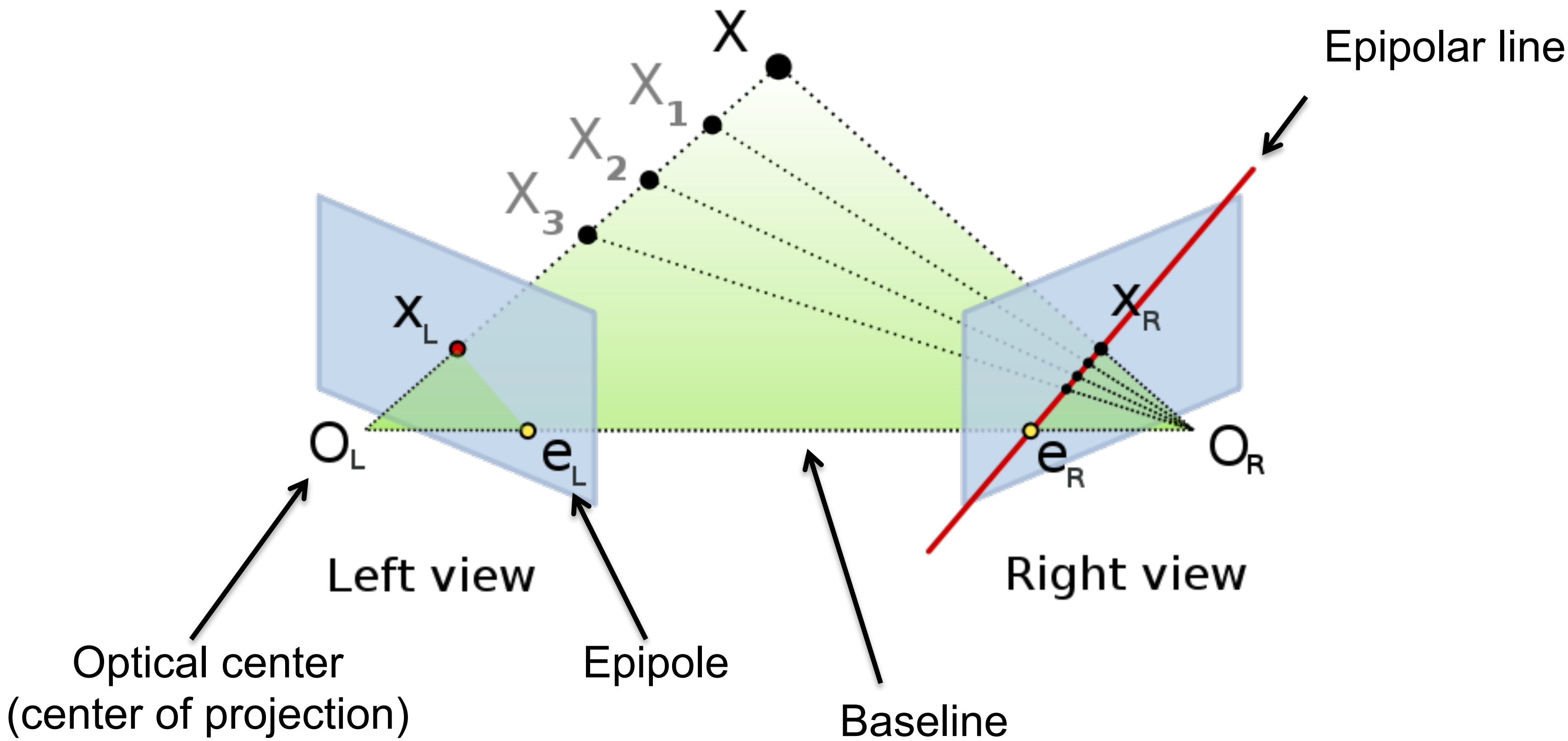
Epipolar Geometry

- Our goal: estimate the unknown 3D point X, from 2 images



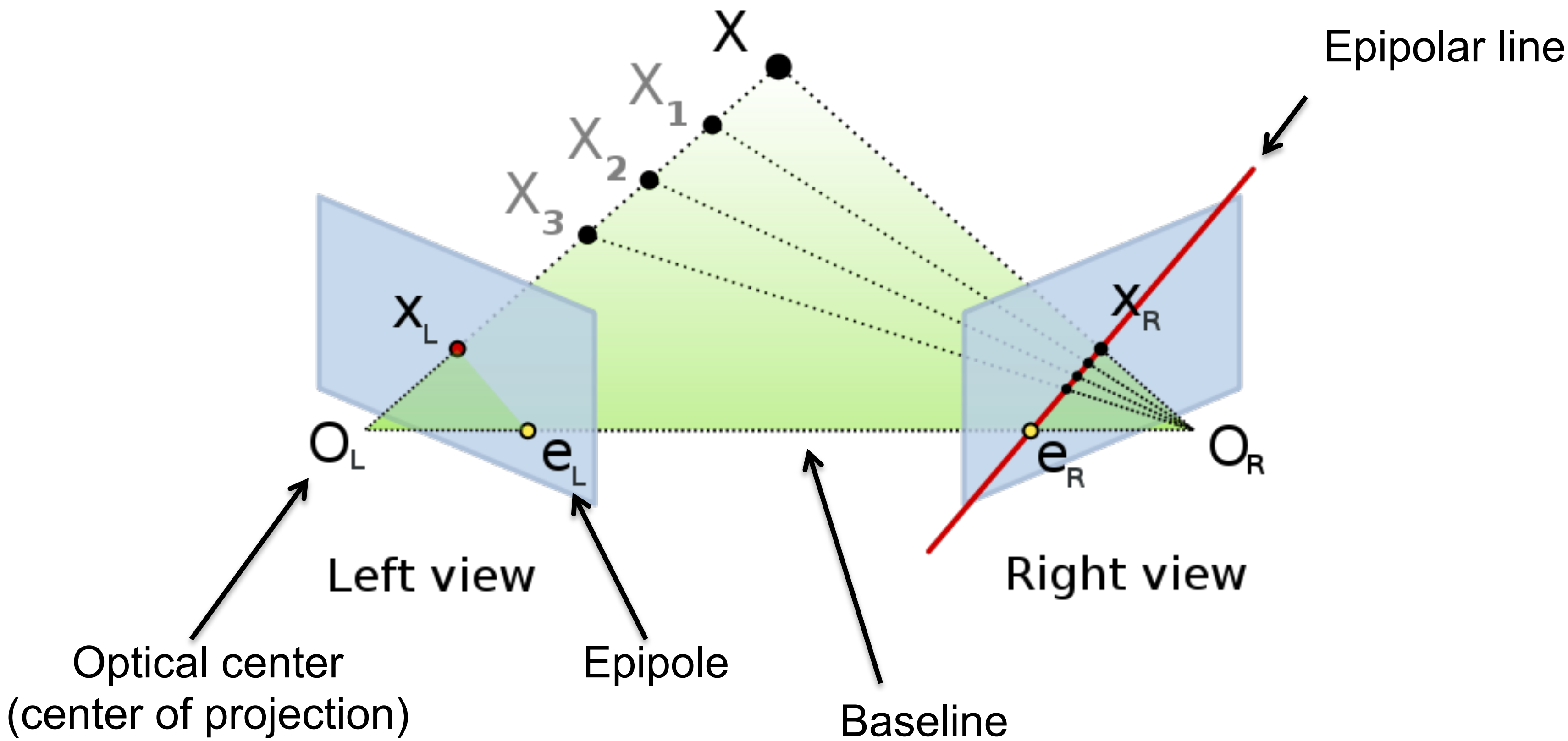
Epipolar Geometry

- Epipole: point of intersection of the line joining the camera centers (the baseline) with the image plane



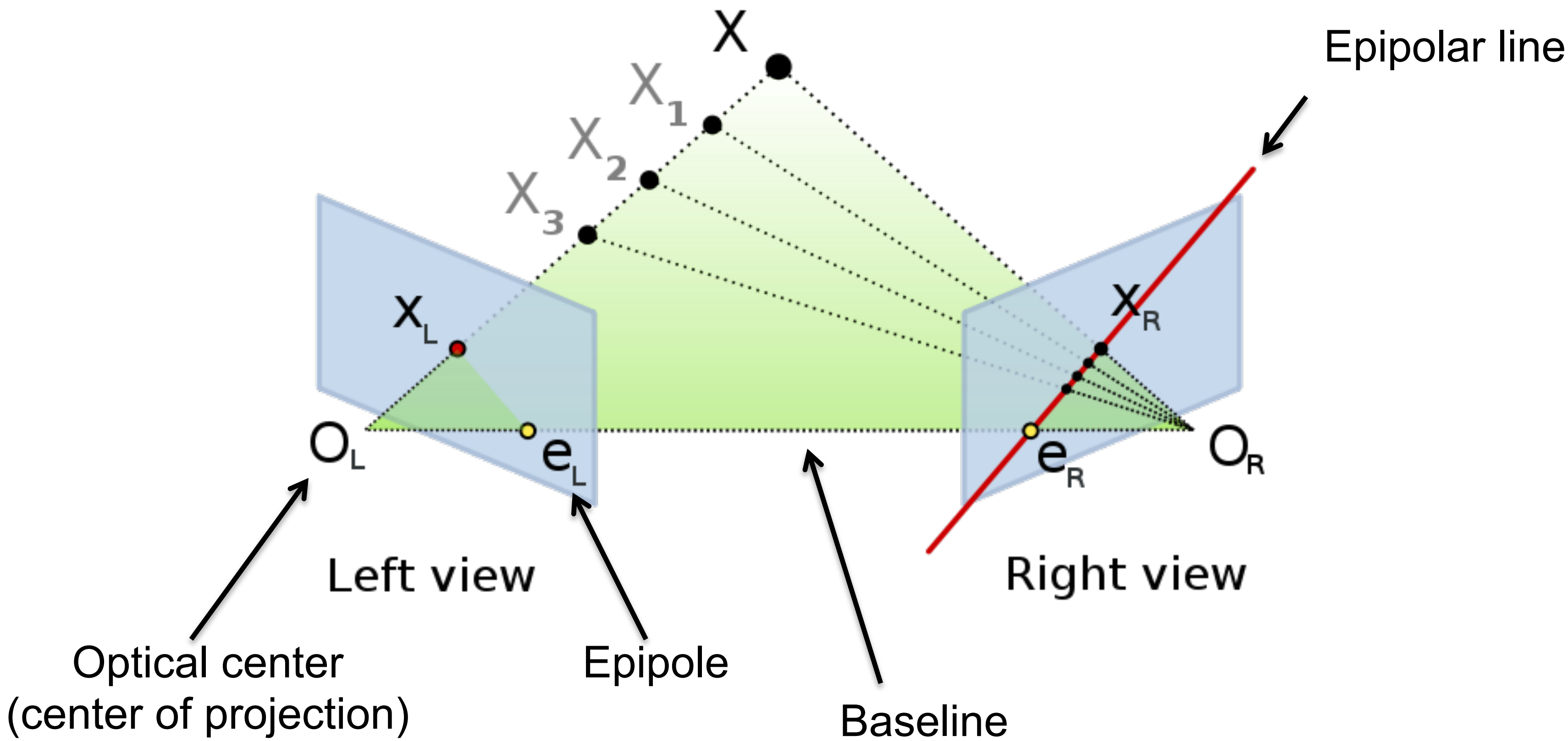
Epipolar Geometry

- O_L-X: a point to left view, but a line to right view
  → Epipolar line (function of X)



