CS 664 Slides #10
Structure From Motion

Prof. Dan Huttenlocher
Fall 2003
Structure From Motion

- Recover 3D coordinates from set of 2D views
  - Rigid body motion
  - Known correspondence of points in views
  - Various camera models

- Consider representative case of
  - Parallel (orthographic) projection
  - All points visible in all views
  - Un-calibrated camera
  - No outliers (least squares ok)
Parallel Projection

- Point \((X,Y,Z)\) in space projects to \((X,Y)\) in image plane
  - Contrast with \((fX/Z, fY/Z)\) in pinhole model
  - Light rays all parallel rather than through principal point
    - Similar when points at same depth, narrow FOV
Recovering 3D Structure

- With enough corresponding points and views can determine 3D locations
  - Redundant information
    - Each view changes only viewing parameters and not point locations
      - 3P unknowns for P points and kF unknowns for F views

- Minimum sufficient correspondences
  - Orthographic projection, three views of four points
  - Central (pinhole) projection, two views of eight points
Sensitive to Measurement Noise

- Solutions based on a small number of points are not stable
  - Errors of the magnitude found in most images yield substantial differences in recovered 3D values

- Method that works in practice called factorization
  - Works on sequence of several frames
  - With correspondences of points
  - Consider case of factorization for orthographic projection, no outliers, can be extended
Input: Sequence of Tracked Points

- Point coordinates
  \[ w'_{fp} = (u'_{fp}, v'_{fp}) \]
  - Where \( f \) denotes frame index and \( p \) denotes point index
  - Points tracked over frames
    * E.g., use corner trackers discussed previously
Centroid Normalized Coordinates

- From observed coordinates \( w'_{fp} = (u'_{fp}, v'_{fp}) \)
  \[ w_{fp} = (u'_{fp} - \bar{u}_{fp}, v'_{fp} - \bar{v}_{fp}) \]
  - Where
  \[ u_{fp} = \frac{1}{P} \sum_{p} u'_{fp} \]
  and
  \[ \bar{v}_{fp} = \frac{1}{P} \sum_{p} v'_{fp} \]
Normalization

- Goal of separating out effects of camera translation from those of rotation
- Subtract out centroid to remove translation effects
  - Assume all points belong to object and present at all frames
  - Centroid preserved under projection
- Left to recover 3D coordinates (shape) of P points from F camera orientations
Measurement Matrix

- $2F \times P$ – 2 rows per frame, one col per point
- In absence of sensor noise this matrix is highly rank deficient
  - Under orthographic projection rank 3 or less

$$W = \begin{bmatrix}
  u_{11} & \ldots & u_{1P} \\
  \vdots & & \vdots \\
  u_{F1} & \ldots & u_{FP} \\
  v_{11} & \ldots & v_{1P} \\
  \vdots & & \vdots \\
  v_{F1} & \ldots & v_{FP}
\end{bmatrix}$$
Structure of W

- World point $s_p' = (x_p', y_p', z_p')$ projects to image points
  
  $$u'_{fp} = m_f^T (s_p' - t_f)$$
  $$v'_{fp} = n_f^T (s_p' - t_f)$$

  - Where $m_f, n_f$ are unit vectors defining orientation of image plane in world
  - And $t_f$ is vector from world origin to image plane origin
Structure of $W$ (Cont’d)

- Can rewrite in centroid normalized coordinates
  - Since centroid preserved under projection
  - Projection of centroid is centroid of projection
    \[ u_{fp} = m_f^T s_p \]
    \[ v_{fp} = n_f^T s_p \]
  - Where
    \[ s_p = s_p' - \bar{s} \]
    \[ \bar{s} = \frac{1}{P} \sum_p s_p' \]
W Factors Into Simple Product

- $W = MS$ where
  - $M$ is $2Fx3$ matrix of camera locations
  - $S$ is $3xP$ matrix of points in world
  - Product is $2Fx3$ matrix $W$
  - Clearly rank at most 3

\[
M = \begin{bmatrix}
m_1^T \\
\vdots \\
m_F^T \\
n_1^T \\
\vdots \\
n_F^T
\end{bmatrix} \quad S = \begin{bmatrix} s_1 & \cdots & s_P \end{bmatrix}
\]
**Factoring W**

- Don’t know $M,S$ only measurements $W$
- When noise or errors in measurements seek least squares approximation
  - Note l.s. assumes no outliers (bad data)
    \[ \text{argmin}_{M,S} \| W - MS \|^2 \]
- The best $M,S$ of this form can be found using the SVD of $W$
  \[ W = U \Sigma V \]
  \[ \Sigma' \text{ contains only three largest singular values} \]
  \[ M^* = U \left( \Sigma' \right)^{1/2} \]
  \[ S^* = \left( \Sigma' \right)^{1/2} V \]
Factorization Not Unique

- Any linear transformation of $M, S$ possible
  \[ W = MS = M(LL^{-1})S = (ML)(L^{-1}S) \]
- Often referred to as “affine shape”
  - Preserves parallelism/coplanarity
- Still haven’t used a constraint on the form of $M$
  - Describes camera plane orientation at each frame
    \[ \begin{bmatrix} m_1^T \\ \vdots \\ m_F^T \\ n_1^T \\ \vdots \\ n_F^T \end{bmatrix} \]
    
    \[ m_i, n_i \text{ all unit vectors} \]
    
    \[ m_i n_i = 0 \]
Factorization Results

1 40
60 80
120 150