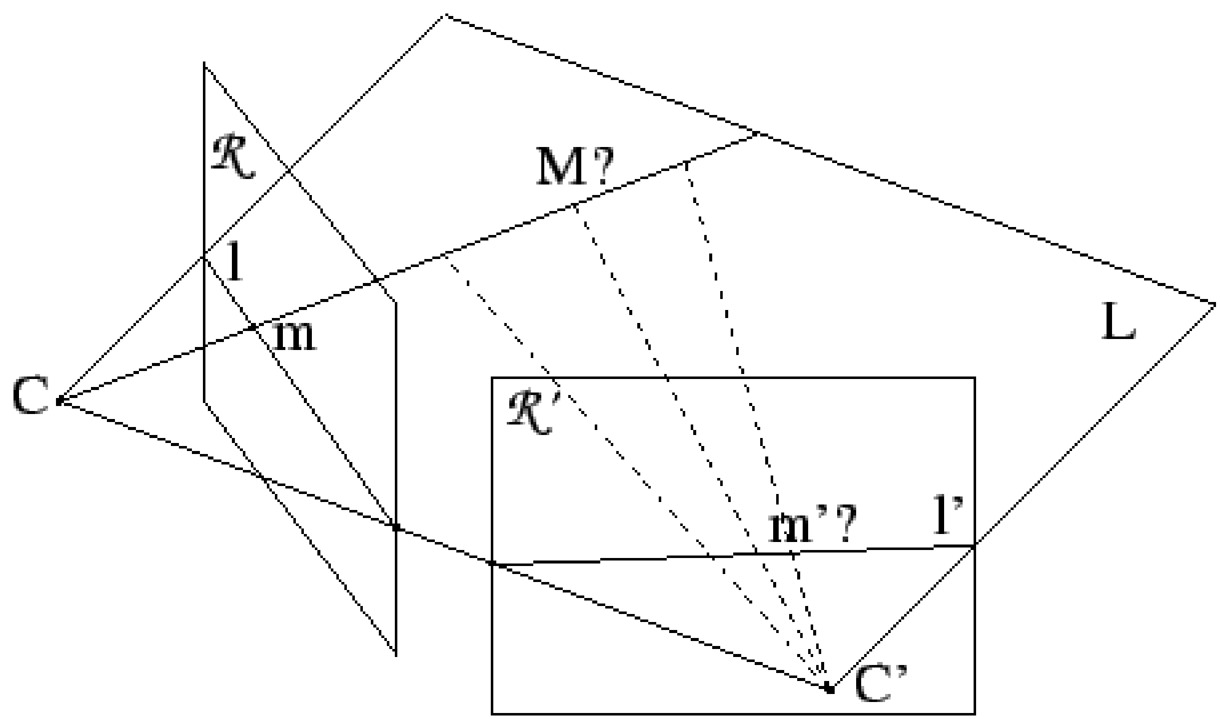
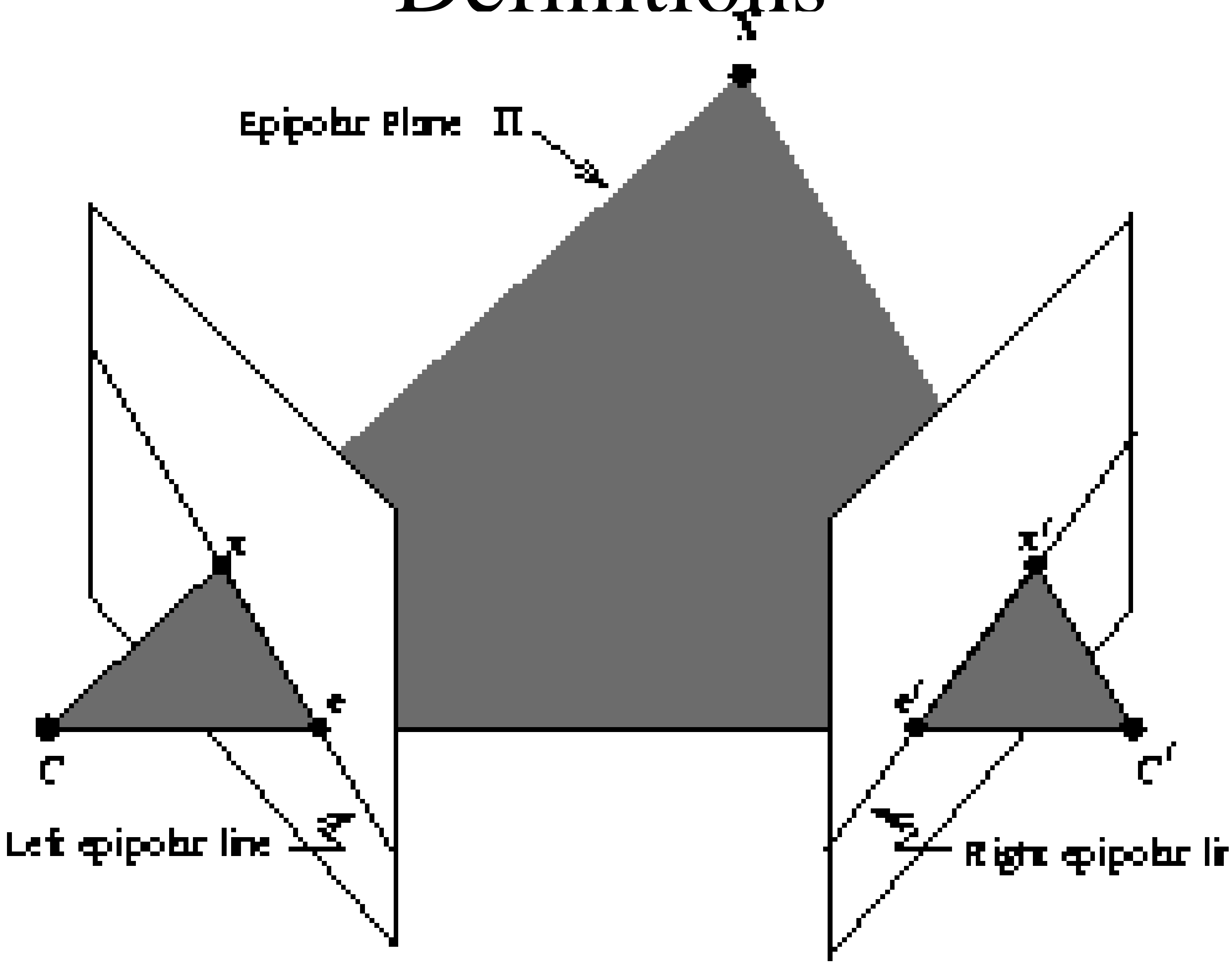
Relations between image coordinates

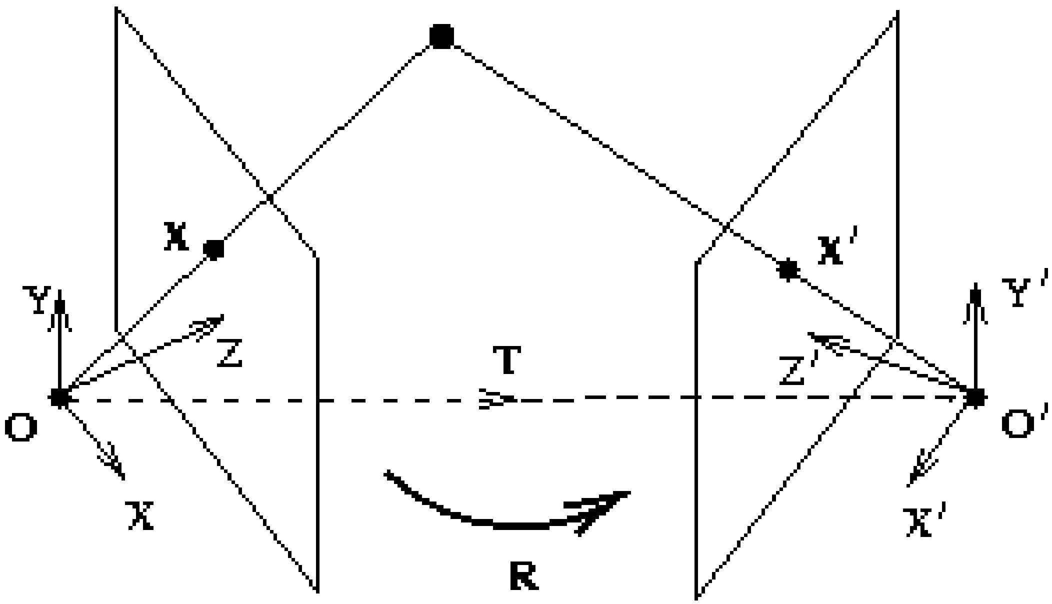
Given coordinates in one image, and the tranformation
Between cameras, T = [R t], what are the image coordinates
In the other camera's image.



