Reconstruction from Multiple Views

Daniel DeMenthon

3D Reconstruction from Image Pairs

- Find interest points
- Match interest points
- Compute fundamental matrix **F**
- Compute camera matrices **P** and **P'** from **F**
- For each matching image points **x** and **x'**, compute point **X** in scene

Computing Scene Point from Two Matching Image Points

- We now have computed **P** and **P'** from **F**
- Problem: find **X** from **x** and **x'**
- **x = P X, x' = P' X**. Combine into a form **A X = 0**
- Solve **A X = 0** using SVD and picking the singular vector corresponding to the smallest singular value
  - Note: Nonlinear methods generally give better results

Computing Scene Point from Two Matching Image Points (Details)

$$\mathbf{x} = \mathbf{P}\,\mathbf{X} \Leftrightarrow \mathbf{x} \times (\mathbf{P}\,\mathbf{X}) = \mathbf{0}$$

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \qquad \mathbf{P}\,\mathbf{X} = \begin{bmatrix} \mathbf{P}_1^{\mathrm{T}} \\ \mathbf{P}_2^{\mathrm{T}} \\ \mathbf{P}_3^{\mathrm{T}} \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{P}_1^{\mathrm{T}}\mathbf{X} \\ \mathbf{P}_2^{\mathrm{T}}\mathbf{X} \\ \mathbf{P}_3^{\mathrm{T}}\mathbf{X} \end{bmatrix}$$

$$y(\mathbf{P}_3^{\mathrm{T}}\mathbf{X}) - (\mathbf{P}_2^{\mathrm{T}}\mathbf{X}) = 0$$

$$x(\mathbf{P}_3^{\mathrm{T}}\mathbf{X}) - (\mathbf{P}_1^{\mathrm{T}}\mathbf{X}) = 0$$

$$x(\mathbf{P}_2^{\mathrm{T}}\mathbf{X}) - y(\mathbf{P}_1^{\mathrm{T}}\mathbf{X}) = 0$$

Linear combination of other 2 equations

Computing Scene Point from Two Matching Image Points (End)

- Homogeneous system **A X** = 0 is

$$\begin{bmatrix} x\mathbf{P}_3^{\mathbf{T}} & -\mathbf{P}_1^{\mathbf{T}} \\ y\mathbf{P}_3^{\mathbf{T}} & -\mathbf{P}_2^{\mathbf{T}} \\ x'\mathbf{P}_3^{\mathbf{T}} & -\mathbf{P'}_1^{\mathbf{T}} \\ y'\mathbf{P}_3^{\mathbf{T}} & -\mathbf{P'}_2^{\mathbf{T}} \end{bmatrix} \mathbf{X} = 0$$

