## CS 4495 Computer Vision

## N-Views (2) - Essential and Fundamental Matrices

Aaron Bobick
School of Interactive Computing


## Administrivia

- Today: Second half of N -Views $(\mathrm{n}=2)$
-PS 3: Will hopefully be out by Thursday
- Will be due October $6^{\text {th }}$.
- Will be based upon last week and today's material
- We may revisit the logistics - suggestions?


## Two views...and two lectures

- Projective transforms from image to image
- Some more projective geometry
- Points and lines and planes
- Two arbitrary views of the same scene
- Calibrated - "Essential Matrix"
- Two uncalibrated cameras "Fundamental Matrix"
- Gives epipolar lines


## Last time

- Projective Transforms: Matrices that provide transformations including translations, rotations, similarity, affine and finally general (or perspective) projection.
- When 2D matrices are $3 \times 3$; for 3D they are $4 \times 4$.



## Last time: Homographies

- Provide mapping between images (image planes) taken from same center of projection; also mapping between any images of a planar surface.



## Last time: Projective geometry



- A line is a plane of rays through origin
- all rays ( $x, y, z$ ) satisfying: $a x+b y+c z=0$

$$
\begin{gathered}
\text { in vector notation: } 0=\left[\begin{array}{lll}
a & b & c
\end{array}\right]\left[\begin{array}{l}
x \\
y \\
z
\end{array}\right] \\
\mathbf{l}
\end{gathered}
$$

- A line is also represented as a homogeneous 3-vector I


## Projective Geometry: lines and points

2D Lines: $\quad a x+b y+c=0$

$$
\left[\begin{array}{lll}
a & b & c
\end{array}\right]\left[\begin{array}{l}
x \\
y \\
1
\end{array}\right]=0
$$

$$
\mathbf{1}^{T} \mathbf{x}=0
$$

$l=\left[\begin{array}{lll}a & b & c\end{array}\right] \Rightarrow\left[\begin{array}{lll}n_{x} & n_{y} & d\end{array}\right]$




$$
\left.\begin{array}{l}
l_{1}=\left[\begin{array}{lll}
a_{1} & b_{1} & c_{1}
\end{array}\right] \\
l_{2}=\left[\begin{array}{lll}
a_{2} & b_{2} & c_{2}
\end{array}\right]
\end{array}\right\} x_{12}=l_{1} \times l_{2}
$$

## Motivating the problem: stereo

- Given two views of a scene (the two cameras not necessarily having optical axes) what is the relationship between the location of a scene point in one image and its location in the other?



## Stereo correspondence

- Determine Pixel Correspondence
- Pairs of points that correspond to same scene point


Epipolar Constraint

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


## Example: converging cameras



Figure from Hartley \& Zisserman

## Epipolar geometry: terms

- Baseline: line joining the camera centers
- Epipole: point of intersection of baseline with image plane
- Epipolar plane: plane containing baseline and world point
- Epipolar line: intersection of epipolar plane with the image plane
- All epipolar lines intersect at the epipole
- An epipolar plane intersects the left and right image planes in corresponding epipolar lines



## From Geometry to Algebra

- So far, we have the explanation in terms of geometry.
- Now, how to express the epipolar constraints algebraically?


## Stereo geometry, with calibrated cameras



Main idea

## Stereo geometry, with calibrated cameras



If the stereo rig is calibrated, we know :
how to rotate and translate camera reference frame 1 to get to camera reference frame 2.

Rotation: $3 \times 3$ matrix $\mathbf{R}$; translation: 3 vector $\mathbf{T}$.

## Stereo geometry, with calibrated cameras



If the stereo rig is calibrated, we know :
how to rotate and translate camera reference frame 1 to get to camera reference frame 2.

$$
\mathbf{X}_{c}^{\prime}=\mathbf{R} \mathbf{X}_{c}+\mathbf{T}
$$

## An aside: cross product

$$
\vec{a} \times \vec{b}=\vec{c}
$$

Vector cross product takes two vectors and returns a third vector that's perpendicular to both inputs.

