Single-view Geometry, Epipolar Geometry, Binocular Stereo

Aykut Erdem
Dept. of Computer Engineering
Hacettepe University

Last time

- Image formation geometry
- Vanishing points and vanishing lines
- Pinhole camera model and camera projection matrix

\[ x = K[R \ t]X \]

- Homogeneous coordinates

\[ (x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \]

Camera calibration

\[ P_i \rightarrow MP_i \rightarrow p_i = \begin{bmatrix} u_i \\ v_i \end{bmatrix} \quad M = K[R \ T] \]

\[ k_0 \]

\[ \kappa = \begin{bmatrix} \alpha & -\alpha \cot \theta & u_0 \\ 0 & \beta \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \]

- 11 unknown
- Need at least 6 correspondences
Once the camera is calibrated...

\[ M = K[R \ T] \]

- Internal parameters \( K \) are known
- \( R, T \) are known - but these can only relate \( C \) to the calibration rig

Can I estimate \( P \) from the measurement \( p \) from a single image?

- No - in general ☹️

\[ P \text{ can be anywhere along the line defined by } C \text{ and } p \]

Recovering structure from a single view

- A. Criminisi, I. Reid and A. Zisserman, Single View Metrology, ICCV 1999
  
  http://www.robots.ox.ac.uk/~vgg/projects/SingleView

Recovering structure from a single view

http://www.robots.ox.ac.uk/~vgg/projects/SingleView
Depth Perception: The inverse problem

Monocular cues to depth

• **Absolute depth cues:** (assuming known camera parameters) these cues provide information about the absolute depth between the observer and elements of the scene.

• **Relative depth cues:** provide relative information about depth between elements in the scene (this point is twice as far at that point, ...)

Relative depth cues

Atmospheric perspective

• Based on the effect of air on the color and visual acuity of objects at various distances from the observer.

• Consequences:
  – Distant objects appear bluer
  – Distant objects have lower contrast.

Simple and powerful cue, but hard to make it work in practice
Atmospheric perspective

http://encarta.msn.com/medias_761571997/Perception_psychology.html

Slide credit: B. Freeman and A. Torralba

Shadows

Slide credit: B. Freeman and A. Torralba, S. Marschner

Shading

Figure from Prados & Faugeras 2006

Slide credit: K. Grauman

Focus/defocus

Images from same point of view, different camera parameters

3d shape / depth estimates

Figure from H. Jin and P. Favaro, 2002

Slide credit: K. Grauman
Texture

From A.M. Loh, The recovery of 3D structure using visual texture patterns, PhD thesis

Slide credit: K. Grauman

Motion

Figures from L. Zhang

http://brainconnection.posiscience.com/illusions-shape-from-motion/

Slide credit: K. Grauman

Linear Perspective

• Based on the apparent convergence of parallel lines to common vanishing points with increasing distance from the observer. (Gibson: "perspective order")
• In Gibson's term, perspective is a characteristic of the visual field rather than the visual world. It approximates how we see (the retinal image) rather than what we see, the objects in the world.
• Perspective: a representation that is specific to one individual, in one position in space and one moment in time (a powerful immediacy).
• Is perspective a universal fact of the visual retinal image? Or is perspective something that is learned?

Simple and powerful cue, and easy to make it work in practice...

Slide credit: B. Freeman and A. Torralba
Linear Perspective

Muller-Lyer
1889

The strength of linear perspective

3D percept is driven by the scene, which imposes its ruling to the objects

Properties of Projection

- Angles are not preserved
- Parallel lines meet

Vanishing point

Lines in a 2D plane

\[ ax + by + c = 0 \]

\[ l = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \]

If \( x = [x_1, x_2]^T \in l \)

\[ x_1 \begin{bmatrix} a \\ b \\ c \end{bmatrix}^T = 0 \]
Lines in a 2D plane

Intersecting lines
\[ \mathbf{x} = \mathbf{l} \times \mathbf{l}' \]

Proof
\[ \mathbf{l} \times \mathbf{l}' \perp \mathbf{l} \rightarrow (\mathbf{l} \times \mathbf{l}') \cdot \mathbf{l} = 0 \rightarrow \mathbf{x} \in \mathbf{l} \]
\[ \mathbf{l} \times \mathbf{l}' \perp \mathbf{l}' \rightarrow (\mathbf{l} \times \mathbf{l}') \cdot \mathbf{l}' = 0 \rightarrow \mathbf{x} \in \mathbf{l}' \]
\[ \mathbf{x} \]
\[ \text{---} \text{ } \mathbf{x} \text{ is the intersecting point} \]

