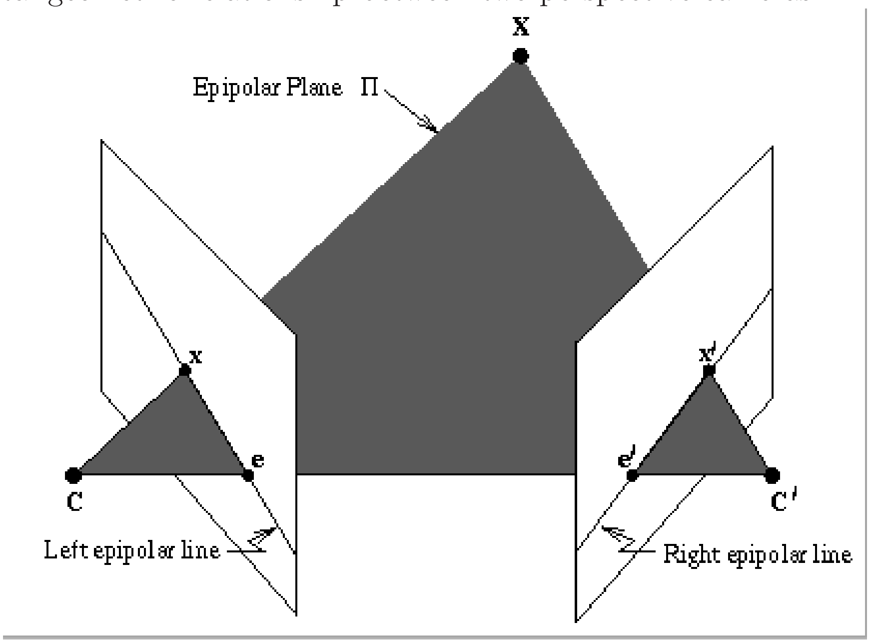
Multiple Views Geometry

Subhashis Banerjee
Dept. Computer Science and Engineering
IIT Delhi
email: suban@cse.iitd.ac.in

January 12, 2008

1 Epipolar geometry

Fundamental geometric relationship between two perspective cameras:



epipole: is the point of intersection of the line joining the optical centers - the *baseline* - with the image plane. The epipole is the image in one camera of the optical center of the other camera.

epipolar plane: is the plane defined by a 3D point and the optical centers. Or, equivalently, by an image point and the optical centers.

epipolar line: is the line of intersection of the epipolar plane with the image plane. It is the image in one camera of a ray through the optical center and the image point in the other camera. All epipolar lines intersect at the epipole.

Epipolar geometry provides a fundamental constraint for the correspondence problem

1.1 Epipolar geometry: uncalibrated case

- Given the two cameras 1 and 2, we have the camera equations:

$$\mathbf{x_1} = \tilde{\mathbf{P}}_\mathbf{1}\mathbf{X} \text{ and } \mathbf{x_2} = \tilde{\mathbf{P}}_\mathbf{2}\mathbf{X}$$

- The **optical center** projects as

$$\tilde{\mathbf{P_i}}\mathbf{X} = \mathbf{0}$$

- Writing

$$\tilde{\mathbf{P_i}} = [\mathbf{P_i} \mid -\mathbf{P_i}\mathbf{t_i}]$$

  where $\mathbf{P_i}$ is 3×3 non-singular we have that $\mathbf{t_i}$ is the optical center.

$$[\mathbf{P_i} \mid -\mathbf{P_i}\mathbf{t_i}] \begin{bmatrix} \mathbf{t_i} \\ 1 \end{bmatrix} = \mathbf{0}$$

- The epipole $\mathbf{e_2}$ in the second image is the projection of the optical center of the first image:

$$\mathbf{e_2} = \tilde{\mathbf{P_2}} \begin{bmatrix} \mathbf{t_1} \\ 1 \end{bmatrix}$$

- The projection of point on infinity along the optical ray $< \mathbf{t_1}, \mathbf{x_1} >$ on to the second image is given by:

$$\mathbf{x_2} = \mathbf{P_2}\mathbf{P_1}^{-1}\mathbf{x_1}$$

- The epipolar line $< \mathbf{e_2}, \mathbf{x_2} >$ is given by the cross product $\mathbf{e_2} \times \mathbf{x_2}$.
- If $[\mathbf{e_2}]_\times$ is the 3×3 antisymmetric matrix representing cross product with $\mathbf{e_2}$, then we have that the epipolar line is given by

