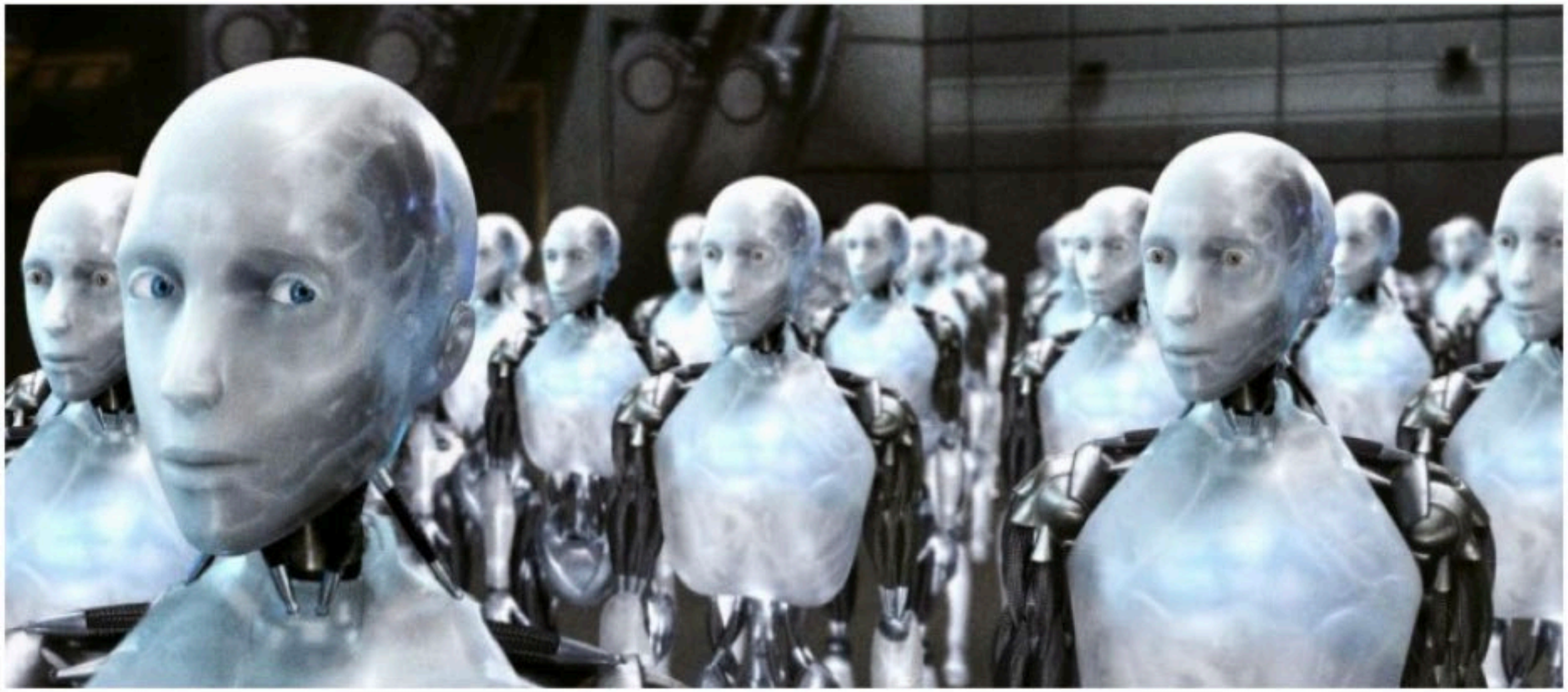


Towards 3D vision

Hansung Kim
h.kim@soton.ac.uk

3D Human Visual System

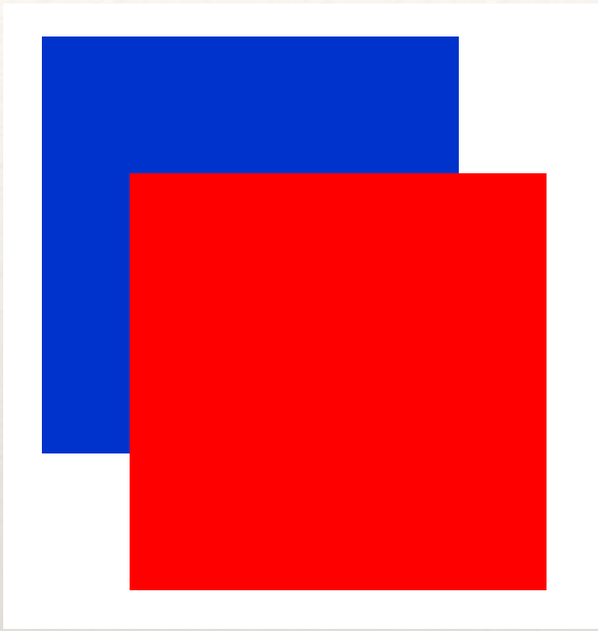
Depth Perception

- How does human perceive the world in three dimensions and the distance of an object?

Depth Perception

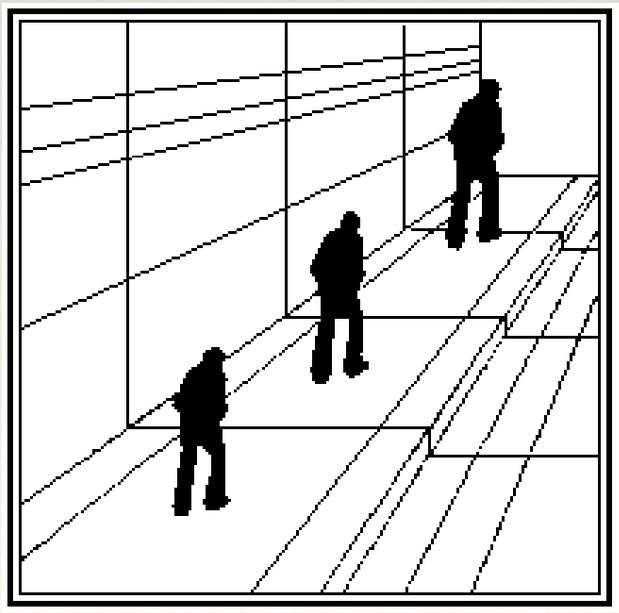
- Monocular cues (Heuristic cues)

<Motion Parallax>



<Occlusion>

<Size>

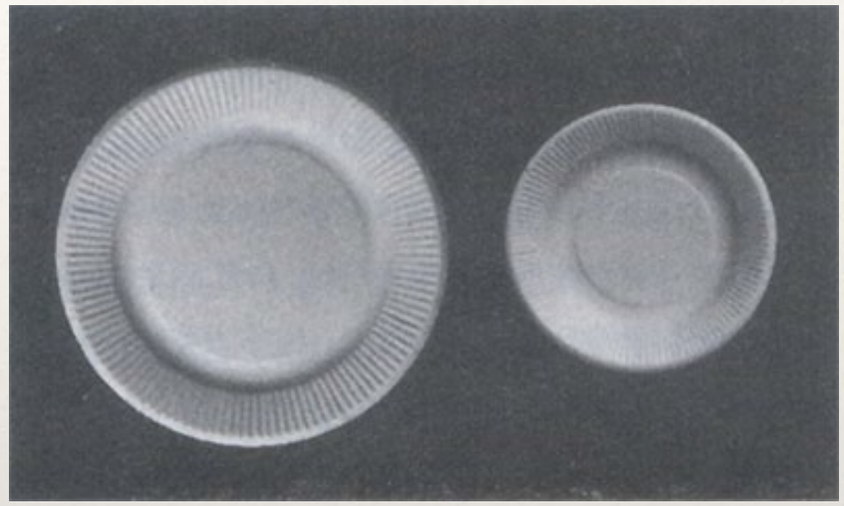


<Texture Gradient>

Depth Perception

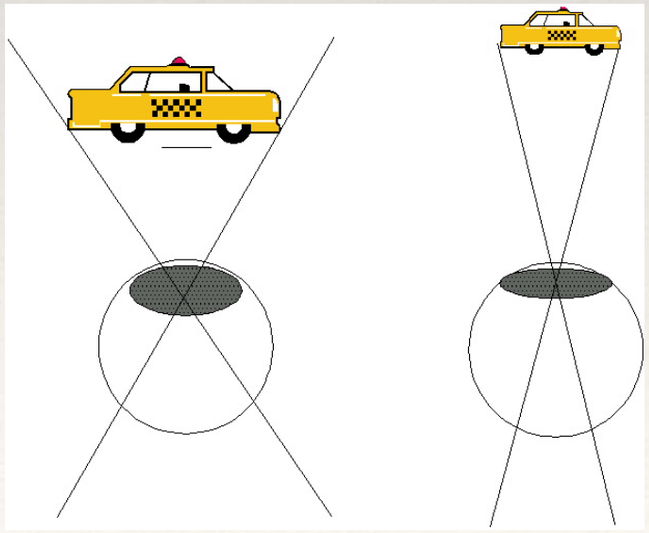
- Monocular cues (Heuristic cues)

< Aerial perspective (Distance fog)> >

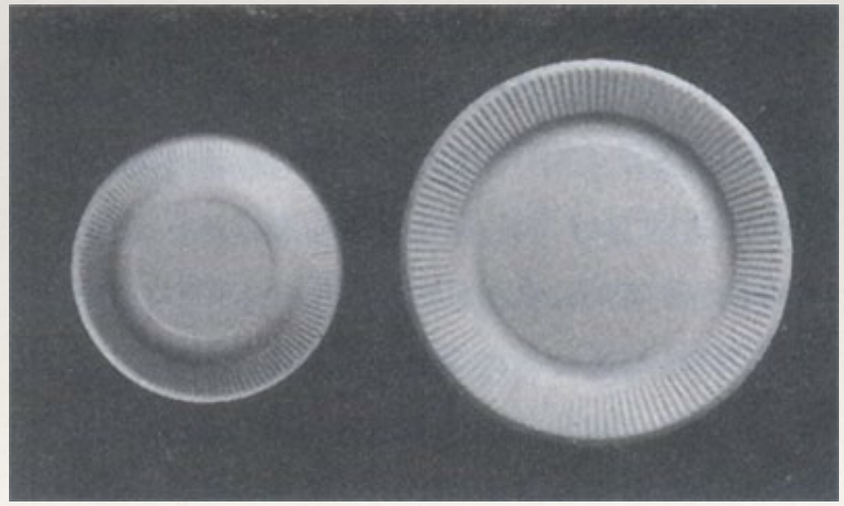


< Shading >

- Monocular cues (Physiological cues)



<Accommodation (Focus)>

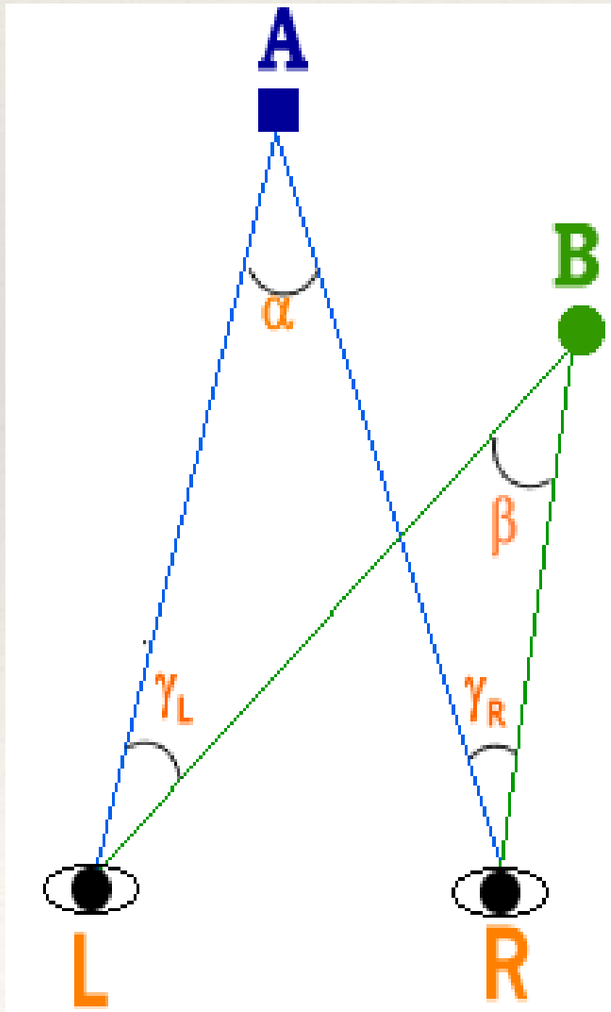


Depth Perception

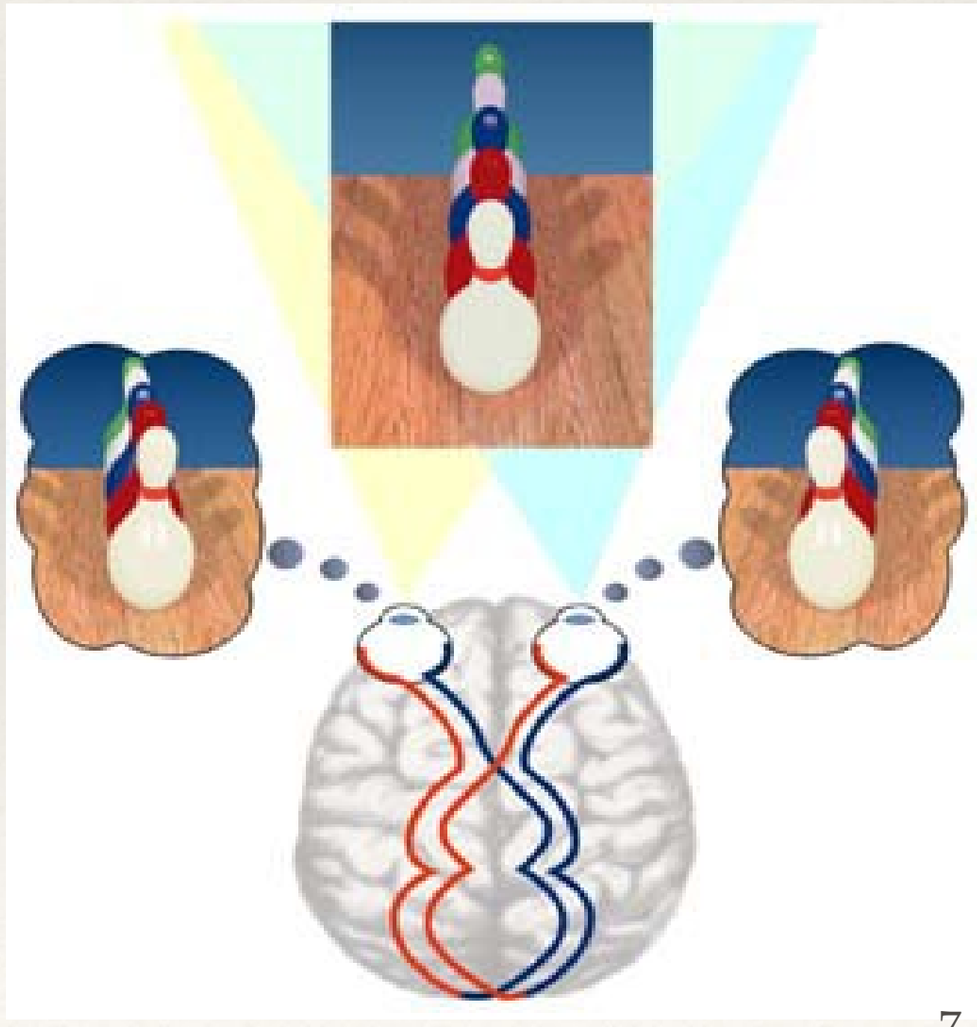
- Monocular cues are enough? Why do we have two eye?

