Dense 3D reconstruction and modelling of faces: addressing the real-world challenges

Anastasios (Tassos) Roussos
Postdoctoral Research Associate,
Intelligent Behaviour Understanding Group (iBUG)

Joint work with:
Patrick Snape, James Booth, Yannis Panagakis, Stefanos Zafeiriou,
Allan Ponniah, David Dunaway
Ravi Garg, Lourdes Agapito

DoC RA Symposium, Imperial College London
14 June 2016

Presentation Outline

Dense 3D Reconstruction & Modelling of Faces

- Model & reconstruct the detailed 3D shape & dynamics of human faces
- Numerous applications:
  - facial expression recognition
  - face recognition
  - human-computer interaction
  - augmented reality
  - performance capture
  - craniofacial surgery

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

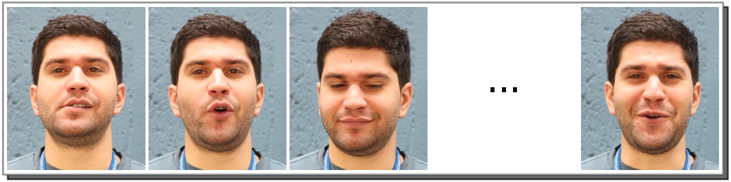
Limitations of Existing Systems

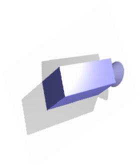
- 3D face reconstruction:
  - reliable only under **restrictive acquisition conditions**
- 3D face modelling:
  - representing **only specific demographic groups**

Presentation Outline

Dense 3D Reconstruction from Monocular Sequences

Input: **monocular** face sequence.

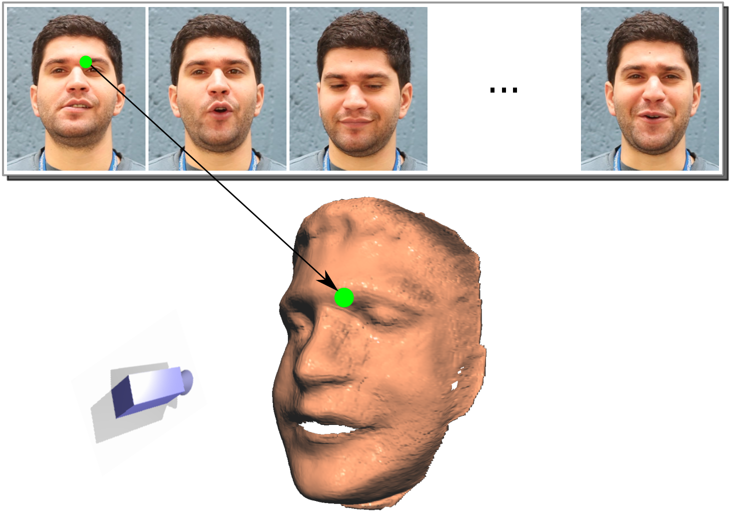




Goal: estimation of 3D location of **every pixel** at **every frame**.

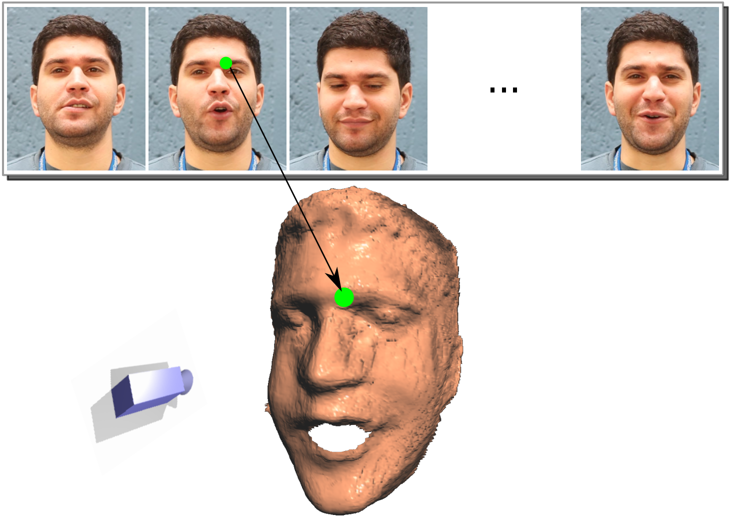
Dense 3D Reconstruction from Monocular Sequences

Input: **monocular** face sequence.