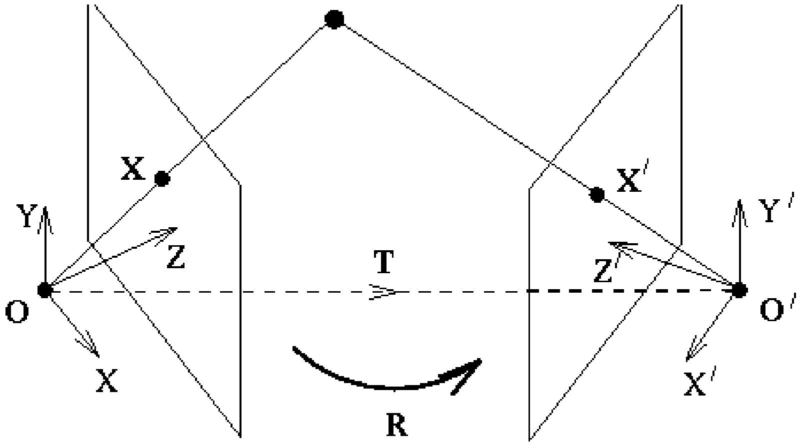
$$[e_2]_\times \mathbf{P_2}\mathbf{P_1}^{-1}\mathbf{x_1} = \mathbf{F}\mathbf{x_1}$$

- Any point $\mathbf{x_2}$ on this epipolar line satisfies

$$\mathbf{x_2}^T\mathbf{F}\mathbf{x_1} = 0$$

- $\mathbf{F}$ is called the **fundamental matrix**. It is of rank 2 and can be computed from 8 point correspondences.
- Clearly $\mathbf{F}\mathbf{e_1} = \mathbf{0}$ (degenerate epipolar line) and $\mathbf{e_2}^T\mathbf{F} = \mathbf{0}$. **The epipoles are obtained as the null spaces of $\mathbf{F}$.**

1.2 Epipolar geometry: calibrated case



- There are two camera coordinate systems related by $\mathbf{R}, \mathbf{T}$

$$\mathbf{X}' = \mathbf{R}\mathbf{X} + \mathbf{T}$$

- Taking the vector product with $\mathbf{T}$ followed by the scalar product with $\mathbf{X}'$

$$\mathbf{X}' \cdot (\mathbf{T} \times \mathbf{R}\mathbf{X}) = 0$$

which expresses that vectors $\mathbf{OX}$, $\mathbf{O'X'}$ and $\mathbf{OO'}$ are coplanar.

- This can be written as

$$\mathbf{X'}^T \mathbf{E}\mathbf{X} = 0$$

where

$$\mathbf{E} = [\mathbf{T}]_\times \mathbf{R}$$

is the **Essential matrix**.

- Image points and rays in Euclidean 3-space are related by:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{C} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \text{ and } \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{C}' \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}$$

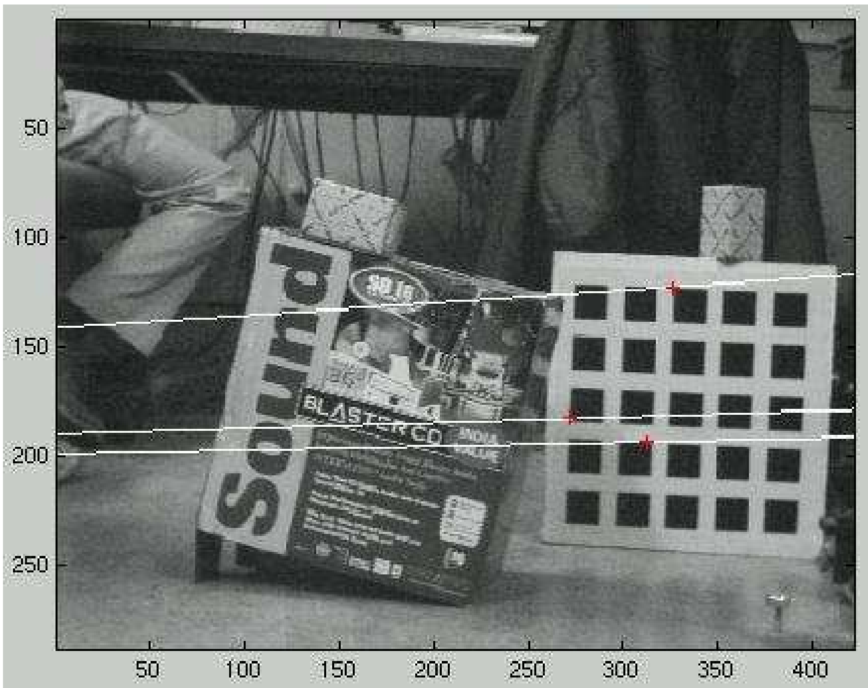
- Hence, we have

$$\mathbf{x'}^T \mathbf{C}'^{-T} \mathbf{E} \mathbf{C}^{-1} \mathbf{x} = 0$$

- Thus, the relation between the essential and fundamental matrix is:

