Stereo Camera System

Lecturer: Sang Hwa Lee

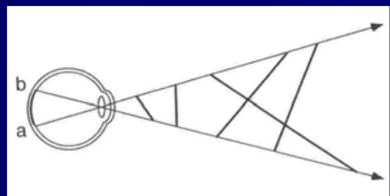
Recovering 3D from Images (1)

- We have relied on a human perception to provide depth cues
  - parallel lines, reference points, etc.

- How do we do this automatically?
  - What cues in the image provide 3D information?

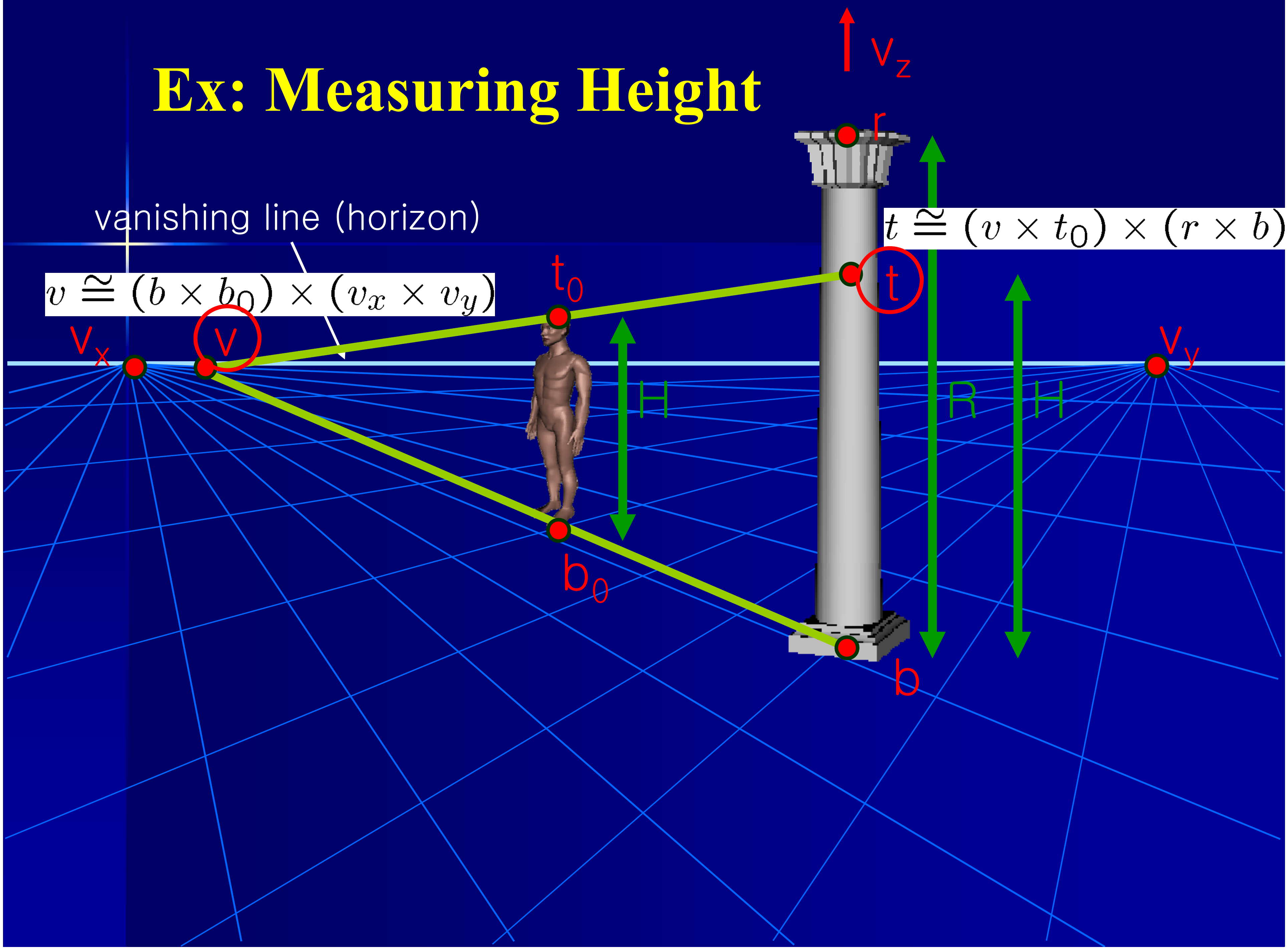
Recovering 3D from Images (2)

Projections



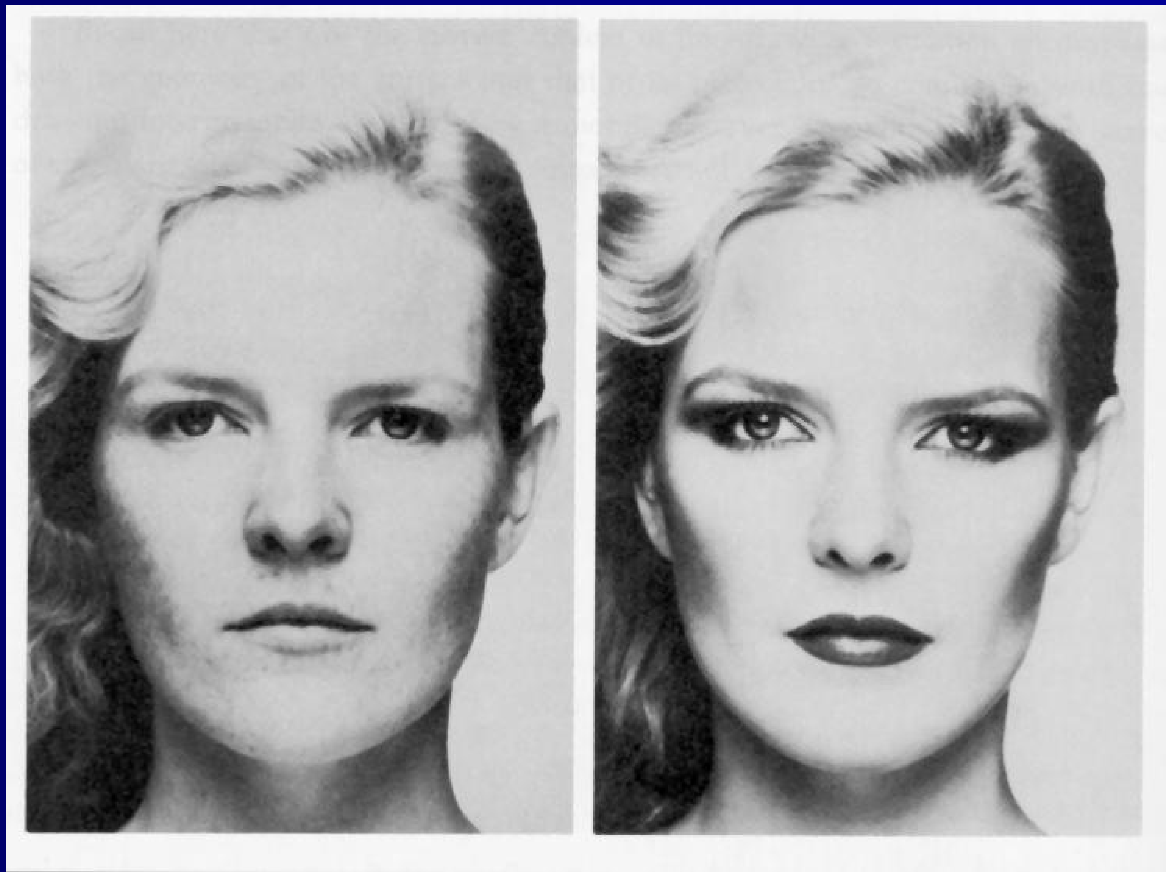
Reconstruction

Ex: Measuring Height



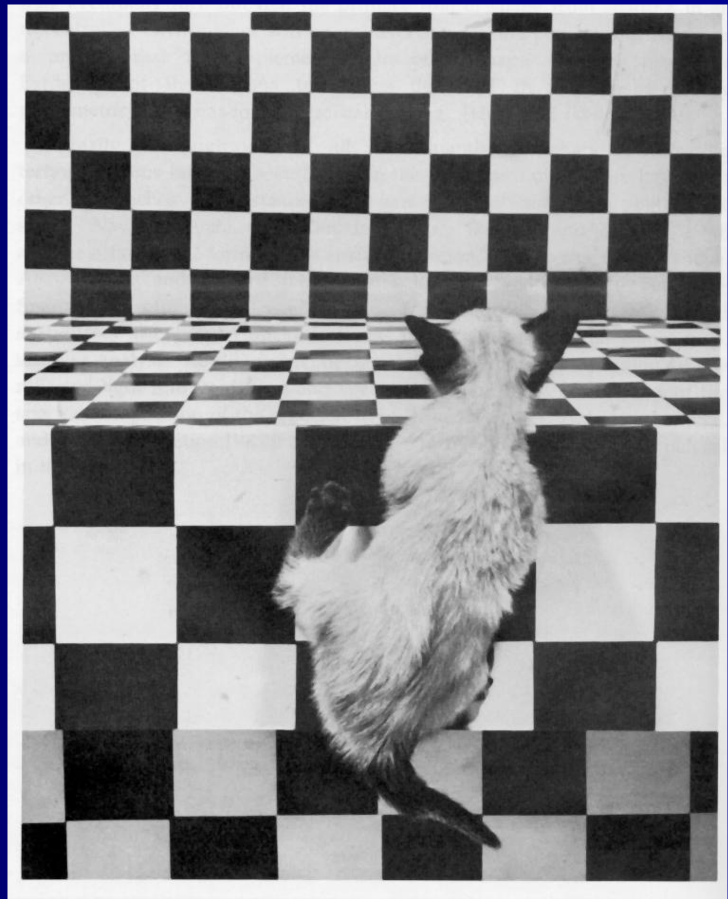
Visual cues for 3D models (1)

- Shading



Visual cues for 3D models (2)

- Shading
- Texture



Visual cues for 3D models (3)

- ➔ Shading
- ➔ Texture
- ➔ Focus

Visual cues for 3D models (4)

- ➔ Shading
- ➔ Texture
- ➔ Focus

- ➔ Motion (moving objects vs moving camera)

