Structure from Motion using Factorization

Sharat Chandran

ViGIL
Indian Institute of Technology Bombay
http://www.cse.iitb.ac.in/∼sharat

March 2014

Note: These slides are best seen with accompanying video

Problem Definition

Can we understand motion using a single camera?

Given 2D point tracks of landmark points from a *single view point*, recover 3D pose and orientation

Assumptions

- 2D tracks of major landmark points are provided
- Scaled-projective/orthographic projection model.

Problem Definition

Can we understand motion using a single camera?

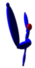
Given 2D point tracks of landmark points from a *single view point*, recover 3D pose and orientation

Assumptions

- 2D tracks of major landmark points are provided
- Scaled-projective/orthographic projection model.

Problem Definition

Can we understand motion using a single camera?



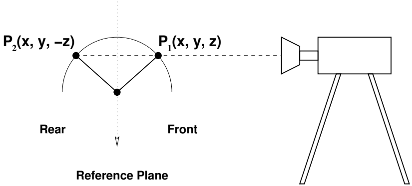
Given 2D point tracks of landmark points from a *single view point*, recover 3D pose and orientation

Assumptions

- 2D tracks of major landmark points are provided
- Scaled-projective/orthographic projection model.

Why is this a hard problem?

The mapping between 2D tracked positions and 3D body pose is many-to-many[1]. This confounds standard regression algorithms.



1

SOATTO, S., AND BROCKETT, R.
1998.
Optimal structure from motion: Local ambiguites and global estimates.
IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

Why this “may not” be such a hard problem after all?

- Human brain perform this *disambiguation* with very little ease.
- Psycho-physical and neuro-physiological imaging experiments have confirmed the fact that we can perceive structure even when we are presented with a video sequence containing only the point tracks of the major joints in the human body[2]

2

JOHANSSON, G.
1976.
Spatio temporal differentiation and integration in visual motion perception.
Psychological Research.

How can we mimic this ability?

Let's observe the trajectories of joint

(a) The top view trajectories of a few dofs plotted (b) One more dof added to the plot (c) All the dofs included

How can we mimic this ability?

Let's observe the trajectories of joint

(a) The top view trajectories of a few dofs plotted (b) One more dof added to the plot (c) All the dofs included

How can we mimic this ability?

Let's observe the trajectories of joint

(a) The top view trajectories of a few dofs plotted

(b) One more dof added to the plot

(c) All the dofs included

How can we mimic this ability?

Let's observe the trajectories of joint

(a) The top view trajectories of a few dofs plotted

(b) One more dof added to the plot

(c) All the dofs included

How can we mimic this ability?

Let's observe the trajectories of joint

(a) The top view trajectories of a few dofs plotted

(b) One more dof added to the plot

(c) All the dofs included

How do we capture these structures?

Matrix Factorization

$$\mathbf{W}_{2F\times P} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ x_{f1} & \cdots & x_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

*If the object in the scene is **rigid** this matrix **W** has a very small rank!!*

How do we capture these structures?

Matrix Factorization

$$\mathbf{W}_{2F \times P} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ x_{f1} & \cdots & x_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

*If the object in the scene is **rigid** this matrix* **W** *has a very small rank!!*

How do we capture these structures?

Matrix Factorization

$$\mathbf{W}_{2F \times P} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ x_{f1} & \cdots & x_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

*If the object in the scene is **rigid** this matrix* **W** *has a very small rank!!*

How do we capture these structures?

Matrix Factorization

$$\mathbf{W}_{2F \times P} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ x_{f1} & \cdots & x_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

*If the object in the scene is **rigid** this matrix **W** has a very small rank!!*

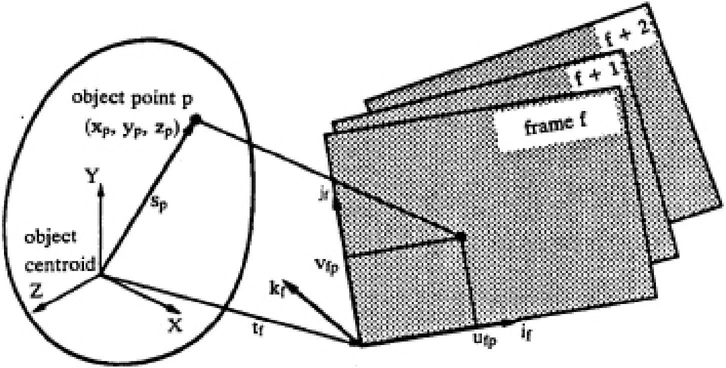
How do we capture these structures?