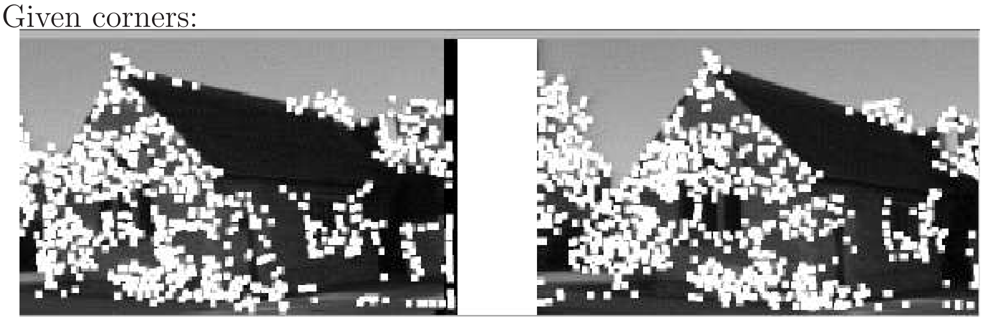
$$\mathbf{F} = \mathbf{C}'^{-T} \mathbf{E} \mathbf{C}^{-1}$$

1.3 Epipolar geometry: examples

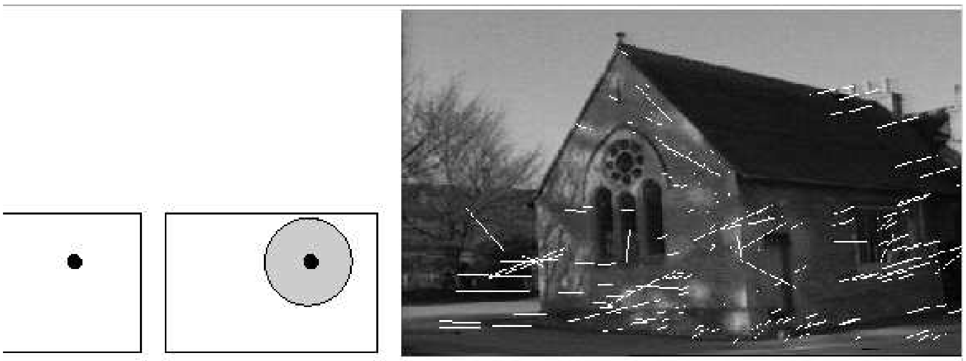


1.4 How to find correspondences

1. Given corners:

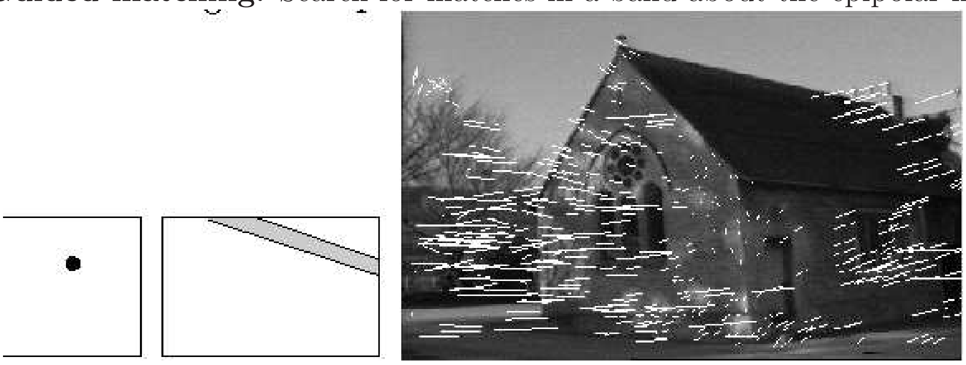


2. **Unguided matching:** Obtain a small number of seed matches using cross-correlation and pessimistic thresholds.



3. Compute epipolar geometry and reject outliers.

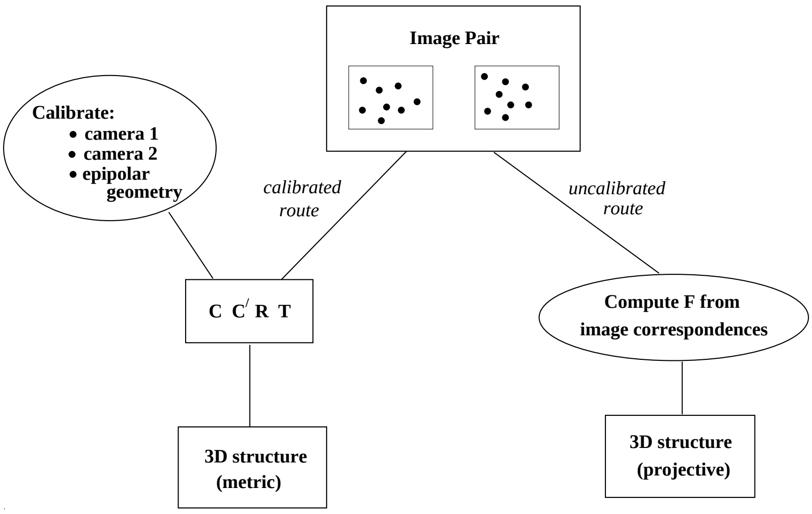
4. **Guided matching:** Search for matches in a band about the epipolar lines.



