EE 576 - Binocular Stereo

H.I. Bozma

Electric Electronic Engineering
Bogazici University

May 27, 2020

Biological Systems

- The world is projected differently onto our two eyes
- Precisely due to this difference $\rightarrow$ Relative distances of objects.
- Close objects - More widely separated on our retinas while the reverse holds for far objects.
- Note: 10% of people do not have true stereo vision, *however* they are still able to determine depth which is based on motion cues.

Stereo Image Pairs

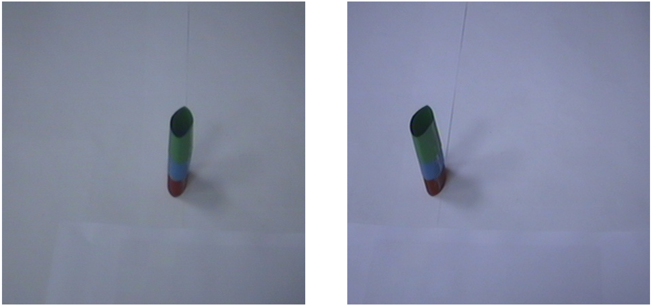
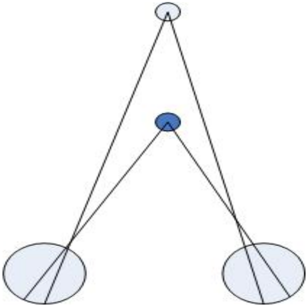


Figure: Stereo image pairs.

Projection of close and far objects on the retina



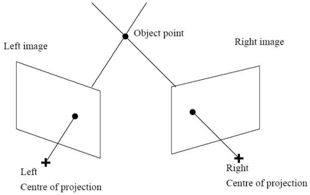
Random Dot Stereograms

Two images of random dots where the second has a portion of the first shifted relative to its original position.

Bela Julesz (1960)

- When viewed appropriately – the views are merged:
- Depth perception - Depends on the amount of shift between the central portions.
- Depending on whether the central portion is shifted to the left or the right, the depth perception is behind or in front of the border

Binocular Vision



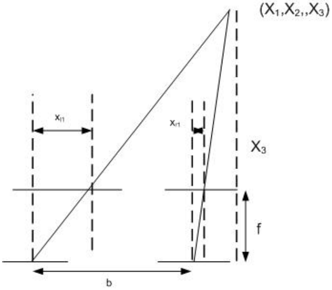
Conjugate pair: Two points in different images that are the projections of the same point in the scene.

Question: Given the conjugate pair x_l and x_r, it is possible to compute the 3D coordinates of the underlying point.

Disparity: Distance between points of a conjugate pair when the two images are superimposed.

From Disparity to 3D Depth

- Consider two cameras, parallel and separated horizontally by a baseline distance b, and having the same focal length



Left image

$$\frac{X_1}{X_3} = \frac{x_{l1}}{f}$$

Right image

$$\frac{X_1 - b}{X_3} = \frac{x_{r1}}{f}$$

$\rightarrow$ Depth: $X_3 = \frac{bf}{x_{l1} - x_{r1}}$

Remaining coordinates:

$$X_1 = \frac{x_{l1} X_3}{f} \qquad X_2 = \frac{x_{l2} X_3}{f}$$

Vergence Angle

- Note in general: Cameras have arbitrary position and orientation with respect to each other.
- The cameras' optical axes may or may not intersect at a point in space.
- **Vergence angle:** The difference in orientations of the two optical axes

Stereo Image Pairs

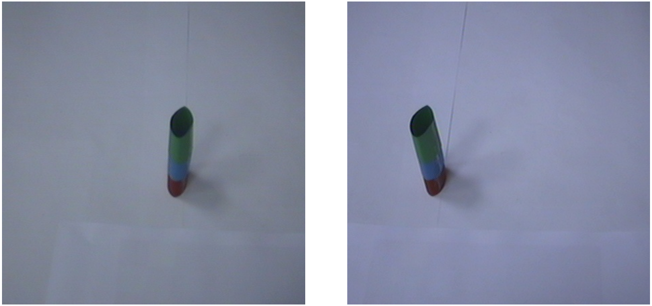


Figure: Stereo image pairs.

