Structure from Motion

Lecture-15

Shape From X

- Recovery of 3D (shape) from one or two (2D images).

Shape From X

- Stereo
- Motion
- Shading
- Photometric Stereo
- Texture
- Contours
- Silhouettes
- Defocus

Applications

- Object Recognition
- Robotics
- Computer Graphics
- Image Retrieval
- Geo-localization
- Archeology
- Sports

Microsoft Kinect sensor

- Data Captured using Microsoft Kinect sensor

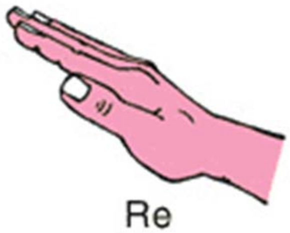
- Approximately 50,000 gesture samples

Gesture Lexicons

Diving Signals

Referee Signals

Nurse Gesture



Music Notes

Gestures from Depth camera

Gestures from RGB camera

Test case 1: Torso motion adds noise (devel 01– 10 gestures)

Instances of Successful Recognition

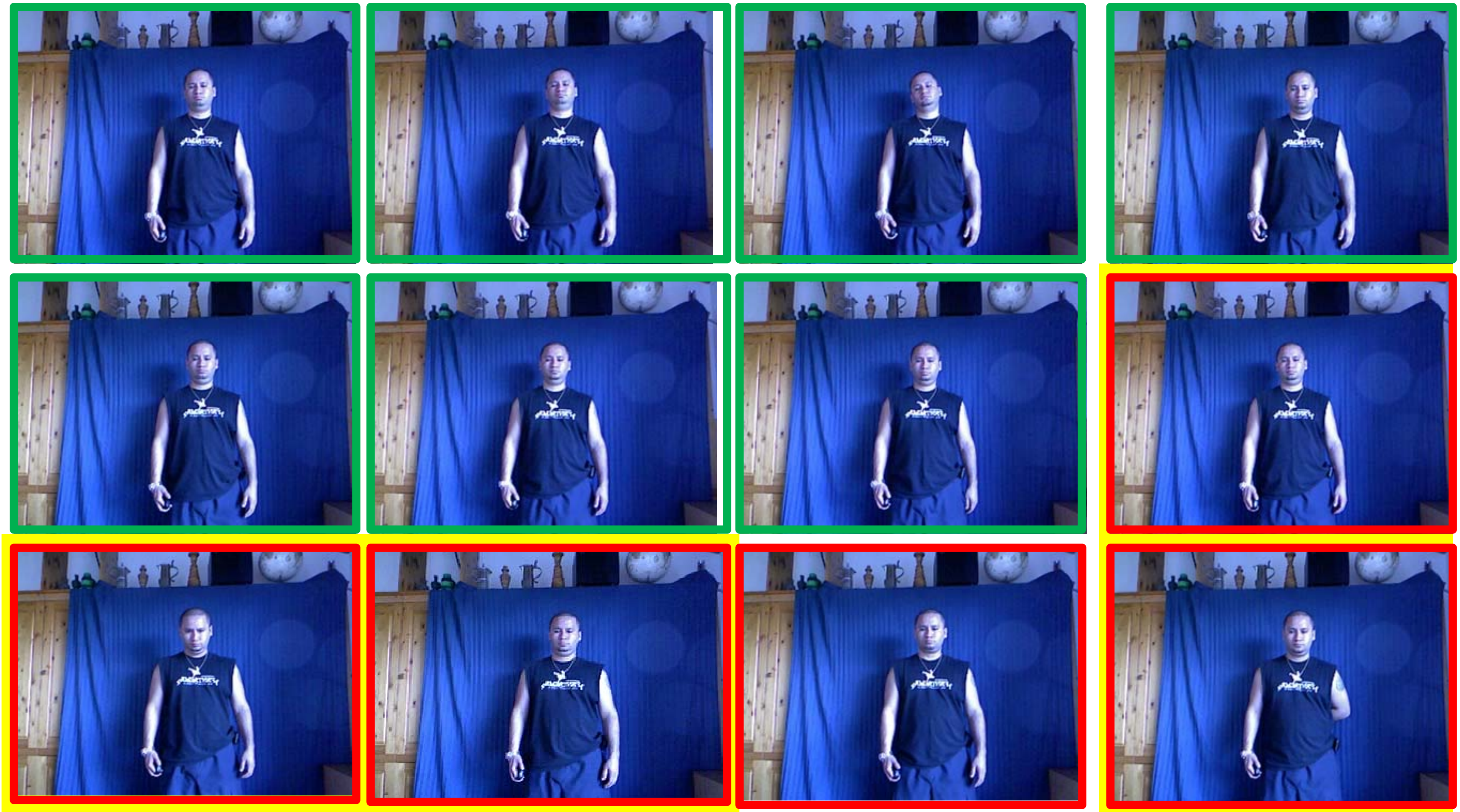
Instances of Failed Recognition

Test Case 2: Improvisations (devel 06 – 9 gestures)