Matrix Factorization

$$\mathbf{W_{2F\times P}} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ x_{f1} & \cdots & x_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

*If the object in the scene is **rigid** this matrix **W** has a very small rank!!*

Rigid Body Geometry and Motion



- Object centroid based World Co-ordinate System (WCS)

Rank Theorem

Define $\tilde{x}_{ij} = x_{ij} - \bar{x}_i$ and $\tilde{y}_{ij} = y_{ij} - \bar{y}_i$ where the bar notation refers to the centroid of the points in the ith frame. We have the *measurement matrix*

$$\bar{\mathbf{W}}_{\mathbf{2F \times P}} = \begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ \tilde{x}_{f1} & \cdots & \tilde{x}_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

The matrix $\bar{\mathbf{W}}$ has rank 3

Rank Theorem

Define $\tilde{x}_{ij} = x_{ij} - \bar{x}_i$ and $\tilde{y}_{ij} = y_{ij} - \bar{y}_i$ where the bar notation refers to the centroid of the points in the ith frame. We have the *measurement matrix*

$$\bar{\mathbf{W}}_{\mathbf{2F\times P}} = \begin{pmatrix} \tilde{x}_{11} & \cdots & \tilde{x}_{1p} \\ y_{11} & \cdots & y_{1p} \\ \vdots & \vdots & \vdots \\ \tilde{x}_{f1} & \cdots & \tilde{x}_{fp} \\ y_{f1} & \cdots & y_{fp} \end{pmatrix}$$

The matrix $\bar{\mathbf{W}}$ has rank 3

Rank Theorem Proof

$$x_{ij} = \mathbf{i}_i^T(\mathbf{P_j} - \mathbf{T_i}), \quad \mathbf{y_{ij}} = \mathbf{j_i^T}(\mathbf{P_j} - \mathbf{T_i}), \quad \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{j=1}}^{\mathbf{n}} \mathbf{P_j} = \mathbf{0}$$

$$\tilde{x}_{ij} = \mathbf{i}_i^T(\mathbf{P_j} - \mathbf{T_i}) - \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{m=1}}^{\mathbf{n}} \mathbf{i_i^T}(\mathbf{P_m} - \mathbf{T_i})$$

$$\tilde{y}_{ij} = \mathbf{j}_i^T(\mathbf{P_j} - \mathbf{T_i}) - \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{m=1}}^{\mathbf{n}} \mathbf{j_i^T}(\mathbf{P_m} - \mathbf{T_i})$$

$$\tilde{x}_{ij} = \mathbf{i}_i^T\mathbf{P_j} \qquad \tilde{\mathbf{y}}_{\mathbf{ij}} = \mathbf{j_i^T}\mathbf{P_j}$$

$$\bar{\mathbf{W}} = \mathbf{RS}$$

$$\mathbf{R} = \begin{bmatrix} \mathbf{i}_1^T \\ \mathbf{j}_1^T \\ \dots \\ \mathbf{i}_N^T \\ \mathbf{j}_N^T \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} \mathbf{P_1} & \mathbf{P_2} & \dots & \mathbf{P_N} \end{bmatrix}$$