Points in 3D

How about lines in 3D?
• Lines have 4 degrees of freedom - hard to represent in 3D-space
• Can be defined as intersection of 2 planes

Points at infinity (ideal points)
\[ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, x_3 \neq 0 \]
\[ \mathbf{x}_\infty = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \]
Let's intersect two parallel lines:
\[ \mathbf{l} \times \mathbf{l}' = (\mathbf{c} - \mathbf{c}') \begin{bmatrix} b \\ -a \\ 0 \end{bmatrix} \]
Agree with the general idea of two lines intersecting at infinity

Points and planes in 3D
\[ \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} \quad \mathbf{\Pi} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \]
\[ \mathbf{x} \in \mathbf{\Pi} \iff \mathbf{x}^T \mathbf{\Pi} = 0 \quad \mathbf{a} \mathbf{x} + \mathbf{b} \mathbf{y} + \mathbf{c} \mathbf{z} + \mathbf{d} = 0 \]

Lines at infinity \( \mathbf{l}_\infty \)
Set of ideal points lies on a line called the line at infinity
How does it look like?
\[ \mathbf{l}_\infty = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \]
Indeed:
\[ \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0 \]
Projective projections of line at infinity (2D)

\[ H = \begin{bmatrix} A & t \\ v & b \end{bmatrix} \]

1' = \( H^{-T} \) l

\[ H_A^{-T} l_\infty = ? = \begin{bmatrix} A & t \\ 0 & 1 \end{bmatrix}^{-T} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} A^{-T} \\ -t^T A^{-T} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]

\[ H^{-T} l_\infty = ? = \begin{bmatrix} A & t \\ v & b \end{bmatrix}^{-T} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \]

no!

Vanishing points (ideal points in 2D)

Vanishing points
- points where parallel lines intersect in 3D

\( d \) = direction of the line

\[ M = K[R \ T] \]

Image of a vanishing point - \( v = K \ d \)

Vanishing points - example

\( v_1, v_2 \): measurements
\( K \) = known and constant

Can I compute R?

\[ d_1 = \frac{K^{-1} v_1}{\|K^{-1} v_1\|} \]

\[ d_2 = \frac{K^{-1} v_2}{\|K^{-1} v_2\|} \]

\[ d_1 = R \ d_2 \]

\[ R \]

Are these two lines parallel or not?
- Recognize the horizon line
- Measure if the 2 lines meet at the horizon
- if yes, these 2 lines are // in 3D

Recognition helps reconstruction!
- Humans have learnt this

Image of a vanishing point = Vanishing points and their image
Vanishing points - example

\[ d_1 = \frac{K^{-1} v_1}{\|K^{-1} v_1\|} \]

\[ d_2 = \frac{K^{-1} v_2}{\|K^{-1} v_2\|} \]

\[ \rightarrow R \]

Planes at infinity & vanishing lines

\[ \Pi_\infty = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \]

- Parallel planes intersect at the plane at infinity
- 2 planes are parallel iff their intersections is a line that belongs to \( P_\infty \)
- Parallel planes intersect the plane at infinity in a common line – the vanishing line (horizon)

Vanishing lines and their images

Parallel planes intersect the plane at infinity in a common line – the vanishing line (horizon)

\[ n = K^T l_{\text{horiz}} \]

Is this parachuter higher or lower than the person taking this picture?
Lower—he is below the horizon
Application of the statistics of edges: Manhattan World

Many scenes of man-made environments are laid out on a 3-D “Manhattan” grid.

This 3-D structure imposes statistical regularities on the edges, and hence the image gradients, in the image.

These regularities allow us to infer the viewer orientation relative to the Manhattan grid and to detect targets unaligned to the grid.

The Structure of Manhattan World

- Assume four kinds of edges in Manhattan-type scenes: x lines, y lines, z lines and random lines.
- Orientations of x,y,z lines given by camera orientation (azimuth, elevation and twist) and knowledge of perspective projection. Orientations also a function of image plane coordinates.

Geometry of the Problem

3-D lines in Manhattan scene project to lines on uv image plane.

Camera axes defined by three orthogonal unit vectors a,b,c which are specified by three Euler angles \( \Psi = (\alpha, \beta, \gamma) \): azimuth, elevation and twist.

Perspective projection:

\[
\begin{align*}
    u &= f \frac{\hat{r} \cdot \hat{a}}{\hat{r} \cdot \hat{c}} \\
    v &= f \frac{\hat{r} \cdot \hat{b}}{\hat{r} \cdot \hat{c}} \\
    f &= \text{focal length of camera}
\end{align*}
\]

Bayesian Model of Manhattan World

Evidence for line edges -- x, y, z or random lines -- provided by the image gradient. Prior on occurrence of these edges.

