Structure from Motion and Multi-view Geometry

Topics in Image-Based Modeling and Rendering

CSE291 J00

Lecture 5

Last lecture

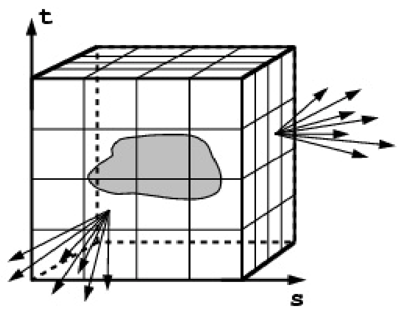
S. J. Gortler, R. Grzeszczuk, R. Szeliski ,M. F. Cohen The Lumigraph, SIGGRAPH, pp 43--54, 1996

M. Levoy, P. Hanrahan, Light Field Rendering , SIGGRAPH, 1996

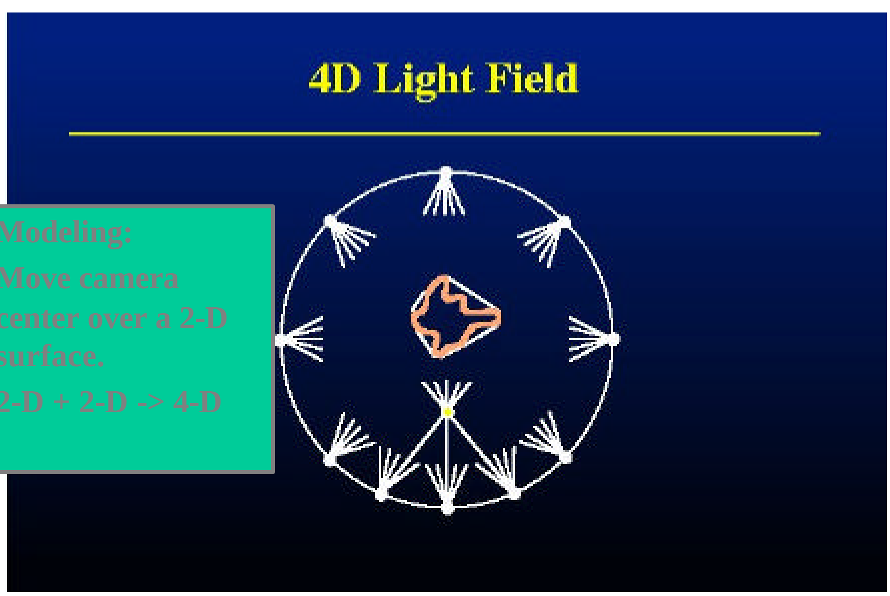
Aaron Isaksen, Leonard McMillan, Steven J. Gortler, Dynamically reparameterized light fields, SIGGRAPH 2000, pp 297 - 306

D. Wood, D. Azuma, W. Aldinger, B. Curless, T. Duchamp, D. Salesin, and W. Steutzle. Surface light fields for 3D photography, SIGGRAPH, 2000.

A cube of light



- All light from an object can be represented as if it were coming from a cube
- Each point on the cube has a a 2-D set of rays emanating from the cube.

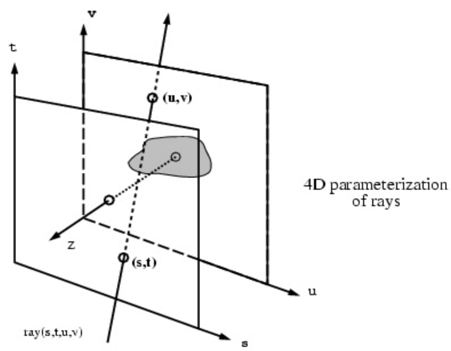


Modeling:

Move camera center over a 2-D surface.

2-D + 2-D -> 4-D

Two plane parameterization



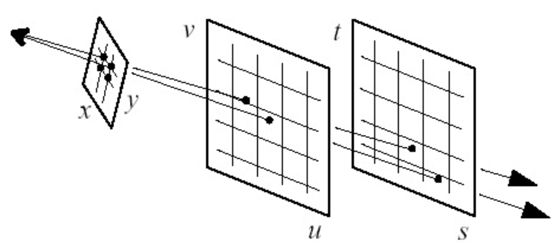
- Rays are parameterized by where they intersect these planes.
- Any point in the 4D Light field/Lumigraph is identified by its coordinates (s,t,u,v)
- Continuous light field

  L(s,t,u,v)

- **Cube around object – six slabs.**

Note that the lumigraph and light field papers interchange roles of (s,t) and (u,v)

Rendering

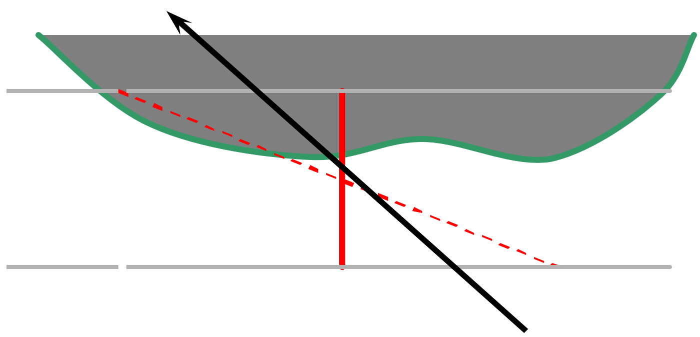


- For each desired ray:
  - Compute intersection with (u,v) and (s,t) planes
  - Take radiance at closest ray
- Can be computed using texture mapping hardware

- Variants: interpolation
  - Bilinear in (u,v) only
  - Bilinear in (s,t) only
  - Quadrilinear in (u,v,s,t)
- Apply band pass filter to remove HF noise that may cause aliasing