Rigid Body Geometry and Motion

- Without noise $\overline{\mathbf{W}}$ is atmost of rank *three*
- Using SVD, $\overline{\mathbf{W}} = \mathbf{O_1} \Sigma \mathbf{O_2}$ where,
  $\mathbf{O_1}$, $\mathbf{O_2}$ are column orthogonal matrices and Σ is a diagonal matrix with singular values in non-decreasing order
- $\mathbf{O_1} \Sigma \mathbf{O_2} = \mathbf{O_1'} \Sigma' \mathbf{O_2'} + \mathbf{O_1''} \Sigma'' \mathbf{O_2''}$ where,
  $\mathbf{O_1'}$ has *first three* columns of $\mathbf{O_1}$, $\mathbf{O_2'}$ has *first three* rows of $\mathbf{O_2}$ and Σ' is 3×3 matrix with 3 largest non-singular values.
- The second term is completely due to noise and can be eliminated
- $\hat{\mathbf{R}} = \mathbf{O_1'} \left[\Sigma'\right]^{1/2}$ and $\hat{\mathbf{S}} = \left[\Sigma'\right]^{1/2} \mathbf{O_2'}$

Rigid Body Geometry and Motion

- Without noise $\overline{\mathbf{W}}$ is atmost of rank *three*
- Using SVD, $\overline{\mathbf{W}} = \mathbf{O_1}\Sigma\mathbf{O_2}$ where,
  $\mathbf{O_1}$, $\mathbf{O_2}$ are column orthogonal matrices and Σ is a diagonal matrix with singular values in non-decreasing order
- $\mathbf{O_1}\Sigma\mathbf{O_2} = \mathbf{O_1'}\Sigma'\mathbf{O_2'} + \mathbf{O_1''}\Sigma''\mathbf{O_2''}$ where,
  $\mathbf{O_1'}$ has *first three* columns of $\mathbf{O_1}$, $\mathbf{O_2'}$ has *first three* rows of $\mathbf{O_2}$ and Σ' is 3×3 matrix with 3 largest non-singular values.
- The second term is completely due to noise and can be eliminated
- $\hat{\mathbf{R}} = \mathbf{O_1'}\left[\Sigma'\right]^{1/2}$ and $\hat{\mathbf{S}} = \left[\Sigma'\right]^{1/2}\mathbf{O_2'}$

Rigid Body Geometry and Motion

- Without noise $\overline{\mathbf{W}}$ is atmost of rank *three*
- Using SVD, $\overline{\mathbf{W}} = \mathbf{O_1} \Sigma \mathbf{O_2}$ where,
  $\mathbf{O_1}$, $\mathbf{O_2}$ are column orthogonal matrices and Σ is a diagonal matrix with singular values in non-decreasing order
- $\mathbf{O_1} \Sigma \mathbf{O_2} = \mathbf{O_1'} \Sigma' \mathbf{O_2'} + \mathbf{O_1''} \Sigma'' \mathbf{O_2''}$ where,
  $\mathbf{O_1'}$ has *first three* columns of $\mathbf{O_1}$, $\mathbf{O_2'}$ has *first three* rows of $\mathbf{O_2}$ and Σ' is 3×3 matrix with 3 largest non-singular values.
- The second term is completely due to noise and can be eliminated
- $\hat{\mathbf{R}} = \mathbf{O_1'} \left[\Sigma'\right]^{1/2}$ and $\hat{\mathbf{S}} = \left[\Sigma'\right]^{1/2} \mathbf{O_2'}$

Rigid Body Geometry and Motion

- Without noise $\overline{\mathbf{W}}$ is atmost of rank *three*
- Using SVD, $\overline{\mathbf{W}} = \mathbf{O_1} \Sigma \mathbf{O_2}$ where,
  $\mathbf{O_1}$, $\mathbf{O_2}$ are column orthogonal matrices and Σ is a diagonal matrix with singular values in non-decreasing order
- $\mathbf{O_1} \Sigma \mathbf{O_2} = \mathbf{O_1'} \Sigma' \mathbf{O_2'} + \mathbf{O_1''} \Sigma'' \mathbf{O_2''}$ where,
  $\mathbf{O_1'}$ has *first three* columns of $\mathbf{O_1}$, $\mathbf{O_2'}$ has *first three* rows of $\mathbf{O_2}$ and Σ' is 3×3 matrix with 3 largest non-singular values.
- The second term is completely due to noise and can be eliminated
- $\hat{\mathbf{R}} = \mathbf{O_1'} \left[\Sigma'\right]^{1/2}$ and $\hat{\mathbf{S}} = \left[\Sigma'\right]^{1/2} \mathbf{O_2'}$

