3D Computer Vision

Dmitry Chetverikov, Levente Hajder

Eötvös Loránd University, Faculty of Informatics

Multi-view reconstruction

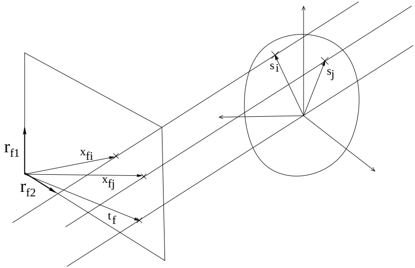
Outline

Possibilities for multi-view reconstruction

1. Concatenation of stereo reconstructions
   - Complicated
   - Reconstruction error cummulated
2. *N*-view solutons
   - Task is non-linear
   - Difficult solutions, implementation challenging
3. Reconstrution by simplified camera models
   - Task is linear if
     - orthogonal or
     - weak-perspective projections applied.

Outline

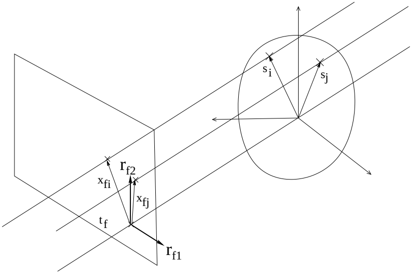
Orthogonal projection



Projection of points

$$\begin{bmatrix} u_{fp} \\ v_{fp} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{\mathbf{f1}}^{\mathbf{T}} \\ \mathbf{r}_{\mathbf{f2}}^{\mathbf{T}} \end{bmatrix} \mathbf{s}_{\mathbf{p}} - \mathbf{t}_{\mathbf{f}} \qquad (1)$$

Orthogonal projection



Projection: origin is the center of gravity.

$$\begin{bmatrix} u_{fp} \\ v_{fp} \end{bmatrix} = \begin{bmatrix} \mathbf{r_{f1}^T} \\ \mathbf{r_{f2}^T} \end{bmatrix} \mathbf{s_p} \qquad (2)$$

Outline

Tomasi-Kanade factorization

- Tracked (matched across multi-frames) coordinates are stacked in measurement matrix **W**.
- It can be factorized into two matrices:

$$\mathbf{W} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1P} \\ v_{11} & v_{12} & \cdots & v_{1P} \\ u_{21} & u_{22} & \cdots & u_{2P} \\ v_{21} & v_{22} & \cdots & v_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ u_{F1} & u_{F2} & \cdots & u_{FP} \\ v_{F1} & v_{F2} & \cdots & v_{FP} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_{11}^T \\ \mathbf{r}_{12}^T \\ \mathbf{r}_{21}^T \\ \mathbf{r}_{22}^T \\ \vdots \\ \mathbf{r}_{F1}^T \\ \mathbf{r}_{F2}^T \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \ldots & \mathbf{s}_P \end{bmatrix}$$

$$\mathbf{W} = \mathbf{MS}$$

Tomasi-Kanade factorization

- As $\mathbf{W} = \mathbf{MS}$, the rank of $\mathbf{W}$ cannot exceed 3 (noiseless-case).
  - Size of $\mathbf{M}$ is $2F \times 3$
  - Size of $\mathbf{S}$ is $3 \times P$
  - Lemma: After factorization, the rank cannot inrease
- Rank reduction of $\mathbf{W}$ by Singular Value Decomposition (SVD)
  - Largest 3 singular values/vectors are kept, other ones are set to zero.
  - $\mathbf{W} = \mathbf{USV^T} \rightarrow \mathbf{W} = \mathbf{U'S'V'^T}$

$$\mathbf{S} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots \\ 0 & \sigma_2 & 0 & 0 & \dots \\ 0 & 0 & \sigma_3 & 0 & \dots \\ 0 & 0 & 0 & \sigma_4 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \rightarrow \mathbf{S'} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}$$

Ambiguity of factorization

- Infinite number of solutions exist:

$$\mathbf{W} = \mathbf{M}\mathbf{S} = \left(\mathbf{M}\mathbf{Q}^{-1}\right)(\mathbf{Q}\mathbf{S}),$$