Depth Perception

- Binocular cues
  - Convergence and Stereopsis (Binocular disparity)

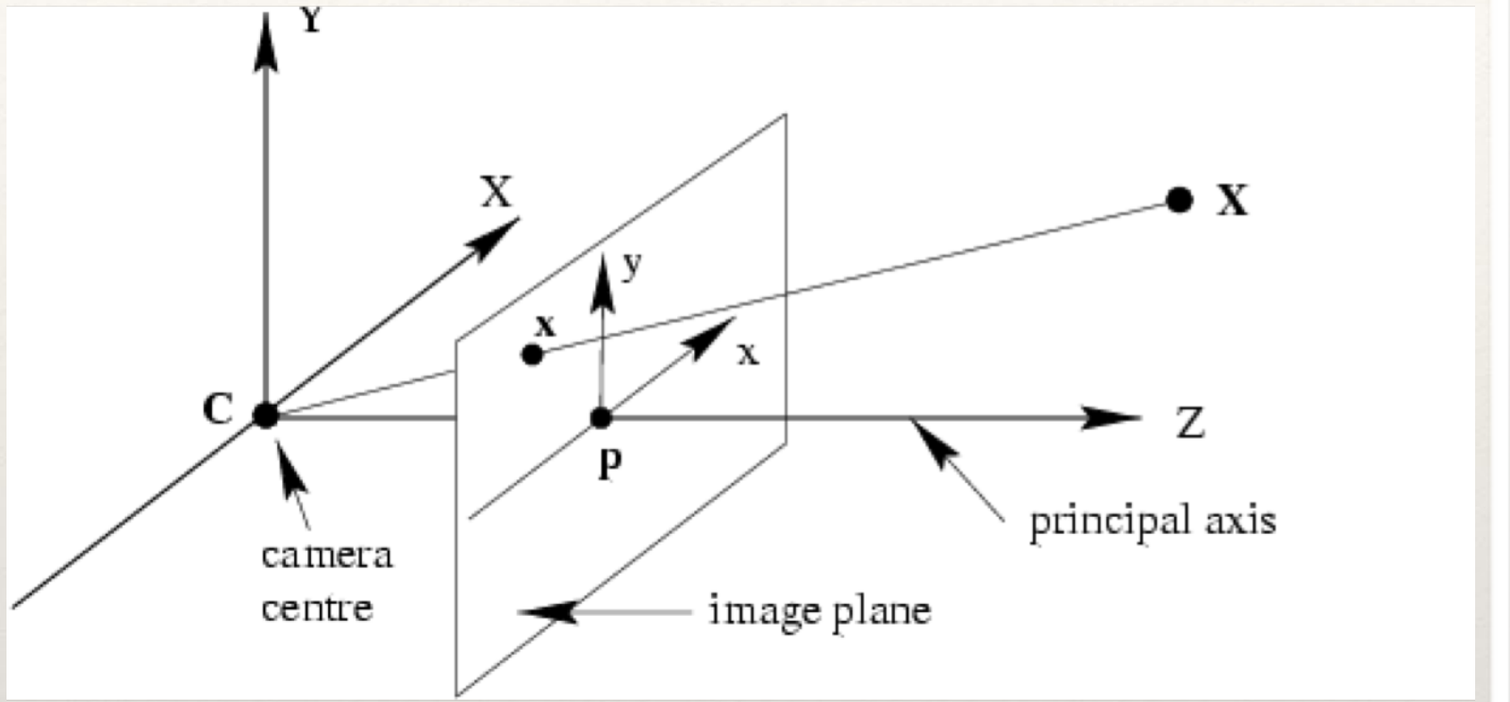


α,β: Convergence
γ_L-γ_R: Disparity



Cameras

Camera Geometry



$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathtt{t} \\ & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Camera Geometry

This is a point in the image

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathtt{t} \\ & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

This is a point in the world

Camera Geometry

These are the "*extrinsic*" parameters

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathtt{t} \\ & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Rotation of the camera in world space

Translation of the camera in world space

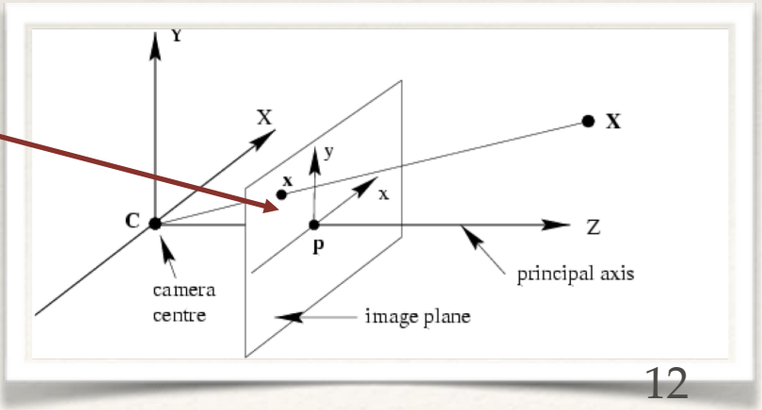
Camera Geometry

These are the "*intrinsic*" parameters

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathtt{t} \\ & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

focal length

position of the *principal point* in the image

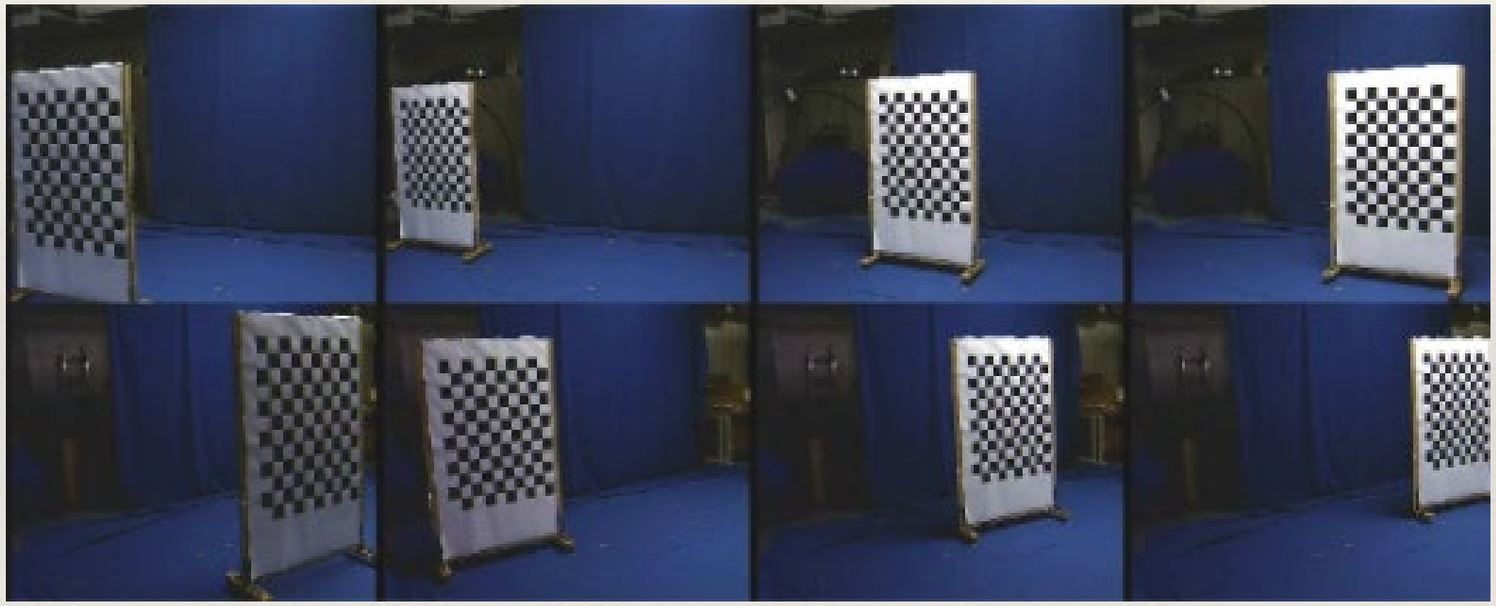


Camera Calibration

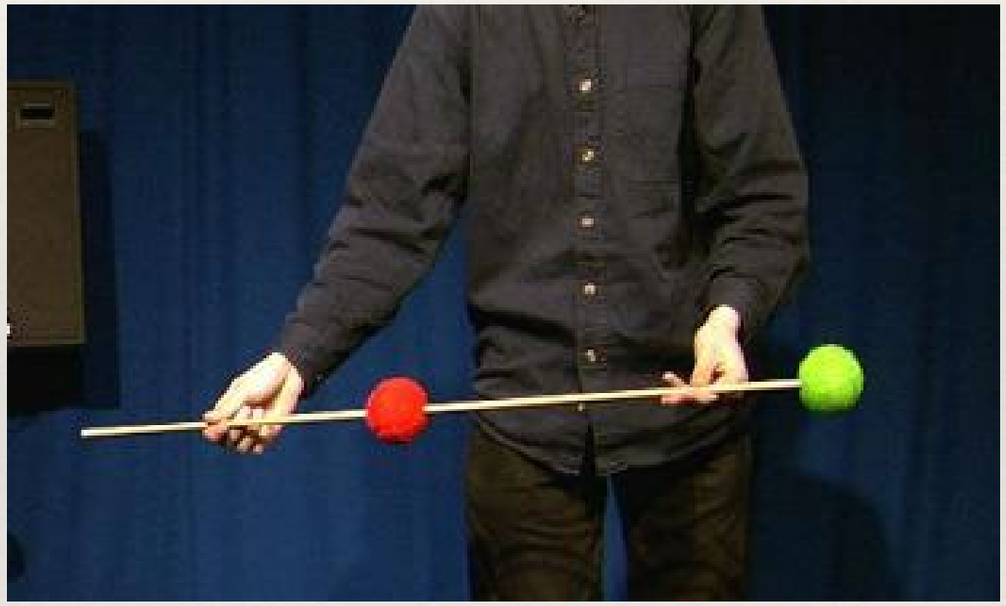
$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \end{bmatrix} \begin{bmatrix} \mathtt{R} & \mathtt{t} \\ & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

- Camera calibration is the process of estimating the intrinsic parameters of a camera
  - Also deals with learning non-linear radial distortion parameters of real camera lenses
  - Typically determined by solving sets of point correspondences from images of “calibration patterns”