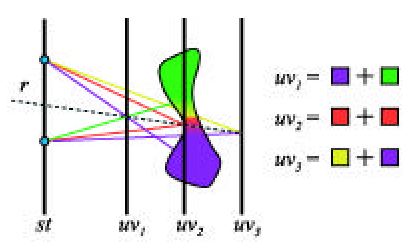
Lumigraph Rendering

- Use rough depth information to improve rendering quality

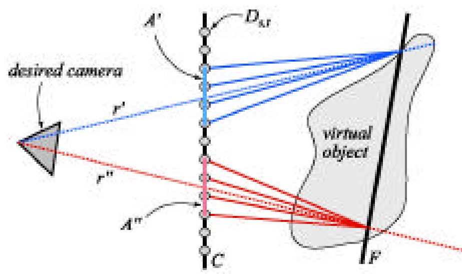


Light Slabs (Again)

- Two-plane parameterization impacts reconstruction filters
- Aperture filtering removes high frequency data
- But if plane is wrong place, only stores blurred version of scene



Variable Aperture



- Real camera aperture produce depth-of-field effects
- Emulate depth-of-field effects by combining rays from several cameras
- Combine samples for each $D_{s,t}$ in the synthetic aperture to produce image

1. Render r' with aperture A': in focus
2. Render r'' with aperature A'': out of focus

This lecture

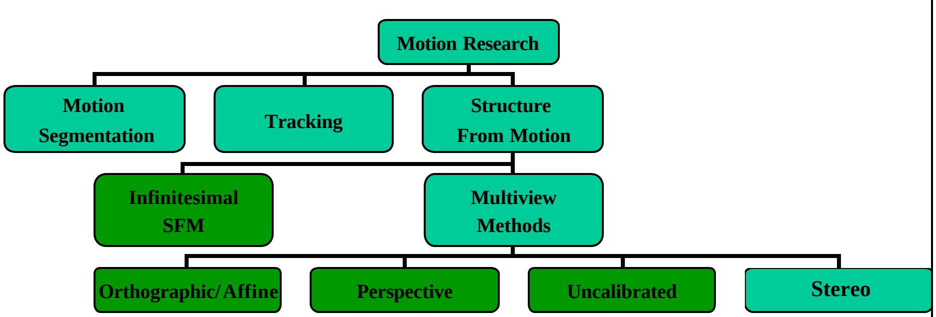
- Motion contains useful information
- Epipolar constraint
- The motion field for a moving camera
- Factorization for orthographic cameras
- Perspective – 8 point algorithm

1. Multiple View Geometry in Computer Vision, by Richard Hartley,Andrew Zisserman, Cambridge University Press, 2000.
2. The Geometry of Multiple Images, by Olivier Faugeras, Quang-Tuan Luong, T. Papadopoulo, MIT Press, 2001.
3. C. Tomasi, T. Kanade, Shape and Motion from Image Streams: A Factorization Method, IJCV, 9(2), 1992, 137-154.
4. Longuet-Higgins, Prazdny, The Interpretation of a Moving Retinal Image, Proc. R. Soc. Long. B, 1980, pp. 385-397.
5. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, Nature, Vol. 293, 1981, pp. 133-135.

Motion Demo

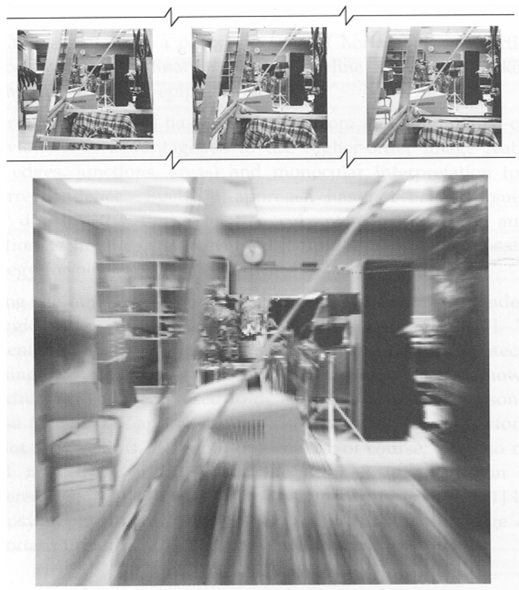
CS348, Fall 2001

Motion Research



CS348, Fall 2001

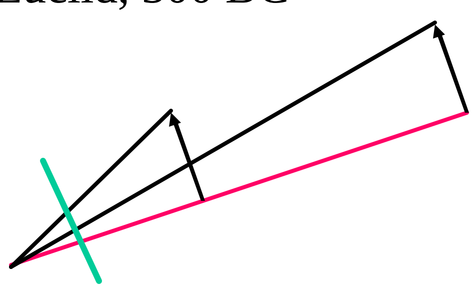
The Motion Field



Motion

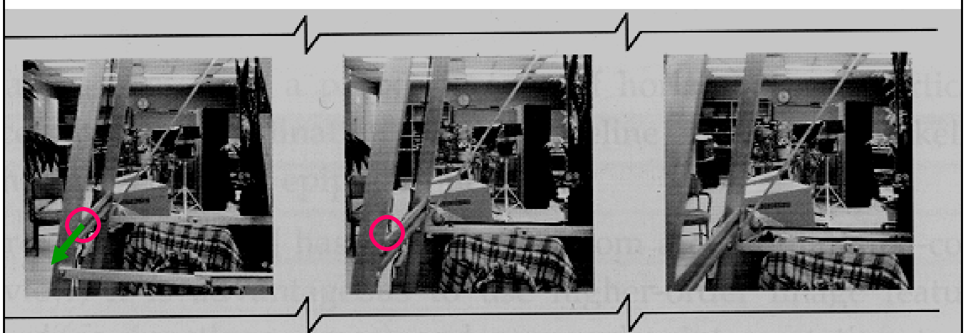
"When objects move at equal speed, those more remote seem to move more slowly."