2 Structure determination

2.1 Overview

Overview



2.2 Projective structure using a five point basis

- Select the projection center of the first camera as the coordinate origin $(0,0,0,1)^T$ and that of the second camera as the unit point $(1,1,1,1)^T$.
- Complete the 3D basis by choosing 3 other visible 3D points $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$ such that no four of the five points are coplanar.
- Let $\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$ and $\mathbf{a_1'}$, $\mathbf{a_2'}$, $\mathbf{a_3'}$ be the image projections of $\mathbf{A_1}$, $\mathbf{A_2}$ and $\mathbf{A_3}$ in the two images.
- Make a projective transformation of each image such that these 3 points and the epipoles form a standard basis

$$\mathbf{a_1} = \mathbf{a_1'} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a_2} = \mathbf{a_2'} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a_3} = \mathbf{a_3'} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{e} = \mathbf{e'} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- Also fix the 3D coordinates as

$$\mathbf{A_1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{A_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{A_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

- It follows that

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{P'} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

- Once $\mathbf{P}$'s are known finding the projective structure of the other points is straightforward.

Localization errors in any of the points affect subsequent computation

2.3 Robust computation of projective structure

- Choose the first camera as

$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$$

  which corresponds to the world coordinate system having its origin at the optical center of the first camera, and its axes aligned with the camera axes. It is always possible to make this choice.

- The choice of the second camera consistent with the fundamental matrix is

$$\mathbf{P}' = [\mathbf{M}' \mid \mathbf{e}']$$

and the fundamental matrix is given as:

$$\mathbf{F} = \mathbf{e}'_{\times}\mathbf{M}'$$

- $\mathbf{e}'$ is the projection of the optical center of the first camera, $(\mathbf{0}^T, 1)^T$, onto the second image

$$\mathbf{P}' \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \mathbf{e}'$$

- Similarly, the optical center of the second camera is is $((\mathbf{M}'^{-1}\mathbf{e}')^T, 1)^T$, and its projection onto the first image is $\mathbf{e}$.

$$\mathbf{P} \begin{bmatrix} \mathbf{M}'^{-1}\mathbf{e}' \\ 1 \end{bmatrix} = \mathbf{M}'^{-1}\mathbf{e}' = \mathbf{e}$$

- Given $\mathbf{P}$ and $\mathbf{P}'$, the 3D point $\mathbf{X_i}$ corresponding to a pair of image point $\mathbf{x_i}$ and $\mathbf{x'_i}$ can be computed by intersecting the back-projected rays

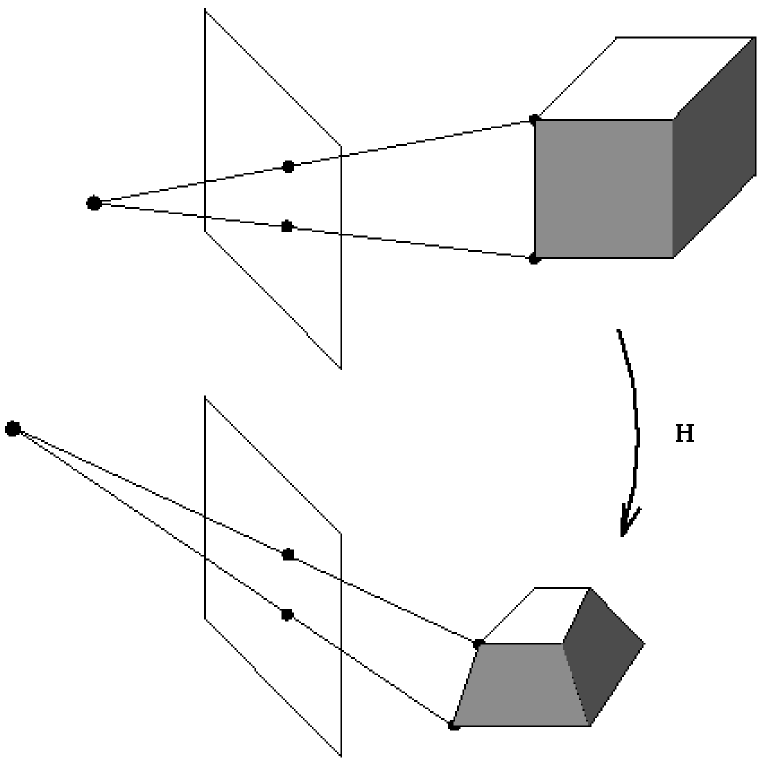
$$\begin{aligned} \mathbf{X_i} &= \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} + \lambda \begin{bmatrix} \mathbf{x_i} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{M}'^{-1}\mathbf{e}' \\ 1 \end{bmatrix} + \mu \begin{bmatrix} \mathbf{M}'^{-1}\mathbf{x_i} \\ 0 \end{bmatrix} \end{aligned}$$

