Using Shape Spaces for Structure from Motion

Sharat Chandran

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

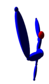
March 2012

Joint Work: Appu Shaji, Yashoteja Prabhu, Pascal Fua, S. Ladha & other ViGIL students

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.

Motivation Factorization Non-Rigid Motion Occlusion

Problem Definition

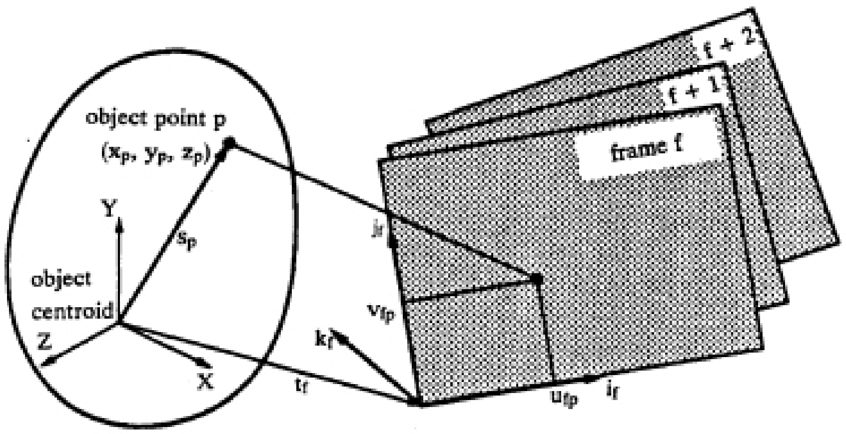
Can we understand motion using a single camera?

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.

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*

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

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*

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

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

Rank Theorem Proof

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

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

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

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

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

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

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 $\Sigma^{'}$ 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 $\hat{\mathbf{S}} = \left[\Sigma^{'}\right]^{1/2}\mathbf{O}_2^{'}$

Motivation Factorization Non-Rigid Motion Occlusion

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 $\Sigma^{'}$ 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 $\hat{\mathbf{S}} = \left[\Sigma'\right]^{1/2}\mathbf{O_2'}$

Motivation Factorization Non-Rigid Motion Occlusion

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 $\mathbf{S}$

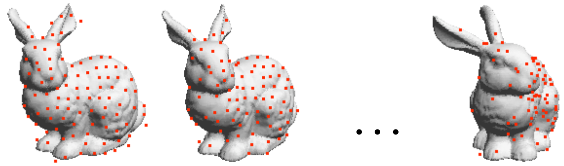
$$\hat{\mathbf{i}}_f^T\mathbf{Q}\mathbf{Q}^\mathbf{T}\hat{\mathbf{i}}_f = 1$$
$$\hat{\mathbf{j}}_f^T\mathbf{Q}\mathbf{Q}^\mathbf{T}\hat{\mathbf{j}}_f = 1$$
$$\hat{\mathbf{i}}_f^T\mathbf{Q}\mathbf{Q}^\mathbf{T}\hat{\mathbf{j}}_f = 0 \qquad (1)$$

Rigid Body Geometry and Motion

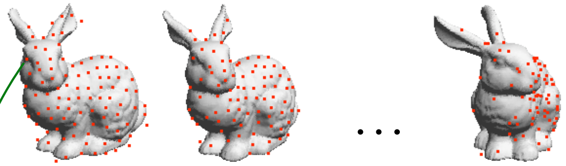
- Solution is not unique any invertible 3×3, $\mathbf{Q}$ matrix can be written as $\mathbf{R} = (\hat{\mathbf{R}}\mathbf{Q})$ and $\mathbf{S} = (\mathbf{Q}^{-1}\hat{\mathbf{S}})$
- $\hat{\mathbf{R}}$ is a linear transformation of $\mathbf{R}$, similarly $\hat{\mathbf{S}}$ is a linear transformation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

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

Tomasi Kanade Factorisation (Recap)



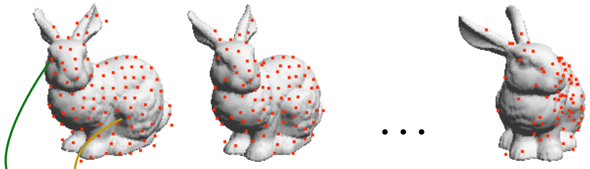
Tomasi Kanade Factorisation (Recap)



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

W

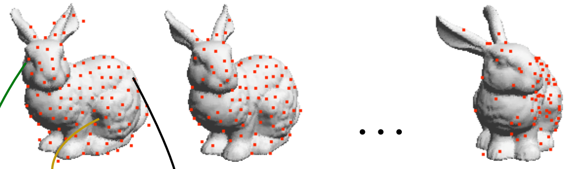
Tomasi Kanade Factorisation (Recap)



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

W

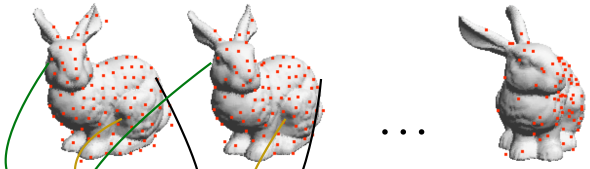
Tomasi Kanade Factorisation (Recap)