So here, c is perpendicular to both a and b , which means the dot product $=0$.

$$
\begin{aligned}
& \vec{a} \cdot \vec{c}=0 \\
& \vec{b} \cdot \vec{c}=0
\end{aligned}
$$

## From geometry to algebra



$$
\mathrm{X}=\mathrm{RX}
$$

$T \times \mathbf{X}^{\prime}=$
Normal to the plane

$$
=\mathbf{T} \times \mathbf{R} \mathbf{X}
$$

## Another aside: Matrix form of cross

 product$$
\vec{a} \times \vec{b}=\left[\begin{array}{ccc}
0 & -a_{3} & a_{2} \\
a_{3} & 0 & -a_{1} \\
-a_{2} & a_{1} & 0
\end{array}\right]\left[\begin{array}{l}
b_{1} \\
b_{2} \\
b_{3}
\end{array}\right]=\vec{c}
$$

Can be expressed as a matrix multiplication!!!

$$
\left[a_{x}\right]=\left[\begin{array}{ccc}
0 & -a_{3} & a_{2} \\
a_{3} & 0 & -a_{1} \\
-a_{2} & a_{1} & 0
\end{array}\right] \quad \begin{aligned}
& \text { Notation: } \\
& \vec{a} \times \vec{b}=\left[\vec{a}_{\times}\right] \vec{b} \\
& \text { Has rank 2! }
\end{aligned}
$$

From geometry to algebra


$$
\mathrm{X}=\mathrm{RX} \mathrm{~T}
$$

$\mathbf{T} \times \mathbf{X}^{\prime}=$
Normal to the plane

$$
=\mathbf{T} \times \mathbf{R X}
$$

$$
0=\mathbf{X}^{\prime} \cdot(\mathbf{T} \times \mathbf{R} \mathbf{X})
$$

## Essential matrix

$$
\begin{aligned}
& \mathbf{X}^{\prime} \cdot(\mathbf{T} \times \mathbf{R X})=0 \\
& \mathbf{X}^{\prime} \cdot\left(\left[\mathrm{T}_{x}\right] \mathbf{R X}\right)=0 \\
& \text { Let } \quad \mathbf{E}= {\left[\mathrm{T}_{x}\right] \mathbf{R} } \\
& \mathbf{X}^{\prime T} \mathbf{E X}=0
\end{aligned}
$$



E is called the essential matrix, and it relates corresponding image points between both cameras, given the rotation and translation.

Note: these points are in each camera coordinate systems.
We know if we observe a point in one image, its position in other image is constrained to lie on line defined by above.

## Essential matrix example: parallel cameras



## Weak calibration

- Want to estimate world geometry without requiring calibrated cameras
- Archival videos (already have the pictures)
- Photos from multiple unrelated users
- Dynamic camera system
- Main idea:
- Estimate epipolar geometry from a (redundant) set of point correspondences between two uncalibrated cameras


## From before: Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates:
where:

$$
\begin{gathered}
\mathbf{\Phi}_{e x t}=\left[\begin{array}{rrr}
r_{11} & & r_{12} \\
r_{21} & r_{22} \\
r_{31} & r_{32}
\end{array}\right. \\
\mathbf{K}_{\text {int }}
\end{gathered}=\left[\begin{array}{ccc}
-f / s_{x} & 0 & o_{x} \\
0 & -f / s_{y} & o_{y} \\
0 & 0 & 1
\end{array}\right] .
$$

Note: Invertible, scale $x$ and $y$, assumes no skew

## From before: Projection matrix

- This can be rewritten as a matrix product using homogeneous coordinates:

$$
\left[\begin{array}{c}
w X_{i m} \\
w y_{i m} \\
w
\end{array}\right]=\mathbf{K}_{i n t} \mathbf{\Phi}_{e x t}\left[\begin{array}{c}
X_{w} \\
Y_{w} \\
Z_{w} \\
1
\end{array}\right]
$$

$$
\mathbf{p}_{i m}=\mathbf{K}_{i n t} \underbrace{\mathbf{\Phi}_{\text {ext }} \mathbf{P}_{w}}_{\mathbf{p}_{c}}
$$