- Euclid, 300 BC



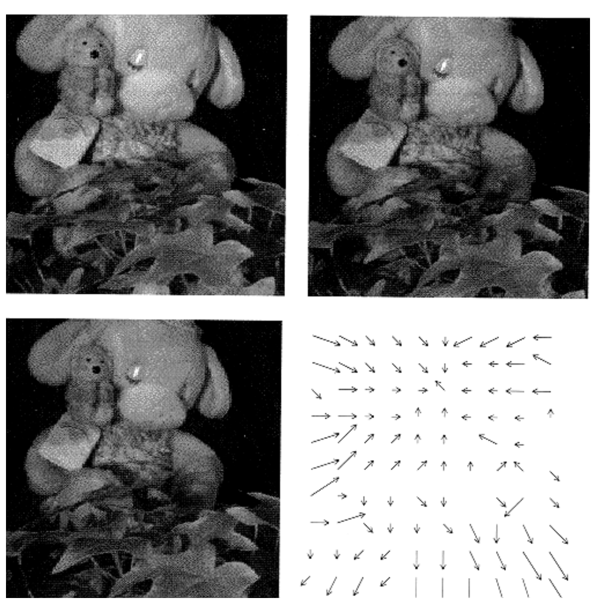
The Motion Field

Where in the image did a point move?



Down and left

The Motion Field



Motion Field Equation

$$v_x = \frac{T_z x - T_x f}{Z} - \omega_y f + \omega_z y + \frac{\omega_x xy}{f} - \frac{\omega_y x^2}{f}$$

$$v_y = \frac{T_z y - T_y f}{Z} + \omega_x f - \omega_z x - \frac{\omega_y xy}{f} + \frac{\omega_x y^2}{f}.$$

- v_x, v_y : Components of image motion
- **T**: Components of 3-D linear motion
- w: Angular velocity vector
- (x,y): Image point coordinates
- Z: depth
- f: focal length

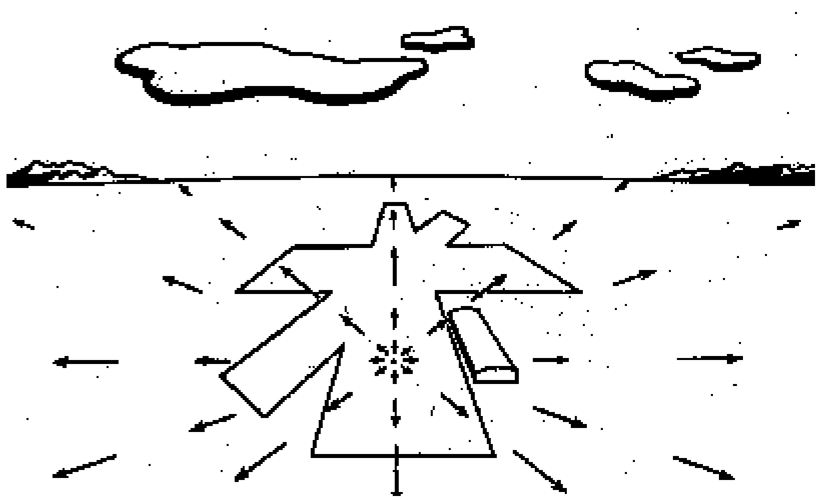
Pure Translation

$$v_x = \frac{T_z x - T_x f}{Z} - \omega_y f + \omega_z y + \frac{\omega_x xy}{f} - \frac{\omega_y x^2}{f}$$

$$v_y = \frac{T_z y - T_y f}{Z} + \omega_x f - \omega_z x - \frac{\omega_y xy}{f} + \frac{\omega_x y^2}{f}.$$

w = 0

Forward Translation & Focus of Expansion [Gibson, 1950]



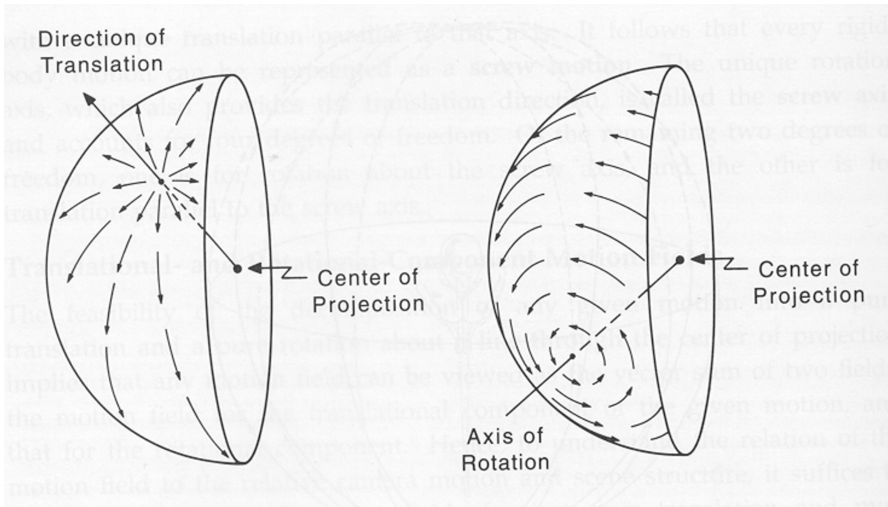
Pure Rotation: **T**=0

$$v_x = \frac{T_z x - T_x f}{Z} - \omega_y f + \omega_z y + \frac{\omega_x x y}{f} - \frac{\omega_y x^2}{f}$$

$$v_y = \frac{T_z y - T_y f}{Z} + \omega_x f - \omega_z x - \frac{\omega_y x y}{f} + \frac{\omega_x y^2}{f}.$$

- Independent of T_x T_y T_z
- Independent of Z
- Only function of (x,y), f and w