Multi Camera Calibration

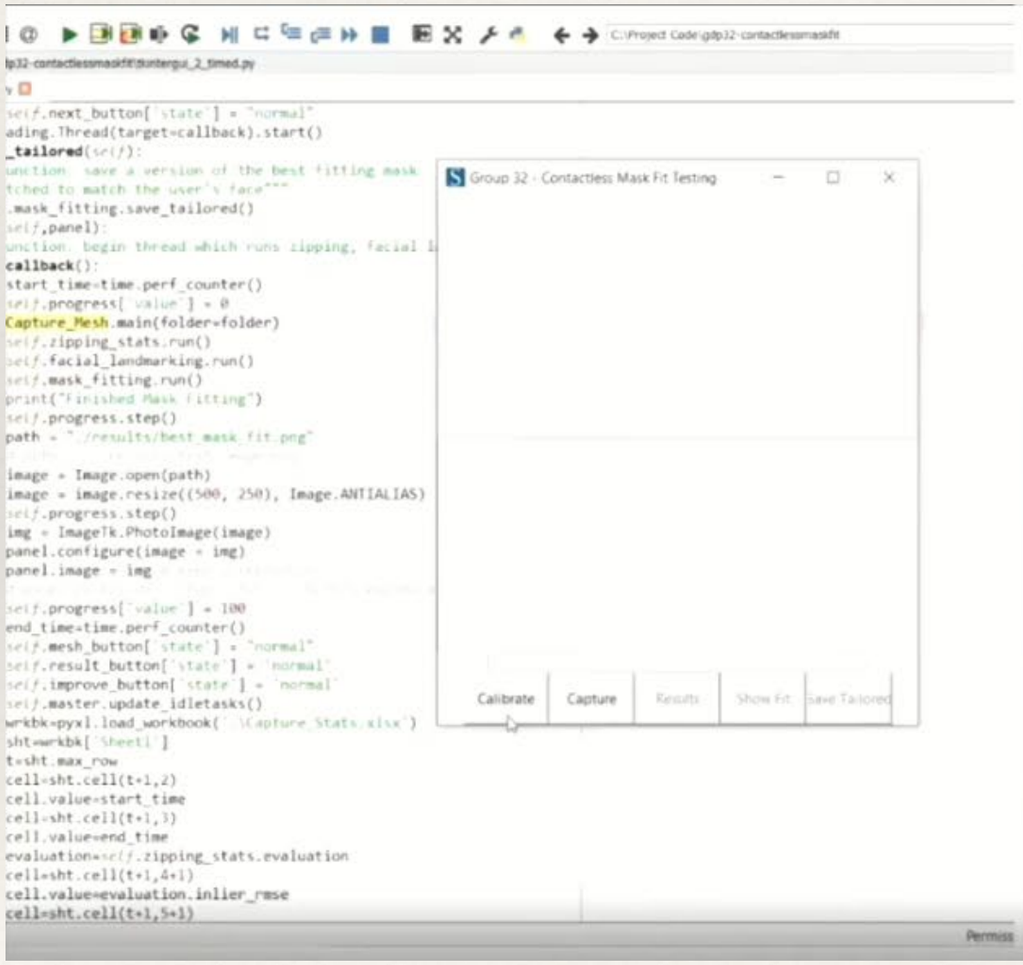


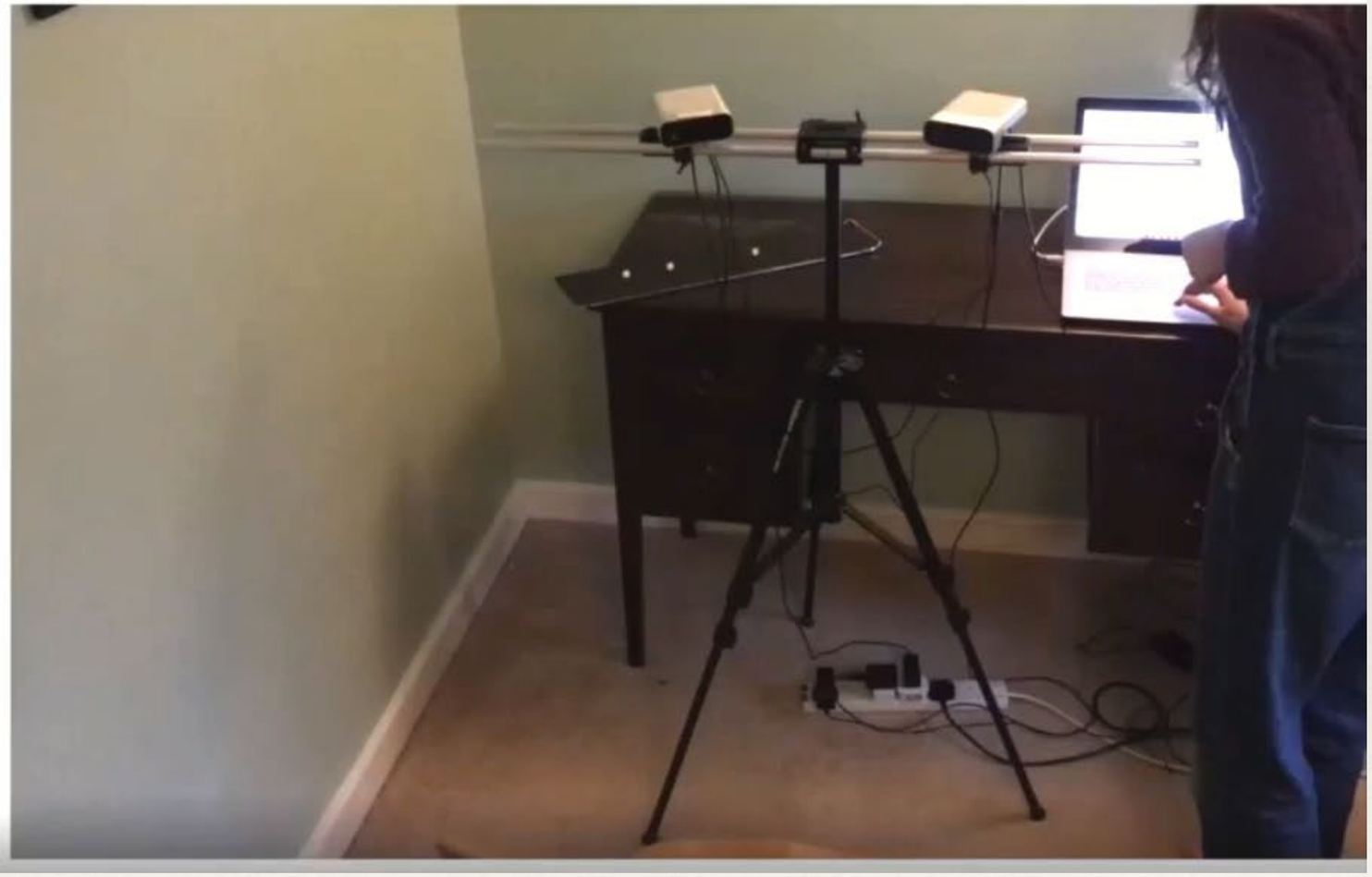
Intrinsic calibration



Extrinsic calibration

Camera Calibration Demo

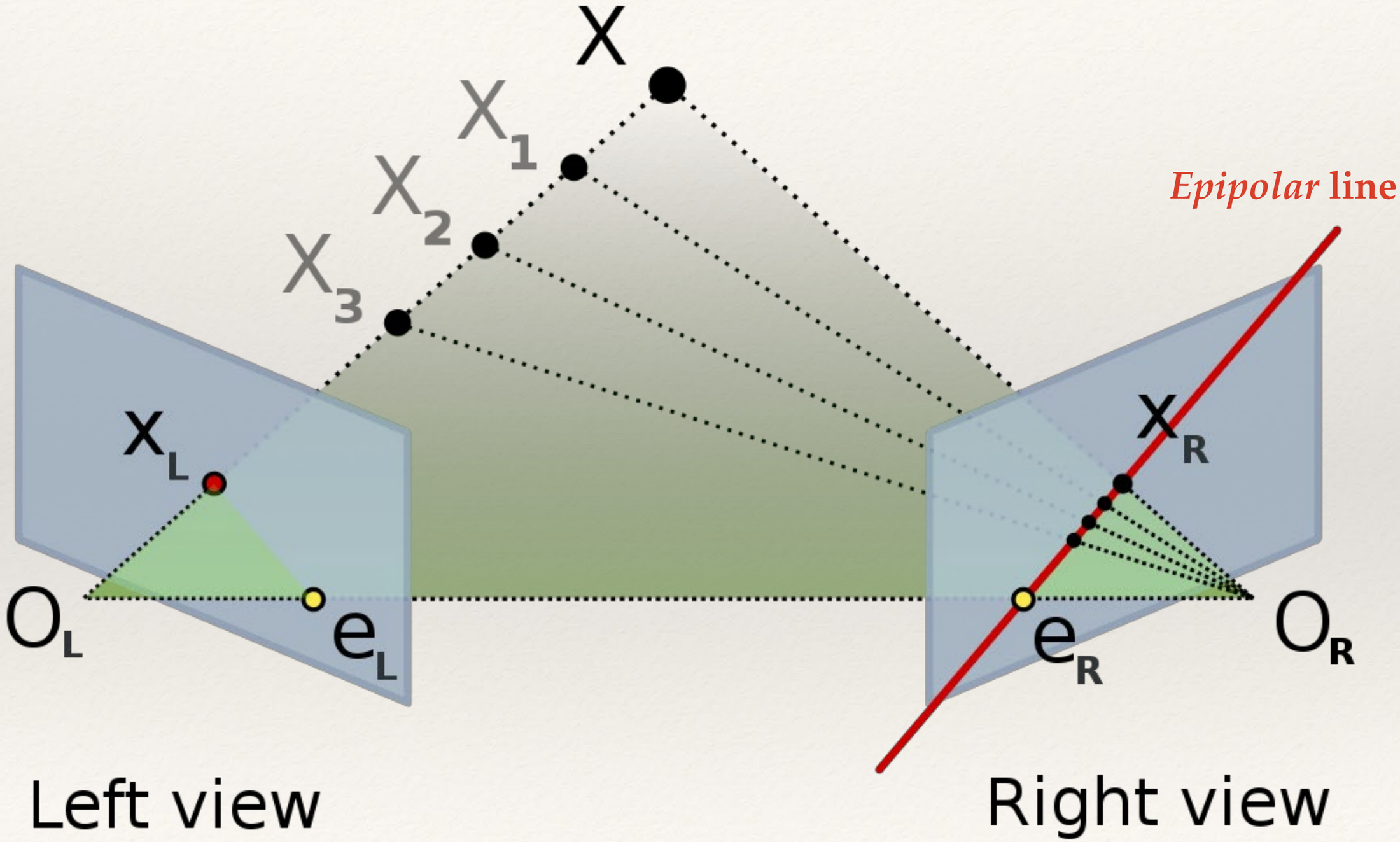




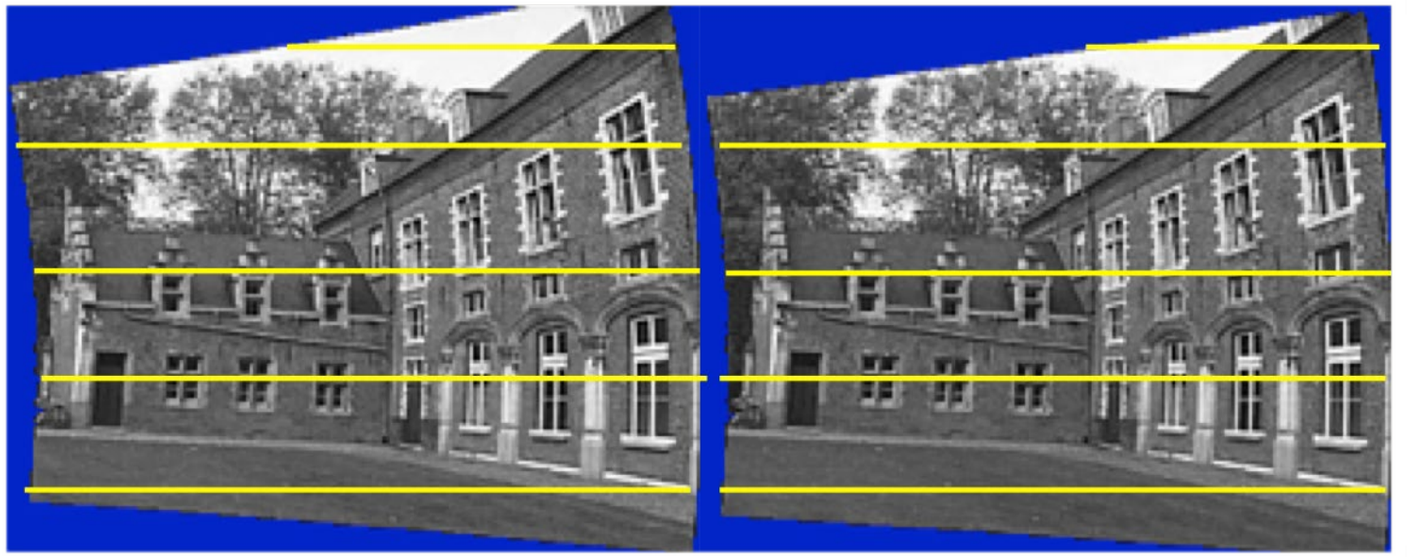
Measuring Depth

Narrow Baseline Stereo

Epipolar geometry



Rectification and matching



Warp images to simplify epipolar geometry

Compute disparity map by matching pixels along the epipolar lines

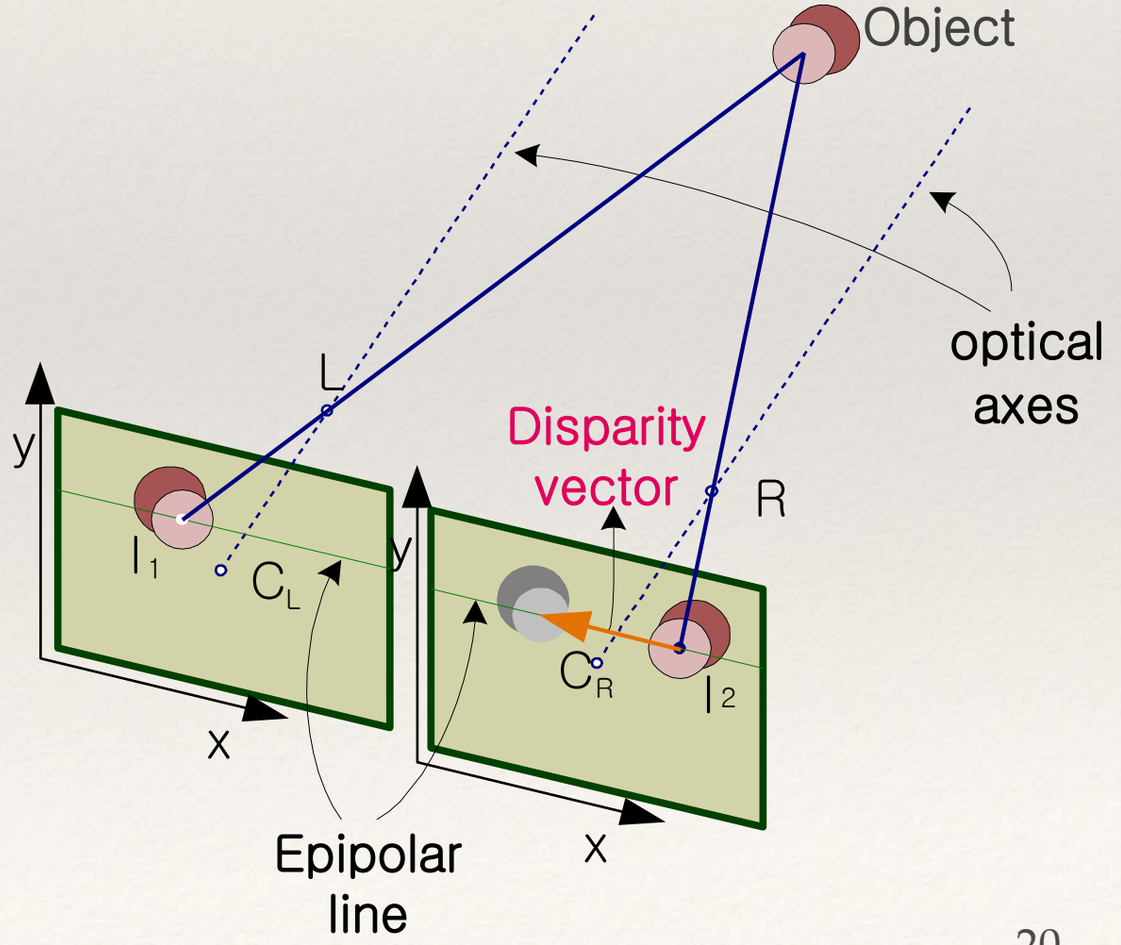
Stereo Camera

- Simple case of a parallel stereo camera system

