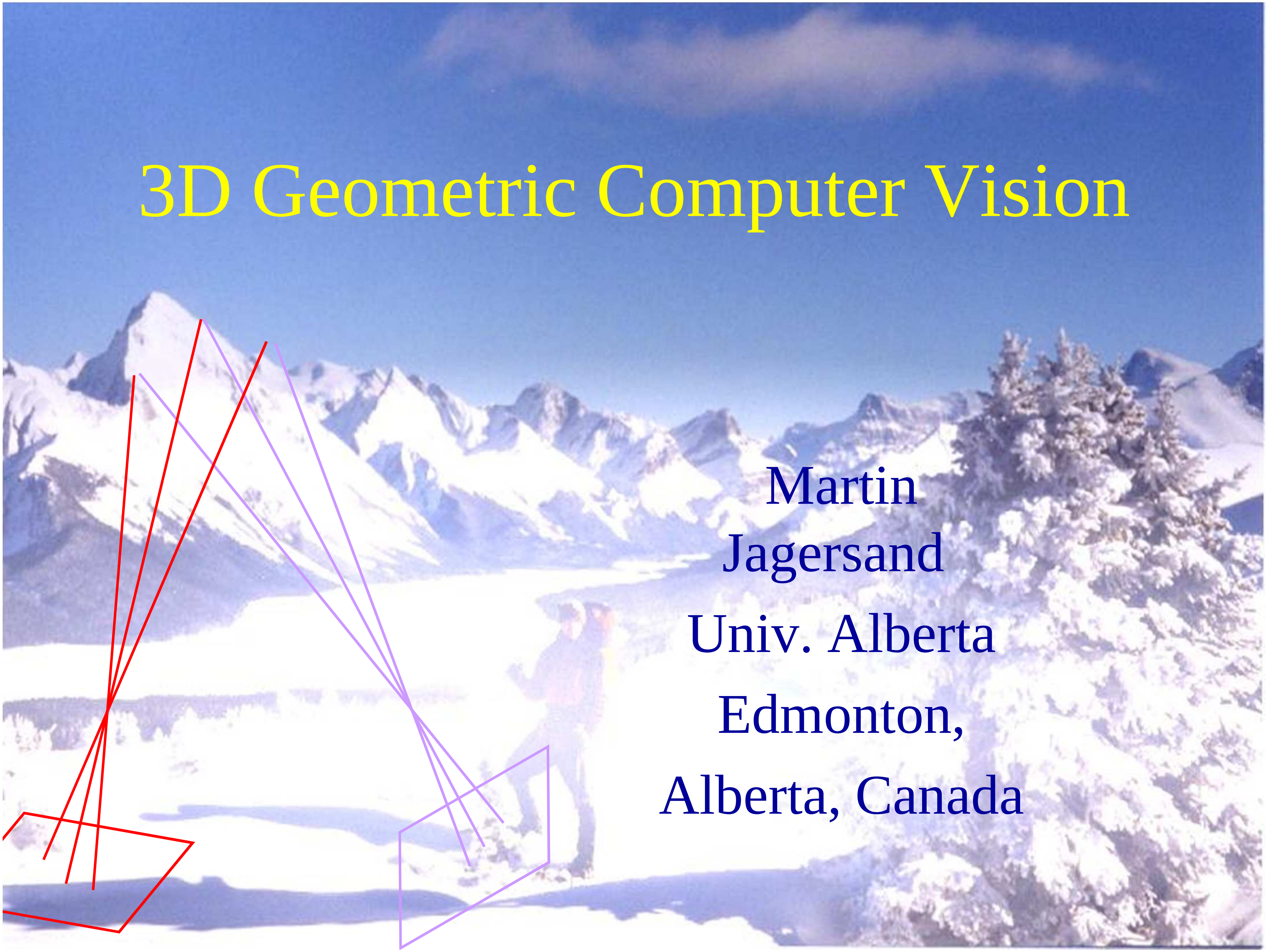
3D Geometric Computer Vision



Martin Jagersand

Univ. Alberta

Edmonton,

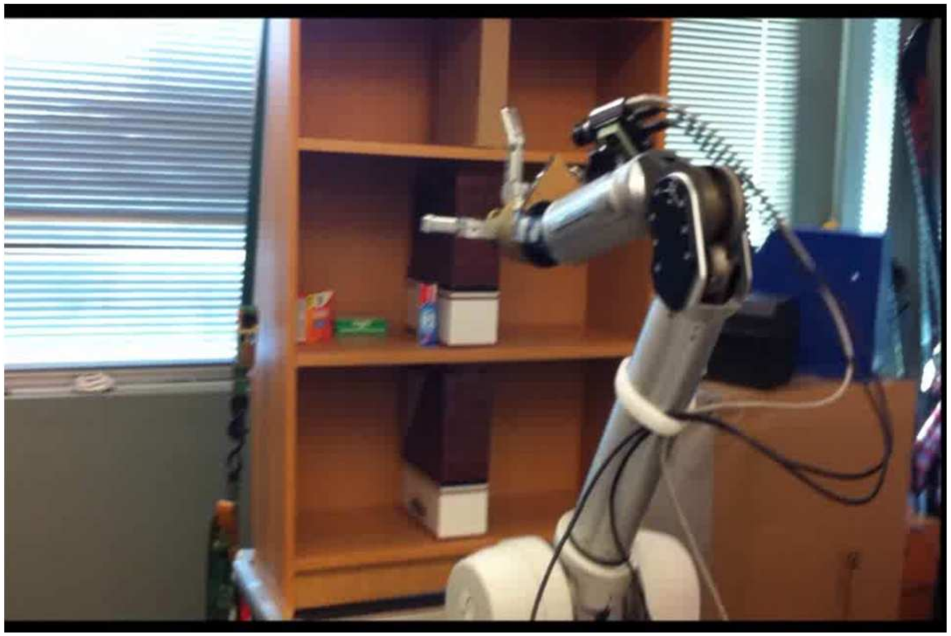
Alberta, Canada

Multi-view geometry

Build 3D models from images

Carry out manipulation tasks

http://webdocs.cs.ualberta.ca/~vis/ibmr/



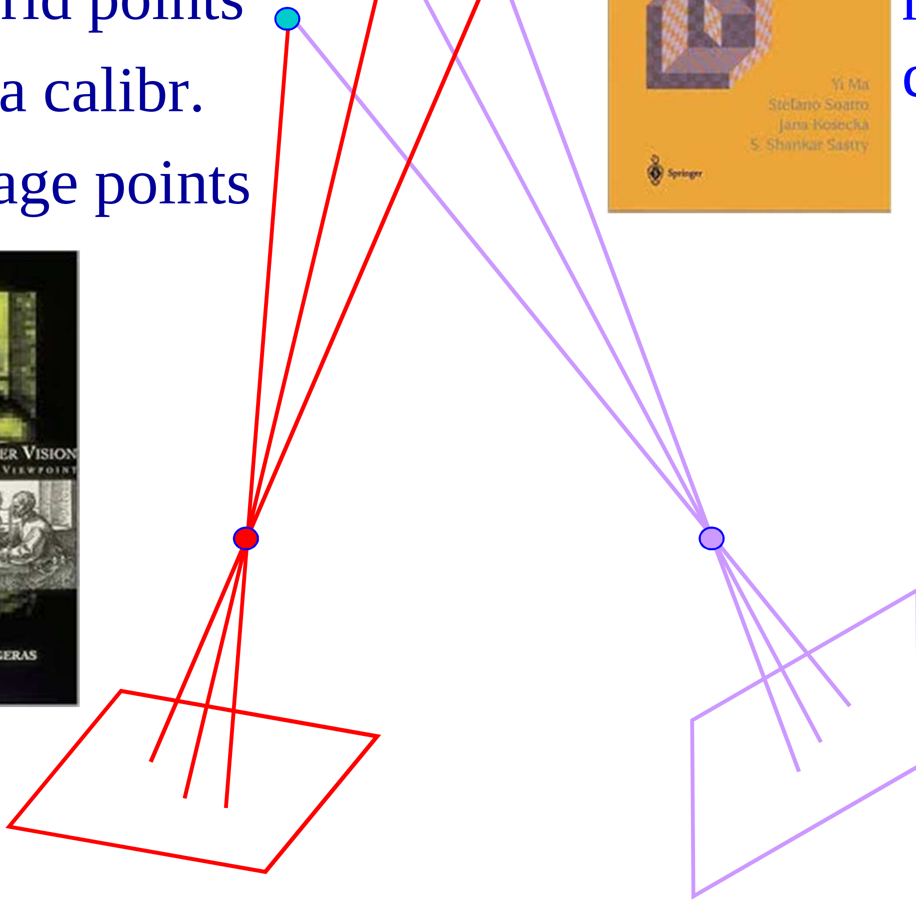
Training for Amazon manipulation challenge

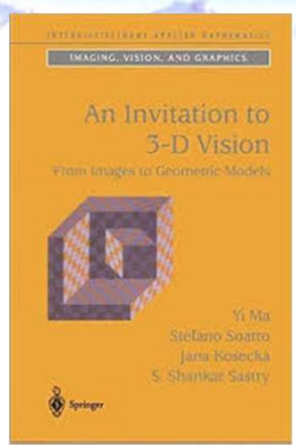
And many other usages…

Multi-view Geometry

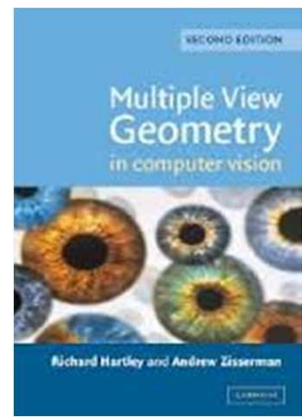
Relates

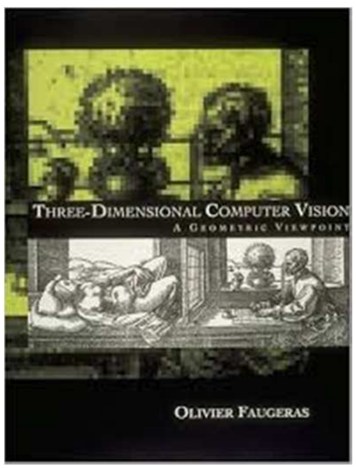
- 3D world points
- Camera calibr.
- 2D image points





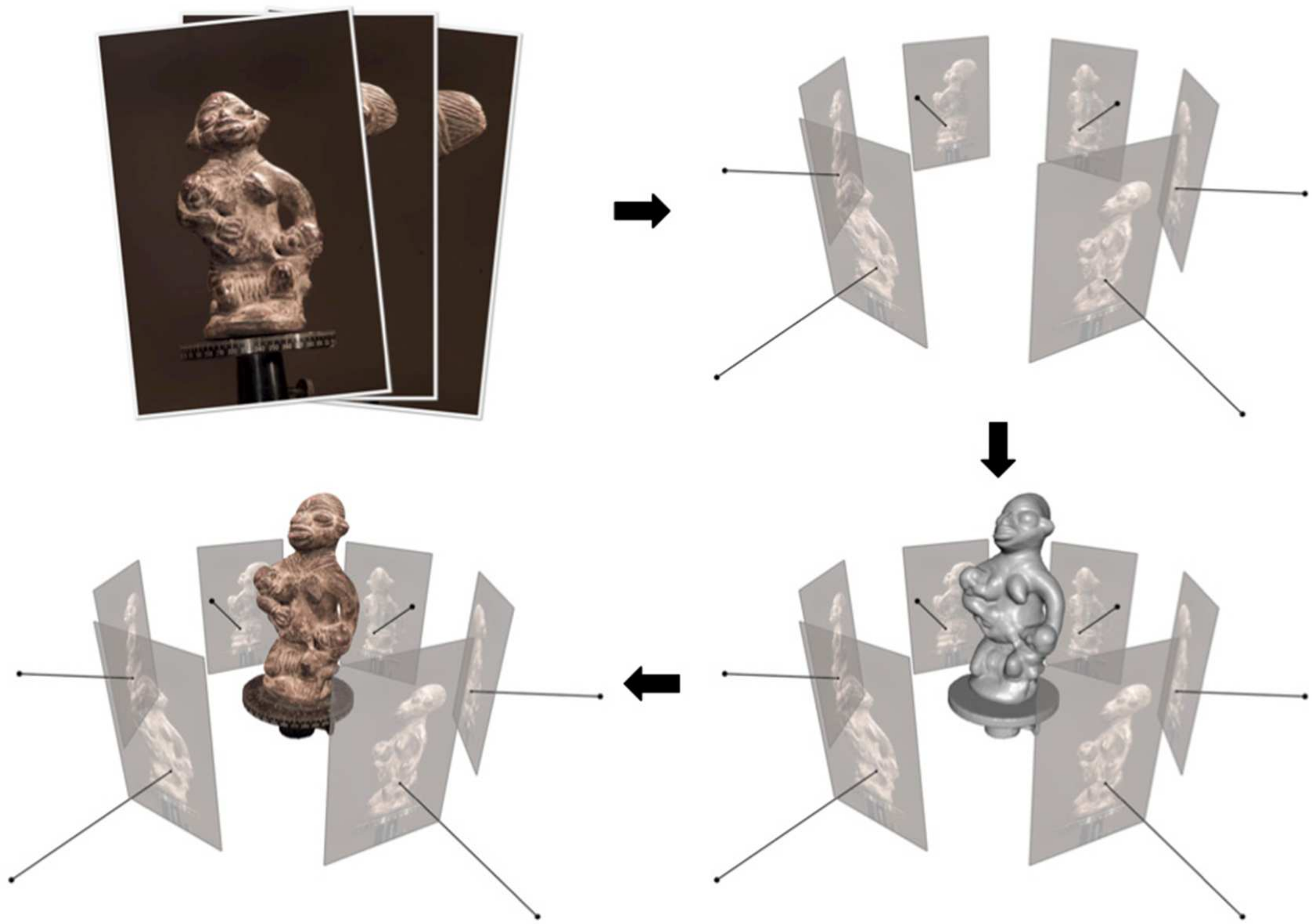
The reconstruction of 3D models of objects from a collection of 2D images



