Dense 3D Modelling and Monocular Reconstruction of Deformable Objects

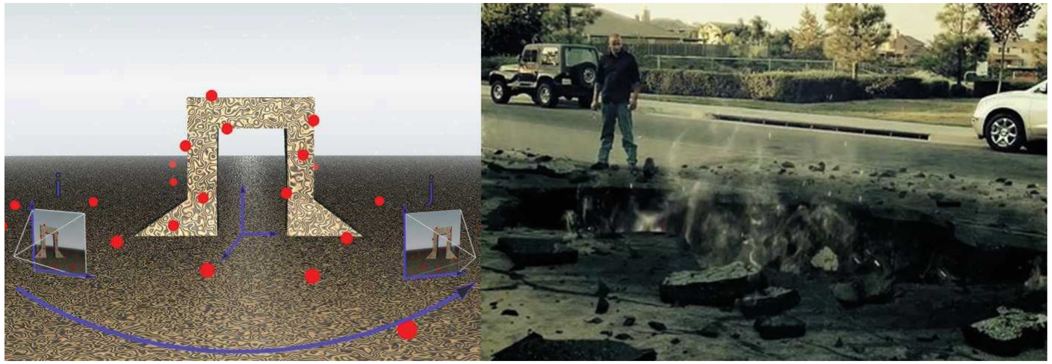
Anastasios (Tassos) Roussos
Lecturer in Computer Science, University of Exeter
Research Associate, Imperial College London

Overview of Recent & Ongoing Research
March 2017

Presentation Outline

3D Computer Vision

- Inference of **3D information** from **2D images**
- Wide variety of **real-world applications**



Match Moving

Kinect

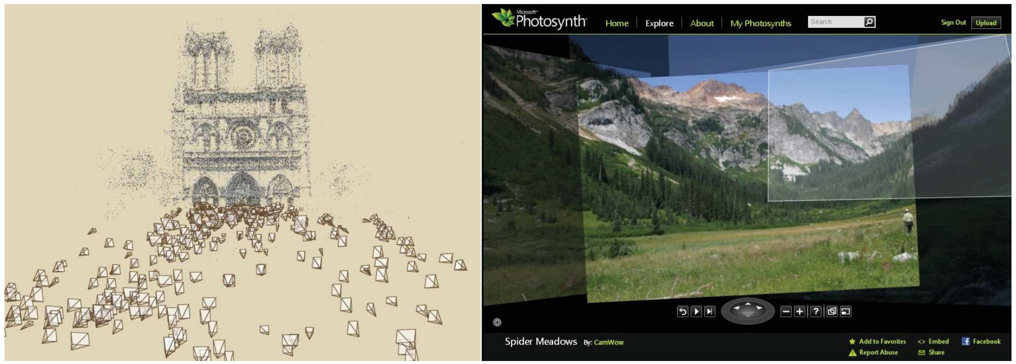


Photo Tourism, Photosynth

Dense 3D models of buildings

3D Computer Vision

- Inference of **3D information** from **2D images**
- Wide variety of **real-world applications**

Faceshift software

Avatar motion capture

MPI & 3dMD 4D scanner

The Digital Emily Project

3D Computer Vision

(Thies et al., Face2Face: Real-time Face Capture and Reenactment of RGB Videos, CVPR'16).

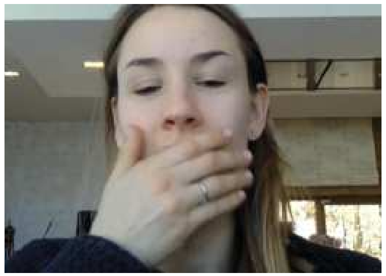
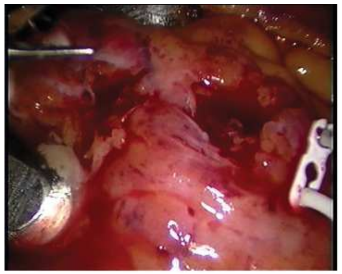
3D Computer Vision: Limitations

- Assumption of **rigidity**:
  - input 2D images: different viewpoints of **exactly the same 3D scene**
- **Multi-camera** systems:
  - equivalent to **single camera** capturing a **rigid** scene
  - **expensive** acquisition setups
- **Active** sensors:
  - limitations in the **acquisition conditions**
- **Specific class** of objects:
  - **unrealistic** shape model priors

3D Computer Vision: Overcoming the Limitations

- Core questions:
  - *How can we make detailed 3D reconstruction work in any real-world scene?*
  - *How can we minimise the acquisition requirements?*
- Vision: **robust** and **fast** systems that:
  - work under almost any condition
  - use practical, low cost acquisition devices

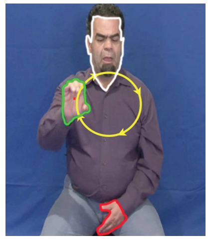
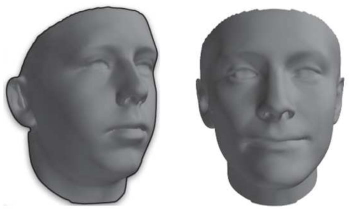


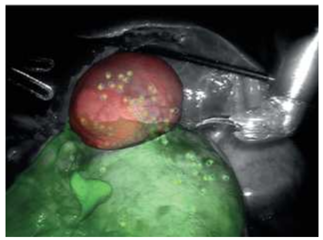





3D Computer Vision: Overcoming the Limitations

- Core questions:
  - *How can we make detailed 3D reconstruction work in any real-world scene?*
  - *How can we minimise the acquisition requirements?*
- Vision: **robust** and **fast** systems that:
  - work under almost any condition
  - use practical, low cost acquisition devices
- Applications:

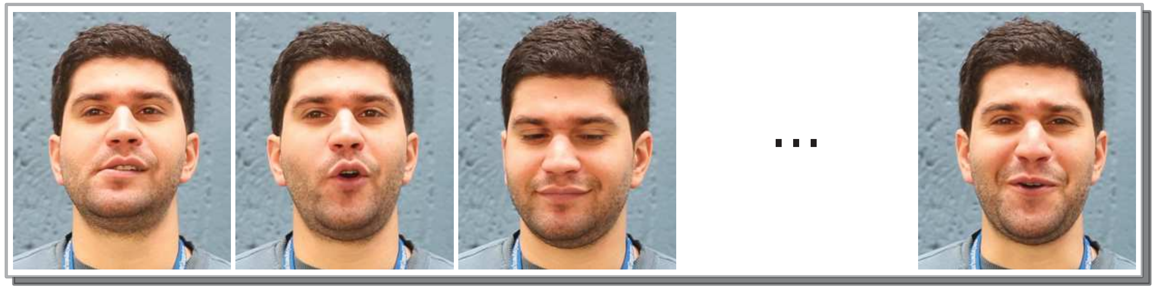


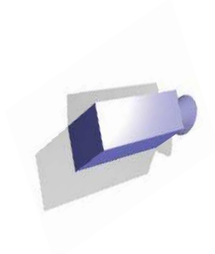





Dense 3D Reconstruction from Monocular Sequences

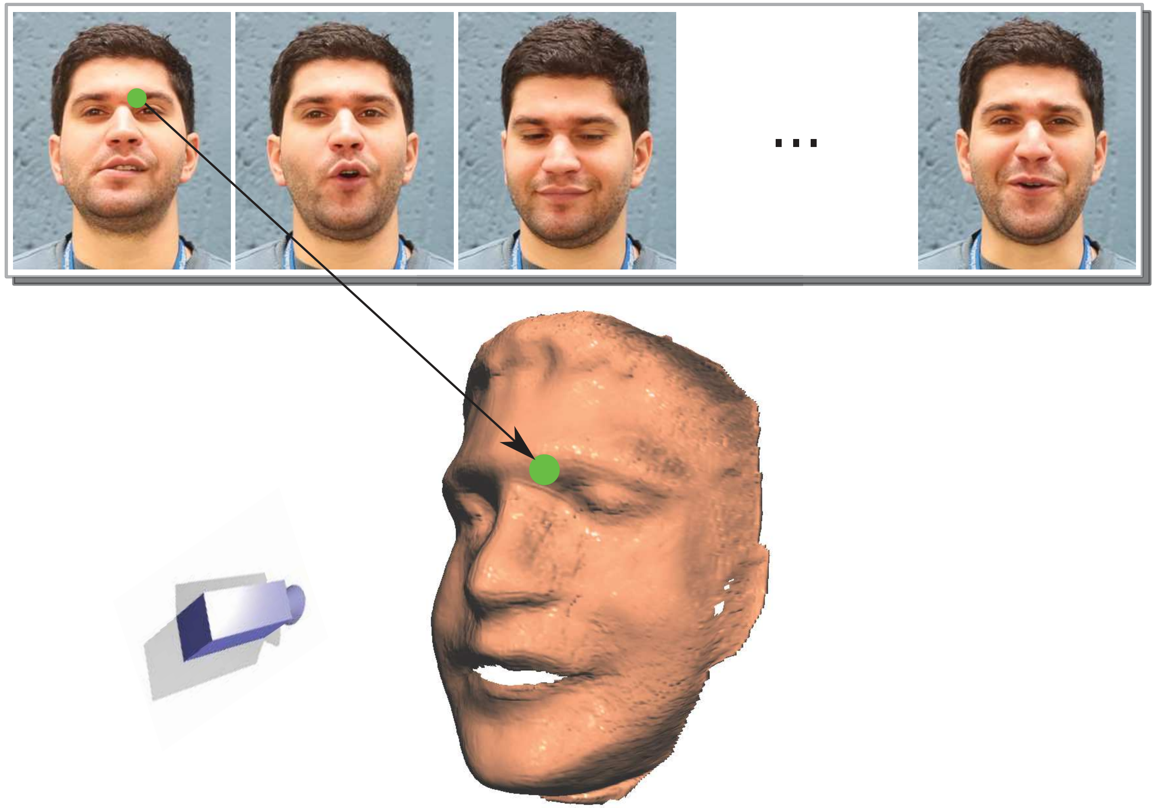
Input: **monocular** sequence of **non-rigid** scene





Dense 3D Reconstruction from Monocular Sequences

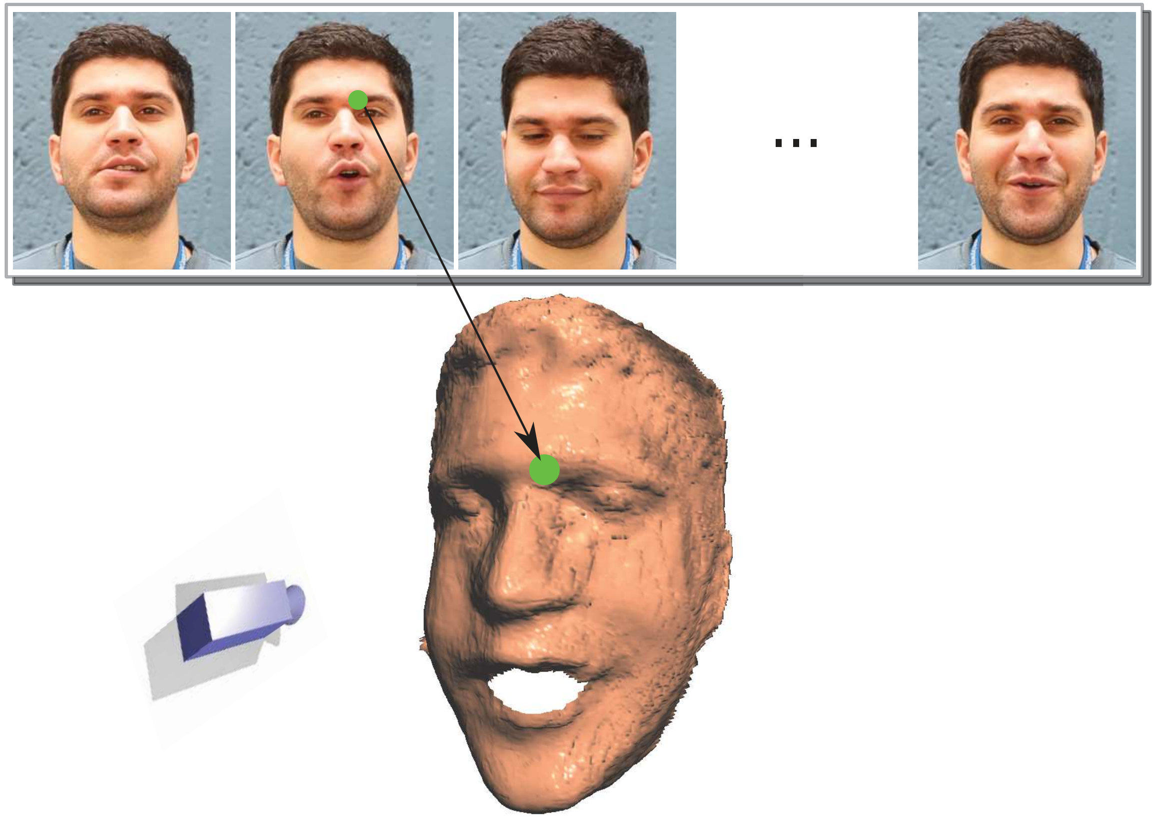
Input: **monocular** sequence of **non-rigid** scene



Goal: estimation of 3D location of **every pixel** at **every frame**

Dense 3D Reconstruction from Monocular Sequences

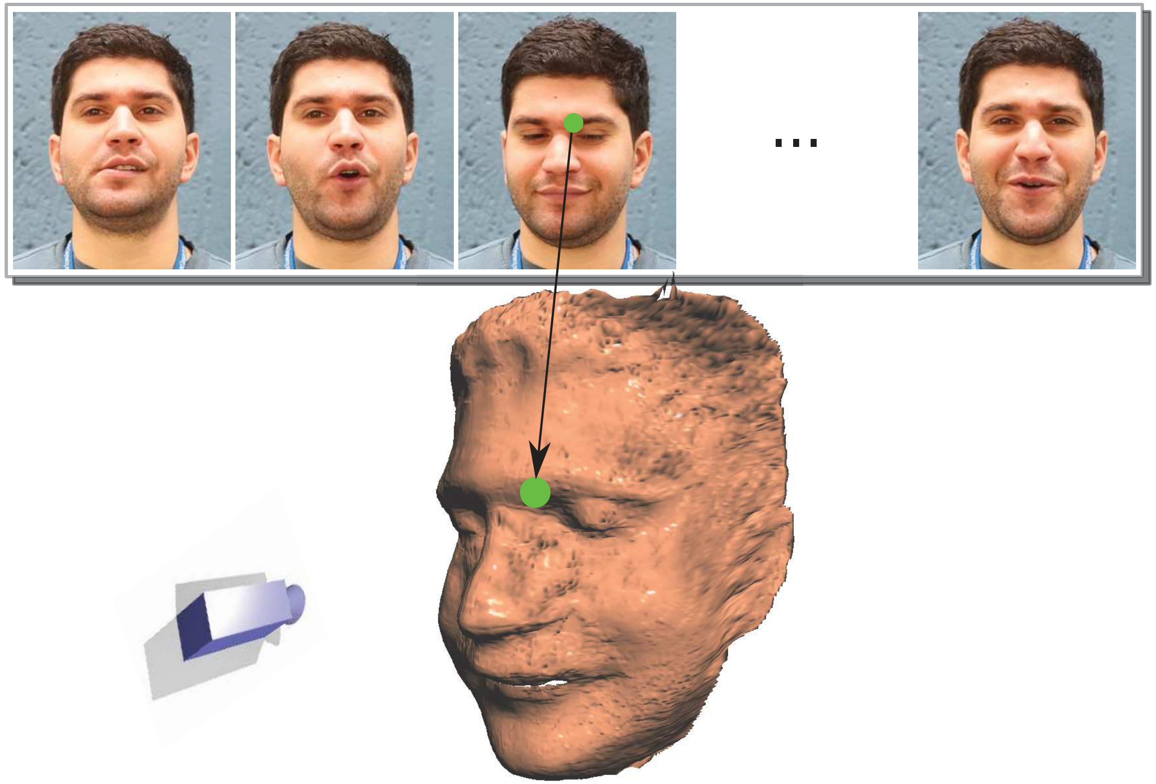
Input: **monocular** sequence of **non-rigid** scene



