Lecture-8

Structure from Motion

Problem

- Given optical flow or point correspondences, compute 3-D motion (translation and rotation) and shape (depth).

3-D Rigid Motion (displacement)

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T = \begin{bmatrix} 1 & -\square & \square \\ \square & 1 & -\square \\ -\square & \square & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$X' = X - \square Y + \square Z + T_X$$
$$Y' = \square X + Y - \square Z + T_Y$$
$$Z' = -\square X + \square Y + Z + T_Z$$

Orthographic Projection (displacement model)

$$X' = X - \square Y + \square Z + T_X$$
$$Y' = \square X + Y - \square Z + T_Y$$
$$Z' = -\square X + \square Y + Z + T_Z$$

$$x' = x - \square y + \square Z + T_X$$
$$y' = \square x + y - \square Z + T_Y$$

Perspective Projection (displacement)

$$X\square = X \square\square Y + \square Z + T_X$$

$$Y\square = \square X + Y \square\square Z + T_Y$$

$$Z\square = \square\square X + \square Y + Z + T_Z$$

$$x\square = \frac{X \square\square Y + \square Z + T_X}{\square\square X + \square Y + Z + T_Z} \qquad x\square = \frac{x \square\square y + \square + \frac{T_X}{Z}}{\square\, \square x + \square y + 1 + \frac{T_Z}{Z}}$$

$$y\square = \frac{\square X + Y \square\square Z + T_Y}{\square\square X + \square Y + Z + T_Z} \qquad y\square = \frac{\square x + y \square\square + \frac{T_Y}{Z}}{\square\, \square x + \square y + 1 + \frac{T_Z}{Z}}$$

Instantaneous Velocity Model

Optical Flow

3-D Rigid Motion

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1 & \square\square & \square \\ \square & 1 & \square\square \\ \square\square & \square & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$\begin{bmatrix} X' - X \\ Y' - Y \\ Z' - Z \end{bmatrix} = \begin{bmatrix} 0 & \square\square & \square \\ \square & 0 & \square\square \\ \square\square & \square & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \begin{bmatrix} 0 & \square\square_3 & \square_2 \\ \square_3 & 0 & \square\square_1 \\ \square\square_2 & \square_1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}$$

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \left(\begin{bmatrix} 0 & \square\square & \square \\ \square & 0 & \square\square \\ \square\square & \square & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$\dot{X} = \square_2 Z \square\square_3 Y + V_1$$
$$\dot{Y} = \square_3 X \square\square_1 Z + V_2$$
$$\dot{Z} = \square_1 Y \square\square_2 X + V_3$$

3-D Rigid Motion

$$\dot{X} = \square_2 Z \square\square_3 Y + V_1$$
$$\dot{Y} = \square_3 X \square\square_1 Z + V_2$$
$$\dot{Z} = \square_1 Y \square\square_2 X + V_3$$

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \qquad \square = \begin{bmatrix} \square_1 \\ \square_2 \\ \square_3 \end{bmatrix}$$

$$\dot{\mathbf{X}} = \square \, \square \mathbf{X} + \mathbf{V}$$

Cross Product

Orthographic Projection

$$\dot{X} = \Omega_2 Z - \Omega_3 Y + V_1$$
$$\dot{Y} = \Omega_3 X - \Omega_1 Z + V_2$$
$$\dot{Z} = \Omega_1 Y - \Omega_2 X + V_3$$

$$y = Y$$
$$x = X$$

$$u = \dot{x} = \Omega_2 Z - \Omega_3 y + V_1$$
$$v = \dot{y} = \Omega_3 x - \Omega_1 Z + V_2$$