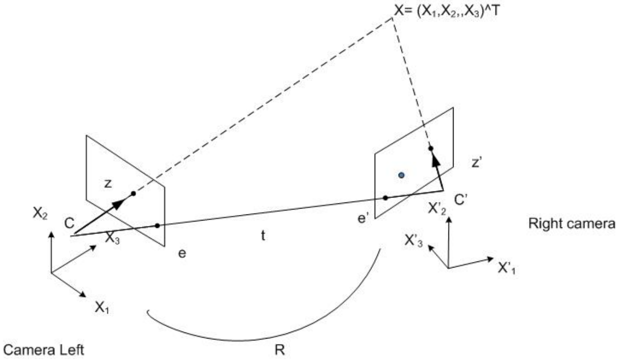
Problem Definition

Correspondence problem:

- Given two images formed in two image planes X_L and X_R,
- For a point x_l in X_L, which point x_r in X_R corresponds to it?
- Two problems:
  1. Where: Given $x_l \in X_L$, where to search for the corresponding pt in X_R
  2. What : How to determine x_l and its conjugate pair? How to establish correspondence?
- Stereo is done at a very low level (Julesz) $\Rightarrow$ Don't "'interpret" the scene before perceiving depth.

Where to Search?

Stereo viewing geometry



Epipolar Geometry

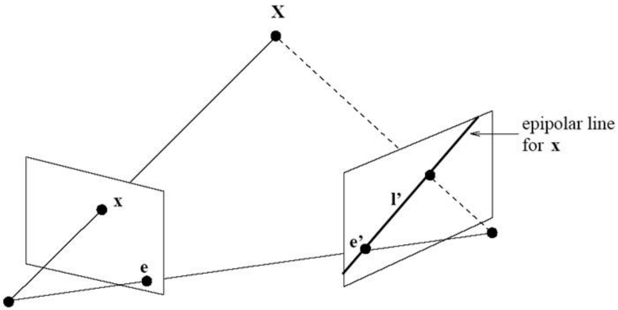
- Epipolar lines, planes & epipole
- Essential matrix & Fundamental matrix

Epipolar Line

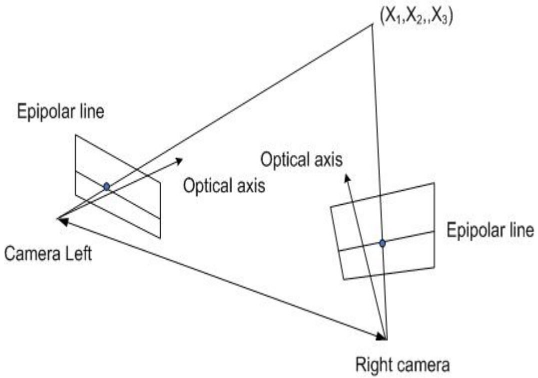
• Possible matching: Correspondence does not require a search through the whole image – rather a single line.

- A feature in one image lies anywhere along the viewing ray.
- Project this viewing ray into the other image.
- A point x in one image generates a line in the other on which its corresponding point x' must lie
- Epipolar line - Line in the second image on which the feature we are trying to match must lie.
- The image in one camera of a ray through the optical center and image point in the other camera.

Epipolar Line



Two Cameras: Epipolar Lines

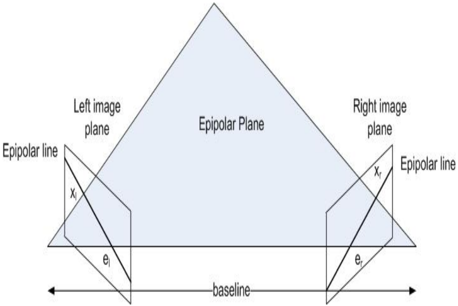


Epipolar Plane

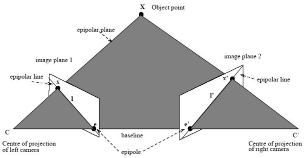
Epipolar plane: The plane defined by a 3D point X and the optical centers C and C'.

- Note: Different 3D points give rise to different epipolar planes.
- Epipolar line - The straight line of intersection of the epipolar plane with the image plane.
- All epipolar planes - Intersect at the baseline of the binocular system.

Epipolar Plane



Epipole



Epipole - Point of intersection of the line joining the optical centers, that is the baseline, with the image plane.

- The image, in one camera, of the optical centre of the other camera.
- All such lines converge at the epipole, which is just the FOE.

Image Formation

- Points: Expressed either in Euclidean or projective coordinates.
  - $X \in R^3 \to x \in R^2$.
  - $x = [x_1 \, x_2]^T$ or $\tilde{x} = \begin{bmatrix} x_1 & x_2 & 1 \end{bmatrix}$.
  - Note that $\tilde{x} = \begin{bmatrix} x_1 & x_2 & 1 \end{bmatrix} = \begin{bmatrix} sx_1 & sx_2 & s \end{bmatrix}$ for any non-zero scalar $s > 0$ represent the same image coordinates.