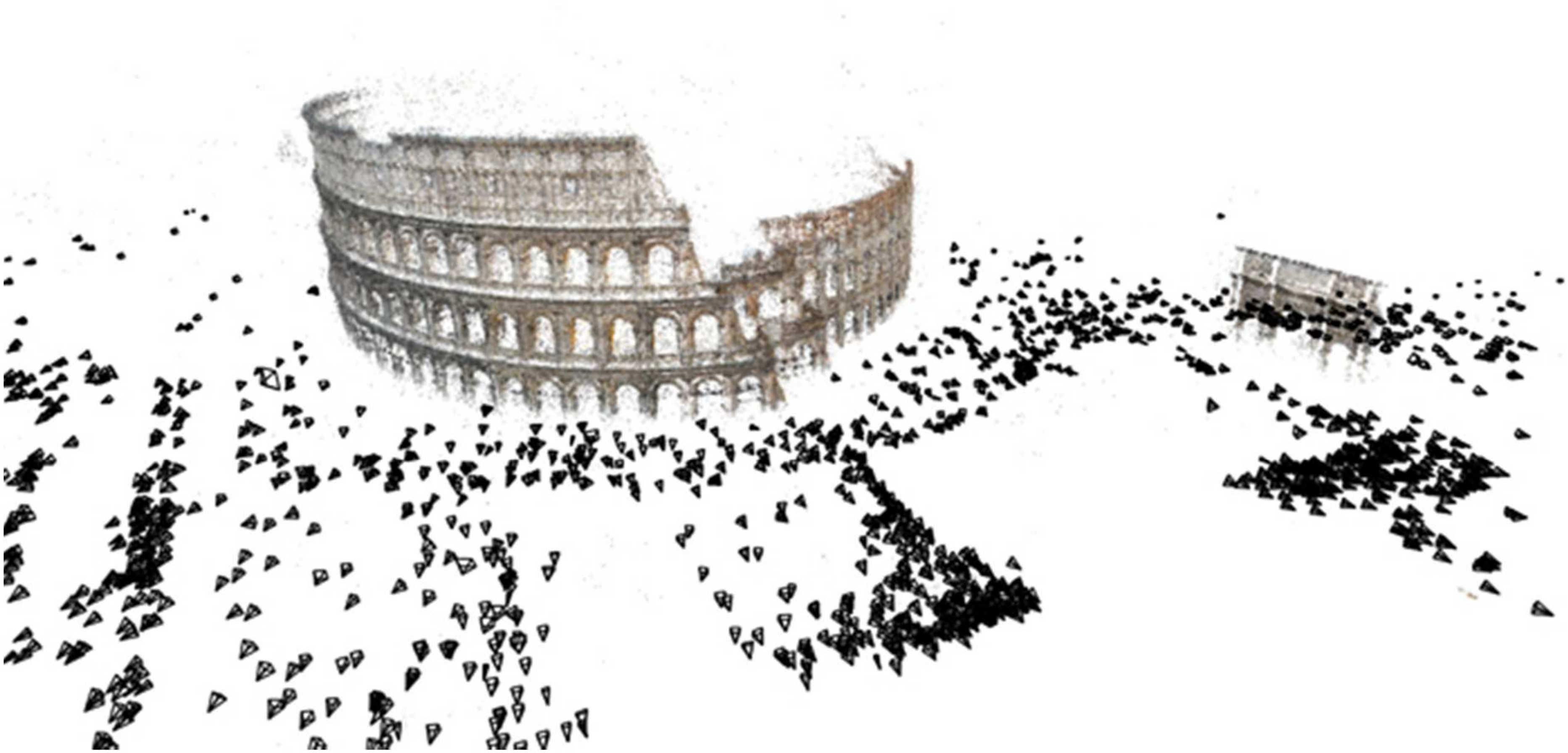


Multi-view Geometry

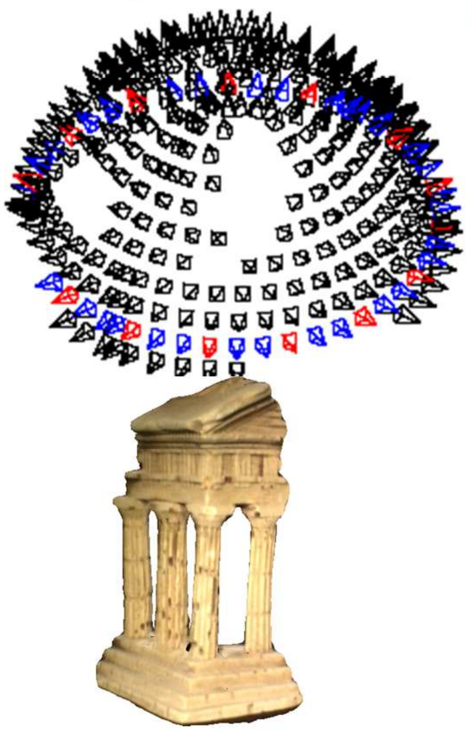
Typical processing pipeline (C. Hernandez MVS tutorial)



Multi-view Geometry



Middlebury multi-view benchmark



ECCV2016_104	0.41	99.6	0.49
ECCV2016_624	0.37	98.9	0.49
Fuhrmann-SG14	0.39	99.4	
Furukawa	0.65	98.7	0.58
Furukawa 2	0.54	99.3	0.55
Furukawa 3	0.49	99.6	0.47
Galliani	0.39	99.2	0.48
Gargallo			0.88
Generalized-SSD	0.53	99.4	0.81
Geodesic GC			
Goesele	0.42	98.0	0.61
Goesele 2007	0.42	98.2	
Guillemaut	0.43	99.0	0.71
Habbecke	0.66	98.0	
Hernandez	0.36	99.7	0.52
Hongxing	0.83	95.7	0.79
Hornung	0.58	98.7	
ICCV2015_1020	0.45	99.2	0.56
ICCV2015_293	0.52	99.2	

A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms, Seitz, Curless, Diebel, Szeliski
CVPR 2006, vol. 1, pages 519-526. Cited by 1479

Calibration accuracy on these datasets appears to be on the order of a pixel (a pixel spans about 1/4mm on the object).

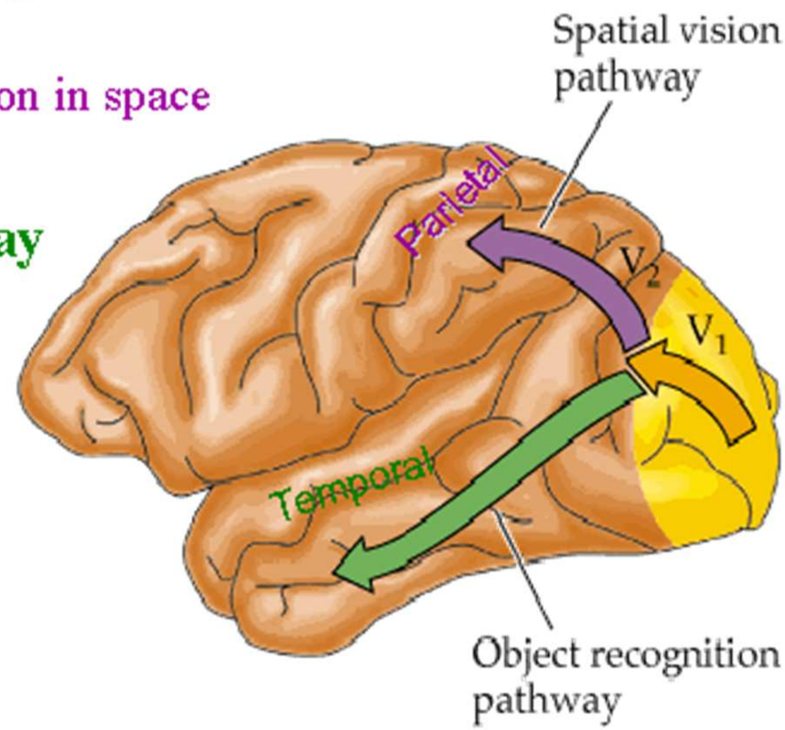
Dorsal and Ventral Pathways Where/What or Action/Perception?

Dorsal (magno) Pathway

- to parietal lobe
- spatial vision – localization in space
- "WHERE"

Ventral (parvo) Pathway

- to temporal lobe
- object recognition
- "WHAT"



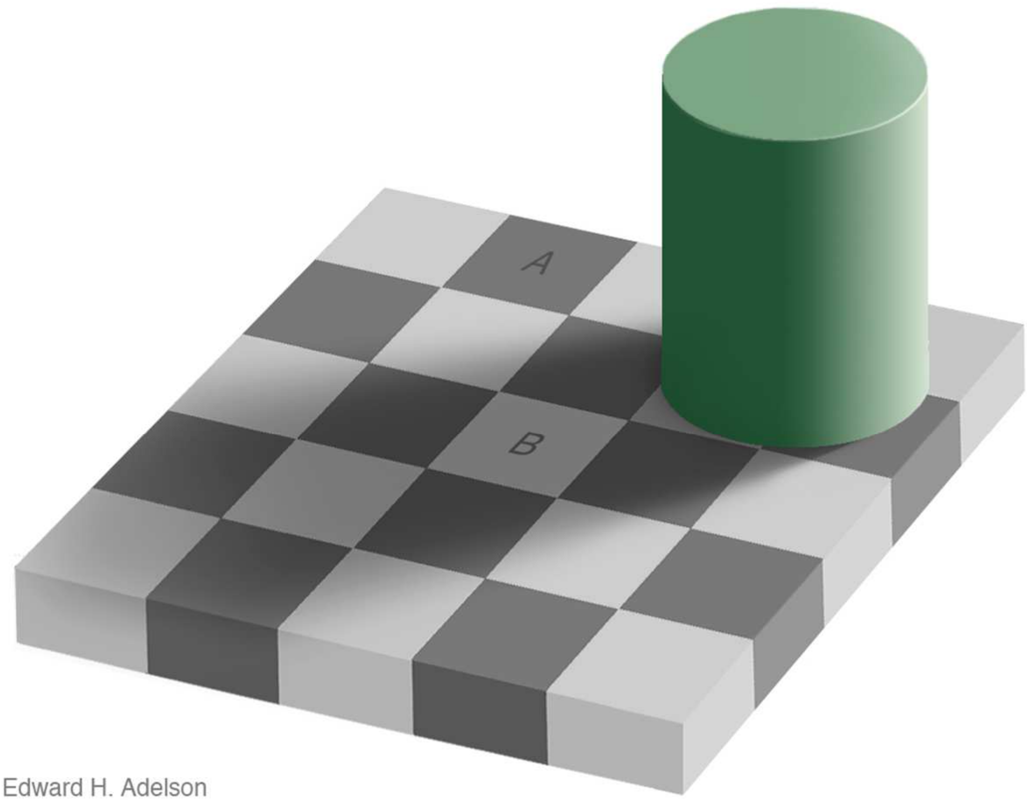
- Humans don't internalize detailed 3D maps!
- Use external world as map. (Hayhoe, Pelz, Rensink, Goodale etc)

Course content

- Video processing: Image motion and tracking
- 2D projective geometry
- 3D projective geometry
- Cameras, their math model and calibration
- Two-camera “stereo” 3D reconstruction
- Multi-view geometry and general 3D reconstruct.
- Light, surface reflectance and math modeling
- Computer vision systems

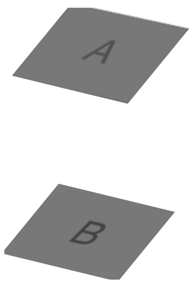
Challenges:
Geometric size ambiguity

Challenges
Light / photometric ambiguity



Edward H. Adelson

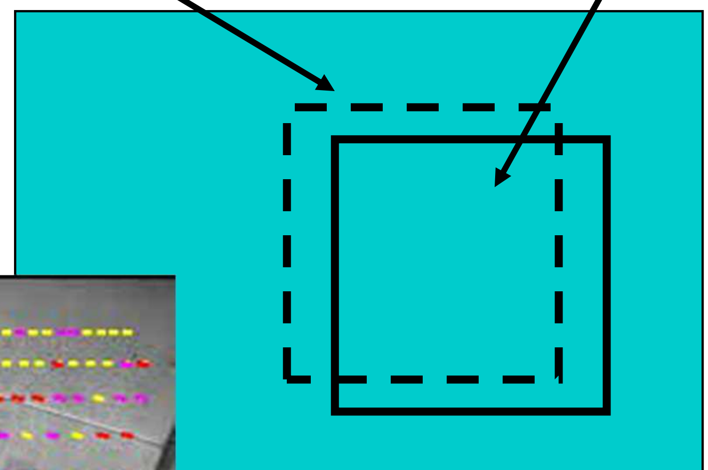
Challenges
Light / photometric ambiguity

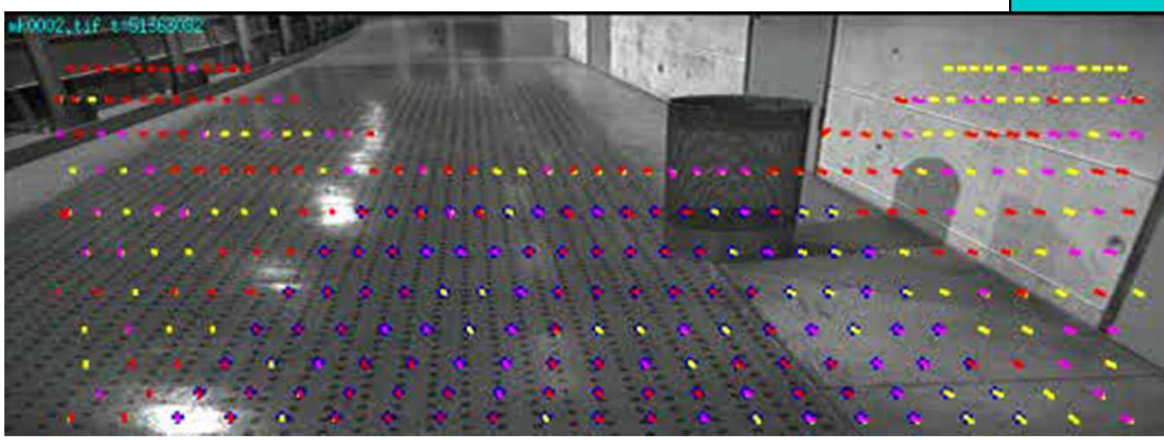


Fundamental types of video processing “Visual motion detecton”

- Relating two adjacent frames: (small differences):

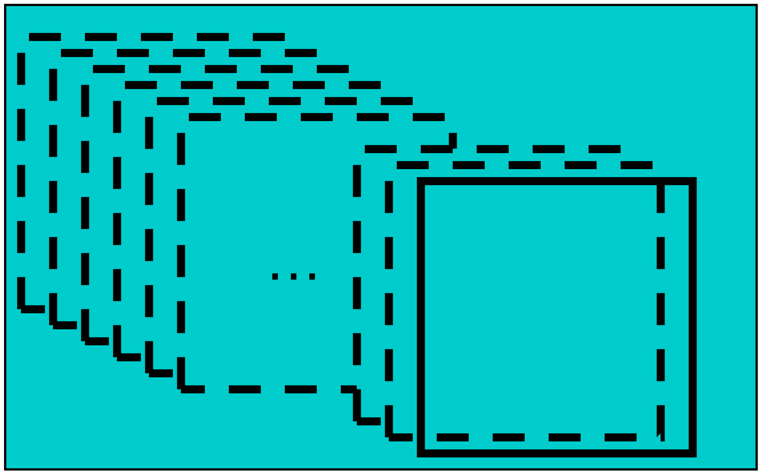
$$\mathrm{Im}(x+\delta x,\ \ y+\delta y,\ \ t+\delta t)\ \ =\ \ \mathrm{Im}(x,\ \ y,\ \ t)$$

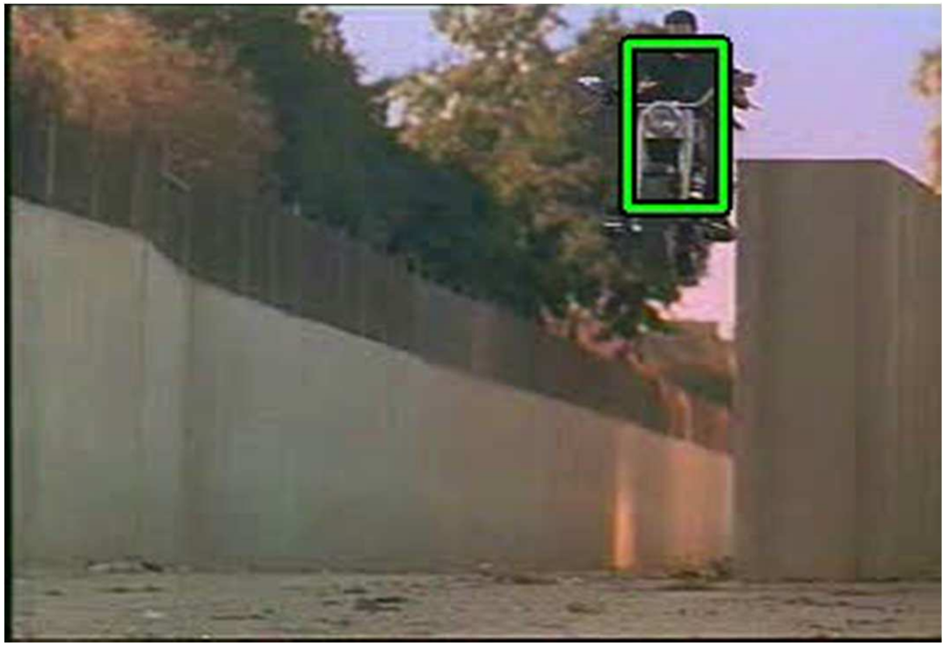




Fundamental types of video processing "Visual Tracking" / Stabilization

- Globally relating all frames: (large differences):





Intensity images / video non-trivial to get information

abs(Image 1 - Image 2) = ?