To summarize:

1. Compute the fundamental matrix $\mathbf{F}$ from $\mathbf{x_i} \leftrightarrow \mathbf{x'_i}$.
2. Decompose $\mathbf{F}$ as $\mathbf{F} = \mathbf{e}'_{\times}\mathbf{M}'$.
3. Compute 3D points $\mathbf{X_i}$ by intersecting back-projected rays using

$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}] \text{ and } \mathbf{P}' = [\mathbf{M}' \mid \mathbf{e}']$$

2.4 Projective ambiguity



- Although the two projection matrices define $\mathbf{F}$ the converse is not true.
- Suppose $\mathbf{P}$ and $\mathbf{P}'$ are two matrices consistent with $\mathbf{F}$, then

$$\mathbf{x} = \mathbf{P}\mathbf{X} \text{ and } \mathbf{x}' = \mathbf{P}'\mathbf{X}'$$

- But, if $\mathbf{H}$ is any arbitrary homography of 3-space, then

$$\mathbf{x} = (\mathbf{P}\mathbf{H}^{-1})(\mathbf{H}\mathbf{X}) \text{ and } \mathbf{x}' = (\mathbf{P}'\mathbf{H}^{-1})(\mathbf{H}\mathbf{X}')$$

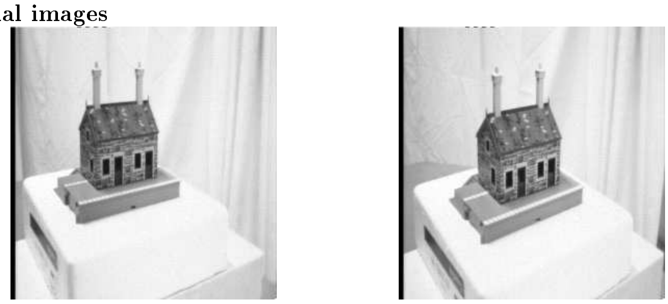
- Thus, we can only recover projective structure modulo an unknown homography.

How to recover $\mathbf{H}_{\mathbf{PE}}$, the homography that upgrades the projective structure to Euclidean?

2.5 Examples of projective reconstruction

Uncalibrated reconstruction example

Original images



Projective/affine reconstructions

Euclidean reconstructions

3 Planes

3.1 Planes and epipolar geometry

- The homography between two planes is represented by a 3×3 matrix

$$\tilde{\mathbf{x}}' = \mathbf{T}\mathbf{x}$$

- The epipoles are image projections of a 3D point lying on all planes, because the line joining the optical centers intersects any plane.

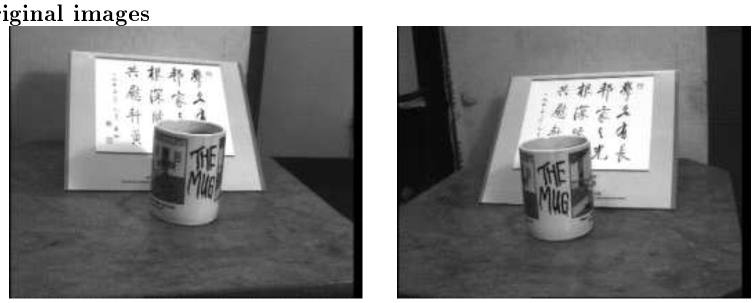
$$\mathbf{e}' = \mathbf{T}\mathbf{e}$$

- Four coplanar points and two points off the plane determine the epipolar geometry.