Left Camera and Right Camera: Projective Coordinates

- Consider left camera with coordinate frame C. Related projective coordinates are: $\tilde{x} = \begin{bmatrix} \frac{x_1 - o_1}{f} \\ \frac{x_2 - o_2}{f} \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 - o_1 \\ x_2 - o_2 \\ f \end{bmatrix}$
- Now, consider the right camera with coordinate system centered at C'. Related projective coordinates

$$\tilde{x}' = \begin{bmatrix} \frac{x_1' - o_1'}{f'} \\ \frac{x_2' - o_2'}{f'} \\ 1 \end{bmatrix} = \begin{bmatrix} x_1' - o_1' \\ x_2' - o_2' \\ f' \end{bmatrix}$$

Relating Right & Left Image Coordinate Systems

Suppose that the global coordinate system is fixed at C'. Then

$$\tilde{x}' = \begin{bmatrix} x_1' - o_1' \\ x_2' - o_2' \\ f' \end{bmatrix} = R \begin{bmatrix} x_1 - o_1 \\ x_2 - o_2 \\ f \end{bmatrix} + t = R\tilde{x} + t$$

R - Rotation matrix btw the two coordinate systems btw C and C'
t - Translation vector btw the two coordinate systems btw C and C'

Matrix Theory Results To be Used

- A line going through two points, m and n - Represented by the cross product $m \times n$ - namely $\alpha_1 m + \alpha_2 n$.
- If x is a point on this line, then $det(x, m, n) = 0$
- Equivalent to $x^T(m \times n) = 0$.
- Note that

$$m \times n = \begin{bmatrix} 0 & -m_3 & m_2 \\ m_3 & 0 & -m_1 \\ -m_2 & m_1 & 0 \end{bmatrix} n = J(m)n$$

where $J(m)$ is the skew-symmetric matrix associated with vector m

Essential Matrix E - Relating Right & Image Coordinates

Recall that $\tilde{x}' = R\tilde{x} + t$ Using the results from previous slide

$$\begin{aligned}
\tilde{x}\prime^T (R\tilde{x} \times t) &= 0 \\
\tilde{x}\prime^T (t \times R\tilde{x}) &= 0 \\
\tilde{x}\prime^T TR\tilde{x} &= 0 \\
\tilde{x}\prime^T E\tilde{x} &= 0
\end{aligned}$$

T - Skew-symmetric matrix formed by t - namely $T = J(t)$.

Estimating Essential Matrix

- Cameras are partially calibrated - - namely no knowledge of R and t, but only image coordinates in the image plane wrt to center of projection. Hence o_1, o_2, f are known.
- Essential matrix E - Can be estimated from a small number of corresponding points.
- In $\tilde{x}\prime^T E\tilde{x} = 0$, image coordinates include offsets and normalization by focal lengths

• A 3×3 matrix with only 5 degrees of freedom. (In fact six parameters, three from rotation and three from translation, but only two of translation is recoverable without additional information!

• To estimate it using corresponding image points, the intrinsic parameters of both cameras must be known.

Fundamental Matrix

- Fundamental matrix - A mathematical construct that encodes the geometric information that relates two different viewpoints of the same scene. The two viewpoints:
  - A pair of stereo images,
  - A temporal pair of images taken at different times with the camera moving between image acquisitions.

Projective Coordinates

Left Image Right image

$$\begin{bmatrix} x_1 - o_1 \\ x_2 - o_2 \\ f \end{bmatrix} \quad \begin{bmatrix} x_1' - o_1' \\ x_2' - o_2' \\ f' \end{bmatrix}$$

Equivalently, left image coordinates:

$$\begin{bmatrix} x_1 - o_1 \\ x_2 - o_2 \\ f \end{bmatrix} = \begin{bmatrix} 1 & 0 & -o_1 \\ 0 & 1 & -o_2 \\ 0 & 0 & f \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

and right image coordinates

$$\begin{bmatrix} x_1' - o_1' \\ x_2' - o_2' \\ f' \end{bmatrix} = \begin{bmatrix} 1 & 0 & -o_1' \\ 0 & 1 & -o_2' \\ 0 & 0 & f' \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ 1 \end{bmatrix}$$

Projective Coordinates → Image Coordinates

Let $x' = \begin{bmatrix} x_1' & x_2' & 1 \end{bmatrix}^T$.

$$\begin{bmatrix} x_1' - o_1' \\ x_2' - o_2' \\ f' \end{bmatrix} \approx \frac{1}{f'} \begin{bmatrix} 1 & 0 & -o_1' \\ 0 & 1 & -o_2' \\ 0 & 0 & f' \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \\ 1 \end{bmatrix}$$

$$\Rightarrow \tilde{x}' \approx \begin{bmatrix} \frac{1}{f'} & 0 & \frac{-o_1'}{f'} \\ 0 & \frac{1}{f'} & \frac{-o_2'}{f'} \\ 0 & 0 & 1 \end{bmatrix} x'$$

$$\Rightarrow \tilde{x}' \approx K'^{-1} x'$$

Similarly, for left camera: $\tilde{x} \approx K^{-1} x$

Recall K: Perspective Projection Equations

A scene point X is projected onto the image plane:

$$\begin{bmatrix} sx_1 \\ sx_2 \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & o_1 & 0 \\ 0 & f & o_2 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & o_1 \\ 0 & f & o_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix}$$

$$= \mathbf{K} \begin{bmatrix} I & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix}$$

- K - Matrix of the intrinsic parameters of the camera!
- o_1, o_2 - Center of the camera image plane,
- I - 3×3 identity matrix.

