Epipolar Geometry

A lot of slides from Noah Snavely + Shree Nayar’s YT series: First principals of Computer Vision

CS180: Intro to Computer Vision and Comp. Photo
Angjoo Kanazawa & Alexei Efros, UC Berkeley, Fall 2023
So how do we get depth?

- Find the disparity! of corresponding points!
- Called: Stereo Matching

Left/Right Camera Images

Disparity Map (Ground Truth)
Where is the corresponding point going to be?

Hint
Epipolar Line

Two images captured by a purely horizontal translating camera (rectified stereo pair)

\[ x_1 - x_2 = \text{the disparity of pixel } (x_1, y_1) \]
Your basic stereo algorithm

For every 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

Improvement: match *windows*, + clearly lots of matching strategies
Your basic stereo algorithm

Determine Disparity using Template Matching

Template Window $T$

Search Scan Line $L$

Left Camera Image $E_l$

Right Camera Image $E_r$
Correspondence problem

Parallel camera example – epipolar lines are corresponding rasters

Source: Andrew Zisserman
Intensity profiles

• Clear correspondence between intensities, but also noise and ambiguity

Source: Andrew Zisserman
Correspondence problem

Neighborhood of corresponding points are similar in intensity patterns
- Use Normalized Cross Correlation (NCC) or a distance in some descriptor within a window

Source: Andrew Zisserman
Correlation-based window matching

Source: Andrew Zisserman
Dense correspondence search

For each epipolar line

For each pixel / window in the left image

• compare with every pixel / window on same epipolar line in right image

• pick position with minimum match cost (e.g., SSD, correlation)

Adapted from Li Zhang
Textureless regions

Target region

Left image band (x)

Source: Andrew Zisserman
Effect of window size

Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

Figures from Li Zhang
Issues with Stereo

- Surface must have non-repetitive texture
- Foreshortening effect makes matching a challenge
Stereo Results

– Data from University of Tsukuba

Scene

Ground truth
Results with Window Search

Window-based matching (best window size)

Ground truth
Better methods exist...

Energy Minimization

Ground truth
Summary

• With a simple stereo system, **how much pixels move, or “disparity”** give information about the depth

• Correspondences to measure the pixel disparity
Many problems in 3D

- 3D Points (Structure)
- Correspondences
- Camera (Motion)
Camera Calibration

3D Points (Structure)

Correspondences

Camera (Motion)
Stereo (w/2 cameras); Multi-view Stereo / Triangulation

3D Points (Structure)

Correspondences

Camera (Motion)

KNOWN

UNKNOWN

(Estimated)
Camera helps Correspondence: Epipolar Geometry

3D Points (Structure)

Correspondences

Camera (Motion)

Relationship
Correspondence gives camera: **Epipolar Geometry**

Correspondences → 3D Points (Structure) → Correspondences

3D Points (Structure) → Camera (Motion) → Relationship
Recap

We covered:

• How to estimate the camera parameters
  – “Calibration”
  – Solve for intrinsics & extrinsics

• With a simple stereo, correspondences lie on horizontal lines

• depth is inversely proportional to disparity (how much the pixel moves)
What Depth Map provides

warping the pixel based on its depth as you change the views

Monocular Depth Prediction [Ranftl et al. PAMI’20]
More cool things with Depth

Niklaus et al. ToG 2019

Shih et al. CVPR 2020

3D photo

AR
Next: General case

- The two cameras need not have parallel optical axes.
- Assume camera intrinsics are calibrated

Same hammer:
Find the correspondences, then solve for structure
Option 1: Rectify via homography

- reproject image planes onto a common plane
  - plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- Two homographies, one for each input image reprojection
Option 1: Rectify via homography

Then find correspondences on the horizontal scan line

Original stereo pair

After rectification
General case, known camera, find depth: Option 2

1. Find correspondences
2. Triangulate
General case, known camera, find depth:

Option 2

1. Find correspondences
2. Triangulate

Can we restrict the search space again to 1D?

What is the relationship between the camera + the corresponding points?
Where do epipolar lines come from?
Stereo correspondence constraints

- Given p in left image, where can corresponding point p' be?
Stereo correspondence constraints

- Given \( p \) in left image, where can corresponding point \( p' \) be?
Where do we need to search?
Epipolar Geometry

If you get confused with the following math, look at this picture again, it just describes this.
Stereo correspondence constraints
• Potential matches for $p$ have to lie on the corresponding epipolar line $l'$. 
• Potential matches for $p'$ have to lie on the corresponding epipolar line $l$. 

Source: M. Pollefeys
Stereo correspondence constraints

• Geometry of two views allows us to constrain where the corresponding pixel for some image point in the first view must occur in the second view.

Epipolar constraint: Why is this useful?