(u,v) is optical flow

Perspective Projection (arbitrary flow)

$$x = \frac{fX}{Z} \qquad y = \frac{fY}{Z}$$

$$u = \dot{x} = \frac{fZ\dot{X} - fX\dot{Z}}{Z^2} = f\frac{\dot{X}}{Z} - x\frac{\dot{Z}}{Z}$$
$$v = \dot{y} = \frac{fZ\dot{Y} - fY\dot{Z}}{Z^2} = f\frac{\dot{Y}}{Z} - y\frac{\dot{Z}}{Z}$$

$$\dot{X} = \Omega_2 Z - \Omega_3 Y + V_1$$
$$\dot{Y} = \Omega_3 X - \Omega_1 Z + V_2$$
$$\dot{Z} = \Omega_1 Y - \Omega_2 X + V_3$$

$$u = \dot{x} = f\frac{\dot{X}}{Z} - x\frac{\dot{Z}}{Z} = f\frac{\Omega_2 Z - \Omega_3 Y + V_1}{Z} - x\frac{\Omega_1 Y - \Omega_2 X + V_3}{Z}$$
$$v = \dot{y} = f\frac{\dot{Y}}{Z} - y\frac{\dot{Z}}{Z} = f\frac{\Omega_3 X - \Omega_1 Z + V_2}{Z} - y\frac{\Omega_1 Y - \Omega_2 X + V_3}{Z}$$

$$u = f\left(\frac{V_1}{Z} + \Omega_2\right) - \frac{V_3}{Z}x - \Omega_3 y - \frac{\Omega_1}{f}xy + \frac{\Omega_2}{f}x^2$$
$$v = f\left(\frac{V_2}{Z} - \Omega_1\right) + \Omega_3 x - \frac{V_3}{Z}y + \frac{\Omega_2}{f}xy - \frac{\Omega_1}{f}y^2$$

Perspective Projection (optical flow)

$$u = f(\frac{V_1}{Z} + \Omega_2) - \frac{V_3}{Z}x - \Omega_3 y - \frac{\Omega_1}{f}xy + \frac{\Omega_2}{f}x^2$$

$$v = f(\frac{V_2}{Z} - \Omega_1) + \Omega_3 x - \frac{V_3}{Z}y + \frac{\Omega_2}{f}xy - \frac{\Omega_1}{f}y^2$$

$$u = \frac{fV_1 - V_3 x}{Z} + f\Omega_2 - \Omega_3 y - \frac{\Omega_1}{f}xy + \frac{\Omega_2}{f}x^2$$

$$v = \frac{fV_2 - V_3 y}{Z} - f\Omega_1 + \Omega_3 x + \frac{\Omega_2}{f}xy - \frac{\Omega_1}{f}y^2$$

Pure Translation (FOE)

$$u^{(T)} = \frac{fV_1 - V_3 x}{Z}$$

$$v^{(T)} = \frac{fV_2 - V_3 y}{Z}$$

$$x_0 = f\frac{V_1}{V_3}, y_0 = f\frac{V_2}{V_3}$$

$$u^{(T)} = (x_0 - x)\frac{V_3}{Z}$$

$$p_0 = (x_0, y_0)$$

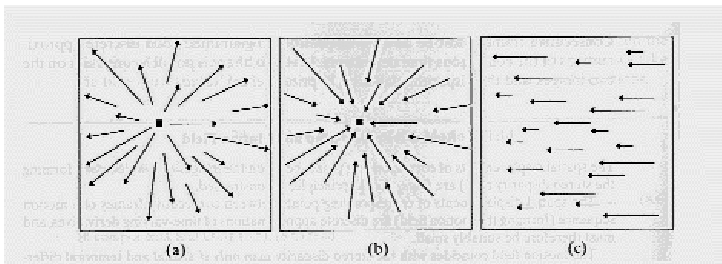
$$v^{(T)} = (y_0 - y)\frac{V_3}{Z}$$

Pure Translation (FOE)

- p_0 is the vanishing point of the direction of translation.
- p_0 is the intersection of the ray parallel to the translation vector with the image plane.

Pure Translation (FOE)

- If V3 is not zero, the flow field is radial, and all vectors point towards (or away from) a single point.
- The length of flow vectors is inversely proportional to the depth, if V_3 is not zero, then it is also proportional to the distance between p and p_0.