Fundamental Matrix

Using results from Essential Matrix calculations,

$$(\tilde{x}')^T TR\tilde{x} = (K^{-1}x')^T TRK^{-1}x = (x')^T \underbrace{(K'^{-1})^T TRK^{-1}}_{F} x = 0$$

$$(x')^T Fx = 0$$

Note $F \approx (K^{-1})^T EK^{-1}$

In $(x')^T Fx = 0$, image coordinates are relative to any arbitrary local image coordinate system!

Epipolar constraint: $(x')^T Fx = 0$ -

It maps image points to their corresponding epipolar lines, that is, $Fx = l'$, since $(x')^T l' = (x')^T Fm = 0$. Similarly, $F^T x' = l$.

Fundamental Matrix

- Fundamental matrix F - 7 degrees of freedom. There are 9 matrix elements, but only their ratio is significant, which leaves 8 degrees of freedom. In addition, the constraint that $\det(F) = 0$ leaves only 7.
- Geometrically, F maps epipoles to the origin of the corresponding image plane.
- Algebraically, the epipoles e and e' are the null vectors of F and F^T respectively.

Essential and Fundamental Matrix Properties

1. E - Encapsulates only the extrinsic parameters while F encodes both the intrinsic and the extrinsic parameters of the camera
2. Both F and E are rank-2 matrices as $\mathrm{rank}(T) = 2$. Thus, $\det(F) = 0$ and $\det(E) = 0$.

Estimating the Fundamental Matrix

$$a^T f = 0 \tag{1}$$

where

$$a = \begin{bmatrix} x_1 x_1' & x_2 x_1' & x_1' & x_1 x_2' & x_2 x_2' & x_2' & x_1 & x_2 & 1 \end{bmatrix}^T$$

$$f = \begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{21} & f_{22} & f_{23} & f_{31} & f_{32} & f_{33} \end{bmatrix}^T$$

If we have at least 8 matches, then we should be able to solve for the f vector upto a scale factor by solving for the null vector of the data matrix.

What to Match?

If the point to be matched is clearly different from its surrounding pixels, this is a simpler task. $\rightarrow$ Find matchable features.

- Raw intensity
- Edges
- Regions
- Features

Similarity measure

Matching Complexity Tradeoff

- Difference in position and orientation of the stereo views small $\rightarrow$ Matching corresponding points is easy! Matching is difficult if the difference is large.
- Accuracy of the 3D reconstruction - Poor if the difference in position and orientation of stereo views is small.

Matching Raw Intensity

- Similarity – Color C dots only match Color C dots.
- Uniqueness– One dot can match no more than one dot.
- Continuity – Disparity values vary smoothly almost everywhere.
- Ordering – If x is to the left y in the left image, then the corresponding points x' and y' are also located similarly in the right image.

Edge Matching

1. Filter each image with Gaussian filters at four filter widths such the each filter is twice as wide as the next via repeated convolution with the smallest filter.
2. Compute edge positions on each row.
3. Match edges in corresponding rows at the coarse resolutions via the comparison of orientations and strengths. Note that horizontal edges can't be matched.
4. Improve the disparity estimates by matching at finer scales.

General features

- Depth values at the edge points $\rightarrow$ A sparse depth map.
- Standard algorithm $\rightarrow$ Implements a multiscale approach, and assumes a parallel geometry
- Meaningless information along occlusions

Region Matching

- An alternative approach is based on region matching.
- Matching can be based on either intensity or some other region features.
- In intensity based matching, one approach to finding interesting regions is to find regions of high variance.

$$F(x) = min_i F_i(x)$$

Conjugate Pairs & Disparity

- Difference in position and orientation of the stereo views is small $\rightarrow$ Matching corresponding points easy.
- Difference is large $\rightarrow$ Matching difficult
- Corresponding image points that have zero disparity values $\rightarrow$ Projected by 3D points at a finite distance from the cameras
- 3D reconstruction accuracy - Poor if the difference in position and orientation of stereo views is large.

Structure From Motion

Stereo with a single camera: Two images taken at time intervals t and $t + \delta t$

Determination of structure: Matching features btw consecutive images

Assuming intrinsic camera parameters are known, determine the relative geometry of the two images through using the 3D motion the camera

Use stereo equations to compute depth

Problematic if the the scene has dynamic entities!