Note: Almost all pixels change!

Constancy: The physical scene is the same

How do we make use of this?

MTF – Modular Tracking Framework

Modular Tracking Framework

A Unified Approach to Registration based Tracking

Abhineet Singh and Martin Jagersand

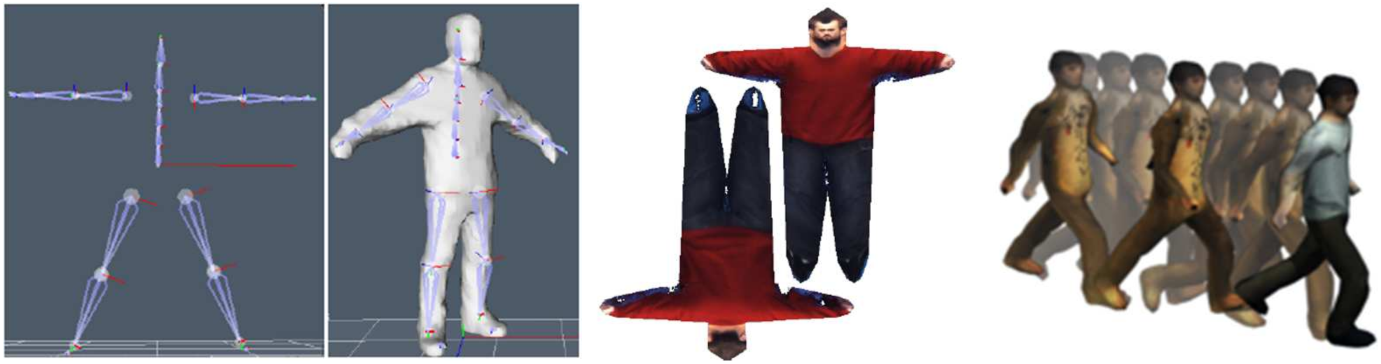
UNIVERSITY OF ALBERTA

- Open source
- C++ implementation
- ROS interface
- Matlab/Pyhton
- Cross platform

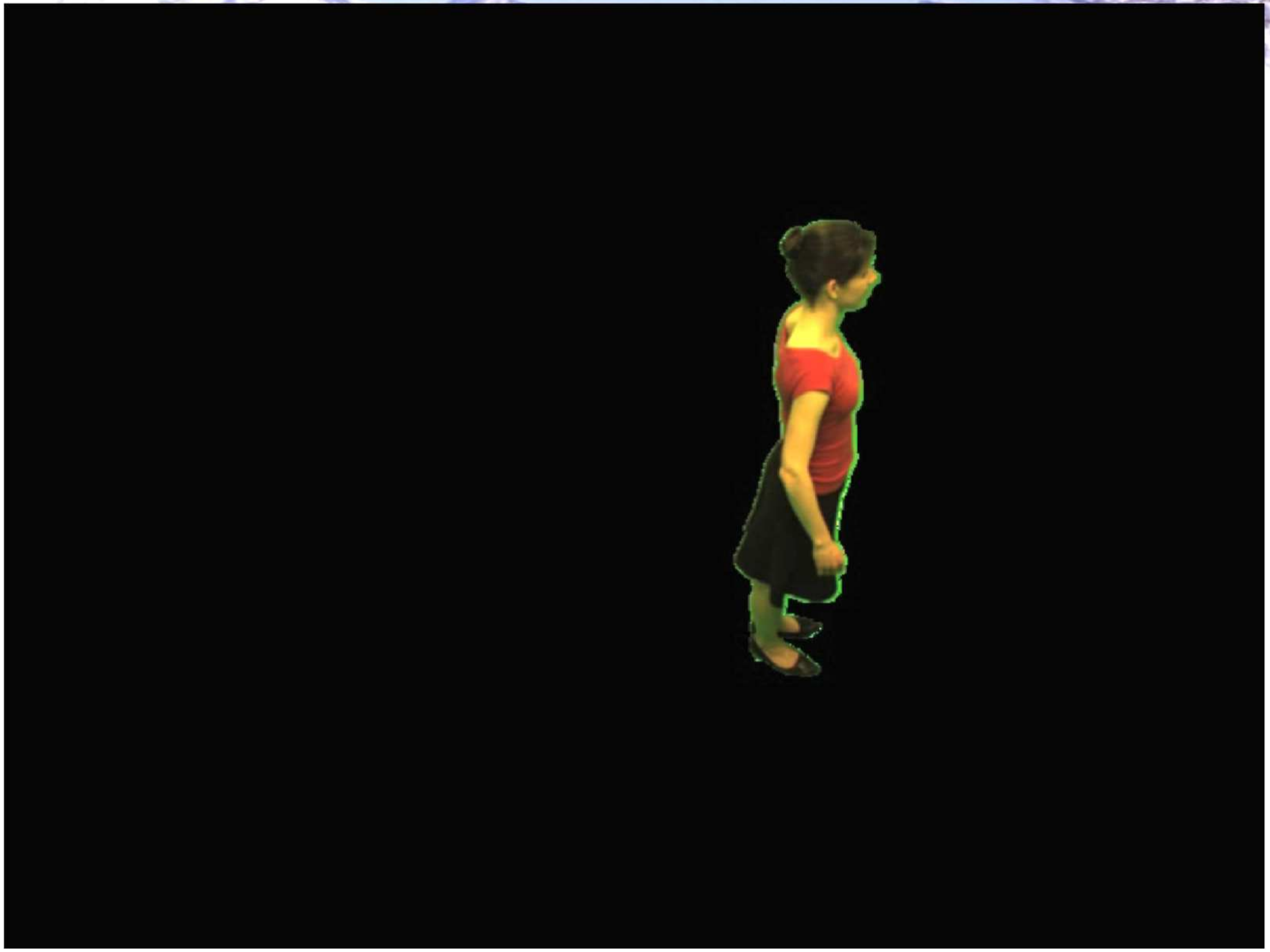
Application: Augmenter Reality

Beyond 3D
Non-rigid and articulated motion

- Humans ubiquitous in graphics applications
- A practical, realistic model requires
  - Skeleton
  - Geometry (manually modeled, laser scanned)
    - Physical simulation for clothes, muscle
  - Texture/appearance (from images)
  - Animation (mocap, simulation, artist)



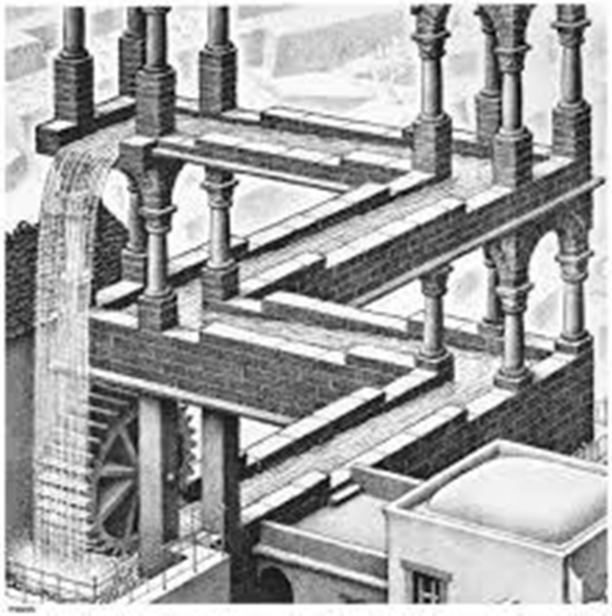
Beyond 3D: Non-rigid and articulated motion



PhD work of Neil Birkbeck, Best thesis prize winner

Uses of partial information from 2D-3D camera geometry

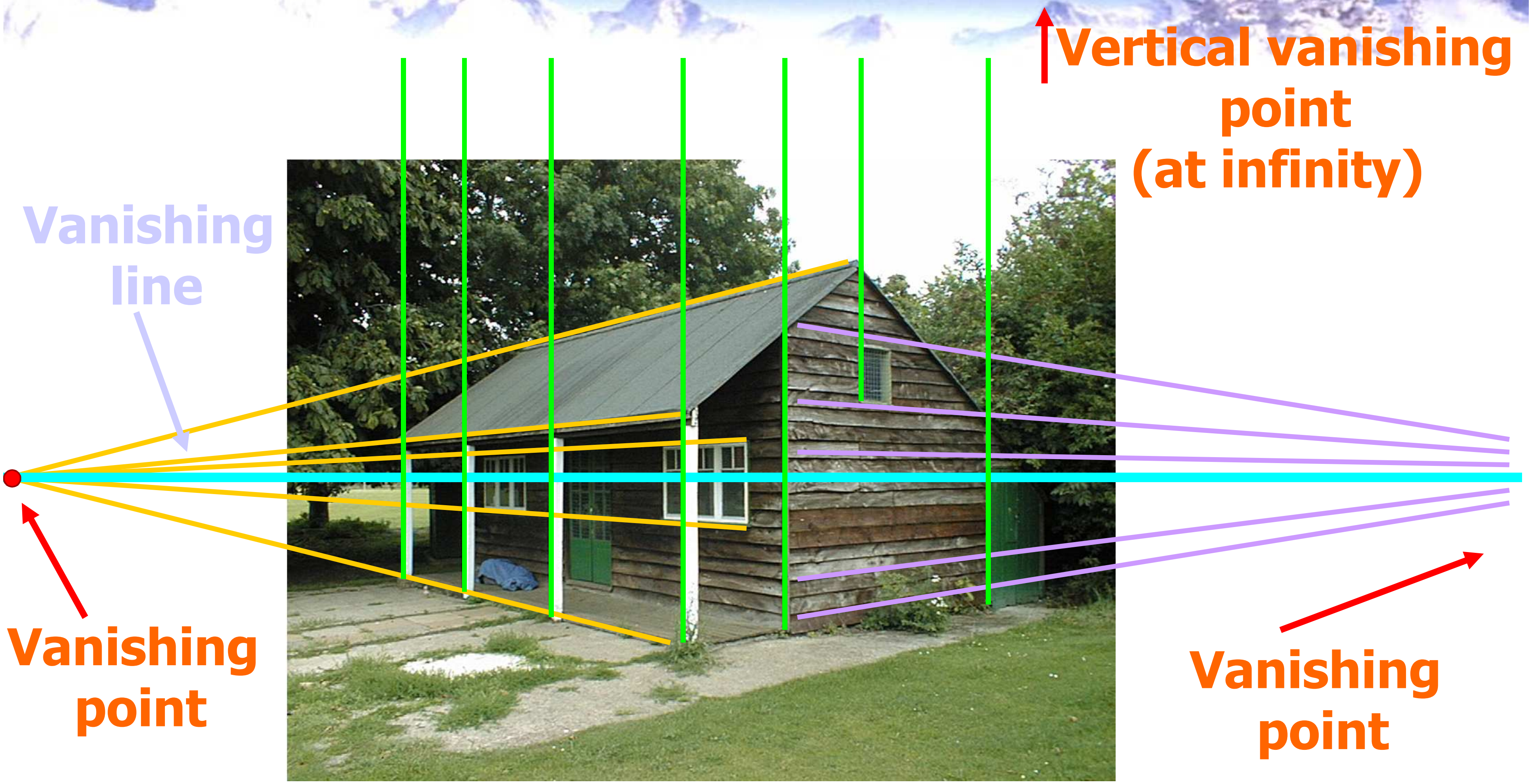
- Measurements in single images
- Visual constraints: verify alignments, detect impossible configurations (Escher paintings)
- Visual servoing
- Video tracking
- Rendering
- … many more



Why? Compact, accurate

Few relative alignments vs. complete global geometry

Geometric Cues



Single View Metrology

Estimating Height

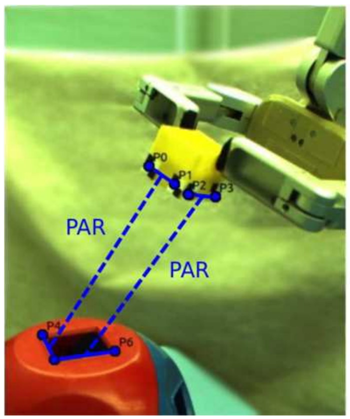
- The distance $\| t_r - b_r \|$ is known
- Used to estimate the height of the man in the scene

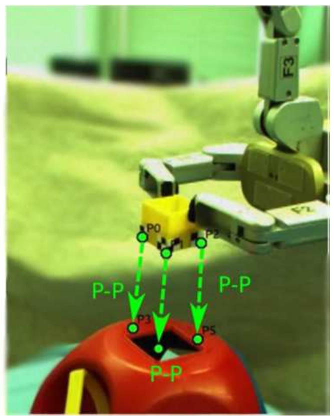
Single View Metrology (559 citations)

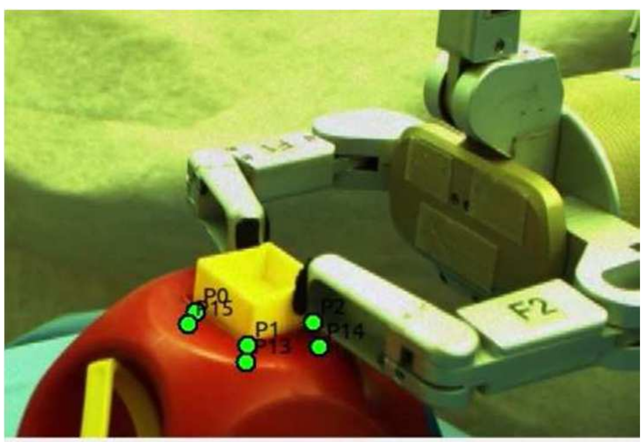
A. Criminisi, I Reid, A Zisserman
International Journal of Computer Vision 40 (2), 123-148, 2000

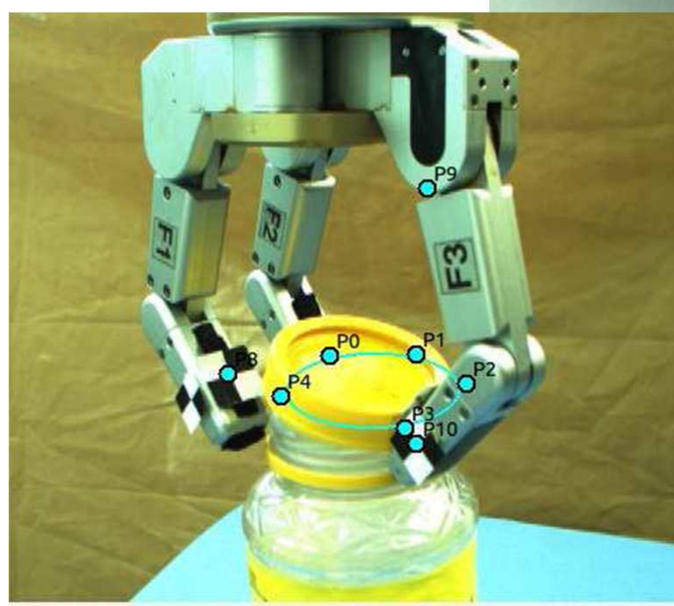
Geometry for hand-eye coordination

Image-Based Visual Servoing (IBVS)