Rigid Body Geometry and Motion

- Without noise $\overline{\mathbf{W}}$ is atmost of rank *three*
- Using SVD, $\overline{\mathbf{W}} = \mathbf{O_1} \Sigma \mathbf{O_2}$ where,
  $\mathbf{O_1}$, $\mathbf{O_2}$ are column orthogonal matrices and Σ is a diagonal matrix with singular values in non-decreasing order
- $\mathbf{O_1} \Sigma \mathbf{O_2} = \mathbf{O_1'} \Sigma' \mathbf{O_2'} + \mathbf{O_1''} \Sigma'' \mathbf{O_2''}$ where,
  $\mathbf{O_1'}$ has *first three* columns of $\mathbf{O_1}$, $\mathbf{O_2'}$ has *first three* rows of $\mathbf{O_2}$ and Σ' is 3×3 matrix with 3 largest non-singular values.
- The second term is completely due to noise and can be eliminated
- $\hat{\mathbf{R}} = \mathbf{O_1'} \left[\Sigma'\right]^{1/2}$ and $\hat{\mathbf{S}} = \left[\Sigma'\right]^{1/2} \mathbf{O_2'}$

Rigid Body Geometry and Motion

- Solution is not unique any invertible 3×3, $\mathbf{Q}$ matrix can be written as $\mathbf{R} = (\hat{\mathbf{R}}\mathbf{Q})$ and $\mathbf{S} = (\mathbf{Q}^{-1}\hat{\mathbf{S}})$
- $\hat{\mathbf{R}}$ is a linear transformation of $\mathbf{R}$, similarly $\hat{\mathbf{S}}$ is a linear transformation of $\mathbf{S}$.
- Using the following orthonormality constraints we can find $\mathbf{R}$ and $\mathbf{S}$

$$\hat{\mathbf{i}}_f^T \mathbf{Q}\mathbf{Q}^{\mathbf{T}} \hat{\mathbf{i}}_f = 1$$

$$\hat{\mathbf{j}}_f^T \mathbf{Q}\mathbf{Q}^{\mathbf{T}} \hat{\mathbf{j}}_f = 1$$

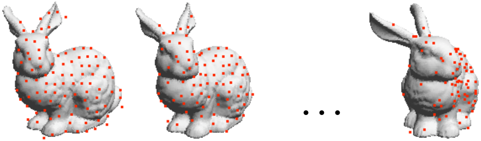
$$\hat{\mathbf{i}}_f^T \mathbf{Q}\mathbf{Q}^{\mathbf{T}} \hat{\mathbf{j}}_f = 0 \tag{1}$$

Rigid Body Geometry and Motion

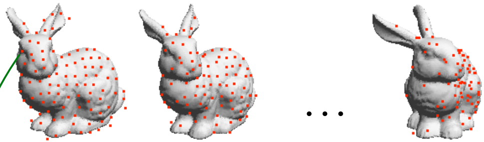
- Solution is not unique any invertible 3×3, $\mathbf{Q}$ matrix can be written as $\mathbf{R} = (\hat{\mathbf{R}}\mathbf{Q})$ and $\mathbf{S} = (\mathbf{Q}^{-1}\hat{\mathbf{S}})$
- $\hat{\mathbf{R}}$ is a linear transformation of $\mathbf{R}$, similarly $\hat{\mathbf{S}}$ is a linear transformation of $\mathbf{S}$.
- Using the following orthonormality constraints we can find $\mathbf{R}$ and $\mathbf{S}$

$$
\begin{aligned}
\hat{\mathbf{i}}_f^T \mathbf{Q}\mathbf{Q}^{\mathbf{T}} \hat{\mathbf{i}}_f &= 1 \\
\hat{\mathbf{j}}_f^T \mathbf{Q}\mathbf{Q}^{\mathbf{T}} \hat{\mathbf{j}}_f &= 1 \\
\hat{\mathbf{i}}_f^T \mathbf{Q}\mathbf{Q}^{\mathbf{T}} \hat{\mathbf{j}}_f &= 0
\end{aligned} \tag{1}
$$