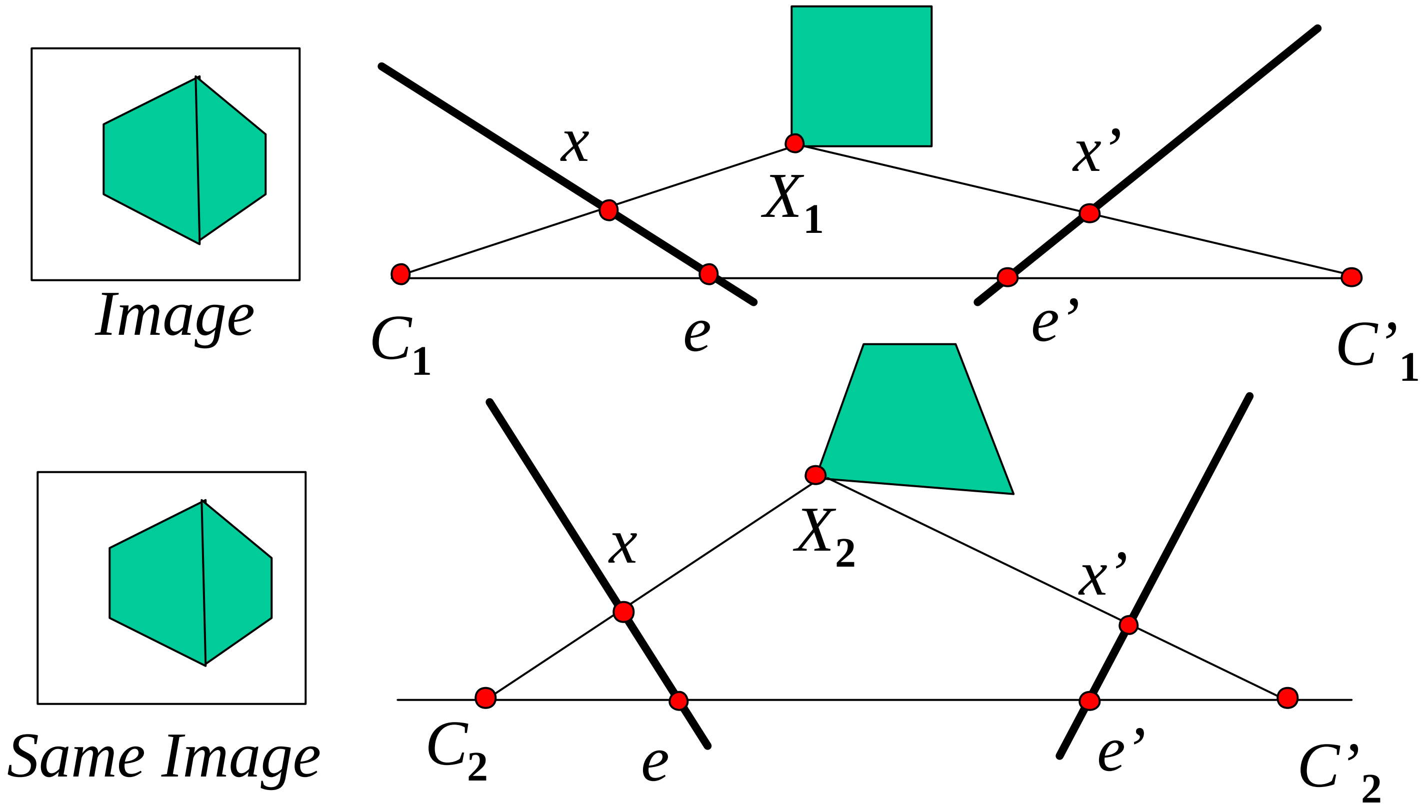
- **X** is the last column of **V** in the SVD of **A**, $\mathbf{A} = \mathbf{U}\,\mathbf{D}\,\mathbf{V}^{\mathbf{T}}$

Projective Reconstruction Theorem

- Assume we determine matching points $\mathbf{x_i}$ and $\mathbf{x'_i}$. Then we can compute a unique fundamental matrix **F**
- The recovered camera matrices are not unique: $(\mathbf{P_1}, \mathbf{P'_1})$, $(\mathbf{P_2}, \mathbf{P'_2})$, etc.
- The reconstruction is not unique: $\mathbf{X_{1i}}$, $\mathbf{X_{2i}}$, etc.
- There exists a projective transformation **H** such that $\mathbf{X_{2i}} = \mathbf{H}\,\mathbf{X_{1i}}$, $\mathbf{P_2} = \mathbf{P_1}\,\mathbf{H}^{-1}$, $\mathbf{P'_2} = \mathbf{P'_1}\,\mathbf{H}^{-1}$

Projective Reconstruction Theorem (Details)

- There exists a projective transformation **H** such that $\mathbf{X}_{2i} = \mathbf{H}\,\mathbf{X}_{1i}$, $\mathbf{P}_2 = \mathbf{P}_1\,\mathbf{H}^{-1}$, $\mathbf{P'}_2 = \mathbf{P'}_1\,\mathbf{H}^{-1}$



$$\mathbf{P}_2\,\mathbf{X}_2 = \mathbf{P}_1\,\mathbf{H}^{-1}\,\mathbf{X}_2 = \mathbf{P}_1\,\mathbf{H}^{-1}\,\mathbf{H}\,\mathbf{x}_1 = \mathbf{P}_1\,\mathbf{x}_1 = \mathbf{x}$$

