ICCV 2011 Tutorial on
Non-rigid registration and reconstruction

Dr. Lourdes Agapito
Prof. Adrien Bartoli
Dr. Alessio Del Bue

Non-rigid Structure From Motion - NRSfM
**Problem statement:** To recover the 3D shape and pose of an object deforming over time from a video sequence.

Input 2D tracks form an image sequence:
Why is this important?

**Motion capture/animation**
Andy Serkins - Rise of the Planet of the Apes

**HCI**
OmniTouch – C. Harrison et al. 2011

**Medical Imaging**
A. Malti et al., MIUA 2011

**Augmented reality**
Pilet et al. 2008

**Cloth animation**
R. White et al., SIGGRAPH 2007
Rigid Structure from Motion


The rigid case

Rigid Structure from Motion: The rigidity prior is enough to make the problem well posed and to introduce specific multiview relations for a rigid object.

Mature problem / Completely solved?
Not quite, still some open problems:
Real time, large scale, dense reconstructions, point matching.

Theory is solid and well-established!
Structure from Motion: Rigid case

\[
\begin{bmatrix}
  w_{11} & w_{12} & \cdots & w_{18} \\
  w_{21} & w_{22} & \cdots & w_{28} \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{41} & w_{42} & \cdots & w_{48}
\end{bmatrix}
\begin{bmatrix}
  R_1 \\
  R_2 \\
  R_3 \\
  R_4
\end{bmatrix}
= 
\begin{bmatrix}
  x_1 & x_2 & x_3 & x_4 \\
  y_1 & y_2 & y_3 & y_4 \\
  z_1 & z_2 & z_3 & z_4
\end{bmatrix}
+ 
\begin{bmatrix}
  t_{1x} & t_{1y} & t_{1z} & \cdots & t_{1x} & t_{1y} & t_{1z} & \cdots \\
  t_{2x} & t_{2y} & t_{2z} & \cdots & t_{2x} & t_{2y} & t_{2z} & \cdots \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\
  t_{4x} & t_{4y} & t_{4z} & \cdots & t_{4x} & t_{4y} & t_{4z} & \cdots
\end{bmatrix}
\]
A global solution to rigid SfM

\[
\begin{pmatrix}
W_1 \\
\vdots \\
W_f \\
\end{pmatrix}_{2f \times n} =
\begin{pmatrix}
R_1 \\
\vdots \\
R_f \\
\end{pmatrix}_{2f \times 3}
\begin{pmatrix}
X \\
\end{pmatrix}_{3 \times n}
\]

The image points expressed as a bilinear product of \( M \) and \( X \) are subject to a rank constraint i.e. \( \text{rank}(W) = 3 \)

**How to solve for \( M \) and \( X \)?**

Find an initial decomposition via truncated SVD:

\[
\text{svd}(W) \mapsto \tilde{M} \tilde{X}
\]

However such solution does not satisfy the metric constraints given by an affine camera.
Affine camera projection matrices

1. Orthographic:

\[\begin{bmatrix} R_i \\ \top \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \]

2. Weak perspective:

\[\begin{bmatrix} R_i \\ \top \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \end{bmatrix} \]

3. Para perspective:

\[\begin{bmatrix} R_i \\ \top \end{bmatrix} = \begin{bmatrix} a & c & 0 \\ 0 & c & b \end{bmatrix} \]

Kanatani et al. Uncalibrated factorization using a variable symmetric affine camera. IEICE 2007
Solution:

Find a corrective transform, the mixing matrix $Q$, that imposes the metric constraints

$$M = \tilde{M}Q = \begin{bmatrix} \tilde{R}_1 \\ \vdots \\ \tilde{R}_f \end{bmatrix} Q$$

How to solve for $Q$:

The correct camera matrices $R_i$ must have pairwise orthonormal rows i.e.

$$\tilde{R}_i QQ^\top \tilde{R}_i^\top = I_2$$

Each image provides 3 constraints on $G = QQ^\top$ (orthographic case).

$3f$ constraints on 6 elements of $G$. It can be solved linearly!

Tomasi & Kanade. “Shape and motion from image streams under orthography: a factorization method”. IJCV 1992
Non-rigid case: camera and 3D shape

A 2D image point is represented as a 2-vector containing the image coordinates at the given frame:

\[
\begin{pmatrix}
W_i \\
2 \times n
\end{pmatrix} = \begin{pmatrix}
R_i \\
2 \times 3
\end{pmatrix} \begin{pmatrix}
X_i \\
3 \times n
\end{pmatrix} + \begin{pmatrix}
T_i \\
2 \times n
\end{pmatrix}
\]

- **Measurement Matrix**
- **Camera Matrix**
- **3D shape Matrix**
- **2D translation**
Affine camera projection matrices

1. Orthographic:

\[ \mathbf{R}_i \quad \mathbf{R}_i^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \]

2. Weak perspective:

\[ \mathbf{R}_i \quad \mathbf{R}_i^\top = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \]

3. Para perspective:

\[ \mathbf{R}_i \quad \mathbf{R}_i^\top = \begin{pmatrix} a & c \\ c & b \end{pmatrix} \]
Multiview relations