Pure Rotation: Motion Field on Sphere

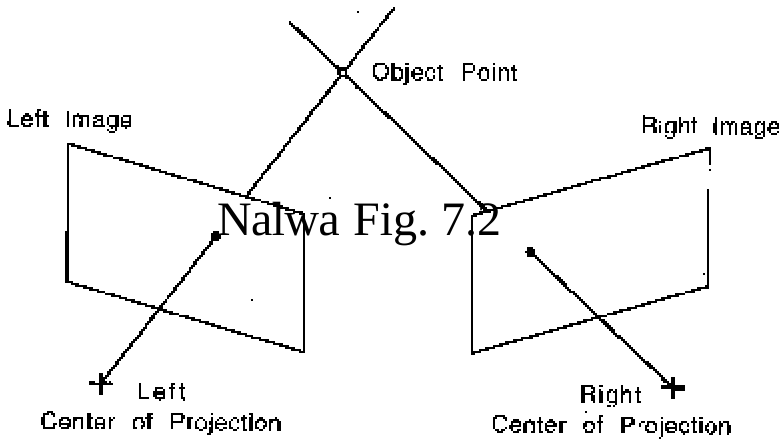


CS348, Fall 2001

Epipolar Geometry

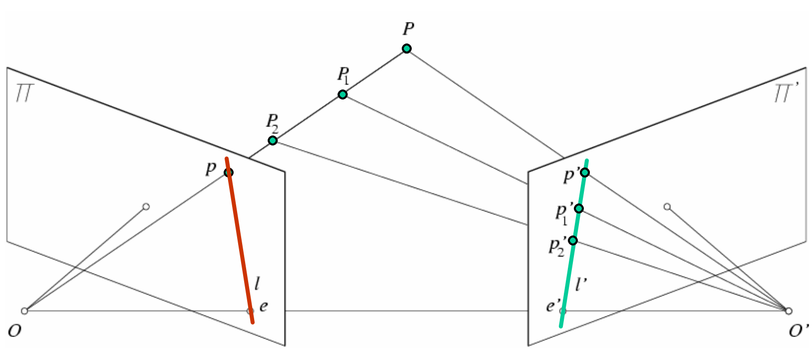
CS348, Fall 2001

Triangulation



Nalwa Fig. 7.2

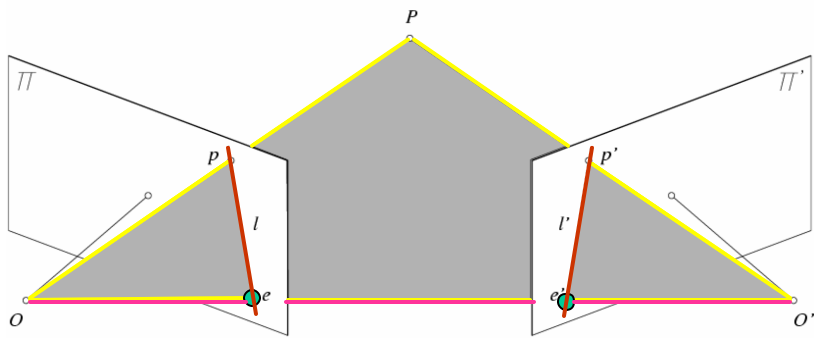
Epipolar Constraint



- 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*.

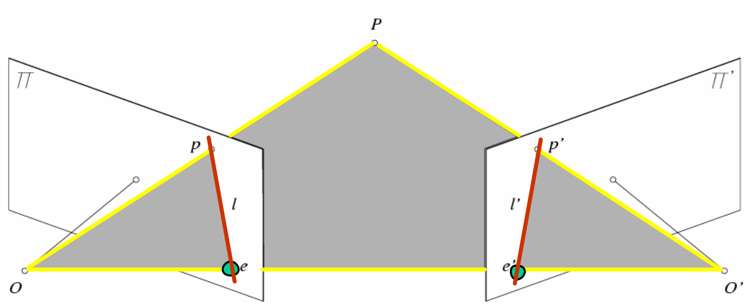
Epipolar Geometry



- Epipolar Plane
- Baseline
- Epipoles
- Epipolar Lines

CS348, Fall 2001

Epipolar Constraint: Calibrated Case



$$\overrightarrow{Op}\cdot[\overrightarrow{OO'}\times\overrightarrow{O'p'}]=0$$

$$\boldsymbol{p}\cdot[\boldsymbol{t}\times(\mathcal{R}\boldsymbol{p}')]=0 \quad \text{with} \begin{cases} \boldsymbol{p}=(u,v,1)^T \\ \boldsymbol{p}'=(u',v',1)^T \\ \mathcal{M}=(\,\mathrm{Id}\quad \boldsymbol{0}\,) \\ \mathcal{M}'=(\,\mathcal{R}^T,-\mathcal{R}^T\boldsymbol{t}\,) \end{cases}$$

$$[t_\times]=\begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

$$\boldsymbol{p}^T\mathcal{E}\boldsymbol{p}'=0 \quad \text{with} \quad \mathcal{E}=[\boldsymbol{t}_\times]\mathcal{R}$$

Essential Matrix
(Longuet-Higgins, 1981)

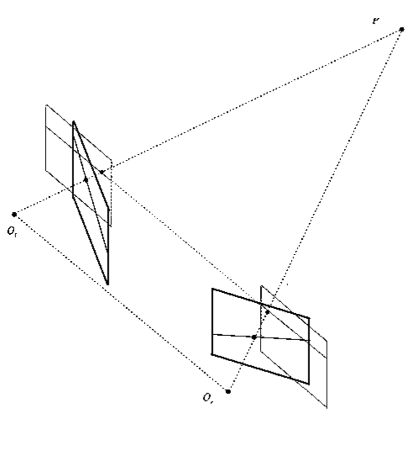
CS348, Fall 2001

Properties of the Essential Matrix

- E p' is the epipolar line associated with p'.
- $E^T p$ is the epipolar line associated with p.
- E e'=0 and $E^T e$=0.
- E is singular.
- E has two equal non-zero singular values (Huang and Faugeras, 1989).

Rectification

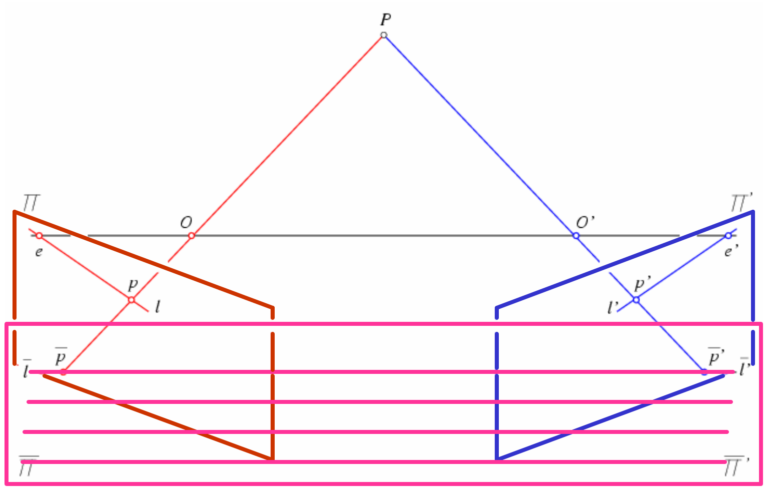
Given a pair of images, transform both images so that epipolar lines are scan lines.