Projective Reconstruction Theorem (Consequences)

- We can compute a projective reconstruction of a scene from 2 views based on image correspondences alone
- We don't have to know anything about the calibration or poses of the cameras
- The true reconstruction is within a projective transformation **H** of the projective reconstruction: $\mathbf{X}_{2i} = \mathbf{H}\,\mathbf{X}_{1i}$

Reconstruction Ambiguities

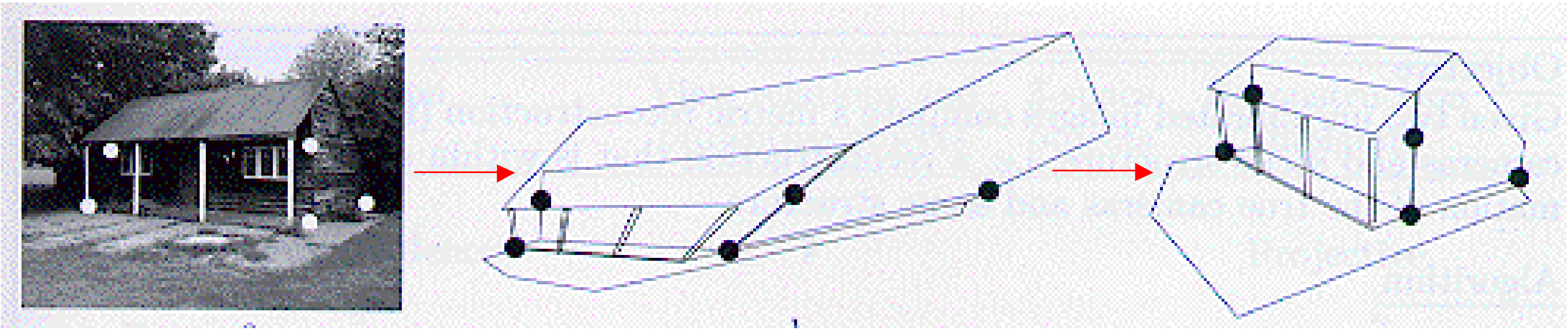
- If the reconstruction is derived from real images, there is a **true** reconstruction that can produce the actual points **X**i of the scene
- Our reconstruction may differ from the actual one
  - If the cameras are calibrated but their relative pose is unknown, then angles between rays are the true angles, and the reconstruction is correct within a similarity (we cannot get the scale)
    - Euclidean or metric reconstruction
  - If we don't use calibration, then we get a projective reconstruction

Rectifying Projective Reconstruction to Metric

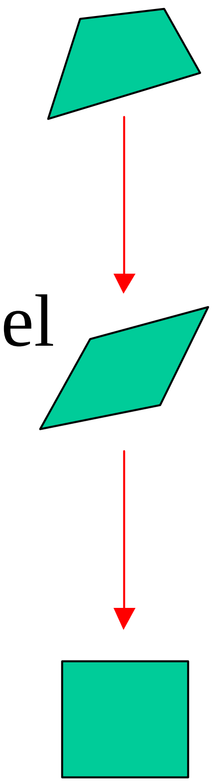
- Compute homography H such that $\mathbf{X}_{\mathbf{Ei}} = \mathbf{H}\,\mathbf{X}_{\mathbf{i}}$ for five or more ground control points $\mathbf{X}_{\mathbf{Ei}}$ with known Euclidean positions
  - H is a 4 x 4 homogeneous matrix
- Then the metric reconstruction is

$$\mathbf{P}_{\mathbf{M}} = \mathbf{P}\,\mathbf{H}^{-1},\ \ \mathbf{P'}_{\mathbf{M}} = \mathbf{P'}\,\mathbf{H}^{-1},\ \mathbf{X}_{\mathbf{Mi}} = \mathbf{H}\,\mathbf{X}_{\mathbf{i}}$$