27	61	⋯	96
97	53	⋯	122
28	62	⋯	97
97	53	⋯	122
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
94	?	⋯	131
109	?	⋯	135

W

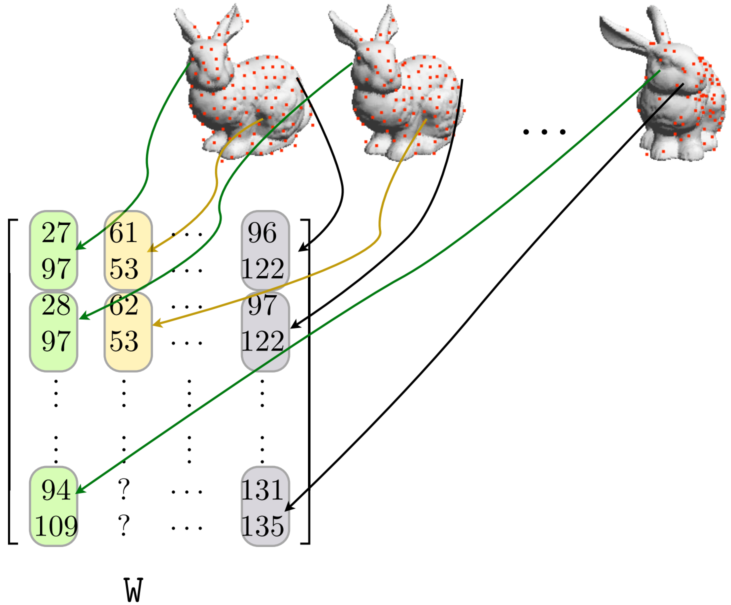
Tomasi Kanade Factorisation (Recap)



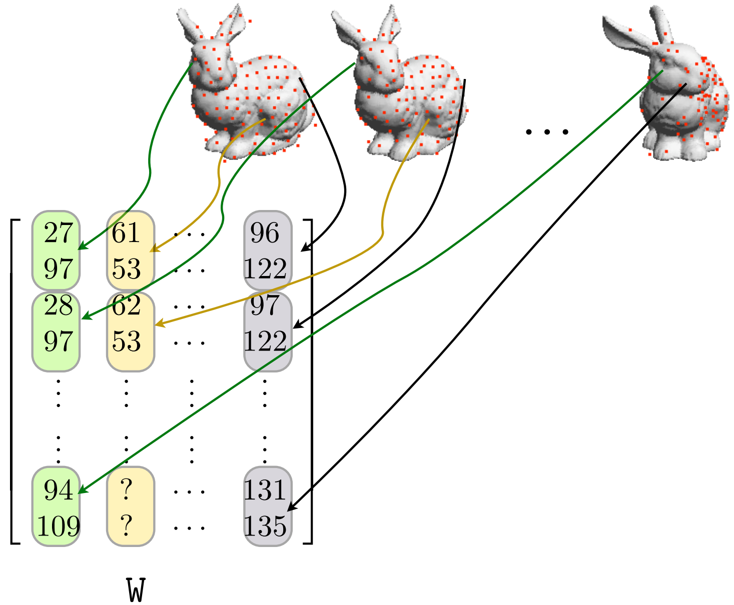
27	61	⋯	96
97	53	⋯	122
28	62	⋯	97
97	53	⋯	122
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
94	?	⋯	131
109	?	⋯	135

W

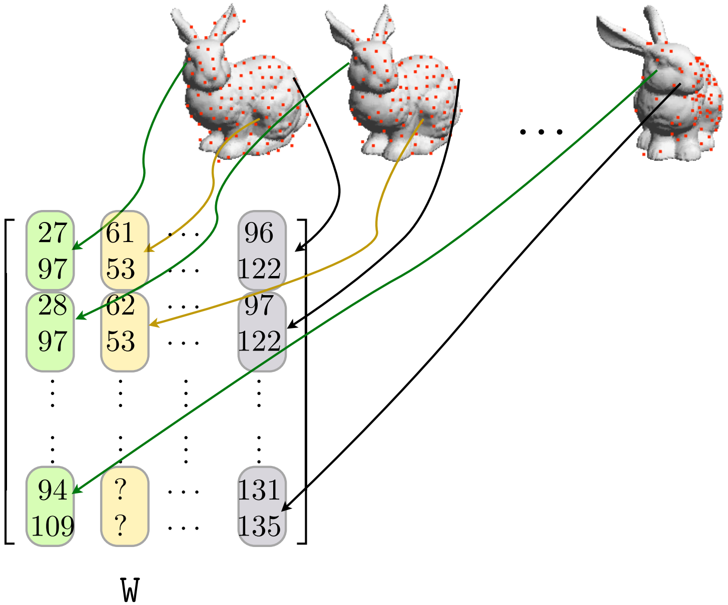
Tomasi Kanade Factorisation (Recap)