• Reduces correspondence problem to 1D search along conjugate epipolar lines

http://www.ai.sri.com/~luong/research/Meta3DViewer/EpipolarGeo.html

Adapted from Steve Seitz
Parts of Epipolar geometry

- **Baseline** – line connecting the two camera centers
- **Epipolar Plane** – plane containing baseline (1D family)
- **Epipoles**
  - intersections of baseline with image planes
  - projections of the other camera center
  - vanishing points of the baseline
The Epipole

Photo by Frank Dellaert
Example
Example: converging cameras

As position of 3d point varies, epipolar lines “rotate” about the baseline

Figure from Hartley & Zisserman
Example: Parallel to Image Plane

Where is the epipole?

Epipoles \textit{infinitely} far away, epipolar lines parallel

Slide credit: David Fouhey
Example: Forward Motion

Epipole is focus of expansion / principal point of the camera.

Epipolar lines go out from principal point
Example: forward motion

Epipole has same coordinates in both images.
Points move along lines radiating from e: “Focus of expansion”

Figure from Hartley & Zisserman
Motion perpendicular to image plane

http://vimeo.com/48425421

Slide credit: David Fouhey
Ok so where were we?

• Setup: Calibrated Camera (both extrinsic & intrinsic)
• Goal: 3D reconstruction of corresponding points in the image
• We need to find correspondences!

→ 1D search along the epipolar line!
Ok so what exactly are $l$ and $l'$?
Step 0: Normalized image coordinates

- Let’s factor out the effect of K (do everything in 3D)
- Since we know the intrinsics K, apply its inverse to x with depth = 1
- This is called the *normalized* image coordinates. It may be thought of as a set of points with K = Identity

\[
x = K[R \ t]X
\]
\[
K^{-1}x = [R \ t]X
\]

\[
x_{\text{norm}} = K^{-1}x_{\text{pixel}} = [I \ 0]X, \quad x'_{\text{norm}} = K'^{-1}x'_{\text{pixel}} = [R \ t]X
\]

- Assume that the points are normalized from here on
Epipolar constraint: Calibrated case

The vectors $x$, $t$, and $x'$ are coplanar.

What can you say about their relationships, given $n = t \times x'$?

- $x' \cdot (t \times x') = 0$
- $x' \cdot (t \times (Rx + t)) = 0$
- $x' \cdot (t \times Rx + t \times t) = 0$
- $x' \cdot (t \times Rx) = 0$
Epipolar constraint: Calibrated case

\[ x' \cdot [t \times (Rx)] = 0 \quad \Rightarrow \quad x'^T [t_x] Rx = 0 \]

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

The vectors \( x, t, \) and \( x' \) are coplanar.
Epipolar constraint: Calibrated case

The vectors $x$, $t$, and $x'$ are coplanar
Epipolar constraint: Calibrated case

- $E \mathbf{x}$ is the epipolar line associated with $\mathbf{x}$ ($l' = E \mathbf{x}$)
  - Recall: a line is given by $ax + by + c = 0$ or
    \[
    \begin{bmatrix}
    a \\
    b \\
    c
    \end{bmatrix}
    \begin{bmatrix}
    x \\
    y \\
    1
    \end{bmatrix} = 0
    \]
Epipolar constraint: Calibrated case

- $E \mathbf{x}$ is the epipolar line associated with $\mathbf{x}$ ($l' = E \mathbf{x}$)
- $E^T \mathbf{x'}$ is the epipolar line associated with $\mathbf{x'}$ ($l = E^T \mathbf{x'}$)
- $E \mathbf{e} = 0$ and $E^T \mathbf{e'} = 0$
- $E$ is singular (rank two)
- $E$ has five degrees of freedom

$$x'^T E x = 0$$
Epipolar constraint: Uncalibrated case

- Recall that we normalized the coordinates
  \[ x = K^{-1} \hat{x} \quad x' = K'^{-1} \hat{x} \]
  where \( \hat{x} \) is the image coordinates
- But in the \textit{uncalibrated} case, \( K \) and \( K' \) are unknown!
- We can write the epipolar constraint in terms of \textit{unknown} normalized coordinates:

\[
x'^{T} E x = 0
\]

\[
(K'^{-1} \hat{x}')^{T} E (K^{-1} \hat{x}) = 0
\]

\[
\hat{x}'^{T} K'^{-T} E (K^{-1} \hat{x}) = 0
\]

\[
\hat{x}'^{T} F \hat{x} = 0
\]

\[
F = K'^{-T} E K^{-1}
\]

\textbf{Fundamental Matrix} \\
(Faugeras and Luong, 1992)
Epipolar constraint: Uncalibrated case

- $F \hat{x}$ is the epipolar line associated with $\hat{x}$ ($l' = F \hat{x}$)
- $F^T \hat{x}'$ is the epipolar line associated with $\hat{x}'$ ($l = F^T \hat{x}'$)
- $Fe = 0$ and $F^Te' = 0$
- $F$ is singular (rank two)
- $F$ has seven degrees of freedom

\[ xx'^T E x = 0 \quad \text{and} \quad \hat{x}'^T F \hat{x} = 0 \quad \text{with} \quad F = K'^{-T} E K^{-1} \]
Where are we? (in the original setup)

We have two images with calibrated cameras, want the 3D points!