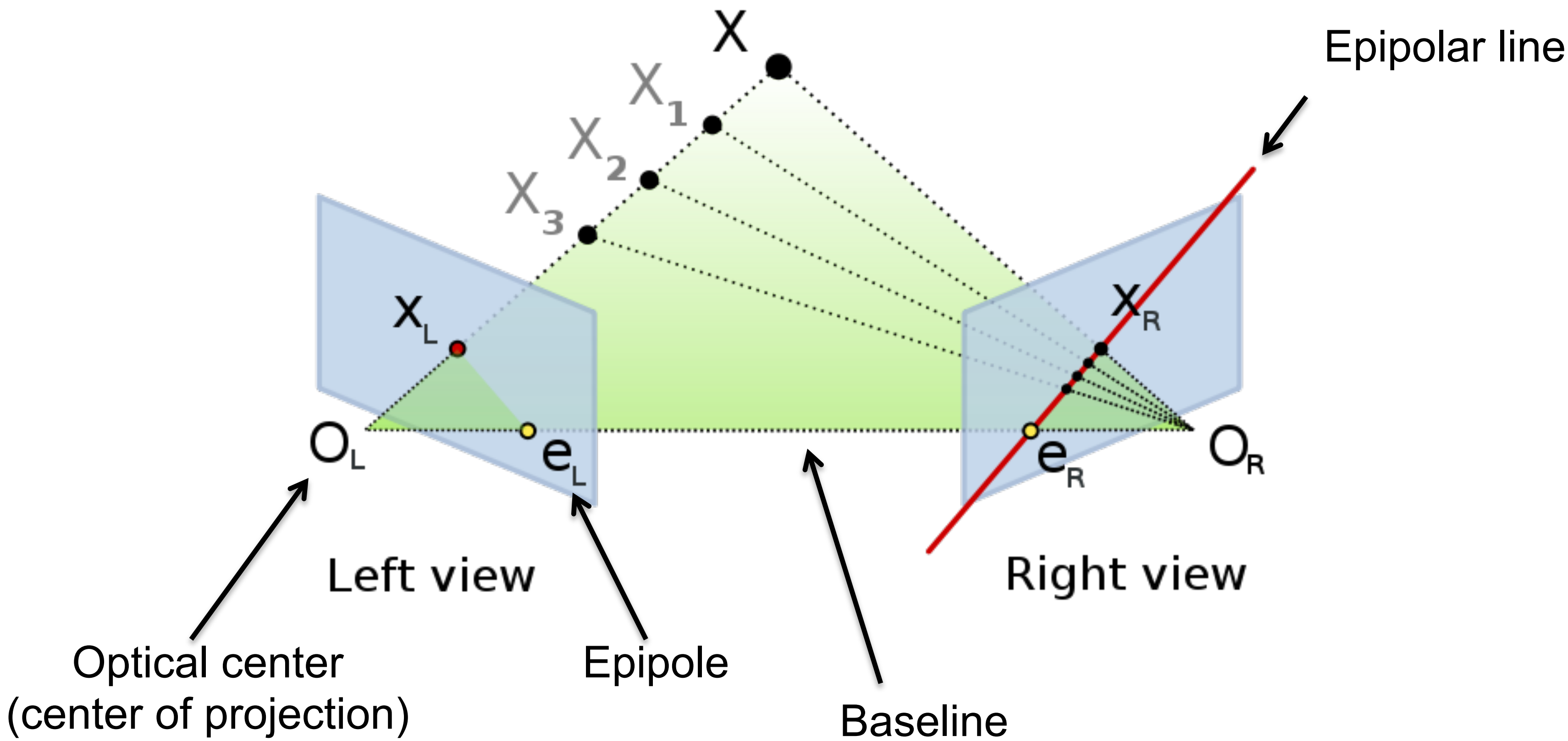
Epipolar Geometry

- O_L-X pass through O_L → Epipolar line pass through epipole at right side



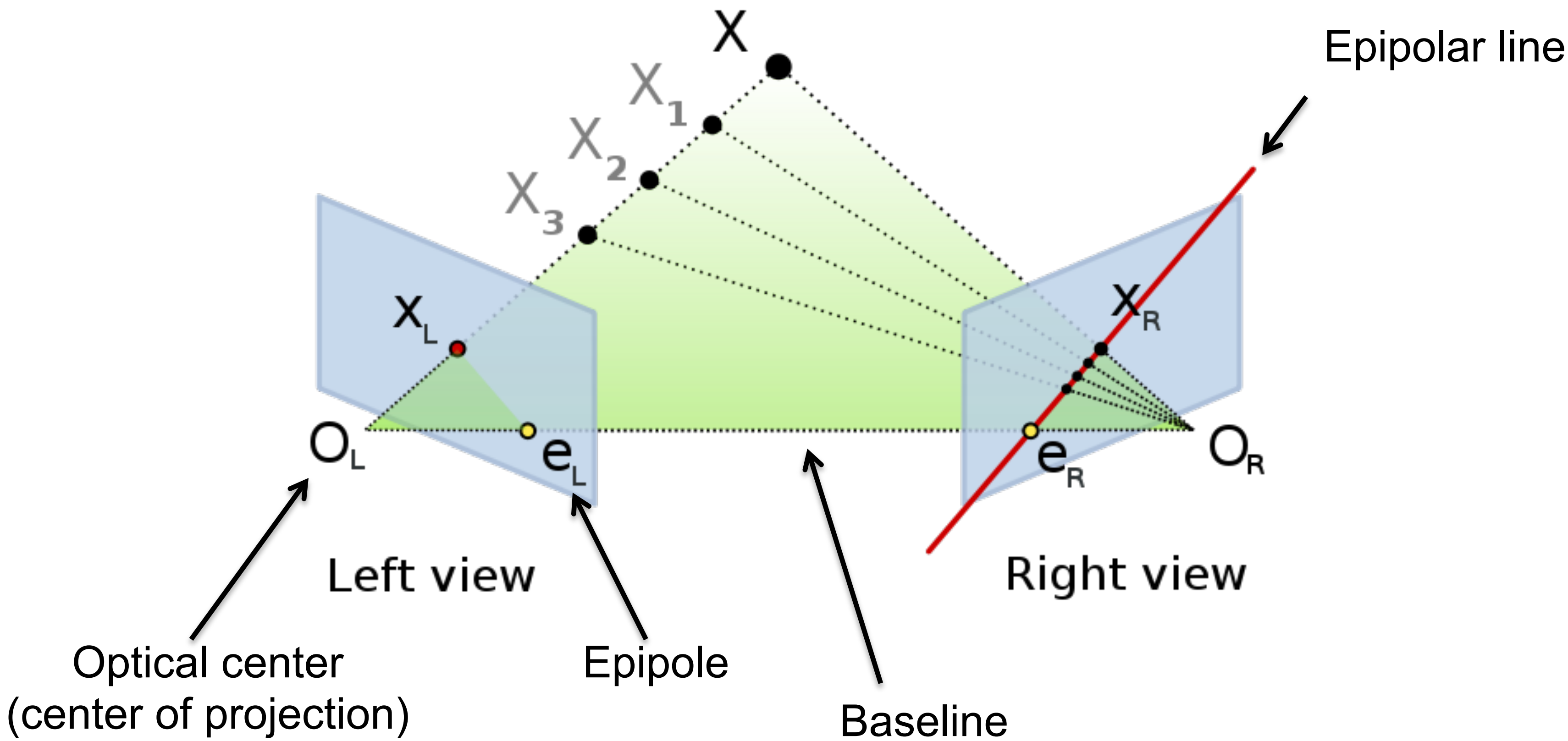
Epipolar Geometry

- Any line intersect epipole is an epipolar line and can be derived from some 3D points X



Epipolar Geometry

- Epipolar constraint: corresponding points must lie on conjugate epipolar lines
  - Search for correspondences becomes a 1-D problem



Epipolar Geometry

- So given X_L → find corresponding X_R → X
- Triangulation

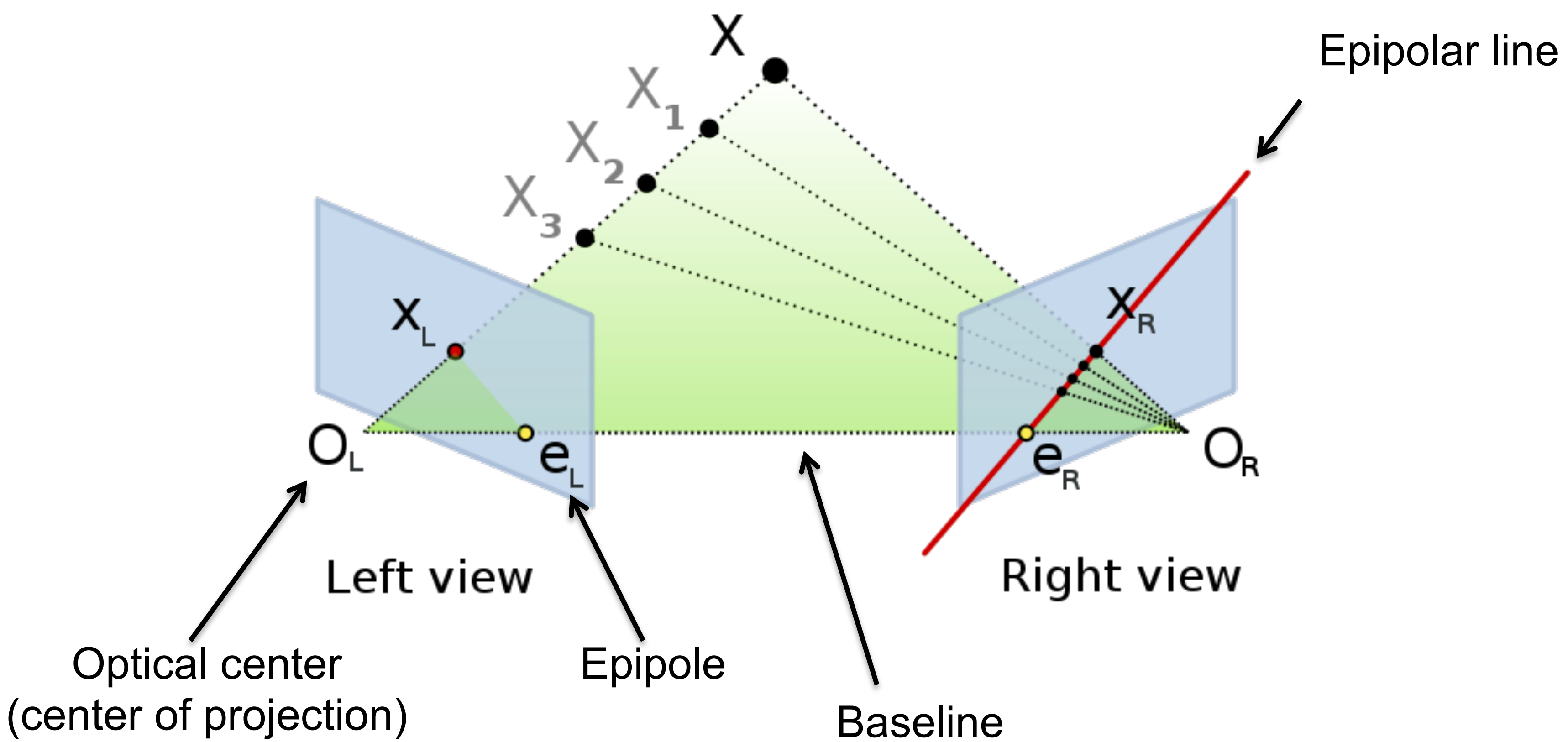
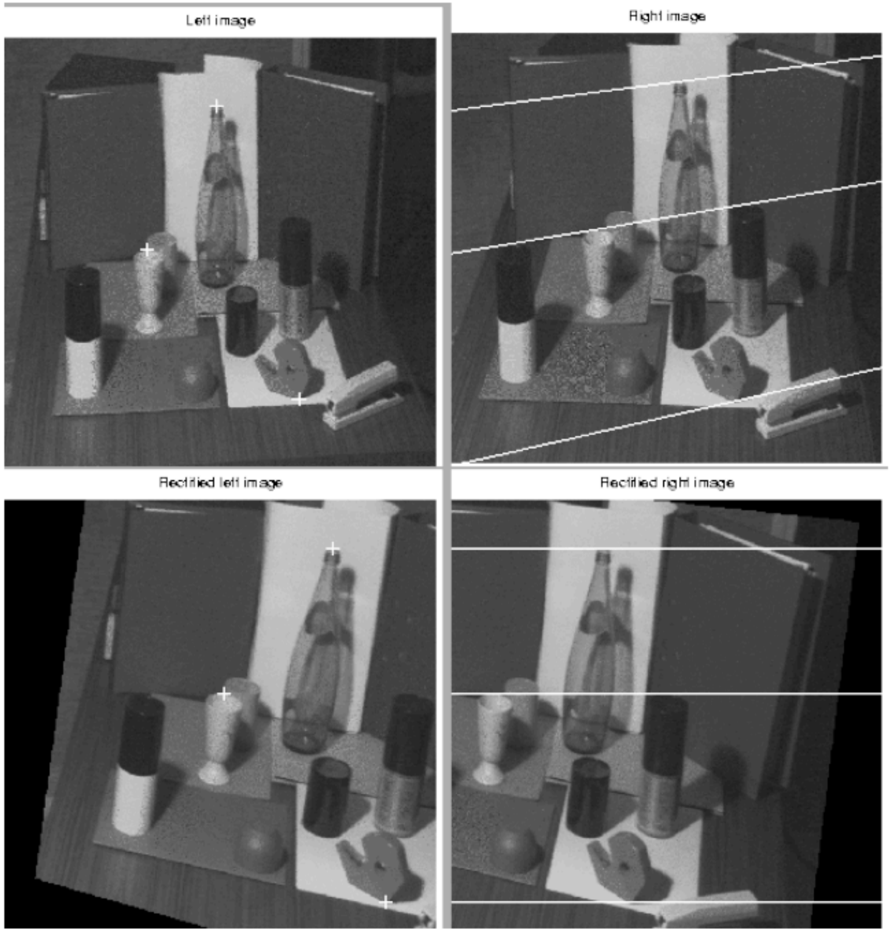