Pure Translation (FOE)

$$u^{(T)} = \frac{fV_1 - V_3 x}{Z}$$

$$v^{(T)} = \frac{fV_2 - V_3 y}{Z}$$

$$u^{(T)} = \frac{fV_1}{Z}$$

$$v^{(T)} = \frac{fV_2}{Z}$$

- If $V_3=0$, the flow field is parallel.

Structure From Motion

ORTHOGRAPHIC PROJECTION

Orthographic Projection (displacement)

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T = \begin{bmatrix} 1 & \square\square & \square \\ \square & 1 & \square \\ \square\square & \square & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_X \\ T_Y \\ T_Z \end{bmatrix}$$

$$x' = x \square \square y + \square Z + T_X$$

$$y' = \square x + y \square \square Z + T_Y$$

Simple Method

- **Two Steps Method**

  -Assume depth is known, compute motion

$$x' = x - \alpha y + \beta Z + T_X$$
$$y' = \alpha x + y - \gamma Z + T_Y$$

$$\begin{bmatrix} x' - x \\ y' - y \end{bmatrix} = \begin{bmatrix} -y & Z & 0 & 1 & 0 \\ x & 0 & -Z & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ T_x \\ T_y \end{bmatrix}$$

Simple Method

-Assume motion is known, refine depth

$$x' = x - \alpha y + \beta Z + T_X$$
$$y' = \alpha x + y - \gamma Z + T_Y$$

$$\begin{bmatrix} \beta \\ -\gamma \end{bmatrix} \begin{bmatrix} Z \end{bmatrix} = \begin{bmatrix} x' - x + \alpha y - T_x \\ y' - y - \alpha x - T_y \end{bmatrix}$$

Tomasi and Kanade

Orthographic Projection

Assumptions

- The camera model is orthographic.
- The positions of “p” points in “f” frames (f>=3), which are not all coplanar, have been tracked.
- The entire sequence has been acquired before starting (batch mode).
- Camera calibration not needed, if we accept 3D points up to a scale factor.

Tomasi & Kanade

Image point $\left\{(u_{fp}, v_{fp}) \mid f = 1,\ldots,F,\ p = 1,\ldots,P\right\}$

$$W = \begin{bmatrix} u_{11} \ldots u_{1P} \\ \vdots \\ u_{F1} \ldots u_{FP} \\ v_{11} \ldots v_{1P} \\ \vdots \\ v_{F1} \ldots v_{FP} \end{bmatrix} \qquad W = \begin{bmatrix} U \\ - \\ V \end{bmatrix}$$

Tomasi & Kanade

$$a_f = \frac{1}{P}\sum_{p=1}^{P} u_p \qquad b_f = \frac{1}{P}\sum_{p=1}^{P} v_p$$