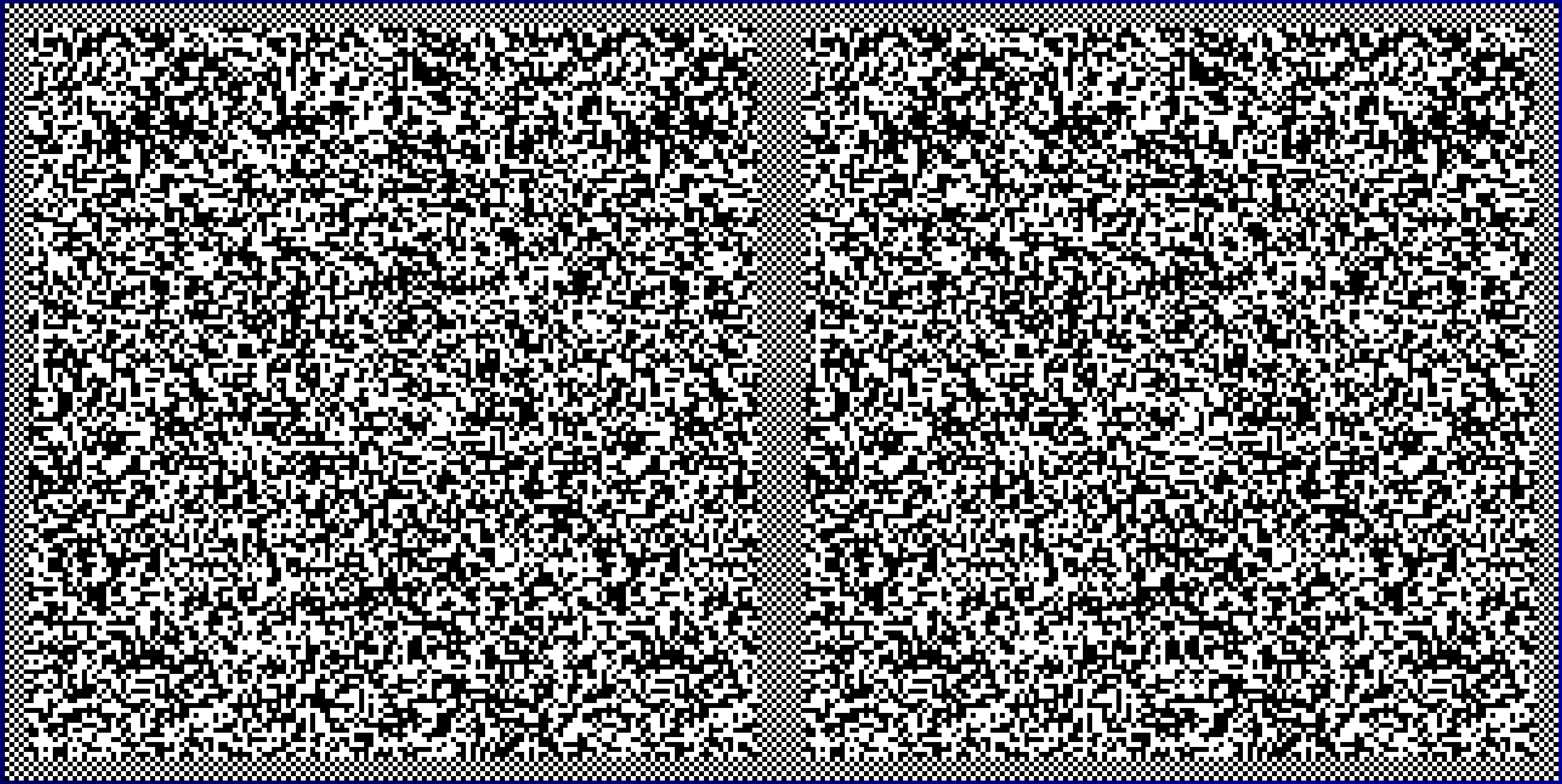
Visual cues for 3D models (5)

Shape From X

- X = shading, texture, focus, motion, ...
- **In this class we'll focus on the motion cue**
- **Stereo system**

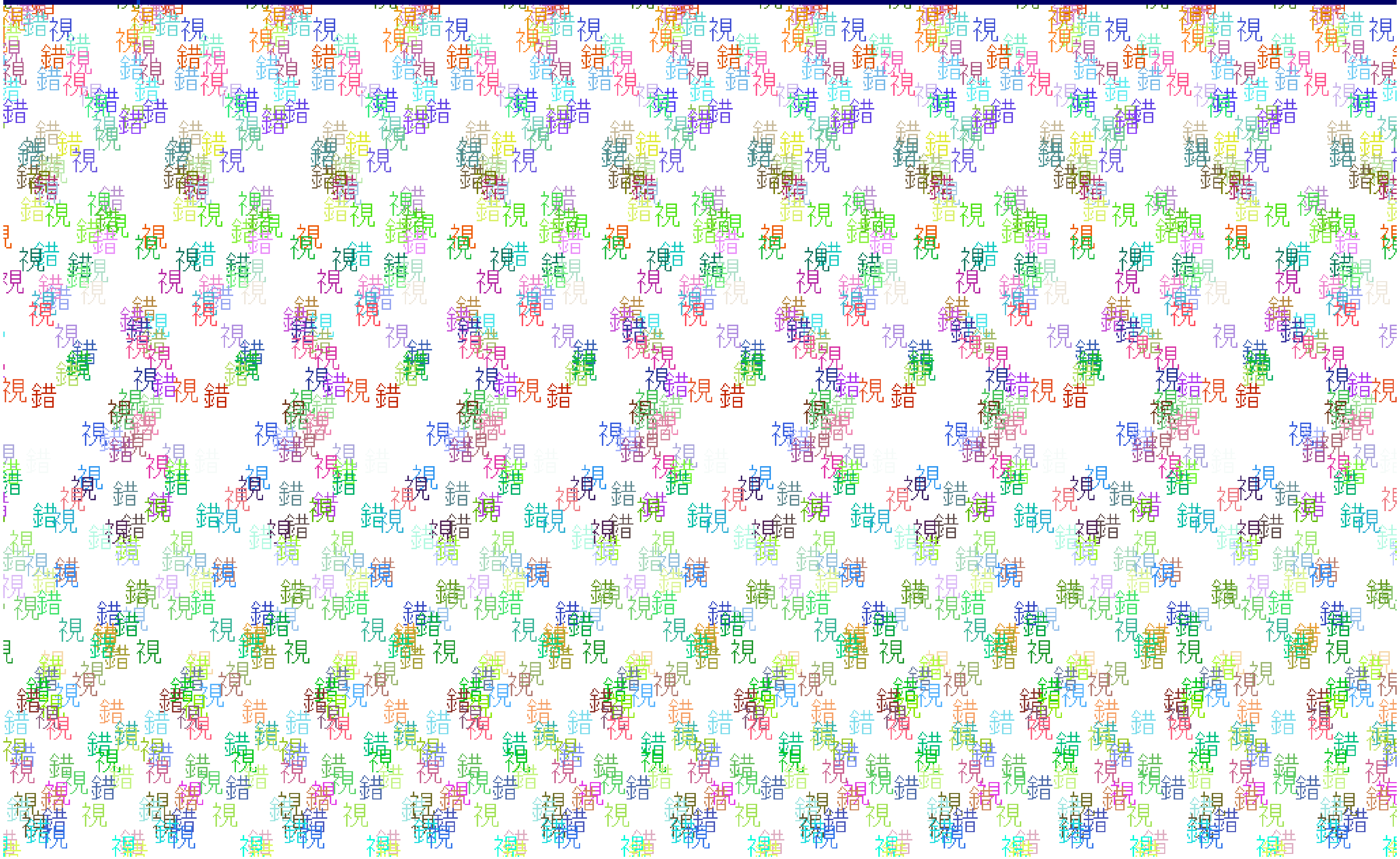
Binocular Stereogram (1)

Bela Julesz (1971)



Binocular Stereogram (2)

Binocular Stereogram (3)

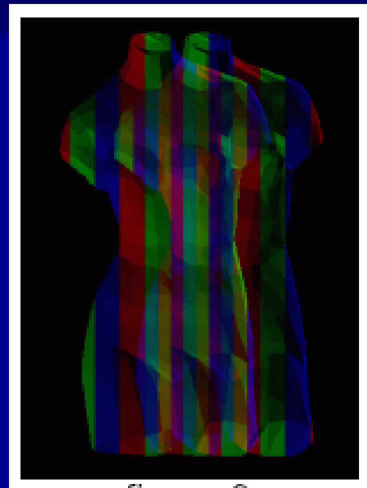


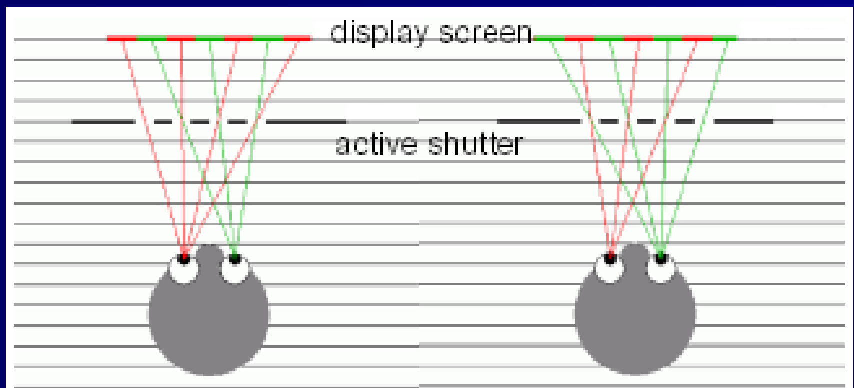
Stereoscopic Display

- Bi-lenticular lens
- Polarization
  - Need of glasses
  - RGB
  - Vertical/Horizontal wave
- Critical to focusing on human eyes
  - Head/eye tracking

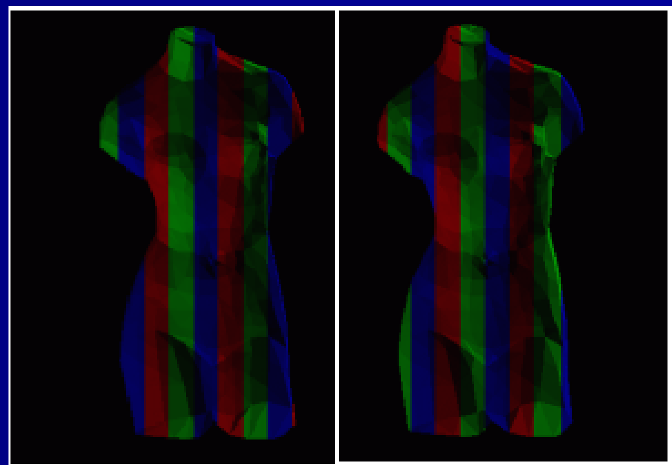
Stereoscopic Display

- Auto-stereoscopic view
  - No need of glasses
  - Frame by frame (temporal)
  - Field by field (interlace)