Definitions



Essential Matrix: Relating between image coordinates



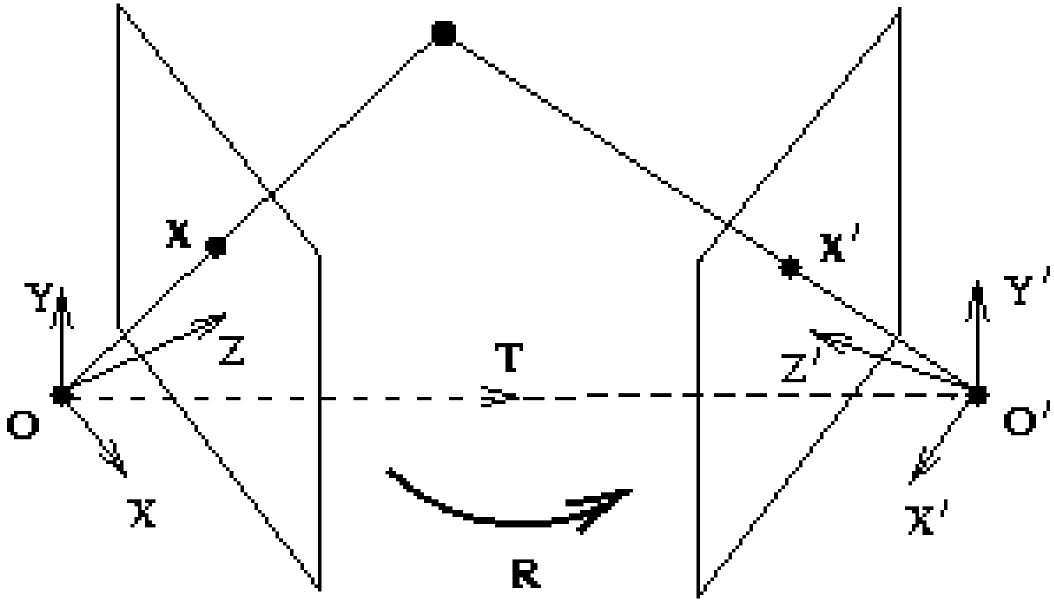
$$\vec{Ox} \quad \vec{O'x'} \quad \vec{OO'}$$

Are Coplanar, so:

$$\vec{O'x'} \cdot \left[\vec{Ox} \times \vec{OO'} \right] = 0$$

camera coordinate systems, related by a rotation **R** and a translation **T**:

$$x' = \begin{bmatrix} R & t \\ 000 & 1 \end{bmatrix} x$$



$$\mathbf{a} \times \mathbf{v} = \begin{bmatrix} a_y v_z - a_z v_y \\ a_z v_x - a_x v_z \\ a_x v_y - a_y v_x \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$$

$$= \mathbf{A}\mathbf{v}$$

$$x' = \begin{bmatrix} R & t \\ 000 & 1 \end{bmatrix} x$$

$$\vec{O'x'} \cdot \left[\vec{Ox} \times \vec{OO'} \right] = 0$$

$$x' \cdot \left(\vec{t} \times Rx \right) = 0$$

$$x' \cdot \left(E x \right) = 0$$

$$x' E x = 0$$

$$E = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} R$$

What does the Essential matrix do?

It represents the normal to the epipolar line *in the other image*

$$n = E\,x$$

The normal defines a line in image 2:

$$x'_{on\ epipolar\ line} \Rightarrow \quad n \cdot x' = 0$$

$$n_1 x_1 + n_2 x_2 + n_3 1 = 0$$

$$(y = mx + b) \Rightarrow \quad b = -n_3, \quad m = -\frac{n_1}{n_2}$$

What if cameras are uncalibrated? Fundamental Matrix

Choose world coordinates as Camera 1.
Then the extrinsic parameters for camera 2 are just **R** and **t**
However, intrinsic parameters for both cameras are unknown.
Let C_1 and C_2 denote the matrices of intrinsic parameters. Then the pixel coordinates measured are not appropriate for the Essential matrix.
Correcting for this distortion creates a new matrix: the Fundamental Matrix.

$$x'_{measured} = C_2 x' \qquad x_{measured} = C_1 x$$

$$(x')^t E\, x = 0 \Rightarrow \left(C_2^{-1} x'_{measured}\right)^t E \left(C_1^{-1} x_{measured}\right) = 0$$

$$(x'_{measured})^t F\, x_{measured} = 0$$

$$F = C_2^{-t} E\, C_1^{-1}$$

$$C = \begin{bmatrix} -f \cdot s_u & 0 & u_0 \\ 0 & -f \cdot s_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Computing the fundamental Matrix

Computing : I Number of Correspondences Given perfect image points (no noise) in general position. Each point correspondence generates one constraint on the fundamental matrix

Constraint for one point

$$\begin{bmatrix} x'_i & y'_i & 1 \end{bmatrix} \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

Each constraint can be rewritten as a dot product.
Stacking several of these results in:

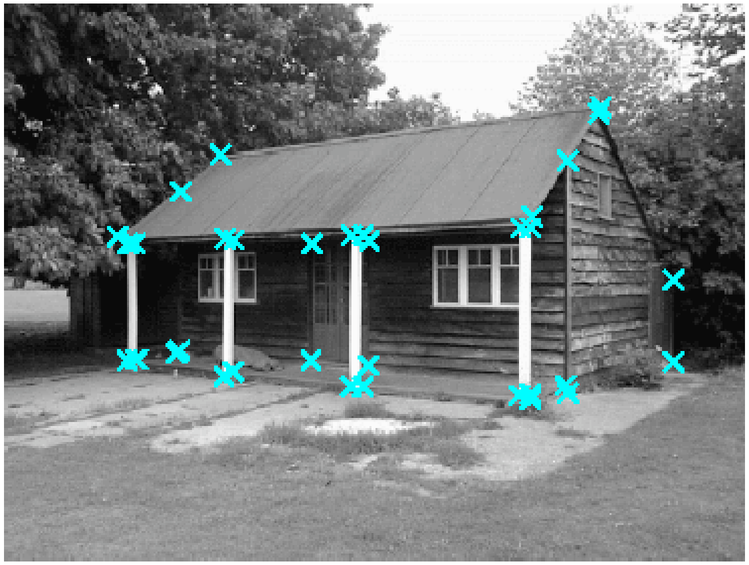
$$\begin{bmatrix} x'_1 x_1 & x'_1 y_1 & x'_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_n x_n & x'_n y_n & x'_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \end{bmatrix} = 0$$