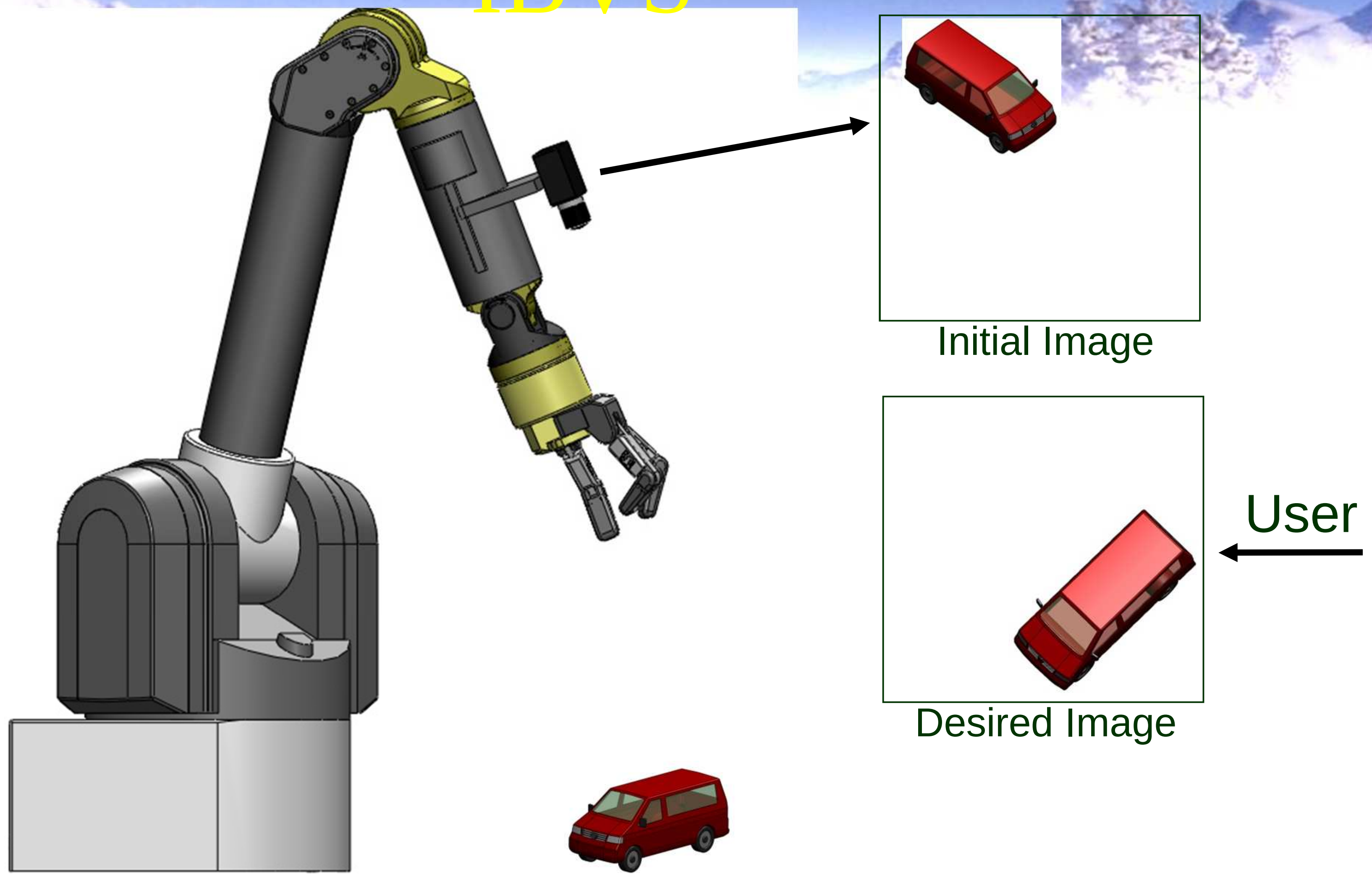




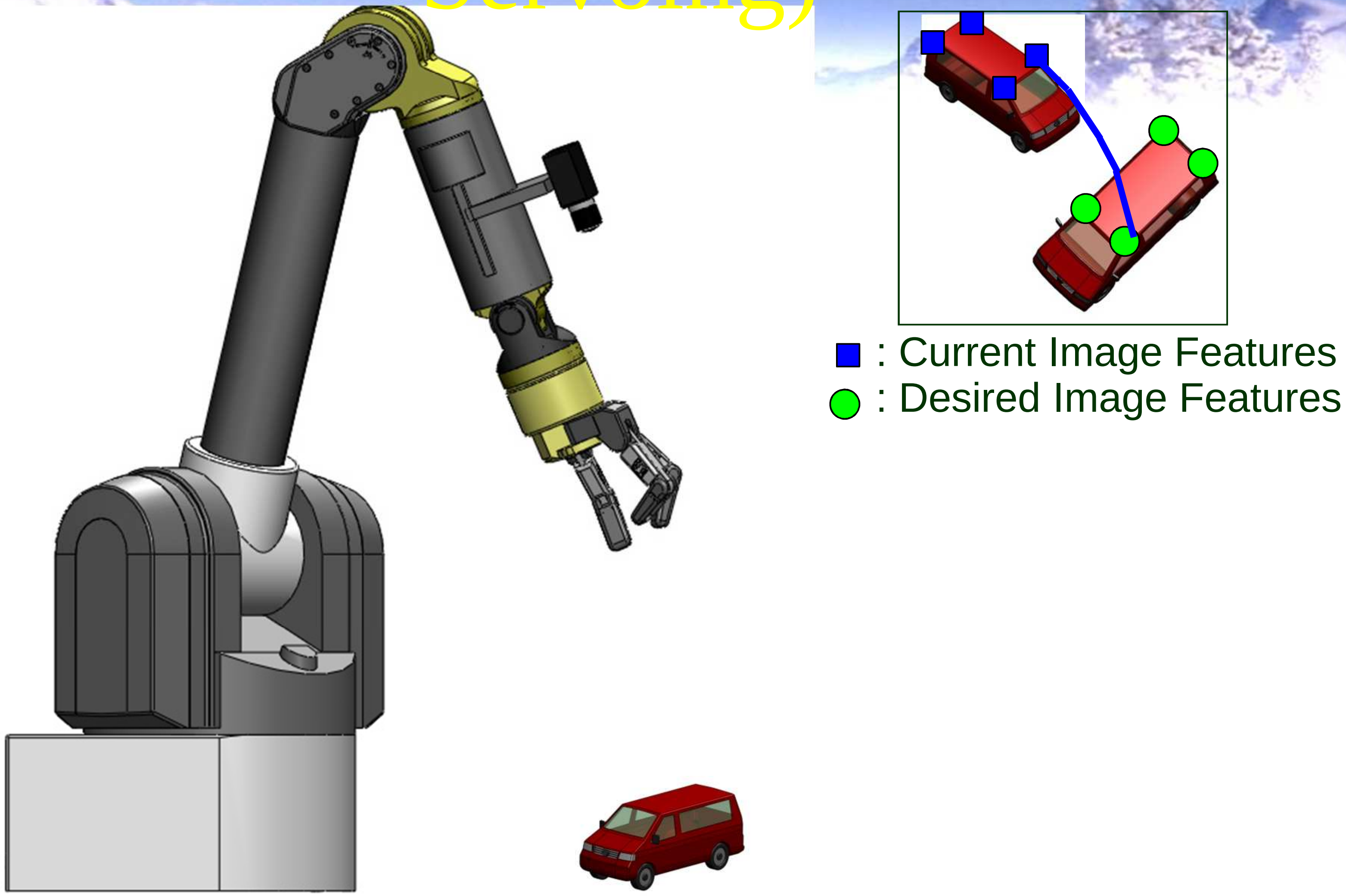




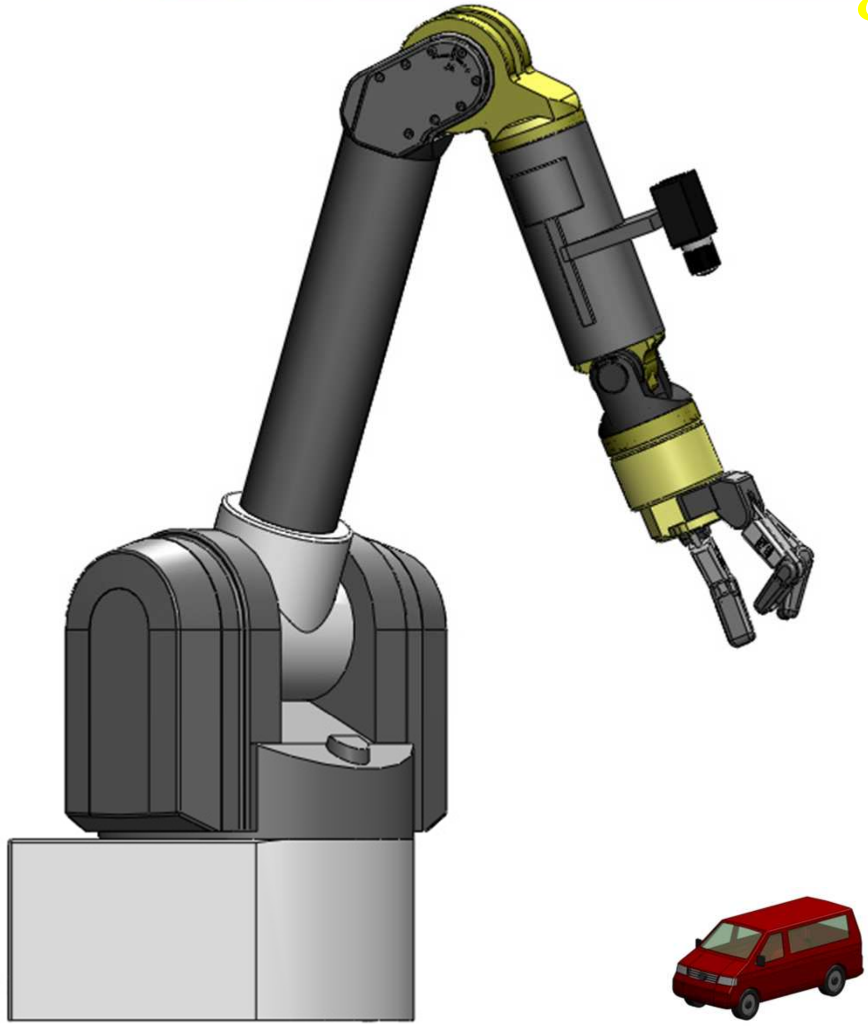
Intro to Image-based Visual Servoing IBVS

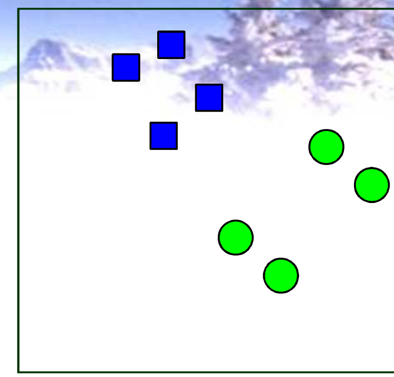


Vision-Based Control (Visual Servoing)



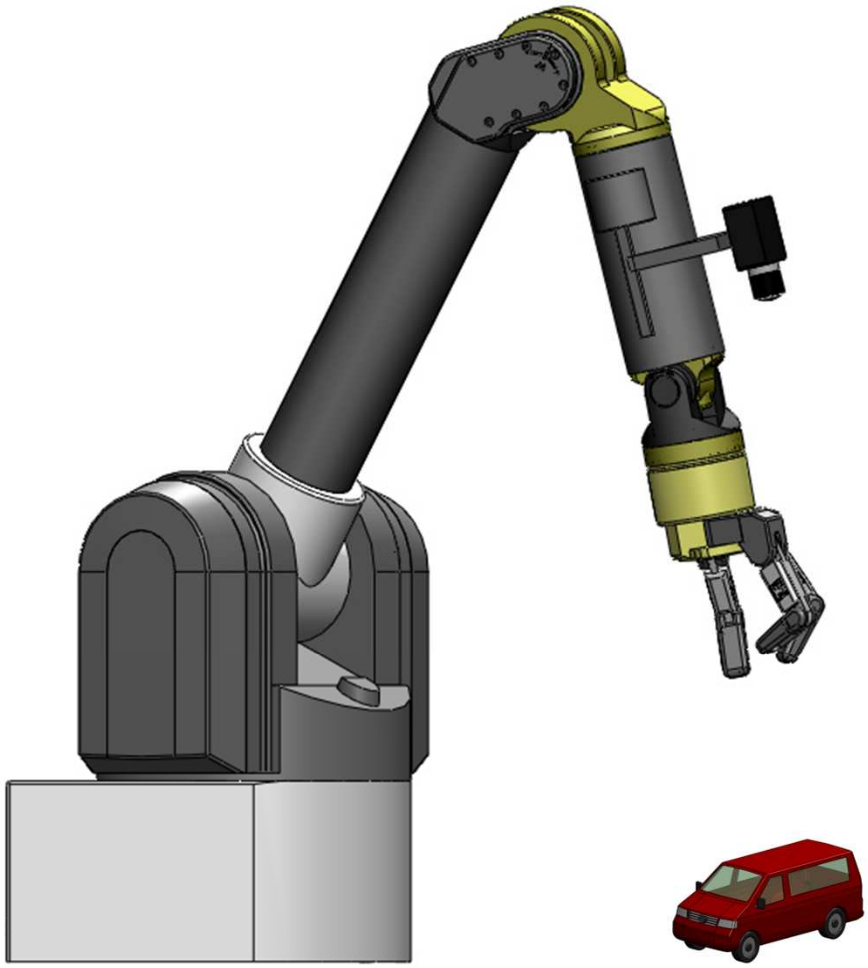
Vision-Based Control (Visual Servoing)