Goal: estimation of 3D location of **every pixel** at **every frame**

Dense 3D Reconstruction from Monocular Sequences

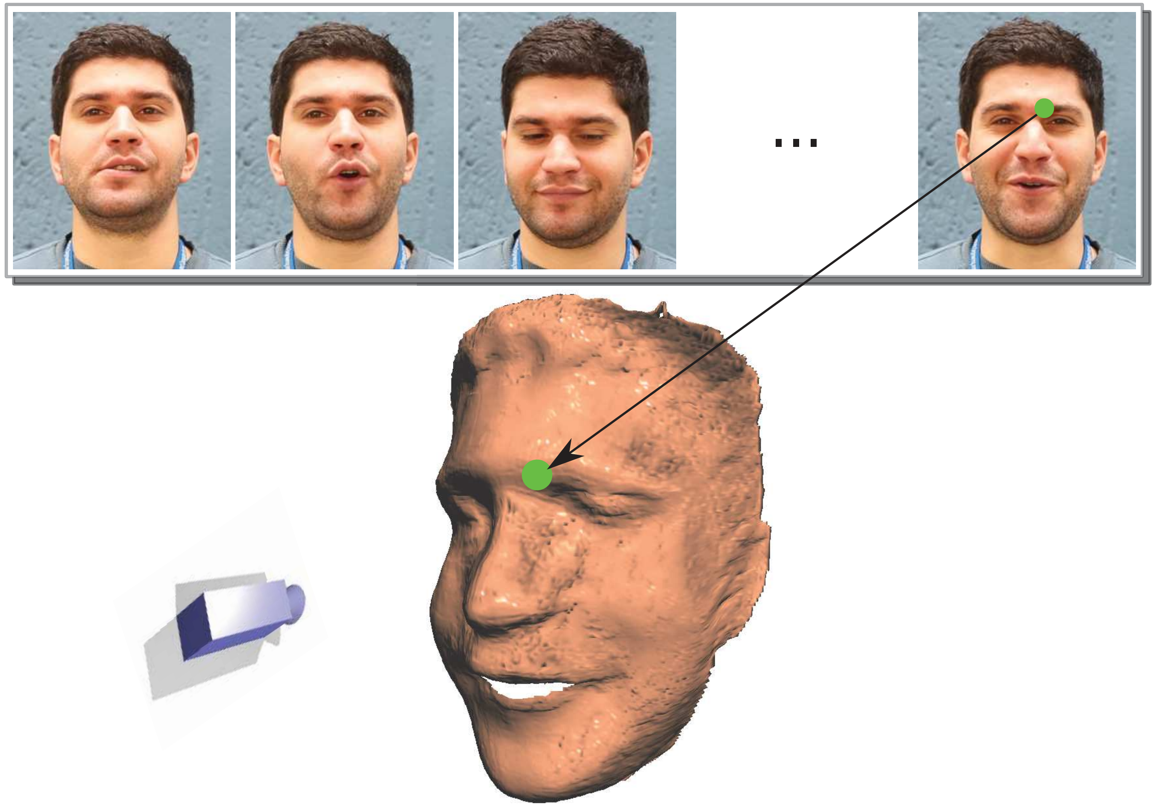
Input: **monocular** sequence of **non-rigid** scene



Goal: estimation of 3D location of **every pixel** at **every frame**

Dense 3D Reconstruction from Monocular Sequences

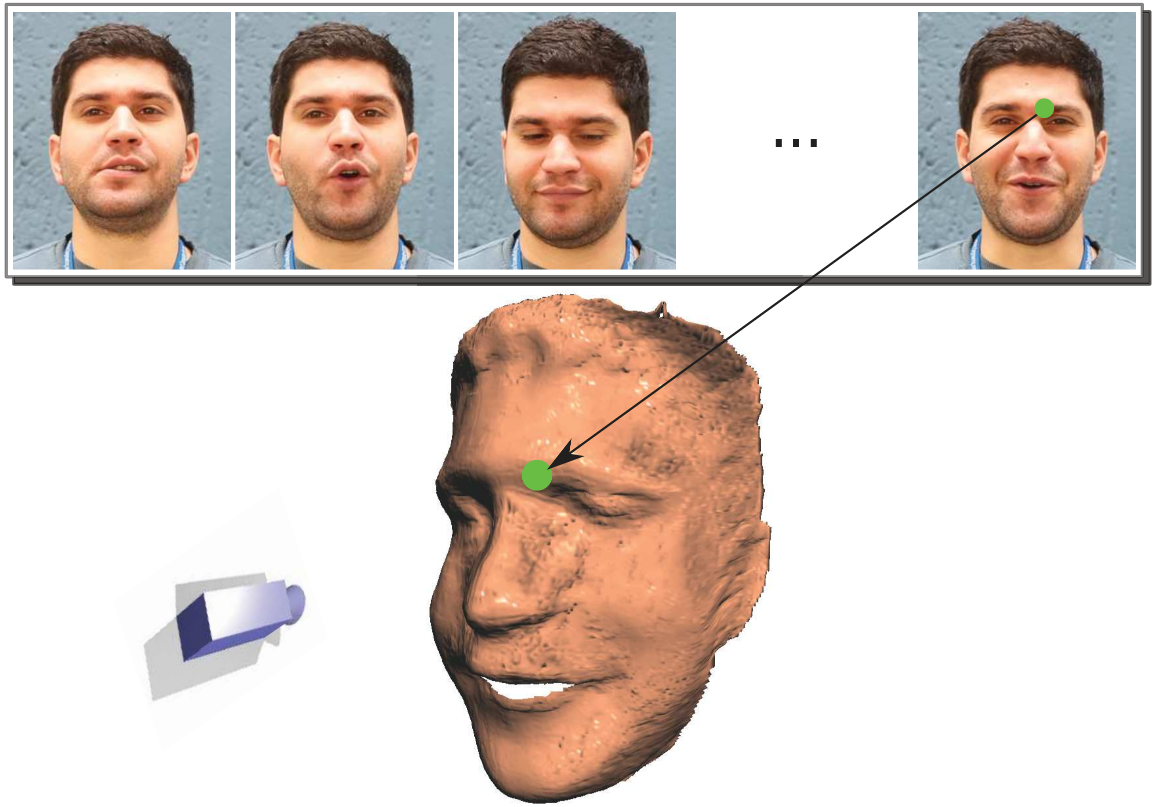
Input: **monocular** sequence of **non-rigid** scene



Goal: estimation of 3D location of **every pixel** at **every frame**

Dense 3D Reconstruction from Monocular Sequences

Input: **monocular** sequence of **non-rigid** scene



Goal: estimation of 3D location of **every pixel** at **every frame**

Approaches:

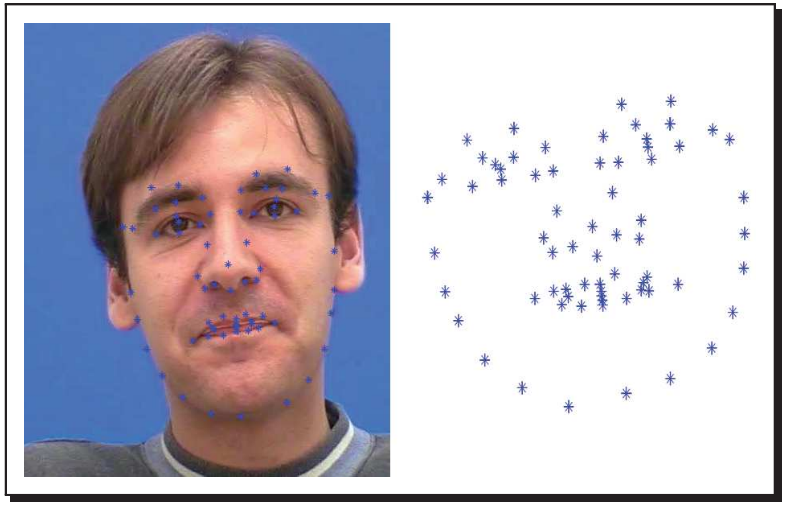
- **model-free**: no prior knowledge about the scene object(s)
- **model-based**: object-specific prior shape models

Presentation Outline

Model-free: Non-rigid Structure from Motion (NRSfM)

- **Leap** from sparse to dense NRSfM

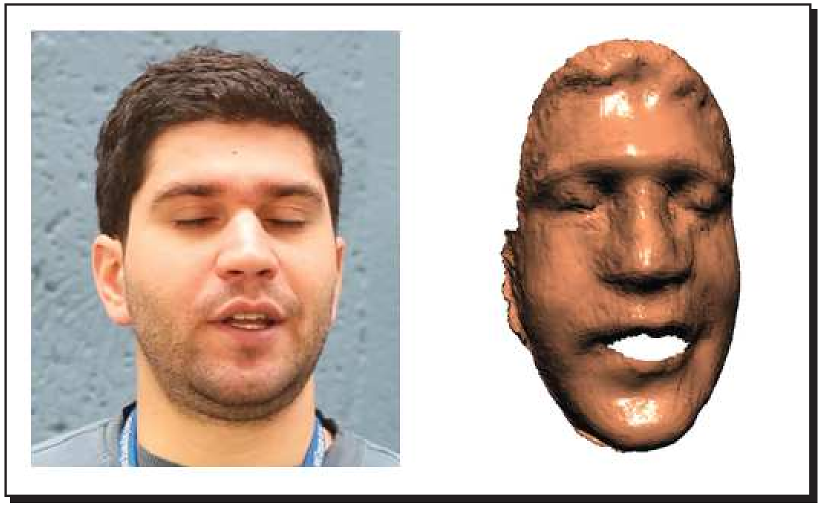
Sparse



(Dai,Li,He, CVPR'12)

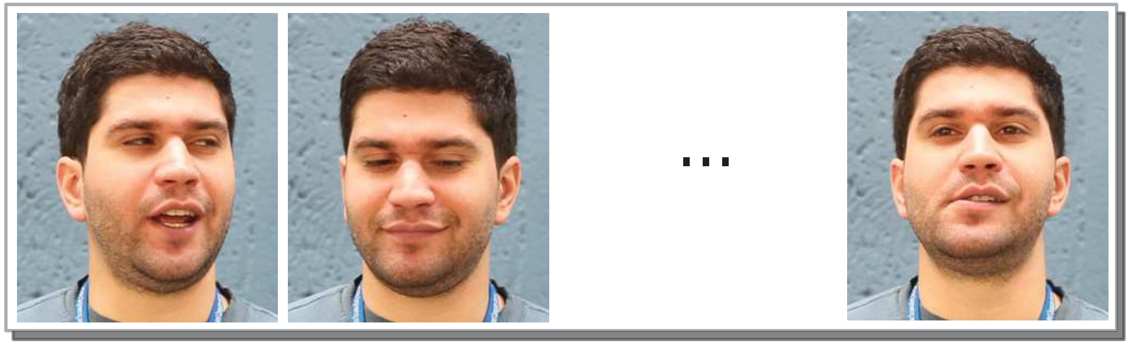
$\rightsquigarrow$

Dense



(Garg,Roussos,Agapito, CVPR'13)

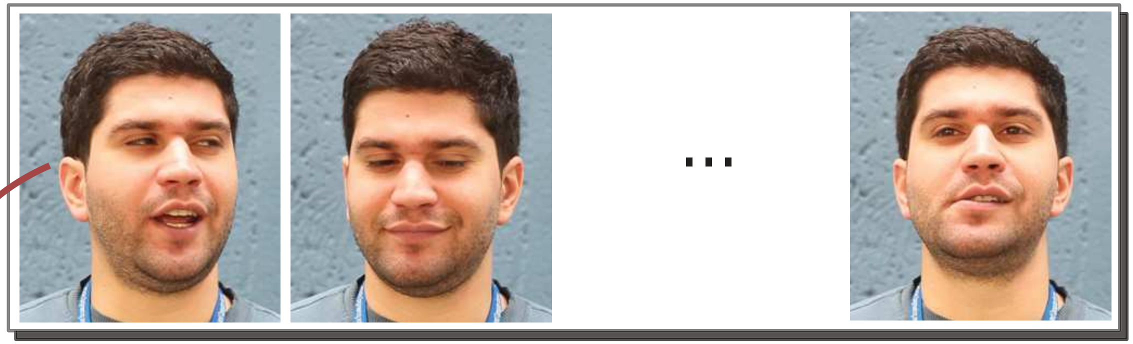
Our Pipeline

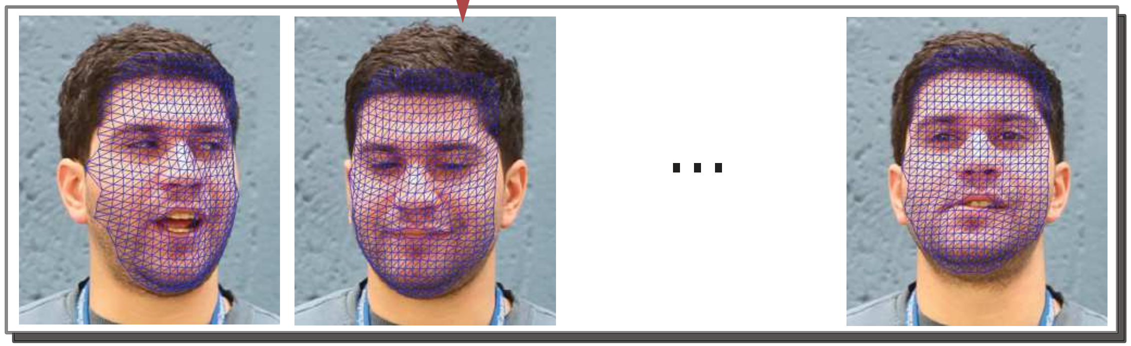


(Roussos, Russell, Garg, Agapito, *IEEE ISMAR 2012*)
(Garg, Roussos, Agapito, *International Journal of Computer Vision 2013*)
(Garg, Roussos, Agapito, *IEEE CVPR 2013*)