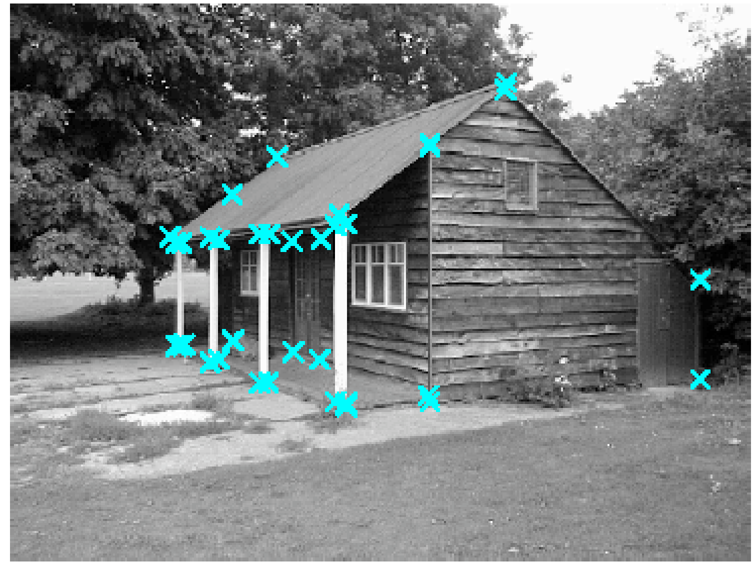
Stereo Reconstruction

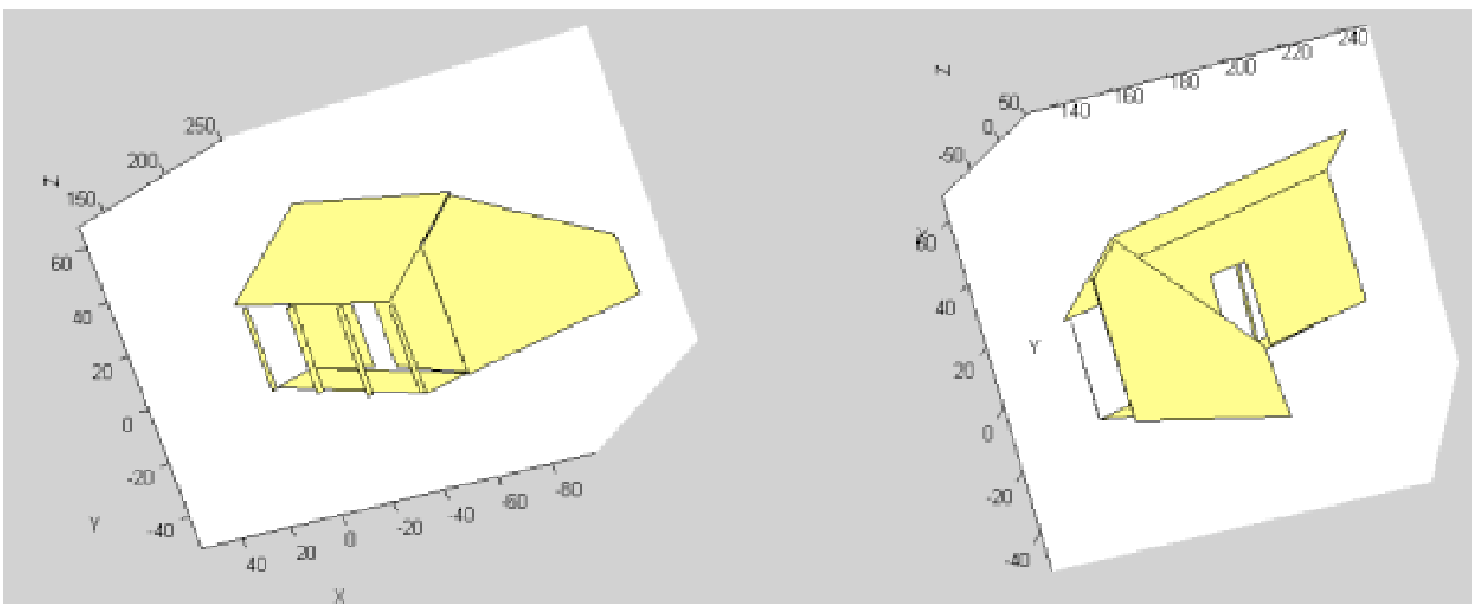
If we know the
- fundamental matrix, and
- internal camera parameters

We can solve for the
- external camera parameters, and
- determine 3D structure of a scene

> 8 Point matches







- This is called *calibrated reconstruction*, since it requires the camera(s) be $\underbrace{\textit{internally calibrated}}$.

  camera calibration matrix C must be known.

- For example:
  - Reconstructing the 3D structure of a scene from multiple views.
  - Determining structure from motion.

Reconstruction Steps

1. Identify a number of (at least 8) point correspondances.
2. Estimate the fundamental matrix using the normalised 8-point algorithm.
3. Determine the external camera parameters (rotation and translation from one camera to the other)
   a. Calculate the essential matrix from the fundamental matrix and the camera calibration matrices.
   b. Extract the rotation and translation components from the essential matrix.
4. Determine 3D point locations.

Determining Extrinsic Camera Parameters

- Want to determine rotation and translation from one camera to the other.
- We know the camera matrices have the form

$$M_1 = C_1\left[I \quad \vec{0}\right]$$

$$M_2 = C_2\left[R \quad \vec{t}\right]$$

> Why can we just use this as the external parameters for camera M1?
> Because we are only interested in the *relative* position of the two cameras.

- First we undo the Intrinsic camera distortions by defining new *normalized* cameras

$$M_1^{norm} = C_1^{-1} M_1 \quad \text{and} \quad M_2^{norm} = C_2^{-1} M_2$$

Determining Extrinsic Camera Parameters

- The *normalized* cameras contain unknown parameters

$$M_1^{norm} = C_1^{-1} M_1 \quad \Rightarrow \quad M_1^{norm} = \begin{bmatrix} I & \vec{0} \end{bmatrix}$$

$$M_2^{norm} = C_2^{-1} M_2 \quad \Rightarrow \quad M_2^{norm} = \begin{bmatrix} R & \vec{t} \end{bmatrix}$$

- However, those parameters can be extracted from the Fundamental matrix

$$\mathcal{F} = C_2^{-t} \mathcal{E} C_1^{-1}$$

$$\mathcal{E} = C_2^{t} \mathcal{F} C_1$$

$$\mathcal{E} = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} R = \vec{t}_{\times} R$$

Extract t and R from the Essential Matrix

How do we recover *t* and *R*? Answer: SVD of $\mathcal{E}$

$$\mathcal{E} = USV^t$$

S diagonal

U,V orthogonal and $\det() = 1$ (rotation)

$$R = UWV^t \quad \text{or} \quad R = UW^tV^t \qquad \vec{t} = u_3 \quad \text{or} \quad \vec{t} = -u_3$$