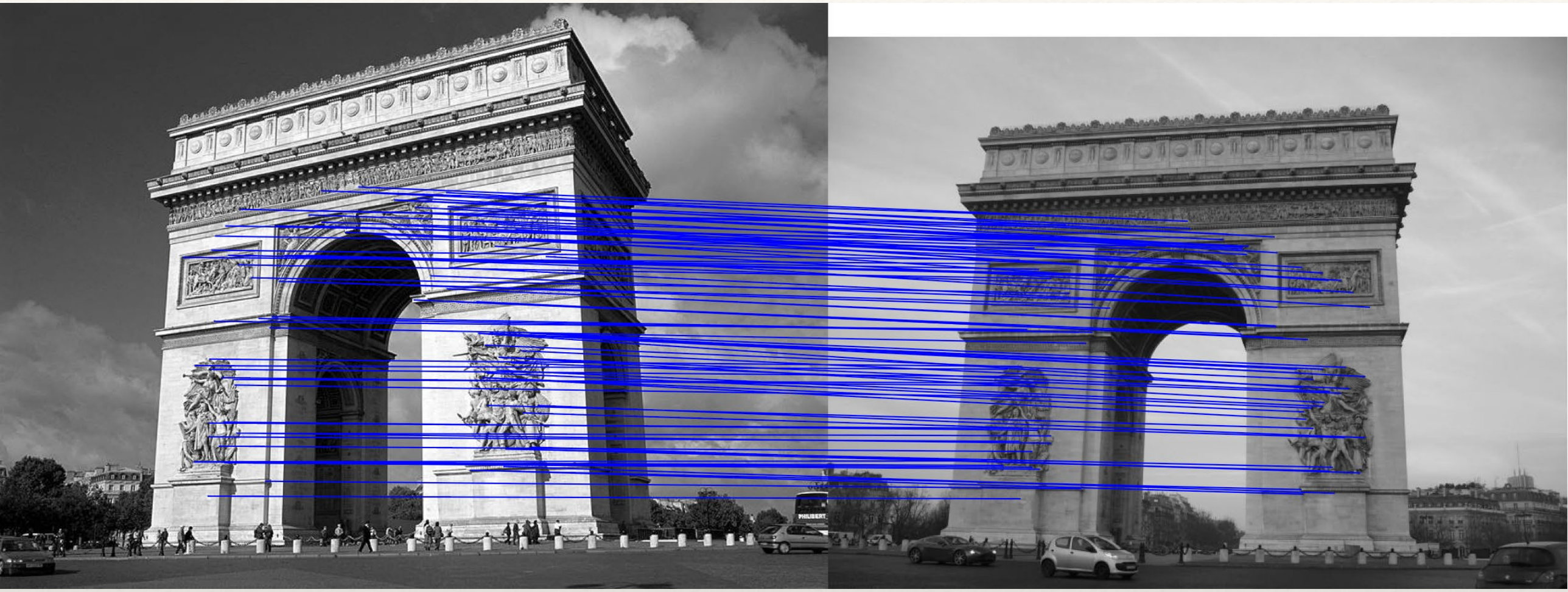
$$Z = \lambda - \frac{\lambda B}{x_{I_1} - x_{I_2}}$$

- Disparity estimation
  - To find the corresponding pair I_1 and I_2 in the stereo image pair

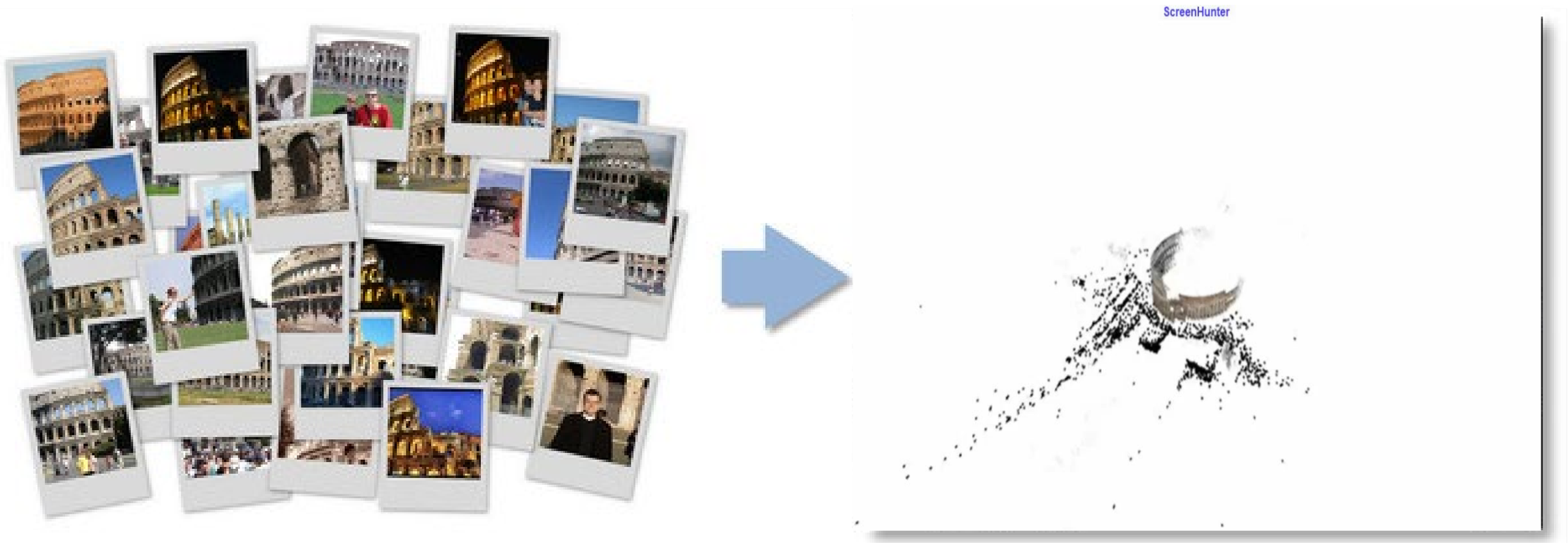
$$I_1(x, y) = I_2(x + d(x, y), y)$$



Wide Baseline Stereo



Point matches (i.e. SIFT) are used as the basis for triangulating 3D points from the 2D images



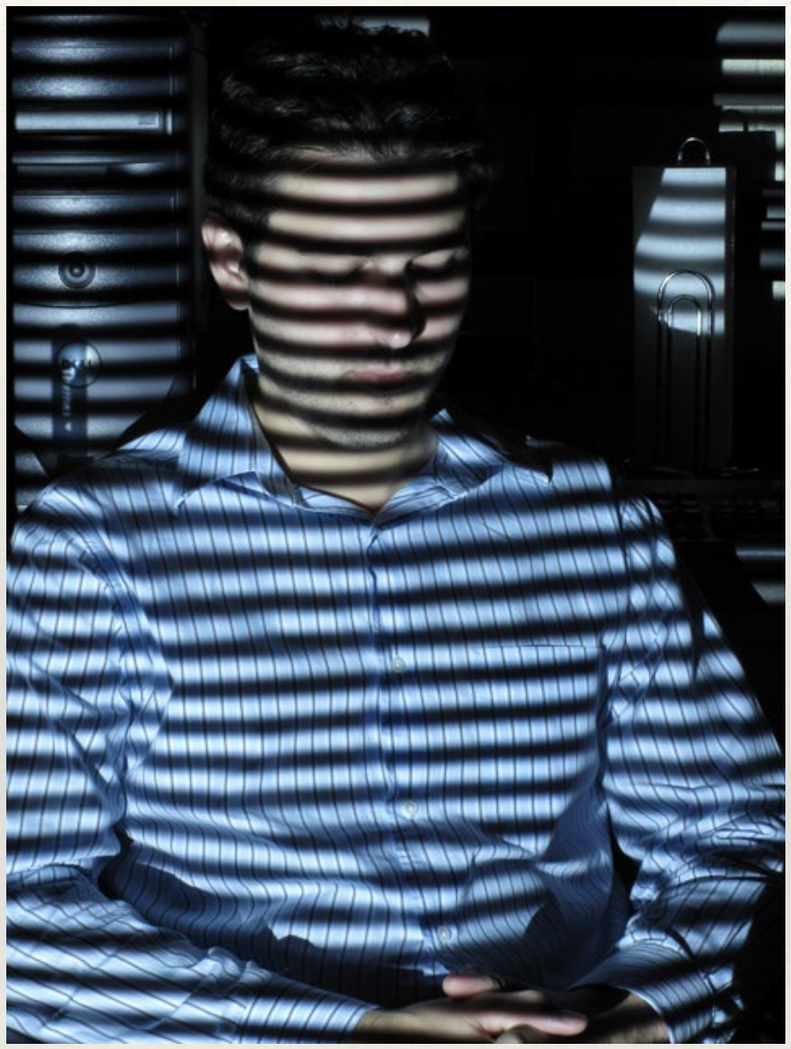
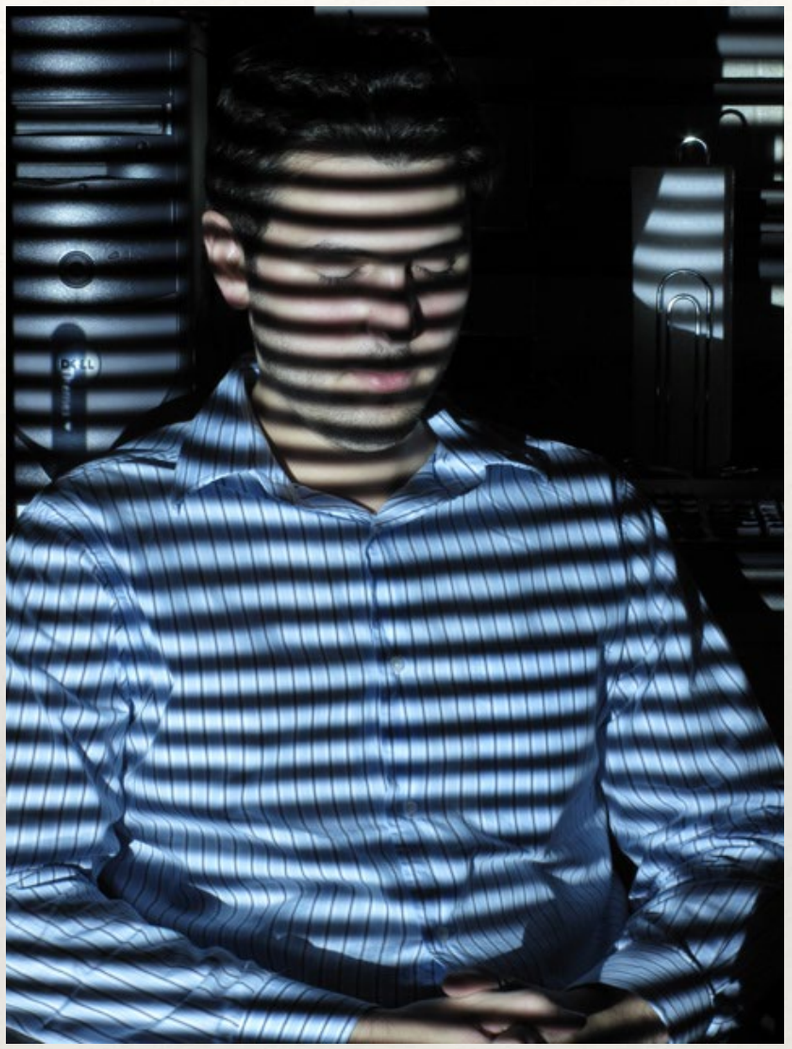
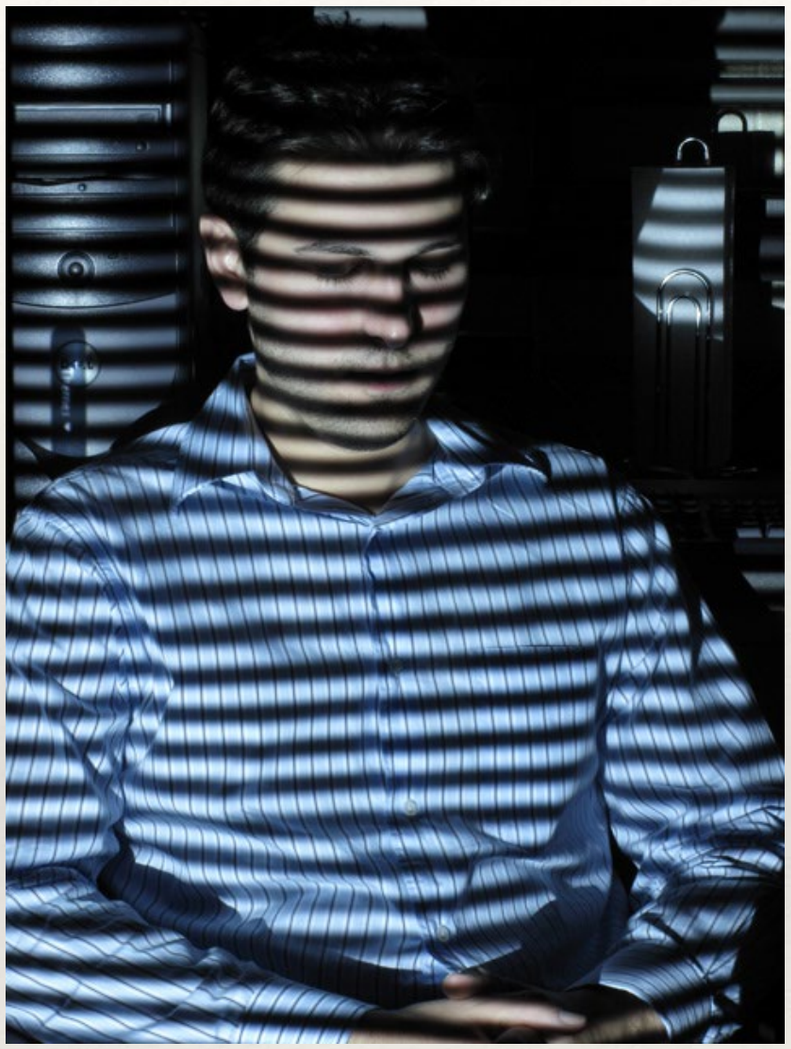
Building Rome in a Day
http://grail.cs.washington.edu/projects/rome/