Each image of a non-rigid body increases the dimensionality of the problem and the number of unknowns.
A severly ill-posed problem

\[
\begin{bmatrix}
\vdots \\
\end{bmatrix}
\begin{bmatrix}
\vdots \\
\end{bmatrix}
= 
\begin{bmatrix}
\vdots \\
\end{bmatrix} + 
\begin{bmatrix}
\vdots \\
\end{bmatrix}
\]

- Measurement matrix is known: 2fn entries
- Camera matrix is unknown: 6f variables
- 3D shape matrix is unknown: 3fn variables

\[
2fn < 6f + 3fn
\]
Formulation of NRSfM

\[
\begin{pmatrix}
W_1 \\
\vdots \\
W_f
\end{pmatrix}
= 
\begin{pmatrix}
R_1 \\
\vdots \\
R_f
\end{pmatrix}
\begin{pmatrix}
X_1 \\
\vdots \\
X_f
\end{pmatrix}
\]

For each image \( i \) we have that:

\[
W_i = R_i X_i \quad \text{subject to} \quad R_i R_i^\top = I_2
\]

The (orthographic) camera matrix enforces non-linear constraints also known as the motion manifold.

\[
\Omega(R, X) = \sum_{i=1}^{f} \|W_i - R_i X_i\|^2 \quad \text{subject to} \quad R_i R_i^\top = I_2
\]

However using solely these constraints is not possible to obtain a unique solution. The only viable option is to use priors or additional constraints.
Several NRSfM methods

Bregler, Hertzmann, Biermann (2000) CVPR

Zhou and Kambhamettu (2000) CVPR
Brand (2001) CVPR
Torresani et al. (2001) CVPR
Tan and Ishikawa (2001) CVIU
Aanæs and Kahl (2002) WMDS
Del Bue and Agapito (2004) ACCV
Xiao et al. (2004) ECCV
Bartoli et al. (2004) CVPR
Brand (2005) CVPR
Del Bue et al. (2006) CVPR
Xiao et al. (2006) IJCV
Del Bue et al. (2007) CVIU
Wang and Wu (2007) LNCS
Torresani et al. (2008) PAMI
Hartley and Vidal (2008) ECCV
Akhter et al. (2008) NIPS
Wang et al. (2008) PRL
Shaji and Chandran (2008) WCVPR
Rabaud and Belongie (2008)
Bartoli et al. (2008) CVPR
Olsen et al, JMIV'08.

Paladini et al. (2009) CVPR
Fayad et al. (2009) BMVC
Ferreira. et al. (2009) BMVC
Ferreira. et al. (2009) ICASSP
Varol et al. (2009) ICCV
Taylor et al. (2010) CVPR
Fayad et al. (2010) ECCV
Del Bue et al. (2010) ECCV
Del Bue (2010) ECCV
Paladini et al. (2010) ECCV
Park et al. (2010) ECCV
Wang et al. (2010) IJCV
Llado et al. (2010) IVC
Gotardo and Martinez (2011) CVPR
Akhter et al. (2011) PAMI
Gotardo and Martinez (2011) PAMI
Russell et al. (2011) CVPR
Fayad et al. (2011) ICCV (now!)
Gotardo and Martínez (2011) ICCV (now!)
Park et al. (2011) ICCV (now!)
Del Bue and Bartoli (2011) ICCV (now!)
Del Bue et al. (2012) PAMI (soon)
A classification of priors for NRSFM: Statistical Priors

Statistical Priors are in general the most used in NRSfM approaches and they can be divided in different categories:

- **Low rank bases model (BHB prior)**
  
  The overall (global) deformations are generated using a finite set of linear 3D bases. These bases span the all possible deformations of the object.

- **Trajectory space bases model**
  
  The camera motion or/and the 3D shape trajectory can be generated by a linear combination of DCT (fixed) bases.

- **Dynamics**
  
  The temporal variations of the deformation parameters in a new observation is given by a linear transformation of the previously observed state.

- **Probabilistic smoothness of deformation**
  
  The parameters controlling the deformation are obtained by sampling a given distribution i.e. Gaussian
Physical Priors can be classified in several categories:

- **Single object**
  The images and/or feature points belong to a single shape

- **Isometry**
  Deformations do not vary the intrinsic distance of points lying on the deforming surface

- **$C_1/ C_0$ continuity**
  Different conditions on the surface continuity/smoothness

- **Elasticity**
  Deformations are subject to elastic constraints given by specific materials

- **Piecewise quadratic/planar**
  A single object can be approximated by a finite series of patches or blocks.

- **Known 3D template**
  The 3D surface is known.
Iterative or Closed Form?