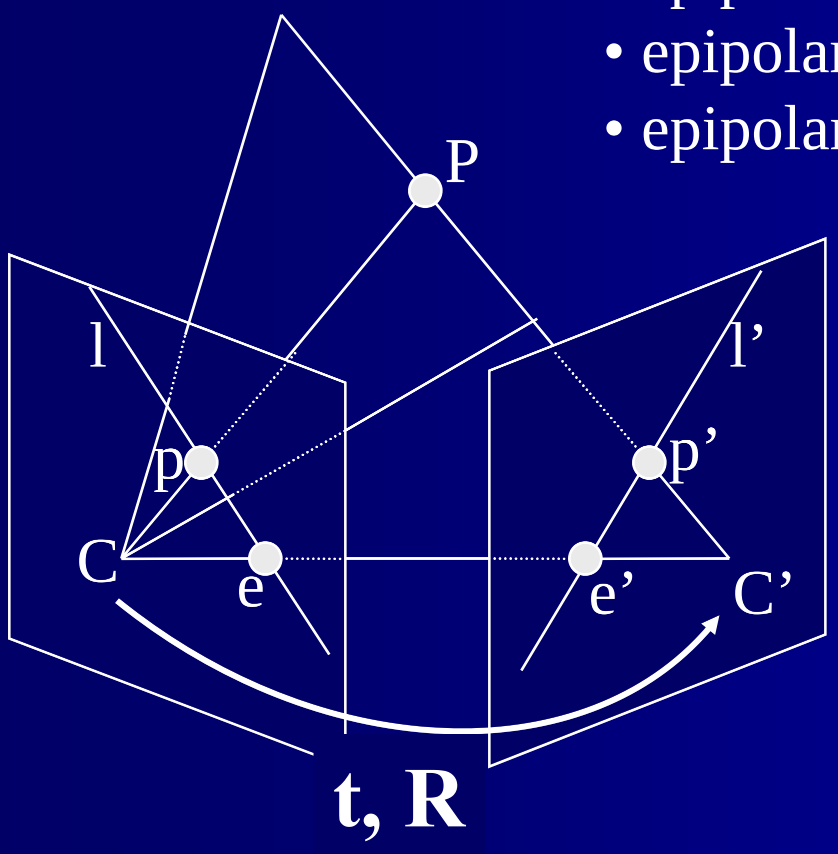


Alternation by active shutter

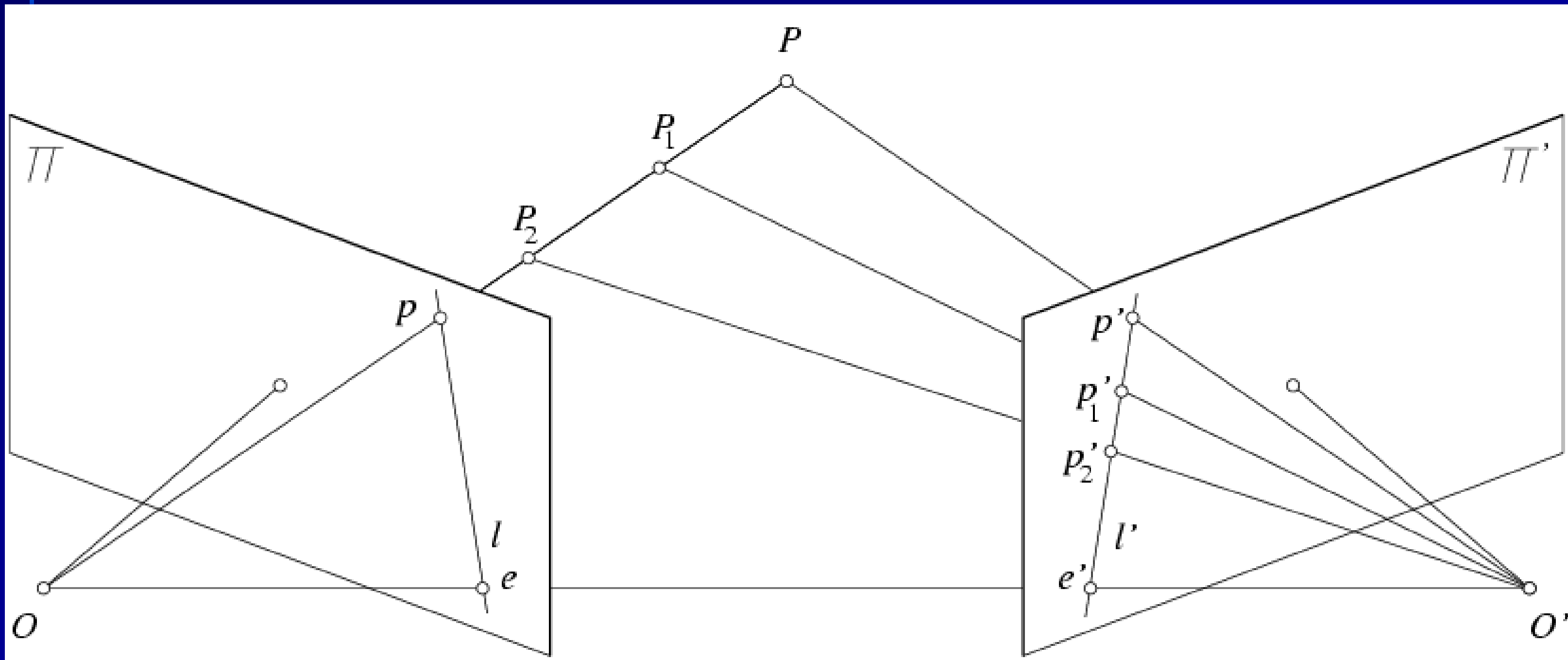


Epipolar Geometry in 2-view (1)

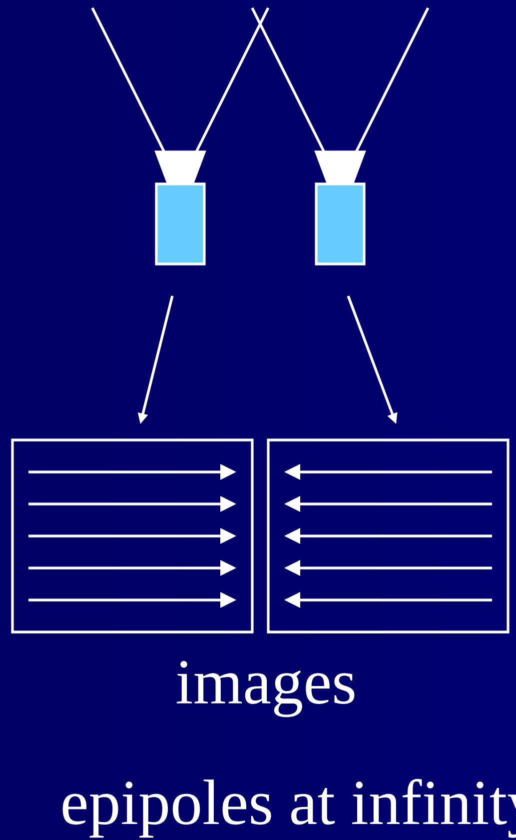
- epipole: e, e’
- epipolar line: l, l’
- epipolar plane: including ΔPCC’

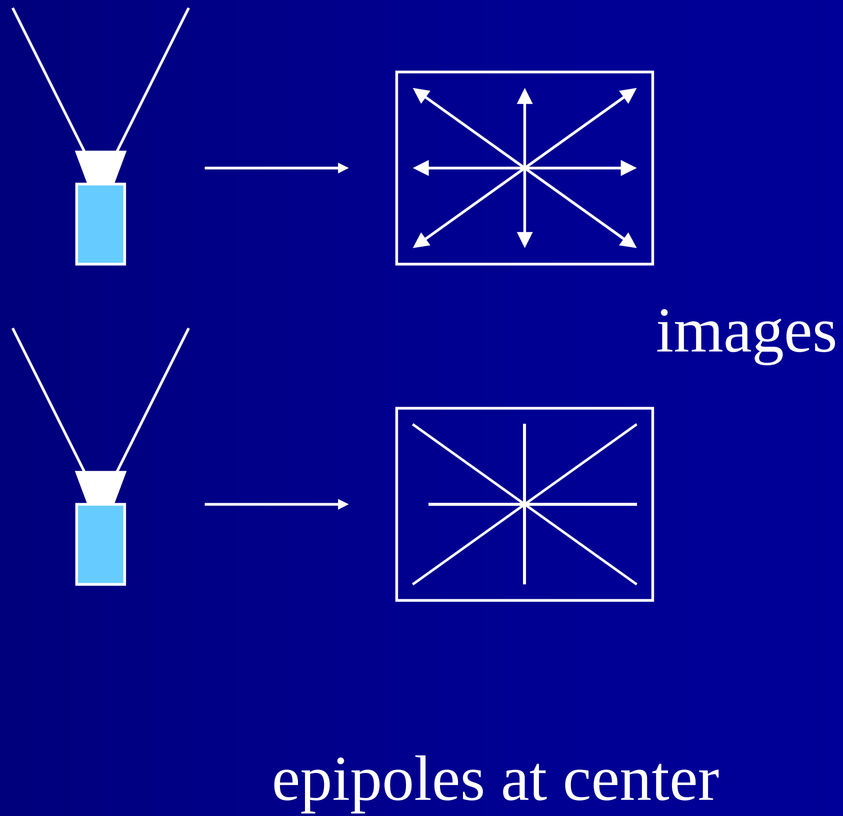


Epipolar Geometry in 2-view (2)



Epipolar Geometry in 2-view (3)





Epipolar Equation (1)

- Normalized image coordinate (pinhole)
  - Epipolar equation

    $\mathbf{m}^T\mathbf{E}\ \mathbf{m}'=0$

    where $\boldsymbol{E} = [\mathbf{t}]\ \mathbf{R}$

    Essential matrix $\boldsymbol{E}$

  - Epipolar line

    $\mathbf{l} = \mathbf{E}\mathbf{m}'$

    $\mathbf{l}' = \mathbf{E}^T\mathbf{m}$

Epipolar Equation (2)