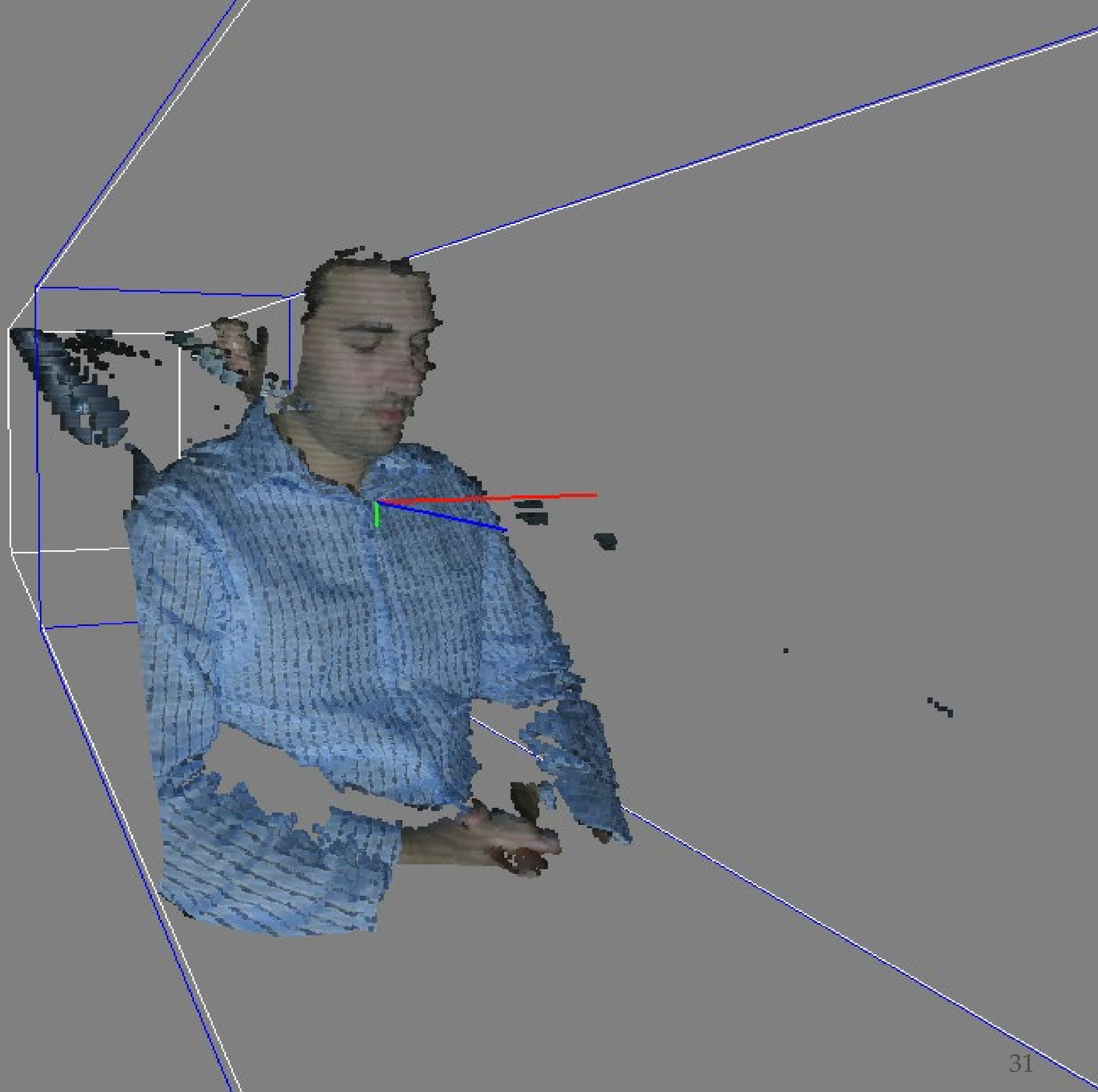
Multiple images can be used to jointly infer 3D structure, and the camera pose and intrinsics of each camera

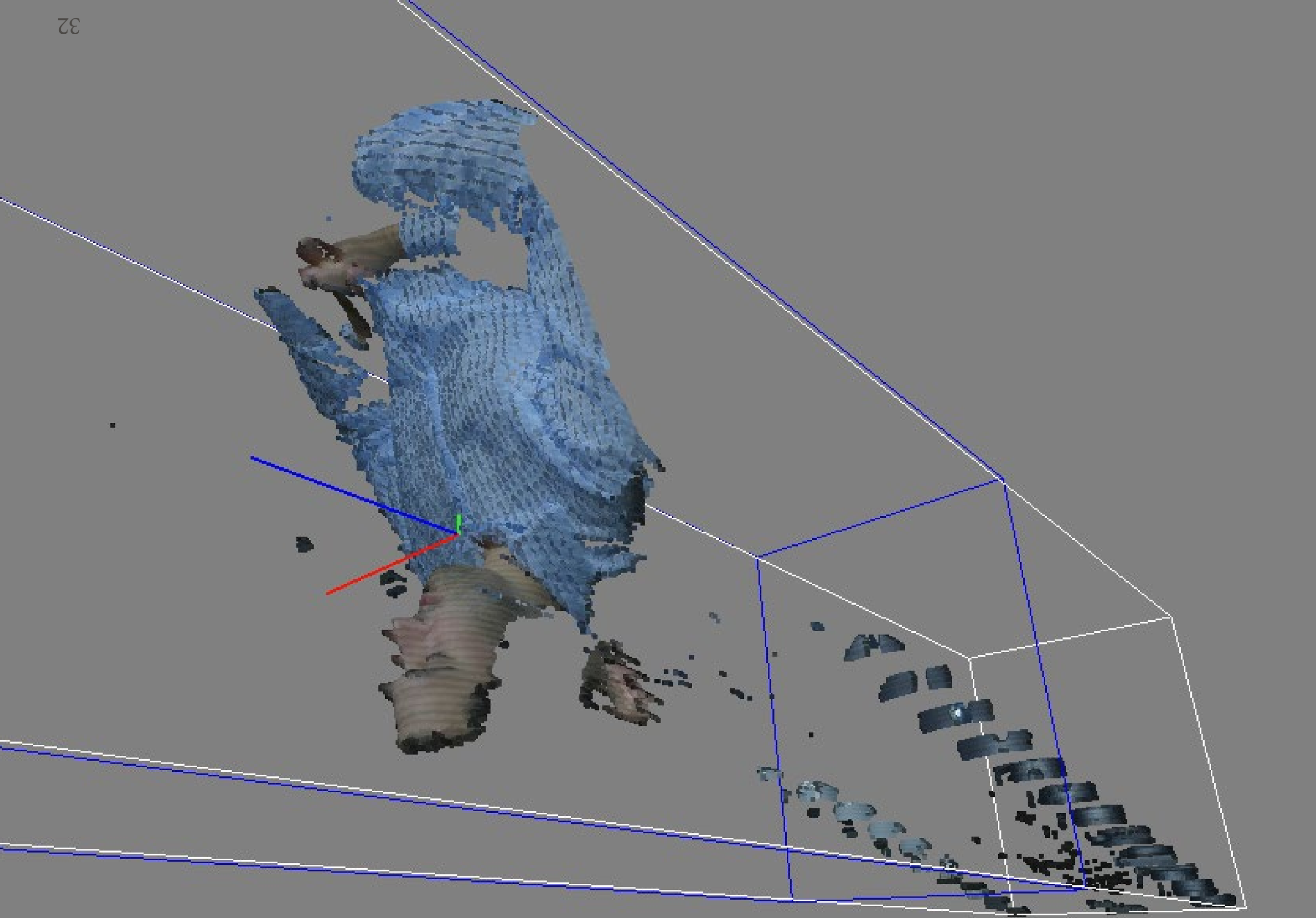
Reconstructing Venice

Monocular Vision

Structured Light Imaging





Non-visible techniques

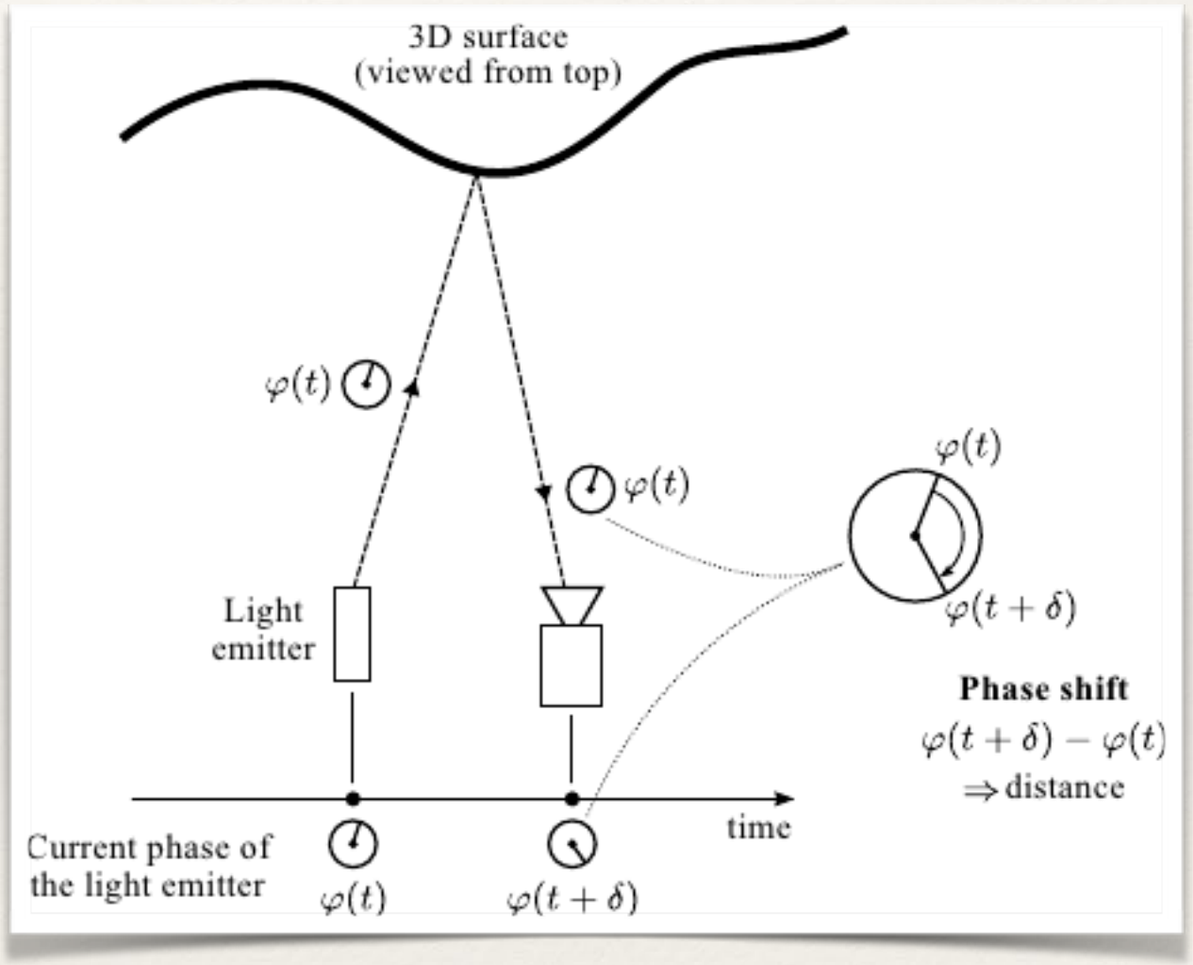
LiDAR

PrimeSense (Kinect 1)