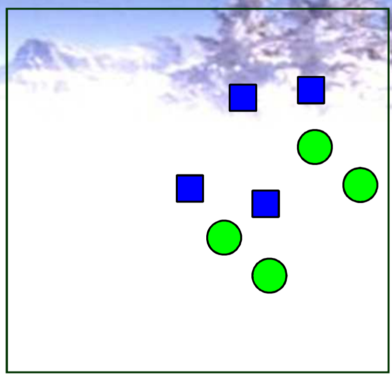




■ : Current Image Features
● : Desired Image Features

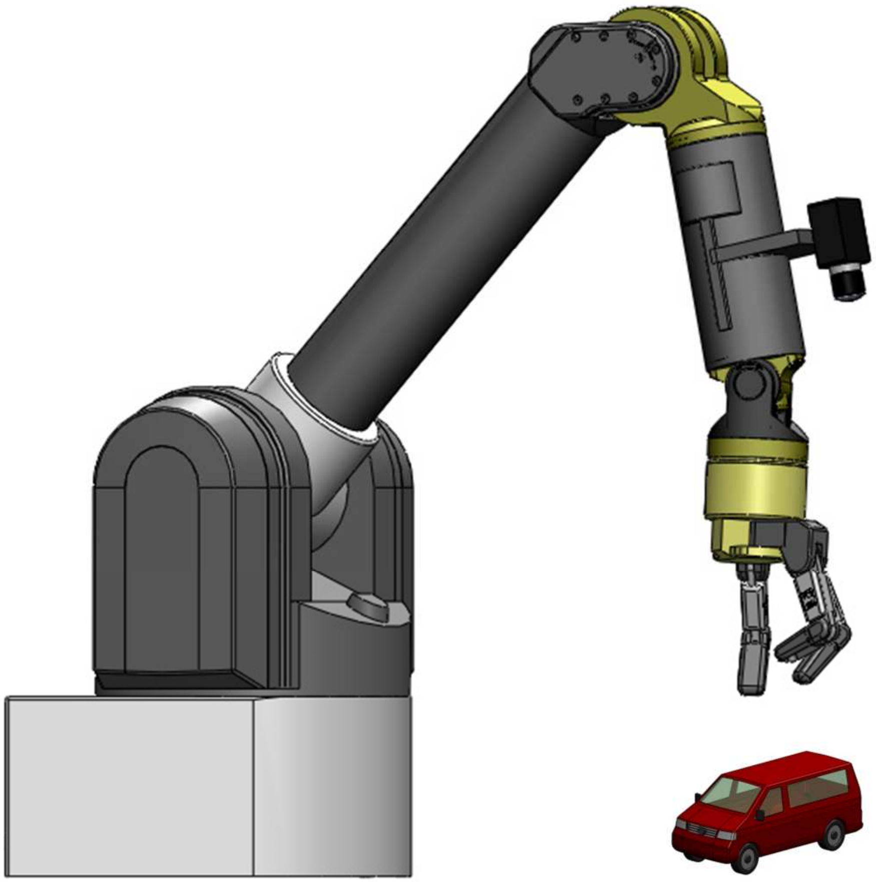
Vision-Based Control (Visual Servoing)

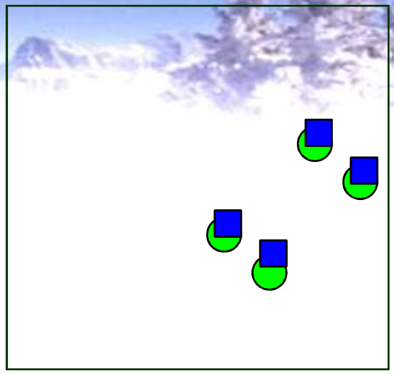




■ : Current Image Features
● : Desired Image Features

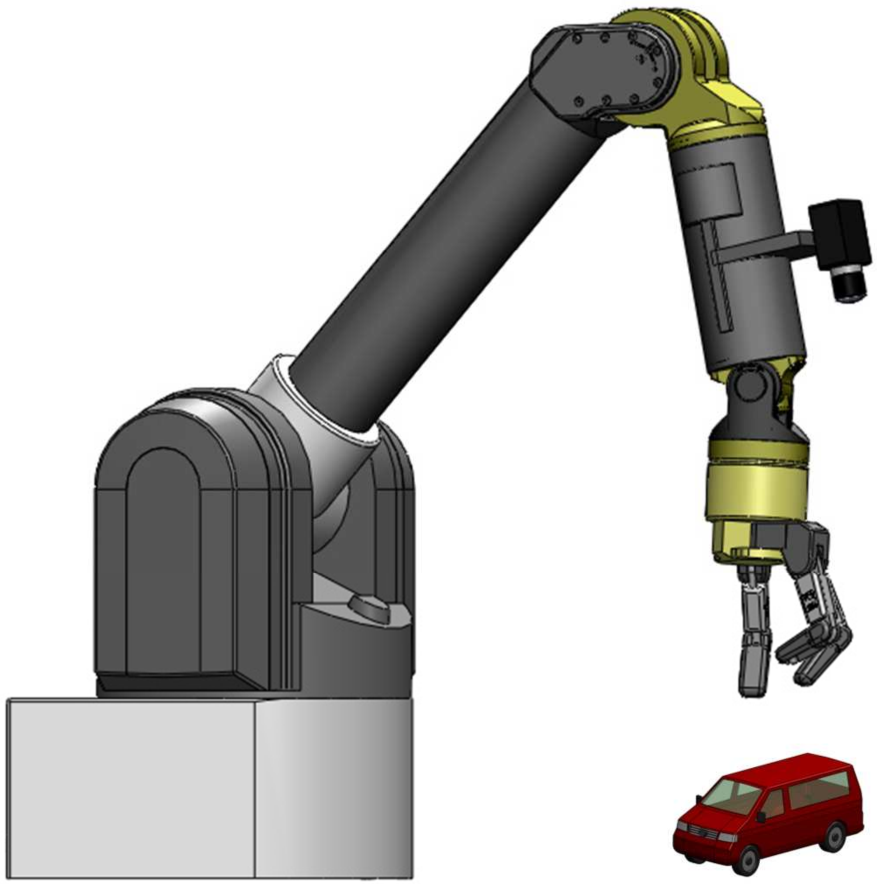
Vision-Based Control (Visual Servoing)

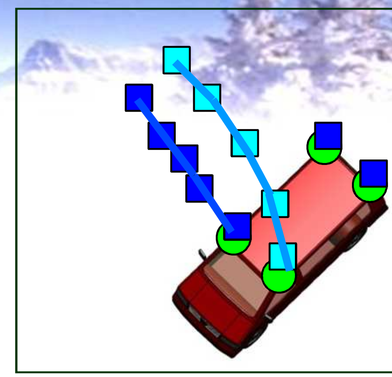




■ : Current Image Features
● : Desired Image Features

Vision-Based Control (Visual Servoing)

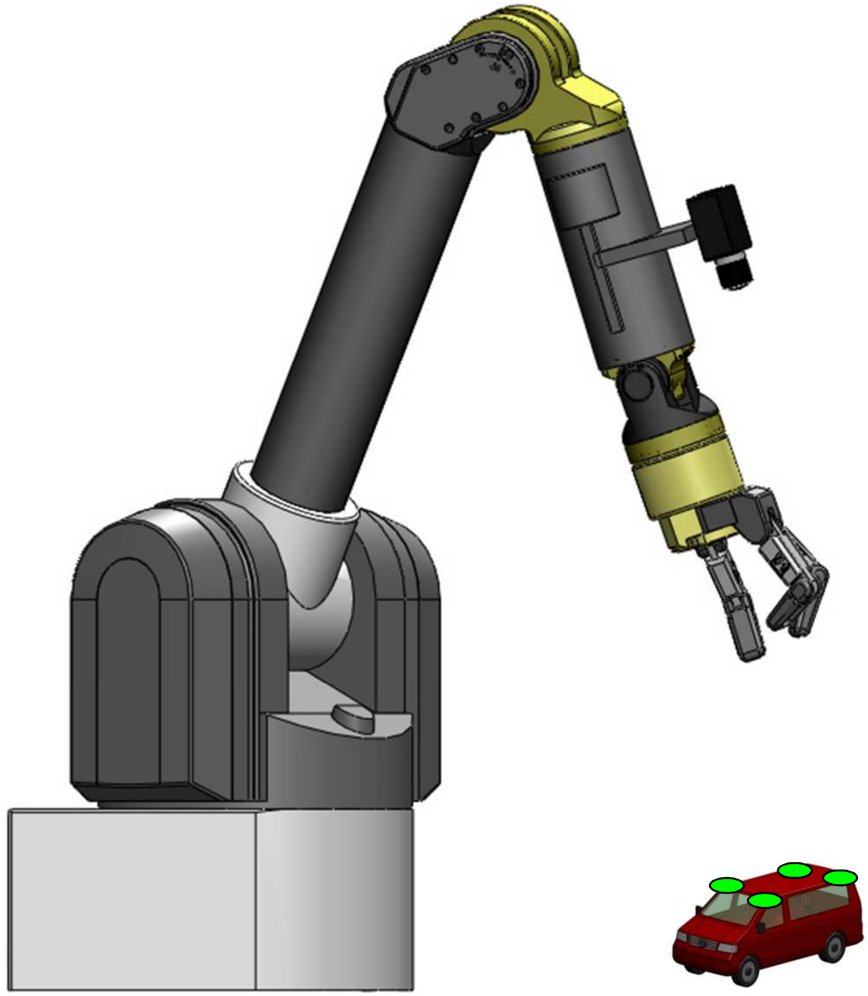


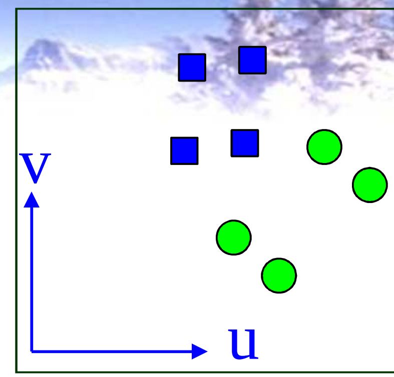


: Current Image Features

: Desired Image Features

u,v Image-Space Error





■ : Current Image Features
● : Desired Image Features

$$\mathbf{E} = [\,\bullet - \blacksquare\,]$$

One point

$$\mathbf{E} = [\mathbf{y}_0 - \mathbf{y}^*]$$

Pixel coord

$$\mathbf{E} = \begin{bmatrix} y_u \\ y_v \end{bmatrix} - \begin{bmatrix} y_u \\ y_v \end{bmatrix}^*$$

Many points

$$\mathbf{E} = \begin{bmatrix} y_1 \\ \vdots \\ y_8 \end{bmatrix}^* - \begin{bmatrix} y_1 \\ \vdots \\ y_8 \end{bmatrix}_0$$

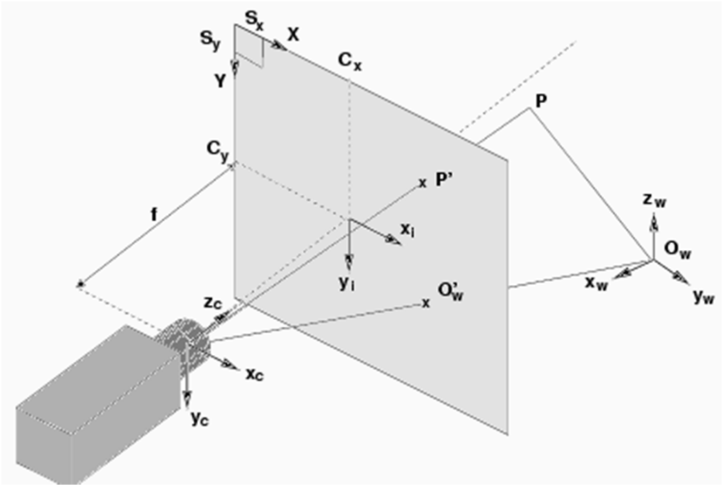
Problem: Lots of coordinate frames to calibrate

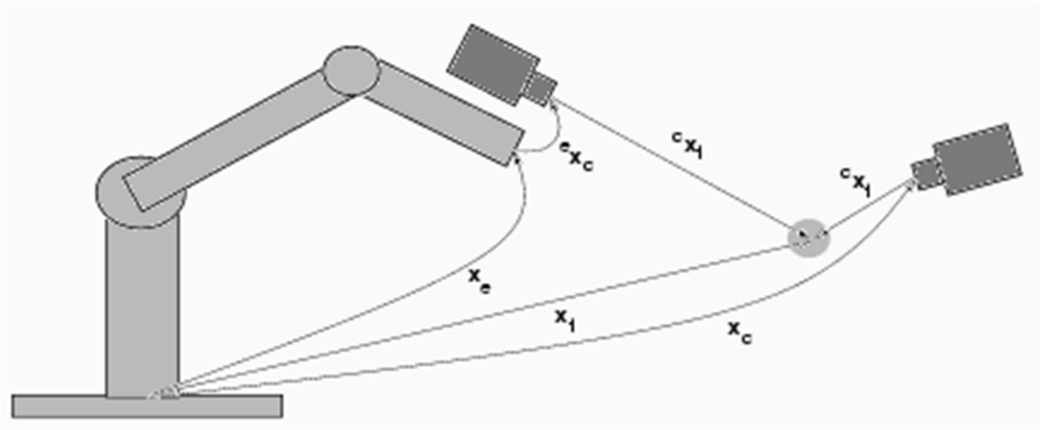
Camera

- Center of projection
- Different models

Robot and scene

- Base frame
- End-effector frame
- Object

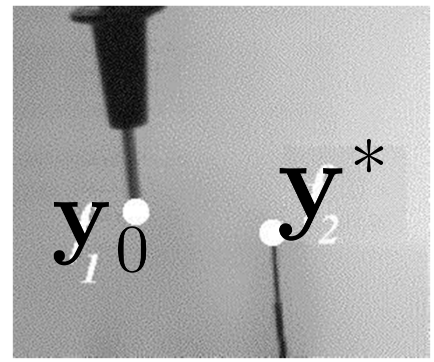




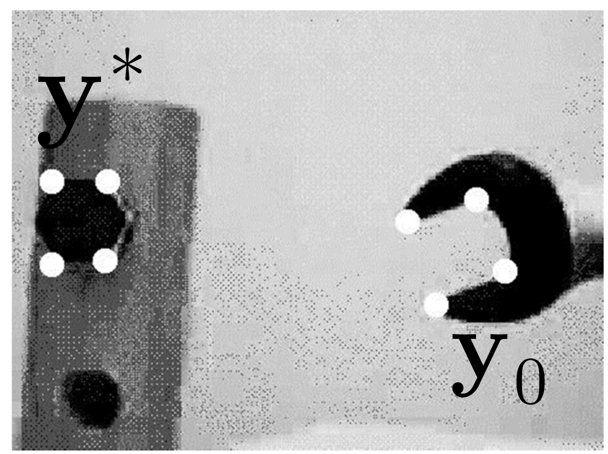
Visual specifications

- Point to Point task “error”:

$$\mathbf{E} = [\mathbf{y}^* - \mathbf{y}_0]$$



$$\mathbf{E} = \begin{bmatrix} y_1 \\ \vdots \\ y_{16} \end{bmatrix}^* - \begin{bmatrix} y_1 \\ \vdots \\ y_{16} \end{bmatrix}_0$$



Why 16 elements?