- Pixel image coordinate
  - Epipolar equation

    $\mathbf{m}^T\mathbf{F}\,\mathbf{m}'=0$

    where $\mathbf{F} = \mathbf{A}^{-T}\,\mathbf{E}\,\mathbf{A}'^{-1}$ (**A** is camera intrinsic matrix)

    Fundamental matrix **F**
  - Epipolar line

    $\mathbf{l} = \mathbf{F}\mathbf{m}'$

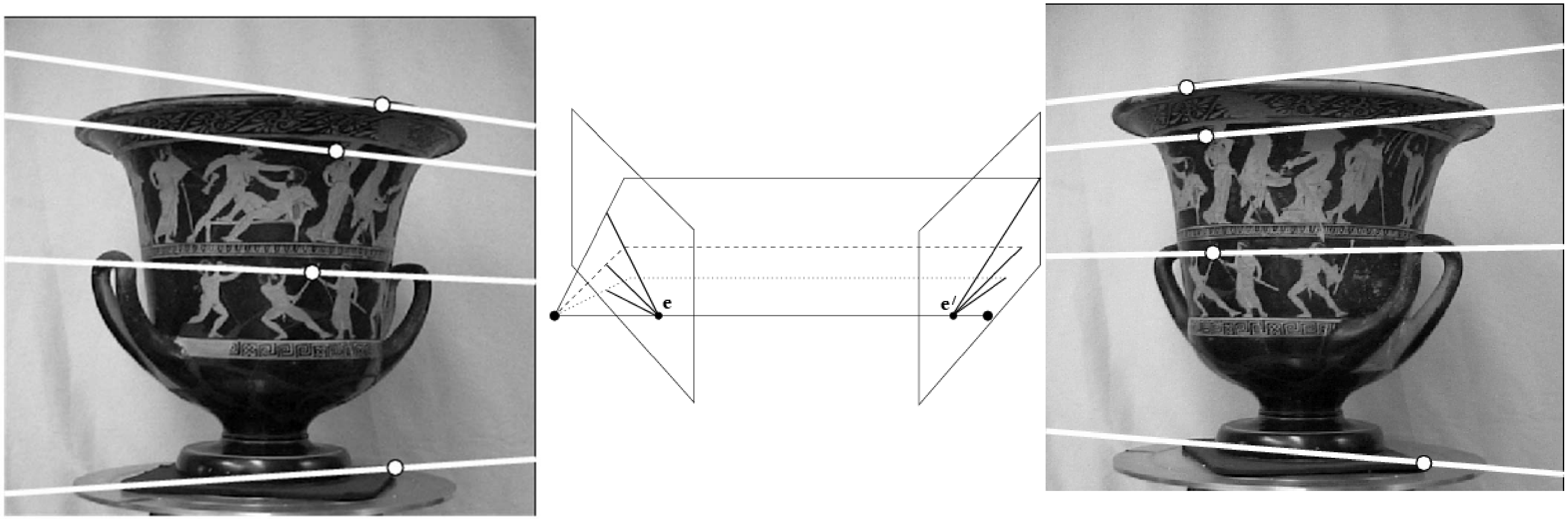
    $\mathbf{l}' = \mathbf{F}^T\mathbf{m}$

Epipolar Equation (3)

- Parallel camera setup

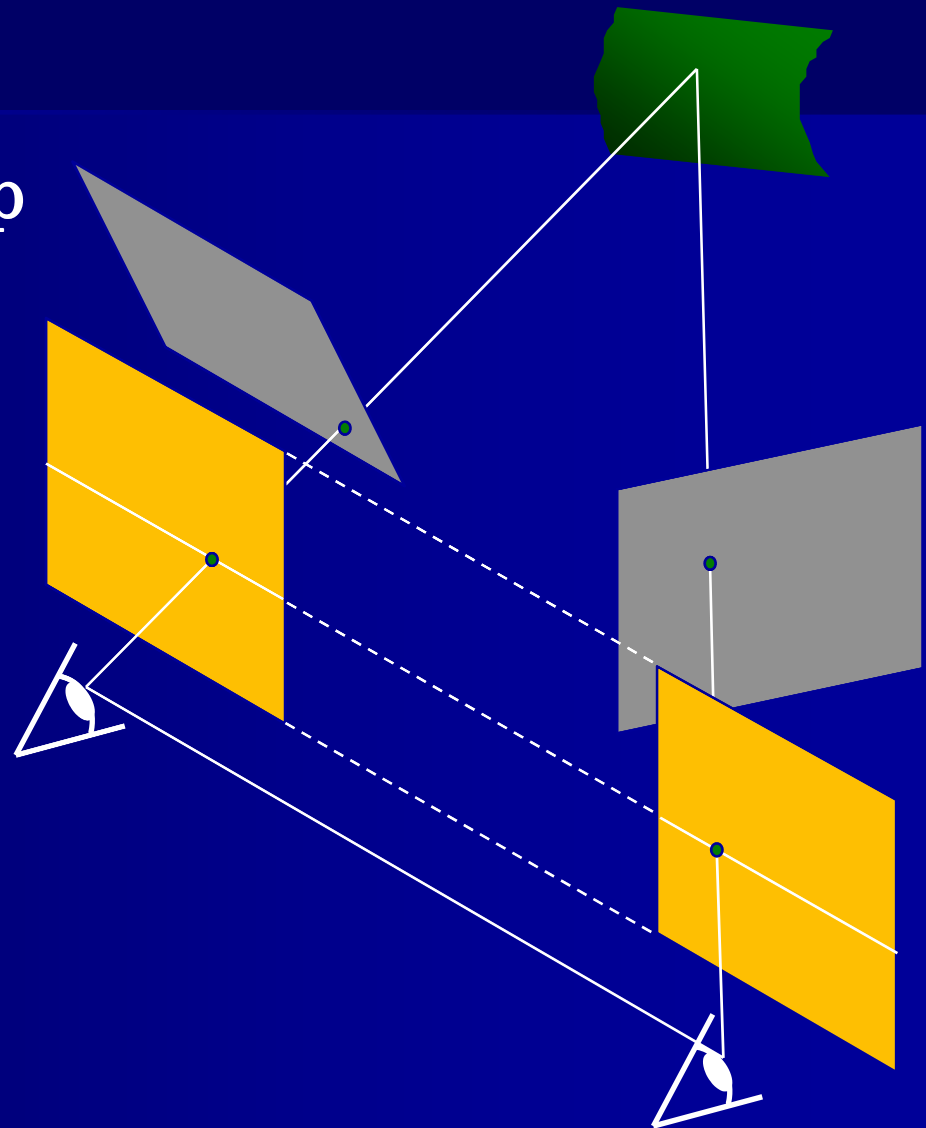
Epipolar Equation (4)

- Convergence camera setup



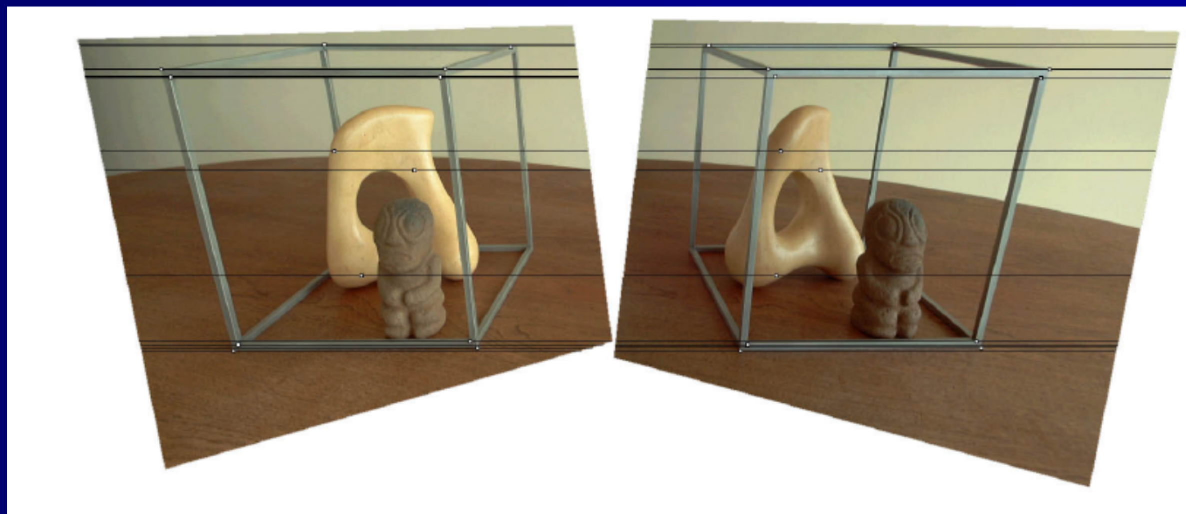
Rectification (1)

- ➔ Making parallel setup
- ➔ Image warping
- ➔ 3x3 homography



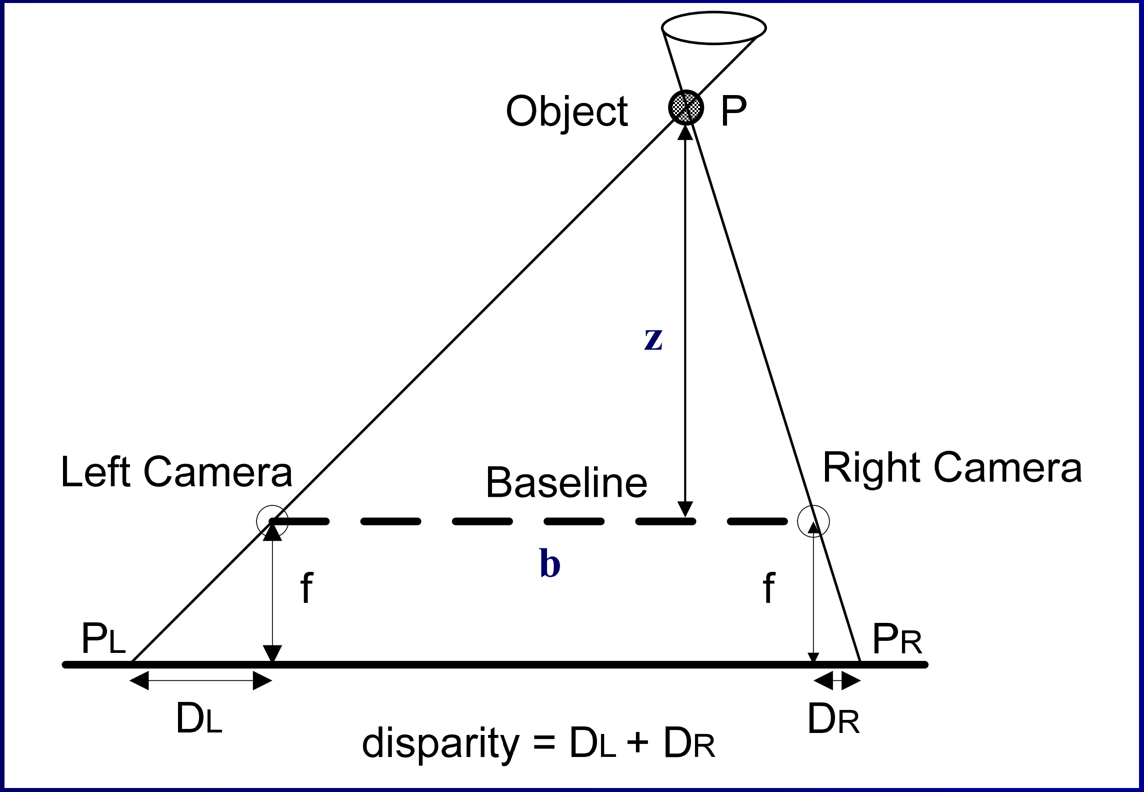
Rectification (2)