Two main classes of algorithms:

**Closed form solutions**

- Bregler, Hertzmann, Biermann (2000) CVPR
- Brand (2001) CVPR
- Del Bue and Agapito (2004) ACCV
- Xiao et al. (2006) IJCV
- Hartley and Vidal (2008) ECCV

**Iterative solutions**

- Torresani et al. (2001) CVPR
- Brand (2005) CVPR
- Del Bue et al. (2007) CVIU
- Torresani et al. (2008) PAMI
- Akhter et al. (2008) NIPS
- Bartoli et al. (2008) CVPR
- Paladini et al. (2009) CVPR
- Taylor et al. (2010) CVPR
- Fayad et al. (2010) ECCV
- Del Bue et al. (2010) ECCV
- Gotardo and Martinez (2011) CVPR
- Russell et al. (2011) CVPR
- Gotardo and Martinez (2011) ICCV (now!)
- Fayad et al. (2011) ICCV (now!)
- Del Bue and Bartoli (2011) ICCV (now!)
- Del Bue et al. (2012) PAMI (soon)

and many more...
Bregler, Hertzmann and Biermann first proposed the **Low-Rank Bases** priors when they defined the non-rigid structure from motion problem in 2000.

The 3D deformations can be modelled as the linear combination of a set of \( k \) 3D basis shapes (BHB prior).

We have that: \( k \ll f \)

\( k \ll n \)

\[
X_i = l_{i1} \times \begin{array}{c} \text{shape} \end{array} + l_{i2} \times \begin{array}{c} \text{shape} \end{array} + \ldots + l_{ik} \times \begin{array}{c} \text{shape} \end{array}
\]

This prior imposes an implicit rank \( 3k \) constraint in the factorization model

*Bregler, Hertzmann, Biermann (BHB), “Recovering non-rigid shapes from image streams”. CVPR 2000*
A NRSfM rank constraint

\[
\begin{pmatrix}
W_1 \\
\vdots \\
W_f
\end{pmatrix}
= 
\begin{pmatrix}
l_{11}R_1 & l_{12}R_1 & \cdots & l_{1k}R_1 \\
\vdots & \vdots & \ddots & \vdots \\
l_{f1}R_f & l_{f2}R_f & \cdots & l_{fk}R_f
\end{pmatrix}
\begin{pmatrix}
B_1 \\
\vdots \\
B_k
\end{pmatrix}
\]

The projection of the linear combination of the basis shapes can be expressed as a matrix product such as:

\[
W_i = R_i \sum_{i=1}^{k} l_{id}B_d 
= \begin{pmatrix}
W_1 \\
\vdots \\
W_f
\end{pmatrix}
= 
\begin{pmatrix}
l_{11}R_1 & l_{12}R_1 & \cdots & l_{1k}R_1 \\
\vdots & \vdots & \ddots & \vdots \\
l_{f1}R_f & l_{f2}R_f & \cdots & l_{fk}R_f
\end{pmatrix}
\begin{pmatrix}
B_1 \\
\vdots \\
B_k
\end{pmatrix}
\]

Now the optimisation problem can be formulated as:

\[
\Omega(R, L, B) = \sum_{i=1}^{f} \left\| W_i - R_i \sum_{i=1}^{k} l_{id}B_d \right\|^2 
\text{subject to } R_iR_i^\top = I_2
\]
A closed form solution?

The BHB prior reintroduce the rank constraint as in the TK solution for the rigid case. Thus the problem can be similarly solved via an initial bilinear decomposition followed by the estimation of a mixing matrix $Q$.

1. Use truncated SVD: 
   \[ \text{svd}(W) \mapsto \tilde{M} \tilde{X} \]

2. Find the mixing matrix: 
   \[ M_{2f \times 3k} = \tilde{M}_{2f \times 3k} Q_{3k \times 3k} \]

However, we pay something in reintroducing the rank constraint:

1. Stronger non-linearities in 
   \[ \begin{bmatrix} l_{i1}R_i & l_{i2}R_i & \cdots & l_{ik}R_i \end{bmatrix} \]
   Difficult to find a closed form solution.

2. Not clear how to estimate the number of bases $k$
   Choosing a different number of bases may give very different results.
Non-linearities in the solution

1. Stronger non-linearities in

\[
\begin{pmatrix}
 l_{i1}R_i & l_{i2}R_i & \ldots & l_{ik}R_i
\end{pmatrix}
\]

Difficult to find a closed form solution.

Now there are two constraints to be satisfied:

1\textsuperscript{st} Constraint:

\( k \) orthogonality constraints at a generic frame \( i \):

\[
\begin{pmatrix}
 l_{id}R_i \\
 l_{id}^TR_i
\end{pmatrix}
\]

\[
\begin{array}{c|c}
2 \times 3 & 3 \times 2 \\
\hline
 l_{id}R_i & l_{id}^TR_i
\end{array}
\]

\[ = \begin{pmatrix} 1 \\ l_{id}^2I_2 \end{pmatrix} \]

2\textsuperscript{nd} Constraint:

Repetitive block structure of the same 2 x 3 camera matrix \( R_i \)

\[
\begin{pmatrix}
 l_{i1}R_i & l_{i2}R_i & \ldots & l_{ik}R_i
\end{pmatrix}
\]

The factorization is coherent to our model if these constraints are valid for every frame in \( M \).
BHB closed form solution

1. Use truncated SVD at rank $3k$:  
   \[ \text{svd}(W) \mapsto \tilde{M}_{2f \times 3k} \tilde{X}_{3k \times n} \]

2. Recover the camera matrices at each frame using rank-1 SVD at each frame (affine solution).

3. Apply standard TK rigid closed form over the computed camera matrices and enforce metric constraints.
   
   \[ \tilde{R}_i Q Q^\top \tilde{R}_i^\top = I_2 \]

The proposed solution is not able to obtain a correct 3D reconstruction for simple toy problems as shown in Xiao et al. (IJCV 2006)

Bregler, Hertzmann, Biermann (BHB), “Recovering non-rigid shapes from image streams”. CVPR 2000
Xiao et al. attempt to solve for the full $Q$ by introducing a novel set of constraints as $k$ independent basis shapes.