Results using 5 points



Stratified Reconstruction



- Begin with a projective reconstruction
- Refine it to an affine reconstruction
  - Parallel lines are parallel; ratios along parallel lines are correct
  - Reconstructed scene is then an affine transformation of the actual scene
- Then refine it to a metric reconstruction
  - Angles and ratios are correct
  - Reconstructed scene is then a scaled version of actual scene

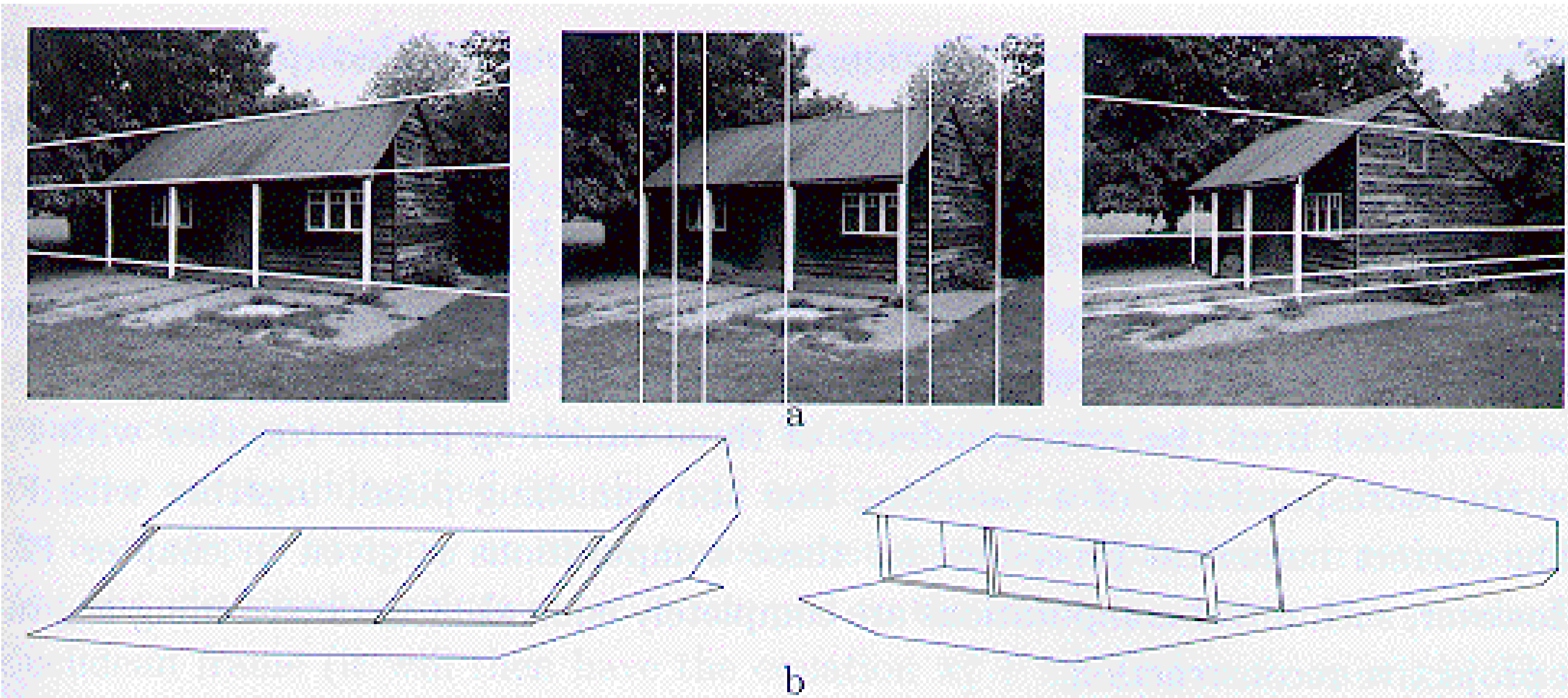
From Projective to Affine Reconstruction