- where $\mathbf{Q}$ is a 3×3 (affine) matrix.
- $\mathbf{M}_{\mathbf{aff}} = \mathbf{M}\mathbf{Q}^{-1}$: affine motion.
- $\mathbf{S}_{\mathbf{aff}} = \mathbf{Q}\mathbf{S}$ affine structure.
- Constraint to resolve ambiguity: motion vectors $\mathbf{r_i}$ are orthnormal.
  - Camera motion vectors is of length 1.0:

$$\mathbf{r}_{\mathbf{i1}}^{\mathbf{T}}\mathbf{r}_{\mathbf{i1}} = 1$$
$$\mathbf{r}_{\mathbf{i2}}^{\mathbf{T}}\mathbf{r}_{\mathbf{i2}} = 1$$

  - They are perpendicular to each other:

$$\mathbf{r}_{\mathbf{i1}}^{\mathbf{T}}\mathbf{r}_{\mathbf{i2}} = 0$$

Ambiguity removal

- Affine $\rightarrow$ real camera:

$$\mathbf{M_{aff}} = \begin{array}{c} \mathbf{M_{aff}Q} \\ \begin{bmatrix} \mathbf{m}_{11}^T\mathbf{Q} \\ \mathbf{m}_{12}^T\mathbf{Q} \\ \vdots \\ \mathbf{m}_{F1}^T\mathbf{Q} \\ \mathbf{m}_{F2}^T\mathbf{Q} \end{bmatrix} \end{array} \begin{array}{c} = \mathbf{M} \\ = \begin{bmatrix} \mathbf{r}_{11}^T \\ \mathbf{r}_{12}^T \\ \vdots \\ \mathbf{r}_{F1}^T \\ \mathbf{r}_{F2}^T \end{bmatrix} \end{array}$$

- Constraints for camera vectors:

$$\begin{aligned} \mathbf{r}_{i1}^T\mathbf{r}_{i1} = 1 \quad &\rightarrow \quad \mathbf{m}_{i1}^T\mathbf{Q}\mathbf{Q}^T\mathbf{m}_{i1} = 1 \\ \mathbf{r}_{i2}^T\mathbf{r}_{i2} = 1 \quad &\rightarrow \quad \mathbf{m}_{i2}^T\mathbf{Q}\mathbf{Q}^T\mathbf{m}_{i2} = 1 \\ \mathbf{r}_{i1}^T\mathbf{r}_{i2} = 0 \quad &\rightarrow \quad \mathbf{m}_{i1}^T\mathbf{Q}\mathbf{Q}^T\mathbf{m}_{i2} = 0 \end{aligned}$$

Estimation of matrix **Q**

- Let us introduce the following notation:

$$\mathbf{L} = \mathbf{Q}\mathbf{Q}^T = \begin{bmatrix} l_1 & l_2 & l_3 \\ l_2 & l_4 & l_5 \\ l_3 & l_5 & l_6 \end{bmatrix}$$

- Important fact: matrix $\mathbf{QQ^T}$ is symmetric
- Constraints can be written in linear form: $\mathbf{A_i l} = \mathbf{b_i}$

$$\mathbf{A_i} = \begin{bmatrix} {m_{i1,x}}^2 & 2m_{i1,x}m_{i1,y} & 2m_{i1,x}m_{i1,z} & {m_{i1,y}}^2 & 2m_{i1,y}m_{i1,z} & {m_{i1,z}}^2 \\ {m_{i2,x}}^2 & 2m_{i2,x}m_{i2,y} & 2m_{i2,x}m_{i2,z} & {m_{i2,y}}^2 & 2m_{i2,y}m_{i2,z} & {m_{i2,z}}^2 \\ m_{i1,x}m_{i2,x} & e_1 & e_2 & m_{i1,y}m_{i2,y} & m_{i1,y}m_{i2,z} + m_{i2,y}m_{i1,z} & m_{i1,z}m_{i2,z} \end{bmatrix}$$

$$\mathbf{l} = [l_1, l_2, l_3, l_4, l_5, l_6]^T \quad \mathbf{b_i} = [1, 1, 0]^T$$

- where $m_{jk,x}$, $m_{jk,y}$ and $m_{jk,z}$ are the coordinates of vector $\mathbf{m_{jk}}$,
- and $e_1 = m_{i1,x}m_{i2,y} + m_{i2,x}m_{i1,y}$, $e_2 = m_{i1,x}m_{i2,z} + m_{i2,x}m_{i2,z}$.