Such constraints boil down in enforcing \textit{a priori} values in the motion matrix:

\[
\begin{bmatrix}
  l_{11}R_1 & l_{12}R_1 & l_{13}R_1 \\
  l_{21}R_2 & l_{22}R_2 & l_{23}R_2 \\
  l_{31}R_3 & l_{32}R_3 & l_{33}R_3
\end{bmatrix}
\rightarrow
\begin{bmatrix}
  R_1 & 0 & 0 \\
  0 & R_2 & 0 \\
  0 & 0 & R_3
\end{bmatrix}
\]

This means that some basis are visible in some frames i.e. they are not linearly combined with the other bases. Such constraints on $M$ can be easily plugged in the computation of the closed form.

However the solution is \textbf{not robust to noise} and highly dependent on the selection of the independent bases.

Akhter et al. CVPR 2009: \textbf{Orthogonality constraints are enough}. Rank 3 of $G$ should be imposed. No practical algorithm for the closed form solution.

\begin{flushright}
J. Xiao, J. Chai and T. Kanade. A Closed-Form Solution to Non-Rigid Shape and Motion Recovery. IJCV 2006
\end{flushright}
Iterative Algorithms

Two main classes of algorithms:

**Closed form solutions**

- Bregler, Hertzmann, Biermann (2000) CVPR
- Brand (2001) CVPR
- Del Bue and Agapito (2004) ACCV
- Xiao et al. (2006) IJCV
- Hartley and Vidal (2008) ECCV

**Iterative solutions**

- Torresani et al. (2001) CVPR
- Brand (2005) CVPR
- Del Bue et al. (2007) CVIU
- Torresani et al. (2008) PAMI
- Akhter et al. (2008) NIPS
- Bartoli et al. (2008) CVPR
- Paladini et al. (2009) CVPR
- Taylor et al. (2010) CVPR
- Fayad et al. (2010) ECCV
- Del Bue et al. (2010) ECCV
- Gotardo and Martinez (2011) CVPR
- Russell et al. (2011) CVPR
- Gotardo and Martinez (2011) ICCV (now!)
- Fayad et al. (2011) ICCV (now!)
- Del Bue and Bartoli (2011) ICCV (now!)
- Del Bue et al. (2012) PAMI (soon)
- and many more…
Non-Rigid Structure from Motion: Current Solutions

Optimization Strategies:

- **Alternation approaches**

  Trilinear: \( \Omega(R, L, B) = \sum_{i=1}^{f} \left\| W_i - R_i \sum_{i=1}^{k} l_{id} B_d \right\|^2 \)
  (Torresani et al. CVPR 2001)

  Bilinear + Manifold Projection: \( \Omega(R, L, B) = \sum_{i=1}^{f} \left\| W_i - R_i \sum_{i=1}^{k} l_{id} B_d \right\|^2 \) subject to \( R_i R_i^T = I_2 \)
  (Paladini et al. IJCV 2011)
  (Del Bue et al. PAMI 2012)

  Expectation Maximization (EM - PPCA):
  \( p(p_{1:T} | G_{1:T}, T_{1:T}, \bar{S}, V, \sigma^2) = \prod_t p(p_t | G_t, T_t, \bar{S}, V, \sigma^2) \)
  (Torresani et al. PAMI 2008)

- **Bundle Adjustment:**

  \( \Omega(R, L, B) = \sum_{i,j}^{f,n} \left\| w_{ij} - R_i \sum_{d=1}^{k} l_{id} B_d \right\|^2 \) + priors
  (Aanaes-Kahl VMDS 2002)
  (Del Bue et al. CVPR 2006)
  (Bartoli et al. CVPR 2008)
Tri-linear approach

\[
\begin{pmatrix}
W_1 \\
\vdots \\
W_f
\end{pmatrix}
\begin{pmatrix}
l_{11}R_1 \\
l_{12}R_1 & \cdots & l_{1k}R_1 \\
\vdots \\
l_{f1}R_f & \cdots & l_{fk}R_f \\
\end{pmatrix}
\begin{pmatrix}
B_1 \\
\vdots \\
B_k
\end{pmatrix}
\]

\[
\Omega(R, L, B) = \sum_{i=1}^{f} \left\| W_i - R_i \sum_{i=1}^{k} l_{id}B_d \right\|^2
\]