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

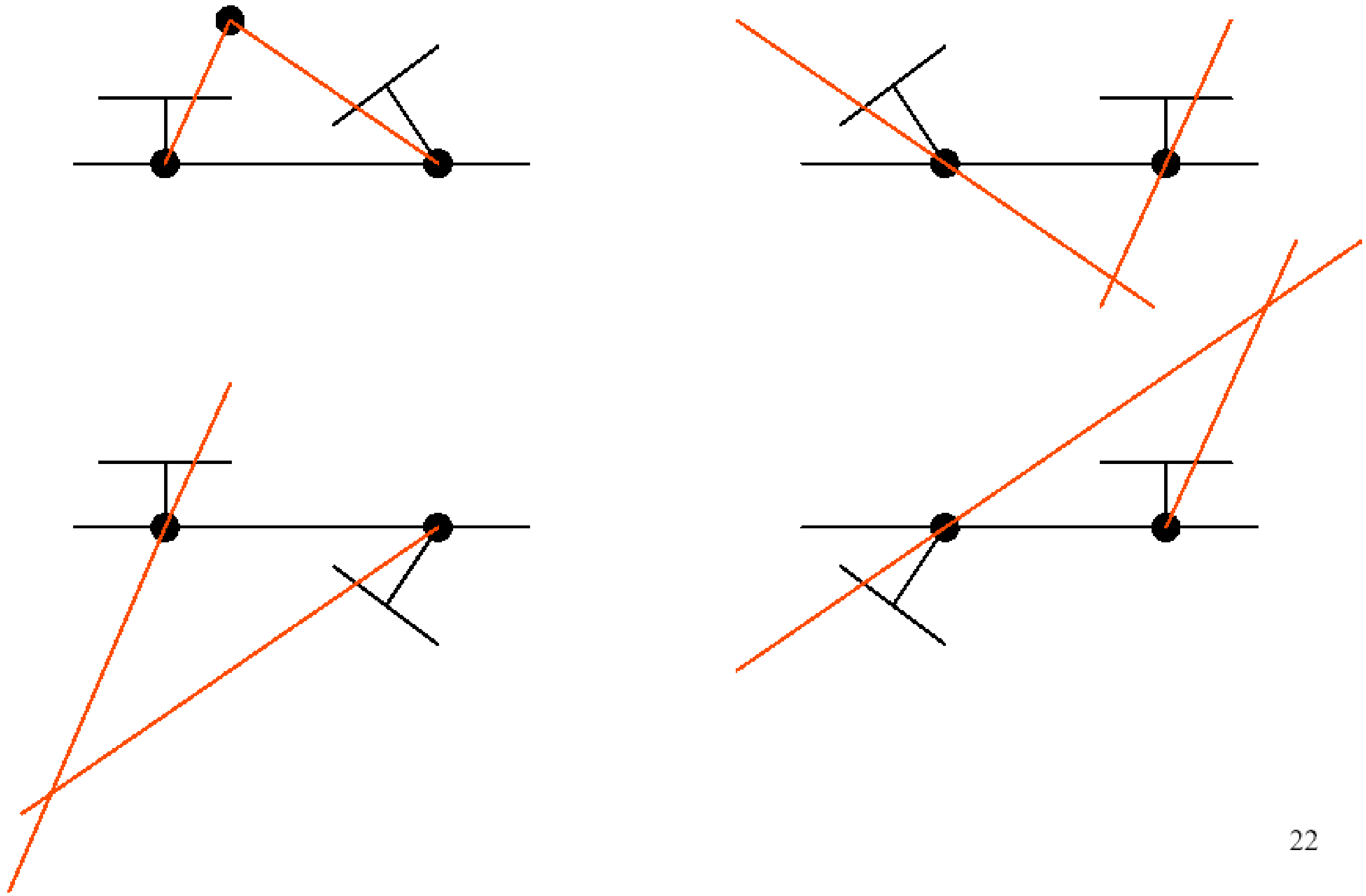
Reconstruction Ambiguity

So we have 4 possible combinations of translations and rotations giving 4 possibilities for $\boldsymbol{M}_2^{norm} = [\boldsymbol{R} \mid \boldsymbol{t}]$

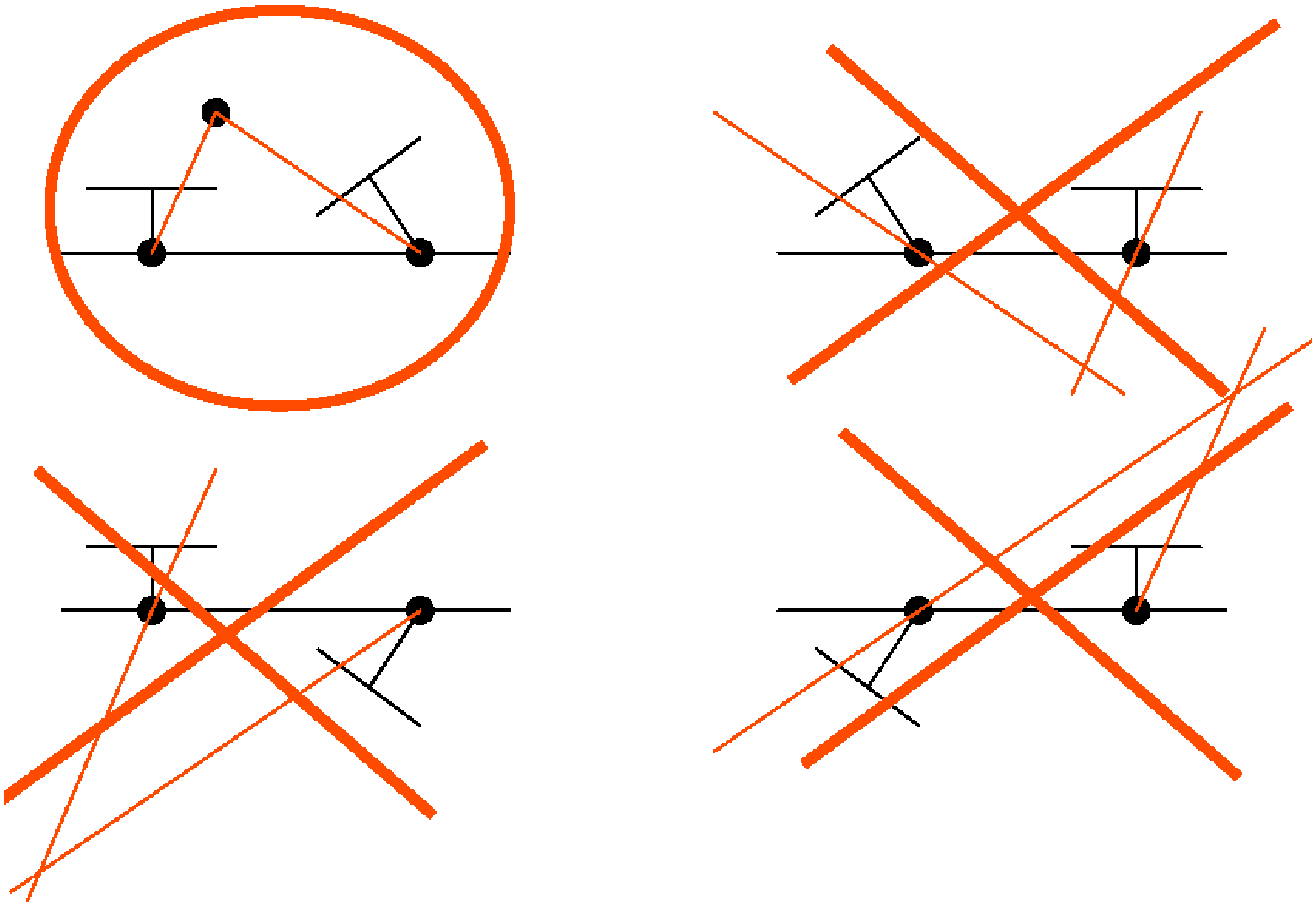
1. $\boldsymbol{M}_2^{norm} = [\boldsymbol{U}\boldsymbol{W}^t\boldsymbol{V}^t \mid \boldsymbol{t}]$
2. $\boldsymbol{M}_2^{norm} = [\boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^t \mid \boldsymbol{t}]$
3. $\boldsymbol{M}_2^{norm} = [\boldsymbol{U}\boldsymbol{W}^t\boldsymbol{V}^t \mid -\boldsymbol{t}]$
4. $\boldsymbol{M}_2^{norm} = [\boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^t \mid -\boldsymbol{t}]$

Which one is right?

- We can determine which of these is correct by looking at their geometric interpretation.



Both Cameras must be facing the same direction



Which one is right?

- The correct pair will have our data points in front of both cameras.
- How do we choose the correct pair?
- Procedure:
  - Take a test point from data
  - Backproject to find 3D location
  - Determine the depth of 3D point in both cameras
  - Choose the camera pair that has a positive depth for both cameras.

How do we backproject?

$$x'_{measured} = C_2 x' \qquad x_{measured} = C_1 x$$

Knowing C_i allows us to determine the undistorted image points:

$$x' = C_2^{-1} x'_{measured} \qquad x = C_1^{-1} x_{measured}$$

Recalling the projection equations allows to relate the world point and the image points.