➔ Example by A. Efros



Stereogram and Disparity

- Concept of the stereogram(2 images) and disparity



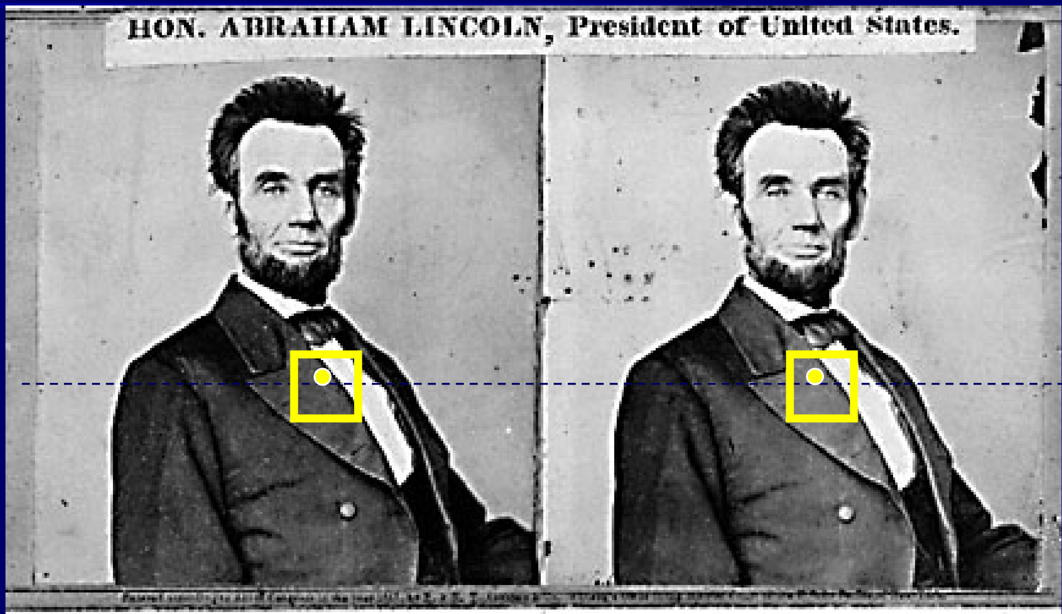
$$Z = fb/d$$

Z: depth

b: baseline length

d: disparity

BM Based Stereo Matching (I)



For each epipolar line

For each pixel in the left image

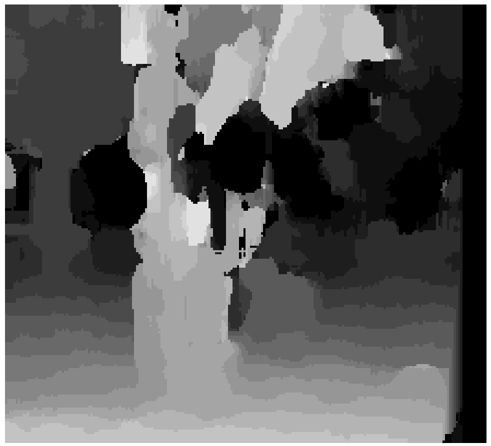
- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

$$E(x, y; d) = \sum_{(x',y') \in N(x,y)} [I_L(x'+d, y') - I_R(x', y')]^2$$

BM Based Stereo Matching (II)

- ➔ Smaller window
  - ➢ more details
  - ➢ more noise
- ➔ Larger window
  - ➢ less noise
  - ➢ less detail

W = 3



W = 20

Topics on Stereo Matching

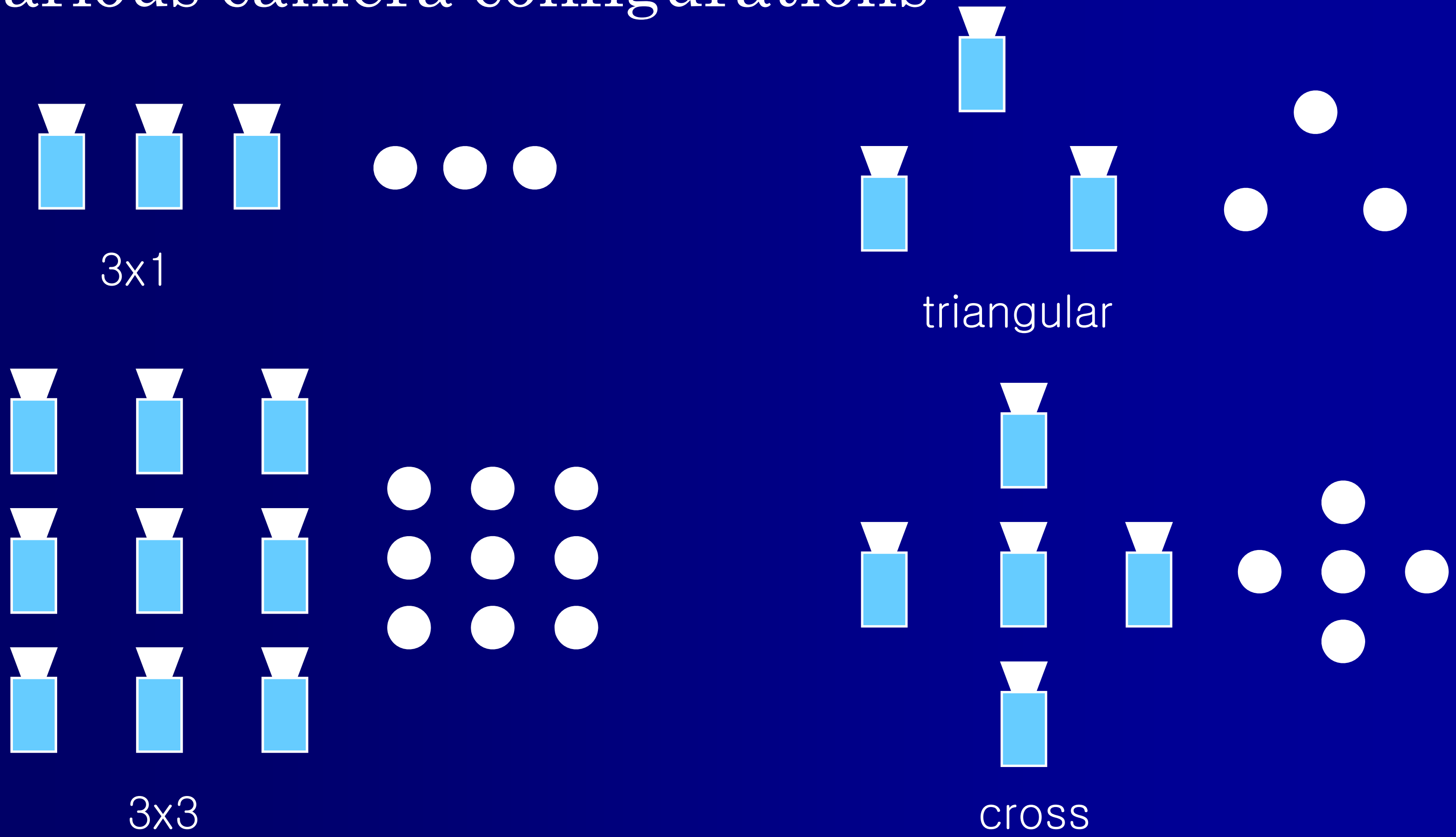
- Fast algorithms
  - GPU implementation
- Optimizations Using MRF mdels
  - Optimizing a cost function to find the best solution
  - Maximum A Posteriori (MAP) algorithm
    - Probabilistic modeling
    - Markov Random Field (MRF) models
  - Dynamic programming
  - Graph-theoretic methods
    - Graph cut, belief propagation
- Combination with various fields modeling
  - Line, occlusion, segmentation

Multi-view Stereo (1)

- Multiple view
  - Three or more views
  - Multi-baseline system
  - Non-planar camera setup
- Advantages
  - Accurate correspondence
  - Occlusion problem
  - 3D reconstruction from multi-view images
    - Panoramic view
    - 3D scene/object reconstruction

Multi-view Stereo (2)

