Factorization Method
Photo synth


http://photosynth.net/
Scene Feature $s_p$

$w = [u_{1p}, v_{1p}]$

$w = [u_{2p}, v_{2p}]$

$w = [u_{Fp}, v_{Fp}]$
Structure from motion under orthographic projection

3D Reconstruction of a Rotating Ping-Pong Ball

First practical 3D reconstruction

Factorization


Goal: combine point correspondence information from multiple points over multiple frames to solve for scene structure and camera motion (structure from motion)

Approach: numerically stable approach based on using SVD to “factor” matrix of observed point positions.

Historical significance: until that time, most SFM work dealt with minimal configurations, and noise-free data. Factorization was one of the first “practical SFM algorithms”
Shape and Motion from Image Streams under Orthography: a Factorization Method

CARLO TOMASI  
Department of Computer Science, Cornell University, Ithaca, NY 14850

TAKEO KANADE  
School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213

Received

Abstract  
Inferring scene geometry and camera motion from a stream of images is possible in principle, but is an ill-conditioned problem when the objects are distant with respect to their size. We have developed a factorization method that can overcome this difficulty by recovering shape and motion under orthography without computing depth as an intermediate step.

An image stream can be represented by the $2F \times P$ measurement matrix of the image coordinates of $P$ points tracked through $F$ frames. We show that under orthographic projection this matrix is of rank 3.

Based on this observation, the factorization method uses the singular-value decomposition technique to factor the measurement matrix into two matrices which represent object shape and camera rotation respectively. Two of the three translation components are computed in a preprocessing stage. The method can also handle and obtain a full solution from a partially filled-in measurement matrix that may result from occlusions or tracking failures.

The method gives accurate results, and does not introduce smoothing in either shape or motion. We demonstrate this with a series of experiments on laboratory and outdoor image streams, with and without occlusions.
Recall: World to Camera Transform

\[ P_C = R \left( P_W - C \right) \]

\[
\begin{pmatrix}
P^C_x \\
P^C_y \\
P^C_z \\
1
\end{pmatrix} =
\begin{pmatrix}
r_{11} & r_{12} & r_{13} & 0 \\
r_{21} & r_{22} & r_{23} & 0 \\
r_{31} & r_{32} & r_{33} & 0 \\
0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
1 & 0 & 0 & -c_x \\
0 & 1 & 0 & -c_y \\
0 & 0 & 1 & -c_z \\
0 & 0 & 0 & 1
\end{pmatrix} \begin{pmatrix}
P^W_x \\
P^W_y \\
P^W_z \\
1
\end{pmatrix}
\]

\[ P_C = M_{ext} \cdot P^W \]
Perspective Projection

\[ x = f \frac{X}{Z} \]
\[ y = f \frac{Y}{Z} \]

- Non-linear equations
- Any point on the ray OP has image p!!
Perspective Projection

\[ x = f \frac{X}{Z} \]

\[ y = f \frac{Y}{Z} \]

**Perspective Projection**: parallel lines appear to meet at a vanishing point; farther objects seem smaller

O. Camps, PSU
Simplification: Weak Perspective

\[ x = \frac{f}{Z_o} X \]

\[ y = \frac{f}{Z_o} Y \]

Weak perspective = Parallel projection (parallel lines remain parallel) + Scaling to simulate change in size due to object distance.
Simpler: Orthographic Projection

\[ x = X \]

\[ y = Y \]

Pure parallel projection. Highly simplified case where we even ignore the scaling due to distance.
Perspective Matrix Equation
(Camera Coordinates)

\[ x = f \frac{X}{Z} \]
\[ y = f \frac{Y}{Z} \]

Using homogeneous coordinates:

\[
\begin{bmatrix}
  x' \\
  y' \\
  z'
\end{bmatrix} = \begin{bmatrix}
  f & 0 & 0 & 0 \\
  0 & f & 0 & 0 \\
  0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  X \\
  Y \\
  Z \\
  1
\end{bmatrix}
\]

\[ x = \frac{x'}{z'} \quad y = \frac{y'}{z'} \]
Weak Perspective Approximation

Using homogeneous coordinates:

\[
x = \frac{f}{Z_o} X
\]

\[
y = \frac{f}{Z_o} Y
\]

\[
\begin{bmatrix}
    x \\
    y \\
    1
\end{bmatrix}
= \begin{bmatrix}
    f/Z_o & 0 & 0 & 0 \\
    0 & f/Z_o & 0 & 0 \\
    0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
    X \\
    Y \\
    Z
\end{bmatrix}
\]
Let’s Consider Orthographic

