Perception for Robots

I. BASICS

Thanks to Peter Corke for the use of some slides
What is a robot

• For the purposes of this class a robot is a goal oriented machine that can sense, reason and act.
BASIC QUESTIONS

• Where am I
• Where are you
• What are you
• How do I get there
• How to achieve a task
Where am I

• Why not use GPS?

GPS is not perfect and has severe limitations in environments where robots are needed:
--- cities, mines, industrial sites, underwater, deep forest.

It only tells where I am
Urban Canyon Problem
Industrial sites, mines
Underwater, deep forest
Humans and animals have a number of senses

- Sight
- Hearing
- Touch
- Smell
- Taste
- Balance

- Echolocation: bats, electric fields: sharks, compass: birds
Vision

www.nvrl.org - Illustration based upon information from National Eye Institute / National Institutes of Health
Hearing

**How Hearing Works?**

Excessive noise exposure is one of the leading causes of hearing loss. The tiny hair cells in the inner ear are easily damaged by loud noise and once you lose them, they never grow back!

- **Outer ear**
  - Pinna
  - Auditory canal

- **Middle ear**
  - Ear drum
  - Hammer, anvil, stirrup
  - Oval window

- **Inner ear**
  - Cochlear
  - Basilar membrane
  - Organ of Corti

**Pre-processing**

- **Spectral analysis**
- **Pattern processing**

Echolocation of bats
Electric field sensing
Magnetic field sensing

Navigating With A Built-In Compass

Researchers are finding creatures in the way grounds navigate using the Earth's magnetic fields.
Vision: most powerful sense

- Essential for survival: finding food, avoiding being food, finding mates
- Long range sensing: beyond our fingertip (vision is our way to touch the world)
Evolution of the eye
\(\frac{1}{2}\) billion years
Climbing mount improbable
10 different designs
A plethora of eyes
Complex Eyes

- **Cup Eye**: detects intensity and direction of light.
- **Compound Eye**: captures detailed images.
- **Camera Eye**: captures detailed images.
Compound eyes
Two kinds of eyes at the top: Camera type or planar Spherical

aperture (virtual camera origin, ≈ eye)

virtual camera (image plane in front)
Many cameras in the market
Catadioptric – panoramic images
How does Vision work?

• Ancient Greeks: Extramission Theory
Descartes got it right
Many theories over the centuries

• The Gestaltists

• Von Helmholtz: Unconscious inference

• David Marr: A reconstruction process that tells us where is what.
Theories influenced by the zeitgeist
Animal perception is active

Free examination.

Estimate material circumstances of the family

Give the ages of the people.

Surmise what the family had been doing before the arrival of the unexpected visitor.

Remember the clothes worn by the people.

3 min. recordings of the same subject

Remember positions of people and objects in the room.

Estimate how long the visitor had been away from the family.
Measuring eye movements
Robots with Vision
PR2 Humanoid
Perception for Robots
3 major problems

• Reconstruction

• Reorganization

• Recognition
Reconstruction
Reorganization: segmentation
Reorganization: flow
Recognition
Images and Videos Contain

- Lines (contours, edges)
- Intensity and Color
- Texture
- movement
Lines
Color, Texture
Motion
Contents of the Class

• **Image Processing**: Images, Light and Color, Filtering, Noise, Convolution, Edge detection, contour finding, texture analysis, segmentation and grouping.

• **3D Geometry**: Stereo, Multiple View Geometry, Epipolar Geometry, Projective Geometry

• **Motion**: Optical Flow, Egomotion, Motion Segmentation, Tracking

• **Navigation**: Map making, SLAM
Break
A theoretical model of an eye

• Pick a point in space and the light rays passing through
Then cut the rays with a plane

- This gives an image
Pinhole cameras

- Abstract camera model - box with a small hole in it
- Pinhole cameras work in practice

(Forsyth & Ponce)
If we change the plane, we get a new image.
How are these images related? (what remains invariant?)
Conics
Projection of circle
Vanishing points

• **Vanishing point**
  – projection of a point at infinity
Vanishing points (2D)

- Image plane
- Vanishing point
- Camera center
- Line on ground plane
Vanishing points

- Properties
  - Any two parallel lines have the same vanishing point \( v \)
  - The ray from \( C \) through \( v \) is parallel to the lines
  - An image may have more than one vanishing point
Parallelism (angles) not invariant
Cross ratio = only invariant

\[
\frac{(CA / CB)}{(DA / DB)} = \frac{(C'A' / C'B')}{(D'A' / D'B')}
\]
Remember that the area of a triangle is 1/2 the base times the height. It is also the product of two sides times the side of the angle between them. Using this, we get:

\[
\begin{align*}
\text{Area}(pAC) &= \frac{h}{2} \langle AC \rangle = \frac{1}{2} \langle pA \rangle \langle pC \rangle \sin(ApC) \\
\text{Area}(pBC) &= \frac{h}{2} \langle BC \rangle = \frac{1}{2} \langle pB \rangle \langle pC \rangle \sin(BpC) \\
\text{Area}(pAD) &= \frac{h}{2} \langle AD \rangle = \frac{1}{2} \langle pA \rangle \langle pD \rangle \sin(ApD) \\
\text{Area}(pBD) &= \frac{h}{2} \langle BD \rangle = \frac{1}{2} \langle pB \rangle \langle pD \rangle \sin(BpD)
\end{align*}
\]