$$\tilde{u}_{fP} = u_{fP} - a_{fP}$$

$$\tilde{v}_{fP} = v_{fP} - b_{fP}$$

$$s_p = (X_p, Y_P, Z_P)$$

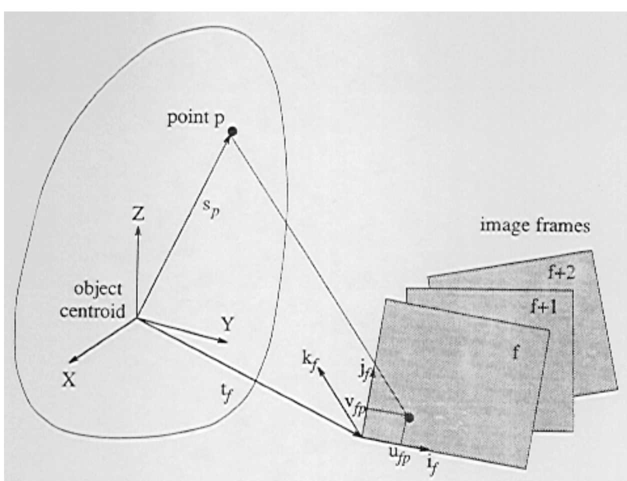
3D world point

$$u_{fP} = i_f^T (s_P - t_f)$$

$$v_{fP} = j_f^T (s_P - t_f)$$

Orthographic projection

$$k_f = i_f \times j_f$$



$$\tilde{u}_{fp} = u_{fP} - a_f$$

$$= i_f^T (s_p - t_f) - \frac{1}{P} \sum_{q=1}^{P} i_f^T (s_q - t_f)$$

$$= i_f^T \left[s_P - \frac{1}{P} \sum_{q=1}^{P} s_q \right]$$

$$= i_f^T s_P$$

Origin of world is at the centroid of object points

$$\tilde{u}_{fP} = i_f^T s_P$$

$$\tilde{v}_{fP} = j_f^T s_P$$

$$\tilde{W} = \begin{bmatrix} \tilde{U} \\ \tilde{V} \end{bmatrix}$$

$$\tilde{u}_{fP} = i_f^T s_P$$

$$\tilde{v}_{fP} = j_f^T s_P$$

$$\tilde{W} = \begin{bmatrix} \tilde{U} \\ \tilde{V} \end{bmatrix}$$

$$\tilde{W} = \underbrace{\begin{bmatrix} i_1^T \\ \vdots \\ i_f^T \\ j_1^T \\ \vdots \\ j_f^T \end{bmatrix}}_{\textbf{2FX3}} \underbrace{\begin{bmatrix} s_1 & \dots & s_P \end{bmatrix}}_{\textbf{3XP}} = RS$$

Rank of ***S*** is 3, because points in 3D space are not Co-planar

Rank Theorem

Without noise, the registered measurement matrix $\tilde{W}$ is at most of rank three.

$$\tilde{W} = \underbrace{\begin{bmatrix} i_1^T \\ \vdots \\ i_f^T \\ j_1^T \\ \vdots \\ j_f^T \end{bmatrix}}_{\textbf{2FX3}} \underbrace{\begin{bmatrix} s_1 & \dots & s_P \end{bmatrix}}_{\textbf{3XP}} = RS$$

Translation

$$\tilde{u}_{fp} = u_{fP} \square a_f$$

$$u_{fp} = \tilde{u}_{fP} + a_f \qquad \tilde{u}_{fp} = i_f^T s_P$$

$$u_{fp} = i_f s_p + a_f \qquad u_{fp} = i_f^T (s_p \square t_f)$$

a_f is projection of camera translation along x-axis

Translation

$$u_{fp} = i_f s_p + a_f \quad v_{fp} = j_f s_p + b_f$$

$$\underset{2FX3}{\mathbf{W}} = \underset{3XP}{\mathbf{R}\mathbf{S}} + \underset{2FX1}{\mathbf{t}}\underset{1XP}{\mathbf{e}_\mathbf{p}^\mathbf{T}}$$

2FX3 3XP 2FX1 1XP

$$\mathbf{t} = (a_1, \ldots, a_f, b_1, \ldots, b_f)^T$$

$$\mathbf{e}_\mathbf{p}^\mathbf{T} = (1, \ldots 1)$$

Translation

Projected camera translation can be computed:

$$\square i_f^T t_f = a_f = \frac{1}{P}\sum_{p=1}^{P} u_p$$

$$\square j_f^T t_f = b_f = \frac{1}{P}\sum_{p=1}^{P} v_p$$

Noisy Measurements

- Without noise, the matrix $\tilde{\mathbf{W}}$ must be at most of rank 3. When noise corrupts the images, however, $\tilde{W}$ will not be rank 3. Rank theorem can be extended to the case of noisy measurements.

Approximate Rank

SVD $$\underset{2F\times P}{O_1}\;\underset{P\times P}{\square}\;\underset{P\times P}{O_2} = \tilde{W}$$

Singular Value Decomposition (SVD)

- For some linear systems Ax=b, Gaussian Elimination or LU decomposition does not work, because matrix A is singular, or very close to singular. SVD will not only diagnose for you, but it will solve it.