Instances of Successful Recognition

Instances of Failed Recognition

Test Case 3: Subtle differences (devel 09 – 10 gestures)

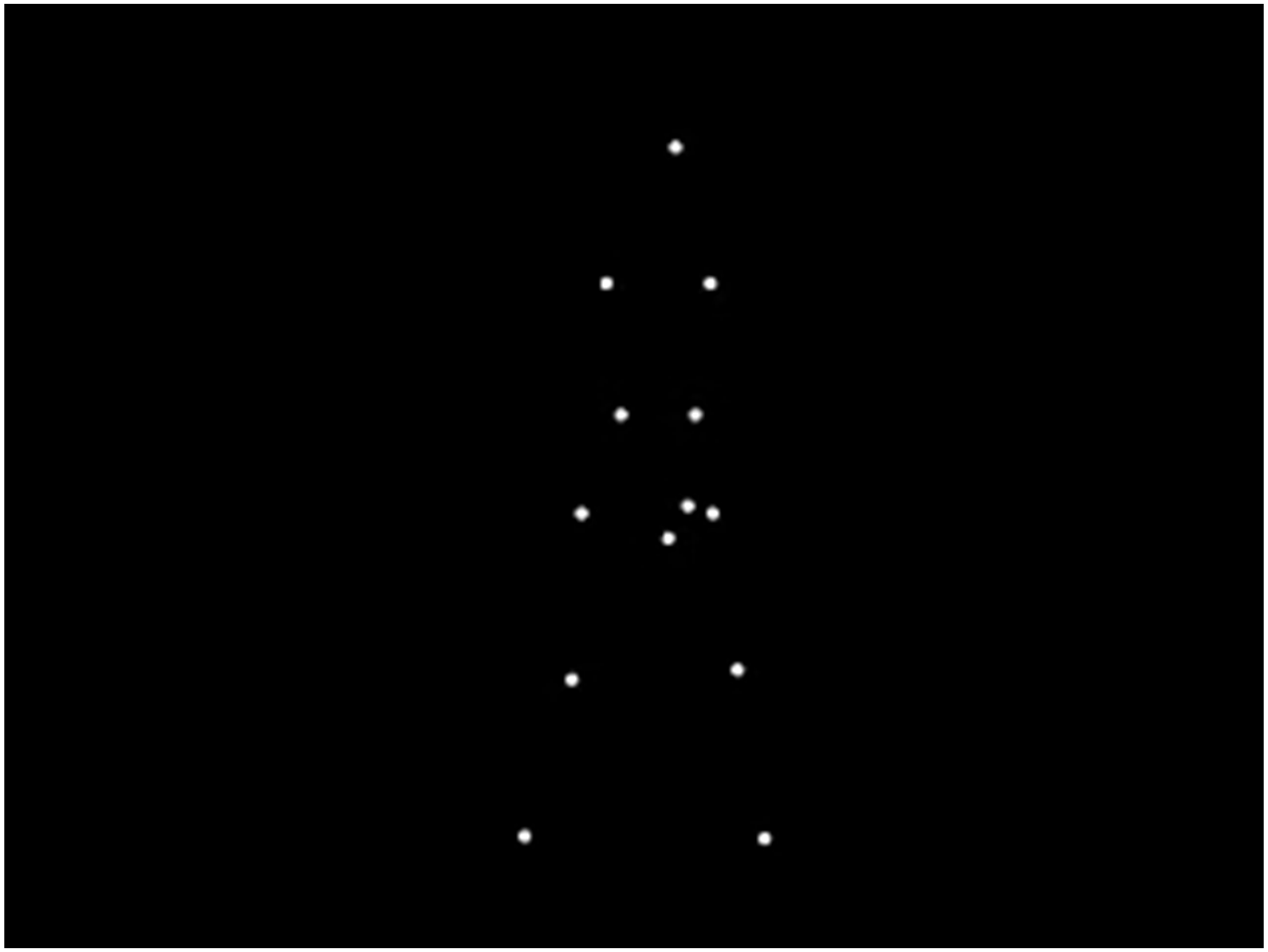


Instances of Successful Recognition

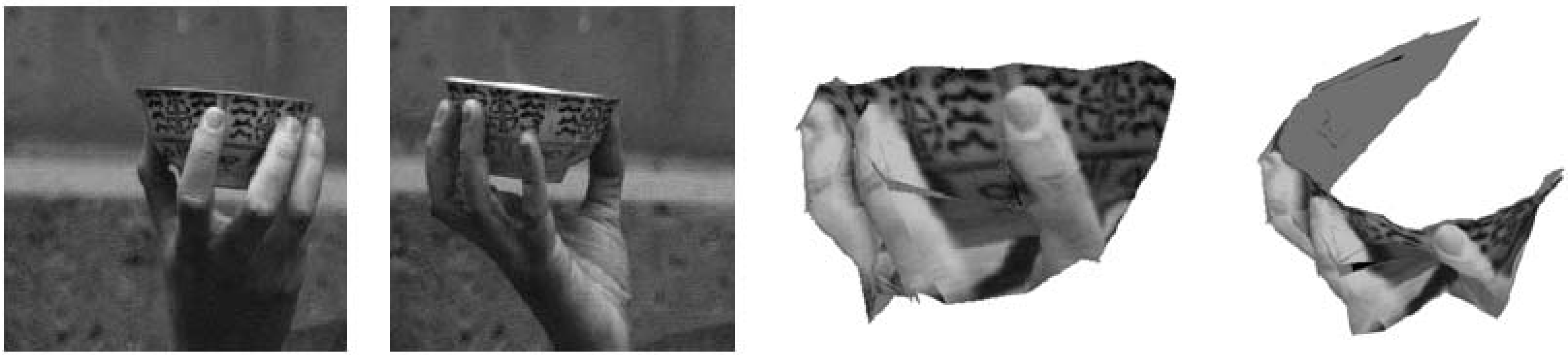
Instances of Failed Recognition

Moving Light Display

Humans are able to recover 3D from motion



Shape from Motion



(a) (b) (c) (d)

Problem

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

Structure from Motion

- S. Ullman
- Hanson & Riseman
- Webb & Aggarwal
- T. Huang