Using homogeneous coordinates:

\[
\begin{bmatrix}
  x \\
  y
\end{bmatrix} =
\begin{bmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
  X \\
  Y \\
  Z \\
  1
\end{bmatrix}
\]

\[x = X\]
\[y = Y\]
Combine with External Params

\[
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
= 
\begin{pmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
  r_{11} & r_{12} & r_{13} & 0 \\
  r_{21} & r_{22} & r_{23} & 0 \\
  r_{31} & r_{32} & r_{33} & 0 \\
  0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  1 & 0 & 0 & -c_x \\
  0 & 1 & 0 & -c_y \\
  0 & 0 & 1 & -c_z \\
  0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  p^W_x \\
  p^W_y \\
  p^W_z \\
  1
\end{pmatrix}
\]

\[
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
= 
\begin{pmatrix}
  r_{11} & r_{12} & r_{13} & 0 \\
  r_{21} & r_{22} & r_{23} & 0
\end{pmatrix}
\begin{pmatrix}
  1 & 0 & 0 & -c_x \\
  0 & 1 & 0 & -c_y \\
  0 & 0 & 1 & -c_z \\
  0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  p^W_x \\
  p^W_y \\
  p^W_z \\
  1
\end{pmatrix}
\]
Combine with External Params

\[
\begin{pmatrix}
  x \\
y
\end{pmatrix}
= 
\begin{pmatrix}
  r_{11} & r_{12} & r_{13} & 0 \\
r_{21} & r_{22} & r_{23} & 0
\end{pmatrix}
\begin{pmatrix}
  1 & 0 & 0 & -c_x \\
0 & 1 & 0 & -c_y \\
0 & 0 & 1 & -c_z \\
0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
P_x^W \\
P_y^W \\
P_z^W \\
1
\end{pmatrix}
\]

\[
\begin{pmatrix}
x \\
y
\end{pmatrix}
= 
\begin{pmatrix}
r_{11} & r_{12} & r_{13} \\
r_{21} & r_{22} & r_{23}
\end{pmatrix}
\begin{pmatrix}
P_x^W - c_x \\
P_y^W - c_y \\
P_z^W - c_z
\end{pmatrix}
\]
Orthographic: Algebraic Equation

\[
\begin{bmatrix}
  x \\
  y
\end{bmatrix} = \begin{bmatrix}
  r_{11} & r_{12} & r_{13} \\
  r_{21} & r_{22} & r_{23}
\end{bmatrix} \begin{bmatrix}
  i^T \\
  j^T
\end{bmatrix} + \begin{bmatrix}
  P^W_x \\
  P^W_y \\
  P^W_z \\
  c_x \\
  c_y \\
  c_z
\end{bmatrix}
\]

\[
x = i^T (P - T)
\]
\[
y = j^T (P - T)
\]
**Figure 2.** World reference and camera frames used in Factorization method derivation. [Trucco & Verri]
Multiple Points, Multiple Frames

Notation (attack of the killer subscripts)

\[ x = i^T (P - T) \]
\[ y = j^T (P - T) \]

\( N \) points
\[ P_1 \ P_2 \ldots \ P_j \ldots \ P_N \]

\( F \) frames
\[ i_1 \ i_2 \ldots \ i_i \ldots \ i_F \]
\[ j_1 \ j_2 \ldots \ j_i \ldots \ j_F \]
\[ T_1 \ T_2 \ldots T_i \ldots T_F \]

\[ x_{ij} = i_i^T (P_j - T_i) \]
\[ y_{ij} = j_i^T (P_j - T_i) \]

Eq 8.31-8.32
T&V book
Factorization Approach

\[ x_{ij} = i_i^T (P_j - T_i) \]
\[ y_{ij} = j_i^T (P_j - T_i) \]

N points
\[ P_1 \ P_2 \ \ldots \ P_j \ \ldots \ P_N \]
(We want to recover these)

Note that absolute position of the set of points is something that cannot be uniquely recovered, so…

**First Trick:** set the origin of the world coordinate system to be the center of pass of the N points!

\[ \frac{1}{N} \sum_{i=1}^{N} P_i = 0 \]
Factorization Approach

\[ x_{ij} = i_i^T (P_j - T_i) \]
\[ y_{ij} = j_i^T (P_j - T_i) \]

Centroid at 0:
\[ \frac{1}{N} \sum_{i=1}^{N} P_i = 0 \]