Visual specifications 2

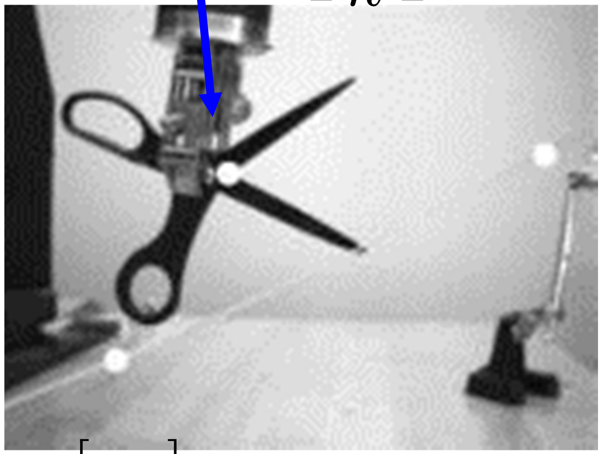
- Point to Line

Note: y's in homogeneous coord.

$$\mathbf{y}_l = \begin{bmatrix} u_l \\ v_l \\ k \end{bmatrix}$$

Line: $$\mathbf{E}_{pl}(\mathbf{y},\mathbf{l}) = \begin{bmatrix} \mathbf{y}_l \cdot \mathbf{l}_l \\ \mathbf{y}_r \cdot \mathbf{l}_r \end{bmatrix}$$

$$\mathbf{l}_l = [\, y_2 \times y_3 \,]_l$$

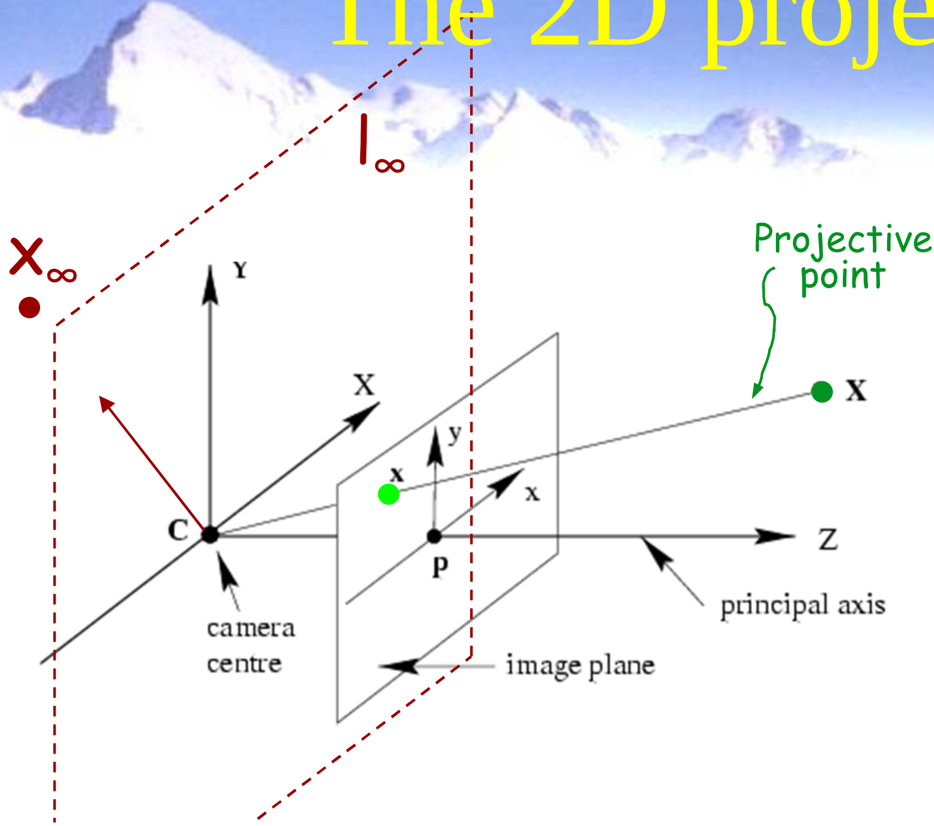


$[y_3]$

$[y_2]$

How to design visual specifications in a principled way?

The 2D projective plane



Homogeneous coordinates

$$\begin{bmatrix}X\\y\\k\end{bmatrix} \equiv s\begin{bmatrix}X\\y\\k\end{bmatrix} \qquad s \neq 0$$

$$\begin{bmatrix}x\\y\end{bmatrix} = \frac{1}{k}\begin{bmatrix}X\\y\end{bmatrix}$$

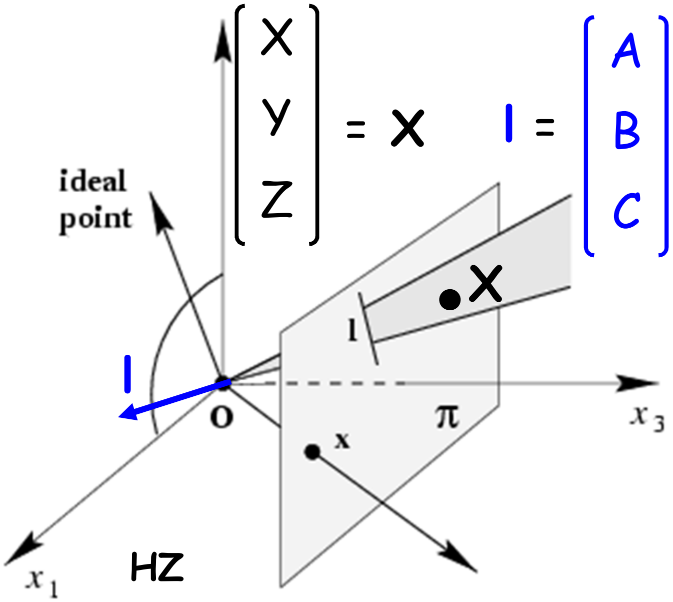
Inhomogeneous equivalent

- 2D projective space models perspective imaging
- Each 3D ray is a point in P^2 : homogeneous coords.
- Ideal points
- P^2 is R^2 plus a “line at infinity” l_∞

$$X_\infty = \begin{bmatrix}X\\y\\0\end{bmatrix}$$

Projective Lines

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{X} \qquad \mathbf{l} = \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$



- Projective line ~ a plane through the origin

$$\mathbf{l}^T\mathbf{X} = \mathbf{X}^T\mathbf{l} = AX + BY + CZ = 0$$

$$\mathbf{X}_\infty = \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix} \qquad \mathbf{l}_\infty = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

"line at infinity"

- Ideal line ~ the plane parallel to the image

<u>Duality</u>: For any 2d projective property, a dual property holds when the role of points and lines are interchanged.

$$\mathbf{l} = \mathbf{X}_1 \times \mathbf{X}_2 \qquad\qquad \mathbf{X} = \mathbf{l}_1 \times \mathbf{l}_2$$

The line joining two points

The point joining two lines

Projective transformations

- Homographies, collineations, projectivities
- 3x3 nonsingular H

  maps $\boldsymbol{P}^2$ to $\boldsymbol{P}^2$
  8 degrees of freedom
  determined by 4 corresponding points

$$\begin{pmatrix} x'_1 \\ x'_2 \\ x'_3 \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$x' = Hx$$

- Transforming Lines?

  subspaces preserved $\mathbf{x}^T\mathbf{l} = 0 \qquad \mathbf{x}'^T\mathbf{l}' = 0$

  substitution $x^T H^T \mathbf{l}' = 0$

  dual transformation $\mathbf{l}' = H^{-T}\mathbf{l}$

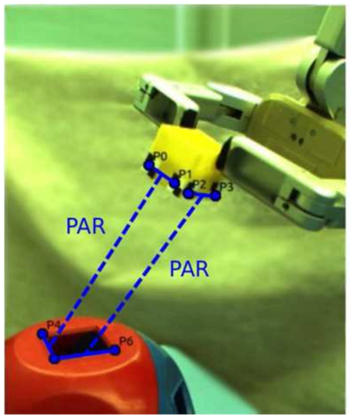
Homographies a generalization of affine and Euclidean transforms

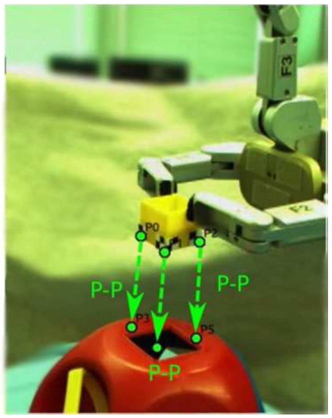
Group	Transformation	Invariants	Distortion
Projective 8 DOF	$H_P = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix}$	• Cross ratio • Intersection • Tangency	
Affine 6 DOF	$H_A = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	• Parallelism • Relative dist in 1d • Line at infinity $\mathbf{l}_\infty$	
Metric 4 DOF	$H_S = \begin{bmatrix} sR & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	• Relative distances • Angles • Dual conic C^*_∞	
Euclidean 3 DOF	$H_E = \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	• Lengths • Areas	

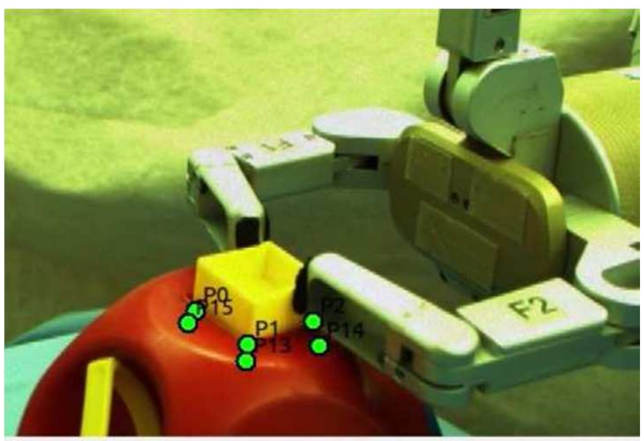
2 dof $\mathbf{l}_\infty$

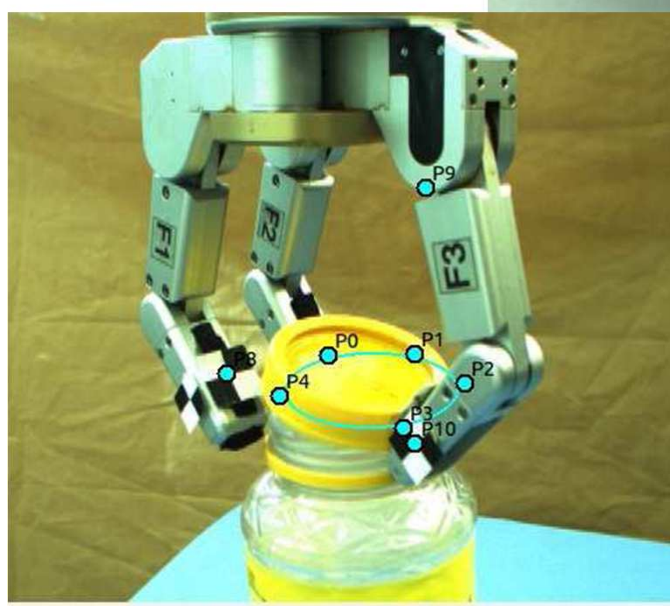
2 dof C^*_∞

How to define a visual task?





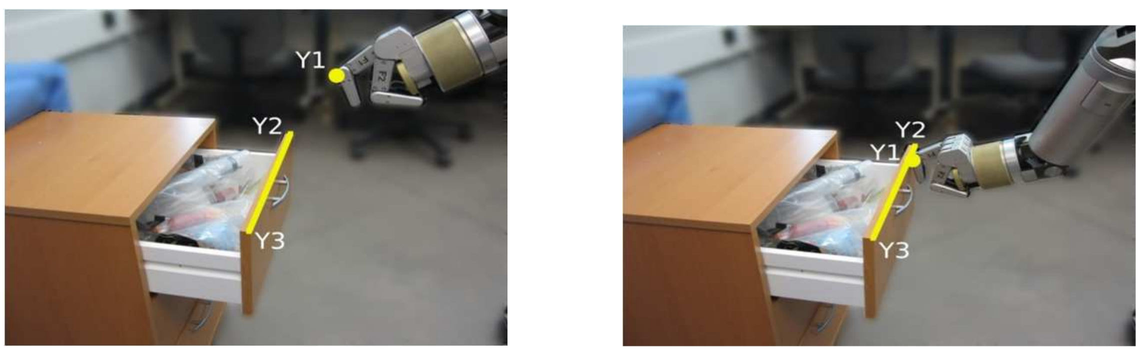




Visually defined alignment: basic

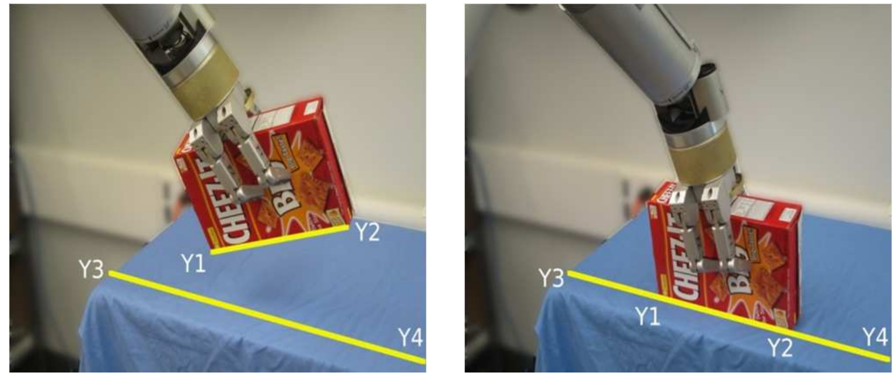
point-to-point: $e_{pp}(y) = y_2 - y_1$

or (homogenous coord) $e_{pp}(y) = y_2 \times y_1$

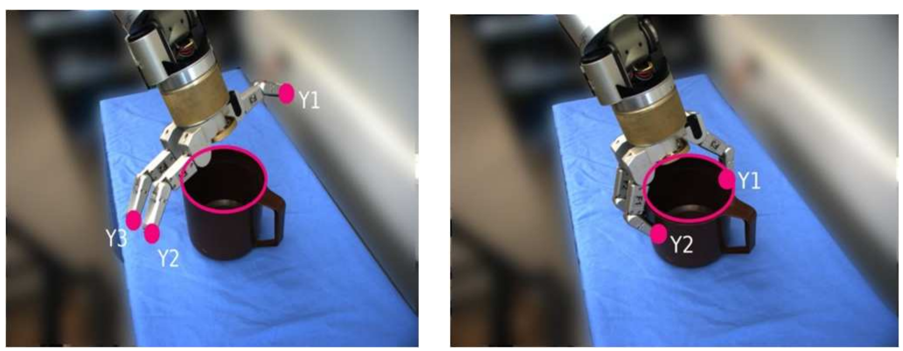


point-to-line: $e_{pl}(y) = y_1 \cdot (y_2 \times y_3)$

Some more visual alignments



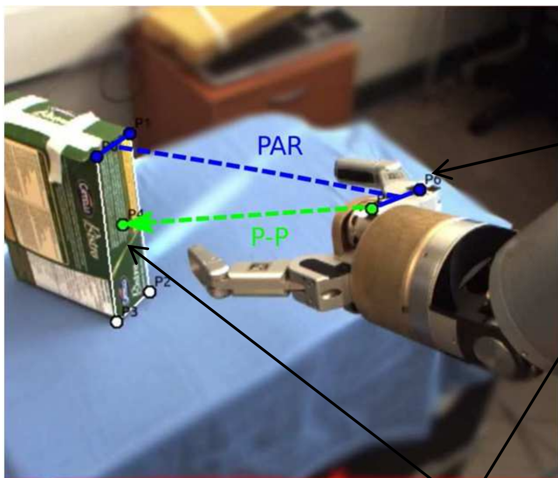
parallel lines: $e_{par}(l) = (l_1 \text{ x } l_2) \text{ x } (l_3 \text{ x } l_4)$ line-to-line: $e_{ll}(y) = y_1 \cdot (y_3 \text{ x } y_4) + y_2 \cdot (y_3 \text{ x } y_4)$

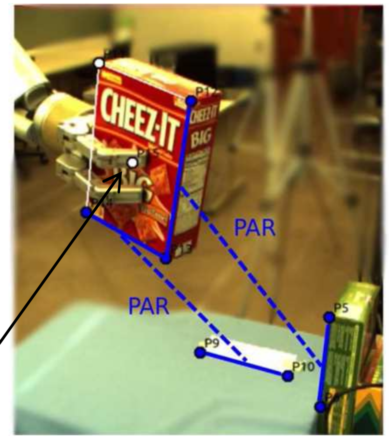


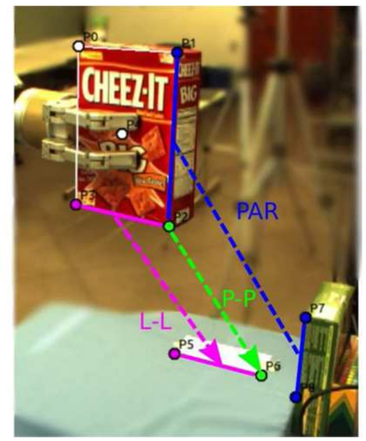
point-to-ellipse: $e_{pe}(y) = y^T{}_1 \, C_{ellipse} \, y_1$

Now: Language for visual alignments What else do we need?