- Heeger and Jepson
- Chellappa
- Faugeras
- Zisserman
- Kanade
- Pentland
- Van Gool
- Pollefeys
- Seitz & Szeliski
- Shahsua
- Irani
- Vidal & Yi Ma
- Medioni
- Fleet
- Tian & Shah
- -

Photosynth

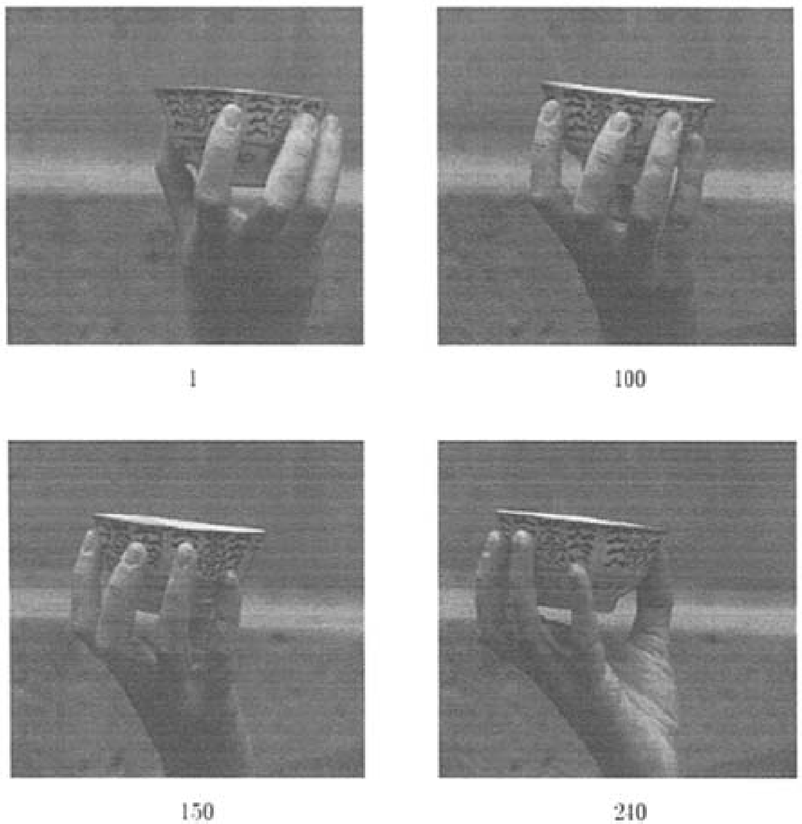
Tomasi and Kanade Factorization

Orthographic Projection

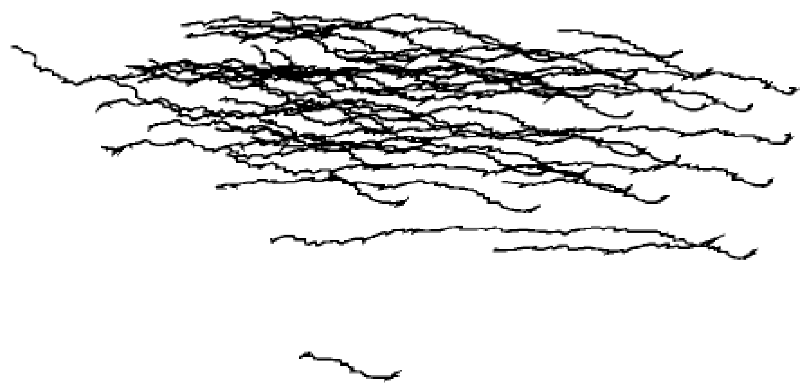
Assumptions

- The camera model is orthographic.
- The positions of “P” points in “F” frames (F>=3), which are not all coplanar, and 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.