Tomasi Kanade Factorisation (Recap)



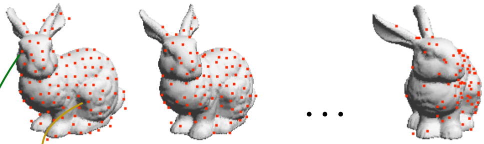
Tomasi Kanade Factorisation (Recap)



$$\begin{bmatrix} 27 & 61 & \cdots & 96 \\ 97 & 53 & \cdots & 122 \\ 28 & 62 & \cdots & 97 \\ 97 & 53 & \cdots & 122 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 94 & ? & \cdots & 131 \\ 109 & ? & \cdots & 135 \end{bmatrix}$$

W

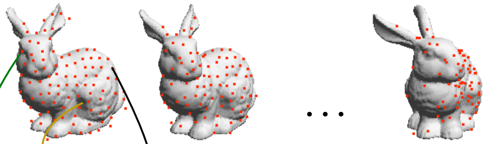
Tomasi Kanade Factorisation (Recap)



$$
\begin{bmatrix}
27 & 61 & \cdots & 96 \\
97 & 53 & \cdots & 122 \\
28 & 62 & \cdots & 97 \\
97 & 53 & \cdots & 122 \\
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
94 & ? & \cdots & 131 \\
109 & ? & \cdots & 135
\end{bmatrix}
$$

W

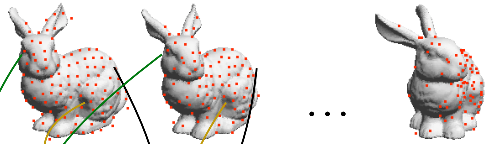
Tomasi Kanade Factorisation (Recap)



$$
\begin{bmatrix}
27 & 61 & \cdots & 96 \\
97 & 53 & \cdots & 122 \\
28 & 62 & \cdots & 97 \\
97 & 53 & \cdots & 122 \\
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
94 & ? & \cdots & 131 \\
109 & ? & \cdots & 135
\end{bmatrix}
$$

W

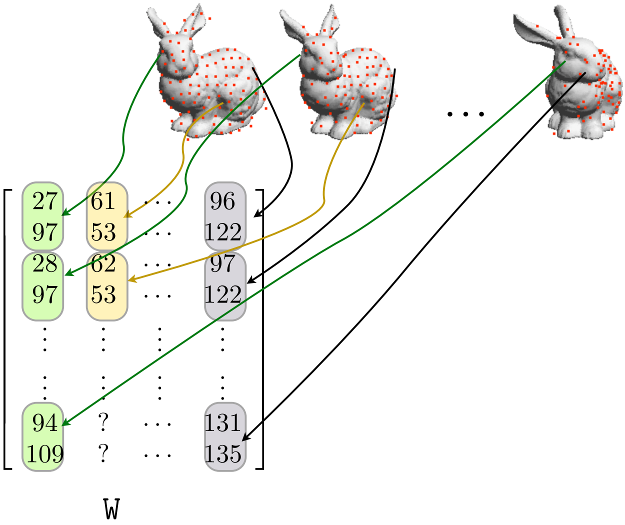
Tomasi Kanade Factorisation (Recap)