Implication:
\[ \bar{x}_{it} = \frac{1}{N} \sum_{i=1}^{N} i_t^T (P_i - T_t) = \frac{1}{N} \sum_{i=1}^{N} i_t^T P_i - \frac{1}{N} \sum_{i=1}^{N} i_t^T T_t = 0 - i_t^T T_t \]

Note: this is the center of mass of x coordinates in frame t
Factorization Approach

\[ \bar{x}_{ti} = \frac{1}{n} \sum_{i=1}^{n} i_t^T (P_i - T_t) = -i_t^T T_t \]

\[ \bar{y}_{ti} = \frac{1}{n} \sum_{i=1}^{n} j_t^T (P_i - T_t) = -j_t^T T_t \]

Second Trick: subtract off the center of mass of the 2D points in each frame. (Centering)

\[ x_{ij} = i_t^T (P_j - T_i) \]

\[ y_{ij} = j_t^T (P_j - T_i) \]

\[ \tilde{x}_{ti} = x_i - \bar{x}_{ti} = i_t^T P_i \]

\[ \tilde{y}_{ti} = y_i - \bar{y}_{ti} = j_t^T P_i \]
Factorization Approach

\[
\begin{align*}
x_{ij} &= i_i^T (P_j - T_i) \\
y_{ij} &= j_i^T (P_j - T_i)
\end{align*}
\]

centering

\[
\begin{align*}
\tilde{x}_{ti} &= x_i - \bar{x}_{ti} = i_t^T P_i \\
\tilde{y}_{ti} &= y_i - \bar{y}_{ti} = j_t^T P_i
\end{align*}
\]

What have we accomplished so far?

1) Removed unknown camera locations from equations.

2) More importantly, we can now write everything as a big matrix equation…
Factorization Approach

Form a matrix of centered image points.

$$\begin{pmatrix}
\tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} & \ldots & \tilde{x}_{1N} \\
\vdots & & & & \\
\tilde{x}_{F1} & \tilde{x}_{F2} & \tilde{x}_{F3} & \ldots & \tilde{x}_{FN} \\
\tilde{y}_{11} & \tilde{y}_{12} & \tilde{y}_{13} & \ldots & \tilde{y}_{1N} \\
\vdots & & & & \\
\tilde{y}_{F1} & \tilde{y}_{F2} & \tilde{y}_{F3} & \ldots & \tilde{y}_{FN}
\end{pmatrix}_{2F \times N}$$

All N points in one frame
Factorization Approach

Form a matrix of centered image points.

$$\begin{pmatrix}
\tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} & \ldots & \tilde{x}_{1N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\tilde{x}_{F1} & \tilde{x}_{F2} & \tilde{x}_{F3} & \ldots & \tilde{x}_{FN} \\
\tilde{y}_{11} & \tilde{y}_{12} & \tilde{y}_{13} & \ldots & \tilde{y}_{1N} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\tilde{y}_{F1} & \tilde{y}_{F2} & \tilde{y}_{F3} & \ldots & \tilde{y}_{FN}
\end{pmatrix}$$

Tracking one point through all F frames.
Factorization Approach

matrix of centered image points:

\[
\begin{pmatrix}
\tilde{x}_{11} & \tilde{x}_{12} & \tilde{x}_{13} & \ldots & \tilde{x}_{1N} \\
\vdots & & & & \\
\tilde{x}_{F1} & W & \tilde{x}_{FN} \\
\tilde{y}_{11} & \tilde{y}_{12} & \tilde{y}_{13} & \ldots & \tilde{y}_{1N} \\
\vdots & & & & \\
\tilde{y}_{F1} & \tilde{y}_{F2} & \tilde{y}_{F3} & \ldots & \tilde{y}_{FN}
\end{pmatrix}
\]

\[
\begin{pmatrix}
i_1^T \\
\vdots \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
i_1^T \\
\vdots \\
\end{pmatrix}
\begin{pmatrix}
\vdots \\
\end{pmatrix}
\begin{pmatrix}
\vdots \\
\end{pmatrix}
\]

\[
\begin{pmatrix}
\tilde{x}_{it} = x_i - \tilde{x}_{it} = i_t^T P_i \\
\tilde{y}_{it} = y_i - \tilde{y}_{it} = j_t^T P_i
\end{pmatrix}
\]

\[
\begin{pmatrix}
P_1 & S & P_N
\end{pmatrix}
\]

\[
\begin{pmatrix}
2FxN \\
2Fx3 \\
3xN
\end{pmatrix}
\]
Factorization Approach

\[ W = M S \]

- Centered measurement matrix
- “Motion” (camera rotation)
- Structure (3D scene points)
Factorization Approach

\[ W = M \times S \]

**Rank Theorem:**

The 2FxN centered observation matrix has at most rank 3.