Estimate a Homography to apply to each image, given at least four corresponding points

$$\begin{bmatrix} u' \\ v' \\ w' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

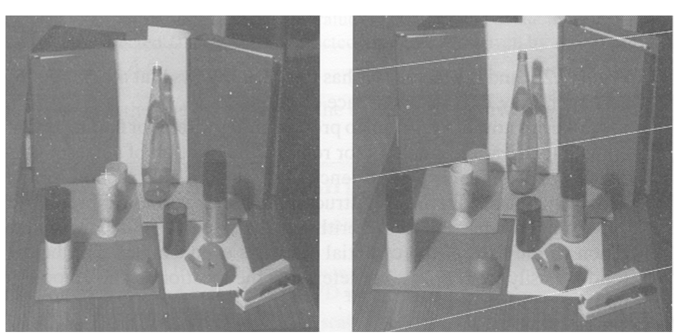
Rectification



All epipolar lines are parallel in the rectified image plane.

Rectification

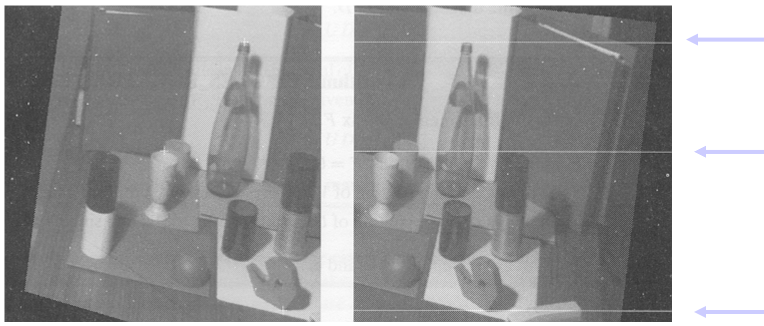
Given a pair of images, transform both images so that epipolar lines are scan lines.



Input Images

Rectification

Given a pair of images, transform both images so that epipolar lines are scan lines.



Rectified Images

CS348, Fall 2001

Features on same epipolar line

CS348, Fall 2001

E vs. F revisited

The Essential Matrix ***E***:

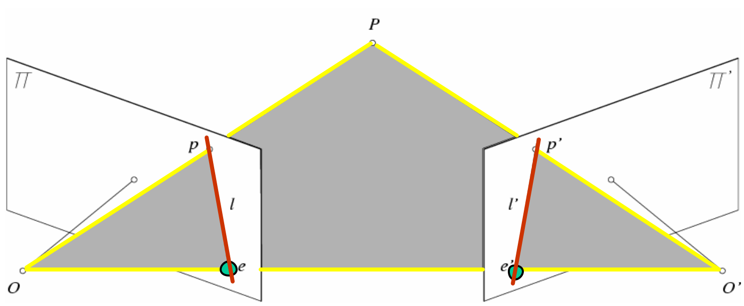
- Encodes information on the extrinsic parameters only
- Has rank 2 since R is full rank and $[T_x]$ is skew & rank 2
- Its two non-zero singular values are equal
- 5 degrees of freedom

The Fundamental Matrix ***F***:

- Encodes information on both the intrinsic and extrinsic parameters
- Also has rank 2 since *E* is rank 2
- 7 degrees of freedom

Thanks Josh

Epipolar Constraint: Uncalibrated Case



$$\hat{\boldsymbol{p}}^T \mathcal{E} \hat{\boldsymbol{p}}' = 0$$

$$\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}$$

$$\boldsymbol{p}' = \mathcal{K}'\hat{\boldsymbol{p}}'$$

$$\boldsymbol{p}^T \mathcal{F} \boldsymbol{p}' = 0 \quad \text{with} \quad \mathcal{F} = \mathcal{K}^{-T}\mathcal{E}\mathcal{K}'^{-1}$$

Fundamental Matrix
(Faugeras and Luong, 1992)

Properties of the Fundamental Matrix

- F p' is the epipolar line associated with p'.
- $F^T p$ is the epipolar line associated with p.
- F e'=0 and $F^T e=0$.
- F is singular.

E vs. F revisited

The Essential Matrix ***E***:

- Encodes information on the extrinsic parameters only
- Has rank 2 since R is full rank and $[T_x]$ is skew & rank 2
- Its two non-zero singular values are equal
- 5 degrees of freedom

The Fundamental Matrix ***F***:

- Encodes information on both the intrinsic and extrinsic parameters
- Also has rank 2 since *E* is rank 2
- 7 degrees of freedom

Thanks Josh

The Eight-Point Algorithm (Longuet-Higgins, 1981)

$$(u, v, 1)\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0 \quad \Rightarrow \quad (uu', uv', u, vu', vv', v, u', v', 1)\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

$$\begin{pmatrix}
u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\
u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\
u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\
u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\
u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\
u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\
u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\
u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8'
\end{pmatrix}
\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Minimize:

$$\sum_{i=1}^{n} (\boldsymbol{p}_i^T \mathcal{F} \boldsymbol{p}_i')^2$$

under the constraint

$|F|^2=1$.

CS348, Fall 2001

Non-Linear Least-Squares Approach

(Luong et al., 1993)

Minimize

$$\sum_{i=1}^{n} [\mathrm{d}^2(\boldsymbol{p}_i, \mathcal{F}\boldsymbol{p}_i') + \mathrm{d}^2(\boldsymbol{p}_i', \mathcal{F}^T\boldsymbol{p}_i)]$$

with respect to the coefficients of F, using an appropriate rank-2 parameterization.