The approach alternates between the 3 components of the problem:
1. Estimate the basis \(B\), keep fixed \(R, L\)
2. Estimate the weights \(L\), keep fixed \(B, R\)
3. Estimate the camera matrices \(R\), keep fixed \(L, B\) (using exponential coordinates)

Problems

- Camera matrices cannot be updated in closed form (use of a single Gauss-Newton step)
- Slow convergence and tendency to zig-zagging slowly towards the minimum

Is there a matrix manifold?

\[
W = \begin{pmatrix}
W_1 \\
\vdots \\
W_f
\end{pmatrix} = 
\begin{pmatrix}
M_i \\
\vdots
\end{pmatrix} = 
\begin{pmatrix}
l_{f1}R_f \\
l_{f2}R_f \\
\vdots \\
l_{fk}R_f
\end{pmatrix} + 
\begin{pmatrix}
B_1 \\
\vdots \\
B_k
\end{pmatrix}
\]

\[
\mathcal{M} = \left\{ l_i \otimes R_i : l_i \in \mathbb{R}^d, R_i \in \mathbb{R}^{2 \times 3}, R_i R_i^\top = I_2 \right\}
\]

where \( l_i = [l_{i1} \ldots l_{ik}] \)

Yes!

Instead of solving for the 3 different unknowns \( R \), \( L \), \( B \) it is possible to reformulate the problem as a bilinear factorization with matrix manifold constraints.

\[
\Omega(M, S) = \|W - MS\|^2 \quad \text{subject to} \quad M_i \in \mathcal{M}
\]

The matrix manifold \( \mathcal{M} \) is also known as the motion manifold of the problem. It can be defined for rigid, deformable and articulated bodies.
The metric projection approach

\[
\begin{bmatrix}
W \\
2f \times n
\end{bmatrix}
= 
\begin{bmatrix}
M \\
2f \times 3k
\end{bmatrix}
\begin{bmatrix}
S \\
3k \times n
\end{bmatrix}
\]

The optimisation approach alternates between:

1. Estimation of the \( S \) matrix via pseudoinverse
2. Estimation of the \( M \) matrix via pseudoinverse
3. An **optimal projection** into the *motion manifold*
4. Repeat until convergence.

The projection is defined as the solution of \( f \) problems:

\[
\min_{N_i} \| M_i - N_i \|_F^2, \quad \text{subject to} \quad N_i \in \mathcal{M}
\]

Paladini et al., “Optimal Metric Projections for Deformable and Articulated Structure-From-Motion”, IJCV 2011
Manifold projection: NRSfM case

\[
\begin{bmatrix}
M_i
\end{bmatrix}
\rightarrow
\begin{bmatrix}
l_1 R_i \\
l_2 R_i \\
\vdots \\
l_k R_i
\end{bmatrix}
\]

\[
\min_{N_i} \| M_i - N_i \|_F^2 \quad \text{subject to} \quad N_i \in \mathcal{M}
\]

How to solve:

**Decouple the problem:** minimum w.r.t \( l_{id} \) given \( R_i \):

\[
l_{id} = \frac{\text{trace}[M_{id}^T R_i]}{\| R_i \|^2} = \frac{1}{2} \text{trace}[M_{id}^T R_i]
\]

**Substituting \( l_{id} \) back,** the optimum \( R_i \) can be found by minimising:

\[
\min_{R_i} \quad r_i^T \left[ - \sum_{d=1}^{k} m_{id} m_{id}^T \right] r_i
\]

such that \( R_i R_i^T = I_{2 \times 2} \)

that is a quadratic minimisation problem presents a non-convex constraint.

---

Paladini et al., “Optimal Metric Projections for Deformable and Articulated Structure-From-Motion”, IJCV 2011
Manifold projection: computation and speedup

\[
\min_{N_i} \|M_i - N_i\|^2_F \quad \text{subject to} \quad N_i \in \mathcal{M} \quad \Rightarrow \quad \min_{\{R_i\}} r_i^\top \left[-\sum_{d=1}^{k} m_{id}m_{id}^\top\right] r_i
\]

such that \( R_i R_i^\top = I_{2 \times 2} \)

We developed a **tight convex relaxation** to give the global minimum that can be solved with off-the-shelf solver such as SeDuMi or CVX.

Even if optimal the computation of the projection might be slow -> speedup

- The projection of one frame is not completely unrelated to the following frames.
- The motion matrix \( M \) will vary slightly frame by frame.
- Warm-start strategy.
- Solved using the Newton method on the **Stiefel** manifold i.e. \( R_i R_i^\top = I_{2 \times 2} \)
- Iterative optimisation directly on the **Stiefel** manifold.

Paladini et al., “Optimal Metric Projections for Deformable and Articulated Structure-From-Motion”, IJCV 2011
Some experimental results

Input 2D data with 32% missing data

MP results
Missing Data Results

Input 2D sequence with 40% missing data

MP results

Torresani et al. results
Missing Data Results

Input 2D sequence with 30% missing data

MP results

Torresani et al. results

Paladini et al., “Optimal Metric Projections for Deformable and Articulated Structure-From-Motion”, IJCV 2011
But why robustness to missing data?