Thus the cross ratio of \( A, B, C, D = \frac{\langle AC \rangle / \langle BC \rangle}{\langle AD \rangle / \langle BD \rangle} \)

\[
= \frac{\langle pA \rangle \langle pC \rangle \sin(ApC) / \langle pB \rangle \langle pC \rangle \sin(BpC)}{\langle pA \rangle \langle pD \rangle \sin(ApD) / \langle pB \rangle \langle pD \rangle \sin(BpD)}
\]

\[
= \frac{\sin(ApC) / \sin(BpC)}{\sin(ApD) / \sin(BpD)}
\]

This last quantity is independent of the line we project to. Thus cross ratios are invariant under projection.

This discussion is based on Courant and Robbins, "What Is Mathematics."
Back to our question: how are the 2 images related to each other

Can we find a map, a function mapping $x'$ to $x$?
**Fundamental Theorem:** If we know how 4 points map to each other in the two planes, then we know how all points map. (if \( a \rightarrow A, b \rightarrow B, c \rightarrow C, d \rightarrow D \), then we can map any point.)
Proof
The projective plane

• Why do we need homogeneous coordinates?
  – represent points at infinity, homographies, perspective projection, multi-view relationships

• What is the geometric intuition?
  – a point in the image is a ray in projective space

• Each point \((x, y)\) on the plane is represented by a ray \((sx, sy, s)\)
  – all points on the ray are equivalent: \((x, y, 1) \equiv (sx, sy, s)\)
Projective lines

• What does a line in the image correspond to in projective space?

• A line is a **plane** of rays through origin
  – all rays \( (x,y,z) \) satisfying: \( ax + by + cz = 0 \)

  \[
  0 = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}
  \]

• A line is also represented as a homogeneous 3-vector \( l \)
Point and line duality
– A line $l$ is a homogeneous 3-vector
– It is $\perp$ to every point (ray) $p$ on the line: $l \cdot p = 0$

What is the line $l$ spanned by rays $p_1$ and $p_2$?
• $l$ is $\perp$ to $p_1$ and $p_2$ $\Rightarrow$ $l = p_1 \times p_2$
• $l$ is the plane normal

What is the intersection of two lines $l_1$ and $l_2$?
• $p$ is $\perp$ to $l_1$ and $l_2$ $\Rightarrow$ $p = l_1 \times l_2$

Points and lines are dual in projective space
• given any formula, can switch the meanings of points and lines to get another formula
Ideal points and lines

- Ideal point ("point at infinity")
  - $p \cong (x, y, 0)$ – parallel to image plane
  - It has infinite image coordinates

Ideal line
- $l \cong (a, b, 0)$ – parallel to image plane
- Corresponds to a line in the image (finite coordinates)
Fundamental Theorem (homography or collineation)

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

or \( x' = Hx \), where \( H \) is a 3 \( \times \) 3 non-singular homogeneous matrix.
## Special Projectivities

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Invariants</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Projectivity</strong></td>
<td>Collinearity, Cross-ratios</td>
</tr>
<tr>
<td>8 dof</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Affine transform</strong></td>
<td>Parallelism, Ratios of areas, Length ratios</td>
</tr>
<tr>
<td>6 dof</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Similarity</strong></td>
<td>Angles, Length ratios</td>
</tr>
<tr>
<td>4 dof</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Euclidean transform</strong></td>
<td>Angles, Lengths, Areas</td>
</tr>
<tr>
<td>3 dof</td>
<td></td>
</tr>
</tbody>
</table>

Projective Geometry

**Projectivity**

\[
\begin{bmatrix}
  h_{11} & h_{12} & h_{13} \\
  h_{21} & h_{22} & h_{23} \\
  h_{31} & h_{32} & h_{33}
\end{bmatrix}
\]

**Affine transform**

\[
\begin{bmatrix}
  a_{11} & a_{12} & t_x \\
  a_{21} & a_{22} & t_x \\
  0 & 0 & 1
\end{bmatrix}
\]

**Similarity**

\[
\begin{bmatrix}
  sr_{11} & sr_{12} & t_x \\
  sr_{21} & sr_{22} & t_y \\
  0 & 0 & 1
\end{bmatrix}
\]

**Euclidean transform**

\[
\begin{bmatrix}
  r_{11} & r_{12} & t_x \\
  r_{21} & r_{22} & t_y \\
  0 & 0 & 1
\end{bmatrix}
\]
Examples of Projective Transformations

- Central projection maps planar scene points to image plane by a projectivity
  - True because all points on a scene line are mapped to points on its image line
- The image of the same planar scene from a second camera can be obtained from the image from the first camera by a projectivity
  - True because $x'_i = H' x_i$, $x''_i = H'' x_i$

so $x''_i = H'' H'^{-1} x'_i$
Projective vs Affine

[Diagram showing projective vs affine transformations]
Rectification