Tomasi Kanade Factorisation (Recap)

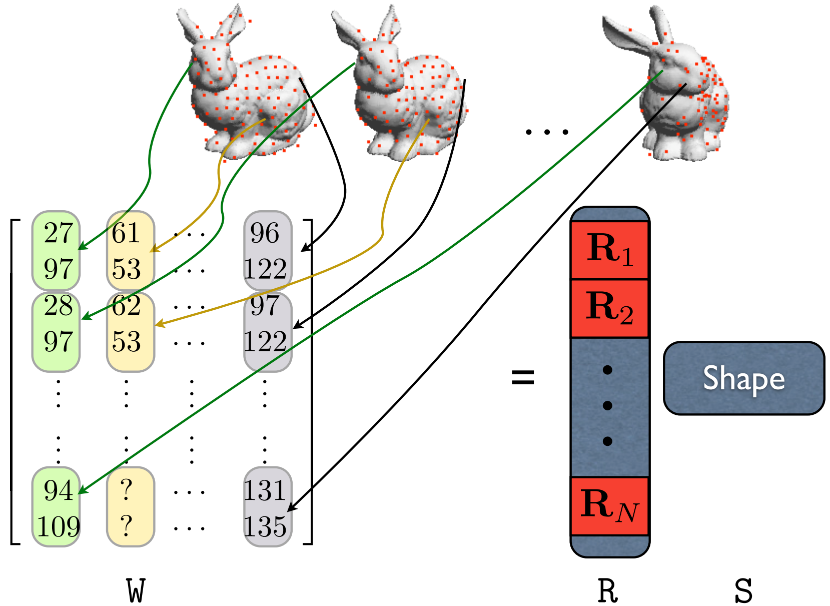


Tomasi Kanade Factorisation (Recap)



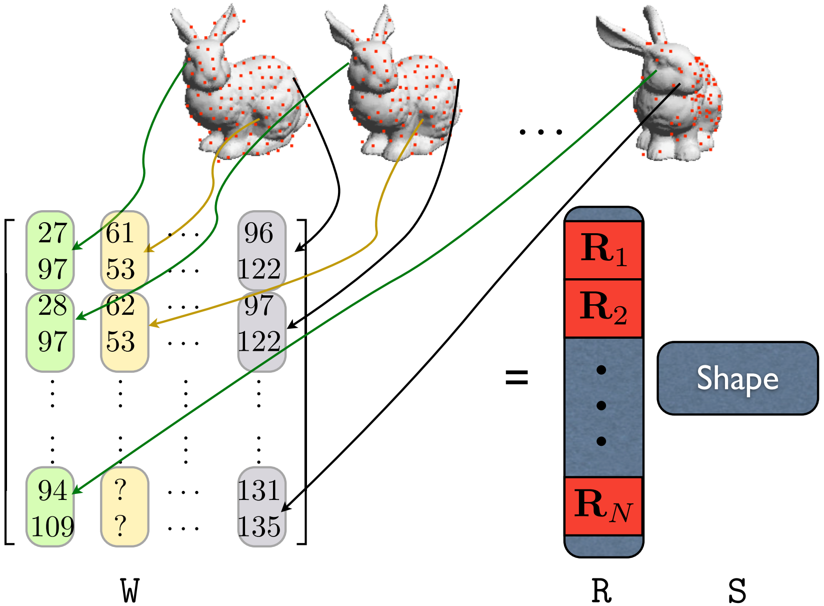
Central Observation: This matrix is rank-limited.
If the object motion is rigid the observation matrix (discounting noise) will have a maximum rank of 4

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)

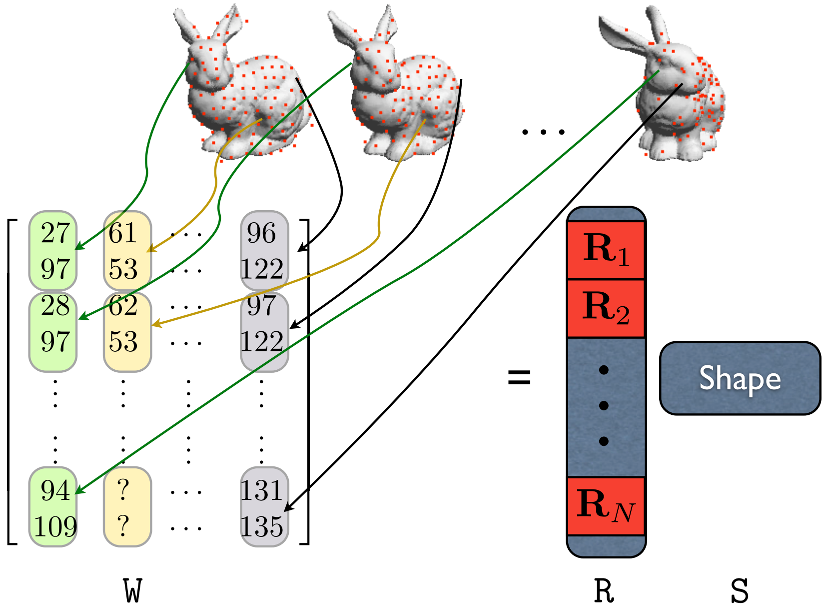


Assumptions

- 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)