In order to understand such robustness we need to go back to the rigid case.

The projection to a motion manifold was first introduced in the rigid case and in the presence of missing data to obtain a robust 3D reconstruction.

Example:

Marques and Costeira. Estimating 3D shape from degenerate sequences with missing data. CVIU 2009
But why robustness to missing data?

In order to understand such robustness we need to go back to the rigid case.

The projection to a motion manifold was first introduced in the rigid case and in the presence of missing data to obtain a robust 3D reconstruction.

Example:

Marques and Costeira. Estimating 3D shape from degenerate sequences with missing data. CVIU 2009
Degenerate sequences and missing data

Example:

- If a single frame contains a **configuration of 2D points that is degenerate**, the accuracy of standard SfM algorithms decrease dramatically.
- In NRSfM, with more degrees of freedom, such effect is even more extreme!

Marques and Costeira. Estimating 3D shape from degenerate sequences with missing data. CVIU 2009
Degenerate sequences and missing data

Example:

- If a single frame contains a configuration of 2D points that is degenerate, the accuracy of standard SfM algorithms decrease dramatically.
- In NRSfM, with more degrees of freedom, such effect is even more extreme!

Marques and Costeira. Estimating 3D shape from degenerate sequences with missing data. CVIU 2009
Advantages:

• Practical optimisation over manifolds
  A change of manifold implies only a change of the projector and not the redesign of the entire algorithm.

• Optimal projection on the motion manifold
  Projections are solved for several cases: rigid, articulated, deformation manifolds.

• Resilience to missing data
  Test on synthetic data can manage missing data ratio up to 80% missing entries. In real test reasonable 3D reconstruction can be obtained up to 40%.

Drawbacks:

• No theoretical guarantee of convergence
  After a projection the error tends to increase, there is no smooth minimization of the overall error.

Paladini et al., “Optimal Metric Projections for Deformable and Articulated Structure-From-Motion”, IJCV 2011
One optimisation problem: BALM (to solve them all)

1 – Create a novel optimisation framework on manifolds with smooth convergence towards a local minima.

\[
\begin{align*}
\text{minimize} & \quad \| W(z) - MS \|^2 \\
\text{subject to} & \quad M_i \in \mathcal{M}, \quad i = 1, \ldots, f,
\end{align*}
\]

Each columns of $M$ lies on a specific manifold given by the specific SfM problem.

2 – Introduce explicitly the missing data as a variable of the problem.

\[
(W(z))_{ij} := \begin{cases} 
W_{ij} & \text{if } (i, j) \in \mathcal{O} \\
Z_{ij} & \text{if } (i, j) \notin \mathcal{O}
\end{cases}
\]

With missing data, we have also to estimate the missing entries $Z_{ij}$.

Several Bilinear Problems

<table>
<thead>
<tr>
<th>Bilinear Problem</th>
<th>S</th>
<th>M</th>
<th>Manifold Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigid SfM [2]</td>
<td>3D points position</td>
<td>Camera motion</td>
<td>Stiefel Matrix</td>
</tr>
<tr>
<td>Non-Rigid SfM [5]</td>
<td>3D basis shapes</td>
<td>Camera motion and deformations</td>
<td>Stiefel Matrix</td>
</tr>
<tr>
<td>Articulated SfM [8]</td>
<td>Two 3D bodies</td>
<td>Camera motion and joint property</td>
<td>Stiefel and sub-Stiefel Matrices</td>
</tr>
<tr>
<td>Surface unfolding problem [9]</td>
<td>Template</td>
<td>Isometry operator</td>
<td>sub-Stiefel Matrix</td>
</tr>
<tr>
<td>Photometric Stereo [1]</td>
<td>Lighting direction</td>
<td>Albedo and surface normals</td>
<td>Normal constraints</td>
</tr>
<tr>
<td>Bilinear 2D-3D registration [10]</td>
<td>Affine camera</td>
<td>Deformation weights</td>
<td>Normal constraints</td>
</tr>
<tr>
<td>Structure from Sound [3]</td>
<td>Microphone location</td>
<td>Sound direction</td>
<td>None</td>
</tr>
<tr>
<td>Head pose and facial expression [4]</td>
<td>Expression</td>
<td>Pose</td>
<td>None</td>
</tr>
<tr>
<td>Face identity and facial expression [12]</td>
<td>Identity</td>
<td>Expressions</td>
<td>None</td>
</tr>
<tr>
<td>Face identity and viewpoint [13]</td>
<td>Identity</td>
<td>Pose</td>
<td>None</td>
</tr>
<tr>
<td>Face identity and illumination [14]</td>
<td>Identity</td>
<td>Illumination</td>
<td>None</td>
</tr>
<tr>
<td>Emotion and speech [15]</td>
<td>Viseme</td>
<td>Expression</td>
<td>None</td>
</tr>
<tr>
<td>{identity, action, viewpoint} [16]</td>
<td>{identity, action}</td>
<td>{view, identity}</td>
<td>Case dependent</td>
</tr>
<tr>
<td>Nonlinear style and content [17]</td>
<td>Content</td>
<td>Style</td>
<td>Custom (learned)</td>
</tr>
<tr>
<td>Factorization of BRDFs [18]</td>
<td>Lighting/Viewpoint</td>
<td>Surface property</td>
<td>Stiefel</td>
</tr>
</tbody>
</table>

The novel concept

Why designing different algorithms if the task is the same?