Computation of matrix $\mathbf{Q}$

- Constraints can be written in linear form: $\mathbf{A}\mathbf{l} = \mathbf{b}$,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_1^T} & \mathbf{A_2^T} & \ldots & \mathbf{A_F^T} \end{bmatrix}^T$$
$$\mathbf{b} = [1, 1, 0, 1, 1, 0, \ldots, 1, 1, 0]^T$$

- Solution by over-determined inhomogeneous linear system of equations
- Matrix $\mathbf{Q}$ can be retrieved from $\mathbf{L}$ by SVD:

$$\mathbf{L} \overset{(SVD)}{=} \mathbf{U}\mathbf{S}\mathbf{U}^\mathbf{T}$$
$$\mathbf{Q} = \mathbf{U}\sqrt{\mathbf{S}}$$

Weak-perspective projection

- Modified constraints:
  - motion vectors are perpendicular to each other:

$$\mathbf{r}_{i1}^T \mathbf{r}_{i2} = 0$$

  - Length of vectors are not unit, but equal:

$$\mathbf{r}_{i1}^T \mathbf{r}_{i1} = \mathbf{r}_{i2}^T \mathbf{r}_{i2}$$

- Equations for affine ambuguity, represented by matrix **Q** as follows:

$$\begin{aligned} \mathbf{m}_{i1}^T \mathbf{Q}\mathbf{Q}^T \mathbf{m}_{i1} - \mathbf{m}_{i2}^T \mathbf{Q}\mathbf{Q}^T \mathbf{m}_{i2} &= 0 \\ \mathbf{m}_{i1}^T \mathbf{Q}\mathbf{Q}^T \mathbf{m}_{i2} &= 0 \end{aligned}$$

- Linear, homogeneous system of equations obtained.

Summary of Tomasi-Kanade factorization

1. Tracked points are stacked in measurement matrix $\mathbf{W}$.
2. Origin is moved to the center of gravity, translated coordinates are stacked in matrix $\tilde{\mathbf{W}}$.
3. SVD computed for $\tilde{\mathbf{W}}$: $\tilde{\mathbf{W}} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.
4. Singular elements are replaced by zero, except the first three values in $\mathbf{S}$: $\mathbf{S} \rightarrow \mathbf{S}'$.
5. Affine factorization: $\mathbf{M}_{\mathbf{aff}} = \mathbf{U}\sqrt{\mathbf{S}'}$ and $\mathbf{S}_{\mathbf{aff}} = \sqrt{\mathbf{S}'}\mathbf{V}^{\mathbf{T}}$.
6. Calculation of matrix $\mathbf{Q}$ by metric constraints.
7. Metric factorization: $\mathbf{M} = \mathbf{M}_{\mathbf{aff}}\mathbf{Q}$ and $\mathbf{S} = \mathbf{Q}^{-1}\mathbf{S}_{\mathbf{aff}}$.

Outline

Multi-view perspective reconstruction

- Three-view geometry
  - Extension of epipolar geometry
  - Relationships can be written for 3D points and lines
  - Trifocal tensor introduced as the extension of the fundamental matrix
  - It has small practical impact
- Perspective Tomasi-Kanade factorization
  - Problem is a perspective auto-calibration
  - Difficulty: projective depths are different for all point/frames
  - Only iterative solutions exist
  - Very complicated
- Viable solution: Concatenation of stereo reconstructions

Outline

Reconstruction by concatenating multiple stereo vision

- For calibrated cameras, stereo reconstruction possible
- Camera calibration:
  - Intrinsic parameters: by chessboard-based Zhang calibration
  - Extrinsic parameters: by decomposition of essential matrix
- Spatial reconstruction: triangulation
- Results:
  - For each stereo image pair, 3D point clouds obtained.
  - Transformation(translation/rotation) between images computed as well