- Find 3 intersections of sets of lines in the scene that are supposed to be parallel
  - These 3 points define a plane *p*
- Find a transformation **H** that maps the plane *p* to the plane at infinity $(0, 0, 0, 1)^T$:
  - This plane contains all points at infinity: $(0, 0, 0, 1)\,(x, y, z, 0)^{\mathbf{T}} = 0$
  - $\mathbf{H}^{-\mathbf{T}} p = (0, 0, 0, 1)^{\mathbf{T}}$, or $\mathbf{H}^{\mathbf{T}}(0, 0, 0, 1)^{\mathbf{T}} = p$

$$\begin{bmatrix} 1 & 0 & 0 & p_1 \\ 0 & 1 & 0 & p_2 \\ 0 & 0 & 1 & p_3 \\ 0 & 0 & 0 & p_4 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} \Rightarrow \mathbf{H} = \begin{bmatrix} \mathbf{I} \,|\, \mathbf{0} \\ \mathbf{p}^{\mathrm{T}} \end{bmatrix}$$

Apply **H** to scene points, and to cameras **P** and **P'**

Example of Affine Reconstruction



From Affine to Metric Reconstruction

- Use constraints from scene orthogonal lines
- Use constraints arising from having the same camera in both images

Direct Metric Reconstruction using Camera Calibration

- Find calibration matrices **K** and **K'** using 3 vanishing points for orthogonal scene lines
  - See homework
- Normalize image points
- Compute fundamental matrix using matched normalized points: *we get the essential matrix* **E**
- Select **P=[I | 0]** and **P'=[R |T]**. Then $\mathbf{E} = [\mathbf{T}]_{\times} \mathbf{R}$
- Find **T** and **R** using SVD of **E**
- From **P** and **P'**, reconstruct scene points

Reconstruction from N Views

- Projective or affine reconstruction from a possibly large set of images
- Problem
  - Set of 3D points $\mathbf{X_j}$
  - Set of cameras $\mathbf{P^i}$
  - For each camera $\mathbf{P^i}$, set of image points $\mathbf{x_j^i}$
  - Find 3D points $\mathbf{X_j}$ and cameras $\mathbf{P^i}$ such that $\mathbf{P^i X_j = x_j^i}$

Bundle Adjustment

- Solve following minimization problem
  - Find $\mathbf{P^i}$ and $\mathbf{X_j}$ that minimize

$$\sum_{\mathbf{i,j}} d(\mathbf{P^i}\ \mathbf{X_j}, \mathbf{x^i_j})^2$$

  - Levenberg-Marquardt algorithm
  - Problems:
    - Many parameters: 11 per camera, 3 per 3D point
      - Matrices (11 m + 3 n) x (11 m + 3 n)
      - Good initialization required
- Mainly used as final adjustment step of the bundle of rays

Initial Solutions: Affine Factorization Algorithm

- Tomasi and Kanade (1992)
- Affine reconstruction
- Affine camera

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & T_1 \\ m_{21} & m_{22} & m_{23} & T_2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{M} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \mathbf{T}$$

- Inhomogeneous coordinates $\mathbf{x} = \mathbf{M}\ \mathbf{X} + \mathbf{T}$

Affine Factorization

- Minimize $\sum_{i,j} (\mathbf{x}_j^i - (\mathbf{M}^i \mathbf{X}_j + \mathbf{T}^i))^2$
- Choose centroid of points as origin of scene coordinate system
- Choose pixel (0, 0) at image of centroid
- Then the problem becomes:
  Minimize $\sum_{i,j} (\mathbf{x}_j^i - \mathbf{M}^i \mathbf{X}_j))^2$
  - Note: This requires the same points to be visible in all views

Affine Factorization

- Consider the measurement matrix (one row per image point)

$$\mathbf{W} = \begin{bmatrix} \mathbf{x}_1^1 & \mathbf{x}_2^1 & \cdots & \mathbf{x}_n^1 \\ \mathbf{x}_1^2 & \mathbf{x}_2^2 & \cdots & \mathbf{x}_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_1^m & \mathbf{x}_2^m & \cdots & \mathbf{x}_n^m \end{bmatrix}$$