Input



(a)



(b)

Images

KLT Tracks

Feature Points

Image points
(This is not optical flow

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

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

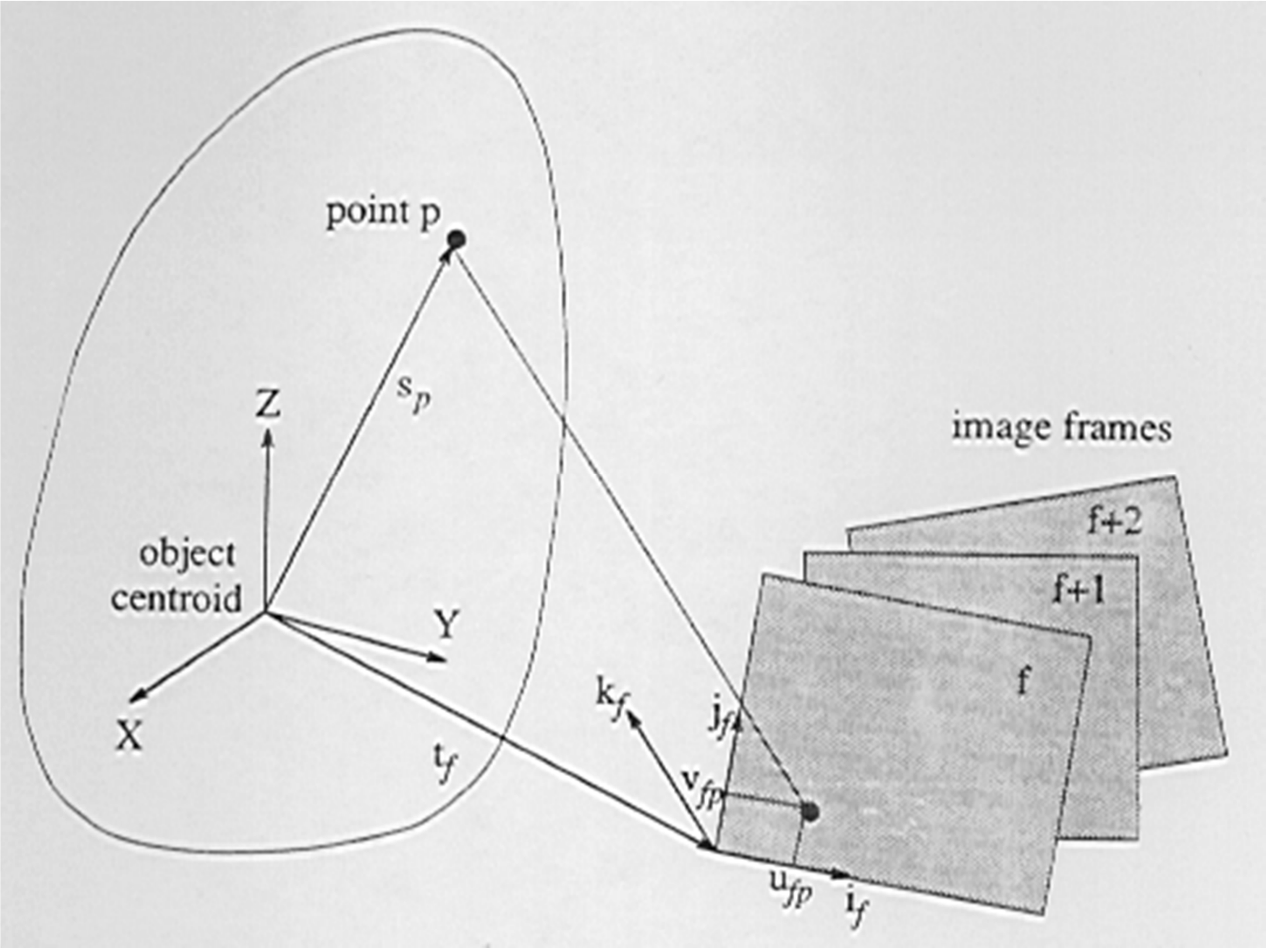
Mean Normalize Feature Points

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

$$\tilde{u}_{fP} = u_{fP} - a_{fP} \qquad \text{(A)}$$

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

Orthographic Projection



Orthographic Projection

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

3D world point

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

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

Orthographic projection

$$k_f = i_f \times j_f$$

i, j, k are unit vectors along X, Y, Z

$$\tilde{u}_{fp} = u_{fP} - a_f \qquad a_f = \frac{1}{P}\sum_{p=1}^{P} u_p$$

$$= i_f^T (s_p - t_f) -$$

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

$$= i_f^T s_P$$

If Origin of world is at the
centroid of object points,
Second term is zero

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

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

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

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

(B)

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

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

$$\tilde{W} = \begin{bmatrix} \tilde{u}_{11} \ldots \tilde{u}_{1p} \\ \vdots \\ \tilde{u}_{F1} \ldots \tilde{u}_{FP} \\ \tilde{v}_{11} \ldots \tilde{v}_{1P} \\ \vdots \\ \tilde{v}_{F1} \ldots \tilde{v}_{FP} \end{bmatrix}$$

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

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.

$$\tilde{W} = \begin{bmatrix} i_1^T \\ \vdots \\ i_F^T \\ j_1^T \\ \vdots \\ j_F^T \end{bmatrix} \begin{bmatrix} s_1 & \dots & s_P \end{bmatrix} = RS$$

Because *W* is a product of two matrices.
The maximum rank of S is 3.