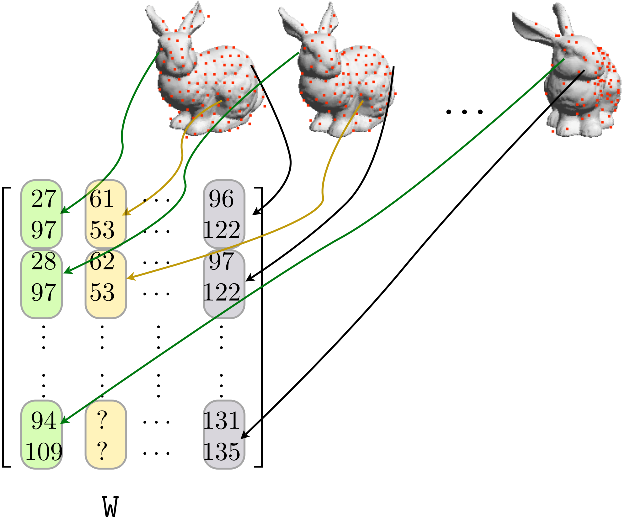
$$
\begin{bmatrix}
27 & 61 & \cdots & 96 \\
97 & 53 & \cdots & 122 \\
28 & 62 & \cdots & 97 \\
97 & 53 & \cdots & 122 \\
\vdots & \vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots & \vdots \\
94 & ? & \cdots & 131 \\
109 & ? & \cdots & 135
\end{bmatrix}
$$

$\mathtt{W}$

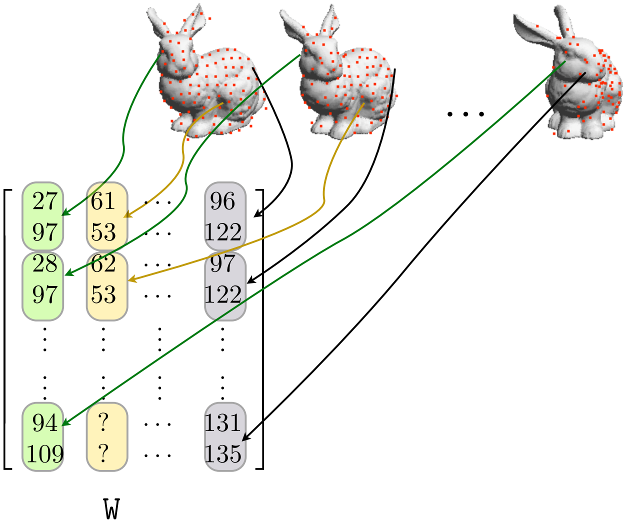
Tomasi Kanade Factorisation (Recap)



Tomasi Kanade Factorisation (Recap)

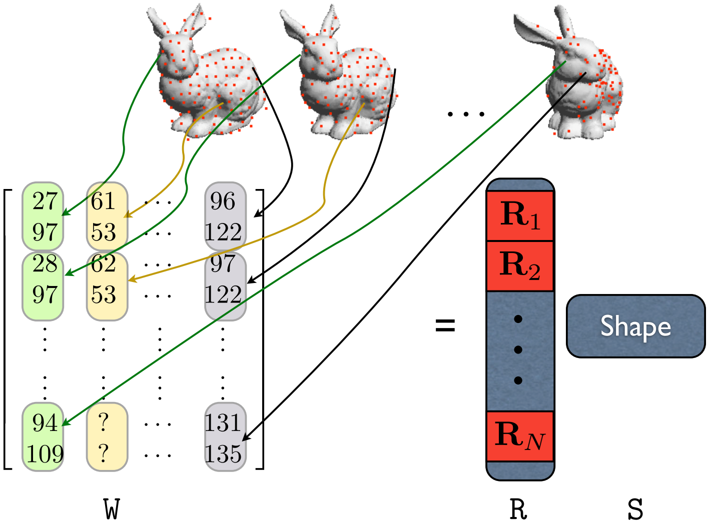


Tomasi Kanade Factorisation (Recap)