1. Solve for correspondences using epipolar constraints from known camera (1D search)
   • Now we know the exact equation of this line
2. Triangulate to get depth!
Finally: computing depth by triangulation

We know about the camera, $K_1$, $K_2$ and $[R\ t]$:

\[ x = KX \]
\[ x' = K'X' \]
\[ X = X' = RX + T \]

and found the corresponding points: $x \leftrightarrow x'$

How many unknowns + how many equations do we have?

only unknowns!

Solve by least squares
Triangulation Disclaimer: Noise

Ray's don't always intersect because of noise!!!

Least squares get you to a reasonable solution but it's not the actual geometric error (it's how far away the solution is from $Ax = 0$)

In practice with noise, you do non-linear least squares, or “bundle adjustment” (more than 2 image case, next lecture..)

Slide credit: Shubham Tulsiani
Summary: Two-view, known camera

0. Calibrate the camera.

1. Find correspondences:
   - Reduce this to 1D search with Epipolar Geometry!

2. Get depth:
   - If simple stereo, disparity (difference of corresponding points) is inversely proportional to depth
   - In the general case, triangulate.
What if we don’t know the camera?

3D Points
(Structure)

Correspondences

Camera
(Motion)
What if we don’t know the camera?

Assume we know the correspondences: \( \hat{x}' \) and \( \hat{x} \) in the image

\[
\hat{x}'^T F \hat{x} = 0
\]

\[
\hat{x} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}
\]

\[
\begin{bmatrix}
  u' & v' & 1 \\
  f_{11} & f_{12} & f_{13} \\
  f_{21} & f_{22} & f_{23} \\
  f_{31} & f_{32} & f_{33}
\end{bmatrix}
\begin{bmatrix}
  u \\
  v \\
  1
\end{bmatrix} = 0
\]

How many correspondences do we need?
Estimating the fundamental matrix
The eight-point algorithm

\[ \mathbf{x} = (u,v,1)^T, \quad \mathbf{x}' = (u',v',1) \]

\[
\begin{bmatrix}
    u' \\
    v' \\
    1
\end{bmatrix} \begin{bmatrix}
    f_{11} & f_{12} & f_{13} \\
    f_{21} & f_{22} & f_{23} \\
    f_{31} & f_{32} & f_{33}
\end{bmatrix} \begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix} = 0
\]

Solve homogeneous linear system using eight or more matches

Enforce rank-2 constraint (take SVD of \( \mathbf{F} \) and throw out the smallest singular value)
The Fundamental Matrix Song

In the other view passing through x-prime

http://danielwedge.com/fmatrix/
https://www.youtube.com/watch?time_continue=8&v=DgGV3I82NTk&feature=emb_title
Going from F to the Camera

Get the essential matrix with K (or some estimates of K)

\[ E = K' T F K. \]

How the 2D lines relate with 3D lines is captured by intrinsics!
Essential matrix can be decomposed

\[ E = T_x R \]

If we know \( E \), we can recover \( t \) and \( R \)

Given that \( T_x \) is a Skew-Symmetric matrix \((a_{ij} = -a_{ji})\) and \( R \) is an Orthonormal matrix, it is possible to "decouple" \( T_x \) and \( R \) from their product using "Singular Value Decomposition".
This completes: Corresp to Camera

Correspondences

3D Points (Structure)

Camera (Motion)
What about more than two views?

The geometry of three views is described by a $3 \times 3 \times 3$ tensor called the *trifocal tensor*.

The geometry of four views is described by a $3 \times 3 \times 3 \times 3$ tensor called the *quadrifocal tensor*.

After this it starts to get complicated…
Putting it all together

Structure-from-Motion: You know nothing!
(except ok maybe intrinsics)
The starting point for all problems where you can’t calibrate actively
(after that): Neural Rendering

A form of multi-view stereo, more on this in the NeRF lecture.
Next: Large-scale structure from motion

Dubrovnik, Croatia. 4,619 images (out of an initial 57,845).
Total reconstruction time: 23 hours
Number of cores: 352

Building Rome in a Day, Agarwal et al. ICCV 2009
Slide courtesy of Noah Snavely
Large-scale structure from motion

Result using COLMAP: Schönberger et al. CVPR ’16