Our Pipeline

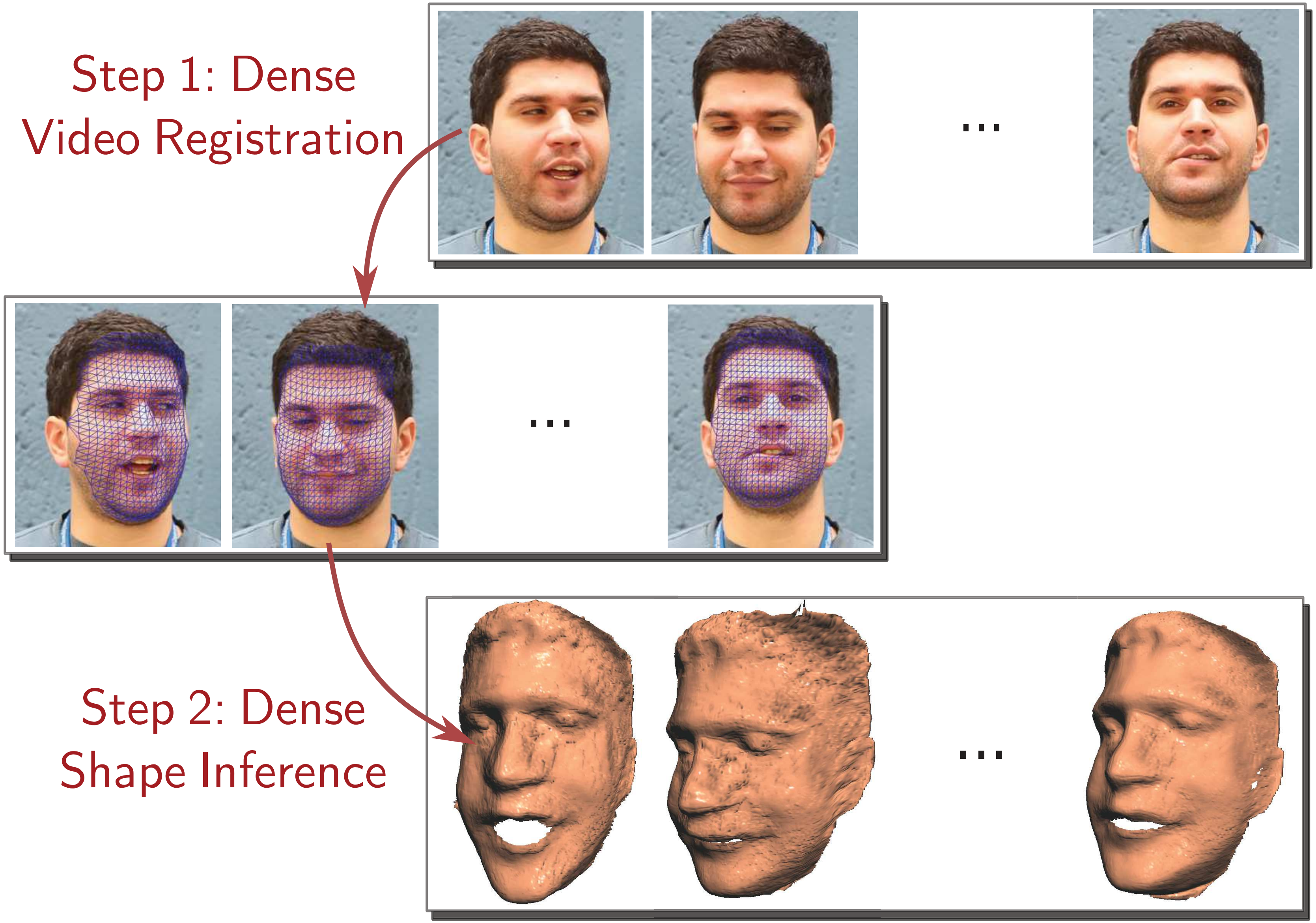
Step 1: Dense Video Registration





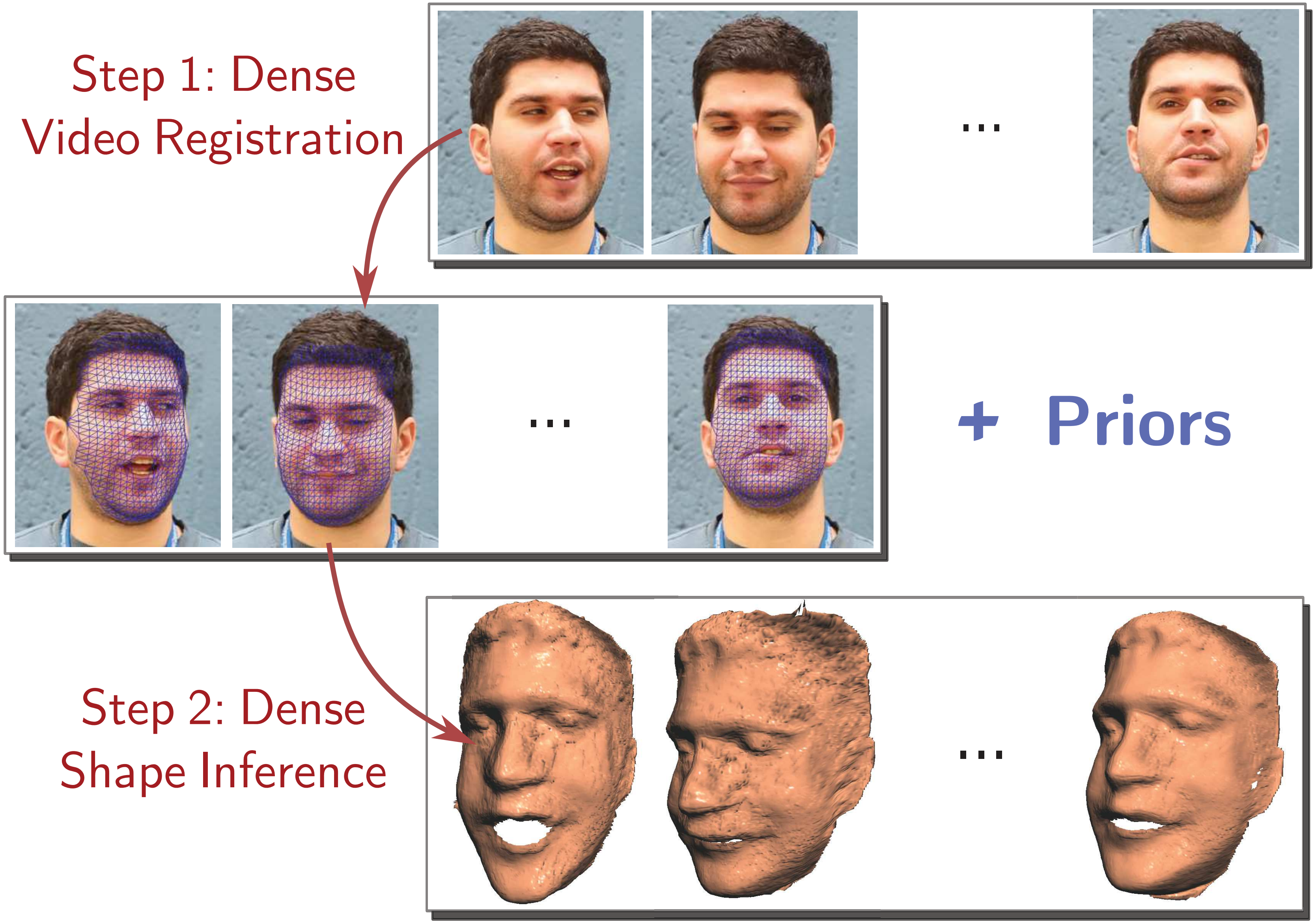
(Roussos, Russell, Garg, Agapito, *IEEE ISMAR 2012*)
(Garg, Roussos, Agapito, *International Journal of Computer Vision 2013*)
(Garg, Roussos, Agapito, *IEEE CVPR 2013*)

Our Pipeline



(Roussos, Russell, Garg, Agapito, *IEEE ISMAR 2012*)
(Garg, Roussos, Agapito, *International Journal of Computer Vision 2013*)
(Garg, Roussos, Agapito, *IEEE CVPR 2013*)

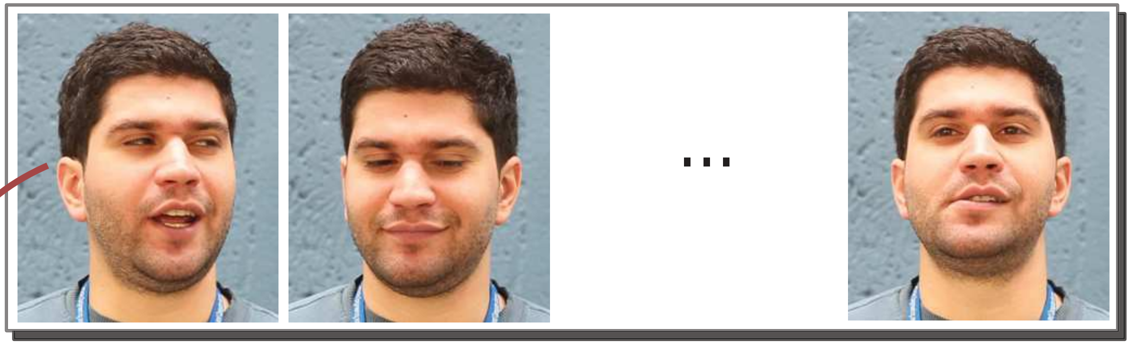
Our Pipeline

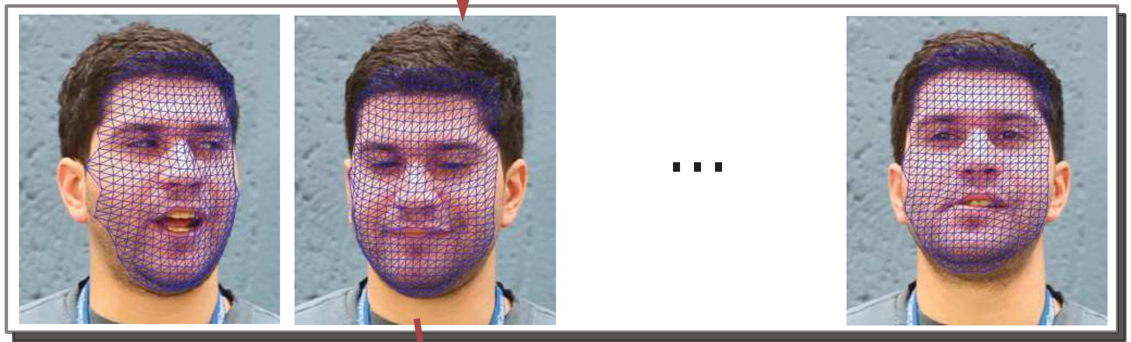


(Roussos, Russell, Garg, Agapito, *IEEE ISMAR 2012*)
(Garg, Roussos, Agapito, *International Journal of Computer Vision 2013*)
(Garg, Roussos, Agapito, *IEEE CVPR 2013*)

Our Pipeline

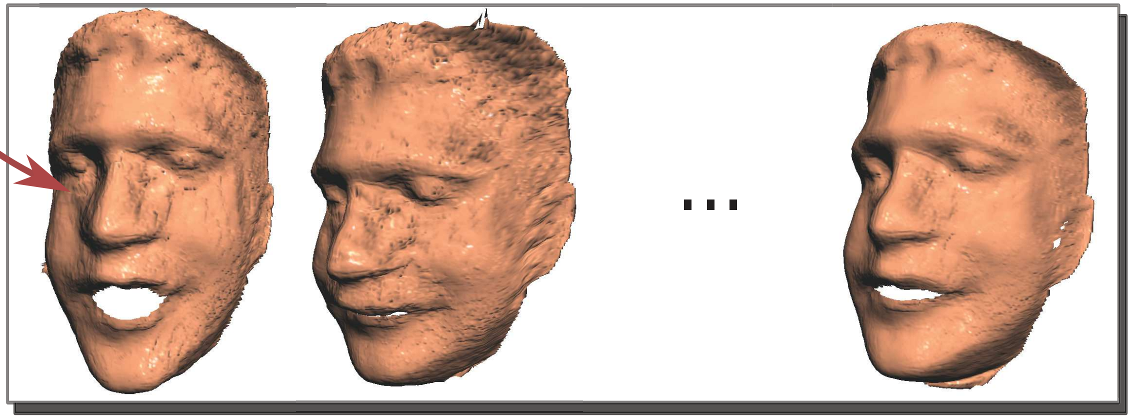
Step 1: Dense
Video Registration





+ Low rank.
Spatial smoothness.

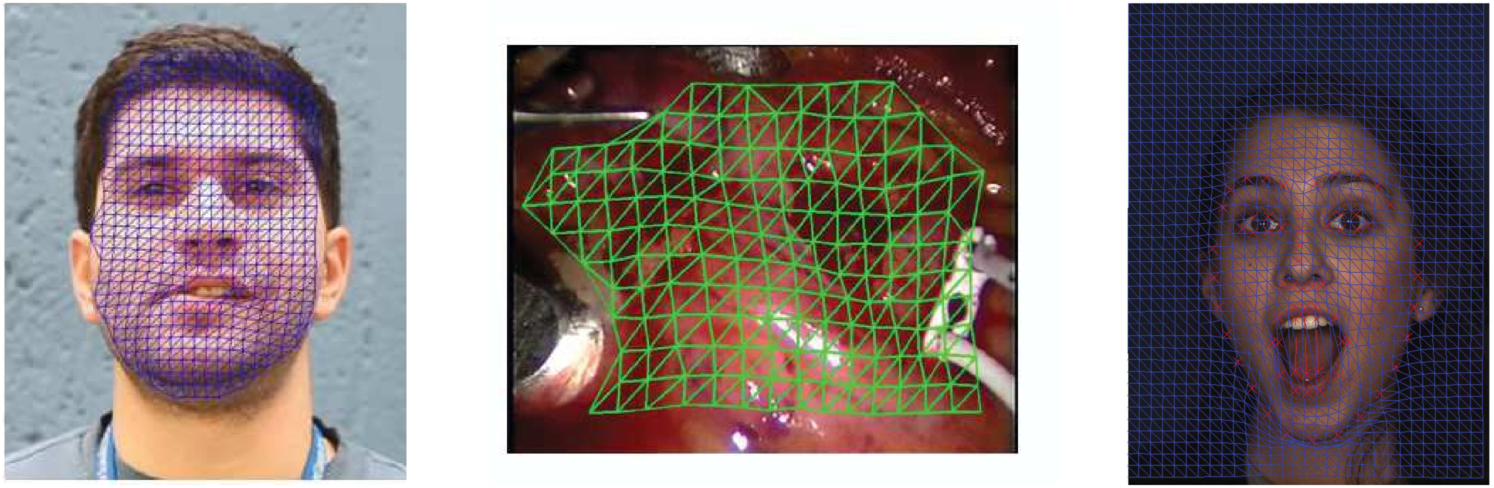
Step 2: Dense
Shape Inference



(Roussos, Russell, Garg, Agapito, *IEEE ISMAR 2012*)
(Garg, Roussos, Agapito, *International Journal of Computer Vision 2013*)
(Garg, Roussos, Agapito, *IEEE CVPR 2013*)

Multi-frame Subspace Flow (MFSF)

- Robust Subspace Constraints for Video Registration

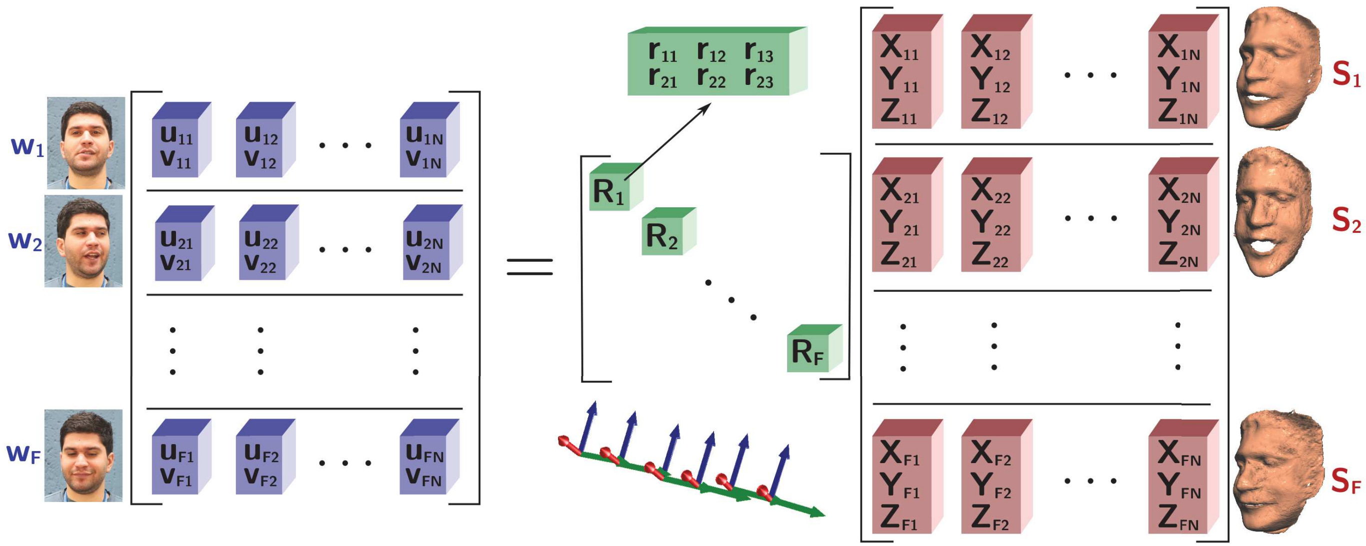


- **The code is now publicly available at**:
  `https://bitbucket.org/troussos/mfsf`

Model-Free Dense 3D Reconstruction from Videos

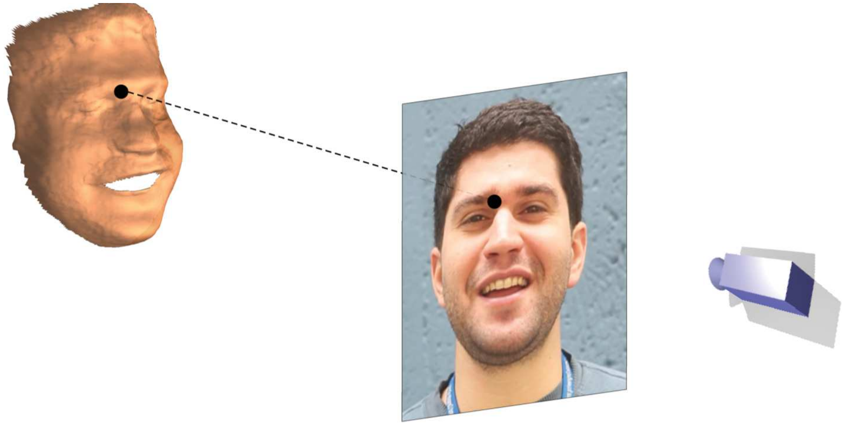
- W: input dense 2D tracks, computed with (Garg,Roussos,Agapito, IJCV'13)
- R, S: the unknown rotations and shapes per frame

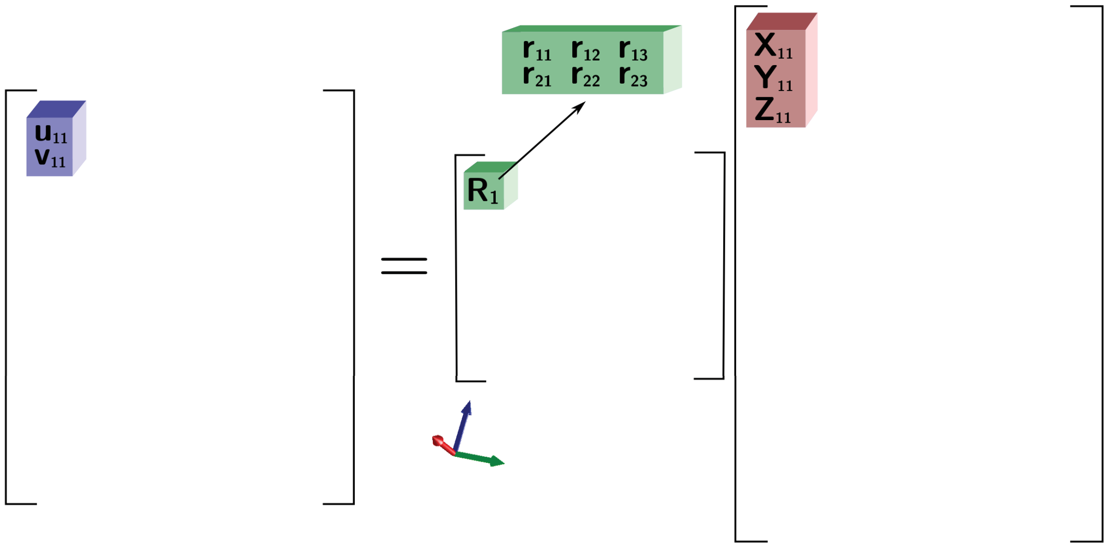
$$W = RS$$



(Garg, Roussos, Agapito, IEEE CVPR 2013)