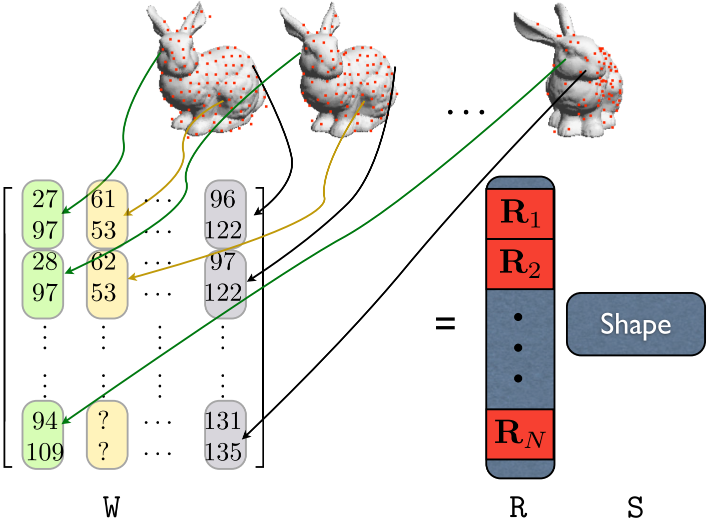
Central Observation: This matrix is rank-limited.
If the object motion is rigid the observation matrix (discounting noise) will have a maximum rank of 4

Tomasi Kanade Factorisation (Recap)



Central Observation: This matrix is rank-limited.
If the object motion is rigid the observation matrix (discounting noise) will have a maximum rank of 4

Tomasi Kanade Factorisation (Recap)



Assumptions

Orthographic Camera Model

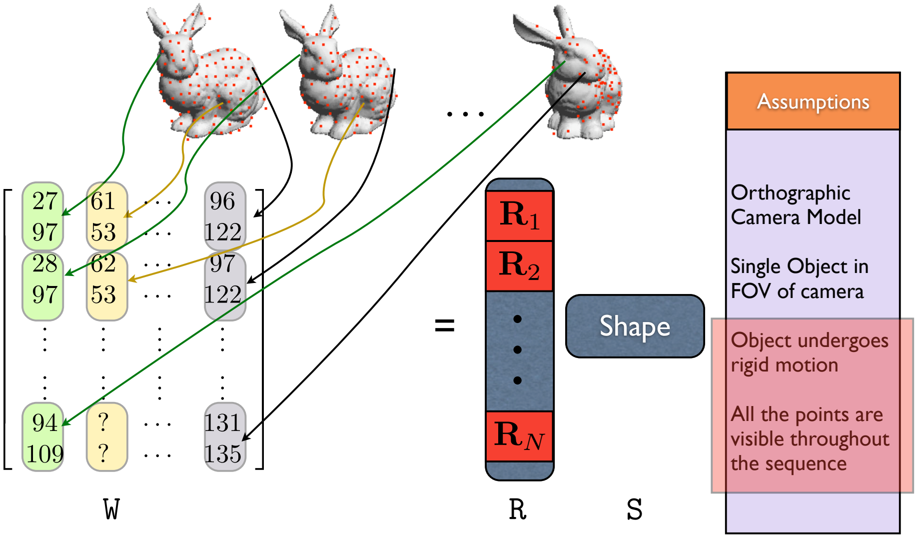
Single Object in FOV of camera

Object undergoes rigid motion

All the points are visible throughout the sequence

Central Observation: This matrix is rank-limited.
If the object motion is rigid the observation matrix (discounting noise) will have a maximum rank of 4

Tomasi Kanade Factorisation (Recap)



Central Observation: This matrix is rank-limited.
If the object motion is rigid the observation matrix (discounting noise) will have a maximum rank of 4

For Further Reading I

G. Golub and A. Loan
Matrix Computations
John Hopkins U. Press, 1996

C. Tomasi and T. Kanade
Shape and motion from image stream: A factorization method
Image of Science: Science of Images, 90:9795–9802,1993

J. Xiao and J. Chai and T. Kanade
A Closed-Form Solution to Non-Rigid Shape and Motion Recovery
ECCV 2004

For Further Reading II

C. Bregler and A. Hertzmann and H. Biermann
Recovering Non-Rigid 3D Shape from Image Streams
CVPR, 2000

M. Brand
Morphable 3D Models from Video
CVPR, 2001

Appu Shaji and Aydin Varol and Pascal Fua and Yashoteja and Ankush Jain and Sharat Chandran
Resolving Occlusion in Multiframe Reconstruction of Deformable Surfaces
NORDIA, *CVPRW*, 2011

M. Kilian, N. Mitra and H. Pottmann. Geometric Modeling in Shape Space. Siggraph, 2008.