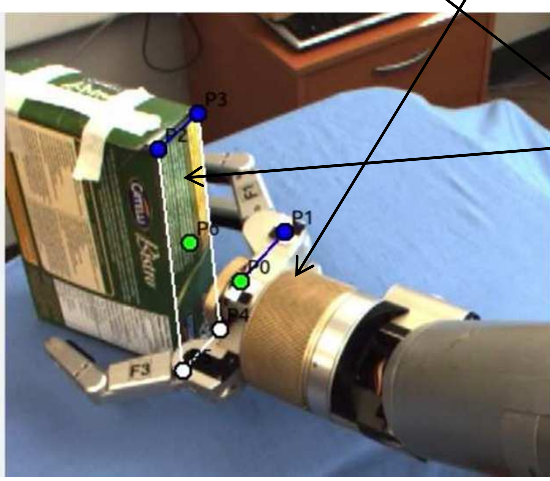
Need: 1. Some way of entering alignments in images
2. video tracking to perform servoing!



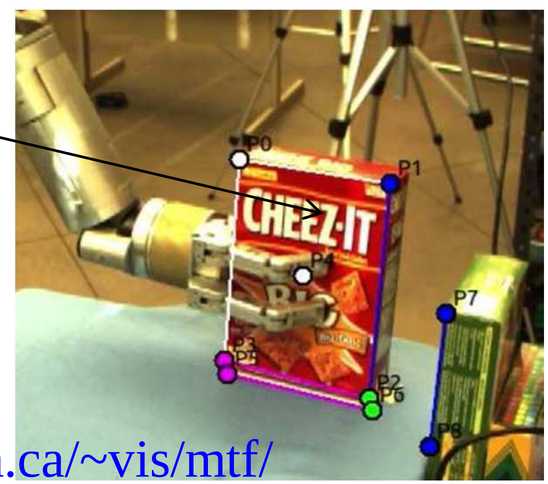




Feature trackers

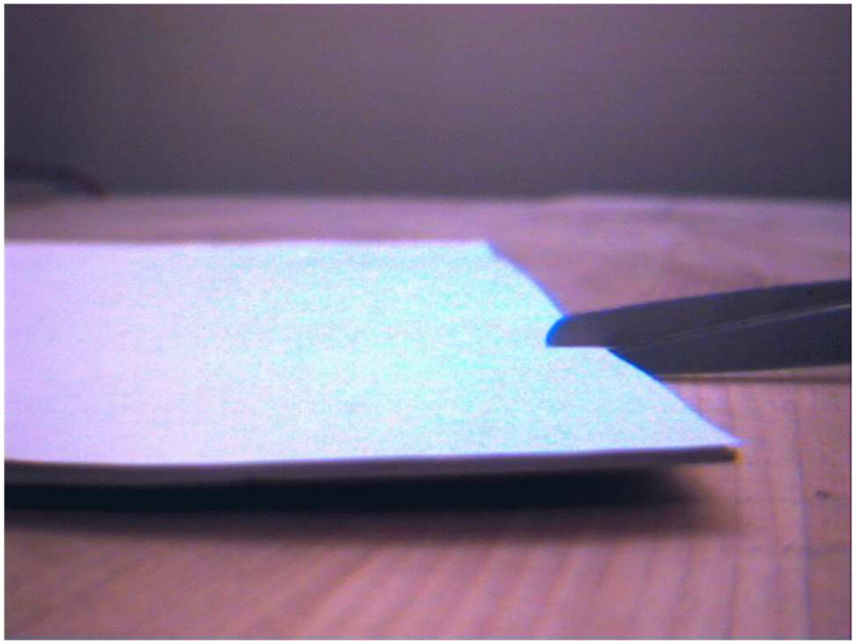


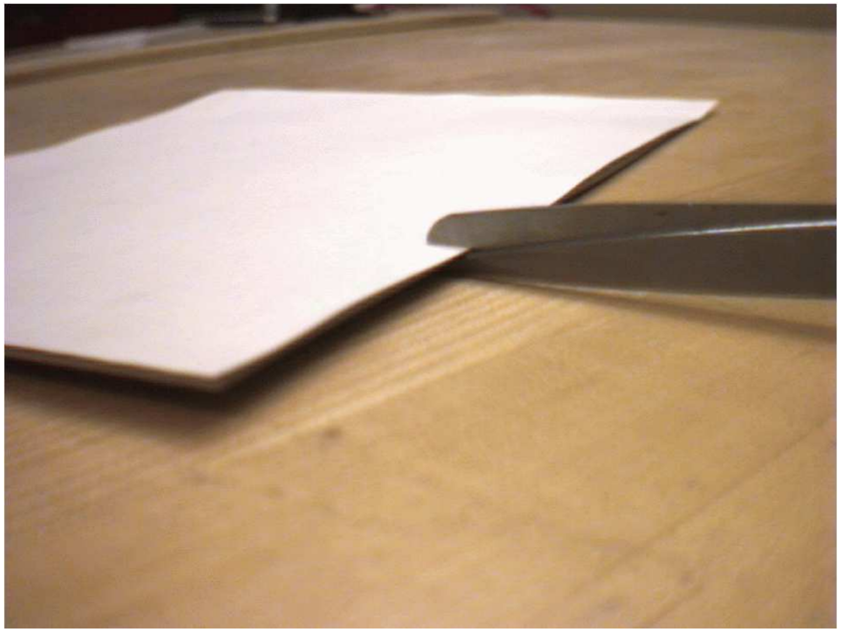
Registration trackers:



Download:
http://webdocs.cs.ualberta.ca/~vis/mtf/

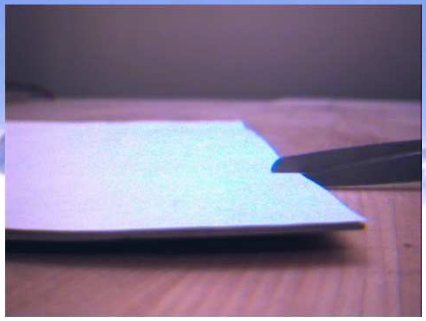
Visual ambiguity





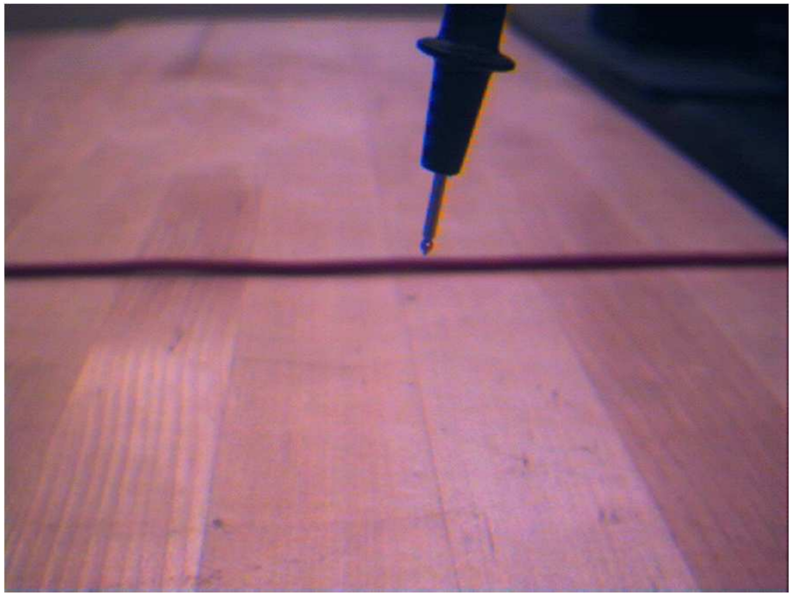
- Will the scissors cut the paper in the middle?

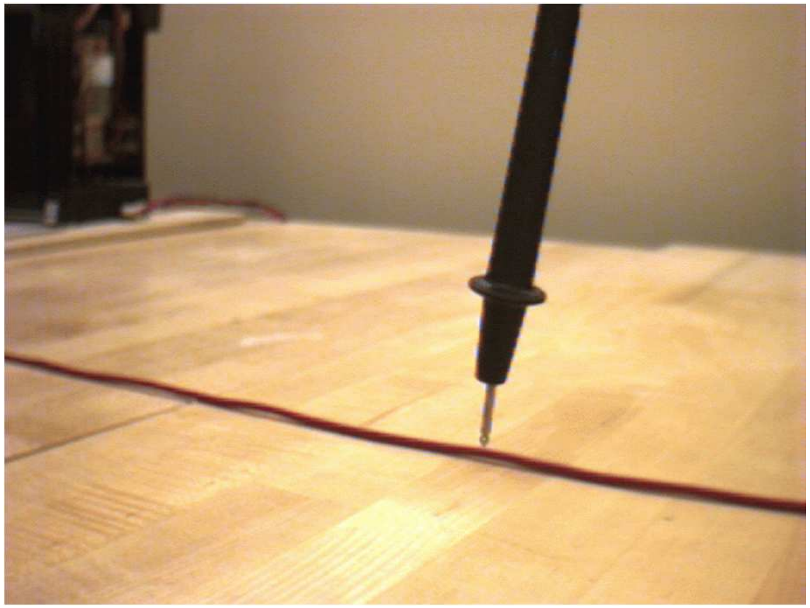
ambiguity



- Will the scissors cut the paper in the middle? **NO!**

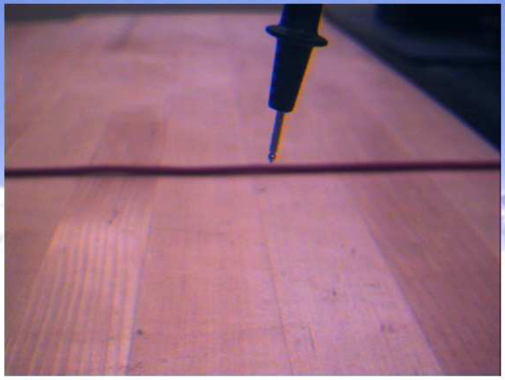
Task Ambiguity

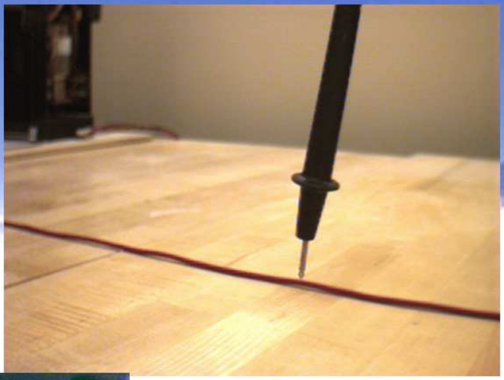


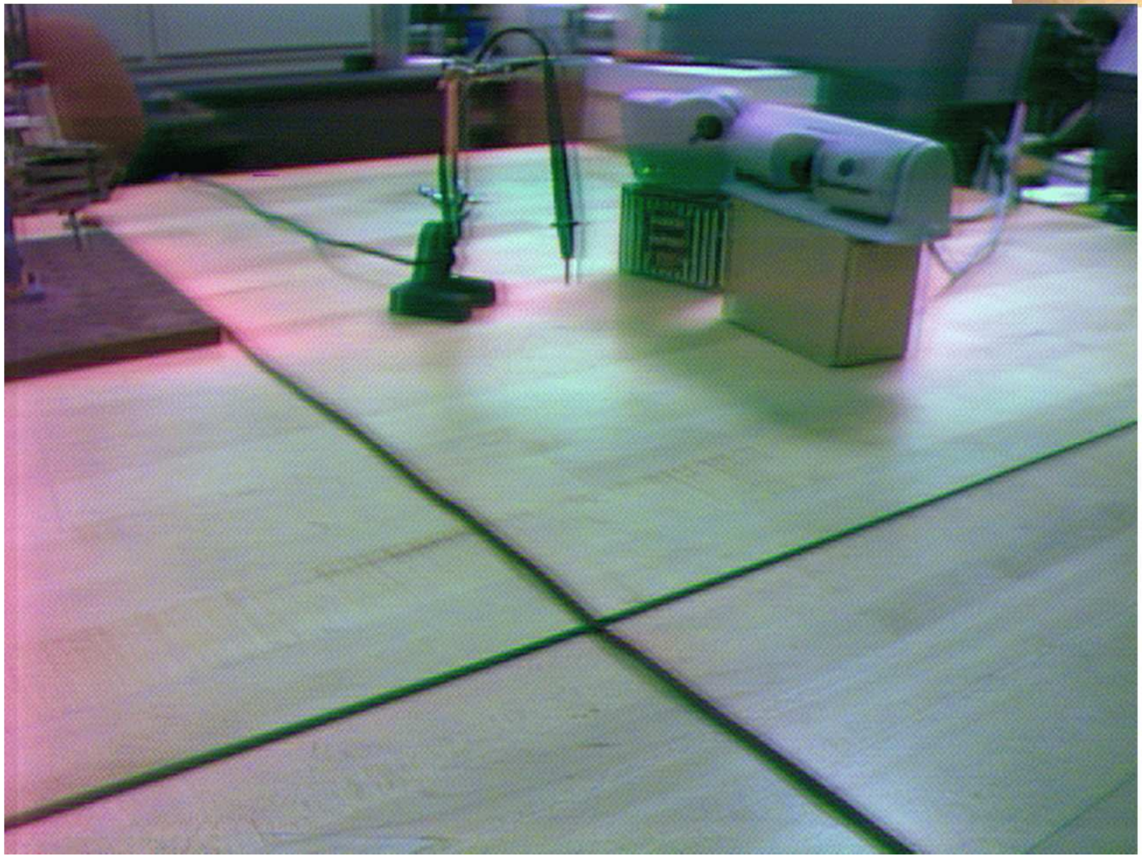


- Is the probe contacting the wire?

Task Ambiguity







- Is the probe contacting the wire? **NO!**

Solve the cut in the middle task?

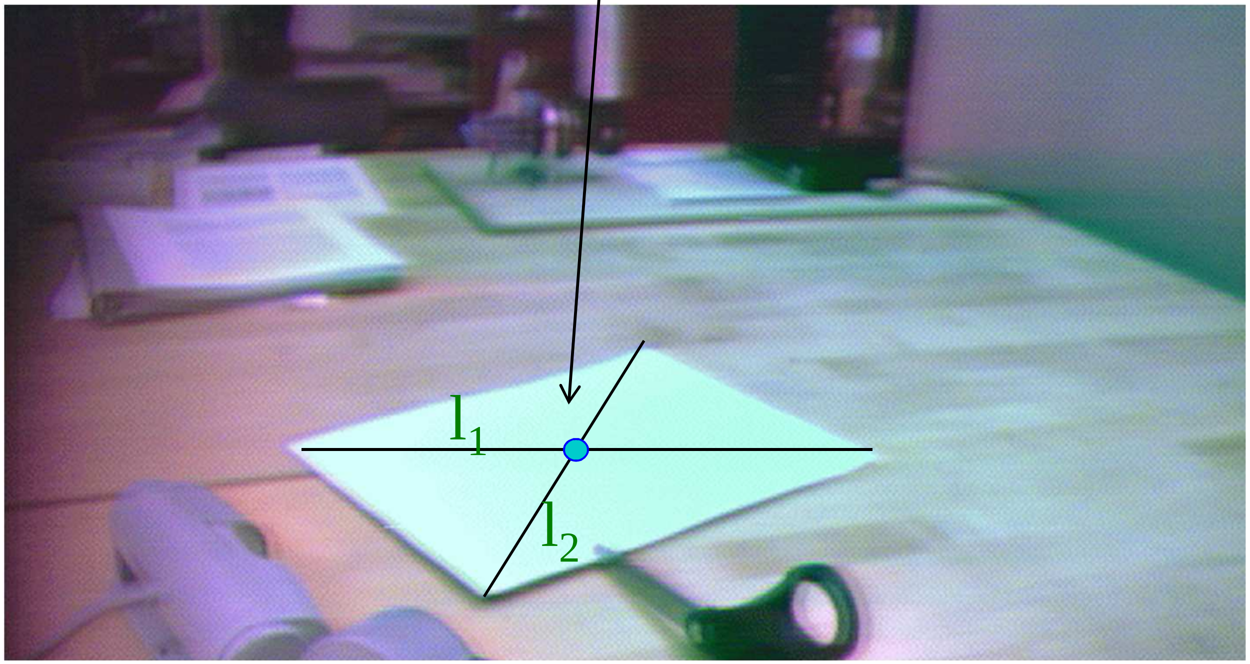
- Compute paper midpoint. How?