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

Non-Rigid Motion

- Many objects are non-rigid
- The parametrisation $\mathbf{S}_{3\times P}$ is no longer valid.
- However, deformable bodies (like human body, face) can be represented using a linear combination of basis shapes

$$\mathbf{S_{morph}} = \sum_{i=1}^{K} c_i \mathbf{S_i} \qquad \mathbf{S_{morph}}, \mathbf{S_i} \in \mathbb{R}^{3\times P}, c_i \in \mathbb{R}$$

where $\mathbf{S_i}$'s are the bases, and c_i are the deformation weights.

Motivation Factorization Non-Rigid Motion Occlusion

Non-Rigid Motion

- Many objects are non-rigid
- The parametrisation $\mathbf{S}_{3\times P}$ is no longer valid.
- However, deformable bodies (like human body, face) can be represented using a linear combination of basis shapes

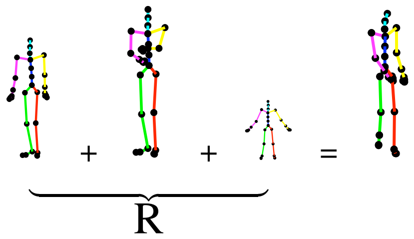
$$\mathbf{S_{morph}} = \sum_{i=1}^{K} c_i \mathbf{S_i} \qquad \mathbf{S_{morph}}, \mathbf{S_i} \in \mathbb{R}^{3\times P}, c_i \in \mathbb{R}$$

where $\mathbf{S_i}$'s are the bases, and c_i are the deformation weights.

Morphable Models

One popular generalisation (used for human faces): linear combination of shapes

$$\mathbf{S}_{\mathbf{morph}} = \sum_{i=1}^{K} c_i \mathbf{S}_{\mathbf{i}} \qquad \mathbf{S}_{\mathbf{morph}}, \mathbf{S}_{\mathbf{i}} \in \mathbb{R}^{3\times P}, c_i \in \mathbb{R}$$



Non-Rigid Framework

- Assume that there are K shape bases $\{\mathbf{B_i} \mid i = 1, ..., K\}$
- The 3D coordinate of point p on frame f is given as,

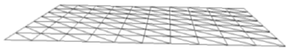
$$\mathbf{X_{fp}} = (x, y, z)^T_{fp} = \sum_{i=1}^{K} c_{fi}\mathbf{b_{ip}} \quad f = 1, ..., F, \; p = 1, ...N \quad (2)$$

- Image coordinate of $\mathbf{X_{fp}}$ under weak perspective projection model is,

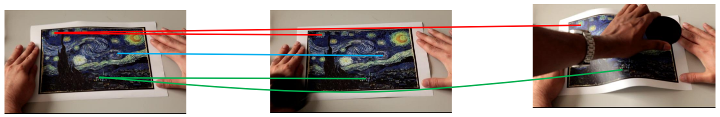
$$\mathbf{x_{fp}} = (u, v)^T_{fp} = s_f(\mathbf{R_f} \cdot \mathbf{X_{fp}} + \mathbf{t_f}) \quad (3)$$

$$\mathbf{x_{fp}} = \begin{pmatrix} c_{f1}\mathbf{R_f} & \dots & c_{fK}\mathbf{R_f} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{b_{1p}} \\ \vdots \\ \mathbf{b_{Kp}} \end{pmatrix} + \mathbf{t_f} \quad (4)$$

The Specific Problem



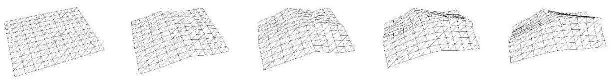
(a) INPUT : 3D mesh



(b) INPUT : Correspondences of Feature Points

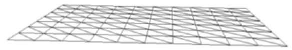
Input/Output:

1. A 3D mesh structure of a deformable object;
2. Location of feature points in a video sequence
3. Recover the object 3D shape for all frames.