- **Image gradient magnitude** provides evidence for presence or absence of edges, using \( P_{on} \) and \( P_{off} \) distributions.
- **Image gradient direction** provides information about edge orientations.

**Hidden assignment variables:** at each pixel, is there an x, y, z or random line, or no edge at all?

If we knew this assignment at each pixel, and the camera orientation \( \Psi \), we could predict likely values of image gradient magnitude and direction, \( \hat{E} \) and \( \hat{g} \).
Evidence over all pixels: Bayes net of full Bayesian model

Box represents entire image, with an image gradient vector and assignment variable at each pixel location $u_{ij}$

Structure of net graphically illustrates assumption of conditional independence across pixels.

Experimental Results

Estimate of most probable camera orientation given image, rendered in terms of the corresponding orientations of x and y lines (drawn in black).

Note how the x lines align with the sides of buildings that are visible and facing left. The y lines align with the other visible sides facing right.

Outlier Detection

Input image:

Log($P_{on}/P_{off}$)

Outliers detected
Classifying edges at each pixel

* x lines in red
  * y lines in green
  * z lines in blue

Distance from the horizon line

- Based on the tendency of objects to appear nearer the horizon line with greater distance to the horizon.
- Objects approach the horizon line with greater distance from the viewer. The base of a nearer column will appear lower against its background floor and further from the horizon line. Conversely, the base of a more distant column will appear higher against the same floor, and thus nearer to the horizon line.

Relative height

- The object closer to the horizon is perceived as farther away, and the object further from the horizon is perceived as closer
- If you know camera parameters: height of the camera, then we know real depth

At which elevation has been taken this picture?
Comparing heights

Intersect the vanishing point (from lines)

Comparing heights

Measuring height

Computing vanishing points (from lines)

Intersect \( p_1 q_1 \) with \( p_2 q_2 \)

\[
v = (p_1 \times q_1) \times (p_2 \times q_2)
\]

Least squares version
- Better to use more than two lines and compute the "closest" point of intersection
- See notes by Bob Collins for one good way of doing this:

Measuring height

vanishing line (horizon)

\[
v = (b \times h_0) \times (u_x \times u_y)
\]

Least squares version
- Better to use more than two lines and compute the "closest" point of intersection
- See notes by Bob Collins for one good way of doing this:
Measuring height

vanishing line (horizon)

What if the point on the ground plane $b_0$ is not known?
- Here the guy is standing on the box
- Use one side of the box to help find $b_0$ as shown above

What if $v_z$ is not infinity?

The cross ratio

A Projective Invariant
- Something that does not change under projective transformations (including perspective projection)

The cross-ratio of 4 collinear points

Can permute the point ordering • $4! = 24$ different orders (but only 6 distinct values)

This is the fundamental invariant of projective geometry
Measuring height

\[
\frac{\|t - b\|}{\|v_z - t\|} = \frac{H}{R}
\]

vanishing line (horizon)

image cross ratio

Single View Metrology

A. Criminisi, I. Reid and A. Zisserman
Department of Engineering Science
University of Oxford
Oxford, UK, OX1 3PJ
[criminisi,ian,az]@robots.ox.ac.uk

• A. Criminisi, I. Reid and A. Zisserman, Single View Metrology, ICCV 1999

Three-dimensional reconstruction of a painting. The Music Lesson (1662-65), by J. Vermeer (1632-1675).
Our goal: Recovery of 3D structure

- Recovery of structure from one image is inherently ambiguous

Multi-view geometry problems

- Structure: Given projections of the same 3D point in two or more images, compute the 3D coordinates of that point
Multi-view geometry problems

- **Stereo correspondence:** Given a point in one of the images, where could its corresponding points be in the other images?

![Stereo Correspondence Diagram](65)

- **Motion:** Given a set of corresponding points in two or more images, compute the camera parameters.

![Motion Diagram](66)

Triangulation

- Given projections of a 3D point in two or more images (with known camera matrices), find the coordinates of the point.

![Triangulation Diagram](67)

- We want to intersect the two visual rays corresponding to $x_1$ and $x_2$, but because of noise and numerical errors, they don’t meet exactly.