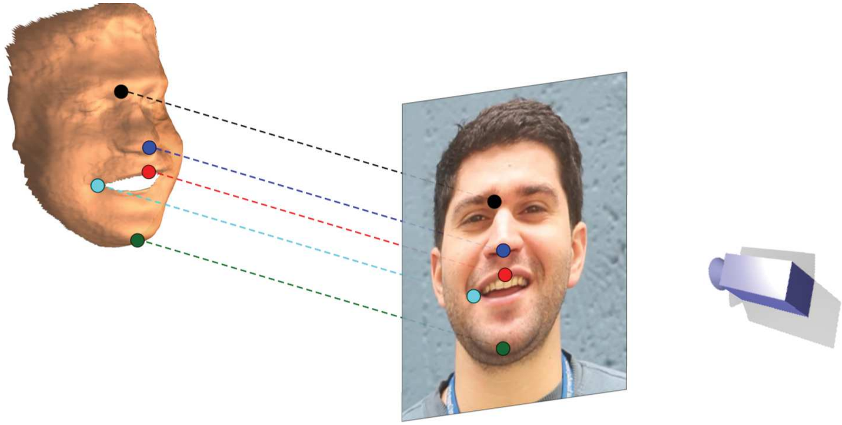
Model-Free Dense 3D Reconstruction from Videos

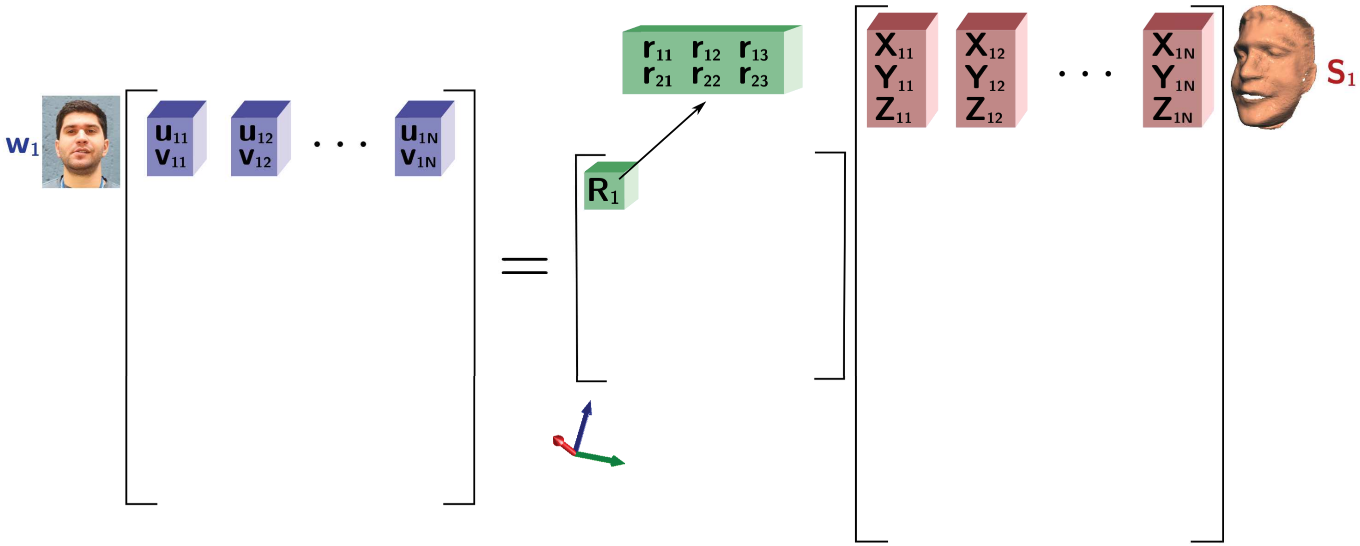




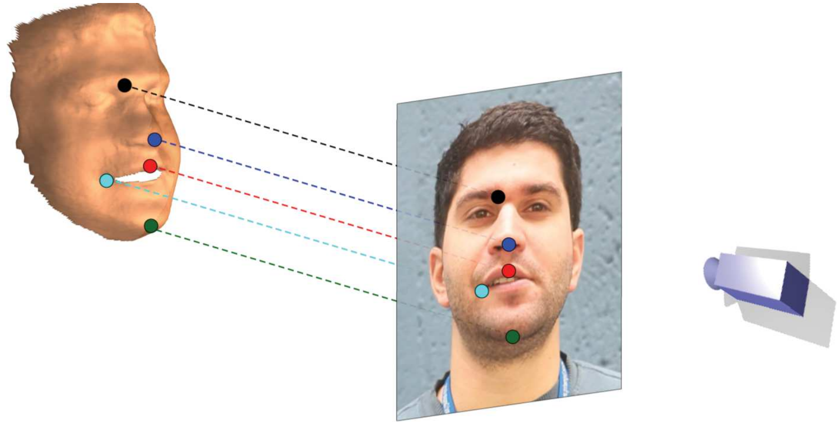
(Garg, Roussos, Agapito, IEEE CVPR 2013)

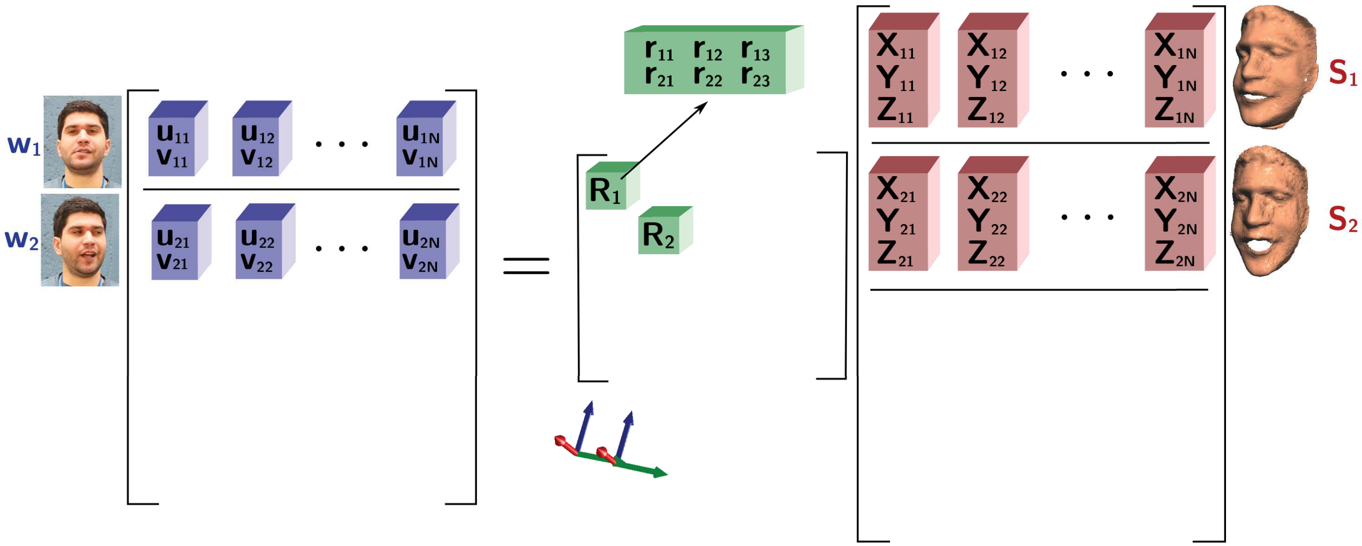
Model-Free Dense 3D Reconstruction from Videos





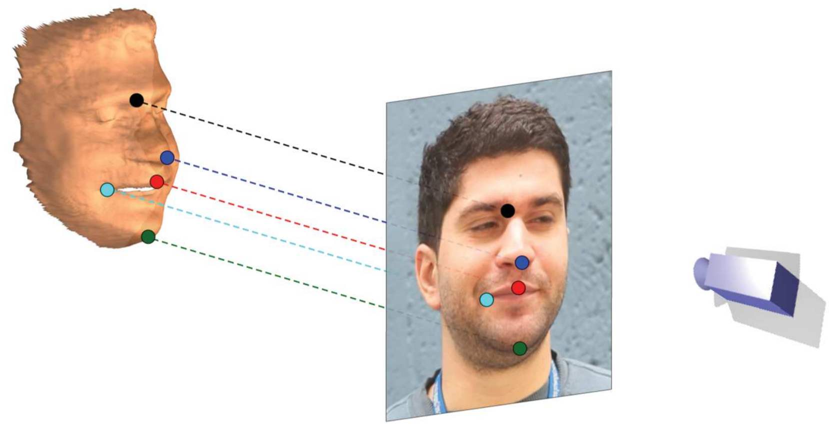
(Garg, Roussos, Agapito, IEEE CVPR 2013)

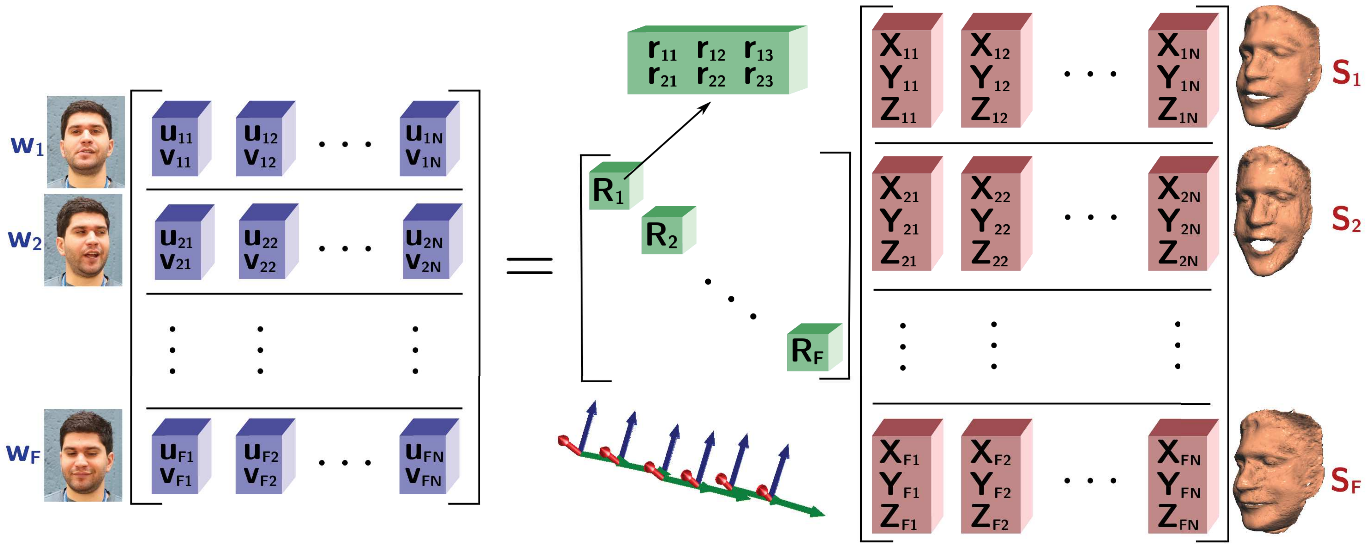




(Garg, Roussos, Agapito, IEEE CVPR 2013)

Model-Free Dense 3D Reconstruction from Videos





(Garg, Roussos, Agapito, IEEE CVPR 2013)

Energy Minimisation Approach to NRSfM

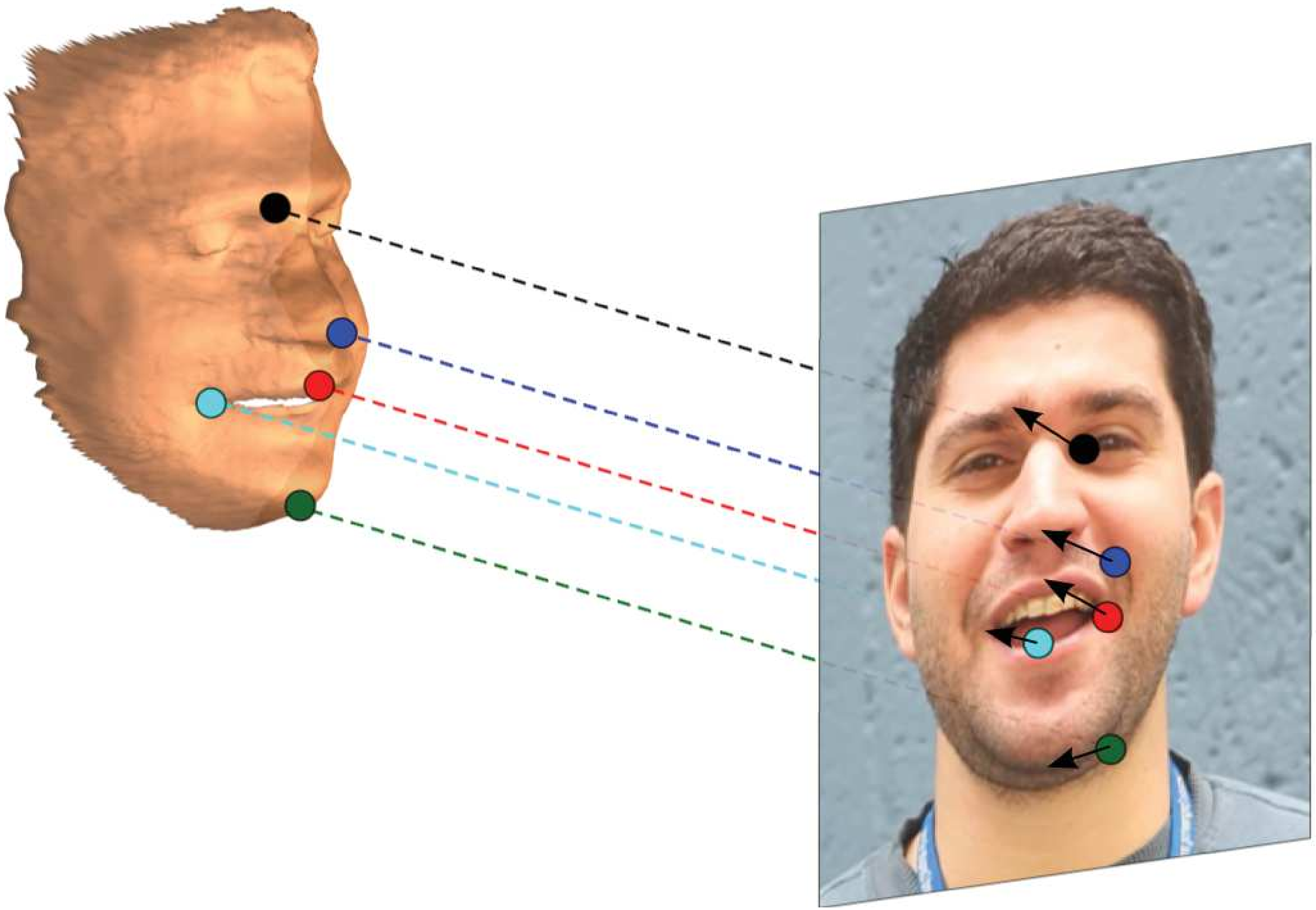
Formulation of a **single unified energy** to estimate:

- Orthographic projection matrices
- 3D shapes for all the frames

$$E(\mathbf{R}, \mathbf{S}) = \lambda E_{data}(\mathbf{R}, \mathbf{S}) + E_{reg}(\mathbf{S}) + \tau E_{trace}(\mathbf{S})$$