“The main difference between different factorization problems is the manifold on which the solution lies. Thus, intuitively, it should be possible to construct an unified optimization framework in which a change of the manifold constraint just implies replacing an inner module of the algorithm”

We introduce a reformulation that decouples the core bilinear aspect of the problem from the manifold specificity.

This can be done by cloning $M$ into a new variable $N$

$$N = \begin{bmatrix} N_1 & \cdots & N_i & \cdots & N_f \end{bmatrix} \in \mathbb{R}^{r \times m}, \quad N_i \in \mathbb{R}^{r \times p},$$

and transfer the manifold constraint to the latter.

By doing so, we separate the bilinear issue from the manifold restriction giving:

$$\text{minimize} \quad \|W(z) - MS\|^2$$

$$\text{subject to} \quad M_i = N_i, \quad i = 1, \ldots, f$$

$$N_i \in \mathcal{M}, \quad i = 1, \ldots, f.$$
Algorithm structure

Algorithm 1 Bilinear factorization via Augmented Lagrange Multipliers (BALM)

1: set $k = 0$ and $\epsilon_{\text{best}} = +\infty$
2: initialize $\sigma^{(0)}$, $R^{(0)}$, $\gamma > 1$ and $0 < \eta < 1$
3: initialize $z^{(0)}$, $S^{(0)}$ and $M^{(0)}$
4: repeat
5: solve
\[
(z^{(k+1)}, S^{(k+1)}, M^{(k+1)}, N^{(k+1)}) = \operatorname{argmin} L_{\sigma^{(k)}}(z, S, M, N; R^{(k)})
\]
subject to $N_i \in \mathcal{M}$, $i = 1, \ldots, f$,

using the iterative Gauss-Seidel scheme described in Algorithm 2
6: compute $\epsilon = ||M^{(k+1)} - N^{(k+1)}||^2$
7: if $\epsilon < \eta \epsilon_{\text{best}}$
8: $R^{(k+1)} = R^{(k)} - \sigma^{(k)} (M^{(k+1)} - N^{(k+1)})$
9: $\sigma^{(k+1)} = \sigma^{(k)}$
10: $\epsilon_{\text{best}} = \epsilon$
11: else
12: $R^{(k+1)} = R^{(k)}$
13: $\sigma^{(k+1)} = \gamma \sigma^{(k)}$
14: endif
15: update $k \leftarrow k + 1$
16: until some stopping criterion

Which can be summarised in few steps...
Optimisation steps

<table>
<thead>
<tr>
<th>Minimise</th>
<th>$|W(z) - MS|^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject to</td>
<td>$M_i = N_i, \quad i = 1, \ldots, f$</td>
</tr>
<tr>
<td></td>
<td>$N_i \in \mathcal{M}, \quad i = 1, \ldots, f.$</td>
</tr>
</tbody>
</table>

The optimisation is based on the Augmented Lagrange Multipliers (ALM) method:

$$L_\sigma(z, S, M, N; R) = \|W(z) - MS\|^2 - \sum_{i=1}^{f} \text{tr} \left( R_i^T (M_i - N_i) \right) + \frac{\sigma}{2} \sum_{i=1}^{f} \|M_i - N_i\|^2$$

1. Solve for the manifold projection given $M$ (custom for each problem)

$$N^{[l+1]} = \arg\min_{N} \sum_{i=1}^{f} \left\| N_i - \left( M_i^{[l]} - \frac{1}{\sigma(k)} R_i^{(k)} \right) \right\|^2$$

Subject to $N_i \in \mathcal{M}, \quad i = 1, \ldots, f$

2. Solve for the bilinear factors given $N$ using Gauss-Siedel

$$(S^{[l+1]}, M^{[l+1]}) = \arg\min_{S, M} L_{\sigma(k)} \left( z^{[l]}, S, M, N^{[l+1]}; R^{(k)} \right)$$

3. Given $M$ and $S$ fill the missing entries in $W$ as $Z_{ij} = M_i S_j$

$$z^{[l+1]} = \arg\min_{z} L_{\sigma(k)} \left( z, S^{[l+1]}, M^{[l+1]}, N^{[l+1]}; R^{(k)} \right)$$

BALM Summary

• Practical optimisation over manifolds (same as MP)
  A change of manifold implies only a change of the projector and not the redesign of the entire algorithm.

• Optimal projection on the motion manifold (more projectors)
  Projections are solved for several cases: rigid, articulated, deformation manifolds (photometric stereo and image registration).

• Resilience to missing data (similar as MP)
  Test on synthetic data can manage missing data ratio up to 80% missing entries. In real test reasonable 3D reconstruction can be obtained up to 40%.

• Theoretical guarantee of convergence (new)
  For a large category of manifolds we have now a proof of convergence to a local minima.