Image Rectification

- Warp images such that conjugate epipolar lines become collinear and parallel to *x* axis

Left image

Right image



Rectified
Left image

Rectified
Right image

Disparity

- With rectified images, disparity is just (horizontal) displacement of corresponding features in the two images
  - Disparity = 0 for distant points
  - Larger disparity for closer points
  - Depth of point proportional to 1/disparity

Correspondence

- Given an element in the left image, find the corresponding element in the right image

- Classes of methods
  - Correlation-based
  - Feature-based

Correlation-Based Correspondence

- Input: rectified stereo pair and a point (u,v) in the first image
- Method:
  - Consider window centered at (u,v)
  - For each potential matching window centered at (u+d,v) in the second image, compute matching score of correspondence
  - Set disparity to value of d giving highest score

Matching score

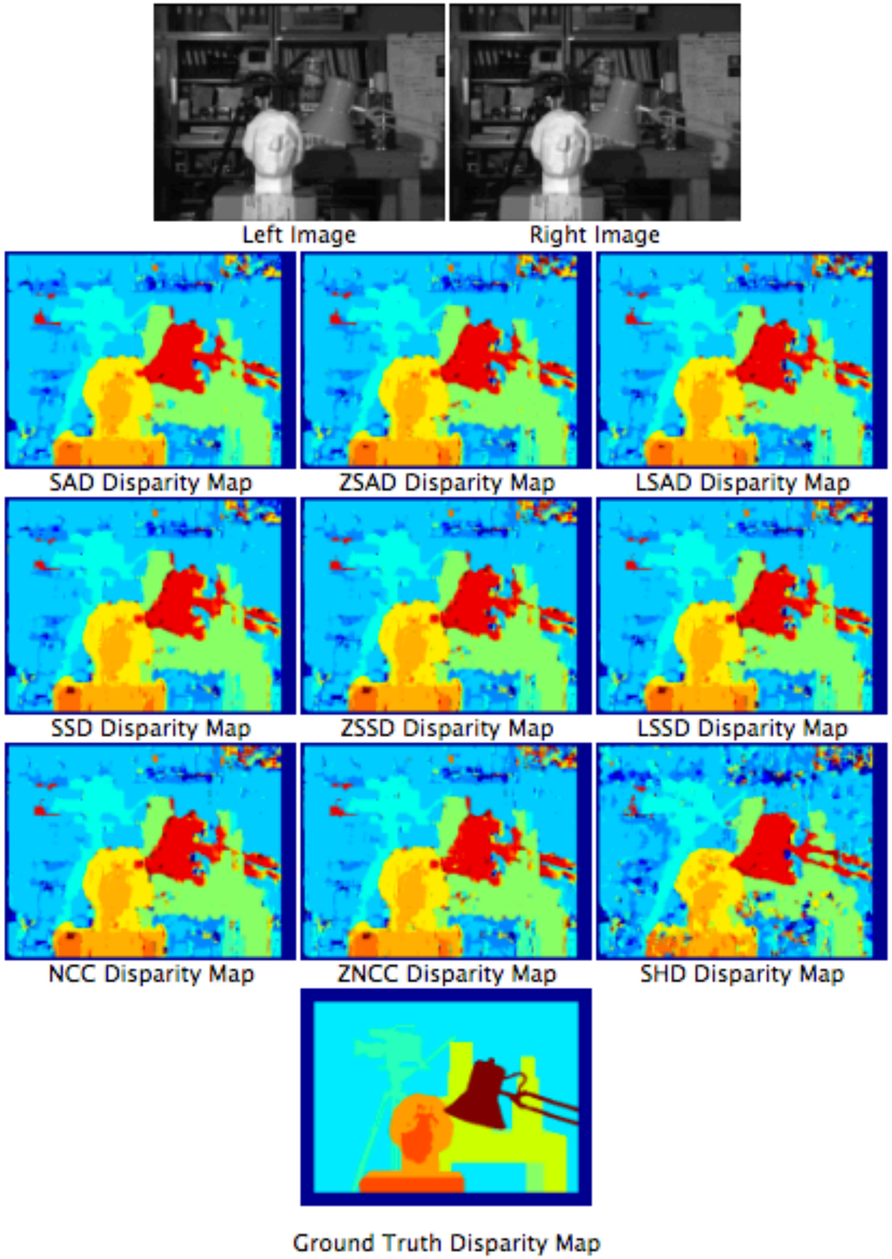
- Many metrics!
  - SSD
  - Normalized Cross-Correlation

Similarity Measure	Formula
Sum of Absolute Differences (SAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - I_2(x+i,y+j)\rvert$
Zero-mean Sum of Absolute Differences (ZSAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y$
Locally scaled Sum of Absolute Differences (LSAD)	$\sum_{(i,j)\in W} \lvert I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+$
Sum of Squared Differences (SSD)	$\sum_{(i,j)\in W} \left(I_1(i,j) - I_2(x+i,y+j)\right)^2$
Zero-mean Sum of Squared Differences (ZSSD)	$\sum_{(i,j)\in W} (I_1(i,j) - \bar{I}_1(i,j) - I_2(x+i,y+j) + \bar{I}_2(x+i,y$
Locally scaled Sum of Squared Differences (LSSD)	$\sum_{(i,j)\in W} \left(I_1(i,j) - \frac{\bar{I}_1(i,j)}{\bar{I}_2(x+i,y+j)} I_2(x+i,y+\right.$
Normalized Cross Correlation (NCC)	$\frac{\sum_{(i,j)\in W} I_1(i,j).I_2(x+i,y+j)}{\sqrt[2]{\sum_{(i,j)\in W} I_1^2(i,j).\sum_{(i,j)\in W} I_2^2(x+i,y+j)}}$
Zero-mean Normalized Cross Correlation (ZNCC)	$\frac{\sum_{(i,j)\in W}(I_1(i,j) - \bar{I}_1(i,j)).(I_2(x+i,y+j) - \bar{I}_2(x+i,y+j)}{\sqrt[2]{\sum_{(i,j)\in W}(I_1(i,j) - \bar{I}_1(i,j))^2.\sum_{(i,j)\in W}(I_2(x+i,y+j) - \bar{I}_2(x+i,y}}$
Sum of Hamming Distances (SHD)	$\sum_{(i,j)\in W} I_1(i,j)\ bitwiseXOR\ I_2(x+i,y+j)$

https://siddhantahuja.wordpress.com/tag/sum-of-squared-differences/

Matching score

- Many metrics!
  - SSD
  - Normalized Cross-Correlation
  - …

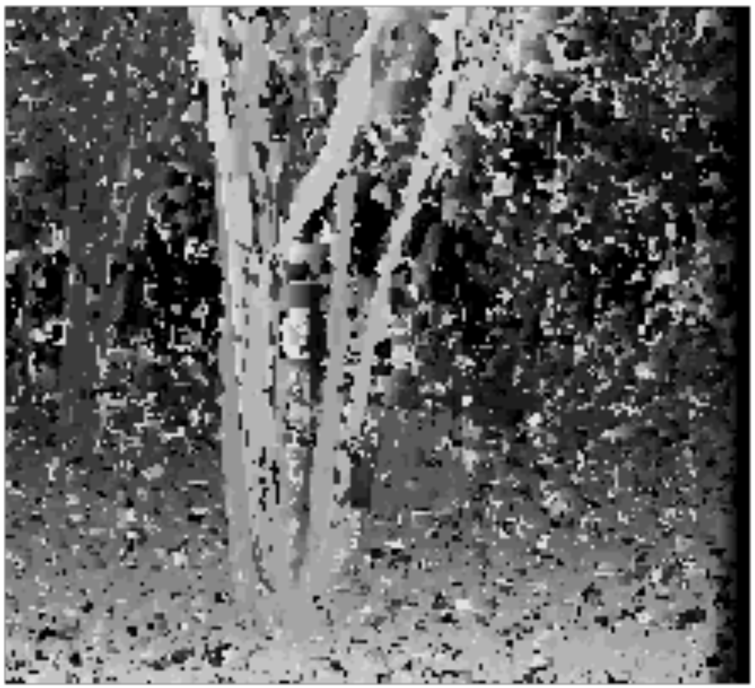


https://siddhantahuja.wordpress.com/tag/sum-of-squared-differences/

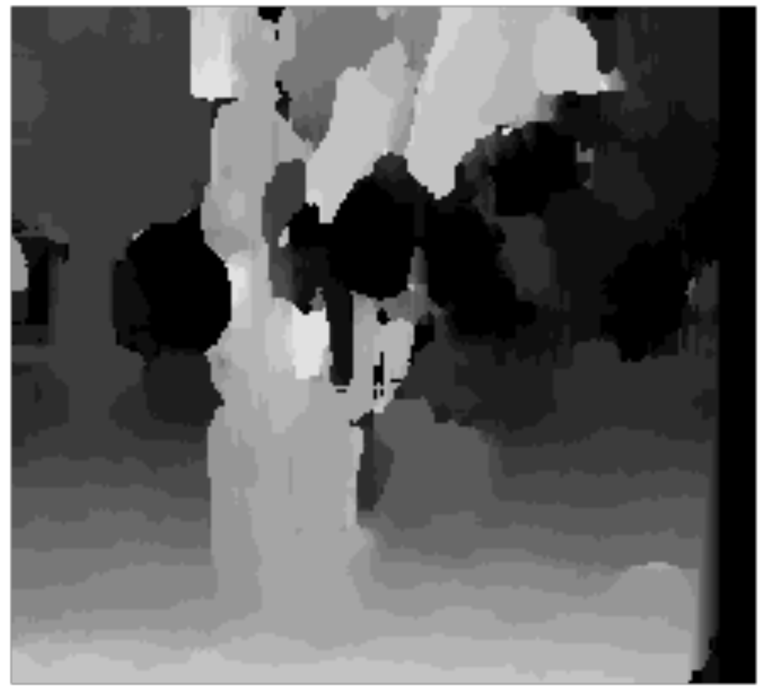
Selecting Window Size *W*

- Small window: more detail, but more noise
- Large window: more robustness, less detail
- Example:

Selecting Window Size *W*



3 pixel window



20 pixel window

Energy Minimization

- Another approach to improve quality of correspondences
- Assumption: disparities vary (mostly) smoothly
- Minimize energy function:

$$E_{data} + \lambda E_{smoothness}$$

- E_{data}: how well does disparity match data
- $E_{smoothness}$: how well does disparity match that of neighbors – regularization