- reprojection error over all frames
- spatial smoothness prior on 3D shapes
- low rank prior on 3D shapes

$$E_{data}(R, S) = \|W - RS\|_{\mathcal{F}}^2$$



$$E_{reg}(S) = \sum_i TV(S_i)$$

Without regularisation

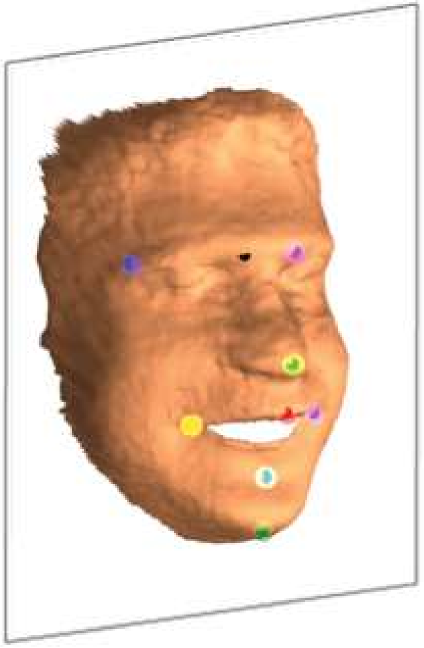
$\longrightarrow$

With regularisation

Low Rank Prior

$$E(\mathbf{R},\mathbf{S}) = \lambda E_{data}(\mathbf{R},\mathbf{S}) + E_{reg}(\mathbf{S}) + \tau E_{trace}(\mathbf{S})$$

$$E_{trace}(\mathbf{S}) = \|\mathbf{S}\|_* = \sum_i \sigma_i(\mathbf{S})$$

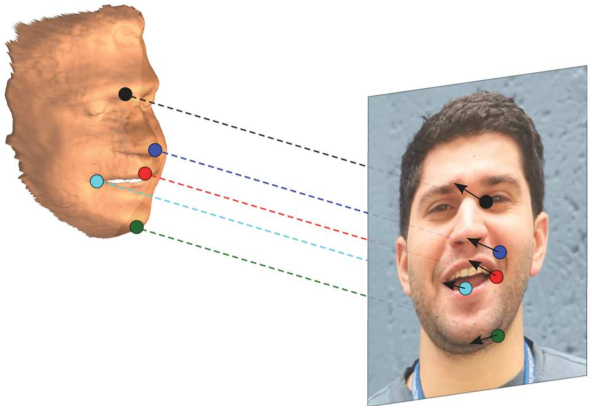


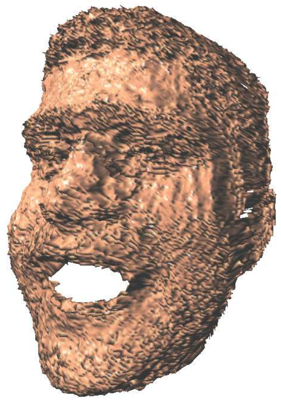
$$\xrightarrow{\textit{lies in}} \textbf{span} \left\{ \quad \right\}$$

$$K \ll F$$

Angst et al. ECCV'12, Dai et al. CVPR'12, Angst et al. ICCV'11, Dai et al. ECCV'10

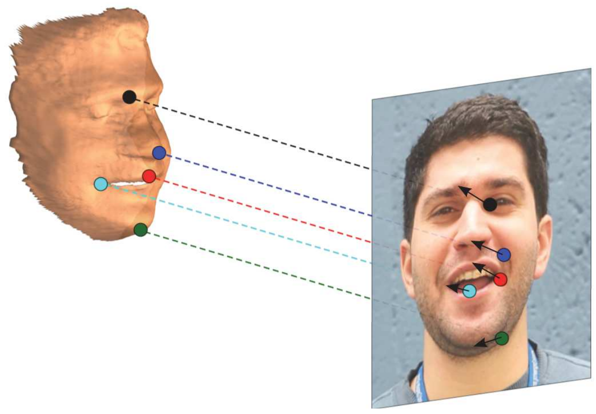
Minimisation of $E(\mathbf{R}, \mathbf{S})$

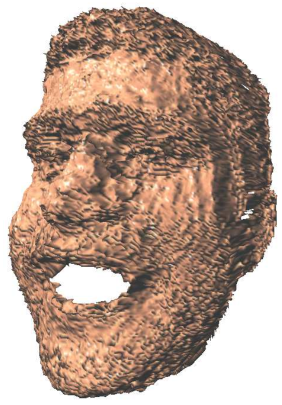




$$\min_{\mathbf{R},\mathbf{S}} \lambda \underbrace{\|W - \mathbf{R}\mathbf{S}\|_{\mathcal{F}}^{2}}_{\substack{\textit{Reprojection}\\ \textit{error}}} + \sum_{i} \underbrace{TV(\mathbf{S}_i)}_{\substack{\textit{Smoothness}\\ \textit{prior}}} + \tau \underbrace{\|\mathbf{S}\|_{*}}_{\substack{\textit{Low rank}\\ \textit{prior}}}$$

Minimisation of $E(R, S)$



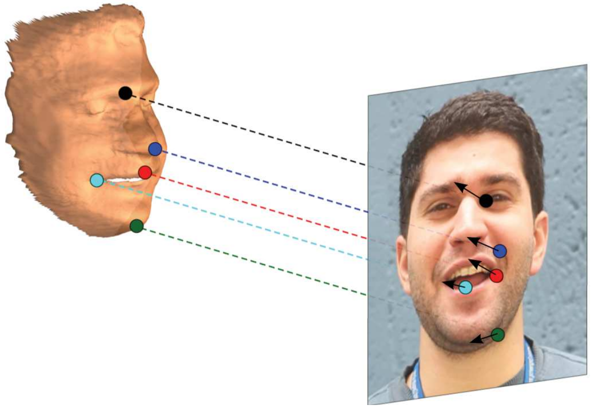


$$\min_{R,S} \lambda \underbrace{\|W - RS\|_{\mathcal{F}}^2}_{\textit{Reprojection error}} + \sum_i \underbrace{TV(S_i)}_{\textit{Smoothness prior}} + \tau \underbrace{\|S\|_*}_{\textit{Low rank prior}}$$

Our Algorithm

- **Initialize** R and S using rigid factorisation.
- Minimize energy via **alternation**:
  - Step 1: Rotation estimation.
  - Step 2: Shape estimation.
- Efficient and highly **parallelizable** algorithm $\rightarrow$ **GPU-friendly**

Minimisation of $E(R, S)$

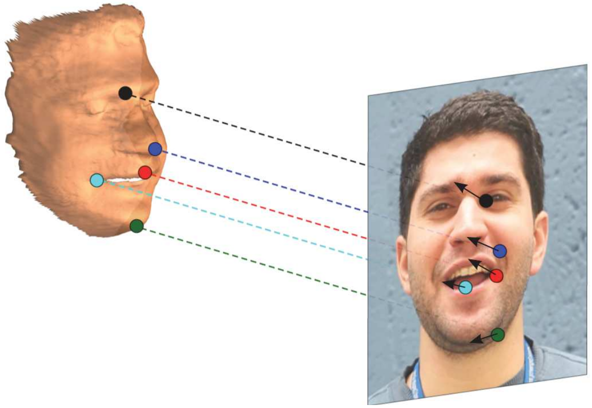


$$\min_{R} \lambda \underbrace{\|W - RS\|_{\mathcal{F}}^{2}}_{\substack{\textit{Reprojection} \\ \textit{error}}}$$

Step 1: Rotation estimation

- Robust estimation by using dense data.
- Solved via Levenberg-Marquardt algorithm.
- Rotations are parametrised as quaternions.

Minimisation of $E(R, S)$

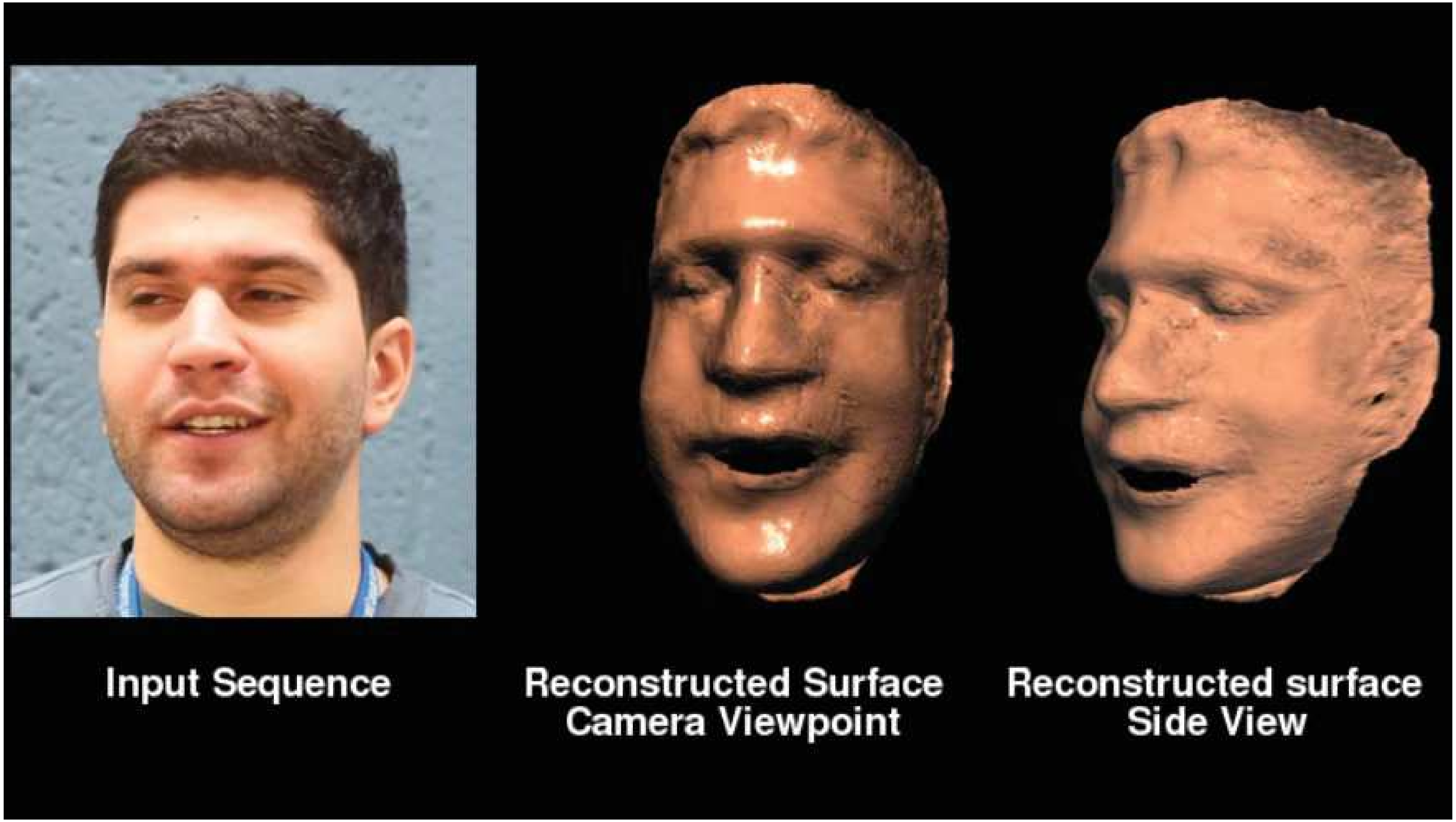


$$\min_{S} \lambda \underbrace{\|W - RS\|_{\mathcal{F}}^{2}}_{\text{Reprojection error}} + \sum_{i} \underbrace{TV(S_i)}_{\text{Smoothness prior}} + \tau \underbrace{\|S\|_{*}}_{\text{Low rank prior}}$$

Step 2: Shape estimation

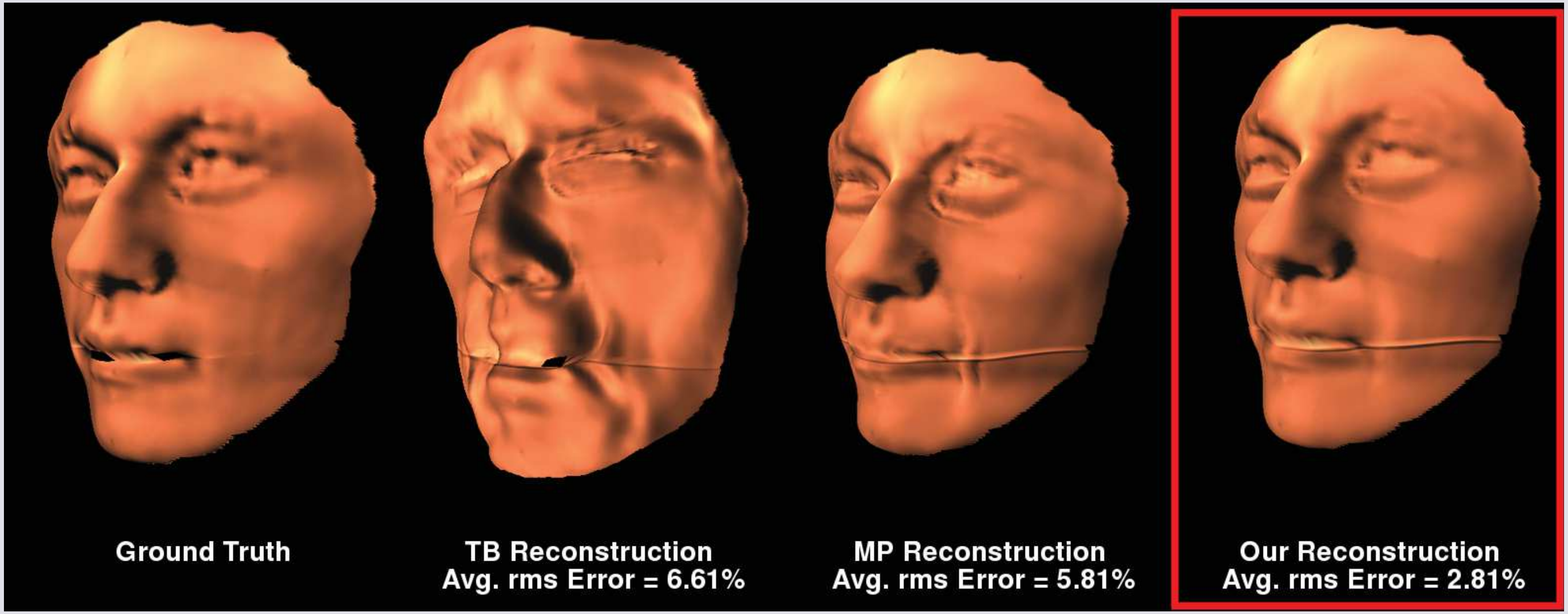
- Convex sub-problem.
- Optimisation via alternation between:
  - Per frame shape refinement: using primal dual algorithm
  - Enforcing low rank: using soft impute algorithm.

Results on real sequences



Quantitative Evaluation

Average RMS 3D reconstruction errors.



Sequence	TB	MP	Ours	Ours($\tau = 0$)
Non-smooth rotations	4.50%	5.13%	**2.60%**	3.32%
Smooth rotations	6.61%	5.81%	**2.81%**	3.89%

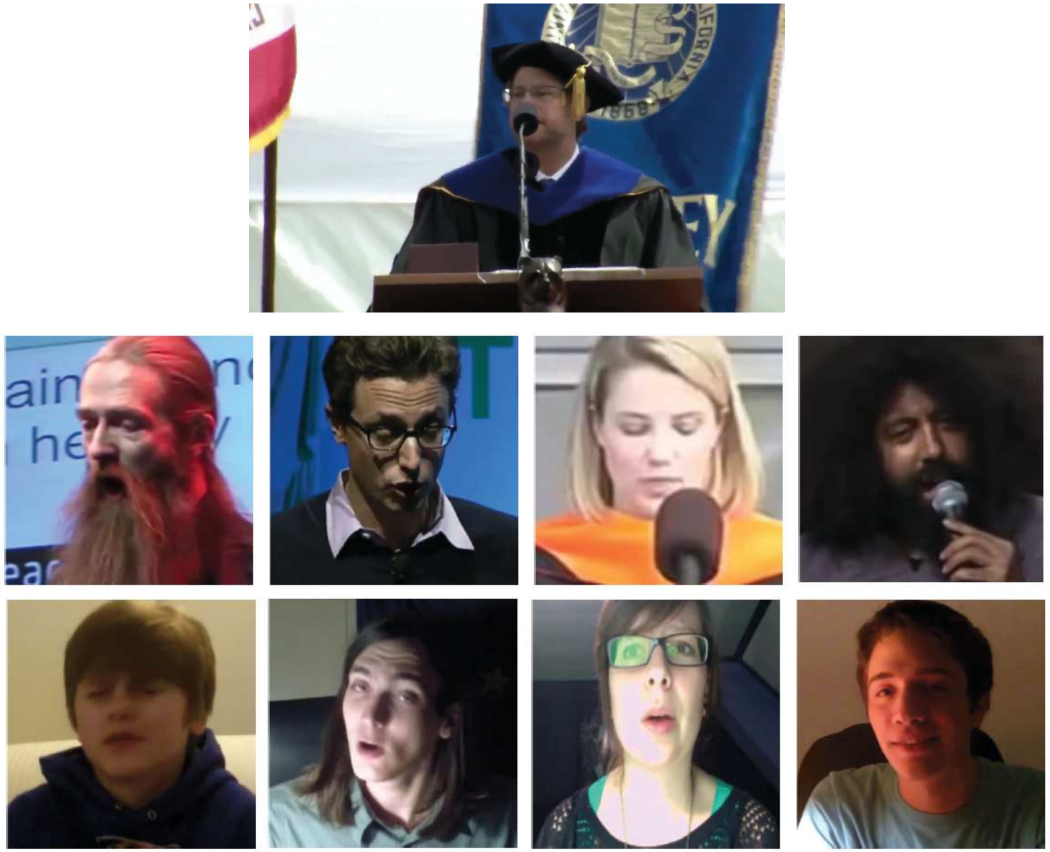
- **TB**: Akhter et al., *Trajectory space: A dual representation for NRSfM*, PAMI'11.
- **MP**: Paladini et al., *Optimal metric projections for deformable and articulated SfM*, IJCV'12.
- Synthetic data generated using (Vlasic et al., SIGGRAPH'05).

