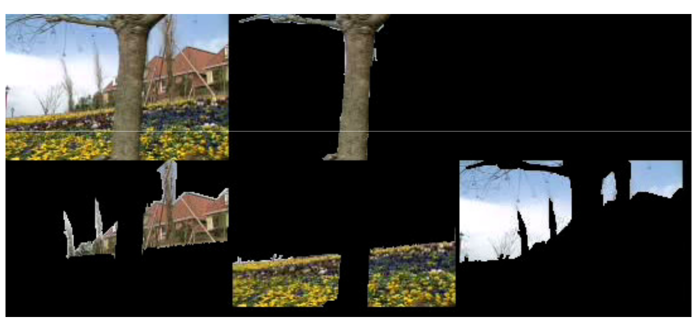
Structure From Motion

- Okay, we have computed motion, what can we do with it?

1. Motion Segmentation…
2. Structure from Motion

Motion Segmentation



Structure From Motion

Structure from Motion

- Problem Definition:
  - Given the motion field estimated from an image sequence, compute the shape, or *structure*, of the visible objects, and their *motion* with respect to the viewing camera.
- Various methods exist using either *sparse* motion field or *dense* motion field
- We will look at one method which uses *sparse* motion field

Structure from Motion: Factorization Method

- Tomasi/Kanade 1992
- Assumptions
  1. The camera model is orthographic
  2. The postion of n image points, corresponding to scene points $\mathbf{p}_1, \mathbf{p}_2, \ldots \mathbf{p}_n$ (not all coplanar) have been tracked in m frames (with $m \geq 3$)

Notations

- n 3D points are seen in m views
- $\mathbf{q}=(u,v,1)$: *2D image point*
- $\mathbf{p}=(x,y,z,1)$: *3D scene point*
- $\mathbf{R}$: projection matrix

SFM under orthographic projection

$$\underset{2\times1}{\overset{\text{2D image point}}{\mathbf{q}}} = \underset{2\times3}{\overset{\text{orthographic projection matrix}}{\mathbf{R}}}\,\underset{3\times1}{\overset{\text{3D scene point}}{\mathbf{p}}} + \underset{2\times1}{\overset{\text{image offset}}{\mathbf{t}}}$$

- Trick
  - Choose scene origin to be centroid of 3D points
  - Choose image origins to be centroid of 2D points
  - Allows us to drop the camera translation:

$$\mathbf{q} = \mathbf{R}\mathbf{p}$$

- Slide by Yaser Sheikh

Factorization (Tomasi & Kanade)

projection of *n* features in one image:

$$\underset{2\times n}{\begin{bmatrix}\mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n\end{bmatrix}} = \underset{2\times 3}{\mathbf{R}}\,\underset{3\times n}{\begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix}}$$

projection of *n* features in *m* images

$$\underset{2m\times n}{\begin{bmatrix}\mathbf{q}_{11} & \mathbf{q}_{12} & \cdots & \mathbf{q}_{1n}\\ \mathbf{q}_{21} & \mathbf{q}_{22} & & \mathbf{q}_{2n}\\ \vdots & \vdots & & \vdots\\ \mathbf{q}_{m1} & \mathbf{q}_{m2} & \cdots & \mathbf{q}_{mn}\end{bmatrix}} = \underset{2m\times 3}{\begin{bmatrix}\mathbf{R}_1\\ \mathbf{R}_2\\ \vdots\\ \mathbf{R}_m\end{bmatrix}}\,\underset{3\times n}{\begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix}}$$

W measurement **M** motion **S** shape

- Slide by Yaser Sheikh

Rank Theorem

- The measurement matrix **W** (without noise) has at most rank 3

> Key Observation: *rank*(**W**) <= 3

- Why?

Factorization

known $\underset{2m\times n}{\mathbf{W}} = \boxed{\underset{2m\times 3}{\mathbf{M}}\ \underset{3\times n}{\mathbf{S}}}$ solve for

- Factorization Technique
  - Goal is to factorize **W** into two matrices **M** and **S**
  - Approach:
    - **W** is at most rank 3 (assuming no noise)
    - We can use *singular value decomposition* to factor **W**:
      - $\mathbf{W} = \mathbf{U}\ \mathbf{D}\ \mathbf{V}^T$
      - D is 2m x 2m, but in the absence of noise, should have only 3 non-zero singular values

Factorization

- **W = U D V^T**
- Make a new **D'** which contains only the first three singular values in D on its diagonal.
- Discard all columns of U other than first 3, and all rows of V other than first 3
- **W'** = **U' D' V'**^{T}
- Then M' = U' D'$^{1/2}$
- And S' = D'$^{1/2}$V'^{T}

Factorization

- Is this factorization unique?
- No, correct only upto a 3x3 linear transformation
- Proof: Consider an arbitrary 3x3 transformation Q whose inverse exists
- Then W' = M' S' = M' Q Q^{-1} S'

Metric Upgrade

- How to find Q so that correct M and S can be recovered?
- Trick: Rows of correct M should be pair-wise orthonormal
- Notation: r_i is the i th row of **M'**
- Then

  $r_i Q Q^T r_i^T = 1$ for all I

  $r_i Q Q^T r_j^T = 0$ for adjacent pairs of rows
- Q has 9 unknowns and these constraints can be used to solve for it.
- Once Q is known, M = M'Q and S = Q^{-1}S'

Limitations

- Orthographic camera is assumed, so will only work for sequences in which distance of camera is large compared to the depth variation of object
- Structure recovery is up to a rotation only
- Translation component in the direction of optical axis cannot be recovered

Complete Algorithm

The input is the registered measurement matrix $\tilde{W}$, computed from n features tracked over N consecutive frames.

1. Compute the SVD of $\tilde{W}$,

$$\tilde{W} = U D V^\top,$$

   where U is a $2N \times 2N$ matrix, V $n \times n$, and D $2N \times n$; $U^\top U = I$, $V^\top V = I$; and D is the diagonal matrix of the singular values.
2. Set to zero all but the three largest singular values in D.
3. Define $\hat{R}$ and $\hat{S}$ as in (8.5.1).
4. Solve (8.39) for Q,for example by means of Newton's method (Exercise 8.8).

The output are the rotation and shape matrices, given by

$$R = \hat{R}Q \quad \text{and} \quad S = Q^{-1}\hat{S}.$$

Structure From Motion