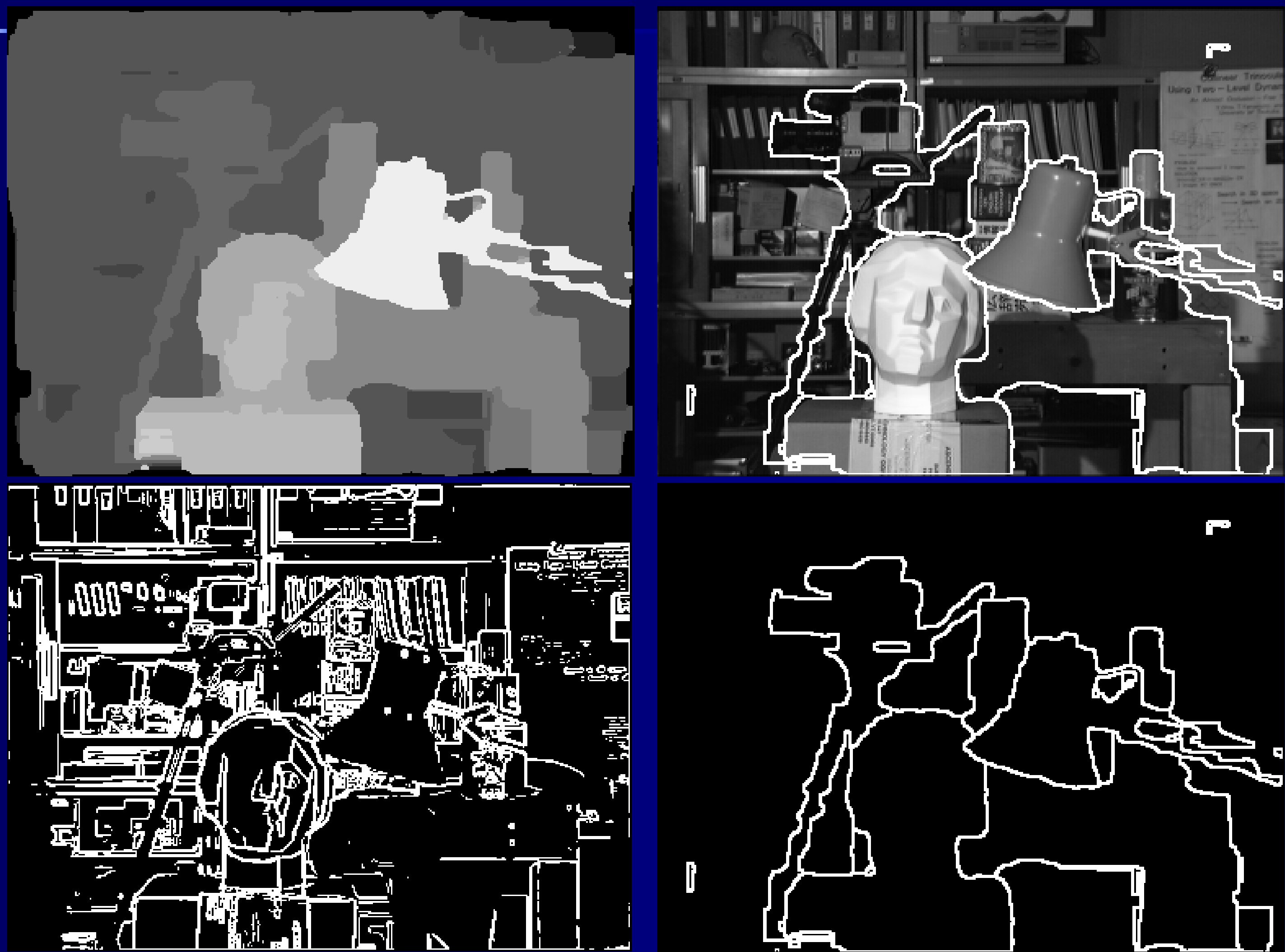
➔ Various camera configurations



Application Issues of Stereo Systems

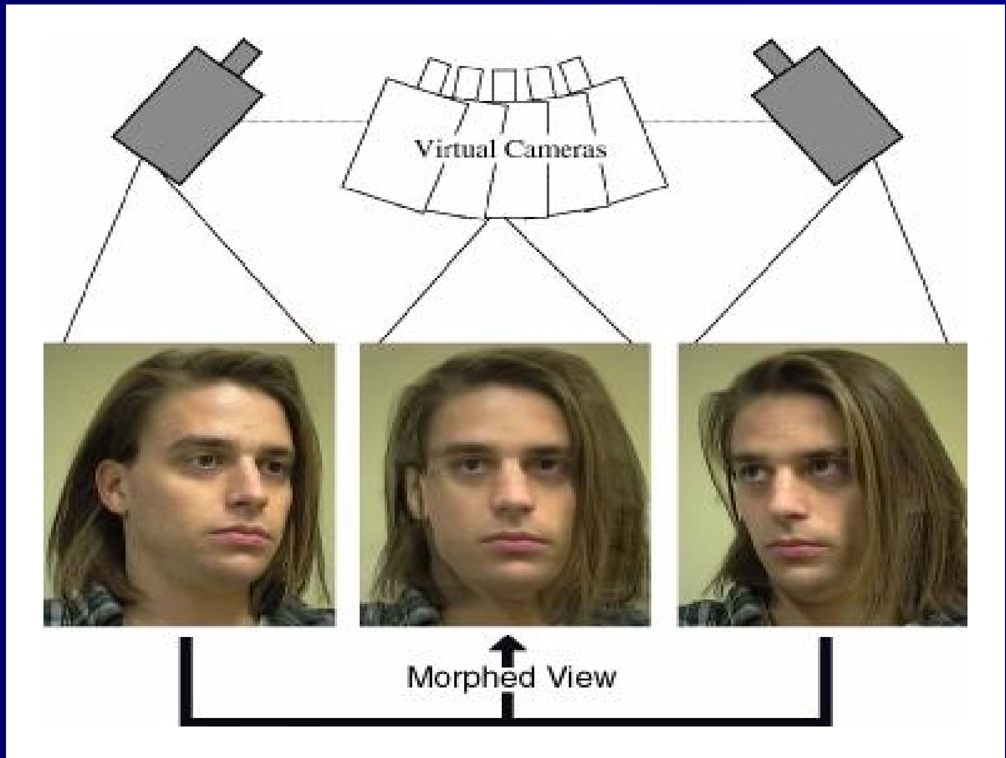
- 3D object/scene reconstruction
  - Image-based rendering, mosaic
- Multiple-view stereoscopic processing
  - Accurate dense disparity field estimation
  - Occlusion problem
- 3D geometrical approaches to stereo matching
- Virtual reality (VR) systems
- Robotics
- Multi-view video coding (MVC)
- 3D display system

Application: Segmentation



Application: View Interpolation

- View morphing between pair of images using epipolar geometry



Application: Depth-Based View Generation (1)

Depth keying + background change

Application: Depth-Based View Generation (2)

- ➔ Z-keying (depth-keying)
  - ➢ mixed synthesys (CMU)

Application: Depth-Based View Generation (3)

Depth keying + object manipulation

Application: Depth-Based View Generation (4)

- Object/background manipulation
  - Depth-based object-background composition
  - Cf: chroma-keying
- Virtual studio system

Application: 3D Modeling (1)

- 3D Modeling using multiple cameras
  - Voxel coloring
  - Space carving

Calibrated 360° rotation (21 images)

Selected Flower Images

Application: 3D Modeling (2)

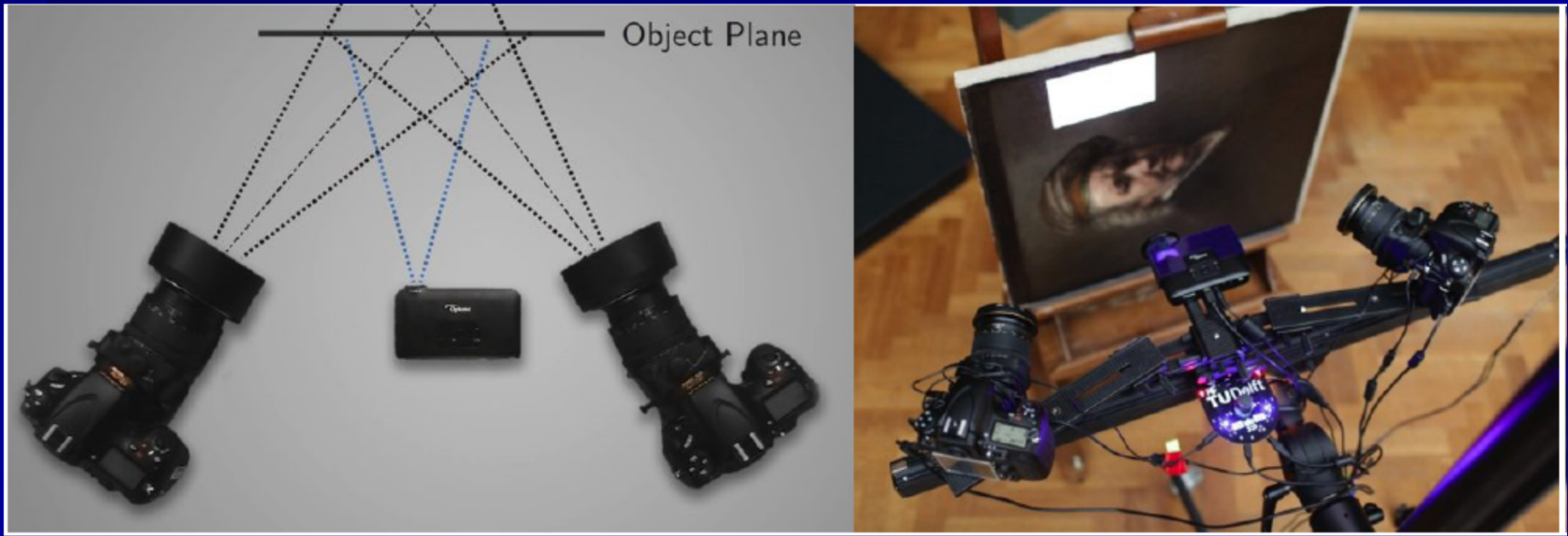
➔ Arbitrary view generation from 3D models

Dinosaur Reconstruction

Flower Reconstruction

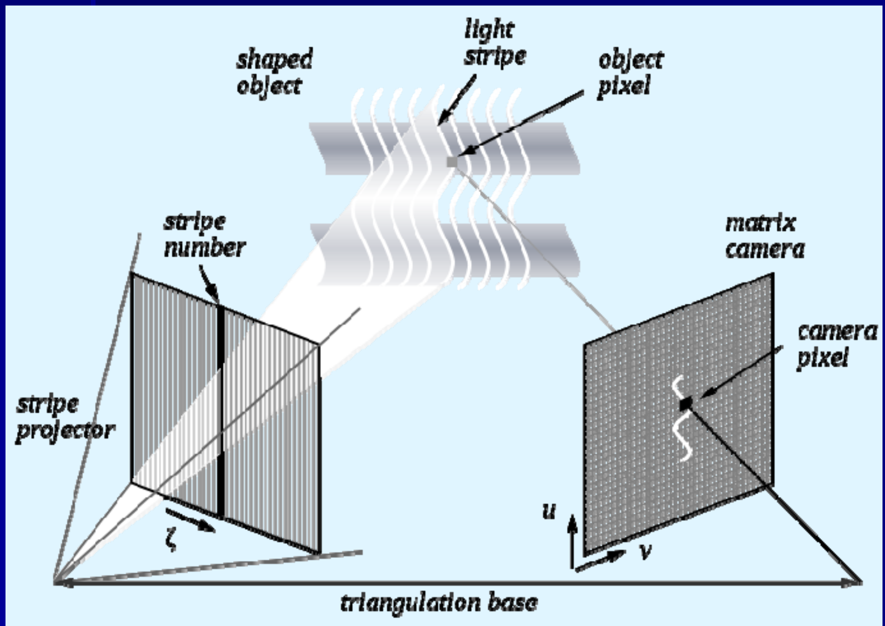
Application: 3D Modeling (3)