Concatenating point clouds

- Two point clouds given
  - For stereo reconstruction, coordinate system is usually fixed to the first camera.
- Point clouds have N common points are stecked in vector sets: $\{\mathbf{p_i}\}$ and $\{\mathbf{q_i}\}$, $(i = 1 \dots N)$.
- Similarity transformation between images has o be estimated.

$$\mathbf{q_i} = s\mathbf{R}\mathbf{p_i} + \mathbf{t}$$

  - s: scale
  - $\mathbf{R}$: rotation
  - $\mathbf{t}$: translation

Concatenating stereo reconstructions

- Task: optimal registration to estimate similarity transformation

$$\sum_{i=1}^{N} ||\mathbf{q_i} - s\mathbf{R}\mathbf{p_i} - \mathbf{t}||^2$$

- Proof given in separate document
  - Optimal translation $\mathbf{t}$: difference of centers of gravity
  - Optimal rotation:

$$\mathbf{H} = \sum_{i=1}^{N} \mathbf{q'_i} \mathbf{p'_i}^T$$

$$\mathbf{R} = \mathbf{V}\mathbf{U}^\mathbf{T} \leftarrow \mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

  - Optimal scale:

$$s = \frac{\sum_{i=1}^{N} \mathbf{q'}_\mathbf{i}^\mathbf{T} R \mathbf{p'_i}}{\sum_{i=1}^{N} \mathbf{p'}_\mathbf{i}^\mathbf{T} \mathbf{p'_i}}$$

Outline

Minimization by a numerical algorithm

- The projected coordinates of j-th point in i-th frame depend on
  - parameters of ith camera and
  - spatial coordinated of j-th point.
- Numerical optimization by Levenberg-Marquardt algorithm.
  - Jacobian matrix of the problem has to be determined.
  - Jacobian is very sparse.
- Thus, a sparse Levenberg-Marquardt algorithm should be applied.
  - It is called bundle Adjustment (BA) in the literature.

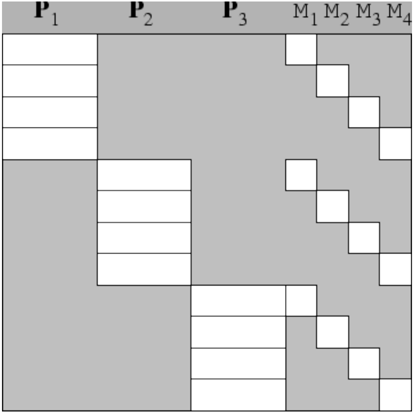
Levenberg-Marquardt for 3D Reconstruction

- LM-rule for parameter tuning:

$$\Delta \mathbf{p} = \left(\mathbf{J}^T \mathbf{J} + \lambda \mathbf{I} \right)^{-1} \mathbf{J}^T \epsilon_p$$

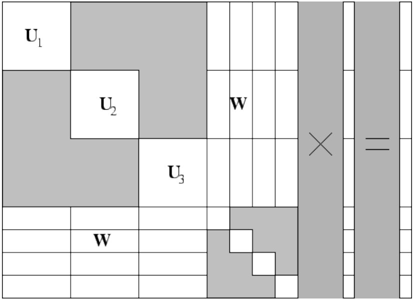
- Parameters to be tuned:
  - camera parameters
  - spatial coordinates
- E.g. for 20 perspective cameras and 1000 3D points: $20 \cdot 11 + 3 \cdot 1000 = 3220$ parameters have to be estimated
  - Dimension of $\mathbf{J}^T \mathbf{J}$ is 3220×3220.
  - Matrix invertion requires very high time demand.
  - Numerical stability of invertion is questionable

Jacobian matrix



Jacobian matrix

Jacobian matrix



Normal equation

Bundle adjustment: normal equation

- Normal equation can be written by block of matrices:

$$\begin{bmatrix} \mathbf{U} & \mathbf{X} \\ \mathbf{X}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{m} \\ \Delta \mathbf{s} \end{bmatrix} = \begin{bmatrix} \epsilon_{\mathbf{m}} \\ \epsilon_{\mathbf{s}} \end{bmatrix}$$