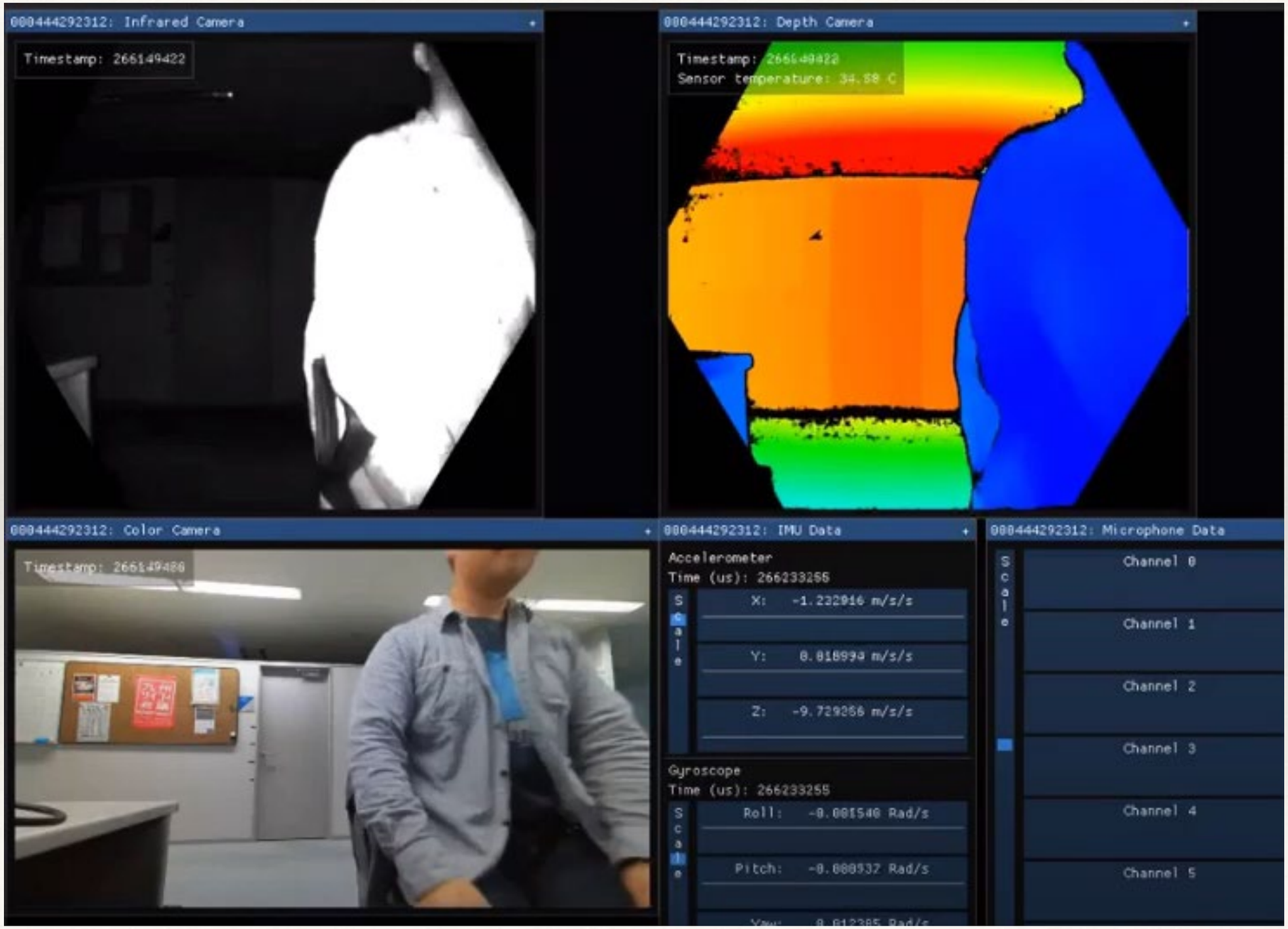
- Uses *coded* structured IR light
  - IR emitter projects a stationary, random pattern of dots
    - Basically shining light through a opaque stencil with holes in it
  - IR camera records those dots
  - Template matching is used to compare the actual location of the dots from the IR sensor to the known location if the dots were projected on a plane perpendicular to the optical axis at a known distance

PrimeSense (Kinect 1)

Time of Flight Imaging (Azure Kinect)



Time of Flight Imaging (Azure Kinect)



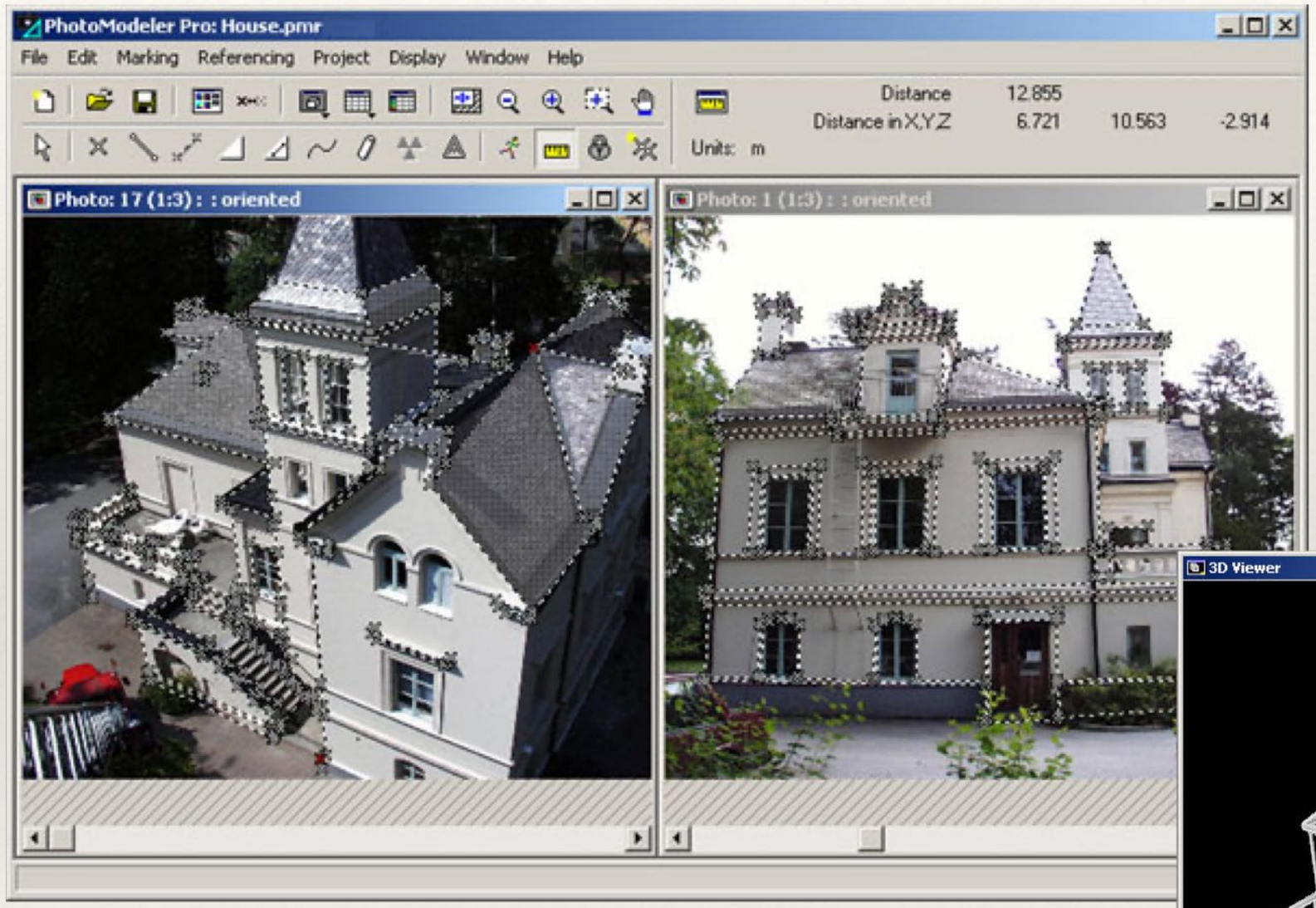
Real-time Multiview reconstruction with 4 Azure Kinect cameras

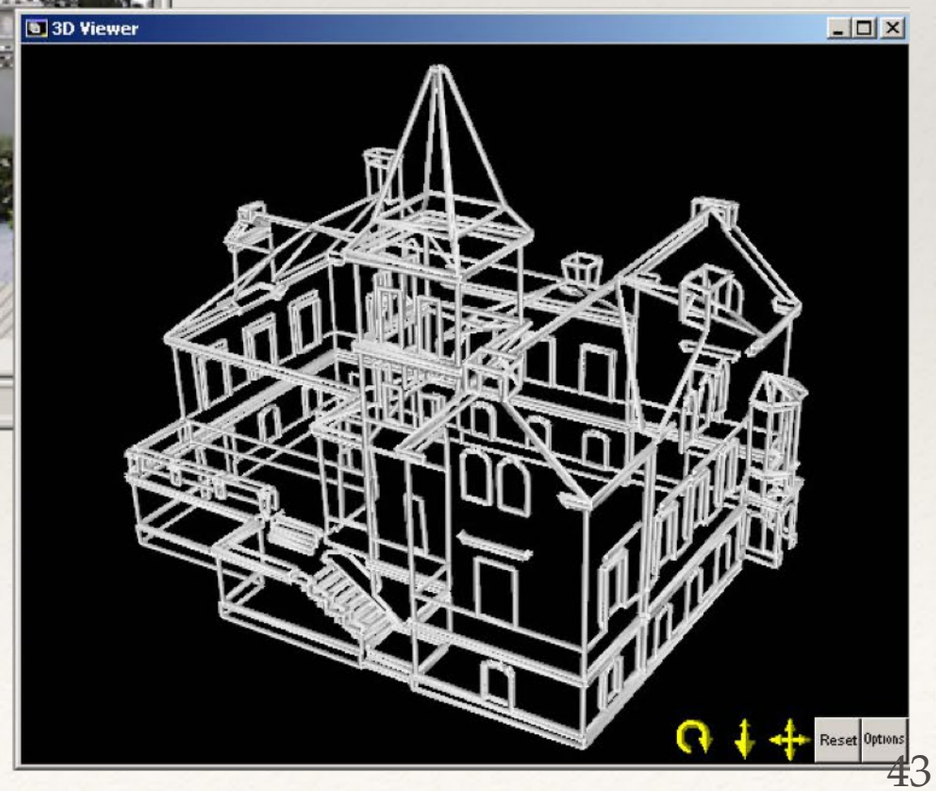
Applications

Self-driving cars

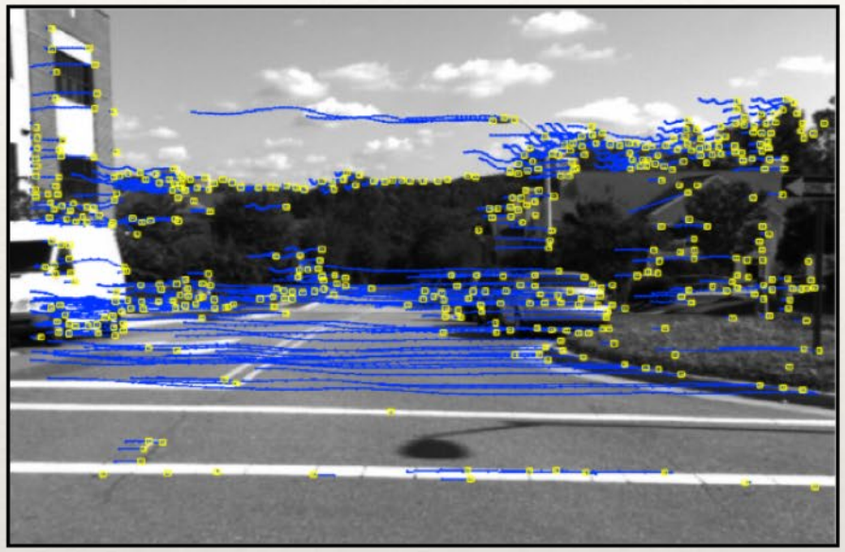
Image from https://www.autonomousvehicleinternational.com/news/adas/enthusiasm-for-self-driving-cars-to-double.html