Singular Value Decomposition (SVD)

Theorem: Any m by n matrix A, for which $m \geq n$, can be written as

$$A = O_1 W O_2$$

mxn mxn nxn nxn

W is diagonal

O_1, O_2 are orthogonal

$O_1^T O_1 = O_2^T O_2 = I$

Singular Value Decomposition (SVD)

If A is square, then O_1, W, O_2 are all square.

$$O_1^{-1} = O_1^T$$

$$O_2^{-1} = O_2^T$$

$$W^{-1} = diag(\frac{1}{w_j})$$

$$A = O_1 W O_2$$

$$A^{-1} = O_2 diag(\frac{1}{w_j}) O_1$$

Singular Value Decomposition (SVD)

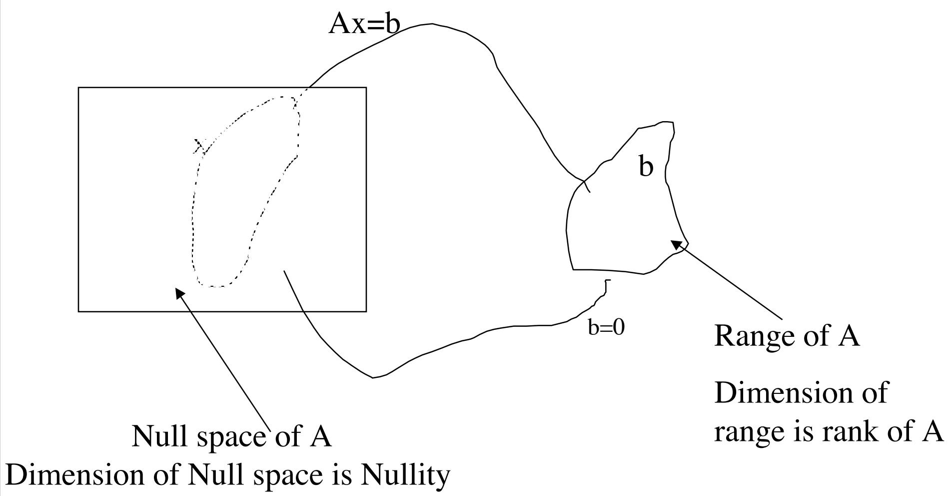
The condition number of a matrix is the ratio of the largest of the w_j to the smallest of w_j. A matrix is singular if the condition number is infinite, it is ill-conditioned if the condition number is too large.

Singular Value Decomposition (SVD)

$$Ax = b$$

- If A is singular, some subspace of “x” maps to zero; the dimension of the null space is called “nullity”.
- Subspace of “b” which can be reached by “A” is called range of “A”, the dimension of range is called “rank” of A.

Range and Null Space



Singular Value Decomposition (SVD)

- If A is non-singular its rank is “n”.
- If A is singular its rank <n.
- Rank+nullity=n

Singular Value Decomposition (SVD)

$$A = O_1 \square O_2$$

- SVD constructs orthonormal basses of null space and range.

- Columns of O_1 with non-zero w_j spans range.

- Columns of O_2 with zero w_j spans null space.

Solution of Linear System

- How to solve Ax=b, when A is singular?

- If “b” is in the range of “A” then system has many solutions.

- Replace $\frac{1}{w_j}$ by zero if $w_j = 0$

$$x = O_2[diag(\frac{1}{w_j})]O_1^T b$$

Solution of Linear System

If b is not in the range of A, above eq still gives the solution, which is the best possible solution, it minimizes:

$$r \equiv | Ax - b |$$

Approximate Rank

$$\tilde{W} = O_1 \Sigma O_2$$

$$O_1 = \begin{array}{cc} 3 & P\text{-}3 \\ \left[O_1' \right. & \left. O_1'' \right] \end{array} \; 2F$$

$$\Sigma = \begin{array}{c} \begin{array}{cc} 3 & P\text{-}3 \end{array} \\ \begin{bmatrix} \Sigma' & 0 \\ 0 & \Sigma'' \end{bmatrix} \end{array} \begin{array}{c} 3 \\ P\text{-}3 \end{array}$$

$$O_1 \Sigma O_2 = O_1' \Sigma' O_2' + O_1'' \Sigma'' O_2''$$

$$O_2 = \begin{array}{c} \begin{bmatrix} O_2' \\ O_2'' \end{bmatrix} \\ P \end{array} \begin{array}{c} 3 \\ P\text{-}3 \end{array}$$

Approximate Rank

$$\hat{W} = O_1' \Sigma' O_2'$$

The best rank 3 approximation to the ideal registered measurement matrix.