- If normal equation is multiplied by $\begin{bmatrix} \mathbf{I} & -\mathbf{X}\mathbf{V}^{-1} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, from the left, normal equation is modified as follows:

$$\begin{bmatrix} \mathbf{U} - \mathbf{X}\mathbf{V}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}} & \mathbf{0} \\ \mathbf{X}^{\mathbf{T}} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{m} \\ \Delta \mathbf{s} \end{bmatrix} = \begin{bmatrix} \epsilon_{\mathbf{m}} - \mathbf{X}\mathbf{V}^{-1}\epsilon_{\mathbf{s}} \\ \epsilon_{\mathbf{s}} \end{bmatrix}$$

Bundle adjustment: solution for normal equation

- Solution:

$$\Delta \mathbf{m} = \left(\mathbf{U} - \mathbf{X}\mathbf{V}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\right)^{-1}\left(\epsilon_{\mathbf{m}} - \mathbf{X}\mathbf{V}^{-1}\epsilon_{\mathbf{s}}\right)$$
$$\Delta \mathbf{s} = \mathbf{V}^{-1}\left(\epsilon_{\mathbf{s}} - \mathbf{X}^{\mathbf{T}}\Delta \mathbf{m}\right)$$

- Inversion required:
  - $\mathbf{V}$:
    - It contains small block matrices, they are inverted separately:
  - $\left(\mathbf{U} - \mathbf{X}\mathbf{V}^{-1}\mathbf{X}^{\mathbf{T}}\right)^{-1}$
    - Its size is relatively small.
    - Is is also a special matrix, sub-blocks can be formed.

Outline

Reconstruction with missing data

- Missing data is not a problem for stereo vision
  - If a feature point visible ony in one image, it cannot be reconstructed.
- Bundle adjustment can cope with missing data
  - Matrices $\mathbf{U_i}$ and $\mathbf{V_j}$ are calculated from less points.
- Tomasi-Kanade factorization requires modification.

Outline: Calibration of weak-perspective camera

- Problem in algebraic form:

$$\begin{bmatrix} u_1 & \dots & u_P \\ v_1 & \dots & v_P \end{bmatrix} = [\mathbf{M}|\mathbf{b}] \begin{bmatrix} \mathbf{X}_1 & \dots & \mathbf{X}_P \\ 1 & \dots & 1 \end{bmatrix} \tag{3}$$

  - where $\mathbf{M} = q\hat{\mathbf{R}}$ and
  - $\hat{\mathbf{R}}$ consists of the first two rows of matrix $\mathbf{R}$
- Camera parameters are unknown, task is a minimization:

$$\arg\min_{q,\hat{\mathbf{R}},\mathbf{b}} \sum_i \left\| \begin{bmatrix} u_i \\ v_i \end{bmatrix} - [q\hat{\mathbf{R}}|\mathbf{b}] \begin{bmatrix} \mathbf{X}_i \\ 1 \end{bmatrix} \right\|_2^2 \tag{4}$$

- This is almost a point registration problem
  - On the left side, only two coordinates are written, not three.

Outline: Calibration of weak-perspective camera

- Trick: left side is extended to have three dimensions
    - Two rows of $\hat{\mathbf{R}}$ can be extended by the third coordinate: if $\mathbf{r}_1^T$ and $\mathbf{r}_2^T$ denote the two rows, the third one can be obtained by cross product:

$$\mathbf{r}_3^T = \mathbf{r}_1^T \times \mathbf{r}_2^T \tag{5}$$

    - Third coordinate of vector **b** is selected to be zero.
    - Third coordinate of left side: $w_i = q\mathbf{r}_3^T\mathbf{X}_i$
- Registration and completion repeated one after the other, until convergence.
- It can be proved that global optimum is reached.

Tomasi-Kanade factorization

- Weak-perspective factorization is written as

$$\mathbf{W} = \mathbf{M}\mathbf{S} \tag{6}$$

→ if the center of gravity is the origin.