Linearly Independence

A finite subset of n vectors, $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$, from the vector space V, is ***linearly dependent*** if and only if there exists a set of n scalars, a_1, a_2, ..., a_n, not all zero, such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}.$$

Rank of a Matrix

- The **column rank** of a matrix A is the maximum number of linearly independent column vectors of A.
- The **row rank** of a matrix A is the maximum number of linearly independent row vectors of A.
- The column rank of A is the dimension of the column space of A
- The row rank of A is the dimension of the row space of A.

Example (Row Echelon)

$$A = \begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{bmatrix} R_2 \to 2r_1 + r_2$$

Rank is 2

How to find Translation?

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

From (A)

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

From (B)

$$u_{fp} = i_f s_p + a_f \quad \text{(E)} \quad u_{fp} = i_f^T (s_p - t_f)$$

From (C)

Comparing above two eqs

$$\boxed{a_f = -t_f i_f^T} \quad \text{(D)}$$

a_f is projection of camera translation along x-axis

How to find Translation

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

$$\mathbf{W} = \mathbf{R}\mathbf{S} + \mathbf{t}\mathbf{e}_{\mathbf{p}}^{\mathbf{T}} \qquad a_f = -t_f i_f^T$$

2FXP 2FX3 3XP 2FX1 1XP From (D)

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

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

How to find Translation

Projected camera translation can be computed:

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

From (D)

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

Noisy Measurements

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

Singular Valued Decomposition

SVD

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

2FXP **2FXP** **PXP** **PXP**

Singular Value Decomposition (SVD)

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

$$A = O_1 \Sigma O_2$$

mxn mxn nxn nxn

Σ is diagonal

O_1 ,O_2 are orthogonal

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

Approximate Rank

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

$$\begin{array}{cc} & \begin{array}{cc} 3 & P-3 \end{array} \\ O_1 = & \begin{bmatrix} O_1' & O_1'' \end{bmatrix} \end{array} \quad 2F$$

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

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

$$\begin{array}{cc} O_2 = \begin{bmatrix} O_2' \\ O_2'' \end{bmatrix} & \begin{array}{c} 3 \\ P-3 \end{array} \\ \quad P & \end{array}$$