Energy Minimization

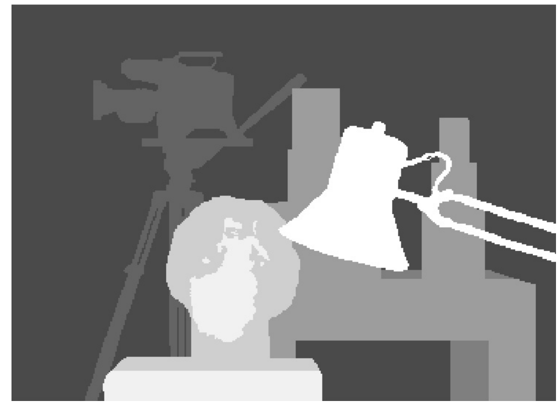
- Hard to find global minima of non-smooth functions
  - Many local minima
  - Provably NP-hard
- Practical algorithms look for approximate minima (e.g., simulated annealing)

Results

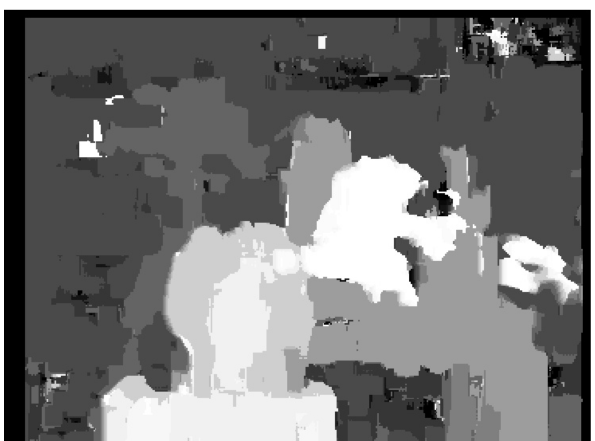
Image sequence

Results

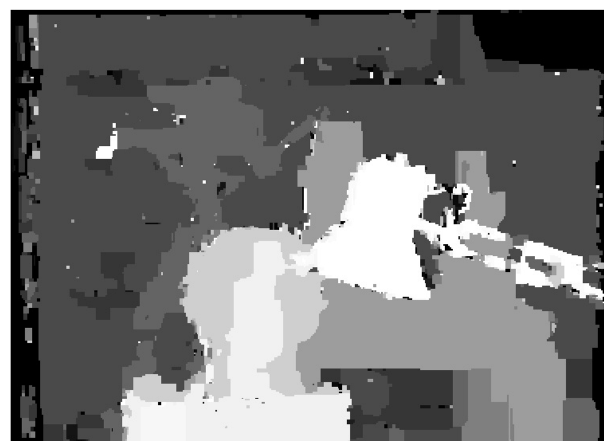
Image



Ground truth



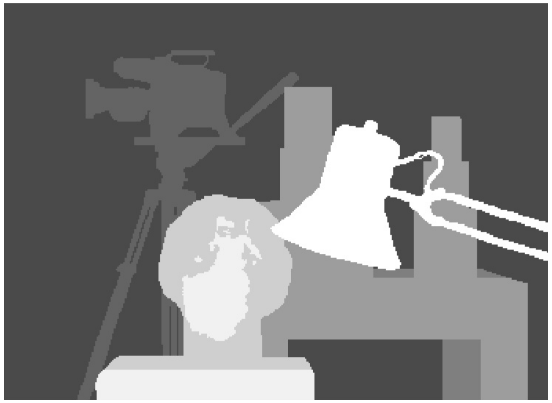
Normalized correlation



Simulated annealing

Results

Image

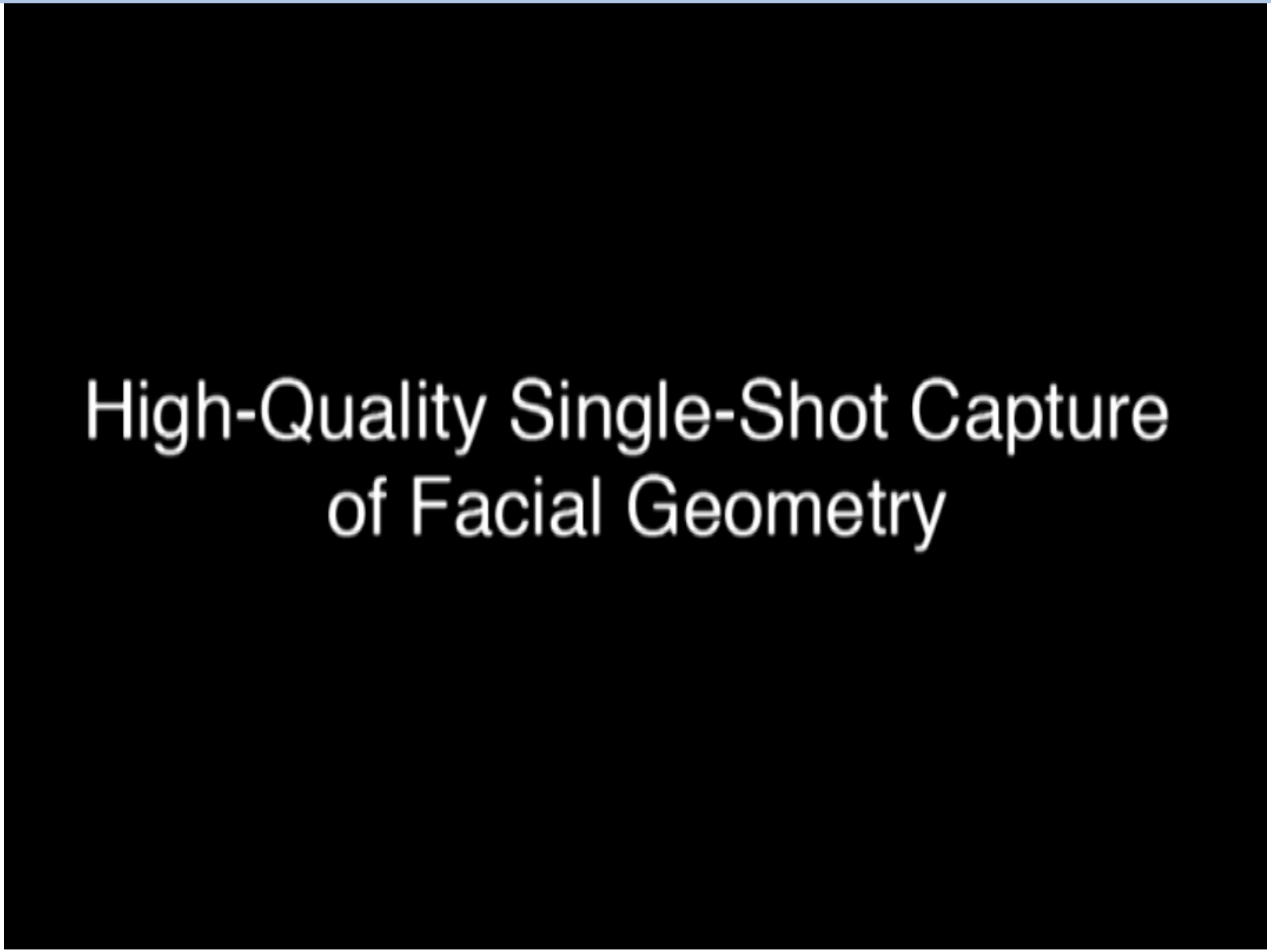


Ground truth

Graph Cut: Swap algorithm

Graph Cut: Expansion algorithm

Results: Scanning Faces



High-Quality Single-Shot Capture of Facial Geometry, Beeler et al. , Siggraph 2011

Results: Large Scale Reconstruction

The Visual Turing Test for Scene Reconstruction, Shan et al. , 3DV 2013

Stereo

- Advantages:
  - Passive
  - Cheap hardware (2 cameras)
  - Easy to accommodate motion
  - Intuitive analogue to human vision
- Disadvantages:
  - Only acquire good data at “features”
  - Sparse, relatively noisy data (correspondence is hard)
  - Bad around silhouettes
  - Confused by non-diffuse surfaces
- Variant: multibaseline stereo to reduce ambiguity

Range Acquisition Taxonomy

- Optical methods
  - Passive
    - Shape from X:
      - stereo
      - motion
      - shading
      - texture
      - focus
      - defocus
  - Active
    - Active variants of passive methods
      - Stereo w. projected texture
      - Active depth from defocus
      - Photometric stereo
    - Time of flight
    - Triangulation

Active Optical Methods

- Advantages:
  - Usually can get dense data
  - Usually much more robust and accurate than passive techniques
- Disadvantages:
  - Introduces light into scene (distracting, etc.)
  - Not motivated by human vision

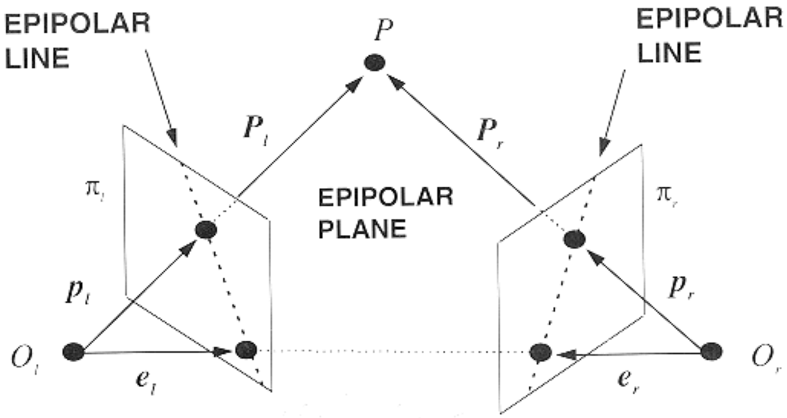
Range Acquisition Taxonomy

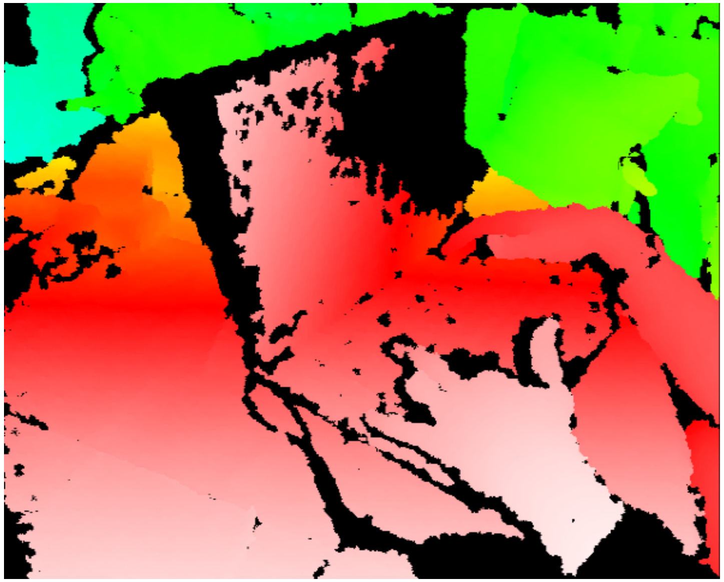
- Optical methods
  - Passive
    - Shape from X:
      - stereo
      - motion
      - shading
      - texture
      - focus
      - defocus
  - Active
    - Active variants of passive methods
      - Stereo w. projected texture
      - Active depth from defocus
      - Photometric stereo
    - Time of flight
    - Triangulation