- Factorization for arbitrary origin:

$$\mathbf{W} = \left[\begin{array}{c|c} \mathbf{M}_1 & \mathbf{b}_1 \\ \vdots & \vdots \\ \mathbf{M}_F & \mathbf{b}_F \end{array}\right] \left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_P \\ 1 & . & 1 \end{array}\right] = [\mathbf{M}|\mathbf{b}] \left[\begin{array}{ccc} \mathbf{X}_1 & \cdots & \mathbf{X}_P \\ 1 & . & 1 \end{array}\right] \tag{7}$$

- 4-rank problem. It can be solved by an alternation:
  - Estimation of camera parameters: M-step
  - Estimation of 3D coordinates : S-step
  - Additional step (completion): extend 2D projected coordinates into 3D
  - All steps can be computed optimally.

Tomasi-Kanade factorization: S-step

- For a spatial point, the following equation can be written for j-th frame:

$$\begin{bmatrix} u_{ij} \\ v_{ij} \\ w_{ij} \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{M}_j & \mathbf{b}_j \end{array} \right] \begin{bmatrix} \mathbf{X}_i \\ 1 \end{bmatrix} = \mathbf{M}_j \mathbf{X}_i + \mathbf{b}_j \quad (8)$$

- Problem is linear, inhomogeneous, $\mathbf{X}_i$ can be estimated by camera matrices:

$$\mathbf{X}_i = \left(\mathbf{M}^T \mathbf{M} \right)^{-1} \mathbf{M}^T \left(\mathbf{W}_i - \mathbf{b} \right) \quad (9)$$

  - where $\mathbf{W}_i$ is the i-t column of measurement matrix $\mathbf{W}$.
- Missing data: if a point is not visible in a frame, the corresponding camera matrix is discarded.

Tomasi-Kanade factorization: M-step

- Estimation of motion matrix is a point registration problem:

$$\begin{bmatrix} u_{ij} \\ v_{ij} \\ w_{ij} \end{bmatrix} = \mathbf{M}_j \mathbf{X}_i + \mathbf{b}_j = q_j \mathbf{R}_j \mathbf{X}_i + \mathbf{b}_j \tag{10}$$

  - Offset vector is denoted by $\mathbf{b}_j$
  - Rotation: $\mathbf{R}_j$
  - Scale: q_j
- Missing data: if a point is not visible in a frame, the corresponding 3D coordinate vector is discarded.

Tomasi-Kanade factorization: completion step

- Third coordinates in measurement matrix have to be recomputed
    - for all frames,
    - for all points,
    - after all steps.

Tomasi-Kanade facotrization with missing data

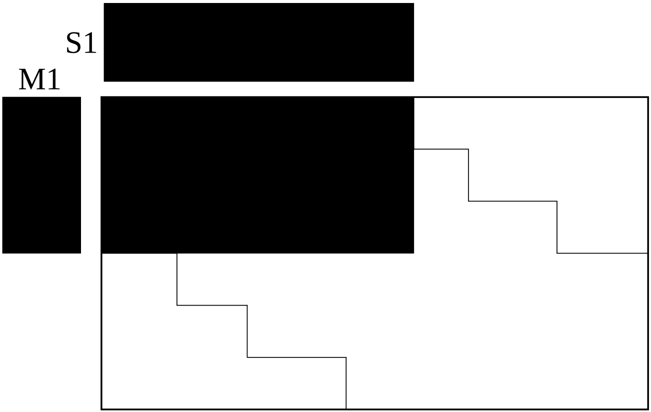
1. Initial motion and 3D parameters are obtained by
   - merging full Tomasi-Kanade factorization, and
   - the third coordinates given by completion-steps.
2. Alternation until convergence:
   - M-step: camera matrix estimation as a 3D-3D point-registration
   - Third coordinates recomputed by completion step
   - 3D coordinates obtained by S-step (linear estimation)
   - Third coordinates recomputed by completion step

- All steps decrease the same least-squares cost function $\rightarrow$ convergence guaranteed.
  - Unfortunately, local minima can occur.