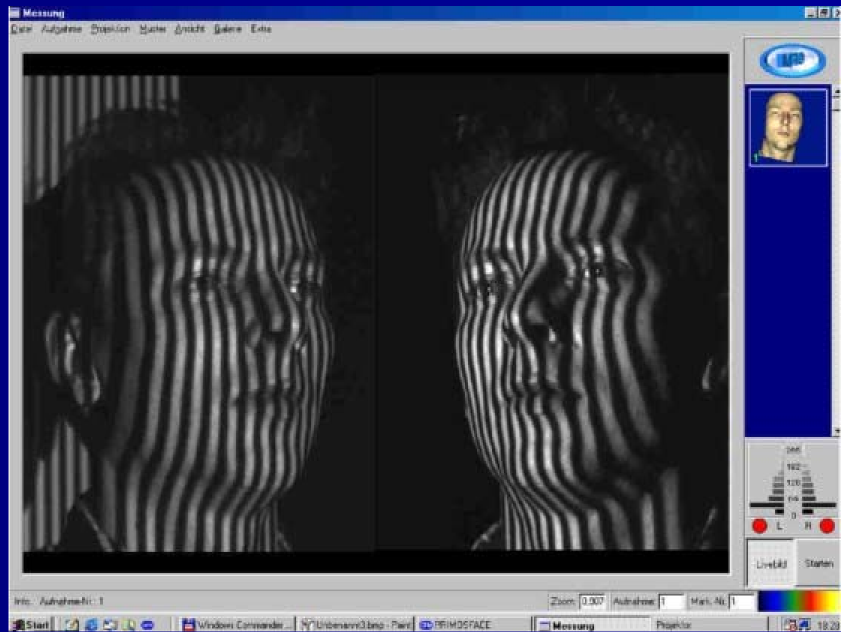
- ➔ Fringe projection and stereo matching
  - ➢ Delft Univ. (Nederland)
  - ➢ Projection fringe patterns onto surface
  - ➢ Capturing by stereo camera
  - ➢ Stereo matching and 3D modeling



Application: 3D Modeling (4)

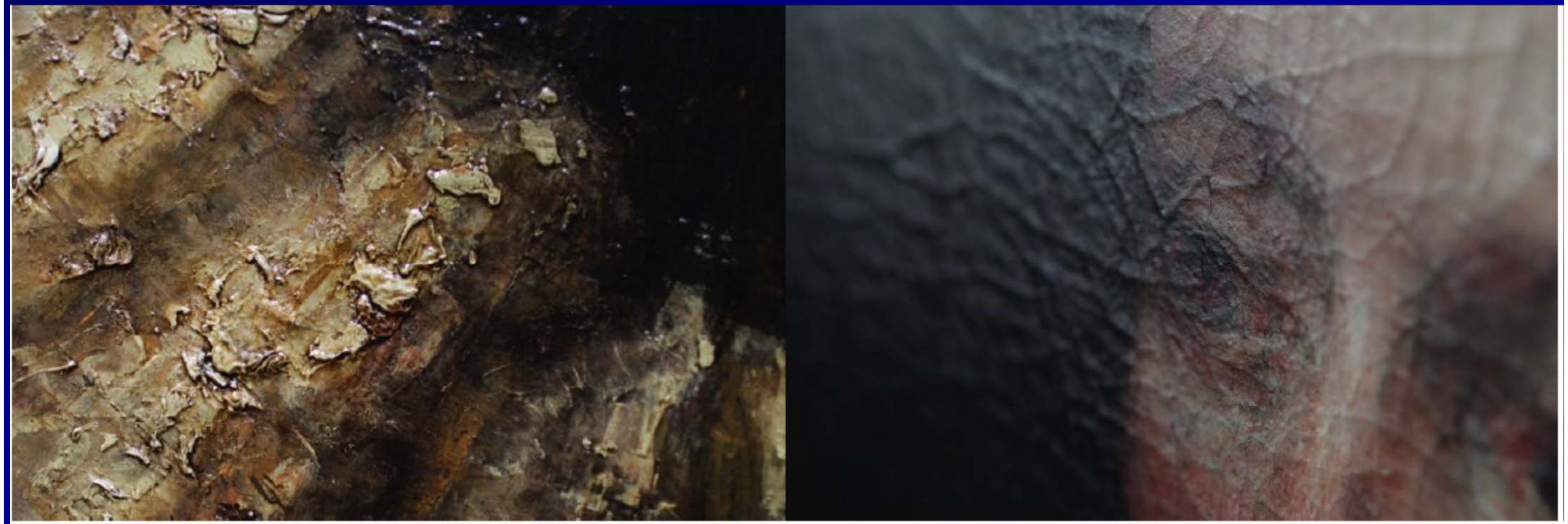
➔ Project structured light patterns
  - Very high resolution and accuracy
  - Matching problem



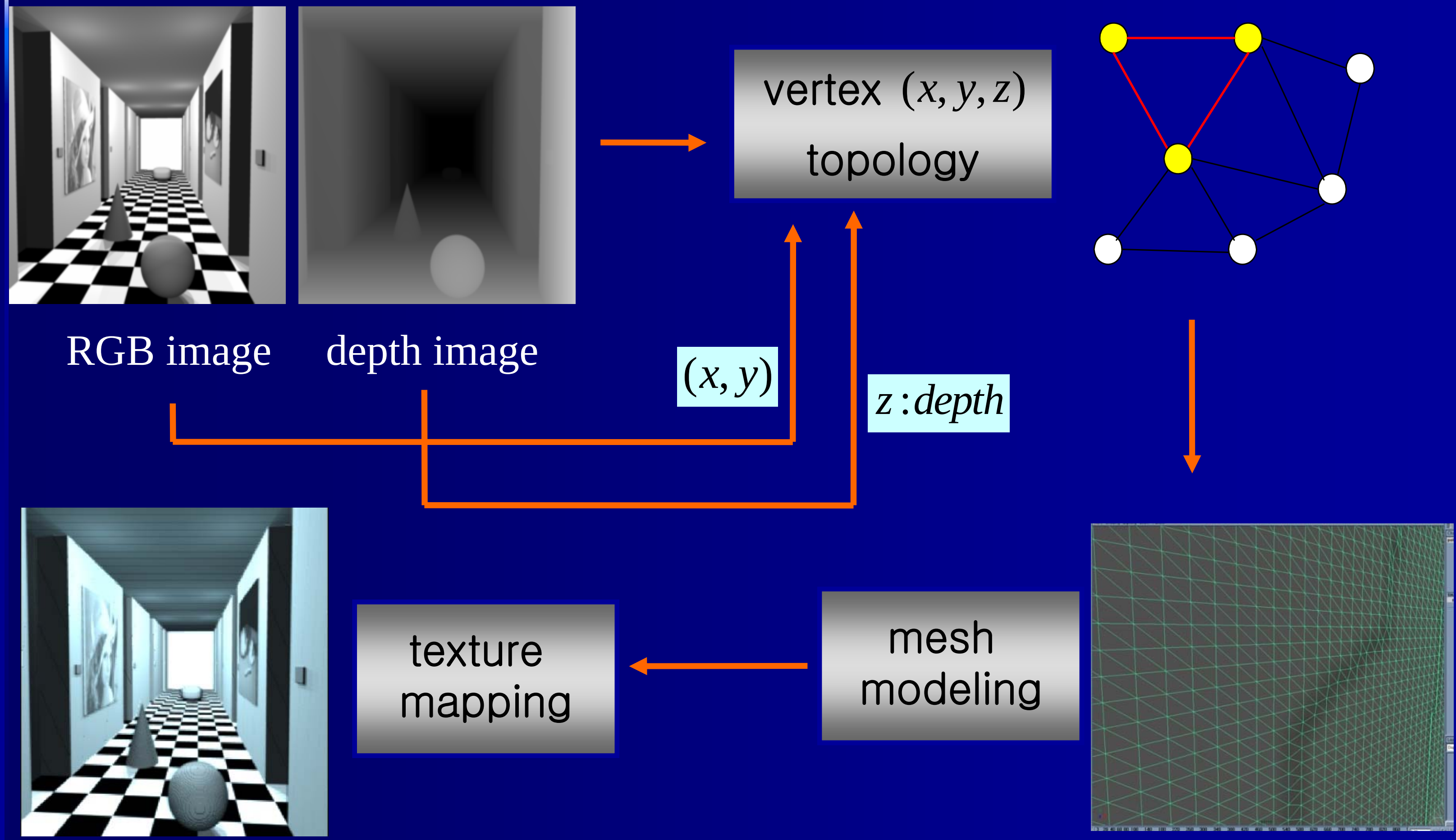


Application: 3D Modeling (5)

- Detail of reconstruction



Application: 3D Model Based Rendering (1)



Application: 3D Model Based Rendering (2)

- Spatial transform can be decomposed into a series of operations.

$$\mathbf{\Pi} = \begin{bmatrix} -fs_x & 0 & x'_c \\ 0 & -fs_y & y'_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{3x3} & \mathbf{0}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{3x3} & \mathbf{T}_{3x1} \\ \mathbf{0}_{1x3} & 1 \end{bmatrix}$$

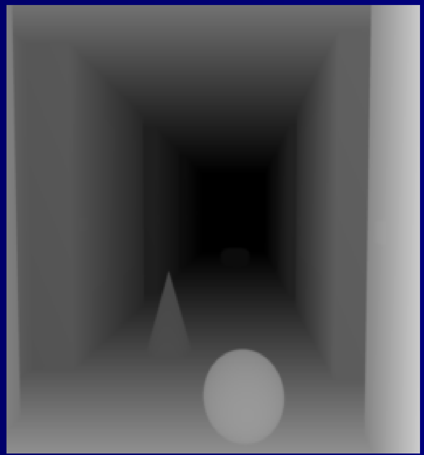
intrinsic projection rotation translation