CS348, Fall 2001

Recovering Motion

Recall epipolar constraint

$$\boldsymbol{p}^T \mathcal{E} \boldsymbol{p}' = 0 \quad \text{with} \quad \mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$$

Where

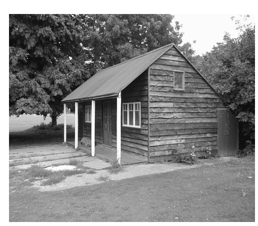
$$[t_\times] = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

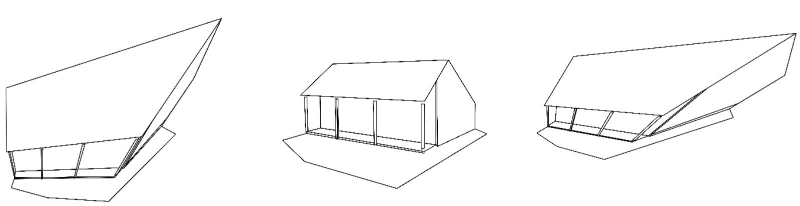
Solve for **R** & **t**

CS348, Fall 2001 © David Kriegman, 2001

Structure from Motion:

Uncalibrated cameras and projective amibguity





CS348, Fall 2001 Thanks Josh © David Kriegman, 2001

The Eight-Point Algorithm (Longuet-Higgins, 1981)

$$(u, v, 1)\begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0$$

$$\Rightarrow (uu', uv', u, vu', vv', v, u', v', 1)\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix} = 0$$

$$\begin{pmatrix} u_1u_1' & u_1v_1' & u_1 & v_1u_1' & v_1v_1' & v_1 & u_1' & v_1' \\ u_2u_2' & u_2v_2' & u_2 & v_2u_2' & v_2v_2' & v_2 & u_2' & v_2' \\ u_3u_3' & u_3v_3' & u_3 & v_3u_3' & v_3v_3' & v_3 & u_3' & v_3' \\ u_4u_4' & u_4v_4' & u_4 & v_4u_4' & v_4v_4' & v_4 & u_4' & v_4' \\ u_5u_5' & u_5v_5' & u_5 & v_5u_5' & v_5v_5' & v_5 & u_5' & v_5' \\ u_6u_6' & u_6v_6' & u_6 & v_6u_6' & v_6v_6' & v_6 & u_6' & v_6' \\ u_7u_7' & u_7v_7' & u_7 & v_7u_7' & v_7v_7' & v_7 & u_7' & v_7' \\ u_8u_8' & u_8v_8' & u_8 & v_8u_8' & v_8v_8' & v_8 & u_8' & v_8' \end{pmatrix}\begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \end{pmatrix} = -\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Minimize:

$$\sum_{i=1}^{n} (\boldsymbol{p}_i^T \mathcal{F} \boldsymbol{p}_i')^2$$

under the constraint

$|F|^2=1.$

Non-Linear Least-Squares Approach
(Luong et al., 1993)

Minimize

$$\sum_{i=1}^{n} [\mathrm{d}^2(\boldsymbol{p}_i, \mathcal{F}\boldsymbol{p}_i') + \mathrm{d}^2(\boldsymbol{p}_i', \mathcal{F}^T\boldsymbol{p}_i)]$$

with respect to the coefficients of F, using an appropriate rank-2 parameterization.

Recovering Motion

Recall epipolar constraint

$$\boldsymbol{p}^T \mathcal{E} \boldsymbol{p}' = 0 \quad \text{with} \quad \mathcal{E} = [\boldsymbol{t}_\times]\mathcal{R}$$

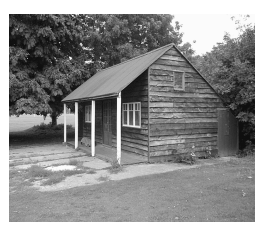
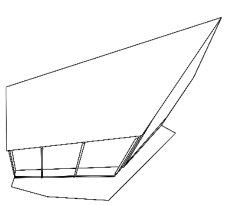
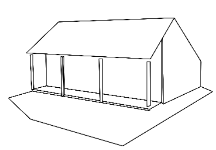
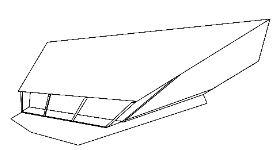
Where

$$[t_\times] = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

Solve for **R** & **t**

Structure from Motion:

Uncalibrated cameras and projective amibguity

Thanks Josh

Projective Ambiguity

Theorem 1 *Let $\{P_1, P_2\}$ and $\{P_1' P_2'\}$ be two pairs of camera transformations. Then $\{P_1, P_2\}$ and $\{P_1' P_2'\}$ correspond to the same essential matrix F if and only if there exists a 4×4 non-singular matrix H such that $P_1 H \approx P_1'$ and $P_2 H \approx P_2'$.*

$$\mathbf{x}_i' = H^{-1}\mathbf{x}_i$$

$$P_j'\mathbf{x}_i' = P_j H H^{-1}\mathbf{x}_i = P_j\mathbf{x}_i = \mathbf{u}_i$$

$$j = 1, 2$$

- We see that the same set of corresponding image points could come from two different sets of real world points and therefore both sets satisfy:

$$\mathbf{u'}_i^{\top} F \mathbf{u}_i = 0$$

The “only if” part

- We would also like to show that if the cameras have the same fundamental matrix then there is a 4 by 4 matrix that relates them.
- If we take the camera matrices

$$\{(M_1| - M_1T_1), (M_2| - M_2T_2)\}$$

and multiply them by:

$$\begin{pmatrix} M_1^{-1} & T_1 \\ \mathbf{0} & 1 \end{pmatrix}$$

we get

$$\{(I|0), (M_2M_1^{-1}| - M_2(T_2 - T_1))\}$$

- The same is true for

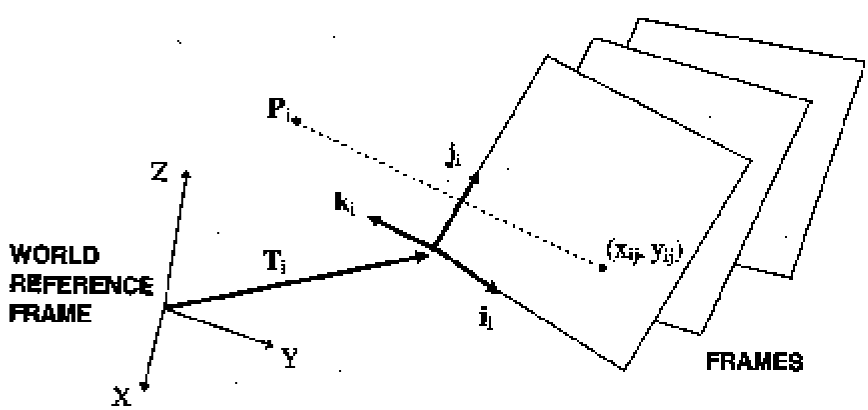
$$\{(M_1'| - M_1'T_1'), (M_2'| - M_2'T_2')\}$$

$$\{(I|0), (M_2'M_1'^{-1}| - M_2'(T_2' - T_1'))\}$$

Factorization method/affine cameras

C. Tomasi, T. Kanade, Shape and Motion from Image Streams: A Factorization Method, IJCV, 9(2), 1992, 137-154.

Factorization Notation



N: Number of images

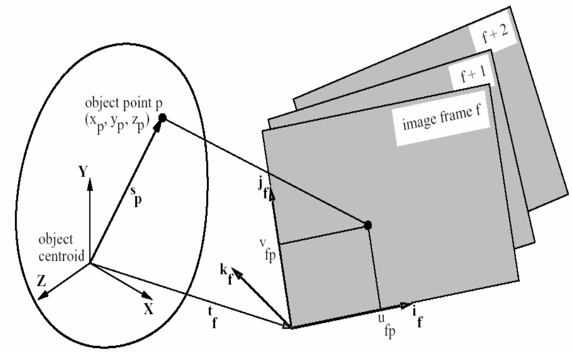
n: Number of points

P_i: i^{th} point in 3-D

R_j, T_j: Rotation & Translation of Camera j

$(x_{i,j}, y_{i,j})$: image of j^{th} point measured in the i^{th} frame.

Origin of 3D Points at 3-D Centroid



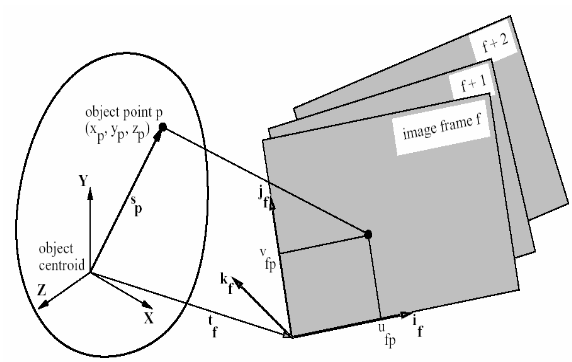
Centroid

$$\overline{P} = \frac{1}{n}\sum_{i=1}^{n} P_j$$

Centered Data Points

$$\widetilde{P} = P_j - \overline{P}$$

Origin of image at image Centroid



Centered Image Points

$$\begin{bmatrix} \tilde{x}_{i,j} \\ \tilde{y}_{i,j} \end{bmatrix} = \begin{bmatrix} x_{i,j} - \overline{x}_i \\ y_{i,j} - \overline{y}_i \end{bmatrix}$$

- Centriod of image points is projection of 3D centroid

Data Matrix

$\tilde{x}_{i,j}, \tilde{y}_{i,j}$ are the i,j-th element of the N by n data matrices X, Y

$$\tilde{W} = \begin{bmatrix} \tilde{X} \\ --- \\ \tilde{Y} \end{bmatrix}$$ 2N by n registered data matrix

Rank Theorem: Without noise, the rank of W is less than or equal to 3.

The Rank Theorem: Factorized

The registered measurement matrix can be expressed in a matrix form: $\widetilde{W} = RS$

$$R = \begin{bmatrix} \mathbf{i}_1^T \\ \vdots \\ \mathbf{i}_F^T \\ \mathbf{j}_1^T \\ \vdots \\ \mathbf{j}_F^T \end{bmatrix}$$ represents the camera rotation

$$S = \begin{bmatrix} \mathbf{s}_1 & \cdots & \mathbf{s}_P \end{bmatrix}$$ is the shape matrix

Factoring

Given a data matrix containing measured feature points, it can be factored using singular value decomposition as:

$$\tilde{W} = UDV^T$$

Where

D: n by n diagonal matrix, non-negative entries, called singular values

U: 2N by n with orthogonal columns

V^T: n by n with orthogonal columns

$$D = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_n \end{bmatrix}, s_1 \geq s_2 \geq s_3 \geq s_n$$

- Without noise, σ_i =0, i>3
- With noise, set σi =0, i>3

Factoring: After setting σ_i=0, i>3

$$\tilde{W}' = U'D'V'^T$$

Where

D': 3 by 3 diagonal matrix, non-negative entries

U: 2N by 3 with orthogonal columns

V^T: 3 by n with orthogonal columns

$$\tilde{W}' = U'D'V'^T = \hat{R}\hat{S}$$

$$where$$

$$\hat{R} = U'D'^{1/2}$$

$$\hat{S} = D'^{1/2}V'^T$$

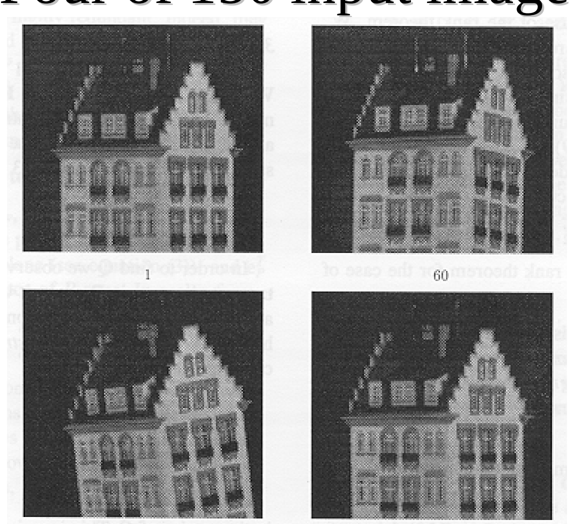
Ambiguity

$$\tilde{W} = \hat{R}\hat{S} = \hat{R}AA^{-1}\hat{S} = (\hat{R}A)(A^{-1}\hat{S})$$