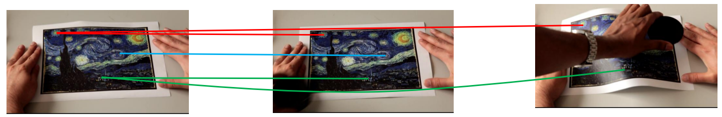


Motivation Factorization Non-Rigid Motion Occlusion

The Specific Problem



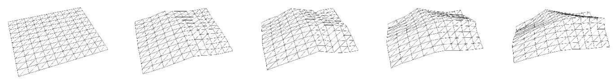
(a) INPUT : 3D mesh



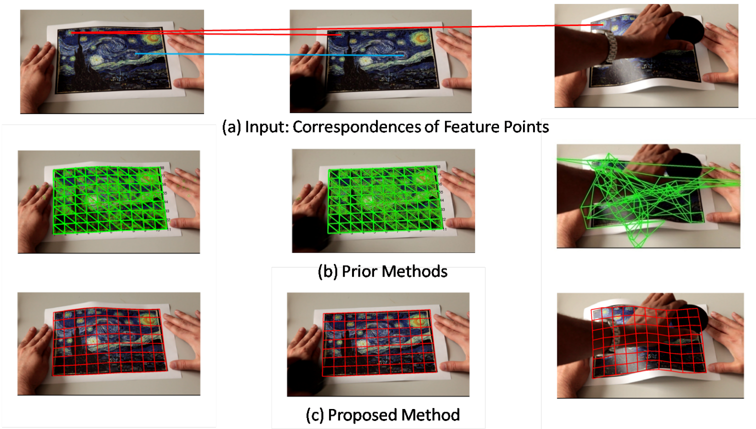
(b) INPUT : Correspondences of Feature Points

Input/Output:

1. A 3D mesh structure of a deformable object;
2. Location of feature points in a video sequence
3. Recover the object 3D shape for all frames.



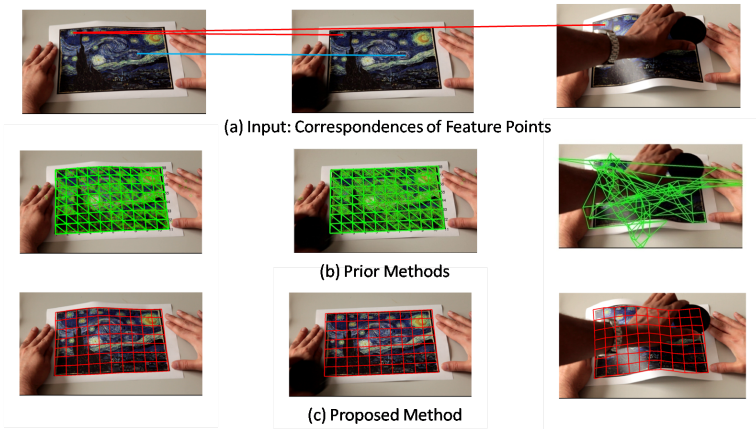
However ...



- Feature points may be partially missing due to occlusions, specular effects, etc. ...
- Reconstruction under occlusions is very troublesome[6] and state-of-the-art algorithms are inadequate.

Motivation Factorization Non-Rigid Motion Occlusion

However ...



- Feature points may be partially missing due to occlusions, specular effects, etc. ...
- Reconstruction under occlusions is very troublesome[6] and state-of-the-art algorithms are inadequate.

Intuition and Idea

Intuition

It should be possible to solve for the missing region in a specific frame, based on the data available in the current, previous and subsequent frames.

Idea

- We assume that the surface is inelastic and deformations should preserve the length of every edge in the mesh.
- We want to find a shape that is consistent with temporal constraints, the deformation model, and one that minimizes the reprojection error.
- This is formulated as an optimization problem on the Riemannian Shape Space.

Motivation Factorization Non-Rigid Motion **Occlusion**

Intuition and Idea

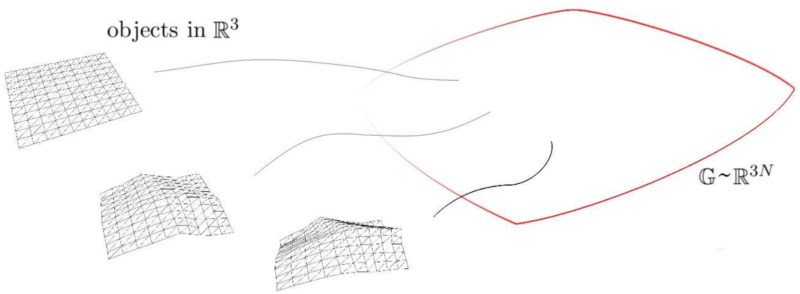
Intuition

It should be possible to solve for the missing region in a specific frame, based on the data available in the current, previous and subsequent frames.

Idea

- We assume that the surface is inelastic and deformations should preserve the length of every edge in the mesh.
- We want to find a shape that is consistent with temporal constraints, the deformation model, and one that minimizes the reprojection error.
- This is formulated as an optimization problem on the Riemannian Shape Space.

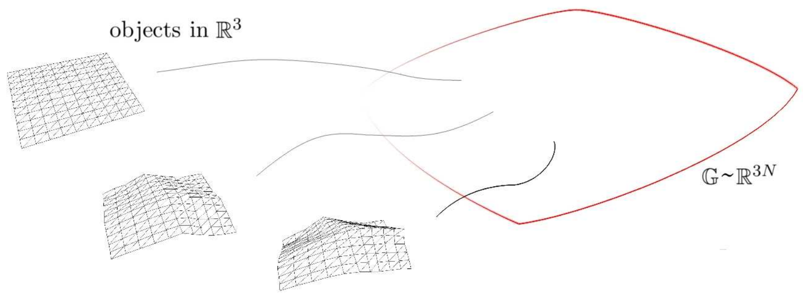
Shape Spaces



- Every point on this space is a 3D mesh.
- A time varying curve in this space corresponds to a deforming shape.
- Technicality: The *local* distance between two neighbouring points is given by the difference in edge lengths of the two meshes.