Architecture

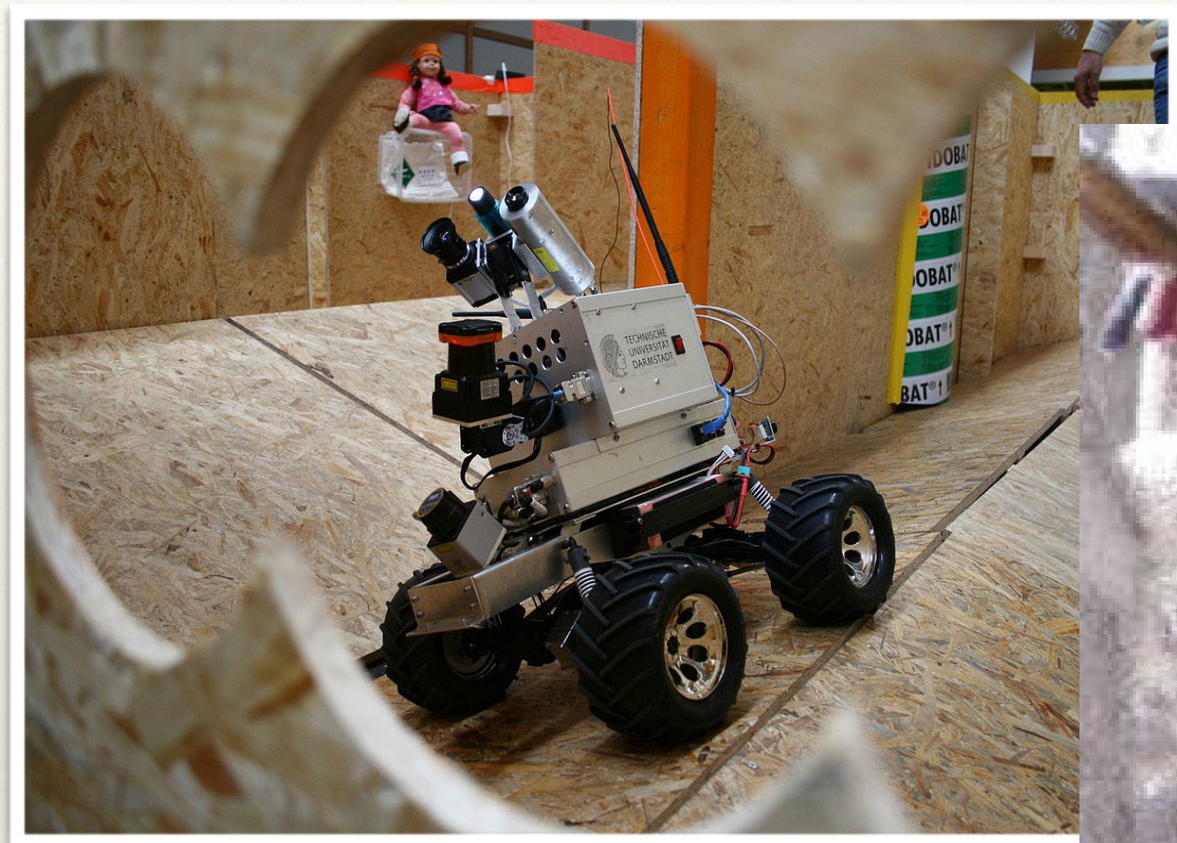


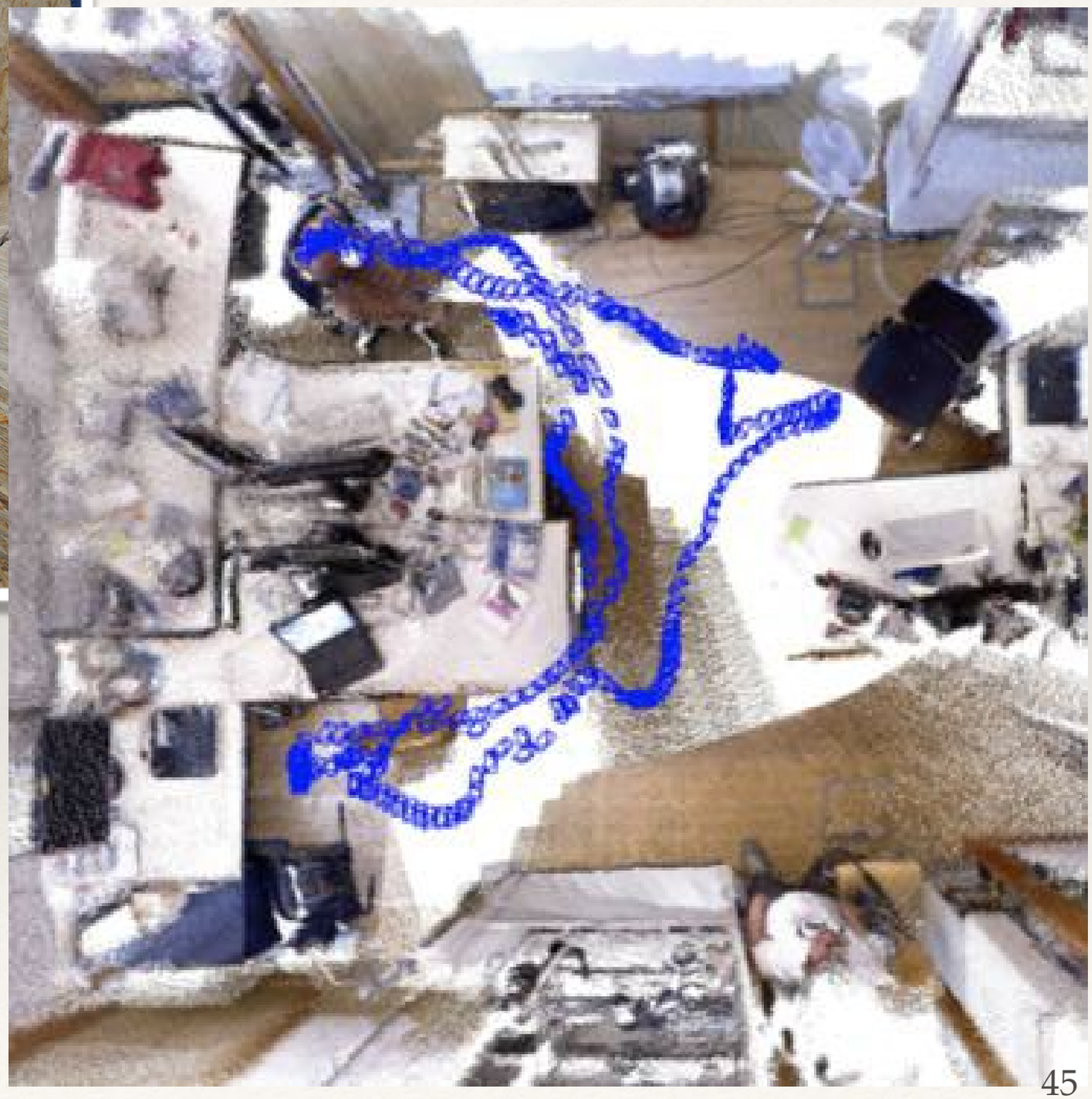


Urban Planning

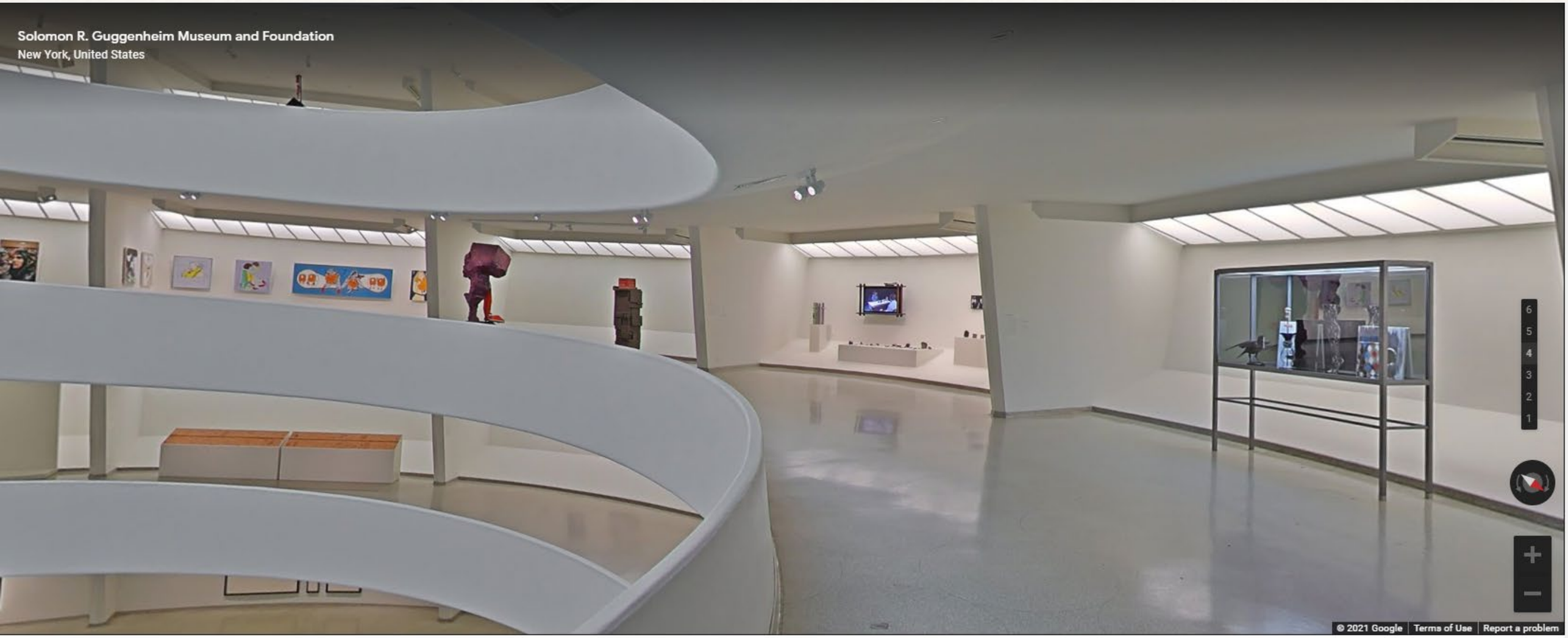


SLAM (Simultaneous localization and mapping)





