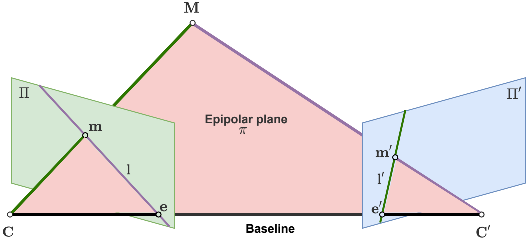
CS-565 Computer Vision

Nazar Khan

Department of Computer Science
University of the Punjab

20. Stereo Reconstruction

So far . . .



- Epipolar geometry relates points in one view to lines in the other view using the fundamental matrix

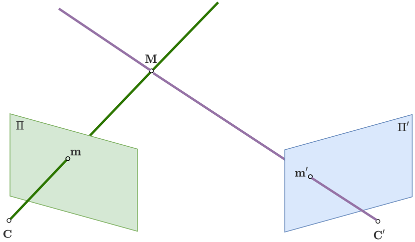
$$\mathbf{l}' = F\tilde{\mathbf{m}}$$

- Fundamental matrix can be estimated using linear methods from at least 8 point correspondences $\mathbf{m}_i \leftrightarrow \mathbf{m}'_i$.

In this lecture . . .

1. **Search**: Given point $\mathbf{m}$ in view 1 and its epipolar line $\mathbf{l}'$ in view 2, how can we find the corresponding point $\mathbf{m}'$ in view 2?

2. **Estimation**: Given corresponding points $\mathbf{m}, \mathbf{m}'$, how can we find the scene point $\mathbf{M}$?



Stereo Reconstruction via F

The fundamental-matrix method for reconstructing the scene is very simple, consisting of the following steps:

1. Given several point correspondences $\mathbf{m}_i \leftrightarrow \mathbf{m}'_i$ across two views, form linear equations in the entries of F based on the epipolar constraint $\tilde{\mathbf{m}}_i'^T F \tilde{\mathbf{m}}_i = 0$.
2. Find F as the solution to a set of linear equations.
3. Compute a pair of camera matrices (P, P') from F according to the simple formula

$$P = \begin{bmatrix} I & \mathbf{0} \end{bmatrix}$$
$$P' = \begin{bmatrix} [\mathbf{e}']_\times F & \mathbf{e}' \end{bmatrix}$$

4. Given the two cameras (P, P') and the corresponding image point pairs $\mathbf{m}_i \leftrightarrow \mathbf{m}'_i$, triangulate the 3$D$ points $\mathbf{M}_i$.

Stereo Reconstruction via F

Caution! Projective Ambiguity

- If H is any 4×4 invertible matrix, representing a projective transformation of $\mathbb{P}^3$, then replacing points $\tilde{\mathbf{M}}_i$ by $H\tilde{\mathbf{M}}_i$ and matrices P and P' by PH^{-1} and $P'H^{-1}$ does not change the image points.
- In other words, given correspondences $\tilde{\mathbf{m}}_i \leftrightarrow \tilde{\mathbf{m}}'_i$, the two different reconstructions of scene points and cameras

  $$\begin{array}{ccc} \tilde{\mathbf{M}}_i & & H\tilde{\mathbf{M}}_i \\ P & \text{and} & PH^{-1} \\ P' & & P'H^{-1} \end{array}$$

  yield the same image points.
- This shows that the points $\tilde{\mathbf{M}}_i$ and the cameras P and P' can be determined *at best* only up to a projective transformation.