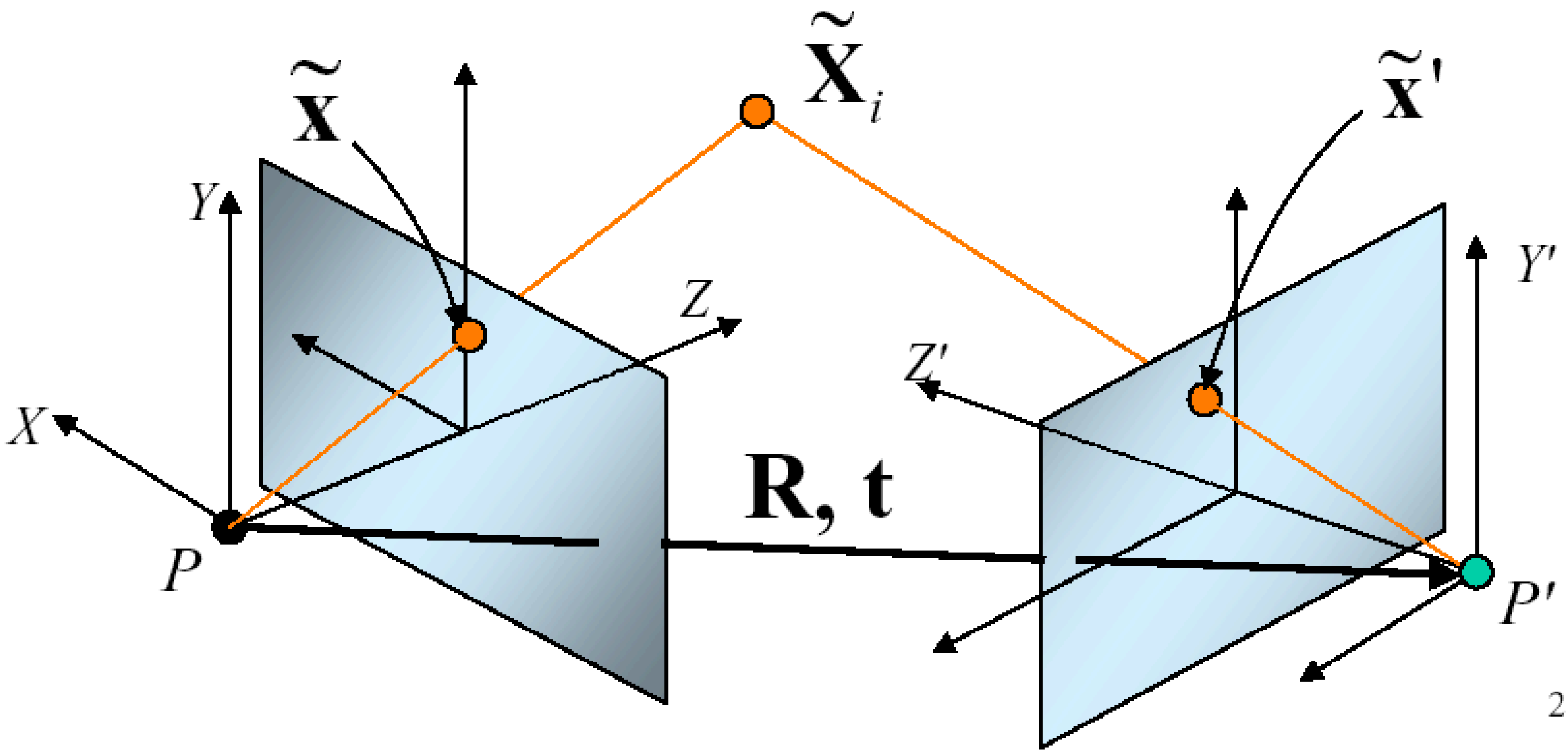
$$x' = C_2^{-1} x'_{measured} \qquad x = C_1^{-1} x_{measured}$$

$$z' x' = C_2^{-1} C_2 M_2^{norm} X \qquad zx = C_1^{-1} C_1 [\mathbf{I} | 0] X$$

$$z' x' = M_2^{norm} X \qquad zx = [\mathbf{I} | 0] X$$

Backprojection to 3D

We now know *x, x'*, *R*, and *t*
Need *X*



Solving…

$$z x_i = M^{norm} X_i$$

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_1^t \\ m_2^t \\ m_3^t \end{bmatrix} X_i \Rightarrow \begin{cases} z u_i = m_1^t \cdot X_i \\ z v_i = m_2^t \cdot X_i \\ z = m_3^t \cdot X_i \end{cases} \Rightarrow \begin{cases} (m_3^t \cdot X_i) u_i = m_1^t \cdot X_i \\ (m_3^t \cdot X_i) v_i = m_2^t \cdot X_i \end{cases}$$

$$\Rightarrow \begin{cases} (m_3^t \cdot X_i) u_i - m_1^t \cdot X_i = 0 \\ (m_3^t \cdot X_i) v_i - m_2^t \cdot X_i = 0 \end{cases}$$

$$\Rightarrow \begin{bmatrix} u_i (m_3^t) - m_1^t \\ v_i (m_3^t) - m_2^t \end{bmatrix} X_i = 0$$

Solving…

- Similarly for the other camera

$$\begin{bmatrix} u'_i({}^2m_3^t) - {}^2m_1^t \\ v'_i({}^2m_3^t) - {}^2m_2^t \end{bmatrix} X_i = 0$$

Where ${}^2m^t_i$ denotes the ith row of the second camera's normalized projection matrix.

Combining 1 & 2:

$$\begin{bmatrix} u_i(m_3^t) - m_1^t \\ v_i(m_3^t) - m_2^t \\ u'_i({}^2m_3^t) - {}^2m_1^t \\ v'_i({}^2m_3^t) - {}^2m_2^t \end{bmatrix} X_i = 0$$

$$AX_i = 0$$

It has a solvable form! Solve using minimum eigenvalue-Eigenvector approach (e.g. Xi = Null(A))

Finishing up

- Now we have the 3D point $\tilde{\mathbf{X}}_i$
- Determine the location of this point for all 4 possible camera configurations
- Next determine the depths of these points in in each camera.

What else can you do with these methods? Synthesize new views

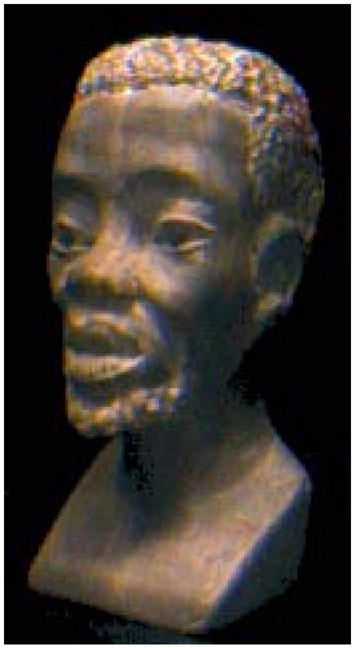


Image 1

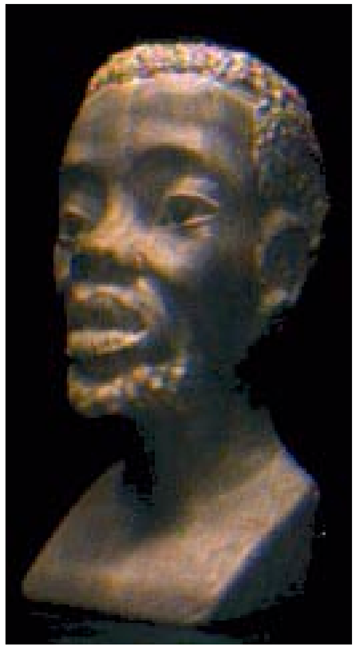
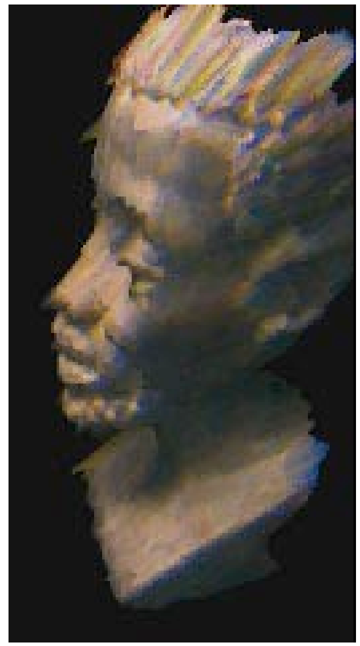
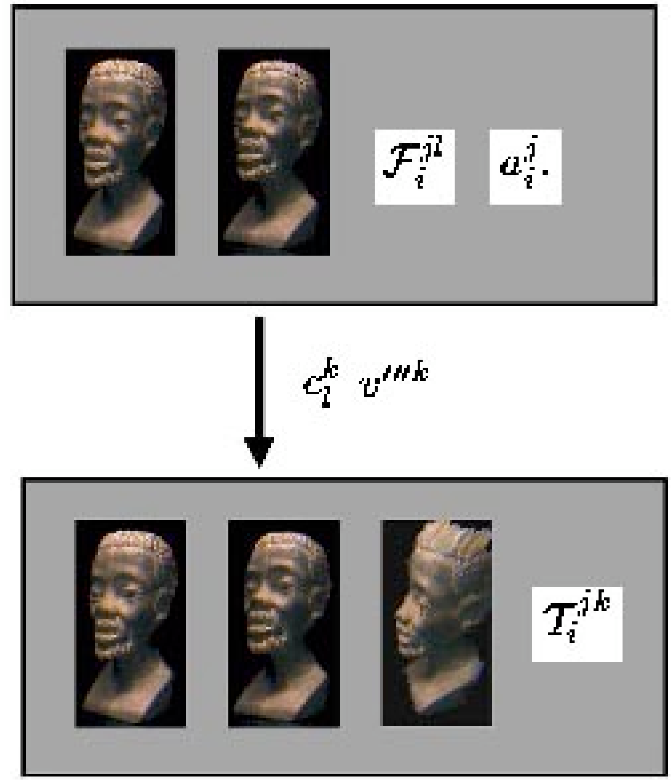