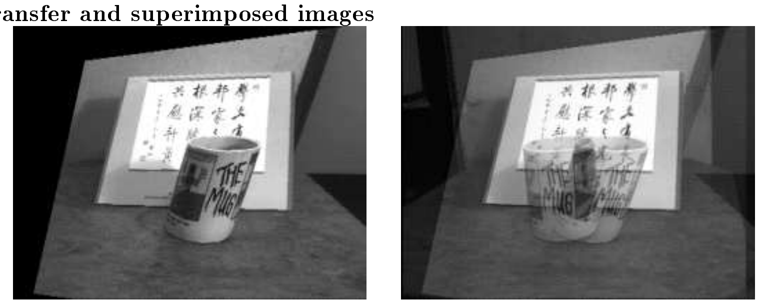
  1. Compute the plane projective transformation $\mathbf{T}$, such that $\tilde{\mathbf{x}}_\mathbf{i}{}' = \mathbf{T}\mathbf{x}_\mathbf{i}, \;\; i = 1 \ldots 4$.
  2. Determine the epipole $\mathbf{e}'$ in the $\mathbf{x}'$ image as the intersection of the lines $(\mathbf{T}\mathbf{x}_\mathbf{5} \times \mathbf{x}'_\mathbf{5})$ and $(\mathbf{T}\mathbf{x}_\mathbf{6} \times \mathbf{x}'_\mathbf{6})$.
  3. The epipolar line in the $\mathbf{x}'$ image of any other point $\mathbf{x}$ is given as $(\mathbf{T}\mathbf{x} \times \mathbf{e}')$.
  4. $\mathbf{F} = \mathbf{e}'_\times \mathbf{T}$

Plane Projective Transfer

Original images



Transfer and superimposed images



3.2 Distinguished planes and F

- Let $\mathbf{T}^*$ be a homography between two images through a plane π. Thus we have (projectively)
$$\mathbf{x}' = \mathbf{T}^*\mathbf{x}$$

- Given $\mathbf{F}$ we know $\mathbf{e}$ and $\mathbf{e}'$.

- Now if $\mathbf{F} = [\mathbf{e}']_\times \mathbf{T}^*$ is a valid decomposition, then so is $\mathbf{F} = [\mathbf{e}']_\times \mathbf{T}$, where
$$\mathbf{T} = \mathbf{T}^* + \mathbf{e}'\mathbf{a}^T$$
where $\mathbf{a}$ is an arbitrary 3 vector.

- Thus, the most general projection matrices consistent with $\mathbf{F}$ are
$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}] \text{ and } \mathbf{P}' = [\mathbf{T} + \mathbf{e}'\mathbf{a} \mid b\mathbf{e}']$$
This is four parameter family, b is an arbitrary scalar.

- The matrix $\mathbf{T} = \mathbf{T}^* + \mathbf{e}'\mathbf{a}^T$ is a projective transformation of the plane π represented by the four vector $(0, 0, 0, 1)^T$, since for points on this plane $\underline{X}_4 = 0$, and
$$\mathbf{x} = \mathbf{P}\begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = [\mathbf{I} \mid \mathbf{0}]\begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$
$$\mathbf{x}' = \mathbf{P}'\begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = [\mathbf{T} \mid b\mathbf{e}']\begin{bmatrix} X \\ Y \\ Z \\ 0 \end{bmatrix} = \mathbf{T}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{T}\mathbf{x}$$

- Thus, the choice of $\mathbf{a}$ corresponds to which actual plane in 3-space is π_∞.

4 Upgrading projective structure to Euclidean: Self Calibration and stratification

4.1 From projective to affine

- Identifying the π_∞ is equivalent to updating the projective structure to affine structure.

- An affine reconstruction is where the point coordinates $\{\mathbf{X^A}\}$ are known up to an affine transformation of their Euclidean values $\{\mathbf{X^E}\}$

$$\mathbf{X^A} = \mathbf{T_A X^E} = \begin{bmatrix} \mathbf{A} & \mathbf{t_A} \\ \mathbf{0} & 1 \end{bmatrix}$$

  $\{\mathbf{A}, \mathbf{t_A}\}$ are unknown and same for all points.

- Choosing $\mathbf{a}$ corresponding to the 'real' plane at infinity updates the projective reconstruction to affine.

- $\mathbf{T}$ then is the **infinite homography**, $\mathbf{H}_\infty$, the homography through the plane at infinity.

4.2 From affine to metric

- Suppose that

$$\mathbf{P_E} = \mathbf{C}\,[\mathbf{I} \mid \mathbf{0}] \text{ and } \mathbf{P'_E} = \mathbf{C}\,[\mathbf{R} \mid \mathbf{t}]$$

  i.e., we assume that the two camera positions have fixed internals.

- Suppose, also, that we have recovered affine structure $\{\mathbf{X^A}\}$ and a pair of projection matrices

$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}] \text{ and } \mathbf{P'} = [\mathbf{T} \mid \mathbf{e'}]$$

- Then we know that $\mathbf{T} = \mathbf{H}_\infty$.

- We also know that

$$\mathbf{P_E} = \mathbf{P}\mathbf{H}^{-1} \text{ and } \mathbf{P'_E} = \mathbf{P'}\mathbf{H}^{-1}$$

  where,

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{C}^{-1} & \mathbf{0} \\ \mathbf{0} & k \end{bmatrix}$$

  is the affine transformation that upgrades the affine reconstruction to Euclidean.