Solve the cut in the middle task?

- Compute paper midpoint.

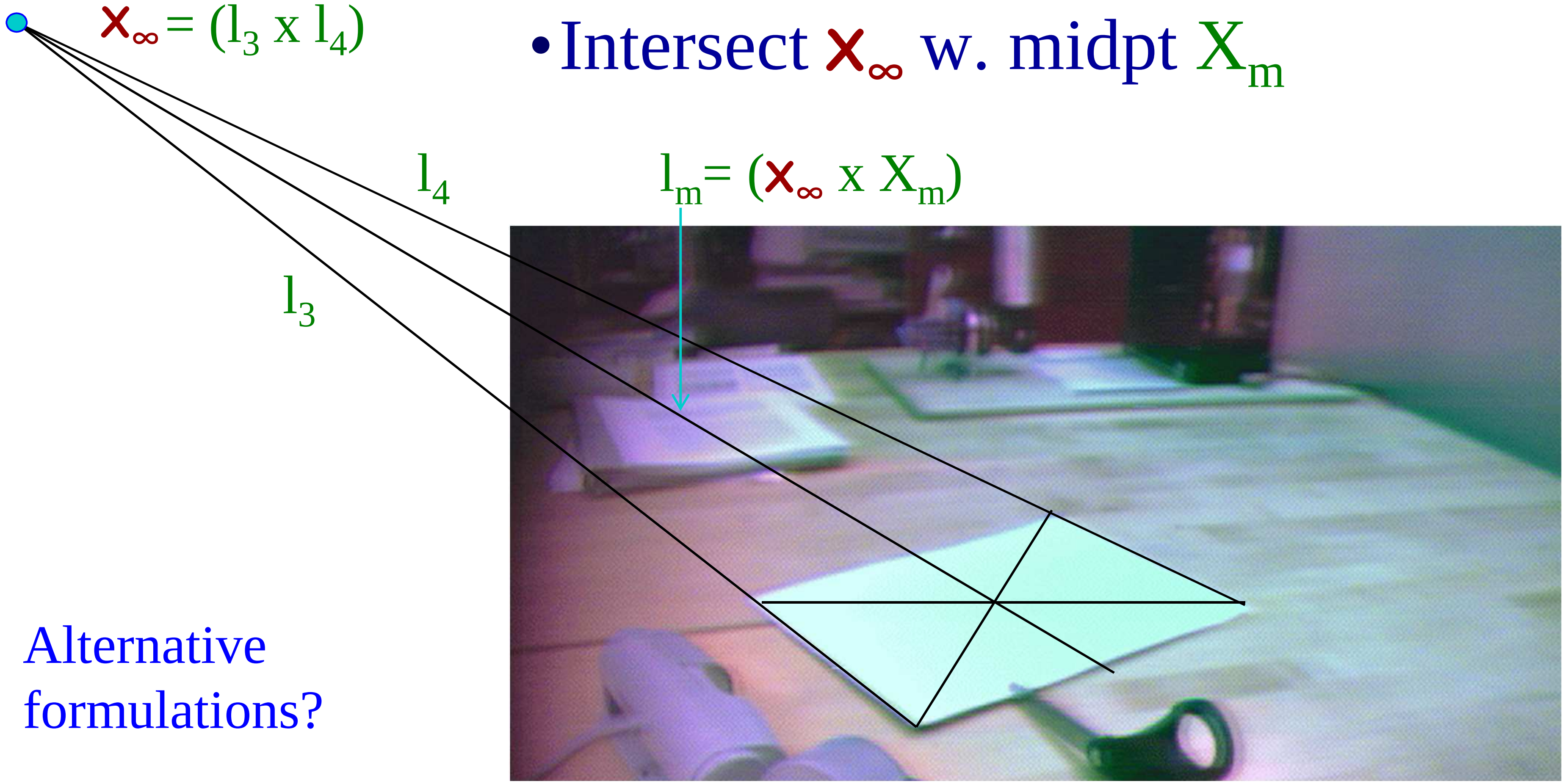
(Are we done yet?)

$x_m = (l_1 \times l_2)$



Solve the cut in the middle task?

- Compute vanishing point x_∞,
- Intersect x_∞ w. midpt X_m

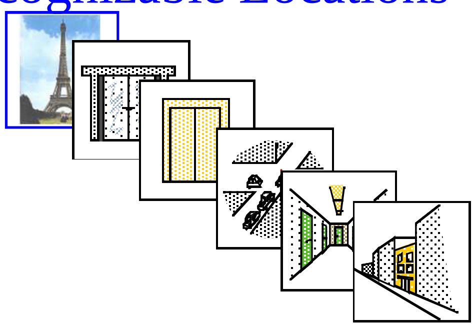


Alternative formulations?

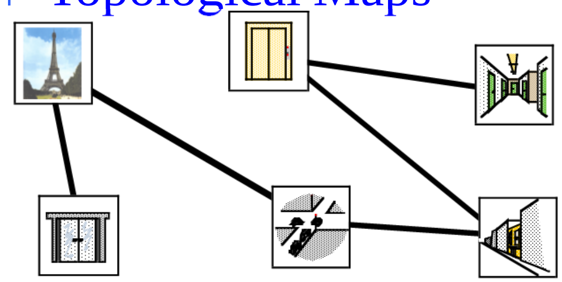
What information do we use to move?

Mobile robot navigation: From appearance to metric SLAM

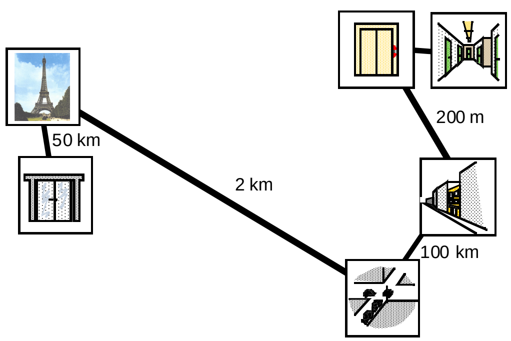
- Recognizable Locations



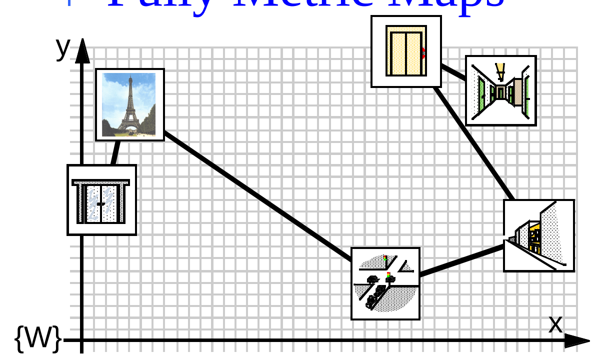
- Topological Maps



- Metric Topological Maps



- Fully Metric Maps



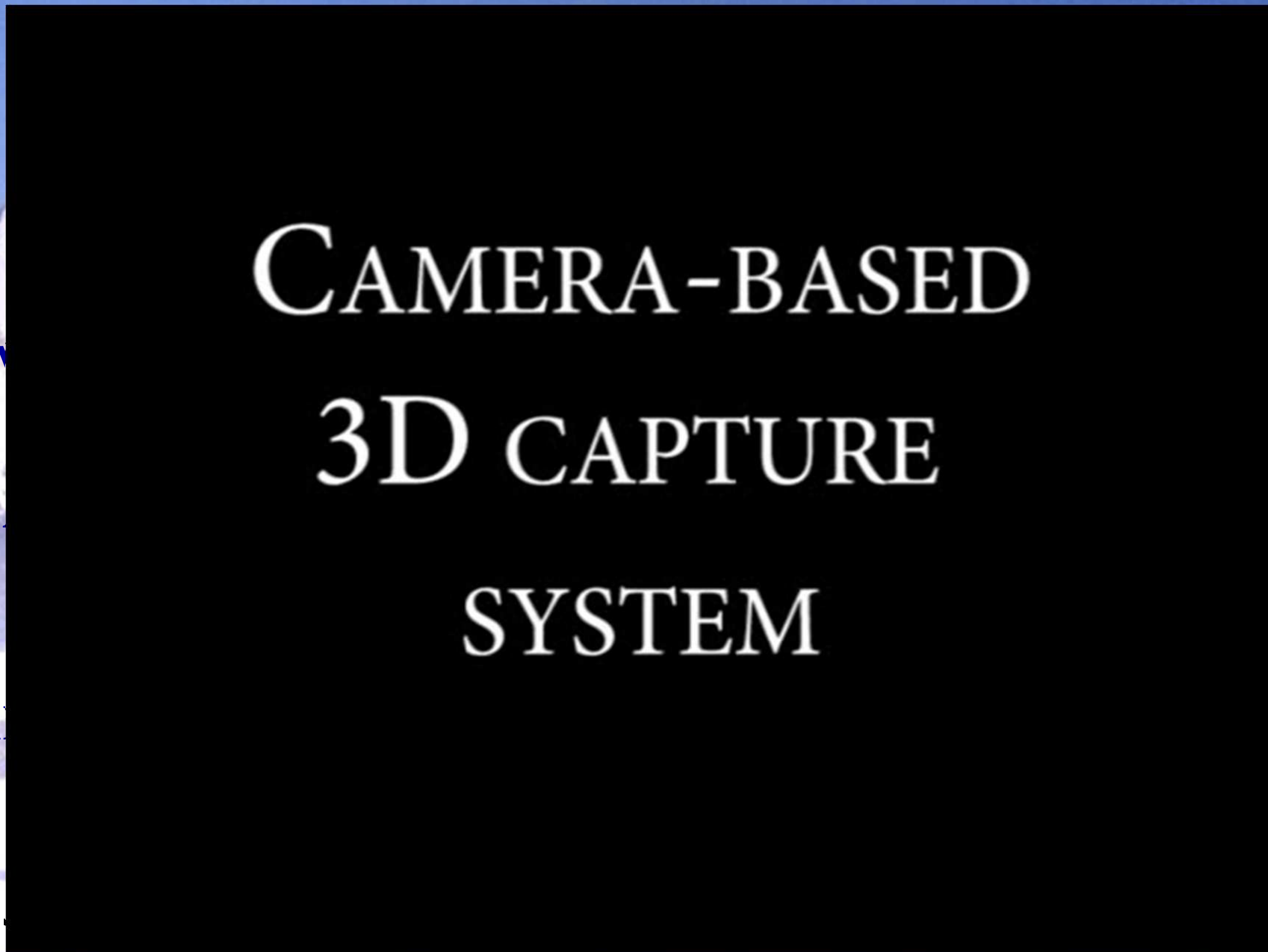
Courtesy
K. Arras

How about other applications?

Beyond projective camera vision and screen GUI: Pointing

The human by pointing instructs the robot which ingredients are to be placed on the pizza

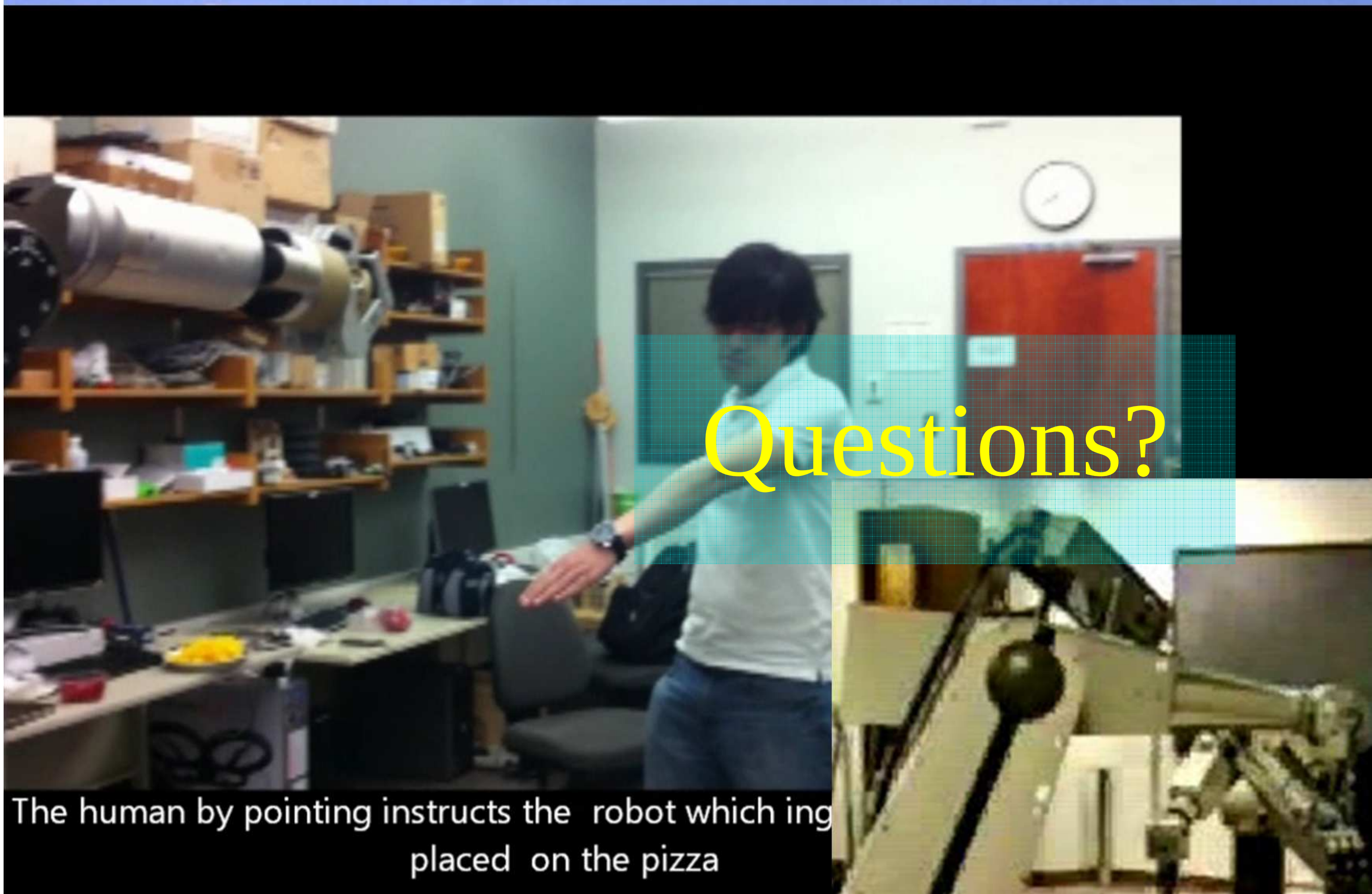
Questions?



- Dow
- Capt
- Mai
- More papers: www.cs.ualberta.ca/~jag

Making Pizza with my robot

3rd Prize ICRA'15 Video competition



Questions?

Change a Lightbulb

ICRA'97 Video