- The projection matrix is $\hat{\mathbf{W}} = \begin{bmatrix} \mathbf{M}^1 \\ \mathbf{M}^2 \\ \vdots \\ \mathbf{M}^m \end{bmatrix} [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \cdots \quad \mathbf{X}_\mathbf{n}]$

- Minimize $\left\| \mathbf{W} - \hat{\mathbf{W}} \right\|$

Affine Factorization

- Minimize $\|\mathbf{W} - \hat{\mathbf{W}}\|$
- Find $\hat{\mathbf{W}}$ as the SVD of **W** truncated to rank 3:

$$\hat{\mathbf{W}} = \mathbf{U}_{\mathbf{2m \times 3}}\, \mathbf{D}_{\mathbf{3 \times 3}}\, \mathbf{V}^{\mathbf{T}}_{\mathbf{n \times 3}}$$

- Then **M** may be chosen as **U D** and **X** as $\mathbf{V}^{\mathbf{T}}$
- This decomposition is not unique:

$$\hat{\mathbf{W}} = \hat{\mathbf{M}}\, \hat{\mathbf{X}} = (\hat{\mathbf{M}}\, \mathbf{A})\,(\mathbf{A}^{-1}\hat{\mathbf{X}})$$

- Reconstruction is defined up to a matrix **A**
- Reconstruction is affine
- To upgrade to a metric reconstruction, see above

Projective Factorization

- $\mathbf{x}_j^i = \mathbf{P}^i \mathbf{X}_j,\ \mathbf{x}_j^i = (u_j^i, v_j^i, w_j^i) = w_j^i (x_j^i, y_j^i, 1) = w_j^i \mathbf{x}'^i_j$

$$\begin{bmatrix} w_1^1 \mathbf{x}_1^1 & w_2^1 \mathbf{x}_2^1 & \cdots & w_n^1 \mathbf{x}_n^1 \\ w_1^2 \mathbf{x}_1^2 & w_2^2 \mathbf{x}_2^2 & \cdots & w_n^2 \mathbf{x}_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ w_1^m \mathbf{x}_1^m & w_2^m \mathbf{x}_2^m & \cdots & w_n^m \mathbf{x}_n^m \end{bmatrix} = \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \vdots \\ \mathbf{P}^m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_1 & \cdots & \mathbf{X}_n \end{bmatrix}$$

- $$\hat{\mathbf{W}} = \begin{bmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \vdots \\ \mathbf{P}^m \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_1 & \cdots & \mathbf{X}_n \end{bmatrix}$$

• The w_j^i are unknown, related to the depths of points in camera coordinates

•We dropped the primes on $\mathbf{x}_j^i$

- $\mathbf{W}$ has rank 4. Assume the w_j^i coefficients known

$$\hat{\mathbf{W}}_{3m\times n} = \mathbf{U}_{3m\times 4}\, \mathbf{D}_{4\times 4}\, \mathbf{V}^{\mathbf{T}}_{n\times 4}$$

Projective Factorization

1. Start with an initial estimate of the depths $w_{\mathbf{j}}^{\mathbf{i}}$
2. From the measurement matrix **W**, find the nearest rank 4 approximation using the SVD and decompose to find the camera matrices and 3D points
3. Reproject the points into each image to obtain new estimates of the depths and repeat from step 2