**Proof:**

Trivial, using the properties:

- rank of mxn matrix is at most min(m,n)
- rank of A*B is at most min(rank(A),rank(B))
Factorization Approach

Form SVD of measurement matrix $W$

$$ W = U \cdot D \cdot V^T $$

Diagonal matrix with eigenvalues sorted in decreasing order:

$$ d_{11} \geq d_{22} \geq d_{33} \geq ... $$
Factorization Approach

Form SVD of measurement matrix $W$

$$
W = \begin{pmatrix}
2F \times N \\
2F \times 2F \\
2F \times N \\
N \times N \\
\end{pmatrix}
$$

Another useful rank property:

Rank of a matrix is equal to the number of nonzero eigenvalues.

$d_{11}, d_{22}, d_{33}$ are only nonzero eigenvalues (the rest are 0).
Factorization Approach

$2F \times N = 2F \times 2F \times 2F \times N$

Eigenvalues in decreasing order
Factorization Approach

2FxN  =  2Fx2F  *  2FxN  *  NxN

Rank theorem says:

These 3 are nonzero

These should be zero

In practice, due to noise, there may be more than 3 nonzero eigenvalues, but rank theorem tells us to ignore all but the largest three.
Factorization Approach

\[ W = U' \ D' \ V'^T \]
Factorization Approach

\[ W = U' \ D' \ V'T \]

\[ W = U' \ D'\frac{1}{2} \ D'\frac{1}{2} \ V'T \]

\[ 2FxN \quad 2Fx3 \quad 3xN \]

\[ W = MS \]

Camera motion

Scene structure
Annoying Details

\[ W = (U' \ D'^{1/2}) (D'^{1/2} \ V'^T) \]

2FxN \hspace{1cm} 2Fx3 \hspace{1cm} 3xN

\[ W = M \ S \]

Problems:

1) This is not a unique decomposition.

   eg: \((M \ Q) (Q^{-1} \ S) = M \ Q \ Q^{-1} \ S = M \ S\)

2) \(i^T, j^T\) pairs (rows of \(M\)) are not necessarily orthogonal
Solving the Annoying Details

Solution to both problems:

Solve for $Q$ such that appropriate rows of $M$ satisfy

$$
\begin{align*}
\hat{i}_i^T Q Q^T \hat{i}_i &= 1 \\
\hat{j}_i^T Q Q^T \hat{j}_i &= 1
\end{align*}
$$

unit vectors

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

orthogonal

3N equations in 9 unknowns

But these are nonlinear equations

linearize and iterate

(see Exercise 8.8 in book for Newton’s method)

(alternative approach is to use Cholesky decomposition – outside our scope)
Factorization Summary

Assumptions
- orthographic camera
- N non-coplanar points tracking in F>=3 frames

Form the centered measurement matrix $W=[\tilde{X} ; \tilde{Y}]$
- where $\tilde{x}_{ij} = x_{ij} - mx_j$
- where $\tilde{y}_{ij} = y_{ij} - my_j$
- $mx_j$ and $my_j$ are mean of points in frame i
- $j$ ranges over set of points

Rank theorem: The centered measurement matrix has a rank of at most 3
Factorization Algorithm

1) Form the centered measurement matrix $W$ from $N$ points tracked over $F$ frames.
2) Compute SVD of $W = U D V^T$
   - $U$ is $2F \times 2F$
   - $D$ is $2F \times N$
   - $V^T$ is $N \times N$
3) Take largest 3 eigenvalues, and form
   - $D' = 3x3$ diagonal matrix of largest eigenvalues
   - $U' = 2Fx3$ matrix of corresponding column vectors from $U$
   - $V'^T = 3xN$ matrix of corresponding row vectors from $V^T$
4) Define
   $$M = U' D'^{1/2} \quad \text{and} \quad S = D'^{1/2} V'^T$$
5) Solve for $Q$ that makes appropriate rows of $M$ orthogonal
6) Final solution is
   $$M^* = M \quad Q \quad \text{and} \quad S^* = Q^{-1} S$$
Sample Results

QuickTime™ and a Cinepak decompressor are needed to see this picture.