$$
\mathbf{p}_{i m}=\mathbf{K}_{i n t} \mathbf{p}_{c}
$$

## Uncalibrated case

For a given camera:

$$
\mathbf{p}_{i m}=\mathbf{K}_{i n t} \mathbf{p}_{c}
$$

So, for two cameras (left and right):

$$
\begin{aligned}
\mathbf{p}_{c, \text { left }} & =\mathbf{K}_{\text {int,left }}^{-1} \mathbf{p}_{\text {im,left }} \\
\mathbf{p}_{c, r i g h t} & =\underbrace{\mathbf{K}_{\text {int,right }}^{-1}}_{\begin{array}{l}
\text { Internal calibration } \\
\text { matrices, one per } \\
\text { camera }
\end{array}} \mathbf{p}_{\text {im,right }}
\end{aligned}
$$

## Uncalibrated case

$\mathbf{p}_{c, \text { right }}=\mathbf{K}_{\text {intr,ight }}^{-1} \mathbf{p}_{\text {im, right }} \quad \mathbf{p}_{c, \text { left }}=\mathbf{K}_{\text {int,left }}^{-1} \mathbf{p}_{\text {im,left }}$
From before, the essential matrix $\mathbf{E}$.

$$
\mathbf{p}_{c, \text { left }}=\mathbf{K}_{i n t, l e f t}^{-1} \mathbf{p}_{i m, \text { left }}
$$

$$
\mathbf{p}_{c, \text { right }}{ }^{\mathrm{T}} \mathbf{E} \mathbf{p}_{c, \text { left }}=0
$$

$$
\left(\mathbf{K}_{\text {int,right }}^{-1} \mathbf{p}_{\text {im,right }}\right)^{\mathrm{T}} \mathbf{E}\left(\mathbf{K}_{\text {int,left }}^{-1} \mathbf{p}_{\text {im,left }}\right)=0
$$

$$
\mathbf{p}_{\text {im,right }}^{\mathrm{T}}(\underbrace{\left.\mathbf{K}_{\text {int,right }}^{-1}\right)^{T} \mathbf{E} \mathbf{K}_{\text {int,left }}^{-1}}) \mathbf{p}_{\text {im,left }}=0
$$

"Fundamental matrix" $\mathbf{F}$

$$
\mathbf{p}_{\text {im,right }}^{\mathrm{T}} \mathbf{F} \mathbf{p}_{i m, l e f t}=0 \quad \text { or } \quad \mathbf{p}^{T} \mathbf{F} \mathbf{p}^{\prime}=0
$$

## Properties of the Fundamental Matrix

$$
\mathbf{p}^{T} F \mathbf{p}^{\prime}=0
$$



- $\mathbf{l}=\mathbf{F} \mathbf{p}^{\prime}$ is the epipolar line associated with $\mathbf{p}^{\prime}$
$\cdot \mathbf{I}^{\prime}=\mathbf{F}^{T} \mathbf{p}$ is the epipolar line associated with $\mathbf{p}$
- Epipoles found by $\mathbf{F p}{ }^{\prime}=\mathbf{0}$ and $\mathbf{F}^{\mathrm{T}} \mathbf{p}=\mathbf{0}$
- You'll see more one these on the problem set to explain
- $\mathbf{F}$ is singular (mapping from 2-D point to 1-D family so rank 2 - more later)


## Fundamental matrix

- Relates pixel coordinates in the two views
- More general form than essential matrix: we remove need to know intrinsic parameters
- If we estimate fundamental matrix from correspondences in pixel coordinates, can reconstruct epipolar geometry without intrinsic or extrinsic parameters.