Image 2

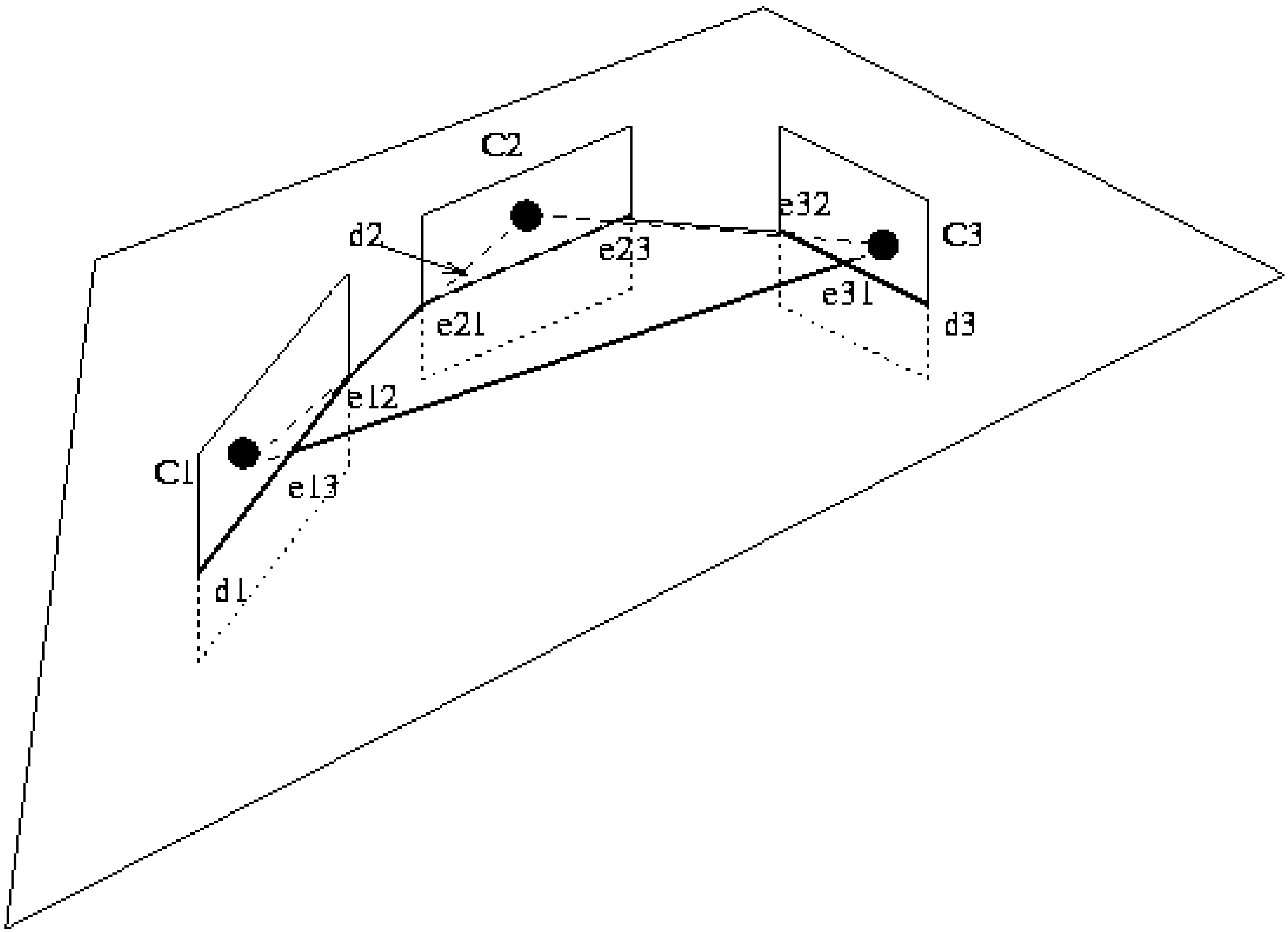


60 deg!



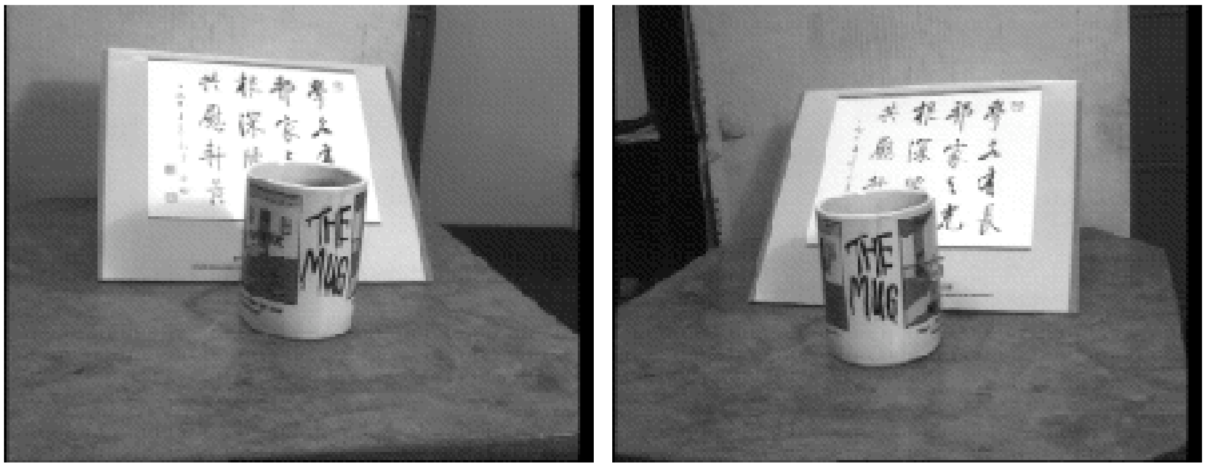
Avidan & Shashua, 1997

Faugeras and Robert, 1996

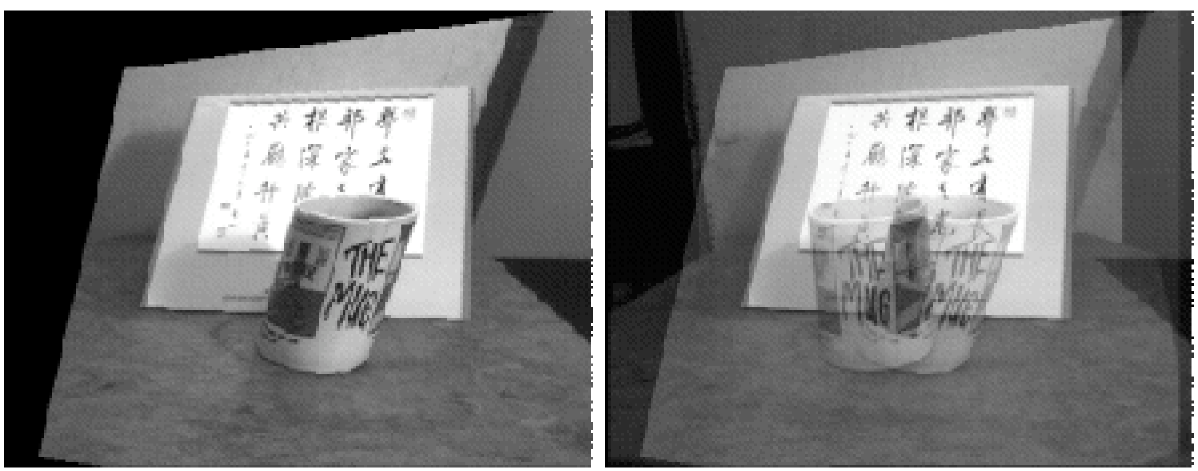


Undo Perspective Distortion (for a plane)