Presentation Outline

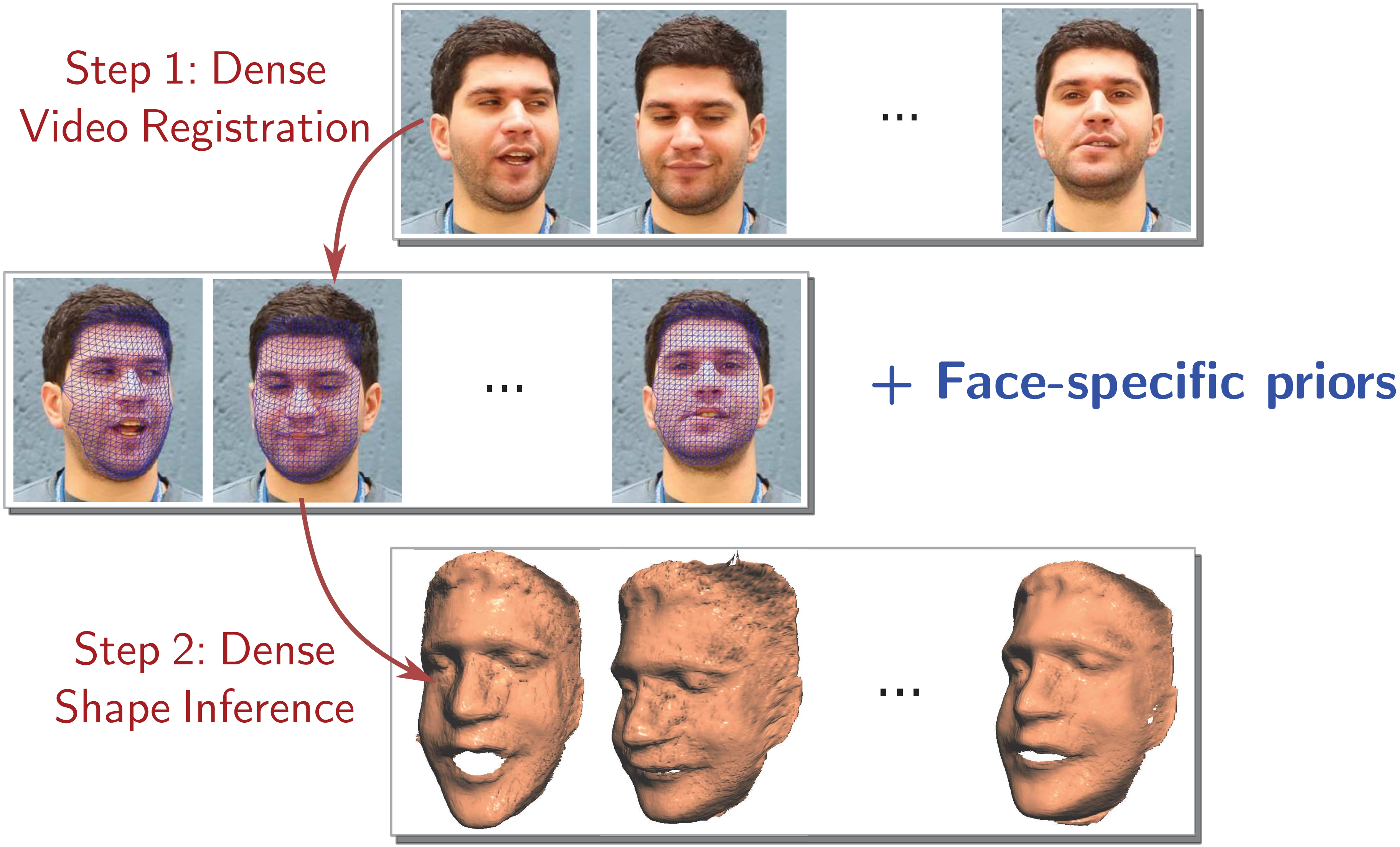
1. Introduction
2. Model-free Dense 3D Reconstruction from Videos
3. Model-based Dense 3D Reconstruction from Videos
4. Craniofacial Surgery Applications
5. Conclusions

But what about **In-the-wild** Videos?

- Addressing the challenges of **unconstrained, everyday-life videos**
- Focusing on **human faces**



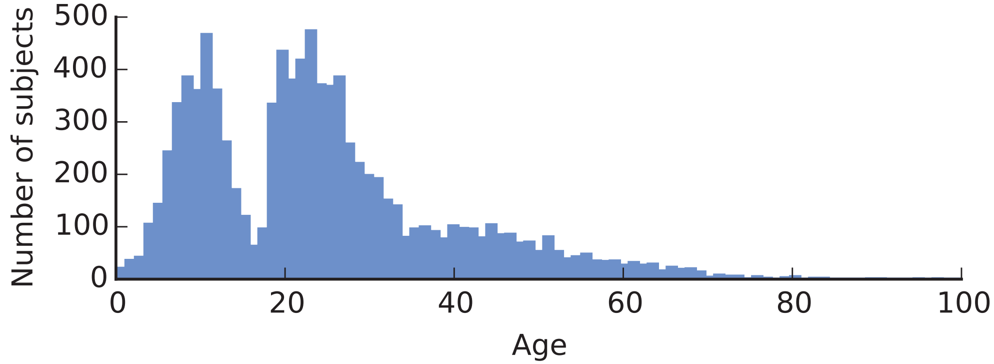
Our Pipeline



(Snape, Roussos, Panagakis, Zafeiriou, *IEEE ICCV 2015*)
(Booth, Roussos, Zafeiriou, Ponniah, Dunaway, *IEEE CVPR 2016*)
(Booth, Roussos, Ponniah, Dunaway, Zafeiriou, *IJCV 2017*, under minor revision)
(Booth, Roussos, et al., *T-PAMI 2017*, submitted)

Constructing Detailed 3D Face Models: Identity Variation

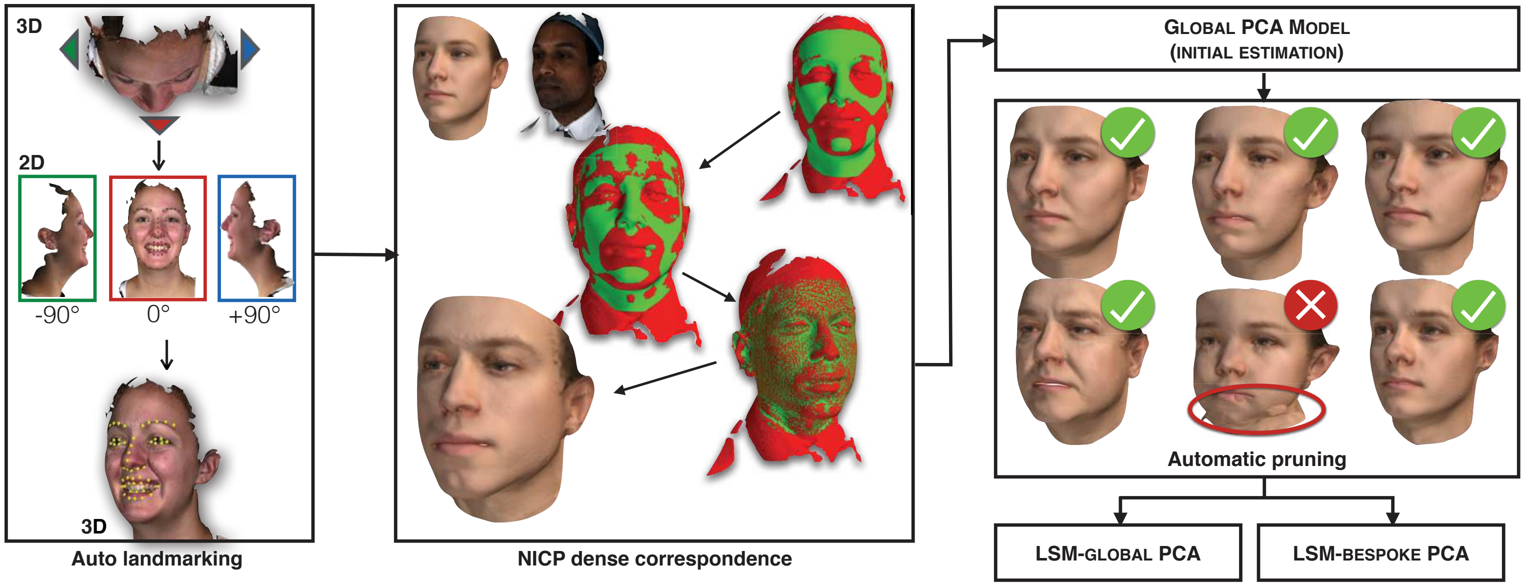
Synthetic faces generated by our **LSFM** model



- High-resolution 3D statistical model
- Automatically built from ~**10,000** 3D scans
- **Largest-scale** Morphable Model ever constructed

(Booth, Roussos, Zafeiriou, Ponniah, Dunaway, IEEE CVPR 2016)

Constructing Detailed 3D Face Models: Identity Variation



- Fully automatic pipeline
- State-of-the-art image localisation on synthetic views
- Natively 3D approach to dense mesh correspondence
- Building **global model** but also **models tailored by age/gender/ethnicity**

(Booth, Roussos, Zafeiriou, Ponniah, Dunaway, IEEE CVPR 2016)

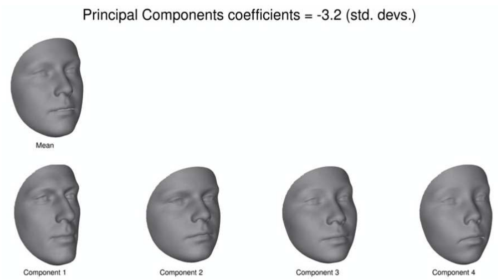
Constructing Detailed 3D Face Models: Identity Variation

Update:

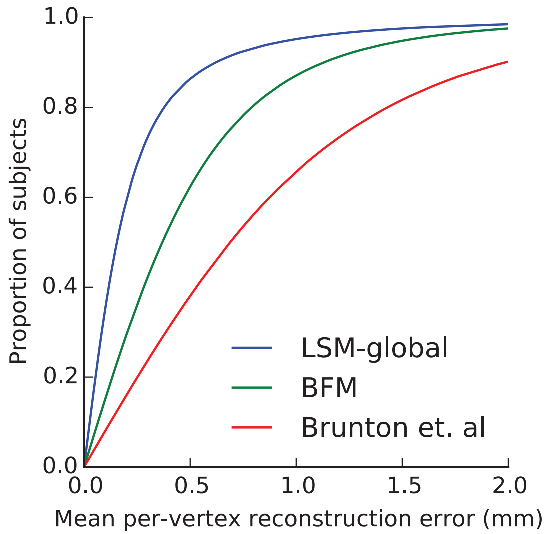
- (Booth, Roussos, Ponniah, Dunaway, Zafeiriou, *Large scale 3D Morphable Models*, IJCV, under minor revision):
  - extended evaluation
  - added texture model

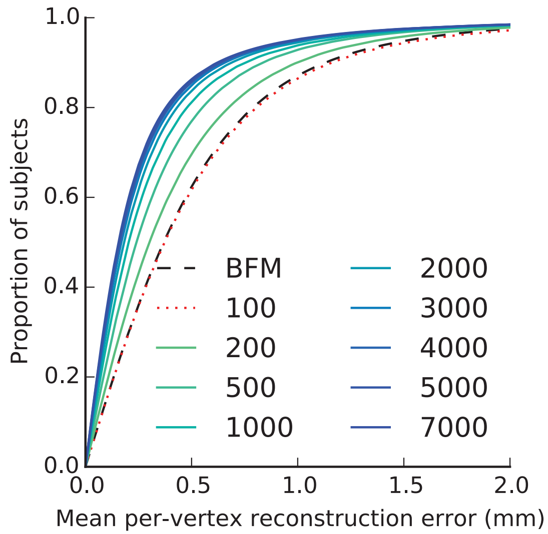
- **source code** for construction pipeline is now **available**: `https://github.com/menpo/lsfm`

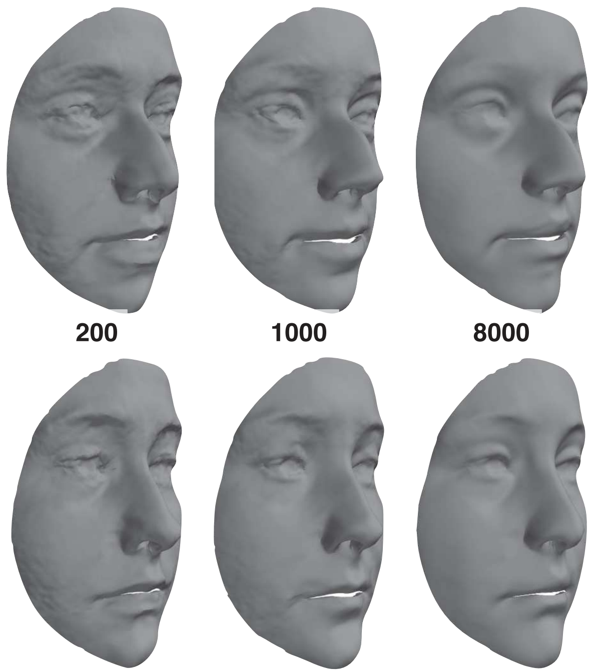
- **shape models** will be available **very soon**:



Evaluation of Model Fitting on 3D Scans







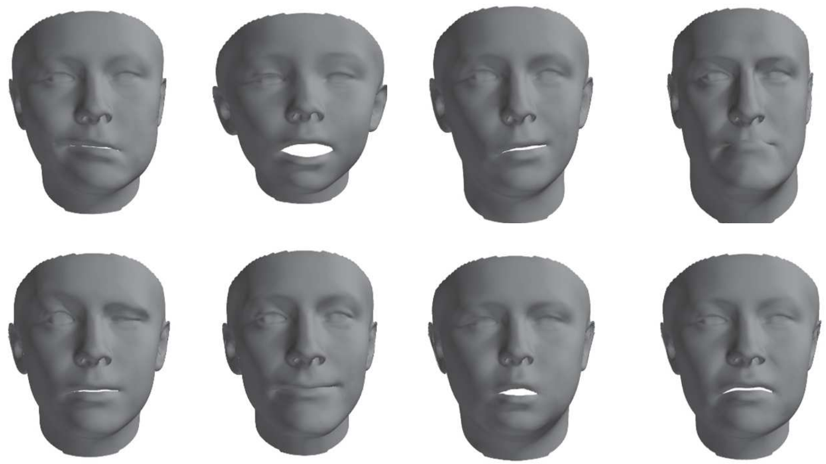
- **BFM**: Basel Face Model (Paysan et al. AVSS'09)
- **Brunton et al.**: PCA model of (Brunton et al., CVIU'14)
- **100-7000**: Proposed LSM, built with **varying size of training set** (100-7000 faces)

Adding Expression to LSFM models

Overall **model of identity & expression** by effectively combining:

- **identity variation** from our LSFM models, with
- **expression variation** from (Cao et al., IEEE T-VG 2014)

$$\mathbf{S}(\mathbf{p}_{\text{id}}, \mathbf{p}_{\text{exp}}) = \mu + \mathrm{U}_{\text{id}}\, \mathbf{p}_{\text{id}} + \mathrm{U}_{\text{exp}}\, \mathbf{p}_{\text{exp}}$$