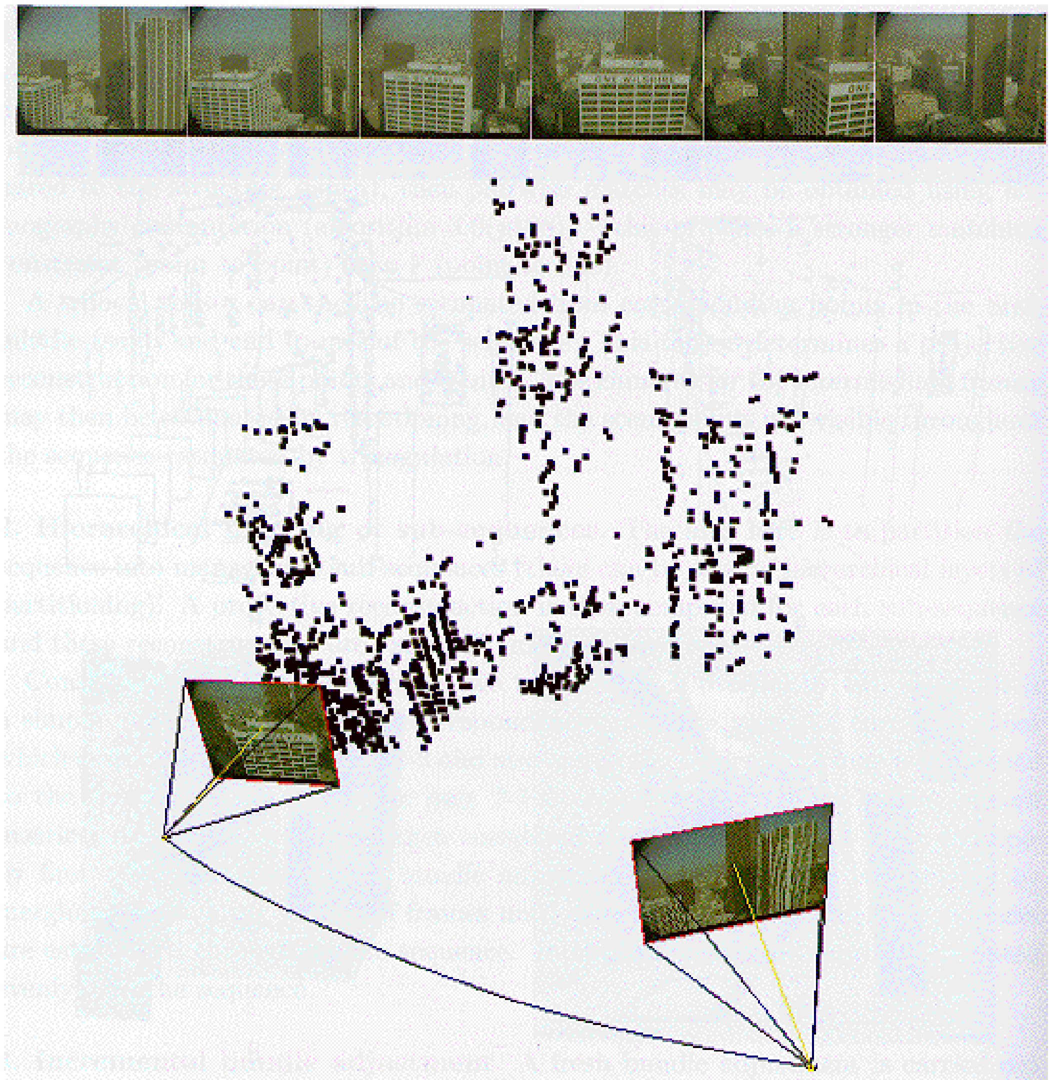
Reconstruction from Video Sequences

- Compute interest points in each image
- Compute interest point correspondences between image pairs
- Compute fundamental matrix **F** for each image pair
- Initial reconstruction
- Bundle-adjust the cameras and 3D structure to minimize projection errors