Active Variants of Passive Techniques

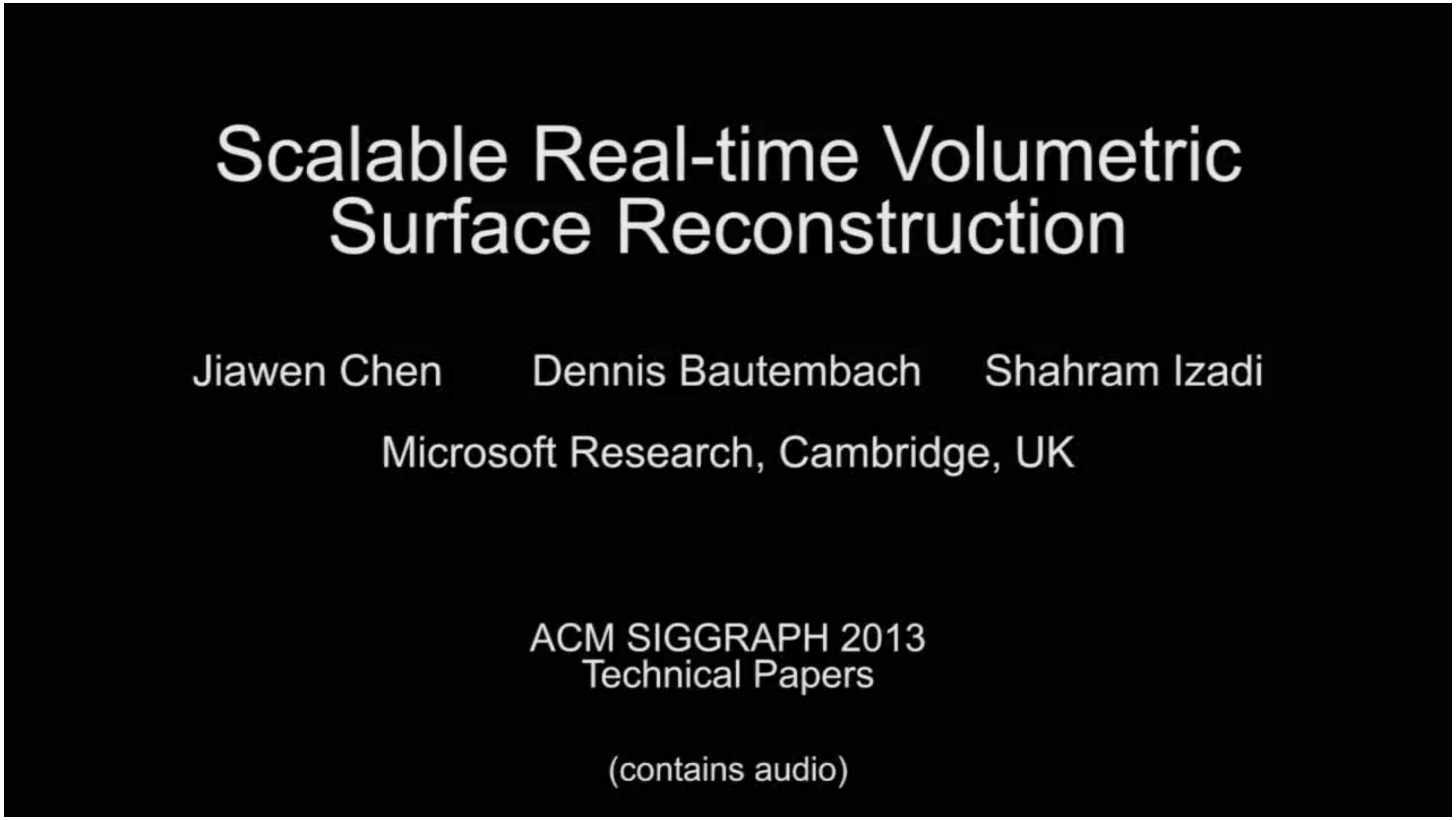
- Regular stereo with projected texture
  - Provides features for correspondence
- Photometric stereo
  - Shape from shading with multiple known lights

MS Kinect: Stereo with Projected Texture





MS Kinect Fusion



http://www.youtube.com/watch?v=NsrmniEvO4s

Photometric Stereo

Dynamic Shape Capture using
Multi-View Photometric Stereo

Daniel Vlasic
Pieter Peers
Ilya Baran
Paul Debevec
Jovan Popović
Szymon Rusinkiewicz
Wojciech Matusik

Dynamic Shape Capture using Multi-View Photometric Stereo, Vlasic et al. , Siggraph Asia 2009

Photometric Stereo

Internet Photometric Stereo

Photometric Stereo using Internet Images.
Boxin Shi, Kenji Inose, Yasuyuki Matsushita, Ping Tan, Sai-Kit Yeung, and Katsushi Ikeuchi, 3DV 2014

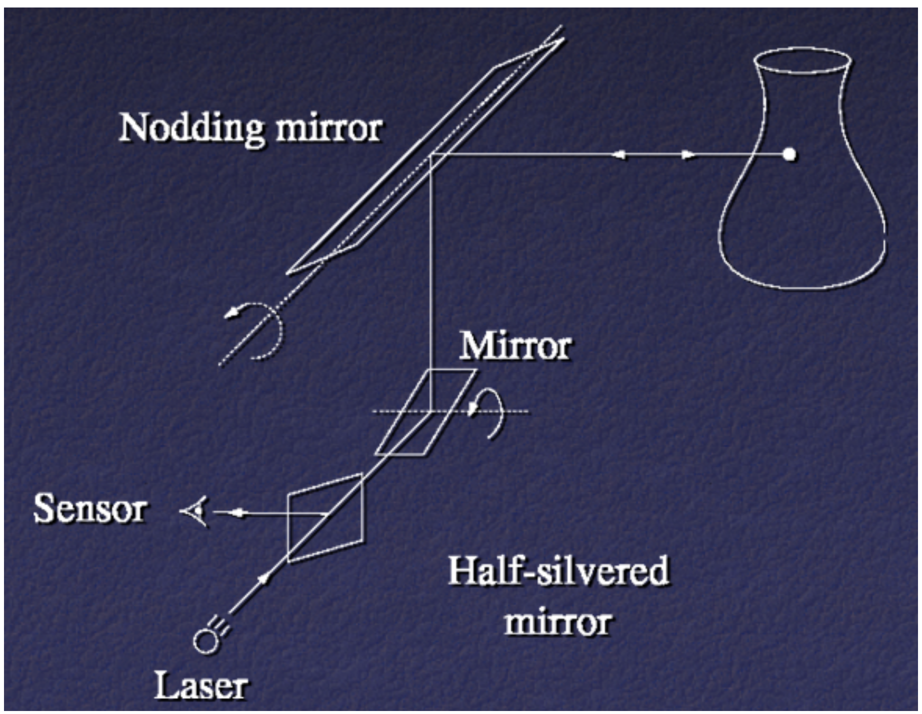
Range Acquisition Taxonomy

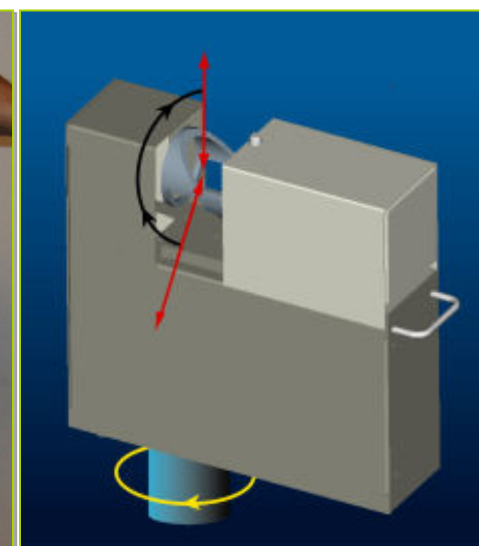
- Optical methods
  - Passive
    - Shape from X:
      - stereo
      - motion
      - shading
      - texture
      - focus
      - defocus
  - Active
    - Active variants of passive methods
      - Stereo w. projected texture
      - Active depth from defocus
      - Photometric stereo
    - Time of flight
    - Triangulation

Pulsed Time of Flight

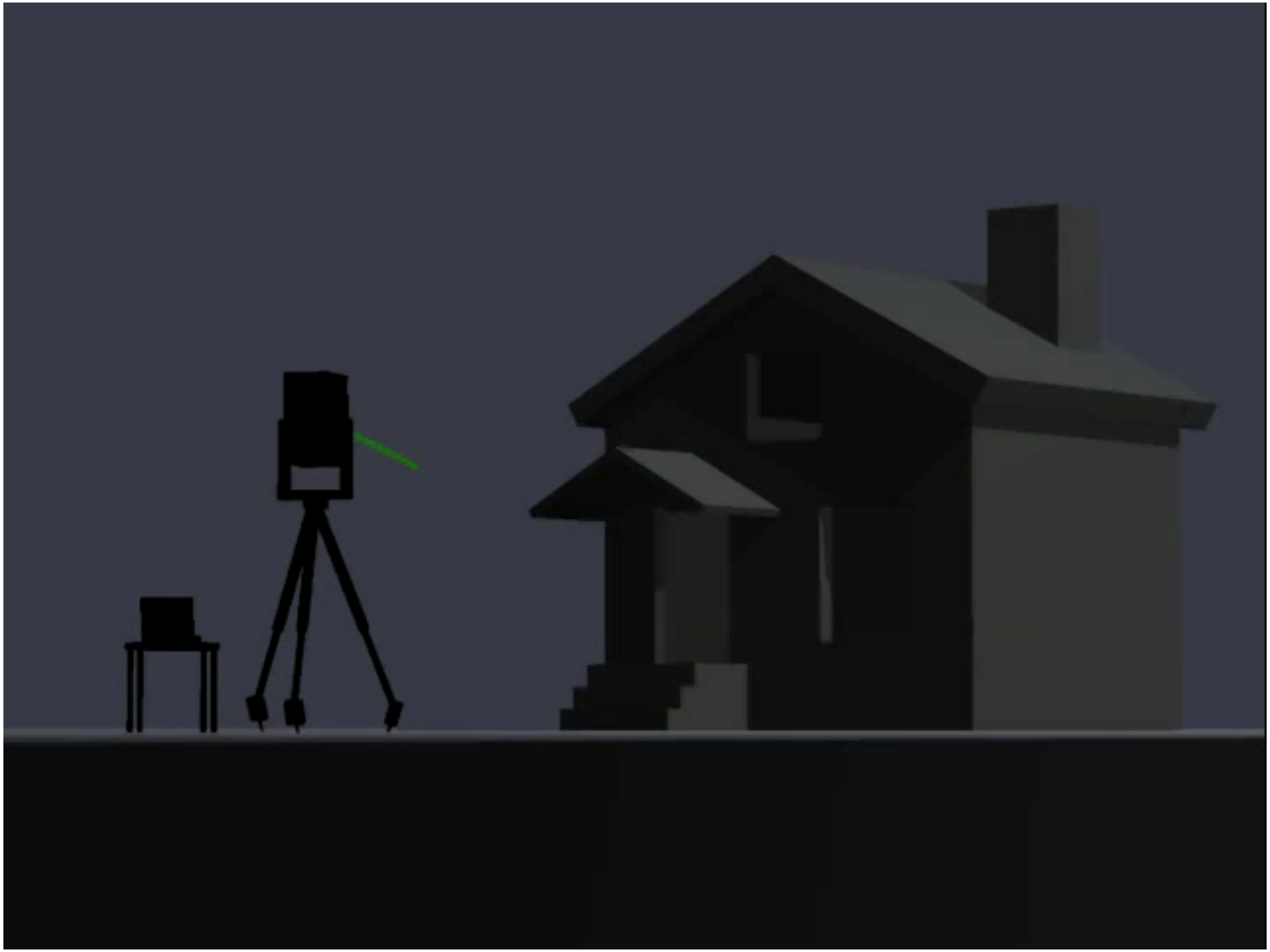
- Basic idea: send out pulse of light (usually laser), time how long it takes to return

$$d = \frac{1}{2} c \Delta t$$





Pulsed Time of Flight



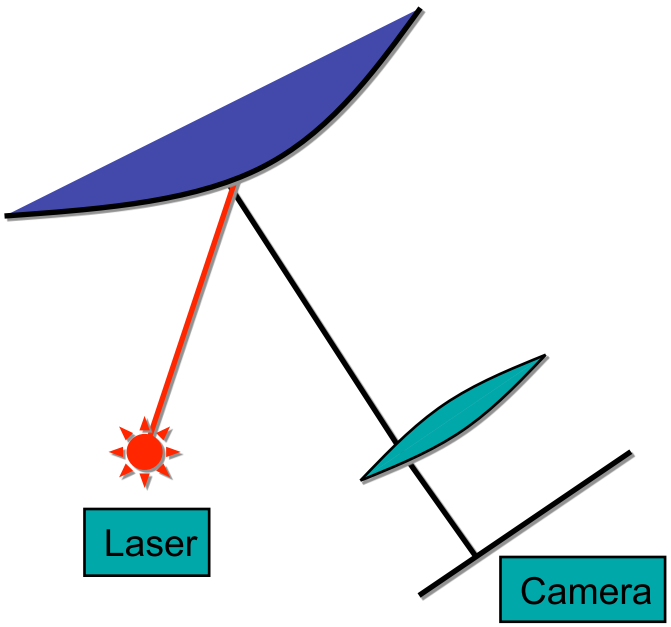
Pulsed Time of Flight

- Advantages:
  - Large working volume (up to 100 m.)
- Disadvantages:
  - Not-so-great accuracy (at best ~5 mm.)
    - Requires getting timing to ~30 picoseconds
    - Does not scale with working volume
- Often used for scanning buildings, rooms, archeological sites, etc.