- The above gives us

$$\mathbf{T} = \mathbf{H}_\infty = \mathbf{C}\mathbf{R}\mathbf{C}^{-1}$$

- Rearranging and using $\mathbf{R}\mathbf{R}^T = \mathbf{I}$, we have

$$\mathbf{K} = \mathbf{T}^{-1}\mathbf{K}\mathbf{T}^T$$

  where

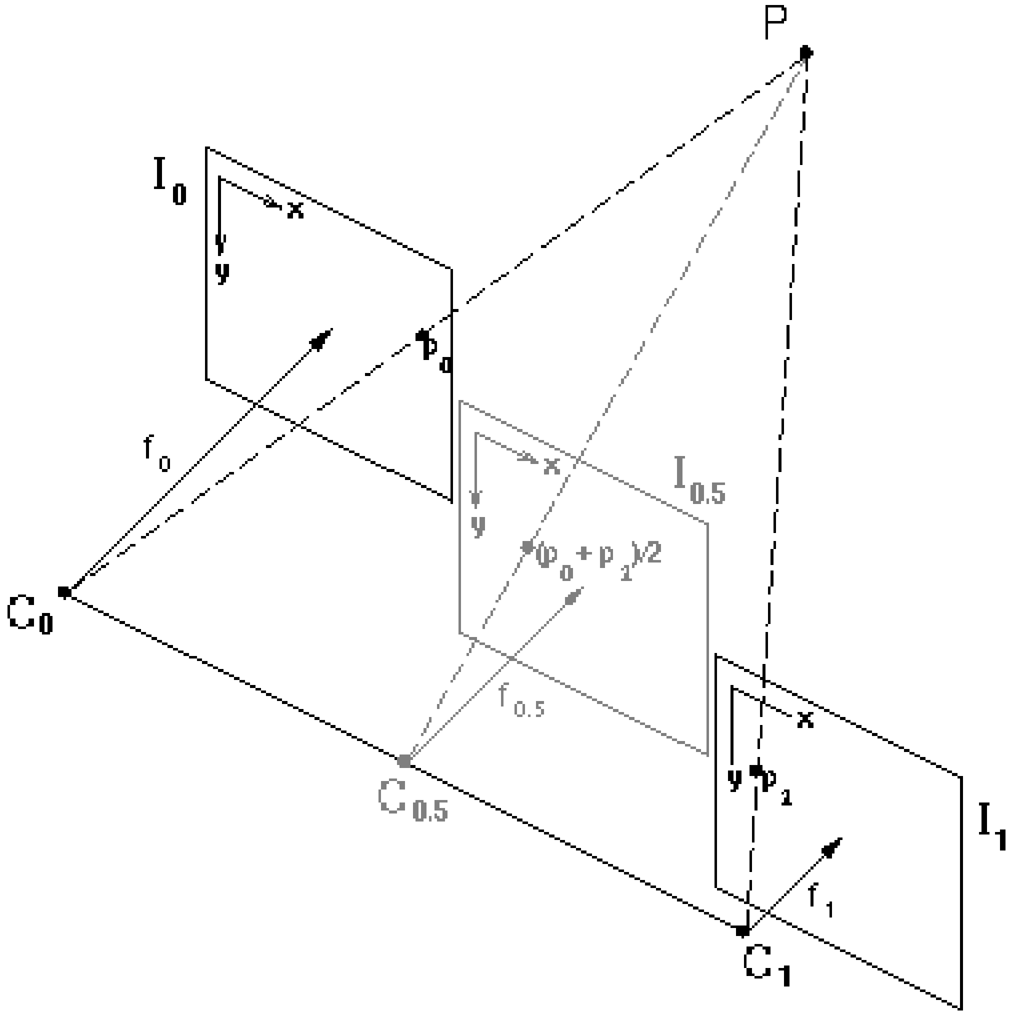
$$\mathbf{K} = \mathbf{C}\mathbf{C}^T$$

- Thus, the internal parameter matrix can be computed from the infinite homography using Cholesky decomposition.

5 Some Applications

5.1 Novel view generation using view morphing

[Seitz and Dyer, 1996]
Parallel views



- Suppose the camera has moved from the world origin to the position $(C_x, C_y, 0)$ and the two projection matrices are

$$\mathbf{\Pi_0} = \begin{bmatrix} f_0 & 0 & 0 & 0 \\ 0 & f_0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and

$$\mathbf{\Pi_1} = \begin{bmatrix} f_1 & 0 & 0 & -f_1 C_x \\ 0 & f_1 & 0 & -f_1 C_y \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- Let $\mathbf{p_0}$ and $\mathbf{p_1}$ are image projections of a scene point $\mathbf{P}$.
- Linear interpolation of $\mathbf{p_0}$ and $\mathbf{p_1}$ yields

$$(1-s)\mathbf{p_0} + s\mathbf{p_1} = (1-s)\frac{1}{Z}\mathbf{\Pi_0 P} + s\frac{1}{Z}\mathbf{\Pi_1 P} = \frac{1}{Z}\mathbf{\Pi_s P}$$