Original images (left and right)



Transfer and superimposed images

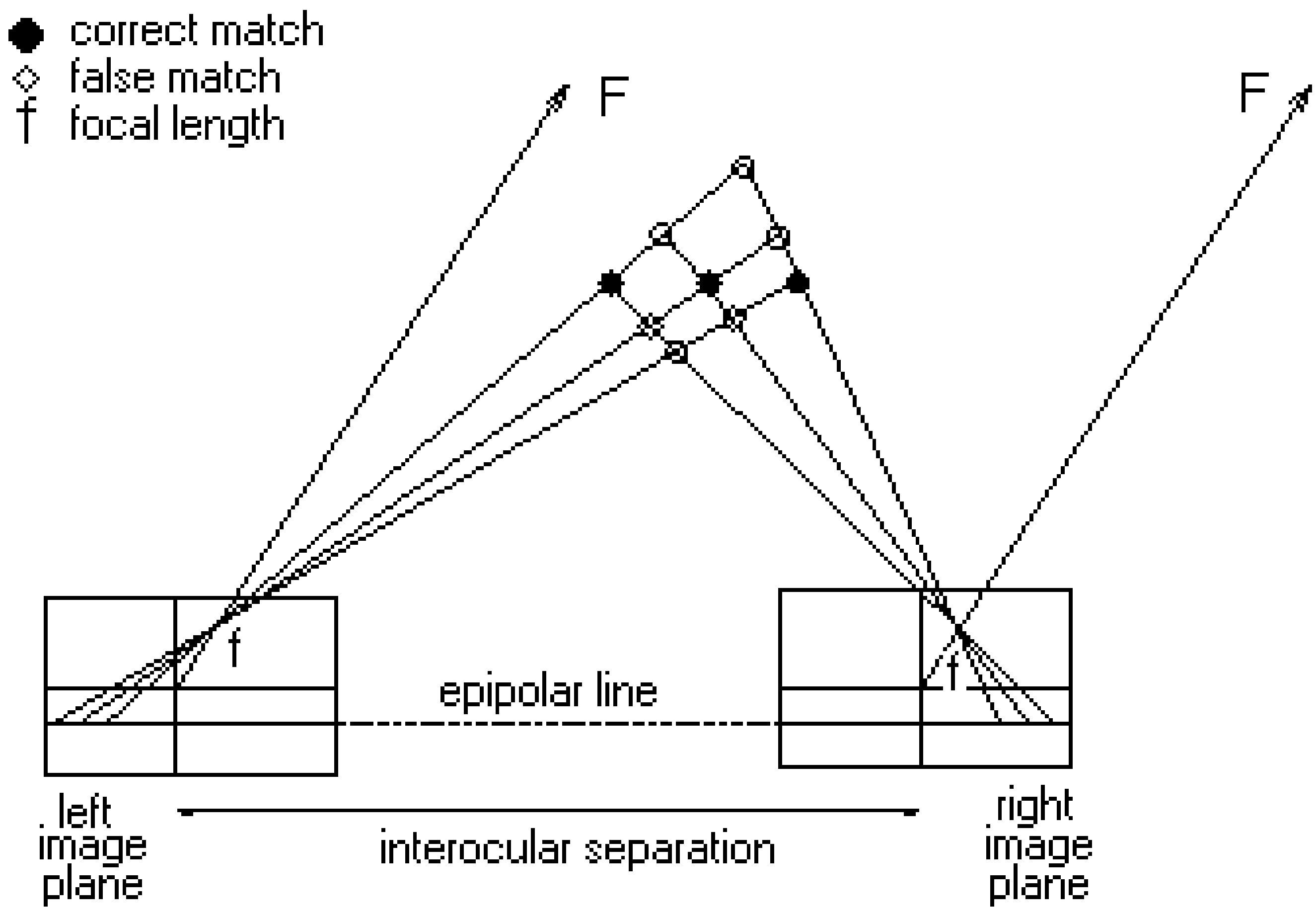


The ``transfer'' image is the left image projectively warped so that points on the plane containing the Chinese text are mapped to their position in the right image.

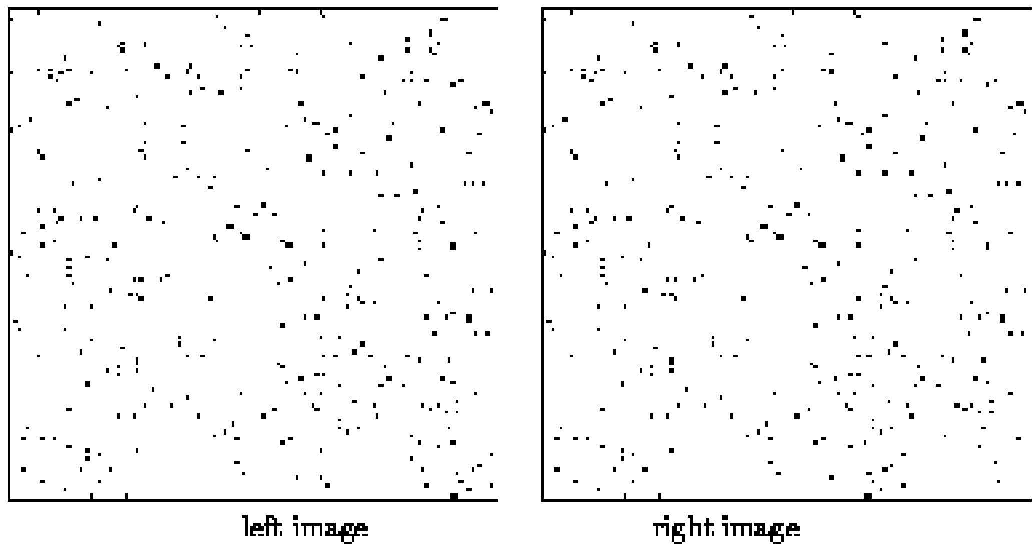
•The ``superimpose'' image is a superposition of the transfer and right image. The planes exactly coincide. However, points off the plane (such as the mug) do not coincide.

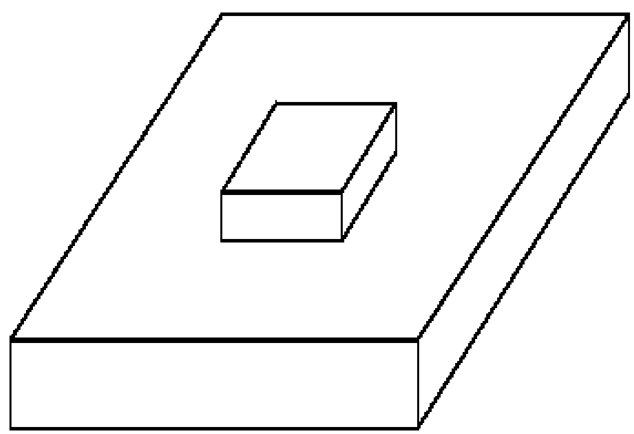
*This is an example of planar projectively induced parallax. Lines joining corresponding points off the plane in the ``superimposed'' image intersect at the epipole.