Spatial transform

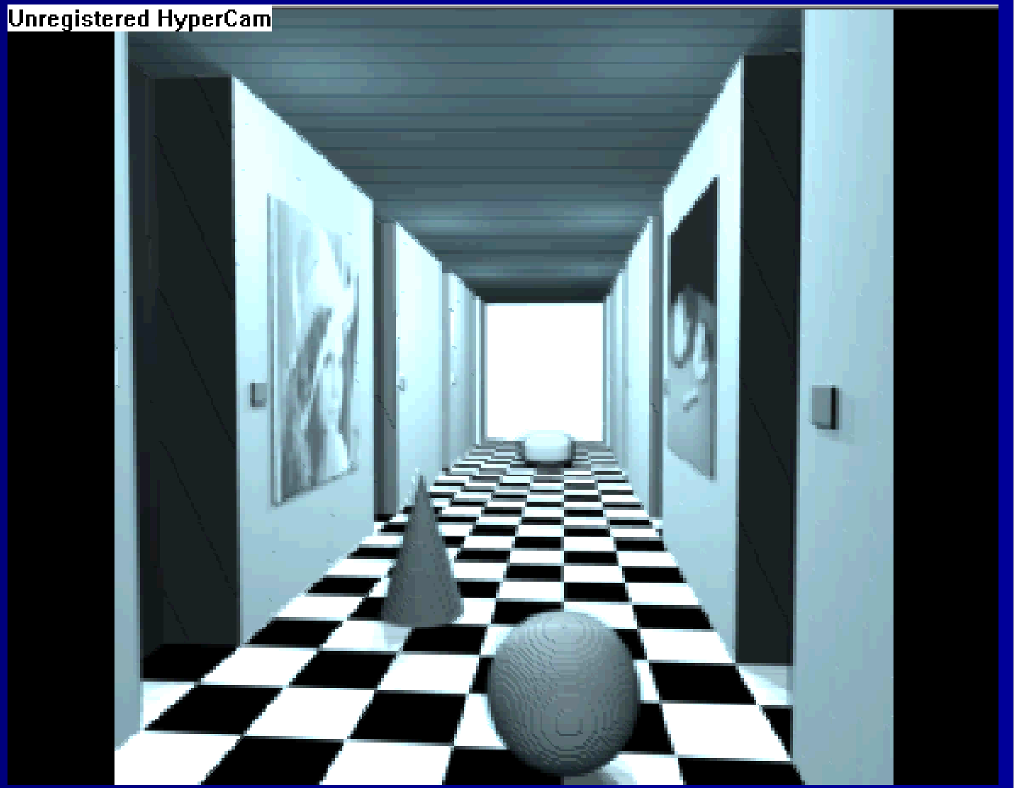
Application: 3D Model Based Rendering (3)

➔ Multi-view rendering

RGB image



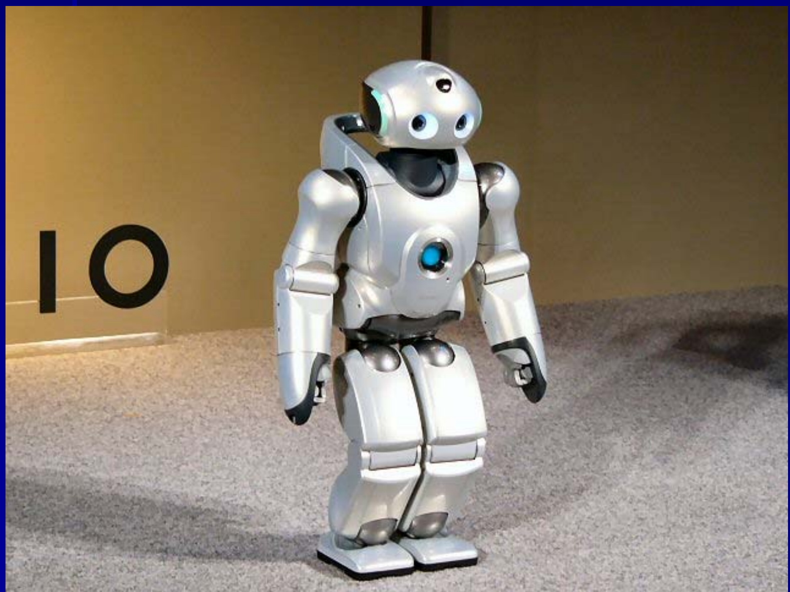
Depth image



3D viewing

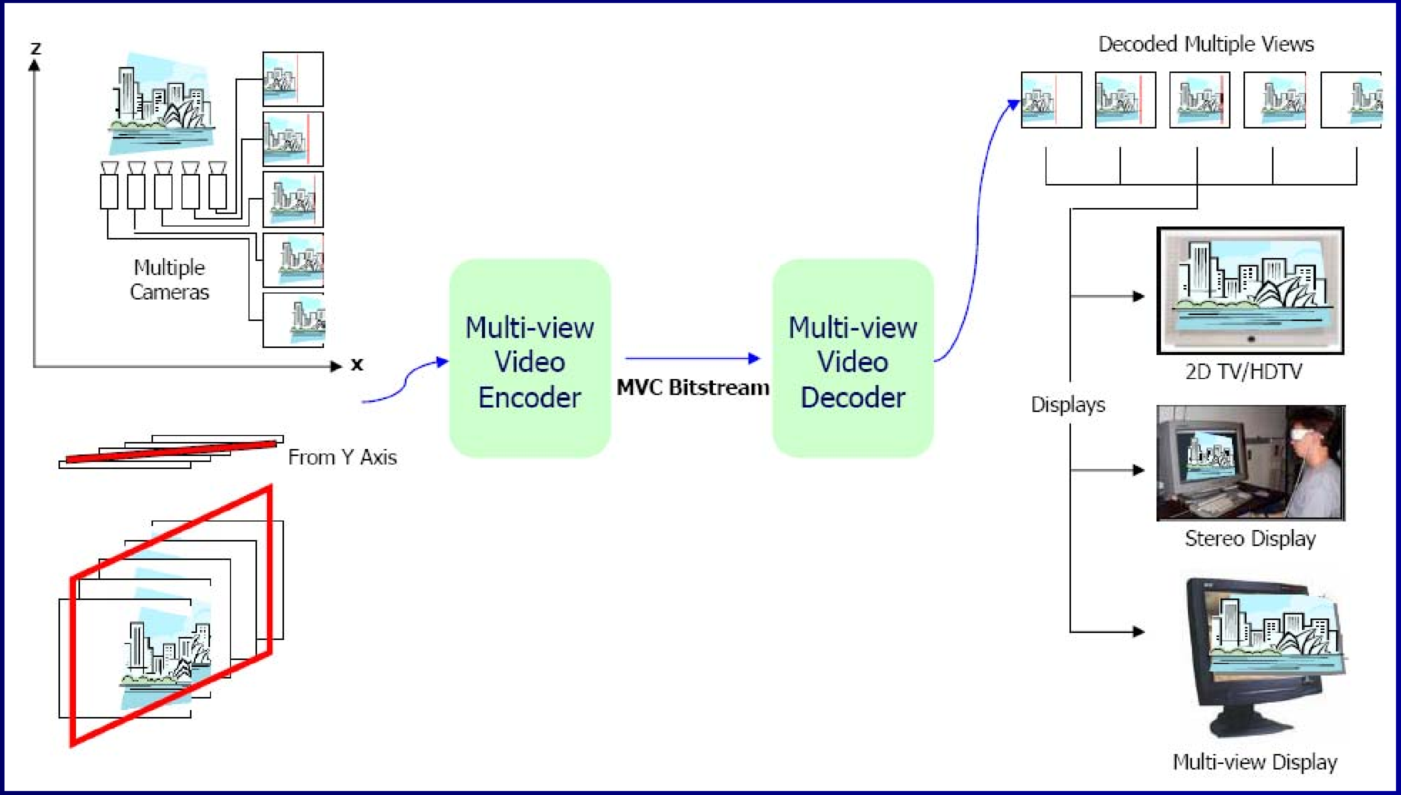
Application: Robotics/Car

- ➔ Stereo camera system for depth estimation
  - ➢ Design of focal length, baseline, and resolution w.r.t the target range



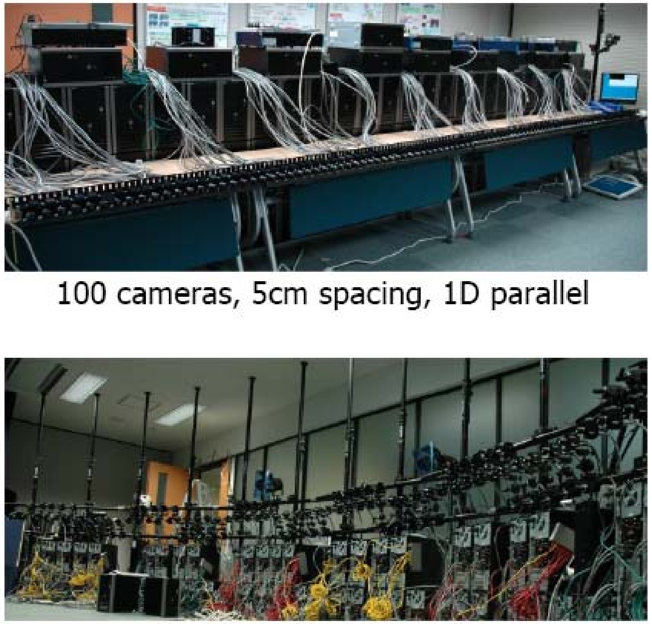
Application: MVC (1)

- ➔ multi-view video display and compression



Application: MVC (2)

- Multi-camera system
  - 2/4/8/16…viewpoints
    - Camera calibration between cameras
  - Compression is most critical.
  - 3D display system



100 cameras, 5cm spacing, 1D parallel

100 cameras, 15.7cm spacing, 1D arc

Comparison od Depth Sensors

	Depth camera	Laser Scanner	Stereo system
Depth resolution	Middle (0.5cm)	High (0.01mm)	Middle (1cm) (depend on cameras)
Depth accuracy	high	Very high	middle
Depth range	variety	variety small (~10m)	Middle (depend on baseline)
Real-time processing	○	○/Δ	○
Application	3D Scene modeling	3D Object modeling	3D Scene modeling