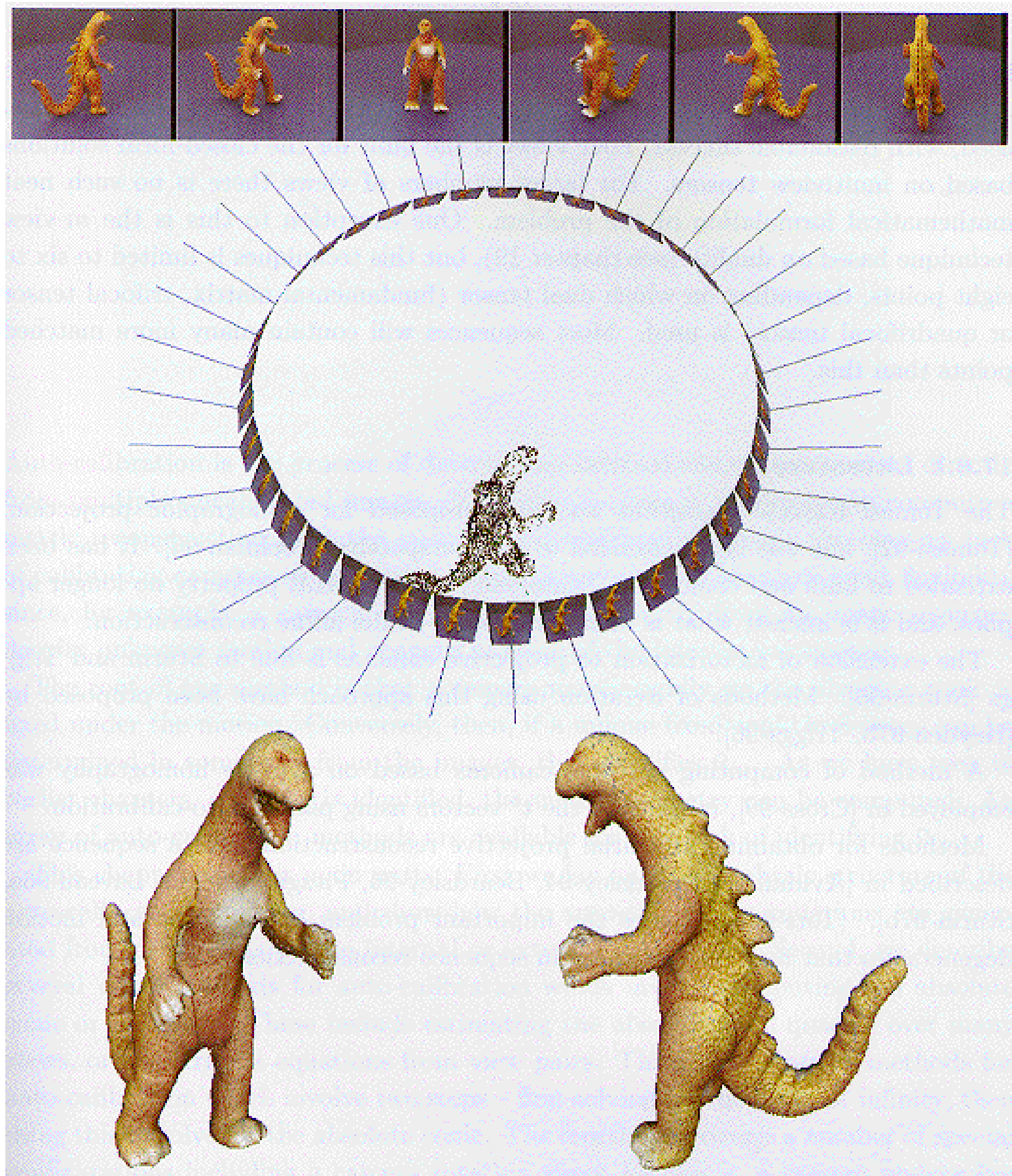
Issues for Videos

- Small baseline between image pairs
  - Advantage: having similar images facilitates finding point correspondences
  - Disadvantage: 3D structure is estimated poorly for each image pair
- Solutions:
  - Use consecutive images for point correspondences, and images far apart for 3D structure reconstruction
  - Make small batches and combine them by least square
  - Use recursive least square method

Examples of 3D Reconstruction



Examples of 3D Reconstruction



References

- Multiple View Geometry in Computer Vision, R. Hartley and A. Zisserman, Cambridge University Press, 2000.
- D. Forsyth and J. Ponce. Computer Vision: A Modern Approach, http://www.cs.berkeley.edu/~daf/book3chaps.html
  - Geometry of Multiple Views (Chapter 12)
  - Stereopsis (Chapter 13)
  - Affine Structure from Motion (Chapter 14)
  - Projective Structure from Motion (Chapter 15)