Approximate Rank

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

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

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

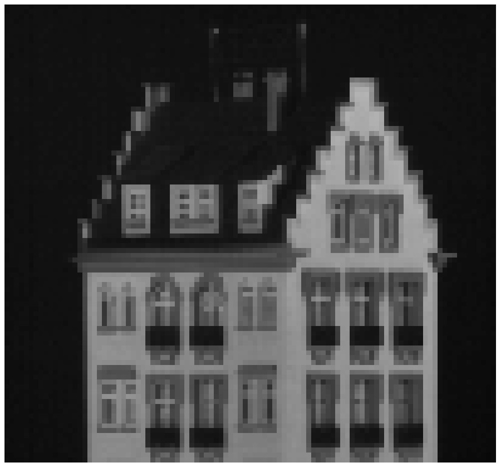
$$\hat{W} = (\hat{R}Q)(Q^{-1}\hat{S})$$

Q is *any* 3X3 invertable matrix

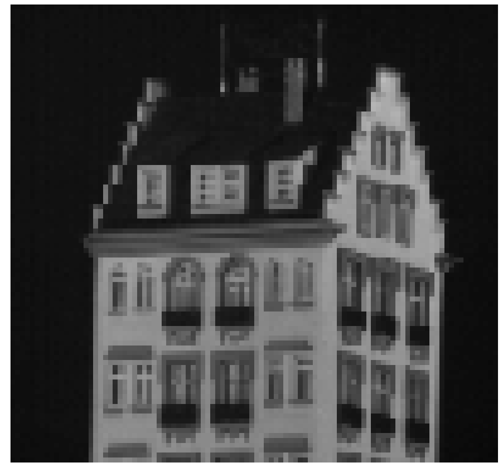
Results

..\..\CAP6411\Fall2002\tomasiTr92Figures.pdf

Hotel Sequence

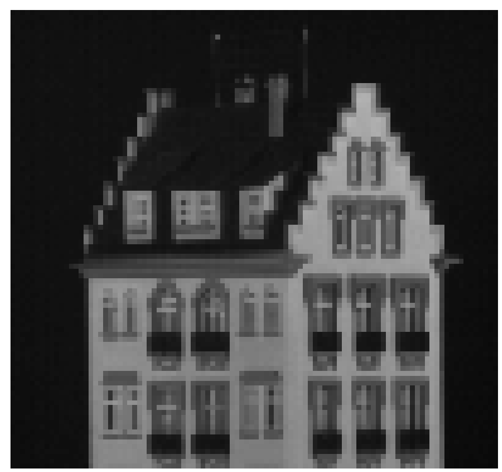


1



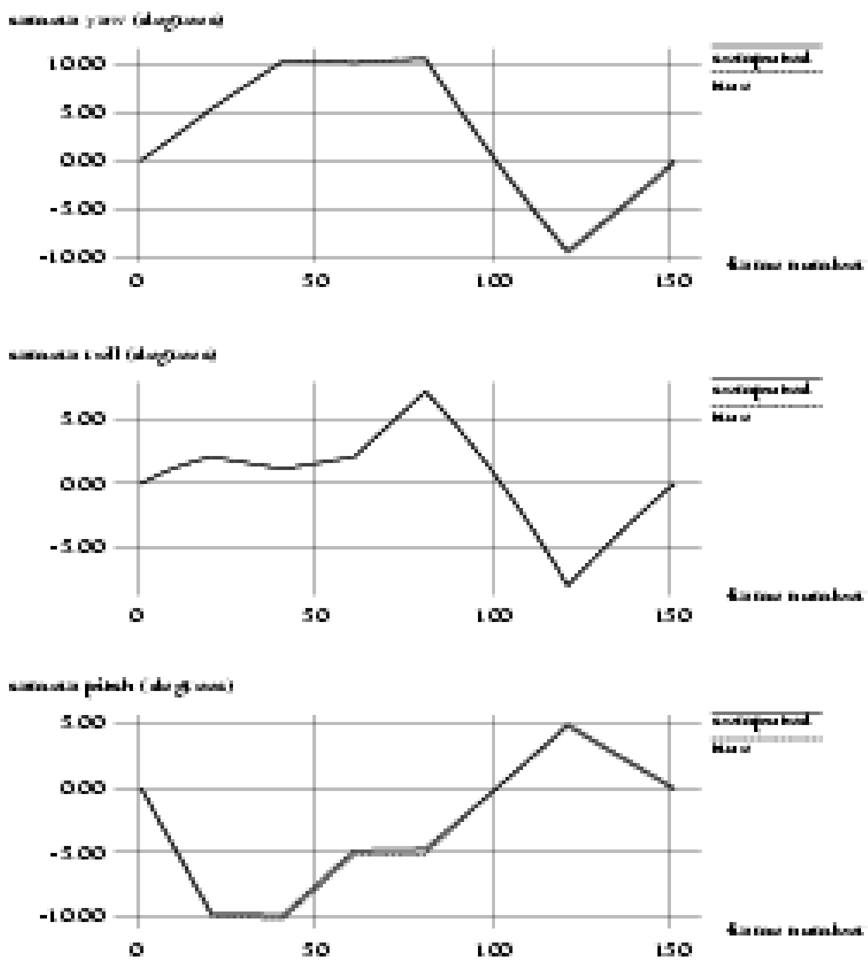
60

120

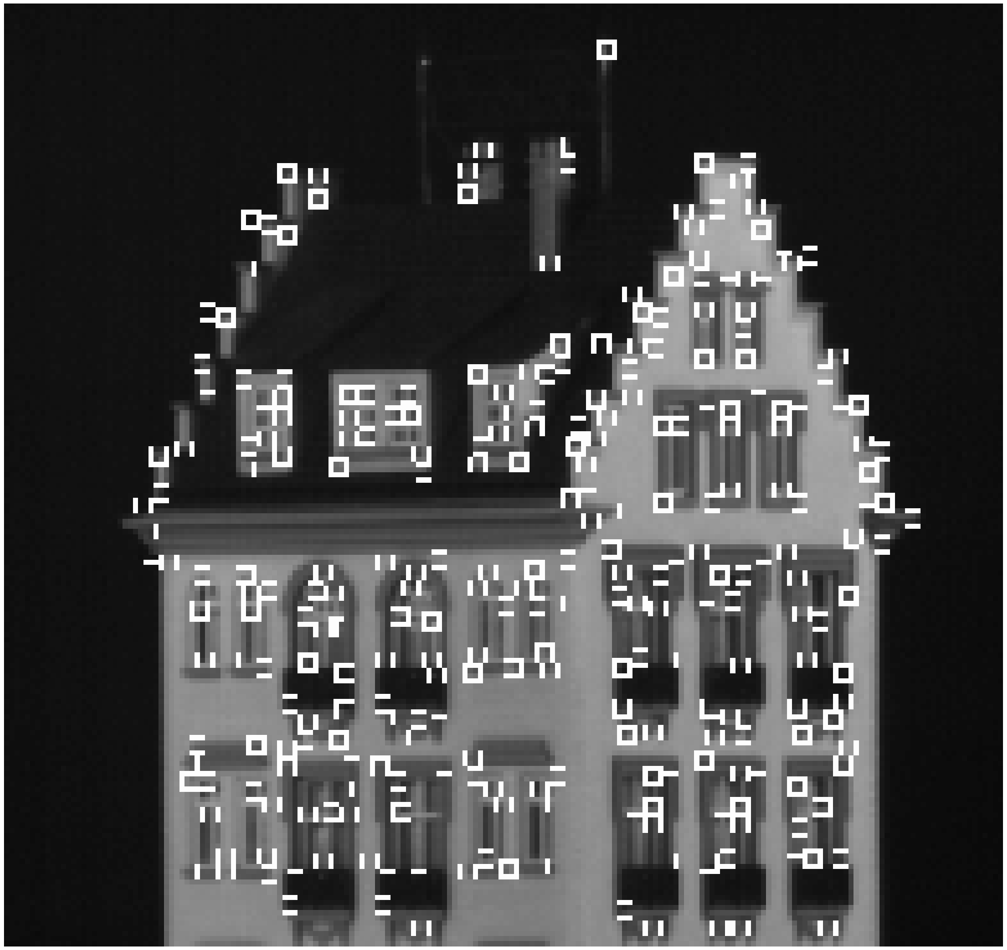


150

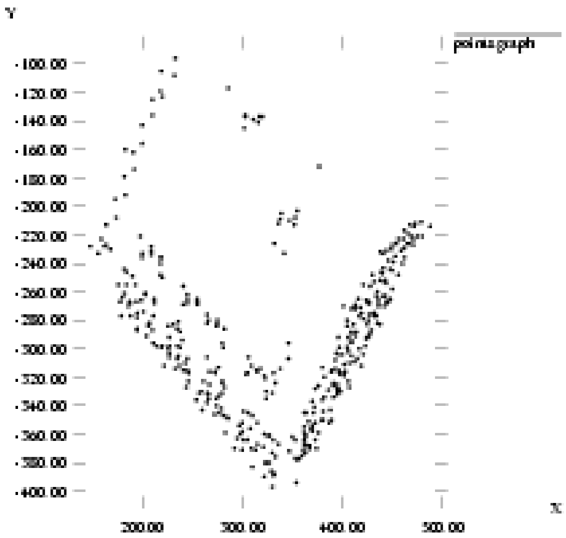
Results (rotations)