Stereo Correspondence

- Given $\mathbf{m}$, describe surrounding region as a vector $\mathbf{x}$ which could be as simple as the raw pixel values in a window around $\mathbf{m}$.
- For every point $\mathbf{n}$ on epipolar line $\mathbf{l}'$, similarly describe surrounding region as a vector $\mathbf{y}$.
- The most deserving corresponding points should have the highest *normalized correlation*

$$\mathbf{m}' = \arg\max_{\mathbf{n}\in\mathbf{l}'} \frac{(\mathbf{x}-\bar{\mathbf{x}})^T}{\|\mathbf{x}-\bar{\mathbf{x}}\|} \frac{(\mathbf{y}-\bar{\mathbf{y}})}{\|\mathbf{y}-\bar{\mathbf{y}}\|}$$

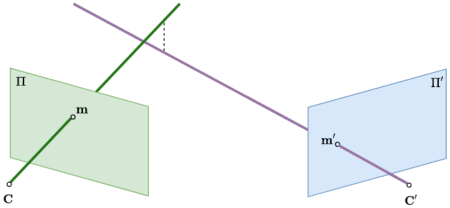
  which is just the dot-product of mean-centered, normalized vectors.
- Normalized correlation lies between -1 (when $\mathbf{x} = -\mathbf{y}$) and $+1$ (when $\mathbf{x} = \mathbf{y}$). It is 0 when $\mathbf{x} \perp \mathbf{y}$.

Stereo Correspondence

- **Weakness**: Using correlation of regions implicitly assumes that observed surface is locally parallel to both image planes.
- If surface is not parallel, then regions around *correct* corresponding points will have different content due to *foreshortening*.
- More sophisticated techniques, such as variational methods, can also be used to find corresponding points.

Triangulation

- In practice, correspondences $\mathbf{m} \leftrightarrow \mathbf{m}'$ are never perfect due to noise and quantization effects.
- Therefore back-projected rays will never intersect in $3D$.



- Instead, we find an *optimal estimate* of the scene point $\mathbf{M}$.

Triangulation

- Since $\tilde{\mathbf{m}} \equiv P\tilde{\mathbf{M}}$, vectors $\tilde{\mathbf{m}}$ and $P\tilde{\mathbf{M}}$ must point in the same direction. This gives us 3 equations

$$\tilde{\mathbf{m}} \times P\tilde{\mathbf{M}} = \mathbf{0}$$

$$\implies \begin{cases} x\left(\mathbf{p}^{3T}\tilde{\mathbf{M}}\right) - \left(\mathbf{p}^{1T}\tilde{\mathbf{M}}\right) = 0 \\ y\left(\mathbf{p}^{3T}\tilde{\mathbf{M}}\right) - \left(\mathbf{p}^{2T}\tilde{\mathbf{M}}\right) = 0 \\ x\left(\mathbf{p}^{2T}\tilde{\mathbf{M}}\right) - y\left(\mathbf{p}^{1T}\tilde{\mathbf{M}}\right) = 0 \end{cases}$$

  that are linear in entries of $\mathbf{M}$ and only 2 equations are linearly independent.
- Similarly, 2 equations constrain $\mathbf{M}$ from the second view.

$$x'\left(\mathbf{p}'^{3T}\tilde{\mathbf{M}}\right) - \left(\mathbf{p}'^{1T}\tilde{\mathbf{M}}\right) = 0$$

$$y'\left(\mathbf{p}'^{3T}\tilde{\mathbf{M}}\right) - \left(\mathbf{p}'^{2T}\tilde{\mathbf{M}}\right) = 0$$

Triangulation

- All 4 equations can be arranged as the homogenous linear system $A\tilde{\mathbf{M}} = \mathbf{0}$

$$\begin{bmatrix} x\mathbf{p}^{3T} - \mathbf{p}^{1T} \\ y\mathbf{p}^{3T} - \mathbf{p}^{2T} \\ x'\mathbf{p}'^{3T} - \mathbf{p}'^{1T} \\ y'\mathbf{p}'^{3T} - \mathbf{p}'^{2T} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = \mathbf{0}$$

- As in the DLT algorithm for homography estimation, $\tilde{\mathbf{M}}$ can be estimated as the eigenvector of $A^T A$ corresponding to the smallest eigenvalue.
- Notice how this method neatly generalizes to more than 2 views. For example, a third view P'' would have added 2 more rows in matrix A.

Summary

- Fundamental matrix is all we need.
- Camera matrices P and P' can be estimated from F upto a projective ambiguity.
- Search for corresponding points along epipolar lines can be performed using normalized correlation.
- Triangulation for scene point $\mathbf{M}$ can be performed via DLT.