![Triangulation Diagram](68)
Triangulation: Geometric approach

- Find shortest segment connecting the two viewing rays and let X be the midpoint of that segment

\[ X = \frac{X_1 + X_2}{2} \]

Cross product as matrix multiplication:

\[
\mathbf{a} \times \mathbf{b} = \begin{bmatrix}
0 & -a_z & a_y \\
-a_z & 0 & -a_x \\
a_y & a_x & 0
\end{bmatrix}
\begin{bmatrix}
b_x \\
b_y \\
b_z
\end{bmatrix} = [\mathbf{a} \times \mathbf{b}]
\]

Two independent equations each in terms of three unknown entries of X

- Find X that minimizes

\[
d^2(x_1, P_1X) + d^2(x_2, P_2X)
\]

Triangulation: Linear approach

\[
\lambda_1 x_1 = P_1X \quad x_1 \times P_1X = 0 \quad [x_1^x]P_1X = 0
\]

\[
\lambda_2 x_2 = P_2X \quad x_2 \times P_2X = 0 \quad [x_2^x]P_2X = 0
\]

Cross product as matrix multiplication:

\[
[a \times b] = \begin{bmatrix}
0 & -a_z & a_y \\
-a_z & 0 & -a_x \\
a_y & a_x & 0
\end{bmatrix}
\begin{bmatrix}
b_x \\
b_y \\
b_z
\end{bmatrix} = [a \times b]
\]
**Binocular stereo**

- Given a calibrated binocular stereo pair, fuse it to produce a depth image

Where does the depth information come from?

---

**Human eye**

Rough analogy with human visual system:

- Pupil/Iris - control amount of light passing through lens
- Retina - contains sensor cells, where image is formed
- Fovea - highest concentration of cones

---

**Human stereopsis: disparity**

From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Human eyes **fixate** on point in space - rotate so that corresponding images form in centers of fovea.
**Human stereopsis: disparity**

*Disparity* occurs when eyes fixate on one object; others appear at different visual angles.

---

**Depth without objects**

Random dot stereograms (Bela Julesz)

- Julesz, 1971

---

**Stereo photography and stereo viewers**

Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

- Invented by Sir Charles Wheatstone, 1838
- Image from fisher-price.com

---

*Disparity: $d = r-l = D-F$.**
Binocular stereo

- Given a calibrated binocular stereo pair, fuse it to produce a depth image.

Stereograms: Invented by Sir Charles Wheatstone, 1838

http://www.johnsonshawmuseum.org

Autostereograms

Exploit disparity as depth cue using single image.

(Single image random dot stereogram, Single image stereogram)

Images from magiceye.com

Autostereograms

Images from magiceye.com
Estimating depth with stereo

- **Stereo**: shape from “motion” between two views
- We’ll need to consider:
  - Info on camera pose (“calibration”)
  - Image point correspondences

Basic Principle: Triangulation
- Gives reconstruction as intersection of two rays
- Requires
  - camera pose (calibration)
  - point correspondence

Epipolar constraint
- Geometry of two views constrains where the corresponding pixel for some image point in the first view must occur in the second view.
  - It must be on the line carved out by a plane connecting the world point and optical centers.

Epipolar Constraint
- Reduces correspondence problem to 1D search along conjugate epipolar lines
What do the epipolar lines look like?

1. 

\[ O_1 \quad O_2 \]

2. 

\[ O_1' \quad O_2' \]

Example: converging cameras

The epipoles are the points of intersection of the line joining the camera centres (the baseline) with the image plane.

Example: parallel cameras

Example: Forward motion

What would the epipolar lines look like if the camera moves directly forward?

Epipole has same coordinates in both images.

Points move along lines radiating from \( e \): “Focus of expansion”
Fundamental matrix

• Let \( p \) be a point in left image, \( p' \) in right image

• Epipolar relation
  – \( p \) maps to epipolar line \( l' \)
  – \( p' \) maps to epipolar line \( l \)

• Epipolar mapping described by a 3x3 matrix \( F \)

\[
l' = Fp \hspace{1cm} l = p'F \]

• It follows that

\[
p'Fp = 0
\]

Slide credit: J. Hays

Fundamental matrix

• This matrix \( F \) is called
  – the “Essential Matrix”
    • when image intrinsic parameters are known
  – the “Fundamental Matrix”
    • more generally (uncalibrated case)

• Can solve for \( F \) from point correspondences
  – Each \((p, p')\) pair gives one linear equation in entries of \( F \)

\[
p'Fp = 0
\]

– 8 points give enough to solve for \( F \) (8-point algorithm)

See Marc Pollefeys’s notes for a nice tutorial
http://cs.unc.edu/~marc/tutorial/node53.html

Slide credit: J. Hays

Next time

• Depth estimation from stereo
• Multi-view geometry