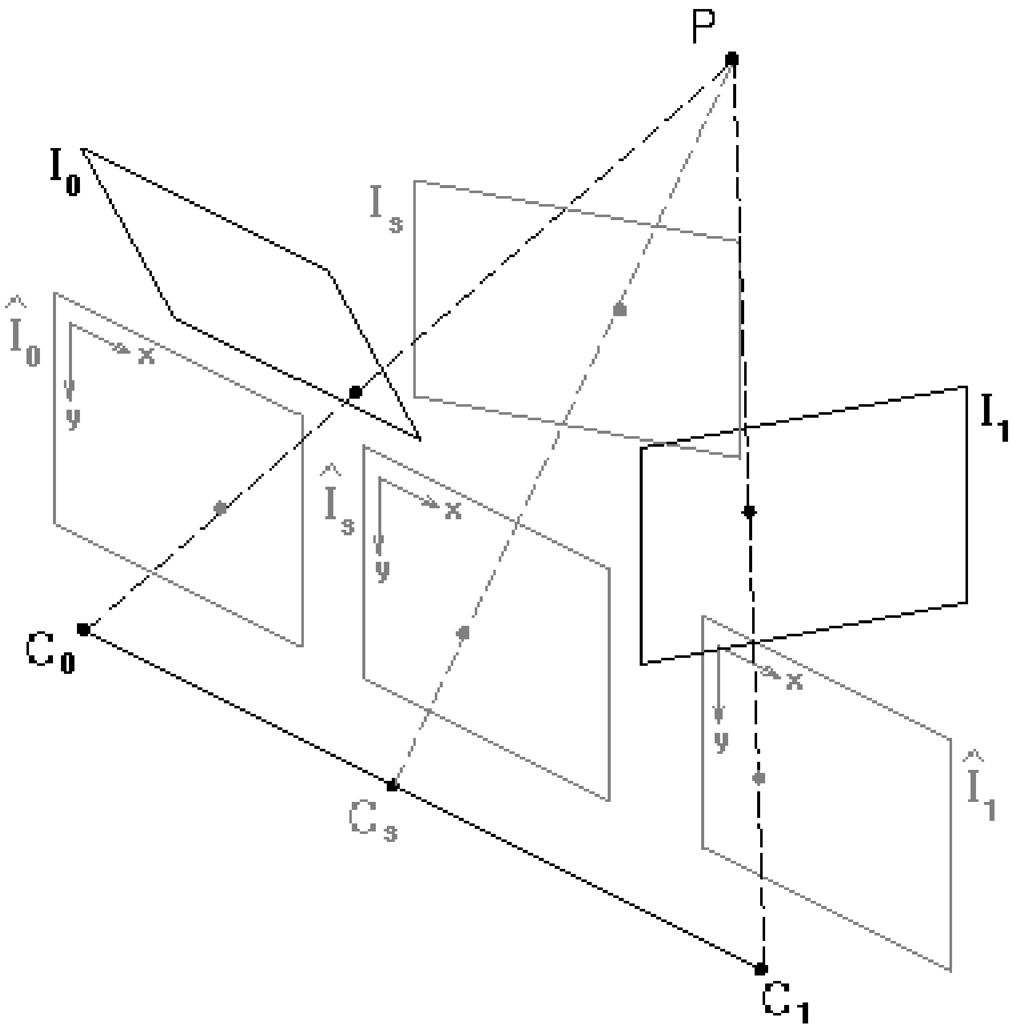
where

$$\mathbf{\Pi_s} = (1-s)\mathbf{\Pi_0} + s\mathbf{\Pi_1}$$

- $\mathbf{\Pi_s}$ represents a camera with center $C_s = (sC_x, sC_y, 0)$ and focal length $f_s = (1-s)f_0 + sf_1$.

Linear morph is consistent with a physical pin-hole camera.

Non-parallel camera



- When two images share the optical center they are related by a planar homography.
- Thus there exist planar homographies between $\mathcal{I}_0 \leftrightarrow \hat{\mathcal{I}}_0$, $\mathcal{I}_1 \leftrightarrow \hat{\mathcal{I}}_1$ and $\mathcal{I}_s \leftrightarrow \hat{\mathcal{I}}_s$.
- View morphing can be achieved as a 3 step process.

The results also hold for uncalibrated cameras.
Demo
Monalisa

5.2 Video compression

Tracking triangulation Sequence 1 Sequence 2

5.3 Panoramas

- When the camera rotates about its optical center or undergoes pure translation then the views are related by a planar homography.
- Recovering the homography from tracking allows us to stitch the images in to an image mosaic.

Demo