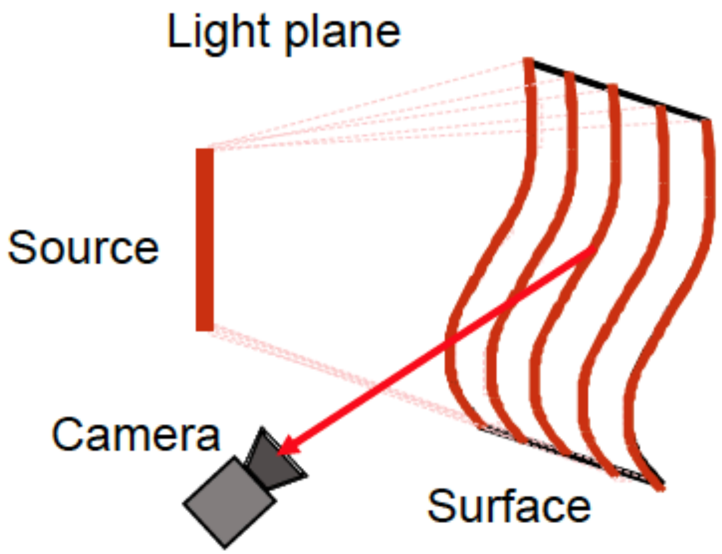
Range Acquisition Taxonomy

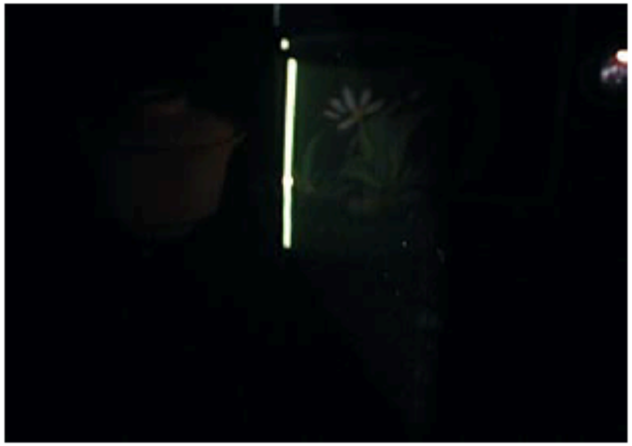
- Optical methods
  - Passive
    - Shape from X:
      - stereo
      - motion
      - shading
      - texture
      - focus
      - defocus
  - Active
    - Active variants of passive methods
      - Stereo w. projected texture
      - Active depth from defocus
      - Photometric stereo
    - Time of flight
    - Triangulation

Triangulation



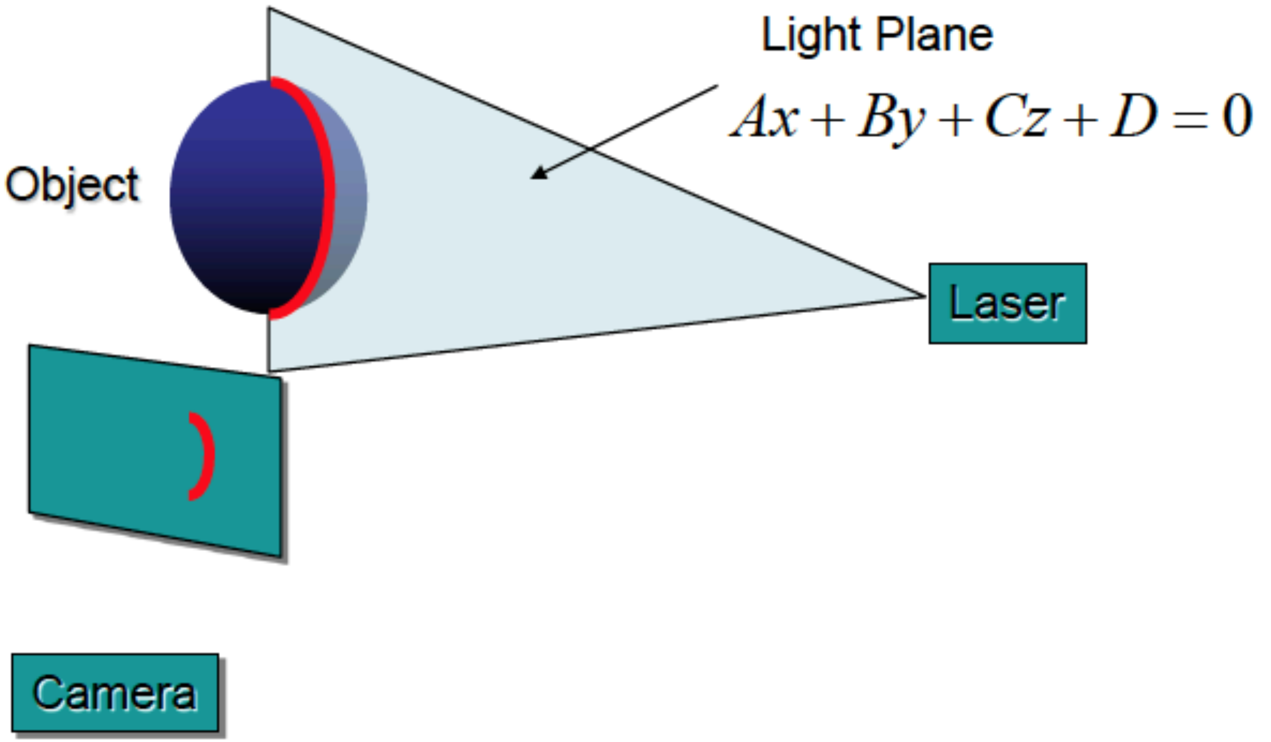
Light stripe scanning - single stripe





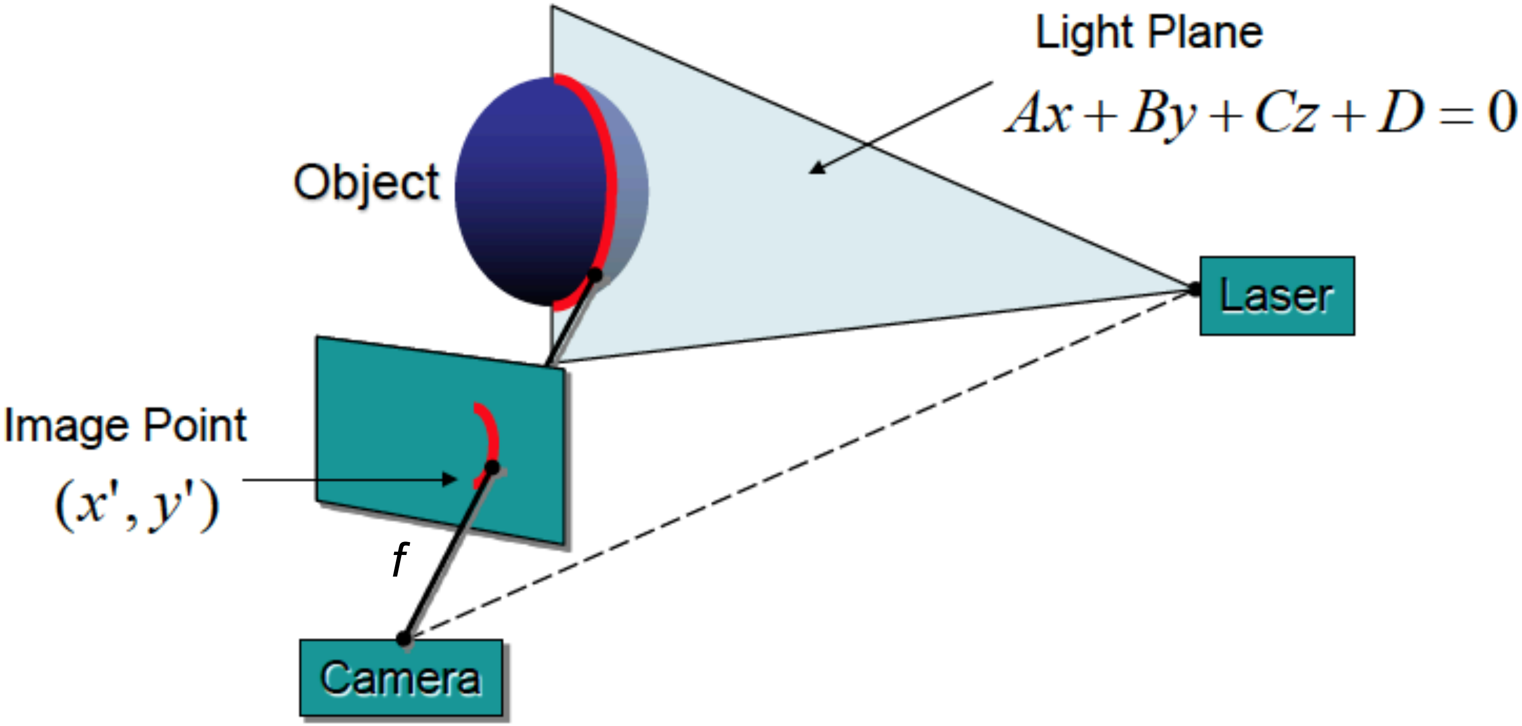
- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning
  - Good for high resolution 3D, but needs many images and takes time

Triangulation



- Project laser stripe onto object

Triangulation



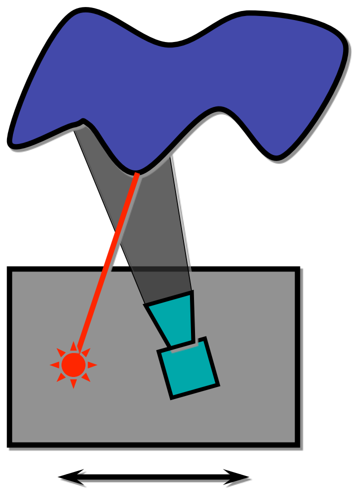
- Depth from ray-plane triangulation:
  - Intersect camera ray with light plane

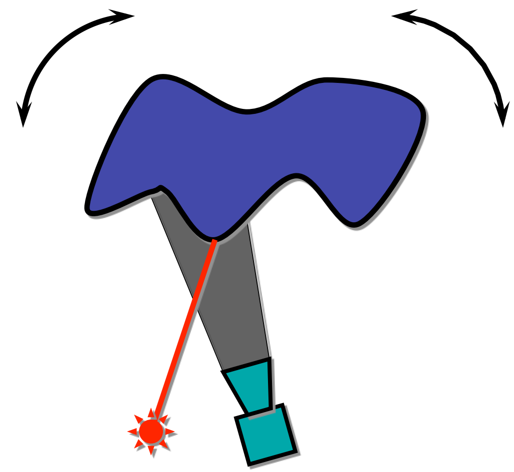
$$x = x'z / f \qquad z = \frac{-Df}{Ax' + By' + Cf}$$
$$y = y'z / f$$