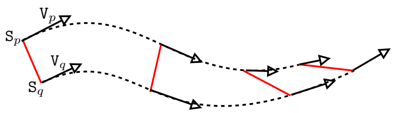
Motivation Factorization Non-Rigid Motion **Occlusion**

Shape Spaces



- Every point on this space is a 3D mesh.
- A time varying curve in this space corresponds to a deforming shape.
- Technicality: The *local* distance between two neighbouring points is given by the difference in edge lengths of the two meshes.

Riemannian Metric



$$\forall \text{ Edge } (p,q) \in \text{Mesh}, \quad \|S_p - S_q\|_2 = \text{const} \tag{5}$$

$$\forall \text{ Edge } (p,q) \in \text{Mesh}, \quad \langle V_p - V_q, S_p - S_q \rangle = 0 \tag{6}$$

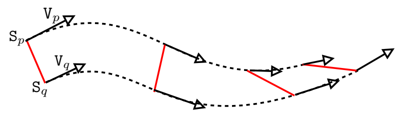
where S_p and S_q are the 3D positions and V_p and V_q are the velocities of vertices p and q respectively.

$$\|V\|_{\text{Iso}} = \sum_{(p,q)\in\text{Mesh}} \langle V_p - V_q, S_p - S_q \rangle$$

A vanishing norm indicates an isometric deformation.

Motivation Factorization Non-Rigid Motion Occlusion

Riemannian Metric



$$\forall \text{ Edge } (p,q) \in \text{Mesh}, \quad \|S_p - S_q\|_2 = \text{const} \tag{5}$$

$$\forall \text{ Edge } (p,q) \in \text{Mesh}, \quad \langle V_p - V_q, S_p - S_q \rangle = 0 \tag{6}$$

where S_p and S_q are the 3D positions and V_p and V_q are the velocities of vertices p and q respectively.

$$\|V\|_{\text{Iso}} = \sum_{(p,q)\in\text{Mesh}} \langle V_p - V_q, S_p - S_q \rangle$$

A vanishing norm indicates an isometric deformation.

Introducing Vision: Reprojection Error

- The 3D coordinates $\mathtt{F}_i^j$ of a feature point j in frame i are given by:

$$\mathtt{F}_i^j = a_j \mathtt{S}_i^{j_1} + b_j \mathtt{S}_i^{j_2} + c_j \mathtt{S}_i^{j_3} \quad (7)$$

where a_j, b_j, c_j are the barycentric coordinates of point j in triangle formed by vertices $\mathtt{S}_i^{j_1}$, $\mathtt{S}_i^{j_2}$, and $\mathtt{S}_i^{j_3}$.

- We have: $\mathtt{f}_i^j = \frac{1}{w_i^j}.\mathbf{C}.\mathtt{F}_i^j$, with $\mathtt{f}_i^j$ the 2D location of feature point j, and $\mathbf{C}$ the perspective projection matrix.
- We can rewrite this equation using Eq. (7) as:

$$\mathbf{m}_i^j.\mathtt{S}_i = 0$$

- By stacking such equation for all feature points, we get the linear system:

$$\mathbf{M}_i.\mathtt{S}_i = 0$$

Therefore, the desired shape $\mathtt{S}_i$ belongs to the null space of $\mathbf{M}_i$.

Introducing Vision: Reprojection Error

- The 3D coordinates $\mathtt{F}_i^j$ of a feature point j in frame i are given by:

$$\mathtt{F}_i^j = a_j \mathtt{S}_i^{j_1} + b_j \mathtt{S}_i^{j_2} + c_j \mathtt{S}_i^{j_3} \quad (7)$$

where a_j, b_j, c_j are the barycentric coordinates of point j in triangle formed by vertices $\mathtt{S}_i^{j_1}$, $\mathtt{S}_i^{j_2}$, and $\mathtt{S}_i^{j_3}$.

- We have: $\mathtt{f}_i^j = \frac{1}{w_i^j}.\mathbf{C}.\mathtt{F}_i^j$, with $\mathtt{f}_i^j$ the 2D location of feature point j, and $\mathbf{C}$ the perspective projection matrix.
- We can rewrite this equation using Eq. (7) as:

$$\mathbf{m}_i^j.\mathtt{S}_i = 0$$

- By stacking such equation for all feature points, we get the linear system:

$$\mathbf{M}_i.\mathtt{S}_i = 0$$

Therefore, the desired shape $\mathtt{S}_i$ belongs to the null space of $\mathbf{M}_i$.