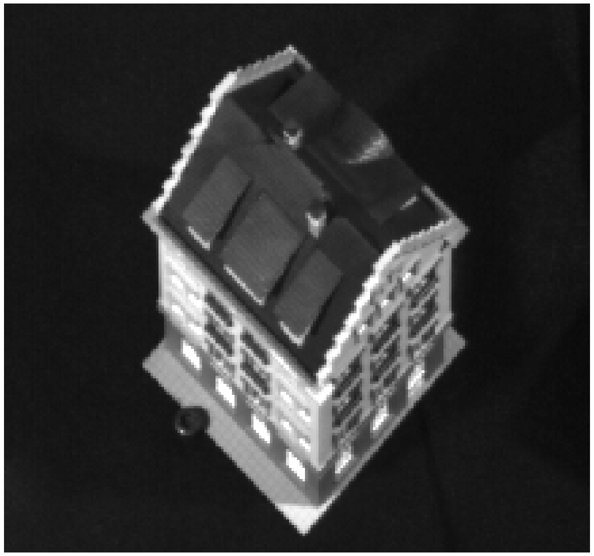


Selected Features

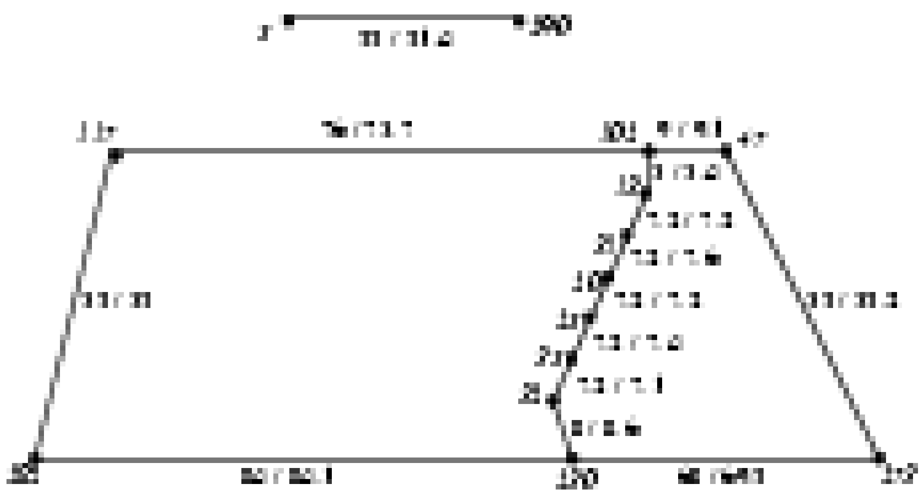


Reconstructed Shape

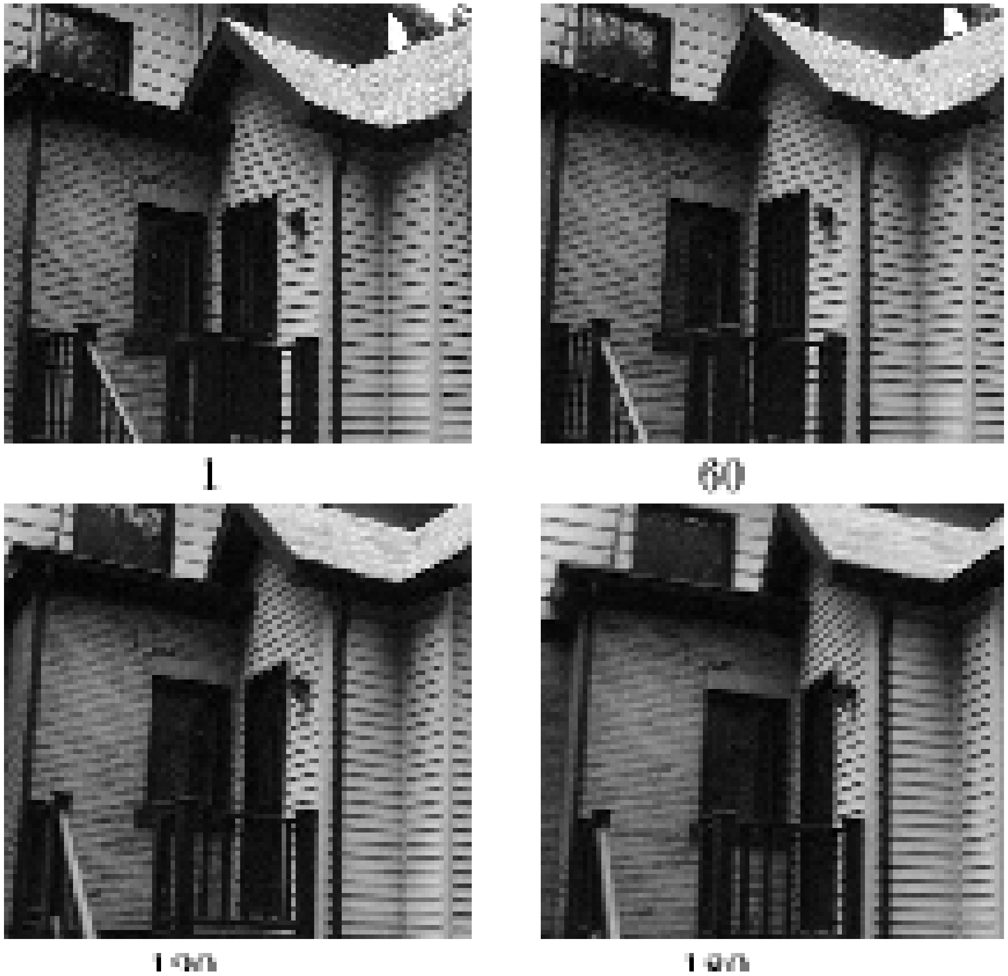




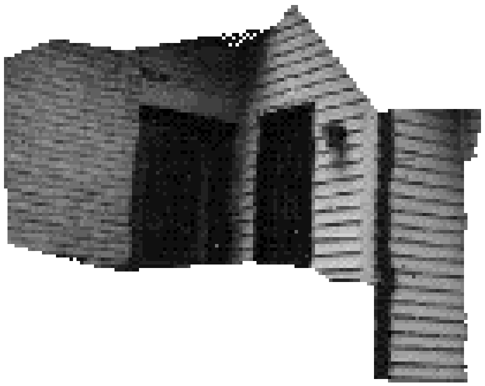
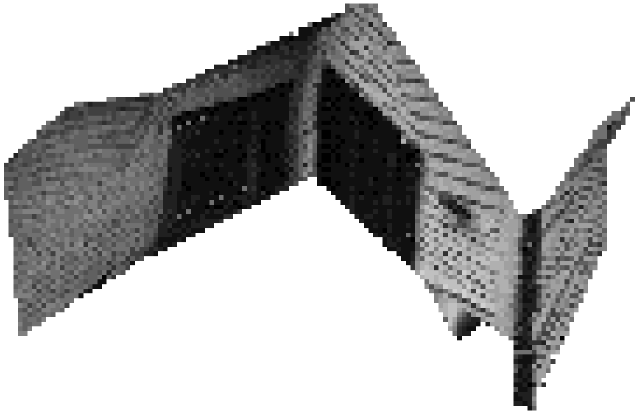
Comparison



House Sequence



Reconstructed Walls

Further Reading

- C. Tomasi and T. Kanade. Shape and motion from image streams under orthography---a factorization method. *International Journal on Computer Vision*, 9(2):137-154, November 1992.
- Computer Vision: Algorithms and Applications, Richard Szeliski, Section 7.3