Rank Theorem for noisy measurement

The best possible shape and rotation estimate is obtained by considering only 3 greatest singular values of $\tilde{W}$ together with the corresponding left, right eigenvectors.

Approximate Rank

$\hat{R} = O_1[\Sigma']^{1/2}$ Approximate Rotation matrix

$\hat{S} = [\Sigma']^{1/2} O_2'$ Approximate Shape matrix

$\hat{W} = \hat{R}\hat{S}$ This decomposition is not unique

$\hat{W} = \left(\hat{R}Q\right)\left(Q^{-1}\hat{S}\right)$ *Q* is *any* 3X3 invertable matrix

Approximate Rank

$R = \hat{R}Q$

$S = Q^{-1}\hat{S}$

R and ***S*** are linear transformation of approximate Rotation and shape matrices

How to determine ***Q?***

$$\hat{i}_f^T Q Q^T \hat{i}_f^T = 1$$

$$\hat{j}_f^T Q Q^T \hat{j}_f^T = 1$$

$$\hat{i}_f^T Q Q^T \hat{j}_f^T = 0$$

Rows of rotation matrix are unit vectors and orthogonal

How to determine *Q:* Newton's Method

$$f_1(\mathbf{q}) = \hat{i}_i^T Q Q^T \hat{i}_i^T - 1 = 0$$
$$f_2(\mathbf{q}) = \hat{j}_1^T Q Q^T \hat{j}_1^T - 1 = 0$$
$$f_3(\mathbf{q}) = \hat{i}_1^T Q Q^T \hat{j}_1^T = 0$$
$$\vdots$$
$$f_{3f-2}(\mathbf{q}) = \hat{i}_f^T Q Q^T \hat{i}_f^T - 1 = 0$$
$$f_{3f-1}(\mathbf{q}) = \hat{j}_f^T Q Q^T \hat{j}_f^T - 1 = 0$$
$$f_{3f}(\mathbf{q}) = \hat{i}_f^T Q Q^T \hat{j}_f^T = 0$$

$$\mathbf{M}\Delta\mathbf{q} = \varepsilon$$

$$\Delta\mathbf{q} = [\Delta q_1, \ldots, \Delta q_9]$$

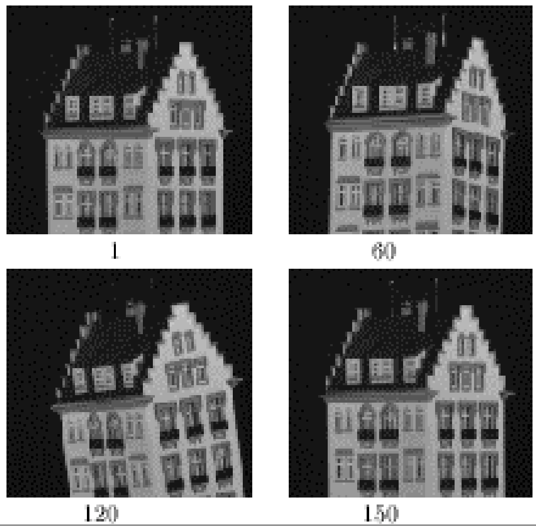
$$\mathbf{M}_{ij} = \frac{\partial f_i}{\partial q_j}$$