Introducing Vision: Reprojection Error

- The 3D coordinates $\mathtt{F}_i^j$ of a feature point j in frame i are given by:

$$\mathtt{F}_i^j = a_j \mathtt{S}_i^{j_1} + b_j \mathtt{S}_i^{j_2} + c_j \mathtt{S}_i^{j_3} \qquad (7)$$

  where a_j, b_j, c_j are the barycentric coordinates of point j in triangle formed by vertices $\mathtt{S}_i^{j_1}$, $\mathtt{S}_i^{j_2}$, and $\mathtt{S}_i^{j_3}$.
- We have: $\mathtt{f}_i^j = \frac{1}{w_i^j}.\mathbf{C}.\mathtt{F}_i^j$, with $\mathtt{f}_i^j$ the 2D location of feature point j, and $\mathbf{C}$ the perspective projection matrix.
- We can rewrite this equation using Eq. (7) as:

$$\mathbf{m}_i^j.\mathtt{S}_i = 0$$

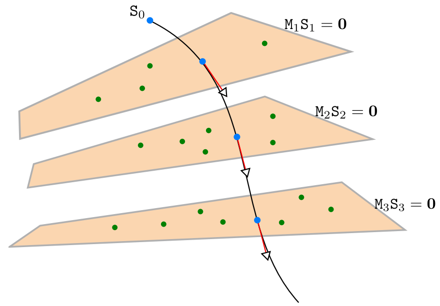
- By stacking such equation for all feature points, we get the linear system:

$$\mathbf{M}_i.\mathtt{S}_i = 0$$

  Therefore, the desired shape $\mathtt{S}_i$ belongs to the null space of $\mathbf{M}_i$.

Motivation Factorization Non-Rigid Motion **Occlusion**

Formulation



The goal is to fit a curve $\{\mathtt{S}_i\}$ in the shape space for the input video sequence.

- The curve should be a geodesic curve to respect the edge length constraint;
- The points on curve should belong to the null space of the $\mathbf{M}_i$ matrices.

Computer Vision Approach: Energies

Deformation Error: $\mathrm{E_D}$ measures the non-isometricity in a deformation sequence.

$$\mathrm{E_{Deform}} = \sum_{i=1}^{F} \sum_{(\mathrm{s}_p, \mathrm{s}_q) \in \mathrm{Mesh}} < \dot{\mathrm{s}_p} - \dot{\mathrm{s}_q}, \mathrm{s}_p - \mathrm{s}_q >^2$$

Reprojection Error:

$$\mathrm{E_{Reproj}} = \sum_{j=1}^{F} ||\mathbf{M}_j \mathrm{s}_j||_2^2$$

Optional Temporal Smoothness Error:

$$\mathrm{E_{Temporal}} = \sum_{i=1}^{F-2} \sum_{\mathrm{v}_i^j \in \mathrm{Vertices}} ||\mathrm{v}_i^j + \mathrm{v}_{i+2}^j - 2\mathrm{v}_{i+1}^j||^2$$

Motivation Factorization Non-Rigid Motion Occlusion

Formulation

Cumulative Cost Function

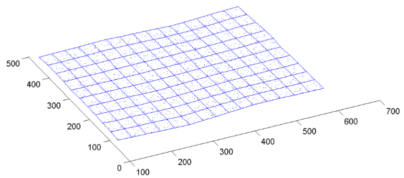
We minimize the following non-convex error term:

$$\min_{s_1 \cdots s_F} \mathrm{E_{Deform}} + \lambda_1 \mathrm{E_{Reproj}} + \lambda_2 \mathrm{E_{Temporal}}$$

- Many commercial softwares can be used, e.g., ‘`fminunc`’ function in matlab.
- However, due to high dimensionality and non-convex nature of problem, we require a reasonable initialization point for the optimization.
- Good initialization leads to faster and better convergence.

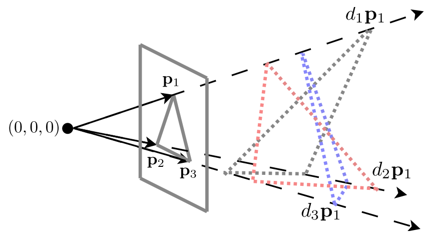
Initialization - Stage 1

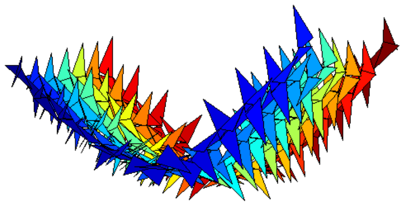
We first recover the *2D projection* of the mesh vertices using weak perspective projection assumption



Initialization - Stage II

- By enforcing mesh length constraints, we recover a maximum of 4 possible shapes for every mesh triangle[1].