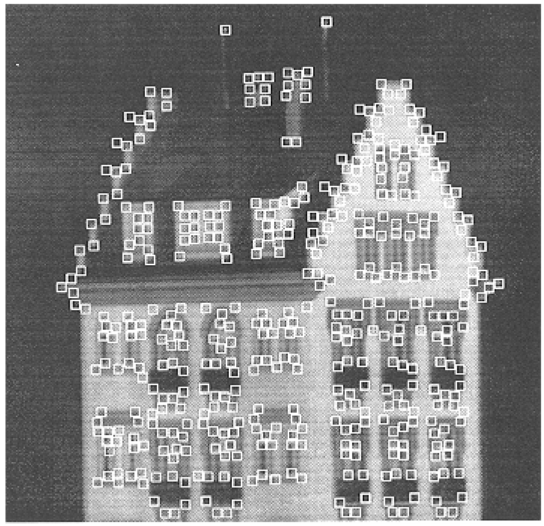
- True for any A.
- So, find an **A** such that rows of **RA** are unit length and pairs corresponding to same image are orthogonal.

$\hat{R}A$	Estimated camera orientation
$A^{-1}\hat{S}$	Estimated 3-D structure

Four of 150 input images

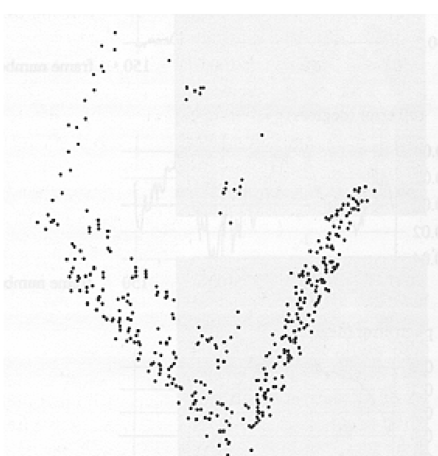


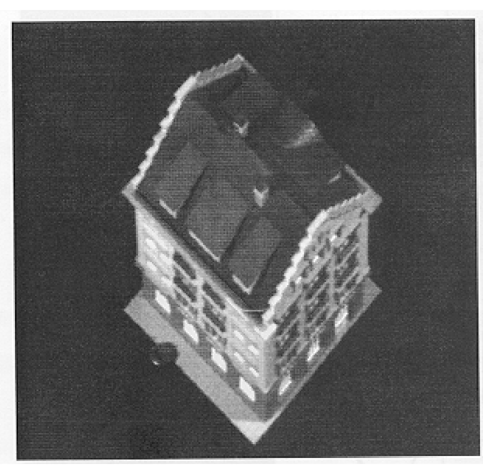
Tracked Corner Features



CS348, Fall 2001

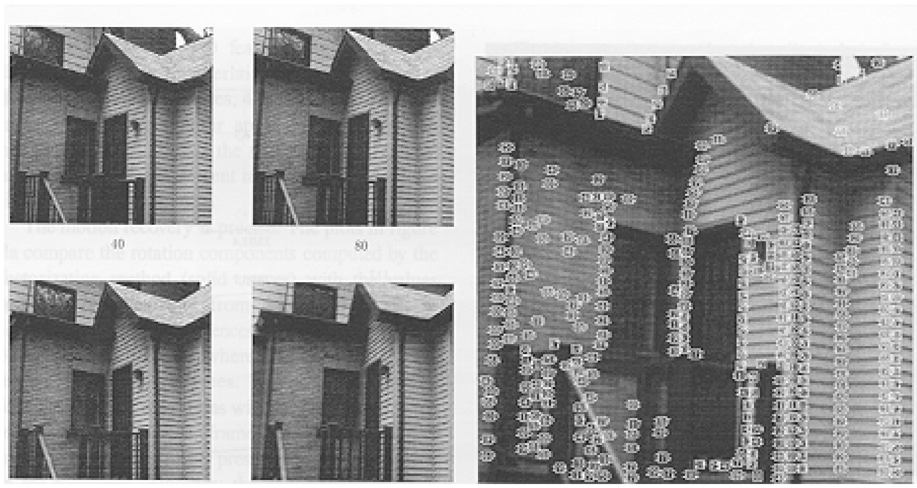
3-D Reconstruction



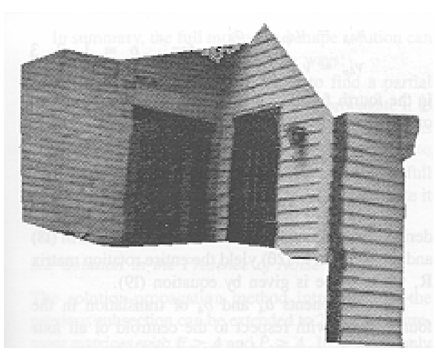


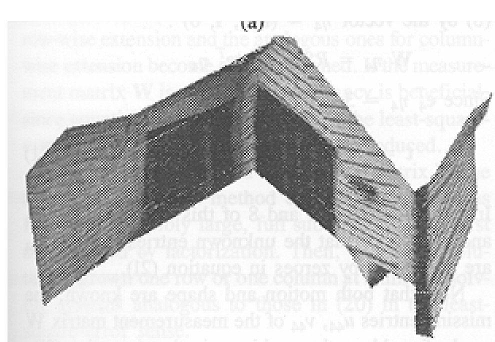
CS348, Fall 2001

Building



Reconstruction





- Triangulate
- Texture Map