## Different Example: forward motion



## Computing F from correspondences

Each point correspondence $\mathbf{p}_{\text {im, right }}^{\mathrm{T}} \mathbf{F} \mathbf{p}_{\text {im,left }}=0$ generates one constraint on F

$$
\left[\begin{array}{lll}u^{\prime} & v^{\prime} & 1\end{array}\right]\left[\begin{array}{llll}f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33}\end{array}\right]\left[\begin{array}{c}u \\ v \\ 1\end{array}\right]=0
$$

Collect n of these $\left\lceil\left[\begin{array}{llllllll}u_{1}^{\prime} u_{1} & u_{1}^{\prime} v_{1} & u_{1}^{\prime} & v_{1}^{\prime} u_{1} & v_{1}^{\prime} v_{1} & v_{1}^{\prime} & u_{1} & v_{1}\end{array}\right]\left[\begin{array}{l}f_{11} \\ \text { constraints } \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33}\end{array}\right]=\mathbf{0}\right.$
Solve for f, vector of parameters.

## The (in)famous "eight-point algorithm"



- In principal can solve with 8 points.
- Better with more - yields homogeneous linear least-squares:
- Find unit norm vector F yielding smallest residual
- Remember SVD or substitute a 1 ?
- What happens when there is noise?


## Doing the obvious thing



## Rank of F



- Assume we know the homography $\mathrm{H}_{\pi}$ that maps from Left to Right (Full $3 \times 3$ )

$$
\mathbf{p}^{\prime}=\mathbf{H}_{\pi} \mathbf{p}
$$

- Let line l' be the epiloar line corresponding to $\mathbf{p}$ - goes through epipole $\mathbf{e}^{\mathbf{\prime}}$
- Rank of $F$ is rank of $\left[\mathbf{e}^{\prime}\right]_{\mathrm{x}}=2$


## Fix the linear solution

- Use SVD or other method to do linear computation for F
- Decompose F using SVD (not the same SVD):

$$
\mathbf{F}=U D V^{T}
$$

- Set the last singular value to zero:

$$
D=\left[\begin{array}{lll}
r & 0 & 0 \\
0 & s & 0 \\
0 & 0 & t
\end{array}\right] \Rightarrow \hat{D}=\left[\begin{array}{lll}
r & 0 & 0 \\
0 & s & 0 \\
0 & 0 & 0
\end{array}\right]
$$

- Estimate new F from the new $\hat{D}$

$$
\hat{\mathbf{F}}=U \hat{D} V^{T}
$$

## That's better...



## Stereo image rectification



## Stereo image rectification

- Reproject image planes onto a common plane parallel to the line between optical centers
- Pixel motion is horizontal after this transformation
- Two homographies ( $3 \times 3$ transform), one for each input image reprojection
> C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.


## Rectification Example

C. Loop and Z. Zhang, Computing Rectifying Homographies for Stereo Vision, IEEE Conf. Computer Vision and Pattern Recognition, 1999.

(a) Orignal image pair overlayed with several epipolar lines.
(b) Image pair transformed by the specialized projective mapping $\mathrm{H}_{H}$ and $\mathrm{H}_{p}^{\prime}$. Note that the epipolar lines are now parallel to each other in each image.
c) Image pair transformed by the similanity $\mathrm{H}_{r}$ and $\mathrm{H}_{\mathrm{r}}^{\prime}$. Note that the image pair is now rectified (the epipolar lines are horizontally aligned)
(d) Final image rectification after shearing transform H , and $\mathrm{H}^{\prime}$. Note that the image pair remains rectified, but the horizontal distortion is reduced

## Some example cool applications...

## Photo synth

Noah Snavely, Steven M. Seitz, Richard Szeliski, "Photo tourism: Exploring photo collections in 3D," SIGGRAPH 2006

http://photosynth.net/

## Photosynth.net



## Based on Photo Tourism

by Noah Snavely, Steve Seitz, and Rick Szeliski

## 3D from multiple images

Building Rome in a Day: Agarwal et al. 2009

## Summary

- For 2-views, there is a geometric relationship that define the relations between rays in one view to rays in the other
- Calibrated - Essential matrix
- Uncalibrated - Fundamental matrix.
- This relation can be estimated from point correspondences - both in calibrated cases and uncalibrated.
- Extensions allow combining multiple views to get more geometric information about scenes
- SLAM (simultaneous localization and mapping) - you'll hear about this (I hope!)