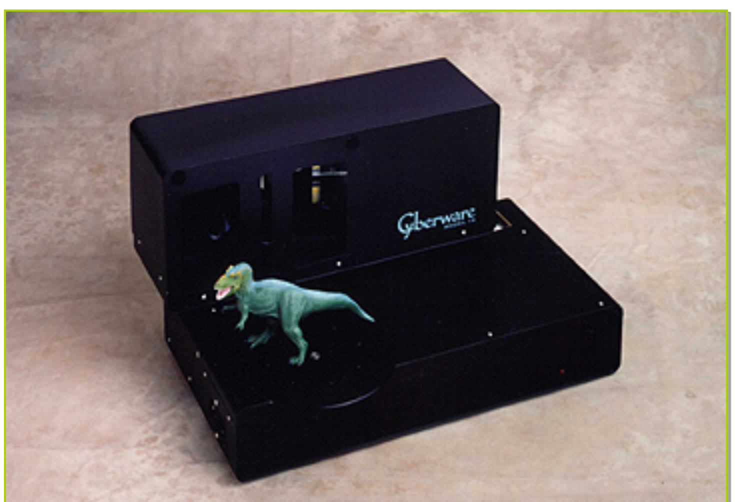
Triangulation: Moving Camera and Illumination

- Moving independently leads to problems with calibration, focus, resolution
- Most scanners mount camera and light source rigidly, move them as a unit

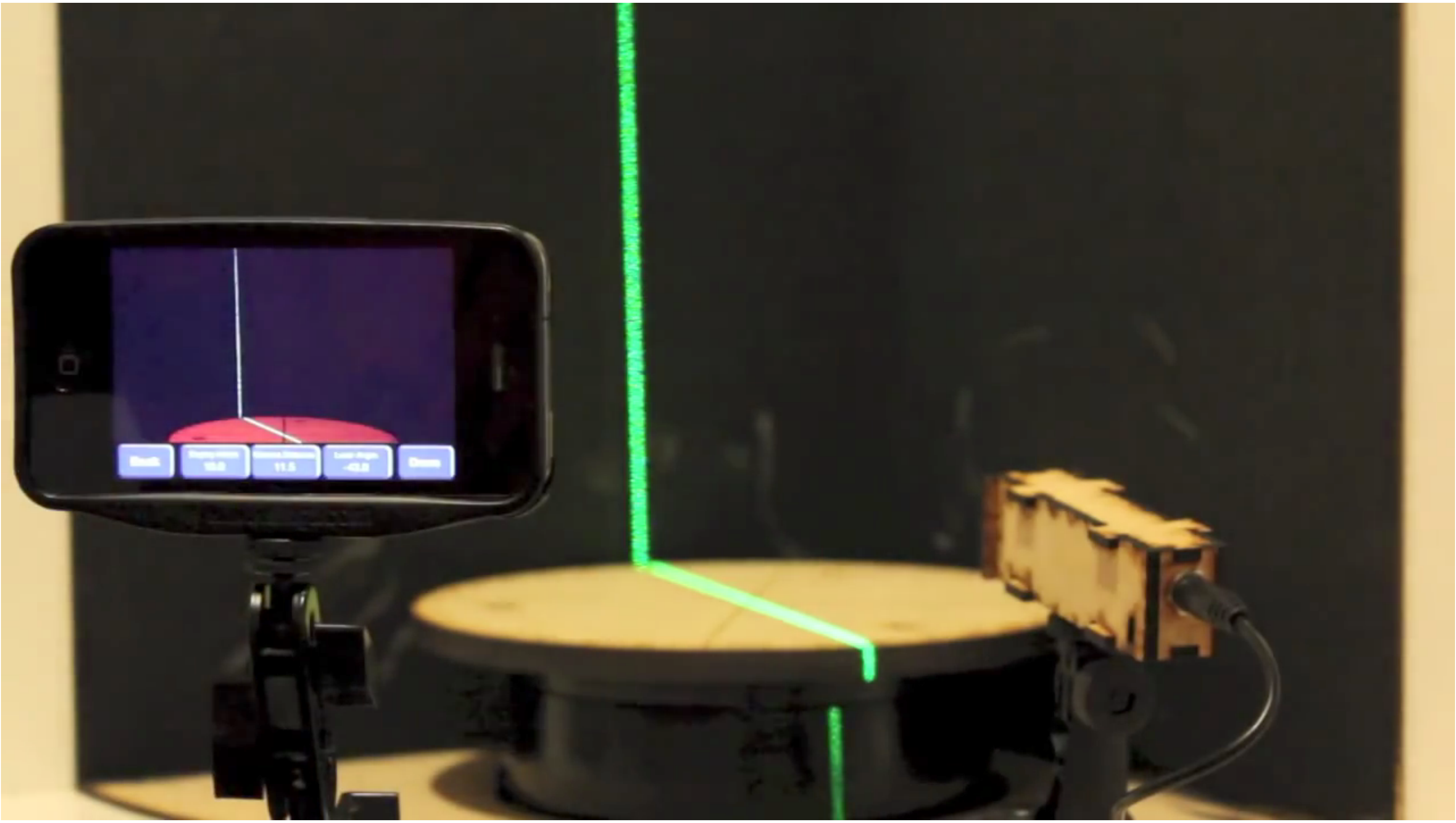
Triangulation: Moving Camera and Illumination



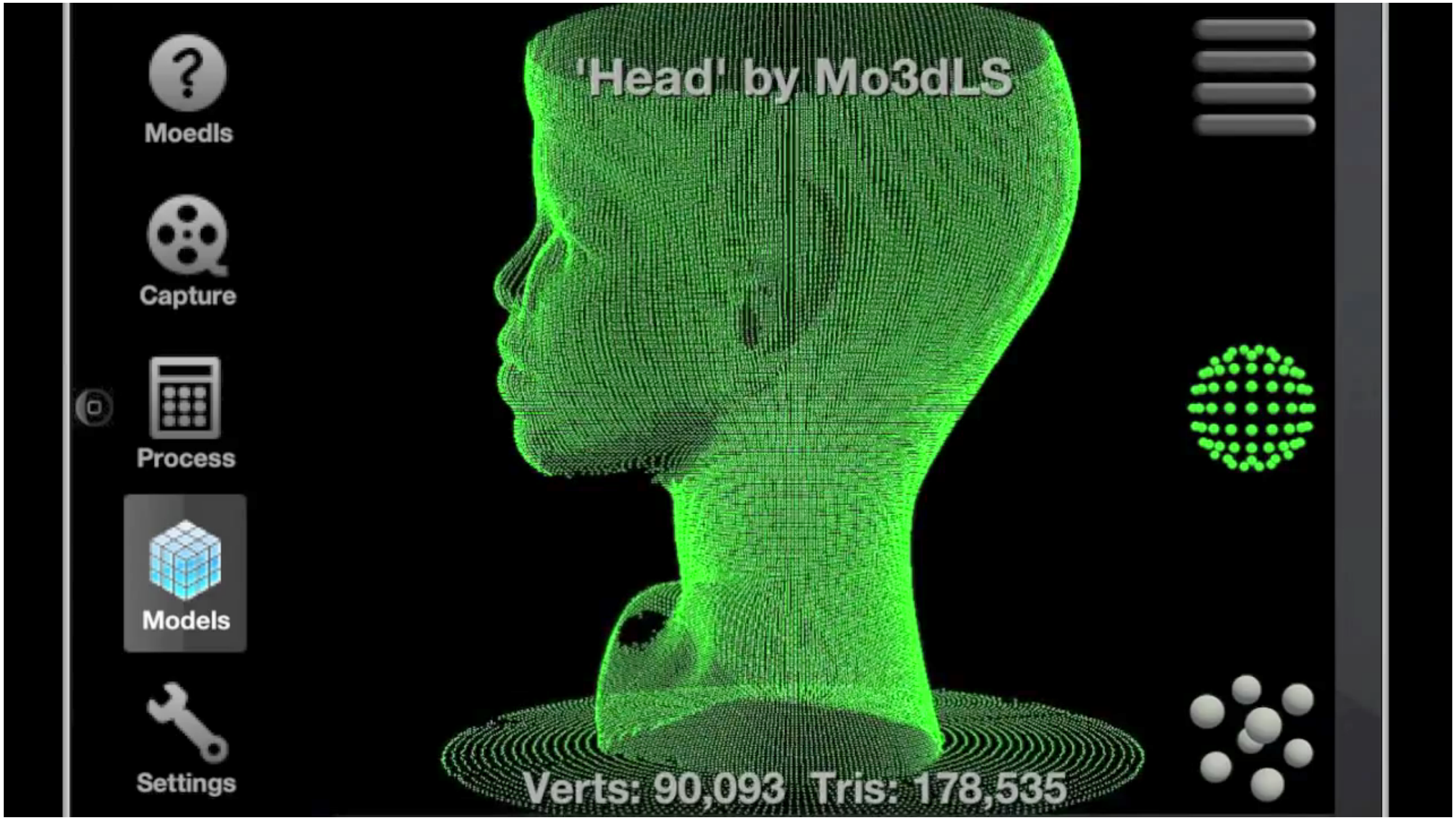




Triangulation: Moving Camera and Illumination



Triangulation: Moving Camera and Illumination

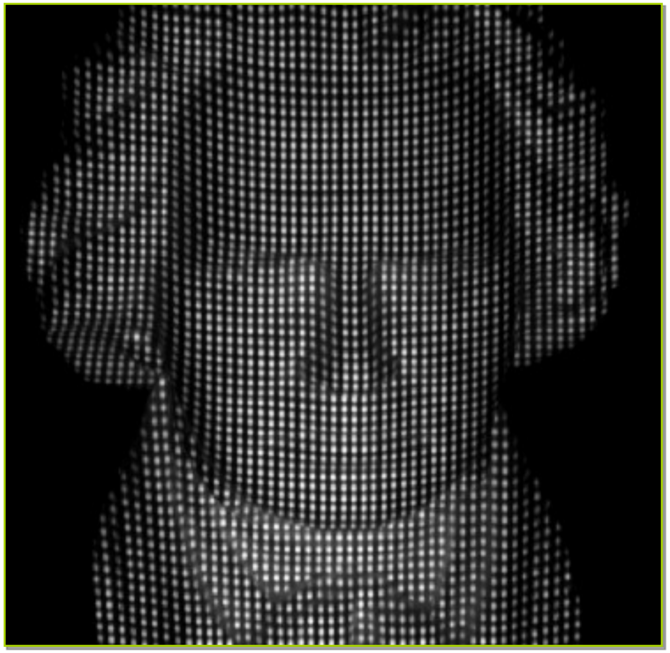


Triangulation Scanner

- Accuracy proportional to working volume (typical is ~1000:1)
- Scales down to small working volume (e.g. 5 cm. working volume, 50 μm. accuracy)

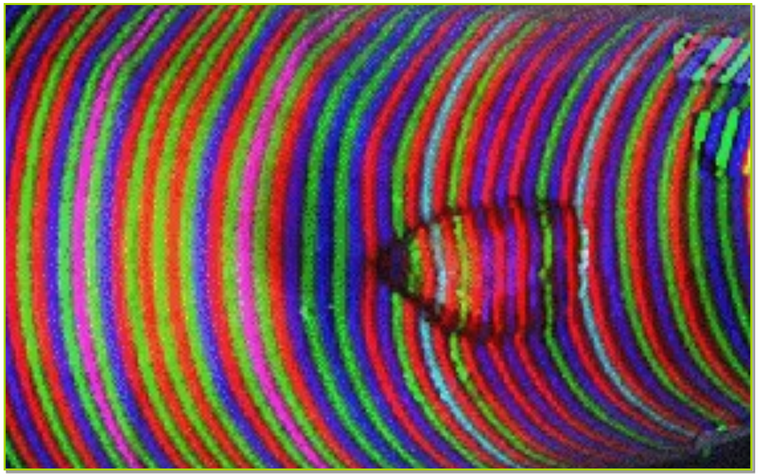
Multi-Stripe Triangulation

- To go faster, project multiple stripes
- But which stripe is which?
- Answer #1: assume surface continuity