ε is error

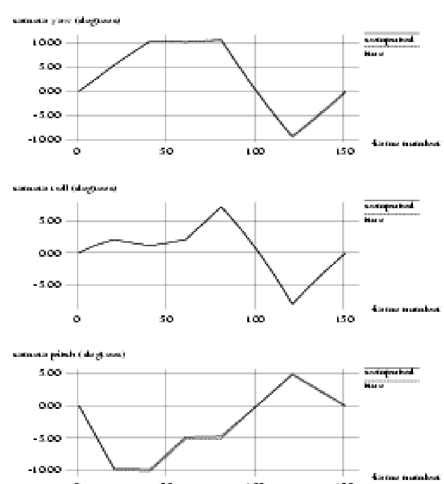
Algorithm

- Compute SVD of $\tilde{W} = O_1 \Sigma O_2$
- define $\hat{R} = O_1' [\Sigma']^{1/2}$ $\hat{S} = [\Sigma']^{1/2} O_2'$
- Compute Q
  - Compute $R = \hat{R} Q$ $S = Q^{-1} \hat{S}$

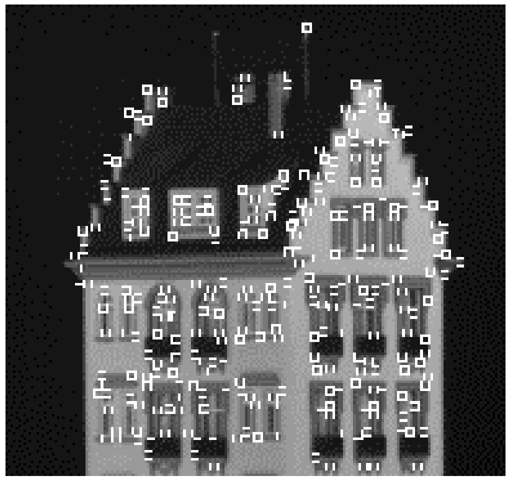
Hotel Sequence



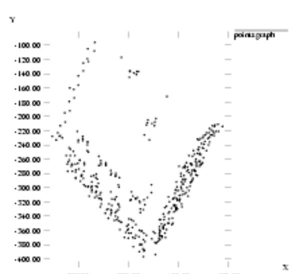
Results (rotations)

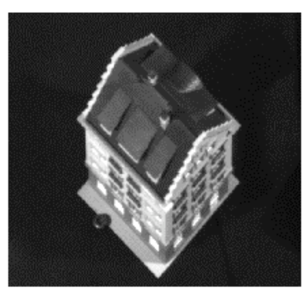


Selected Features

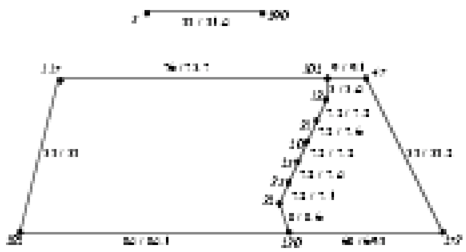


Reconstructed Shape

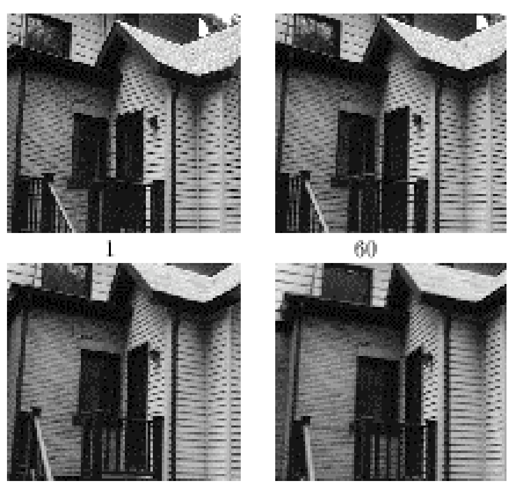




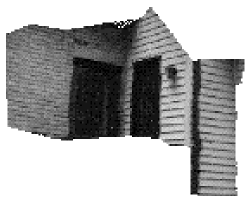
Comparison

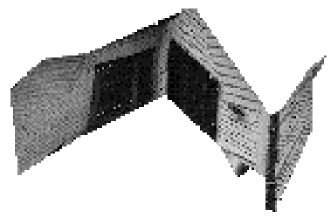


House Sequence



Reconstructed Walls





../tomasiTr92Figures.pdf

Web Page

- http://vision.stanford.edu/cgi-bin/svl/publication/publication1992.cgi