Synthetised faces, with random identity and expression

Adding Expression to LSFM models

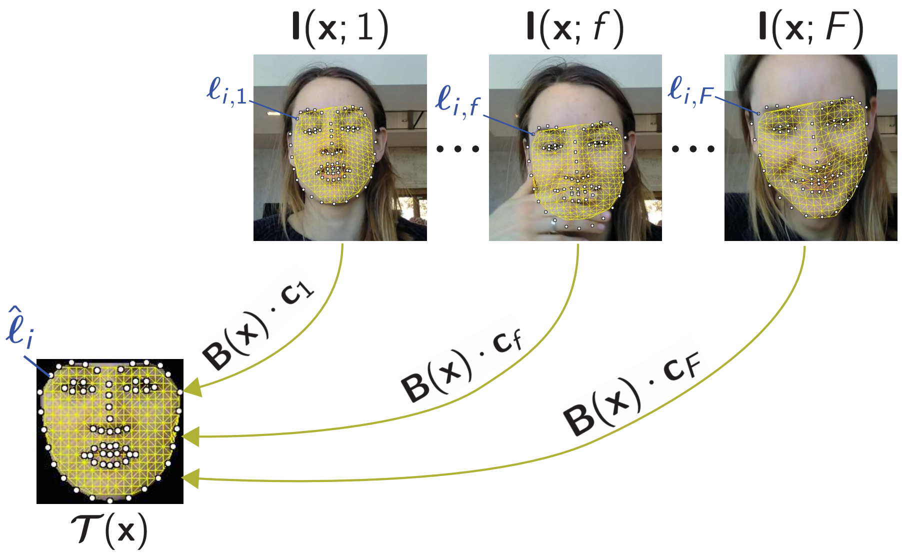
Adding Expression to LSFM models

LSFM-bespoke for *(White ; over 50 years)* with first 4 expression coefficients

Adding Expression to LSFM models

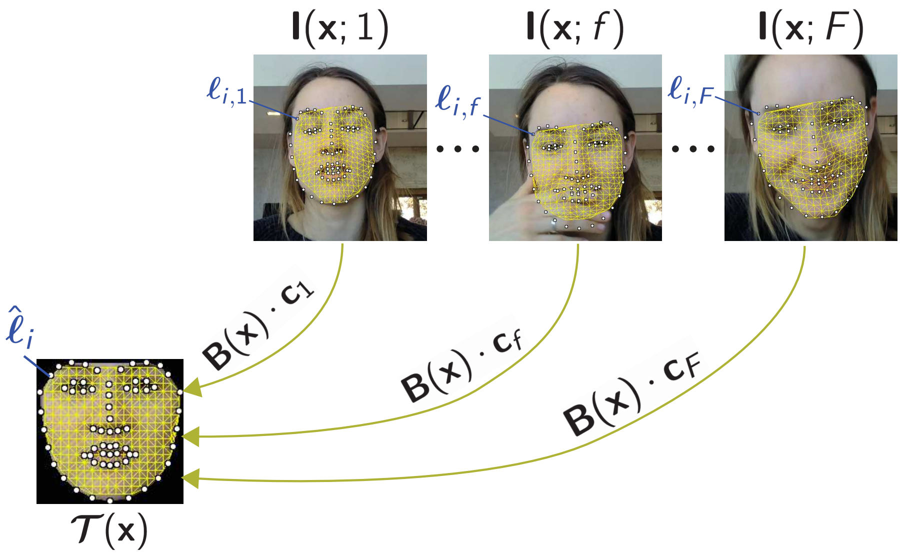
LSFM-bespoke for *(Black)* with first 4 expression coefficients

Face Flow: Face-Specific Video Registration

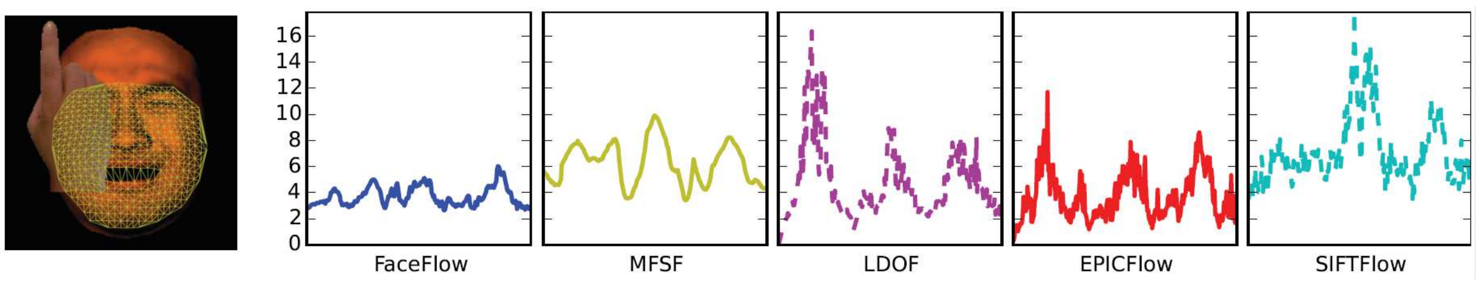


(Snape, Roussos, Panagakis, Zafeiriou, IEEE ICCV 2015)

Face Flow: Face-Specific Video Registration



- Evaluation on **synthetic videos** with **challenging conditions**:



(Snape, Roussos, Panagakis, Zafeiriou, IEEE ICCV 2015)

3DMM Fitting "In-The-Wild" (ITW)

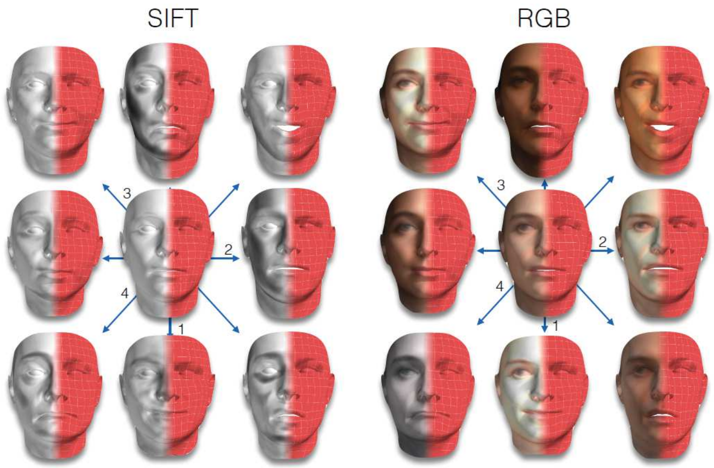
- Fitting on single images, under unconstrained conditions
- 3D shape model of identity + expression
- Texture models for in-the-wild images

3DMM Fitting “In-The-Wild” (ITW)

- Dense image features
- Simplified fitting: no need to estimate lighting
- Robust to illumination changes, occlusions, etc.

3DMM Fitting "In-The-Wild" (ITW)

- Robust PCA with missing values:



3DMM Fitting "In-The-Wild" (ITW)

- Fitting on images:

$$\underset{\mathbf{p},\mathbf{c},\boldsymbol{\lambda}}{\arg\min} \; \left\|\mathbf{F}\left(\mathcal{W}(\mathbf{p},\mathbf{c})\right) - \mathcal{T}(\boldsymbol{\lambda})\right\|^2 + c_l \left\|\mathcal{W}_l(\mathbf{p},\mathbf{c}) - \mathbf{s}_l\right\|^2 + c_s \left\|\mathbf{p}\right\|^2_{\boldsymbol{\Sigma}_s^{-1}} + c_t \left\|\boldsymbol{\lambda}\right\|^2_{\boldsymbol{\Sigma}_t^{-1}},$$

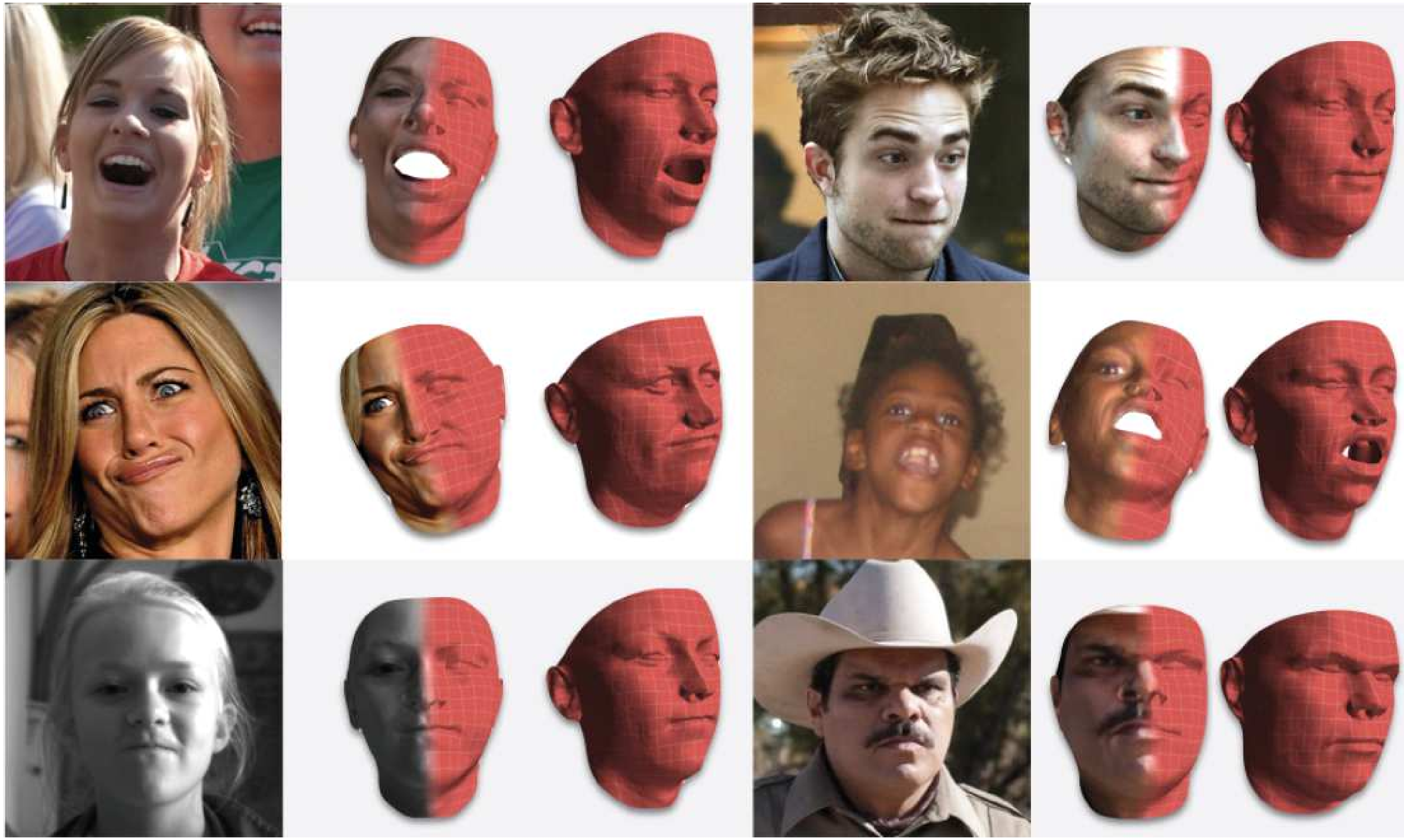
$$\mathcal{W}(\mathbf{p},\mathbf{c}) \equiv \mathcal{P}\left(\mathcal{S}(\mathbf{p}),\mathbf{c}\right) \quad , \quad \mathcal{T}(\boldsymbol{\lambda}) = \bar{\mathbf{t}} + \mathbf{U}_t\boldsymbol{\lambda}$$

- Fast algorithm, AAM-style

- Source code will be available

3DMM Fitting “In-The-Wild” (ITW)

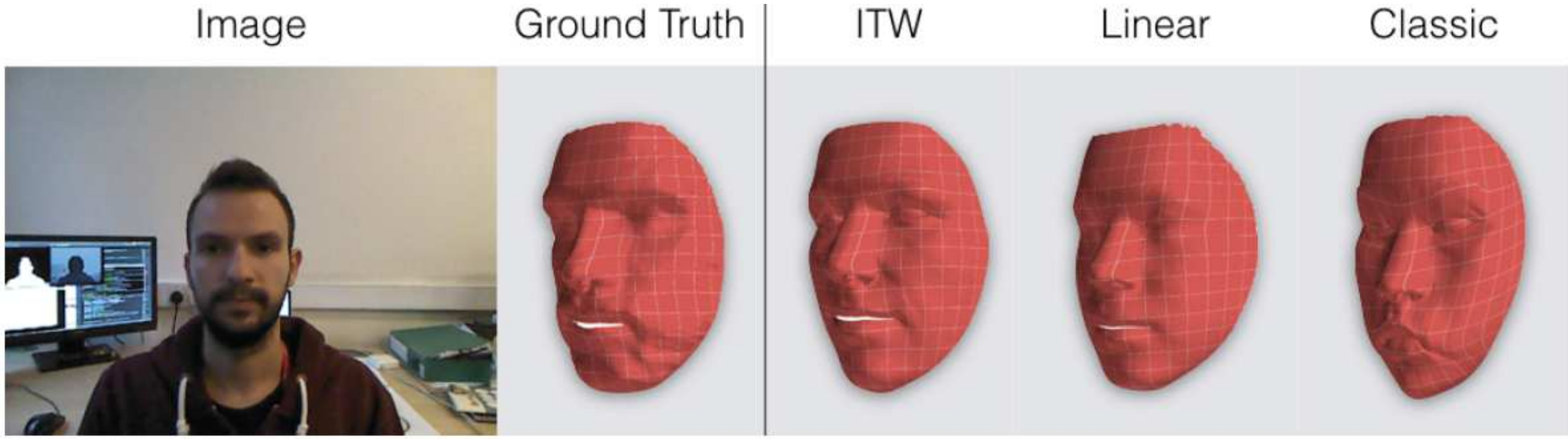
- Results on 300W:



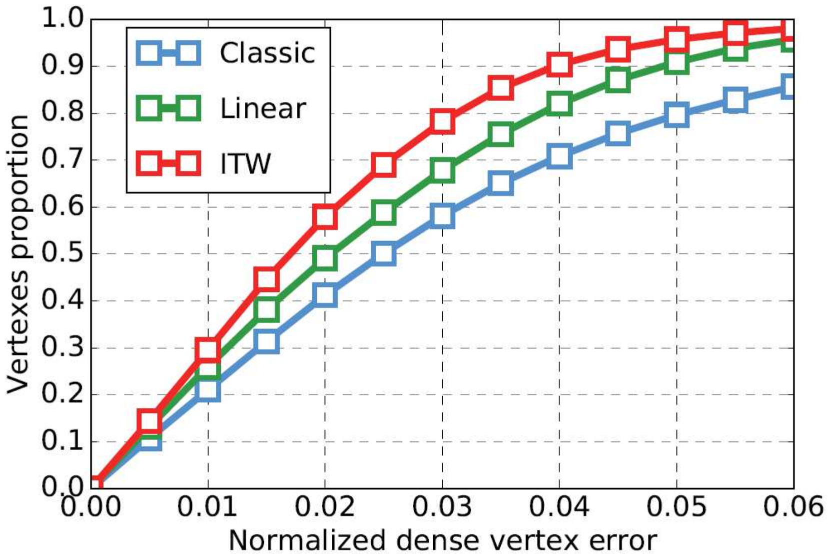
- Results on 300W:

3DMM Fitting "In-The-Wild" (ITW)

- New benchmark:



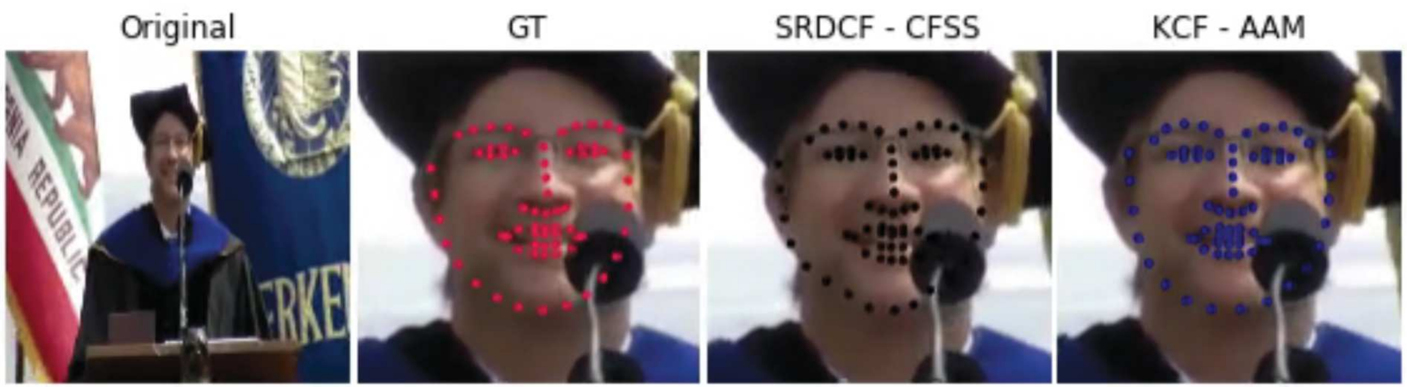
- Quantitative comparisons:



Method	*AUC*	*Failure Rate (%)*
ITW	**0.678**	**1.79**
Linear	0.615	4.02
Classic	0.531	13.9

3DMM Fitting on "In-The-Wild" Videos

- Robust facial landmark tracking
- Valuable for:
  - initialisation
  - constraints on the dense solution

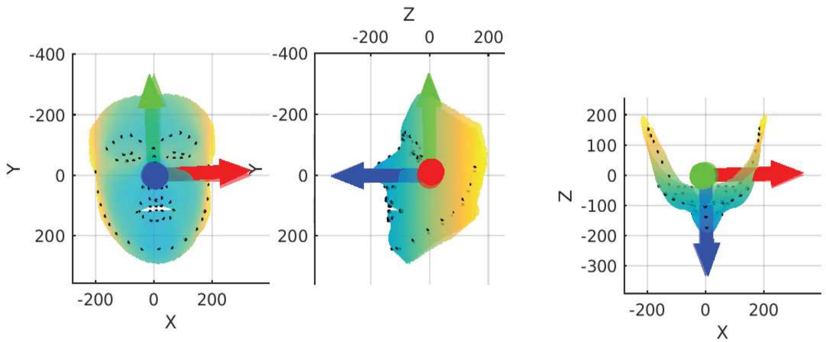


3DMM Fitting on "In-The-Wild" Videos

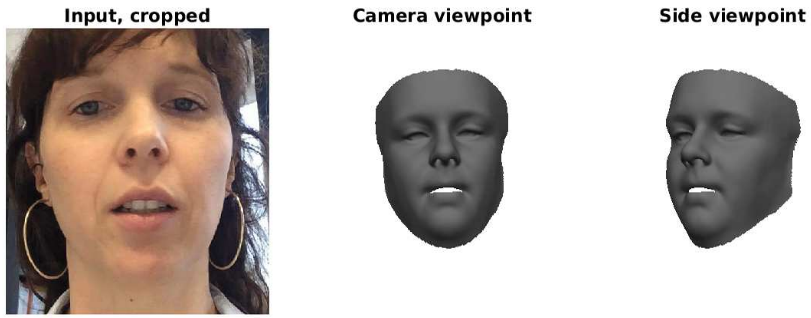
- Initialisation via fitting on the sparse tracks:
  - formulate cost function that combines:
    - reprojection error
    - temporal smoothness over expression
    - quadratic priors on identity & expression coefficients
  - minimise wrt camera, identity and expression coefficients
  - simultaneous estimation over all frames
  - automatic fine-tuning of balancing weights of the cost function

3DMM Fitting on “In-The-Wild” Videos

- Initialisation via fitting on the sparse tracks:
  - estimation of **camera parameters** via rigid **Structure from Motion**

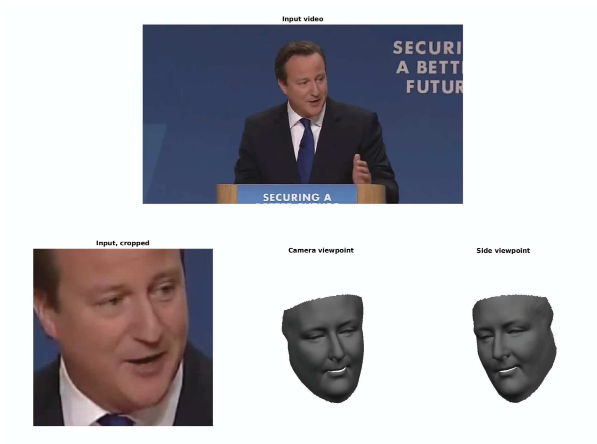


  - **large-scale quadratic optimisation** for identity & expression coefficients



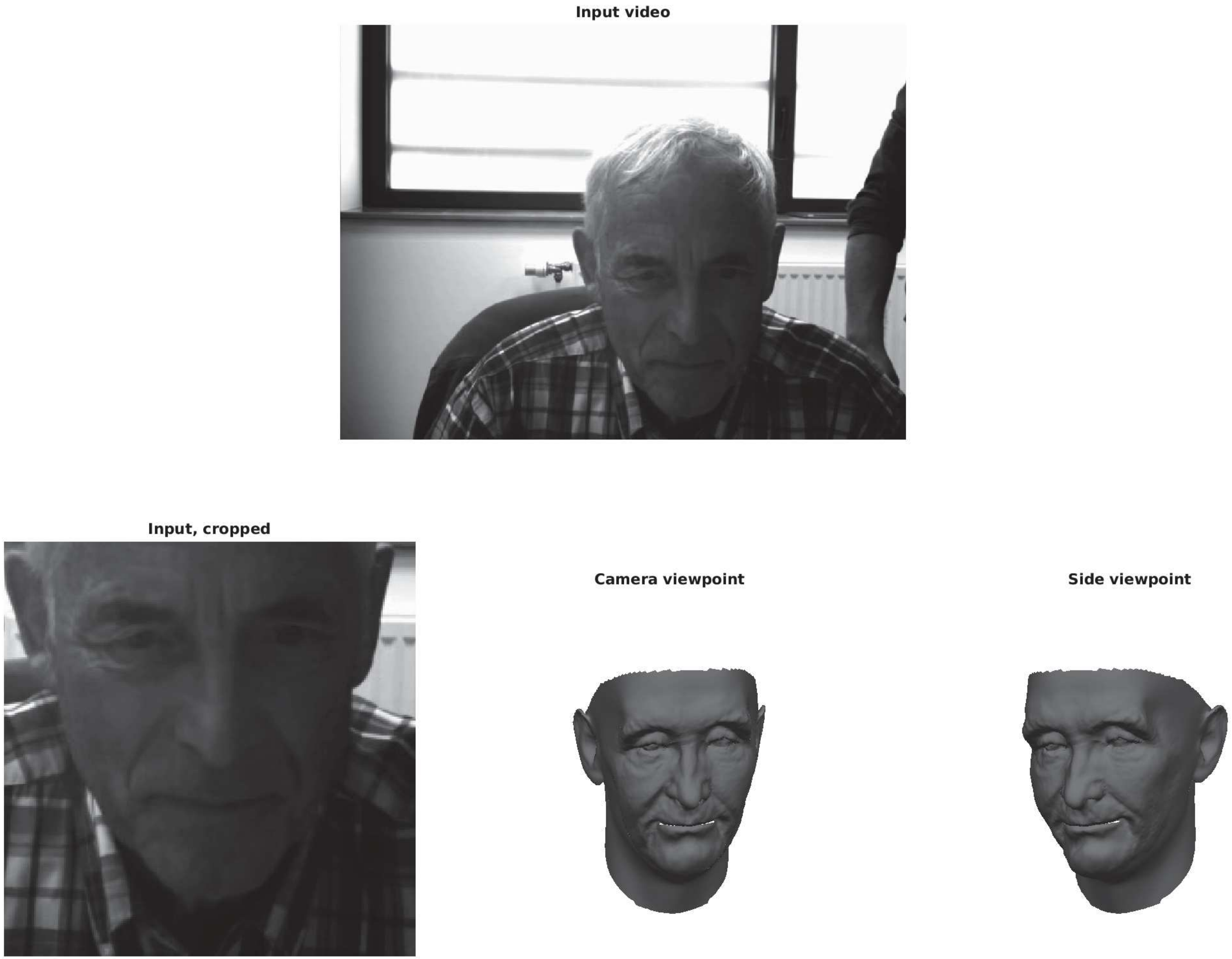
3DMM Fitting on "In-The-Wild" Videos

- Results on 300VW database:



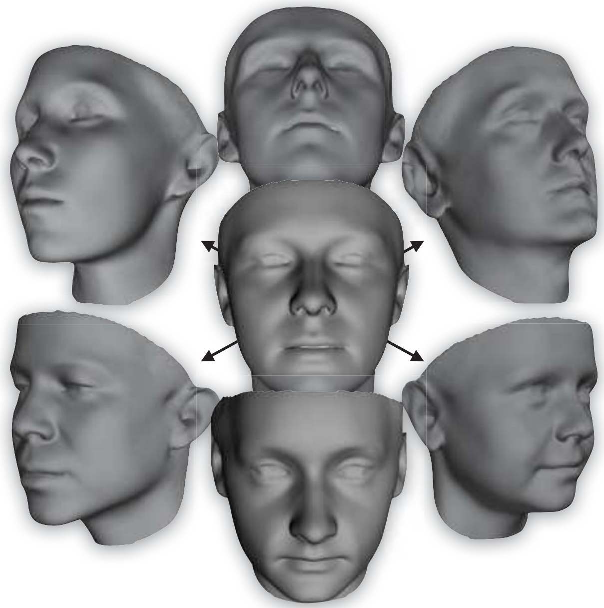
3DMM Fitting on “In-The-Wild” Videos

- Using LSFM-bespoke models:



Presentation Outline

Craniofacial Applications



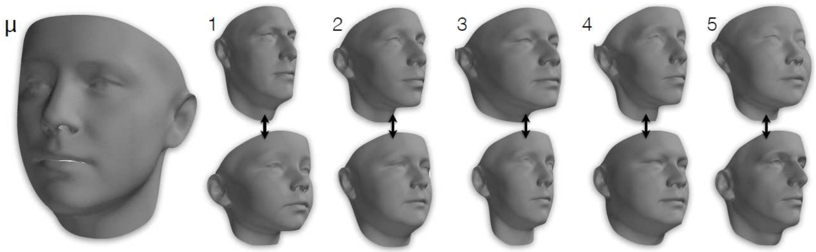
Synthetic faces generated by our **LSM** model

- Useful for: craniofacial surgery **planning and assessment**

before surgery after surgery

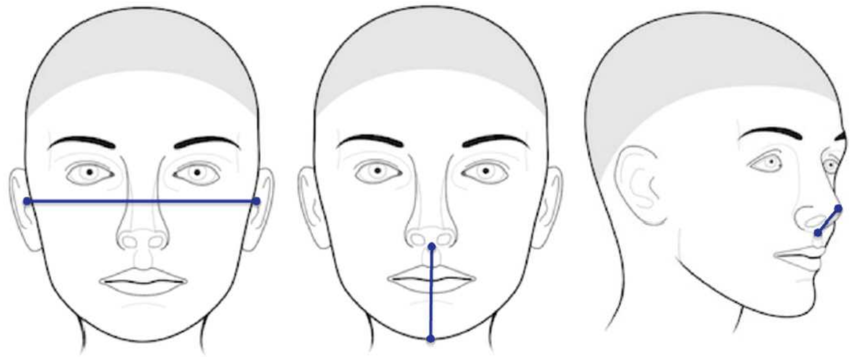
Comparing Facial Morphology Representations

- Representations of facial morphology:



$$\mathbf{S} = \mu + \sum_{i=1}^{N} \mathbf{p}_i U_i$$

dense modelling

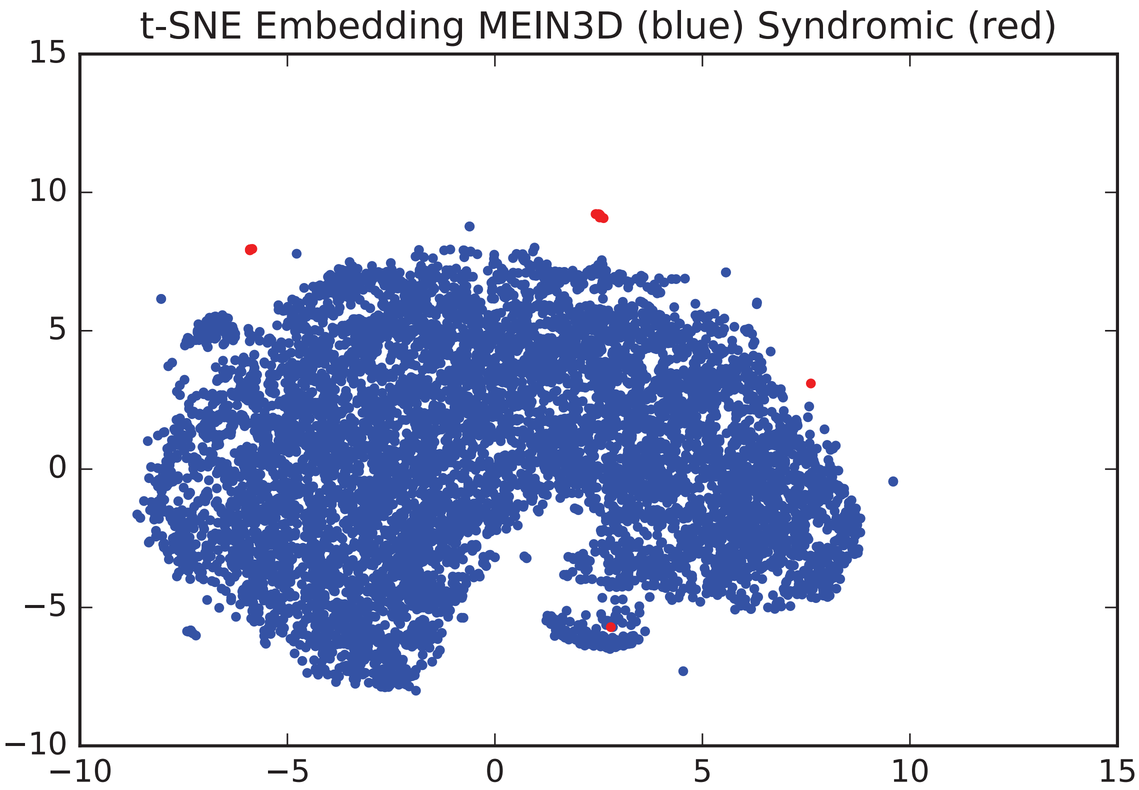


sparse anthropometry

- **Ideally**:
  - **different** shapes $\Rightarrow$ **different** parameters
  - **similar** shapes $\Rightarrow$ **similar** parameters

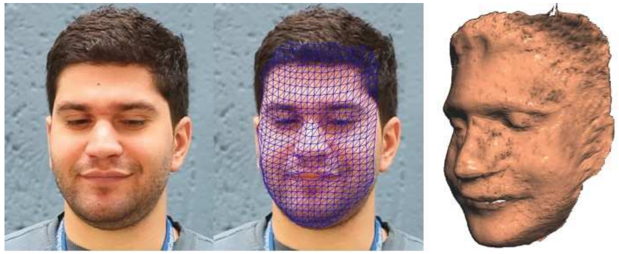
Facial Manifold Visualisation

- Including **syndromic faces**
- 46 scans of patients, including manually annotated landmarks



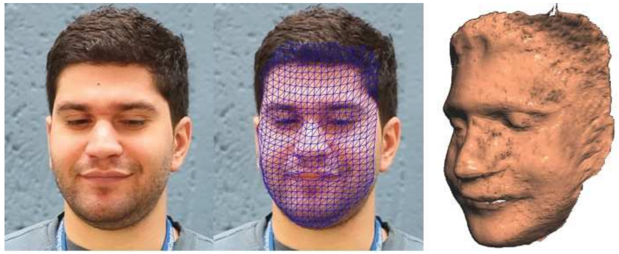
Presentation Outline

Conclusions



- **Pioneering methodologies** for dense 3D reconstruction from non-rigid videos
- Non-rigid videos contain **extremely rich information**
    - most existing methods exploit only part of it

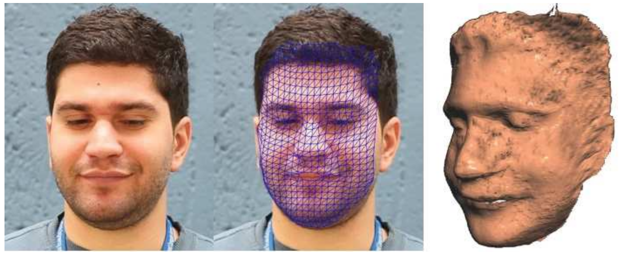
Conclusions



- **Pioneering methodologies** for dense 3D reconstruction from non-rigid videos
- Non-rigid videos contain **extremely rich information**
  - most existing methods exploit only part of it

- Using **monocular input only**, our methods yield **state-of-the-art results** on estimating:
  - multiframe optical flow
  - dense dynamic 3D shape
  - joint dense multibody segmentation, tracking and 3D reconstruction

Conclusions



- **Pioneering methodologies** for dense 3D reconstruction from non-rigid videos
- Non-rigid videos contain **extremely rich information**
  - most existing methods exploit only part of it

- Key components:
  - **dense variational** methods
  - **robust** penalisers and **low-rank** matrix priors
  - **efficient optimisation** approaches
  - **highly-detailed** and **realistic** shape priors

Conclusions

- **Dense 3D face modelling** with **unprecedented quality**
  - large-scale datasets are extremely valuable
  - fully-automated construction pipeline
  - far more diverse than existing models