Its all about point matches



Point match ambiguity in human perception

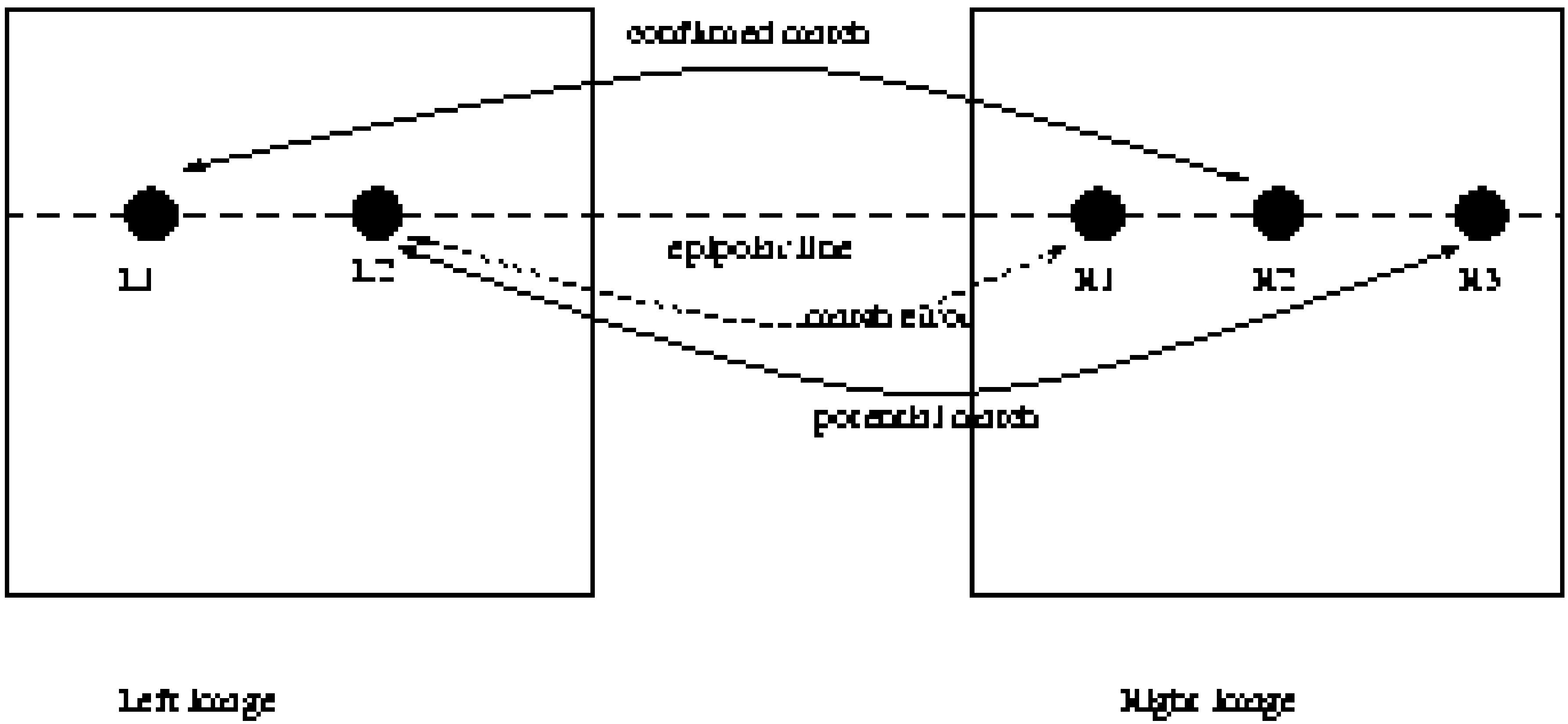




Traditional Solutions

- Try out lots of possible point matches.
- Apply constraints to weed out the bad ones.

Find matches and apply epipolar uniqueness constraint

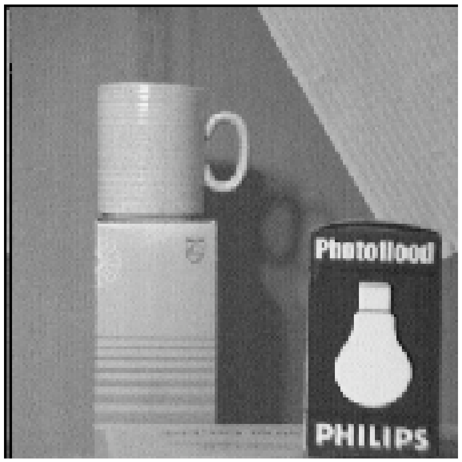


Compute lots of possible matches

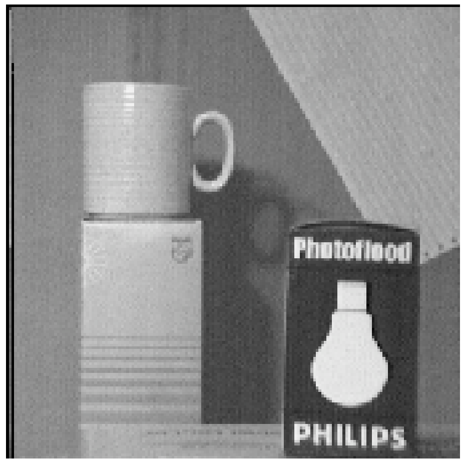
1) Compute "match strength"
2) Find matches with highest strength

Optimization problem with many possible solutions

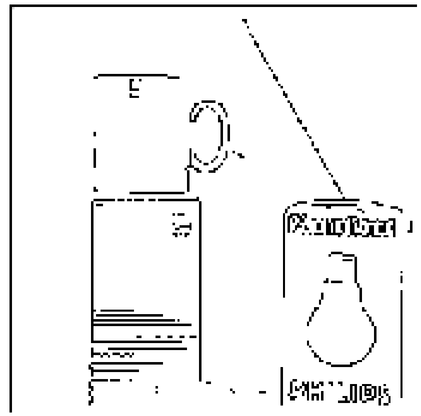
Example



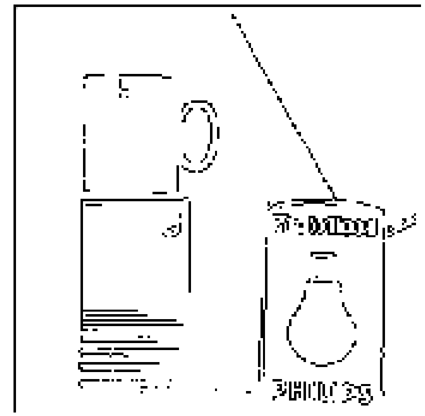
left image



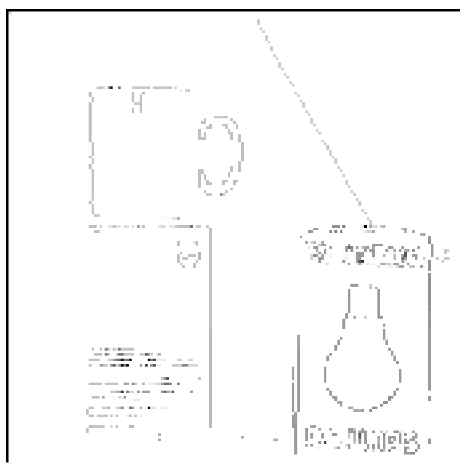
right image



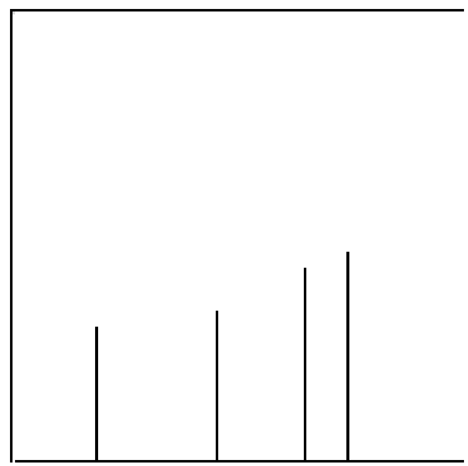
left image



right image



Cyclopean depth image



A scan along row 185 (across the bulb)