Problem 1: Rigid Structure from Motion

60% missing data

90% missing data

Del Bue et al., “Bilinear modelling via Augmented Lagrange Multipliers (BALM)”, PAMI 2012
Problem 2: Articulated Structure from Motion

Del Bue et al., “Bilinear modelling via Augmented Lagrange Multipliers (BALM)”, PAMI 2012
Problem 3: Non-rigid Structure from Motion

Del Bue et al., “Bilinear modelling via Augmented Lagrange Multipliers (BALM)”, PAMI 2012
Problem 4: Non-rigid Image Registration

Synthetic tests against a Branch & Bound (B&B) based approach are surprising. Even if local, BALM can achieve the global optimum. The better performance of BALM are given by B&B discrete nature.

Del Bue et al., “Bilinear modelling via Augmented Lagrange Multipliers (BALM)”, PAMI 2012
Problem 5: Photometric Stereo

Del Bue et al., “Bilinear modelling via Augmented Lagrange Multipliers (BALM)”, PAMI 2012
EM based NRSfM
Torresani et al.

The algorithm is a reformulation of Expectation-Maximization (EM) algorithm adapted to the NRSfM problem.

**E-STEP:**

Estimate the latent \( L_i \) variables for each frame \( i = 1 \cdots f \)

\[
L_i = [l_{i1} \cdots l_{ik}]
\]

**M-STEP:**

Estimate the remaining variables (camera motion and bases)

Notice that this step sums up to a non-linear cost function because the orthonormality constraints. This is solved with a single Gauss-Newton step as in the trilinear approach.
Statistical Priors for EM

Two contributions:

• The standard linear basis model (PCA) is upgraded to a probabilistic PCA (PPCA) model. This results in placing a statistical prior with a Gaussian distribution over the deformation weights $l_{id}$.

$$L_i = [l_{i1} \cdots l_{ik}] \sim \mathcal{N}(0, I)$$

• A linear dynamics model is introduce which enforce temporal consistency on the deformation weights $l_{id}$.

$$L_i = GL_{i-1} + n_i \quad \quad n_i \sim \mathcal{N}(0, Z)$$

Results

Coarse to fine approach

A coarse to fine approach uses a series of nested minimisation problems which iteratively adds 3D deformation modes. The idea is that the modes capture decreasingly important details in the deformation.

Advantages:

- Automatically selects the number of deformation modes ($k$).
- It can handle missing data.
- It uses priors during the incremental modes estimation: temporal smoothness, proximity measure, and the ordering of deformations modes.

Modes estimation

1. First estimate the mean shape using Rigid SfM (Tomasi and Kanade, IJCV 1992). This gives the mean shape with respect to the group of Euclidean transformations and perspective projection (see the Deformotion paper (Yezzi and Soatto, IJCV’03)).

\[ W_i^1 = R_i B_1 + t_i \]

Set \( d = 1 \)

2. Find the deformation mode \( S_{k+1} \) and configuration weight \( l_{k+1} \) with Linear and Nonlinear Least Squares. The algorithm uses surface shape and temporal smoothness priors.

\[ W_i^{d+1} = W_i^d + l_{i,d+1} R_i B_{d+1} + t_i \]

3. Stop or increase \( d \) and loop to step 2
Stopping Criterion

• The reprojection error always decreases with the number of modes

• The method computes a $v$-fold cross-validation score

• This score either stabilizes or increases when ‘superfluous’ modes are added

Results of Coarse-to-Fine Low-Rank Shape Model

0.84 pixels reprojection

Videos: warp visualization grid (left), reconstructed surface (middle) and augmented video (right)
Reconstructed Coarse-to-Fine Low-Rank Shape Model

Video: tracked face (bottom) and reconstructed face (top)
Augmented Video

Video: augmented face
Evaluation setup

- Ground truth data given by a motion capture system (VICON).
- Several camera configurations are created randomly but with smooth trajectories.
- Different ratio of missing data where created up to 80%
- Tested on 1000 trials per configuration
Evaluation
Low-rank bases priors

![Graphs showing 3D error vs. missing data with different noise levels.](image-url)
Summary on NRSfM: Low-rank Bases

• Closed form solutions
  However there exists only approximate solutions or subjective to specific constraints (i.e. indipendent bases of Xiao et al.). They are in general sensible to noise in the measurements).

• Iterative solutions
  Given a right choice of the number of bases $k$ and a correct tuning of the priors, they achieve low reprojection error and accurate 3D reconstructions in respect to the ground truth.

• Few points, few frames
  As long as we have $f > 3k$, $n > 3k$ we can reconstruct the non-rigid body. Of course, more measurements we have, better the reconstruction will be.

• Missing data
  This is a real challenge. The actual limit is 80% - 90% missing data in the measurements for a reasonable 3D reconstruction (Metric Projections and BALM). Can we do more?

• Are low rank bases sufficient?
  Deformations may have highly non-linear motion patterns. Is an approximation in terms of linear bases enough? Could we add physical constraints on the problem?
Grazie!

...and now Lourdes presenting trajectory space and piecewise methods