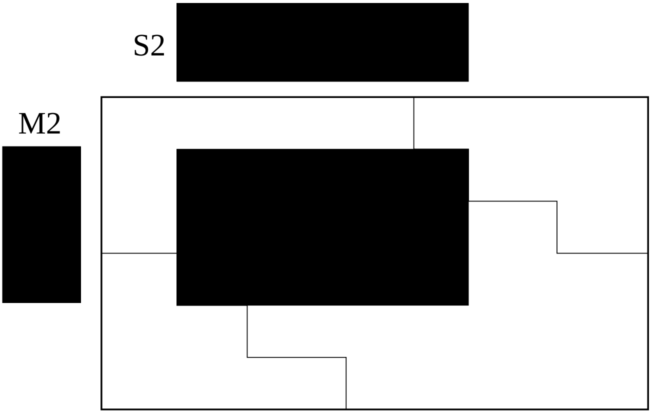
Tomasi-Kanade factorization: initialization

- Mergin full sub–factorization
  - At least 3 images required for a weak-perspective full factorization.
  - Frames 1–3, 2–4, 3–5, etc. have to be processed
- Factorization is carried out for frame-triplets
- Results are merged.

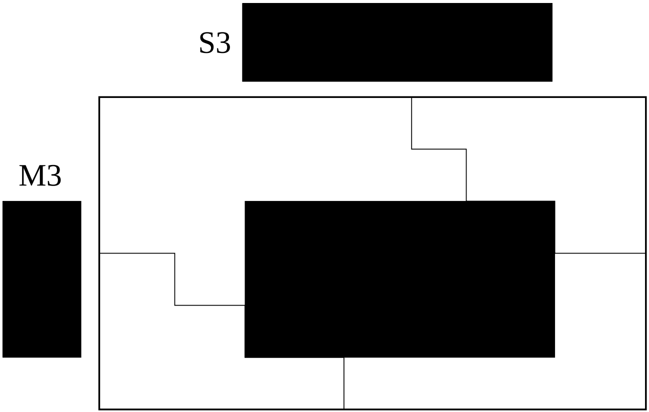
Tomasi-Kanade factorization: partial reconstructions



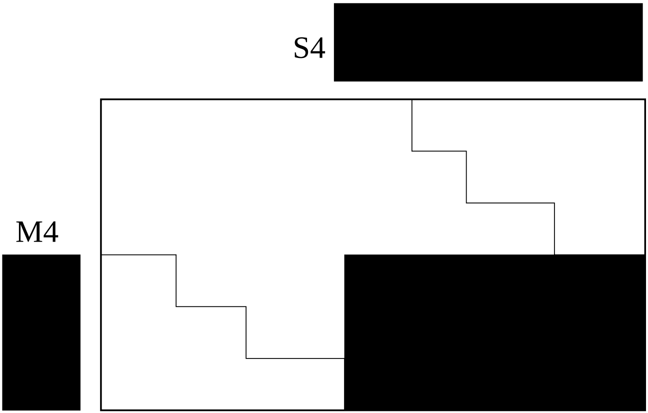
Tomasi-Kanade factorization: partial reconstructions



Tomasi-Kanade factorization: partial reconstructions



Tomasi-Kanade factorization: partial reconstructions



Tomasi-Kanade factorization: initialization

- Concatenation of subfactorizations is not trivial.
- Let us assume that two factorizations are computed
  - Scalar q_j, matrices $\mathbf{R}_j$, $\mathbf{S}_j$, and vector $\mathbf{b}_j$ known for j-th image,
  - as well as q_{j+1}, $\mathbf{R}_{j+1}$, $\mathbf{b}_{j+1}$, and $\mathbf{S}_{j+1}$ for $(j+1)$-th frame.
- Concatenation of 3D point clouds is a point registration problem
  - Obtained parameters after point registration: rotation $\mathbf{R}$, scale q, and offset $\mathbf{t}$.
  - The the registraton for the i-th point:

$$\mathbf{s}_i' = q\mathbf{R}\left(\mathbf{s}_i - \mathbf{o}_1\right) + \mathbf{o}_2 \tag{11}$$

  - Concatenation of motion matrices
    - $\mathbf{M}_{j+1} \leftarrow \frac{1}{q}\mathbf{M}_{j+1}\mathbf{R}^T$
    - $\mathbf{b}_{j+1} \leftarrow \mathbf{b}_{j+1} + q\mathbf{M}_{j+1}\mathbf{R}\mathbf{o}_1 - \mathbf{M}_{j+1}\mathbf{o}_2$