1 M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography. Communications ACM, 24(6):381-395, 1981.

Picking the right triangle

We pick the solution that is the most consistent with its neighbours and minimizes the reprojection error.
This can be expressed by the following quadratic program :

$$\min_{\alpha_i,\beta_i,\gamma_i,\delta_i,s_k,k\in\{1\cdots N_v\}} \quad \lambda_1. \sum_{i\in\mathcal{T}(s_j)} \left(\sum_{j=1}^{N_v} \| T_i^*(s_j) - s_j \|^2 \right) + \lambda_2.\|\mathbf{M}.s\|^2$$

$$\text{subject to} \quad : \quad T_i^* = \alpha_i T_i^{(1)} + \beta_i T_i^{(2)} + \gamma_i T_i^{(3)} + \delta_i T_i^{(4)}$$

$$\alpha_i + \beta_i + \gamma_i + \delta_i = 1, \text{with } \alpha_i, \beta_i, \gamma_i, \delta_i \in [0,1]$$

$$\forall\ i \in \{1, N_{\text{facets}}\}$$

where $\mathcal{T}(s_j)$ is the list of facets to which Vertex s_j can belong.
In practice we relax the integer constraints on α, β, γ and δ to a linear one, and change the equality constraint into an inequality one: $\alpha_i, \beta_i, \gamma_i, \delta_i \leq 1$

Motivation Factorization Non-Rigid Motion **Occlusion**

Picking the right triangle

We pick the solution that is the most consistent with its neighbours and minimizes the reprojection error.
This can be expressed by the following quadratic program :

$$\min_{\alpha_i,\beta_i,\gamma_i,\delta_i,s_k,k\in\{1\cdots N_v\}} \quad \lambda_1. \sum_{i\in\mathcal{T}(s_j)} \left(\sum_{j=1}^{N_v} \| T_i^*(s_j) - s_j \|^2 \right) + \lambda_2.\|\mathbf{M}.s\|^2$$

$$\text{subject to} \quad : \quad T_i^* = \alpha_i T_i^{(1)} + \beta_i T_i^{(2)} + \gamma_i T_i^{(3)} + \delta_i T_i^{(4)}$$

$$\alpha_i + \beta_i + \gamma_i + \delta_i = 1, \text{with } \alpha_i, \beta_i, \gamma_i, \delta_i \in [0,1]$$

$$\forall\ i \in \{1, N_{\text{facets}}\}$$

where $\mathcal{T}(s_j)$ is the list of facets to which Vertex s_j can belong.
In practice we relax the integer constraints on α, β, γ and δ to a linear one, and change the equality constraint into an inequality one: $\alpha_i, \beta_i, \gamma_i, \delta_i \leq 1$

Picking the right triangle

We pick the solution that is the most consistent with its neighbours and minimizes the reprojection error.
This can be expressed by the following quadratic program :

$$\min_{\alpha_i,\beta_i,\gamma_i,\delta_i,s_k,k\in\{1\cdots N_v\}} \quad \lambda_1 . \sum_{i\in\mathcal{T}(\mathrm{s}_j)} \left(\sum_{j=1}^{N_v} \| T_i^*(\mathrm{s}_j) - \mathrm{s}_j \|^2 \right) + \lambda_2 . \|\mathbf{M}.\mathrm{s}\|^2$$

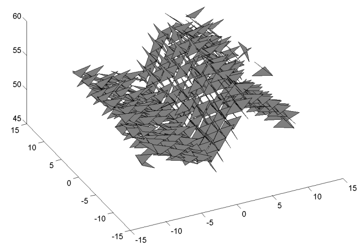
$$\text{subject to} \quad : \quad T_i^* = \alpha_i T_i^{(1)} + \beta_i T_i^{(2)} + \gamma_i T_i^{(3)} + \delta_i T_i^{(4)}$$

$$\alpha_i + \beta_i + \gamma_i + \delta_i = 1, \text{with } \alpha_i, \beta_i, \gamma_i, \delta_i \in [0, 1]$$

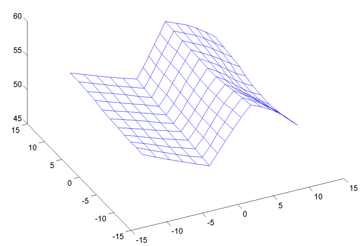
$$\forall\, i \in \{1, N_{\text{facets}}\}$$

where $\mathcal{T}(\mathrm{s}_j)$ is the list of facets to which Vertex s_j can belong.
In practice we relax the integer constraints on α, β, γ and δ to a linear one, and change the equality constraint into an inequality one: $\alpha_i, \beta_i, \gamma_i, \delta_i \leq 1$

Picking the right triangle: Example

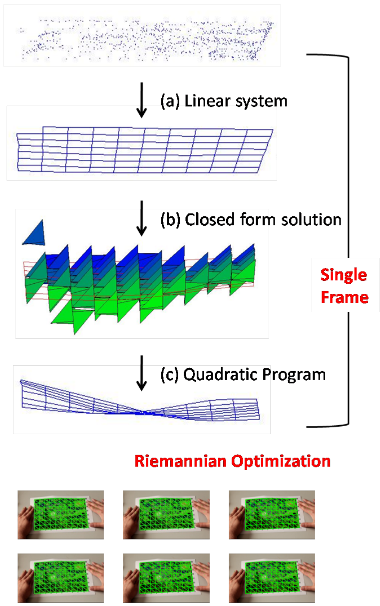


Set of potential triangles



Retrieved (initial) shape

Overall



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.