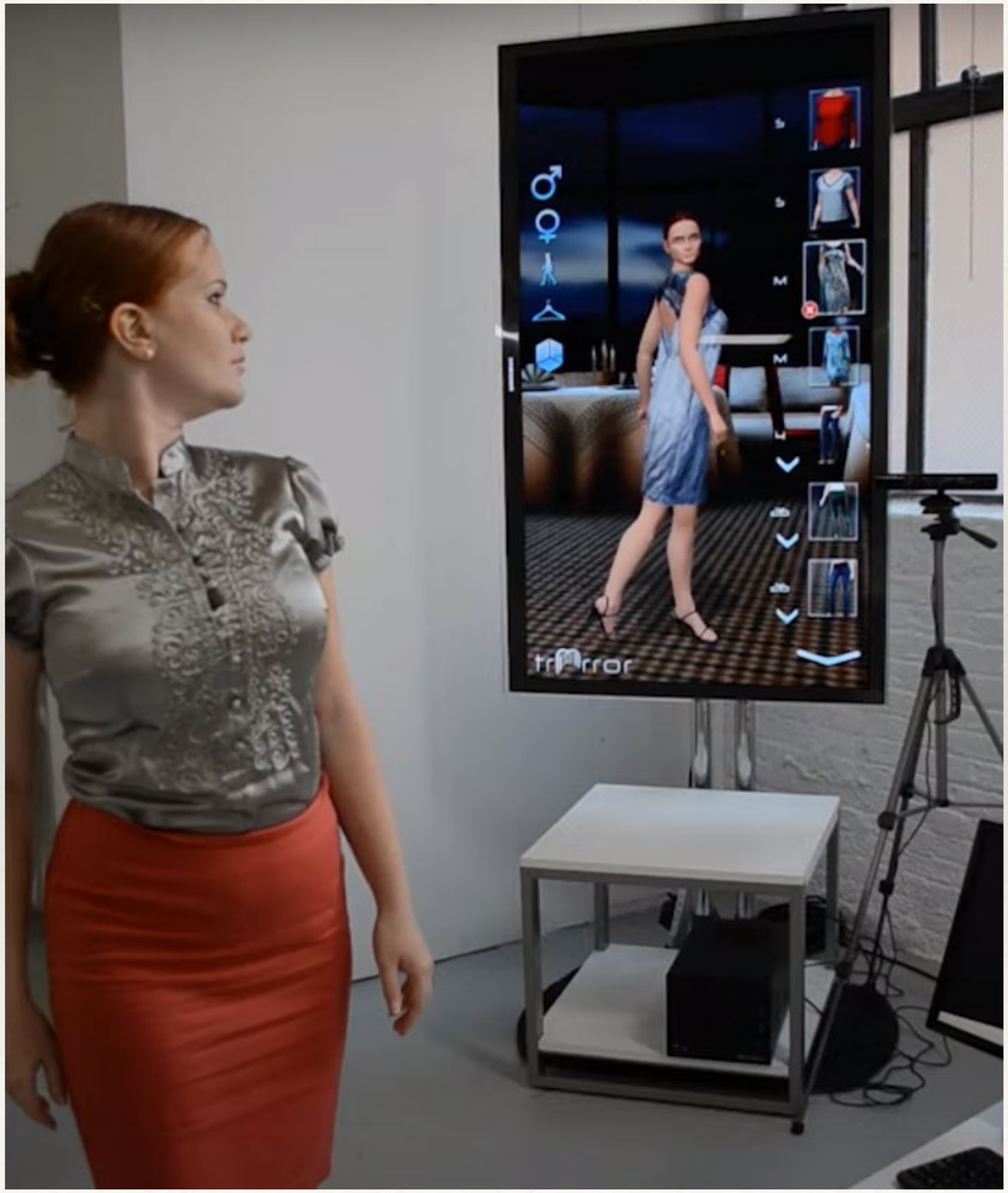
Virtual Tourism



https://artsandculture.google.com/?hl=en

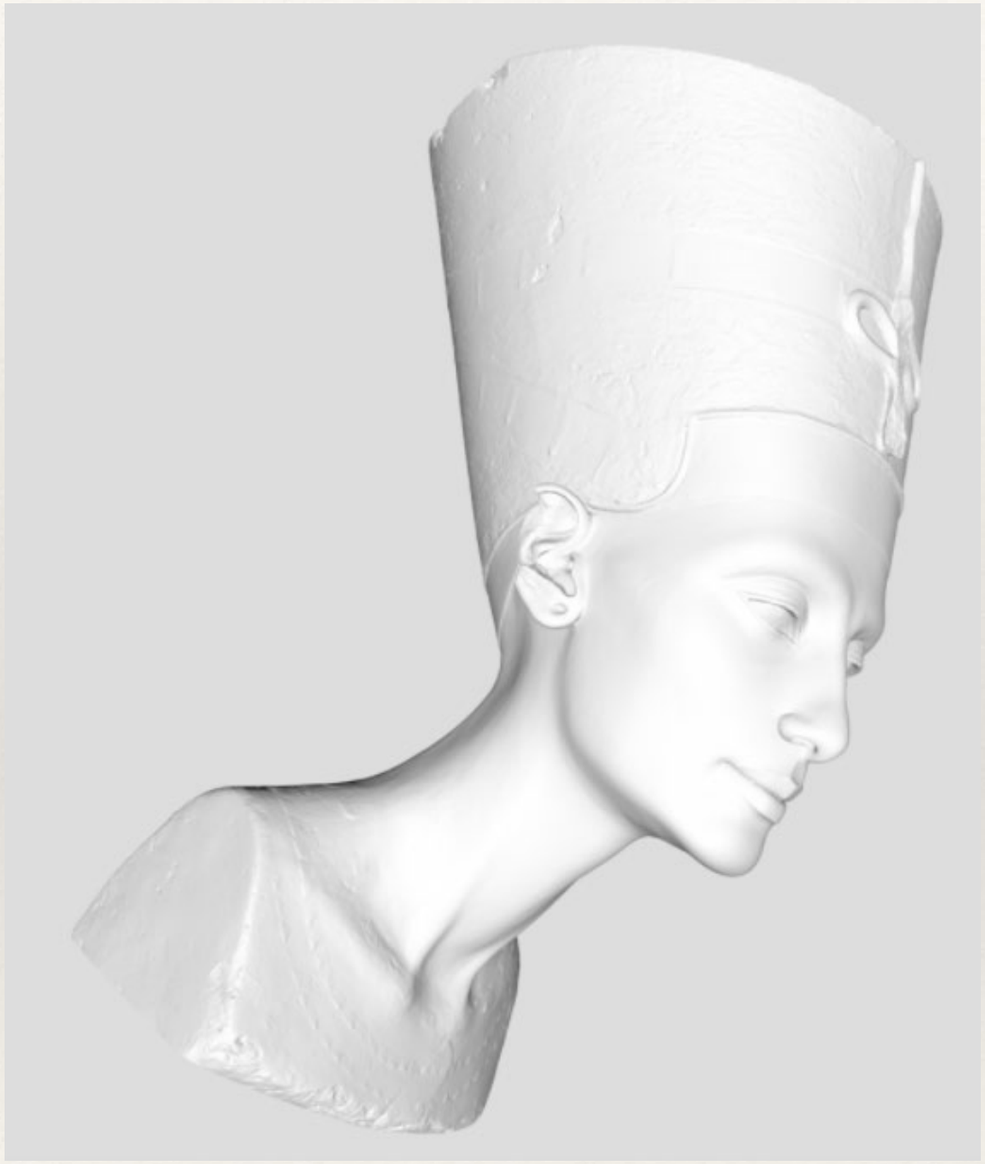
Body measurement and fitting

https://youtu.be/gNH9ba9aWd0



www.trimirror.com

Art & history



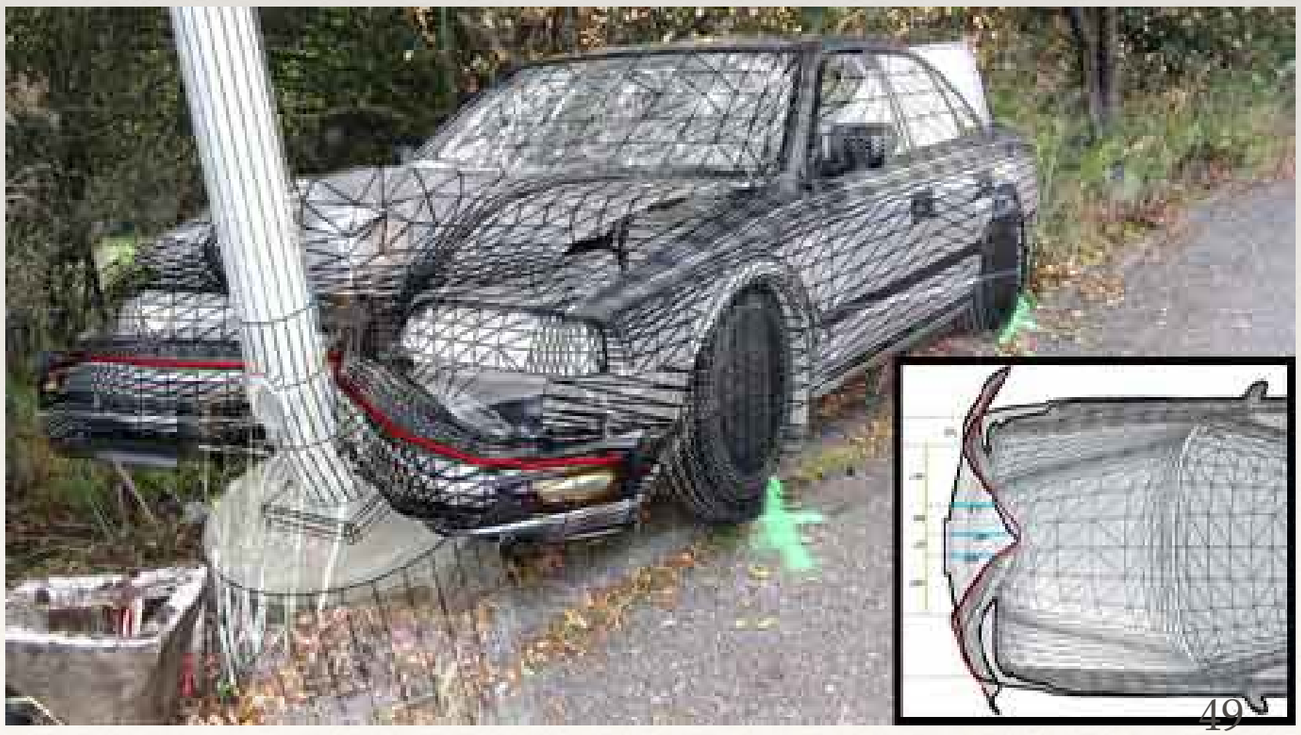
https://www.myminifactory.com/object/3d-print-bust-of-nefertiti-at-the-egyptian-museum-berlin-2951

Cyberarchaeology by Univ. of Tokyo

Forensics


FARO
1
2
3


Droits réservés sur toutes les images.
Philippe Froesch
(Bidabat, Philippe Froesch VISUALFORENSIC)
TERRA

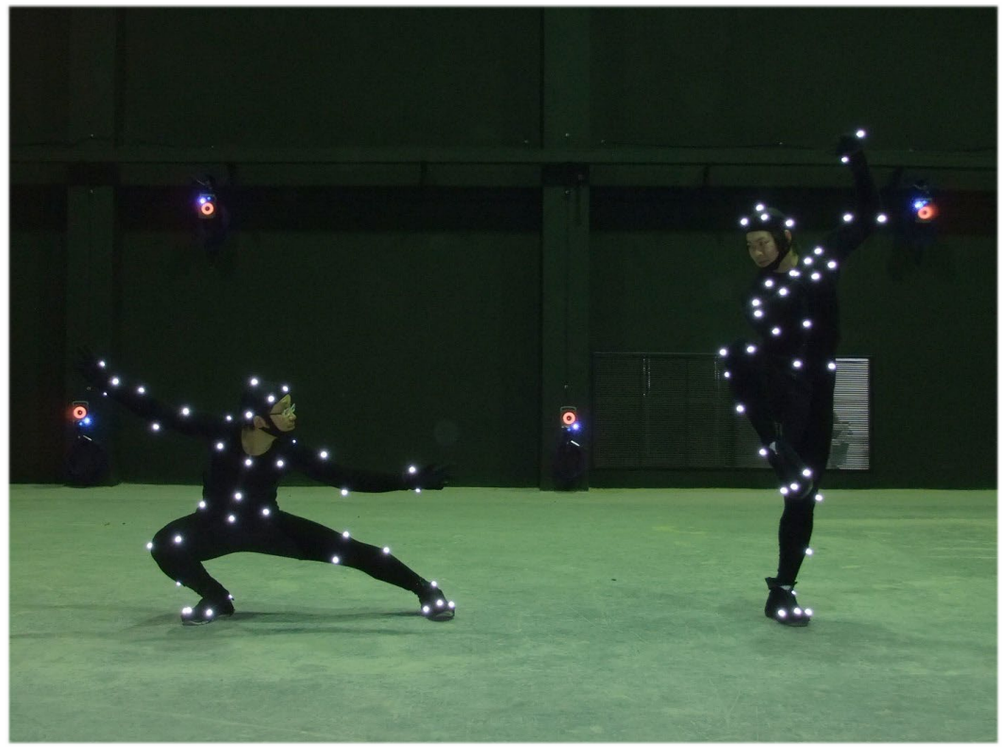


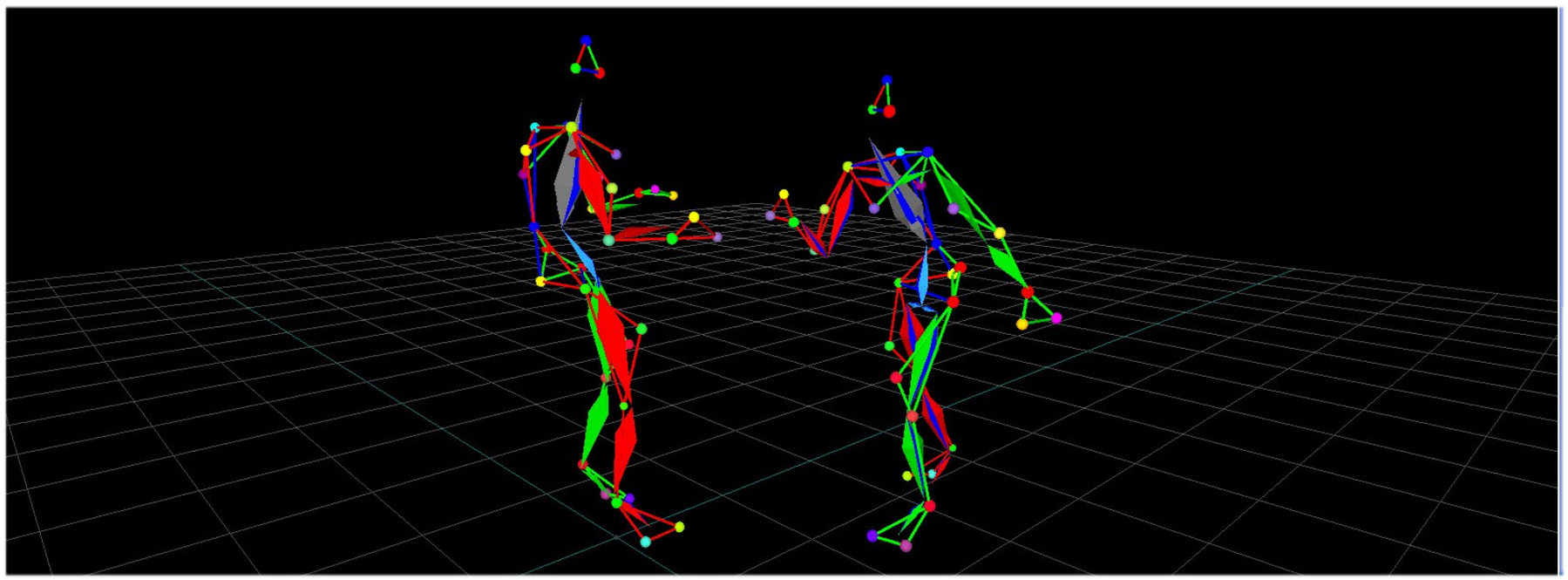
Surveillance

GAIT

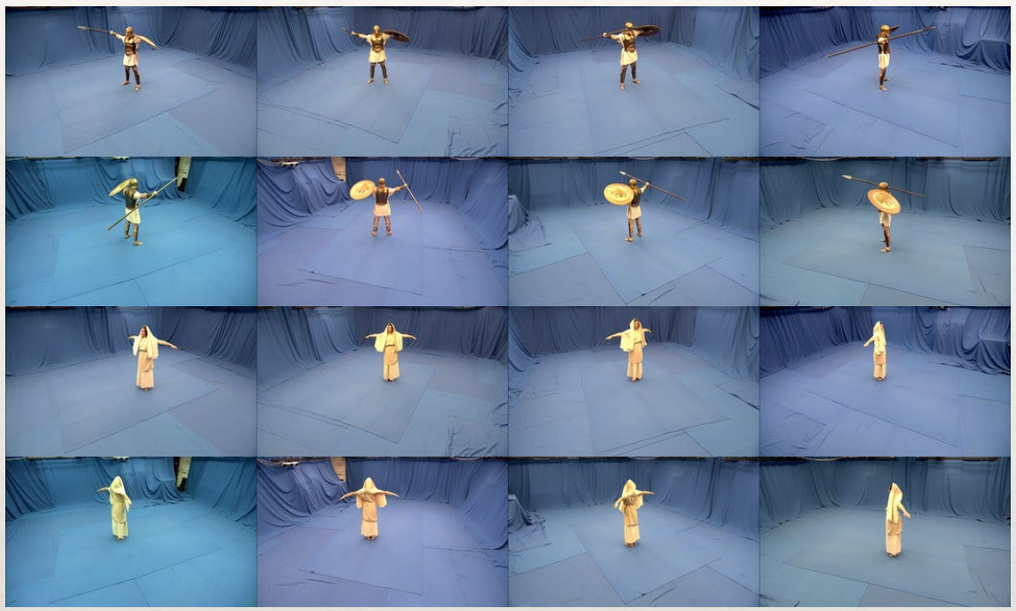
3D Openpose

Motion Capture (Films & Games)





Film production



Summary

- 3D computer vision has lots of practical applications
- Camera models give a mathematical description of how a pixel in a 2D image is related to a point in a 3D scene
  - Camera calibration can be used to find the parameters of a camera
- Multiple views of a scene can be used to infer depth
- There are lots of other techniques for capturing depth that only require a single sensor

* *Acknowledgements: B*ased on earlier Computer Vision lecture slides by Dr. Jon Hare.