Multi-Stripe Triangulation

- To go faster, project multiple stripes
- But which stripe is which?
- Answer #2: colored stripes (or dots)



Multi-Stripe Triangulation

- To go faster, project multiple stripes
- But which stripe is which?
- Answer #3: time-coded stripes

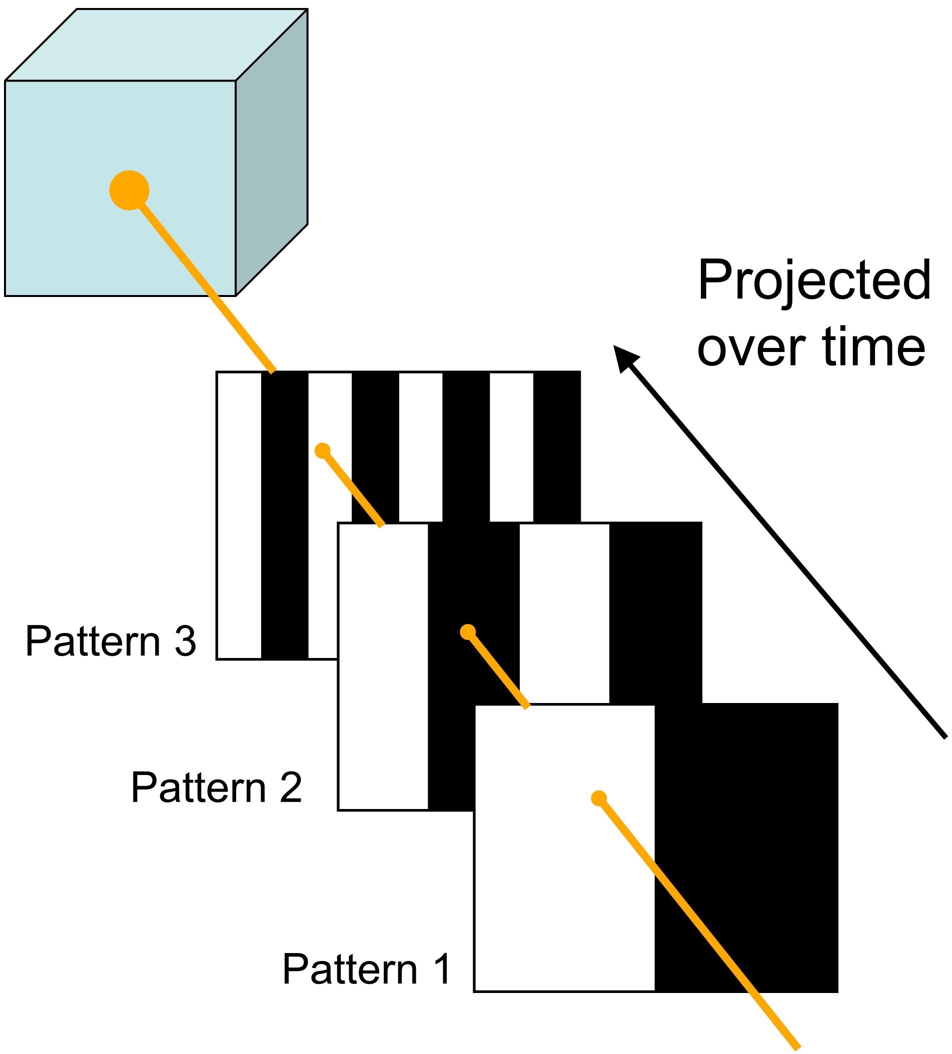
Binary Coding

Faster:

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

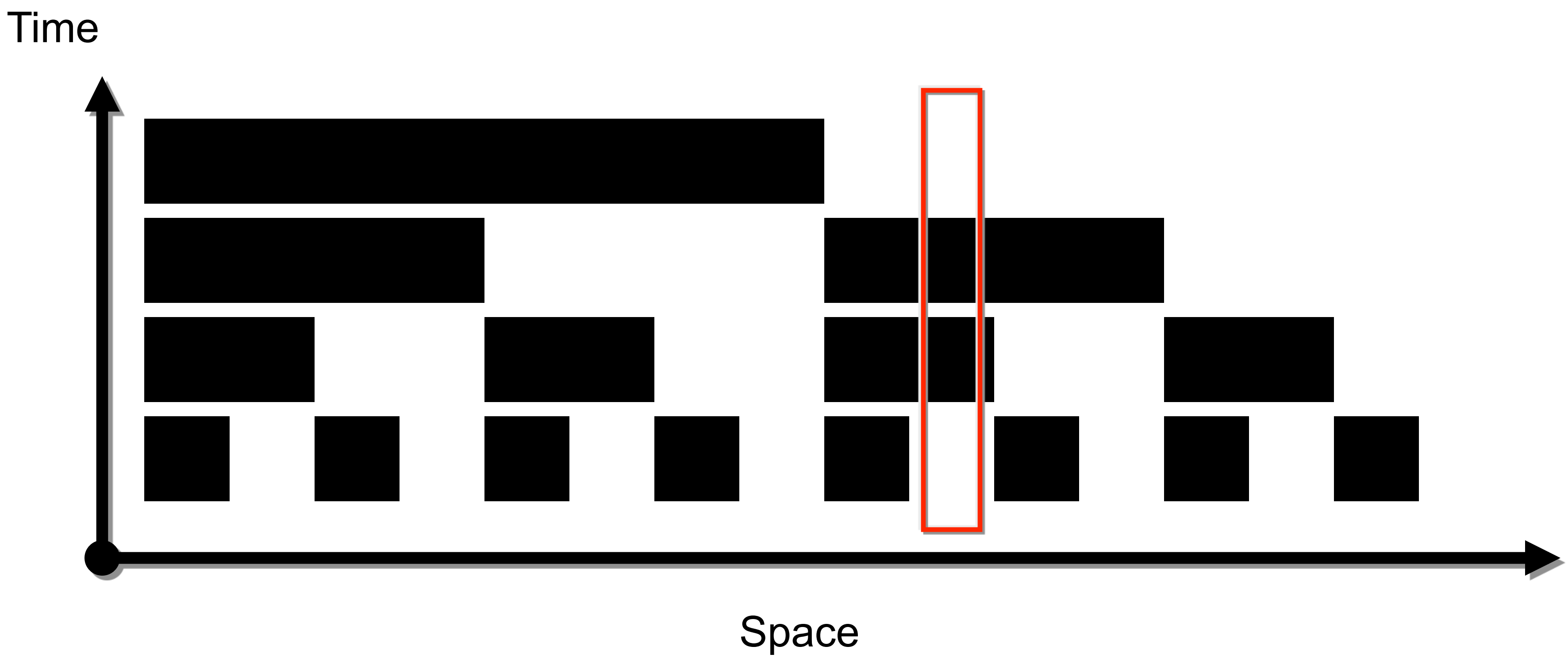
Example:

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



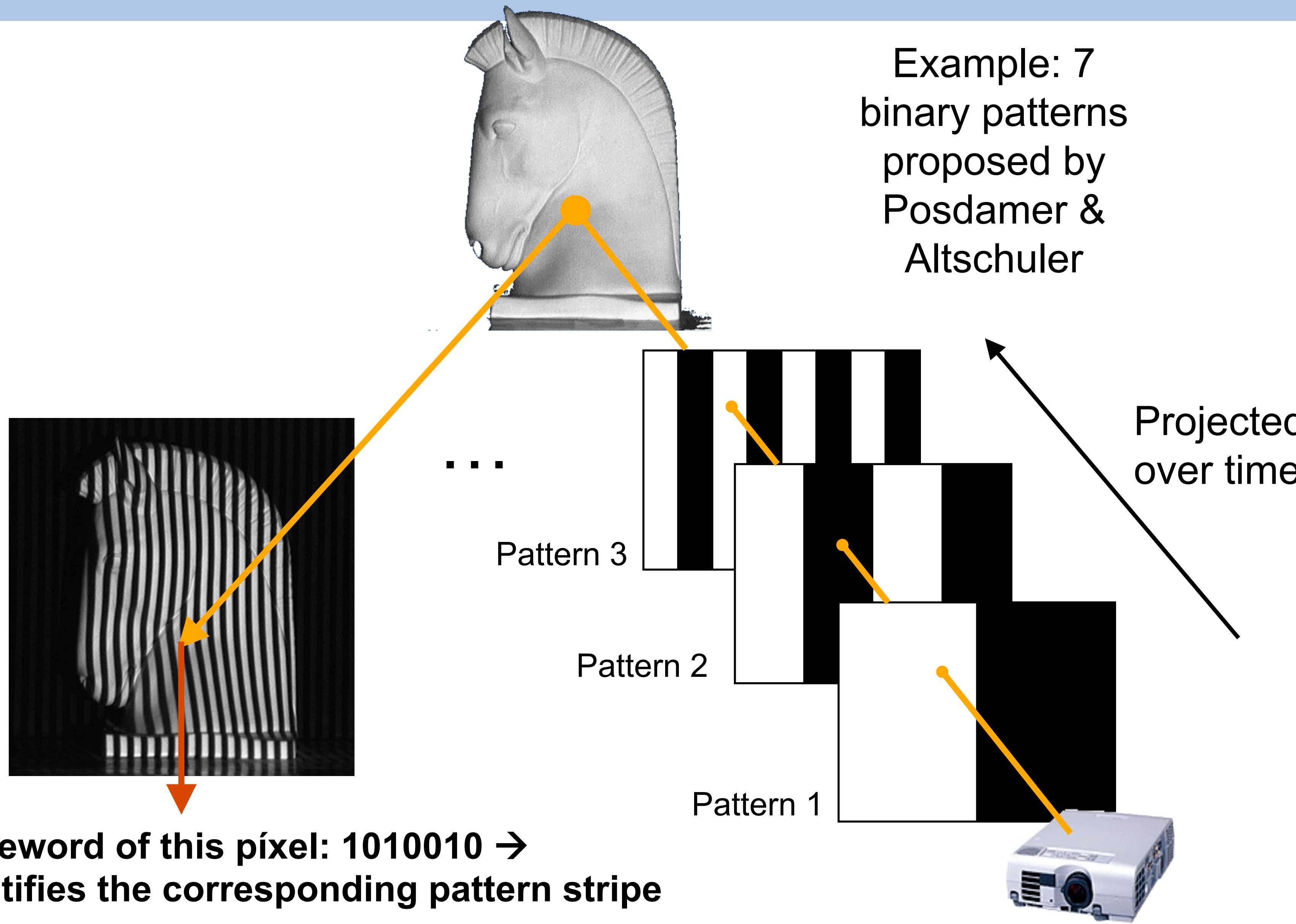
Time-Coded Light Patterns

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



Binary Coding

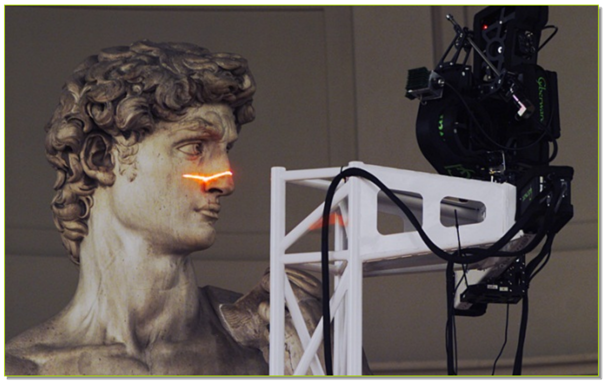
Example: 7 binary patterns proposed by Posdamer & Altschuler



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

Binary Coding

Continuum of Triangulation Methods



Multi-stripe
Multi-frame

Single-stripe

Single-frame

Slow, robust

Fast, fragile

That’s All for Today