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

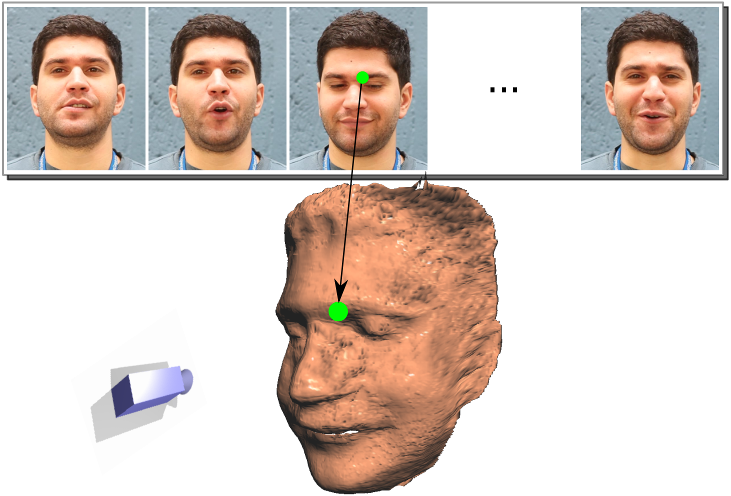
Input: **monocular** face sequence.



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

Dense 3D Reconstruction from Monocular Sequences

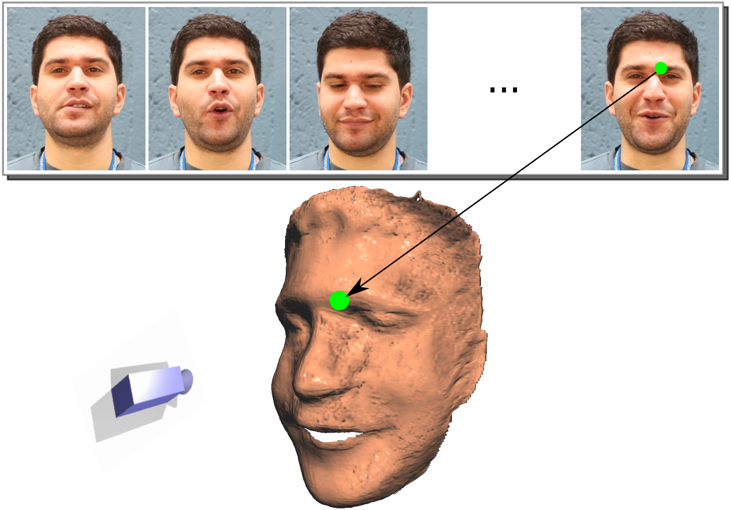
Input: **monocular** face sequence.



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

Dense 3D Reconstruction from Monocular Sequences

Input: **monocular** face sequence.

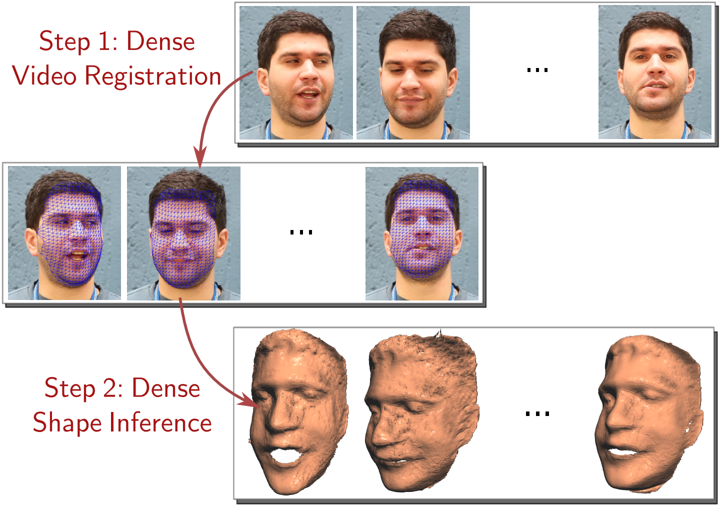


Goal: estimation of 3D location of **every pixel** at **every frame**.

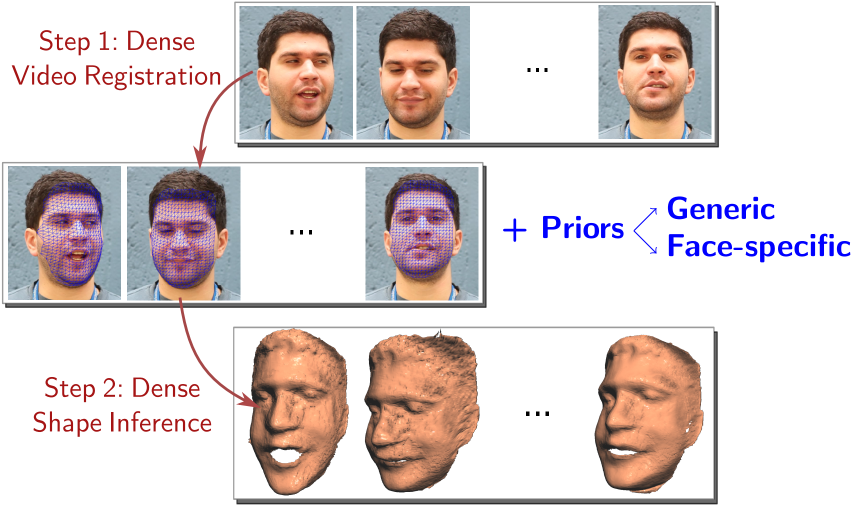
Our typical pipeline

Our typical pipeline

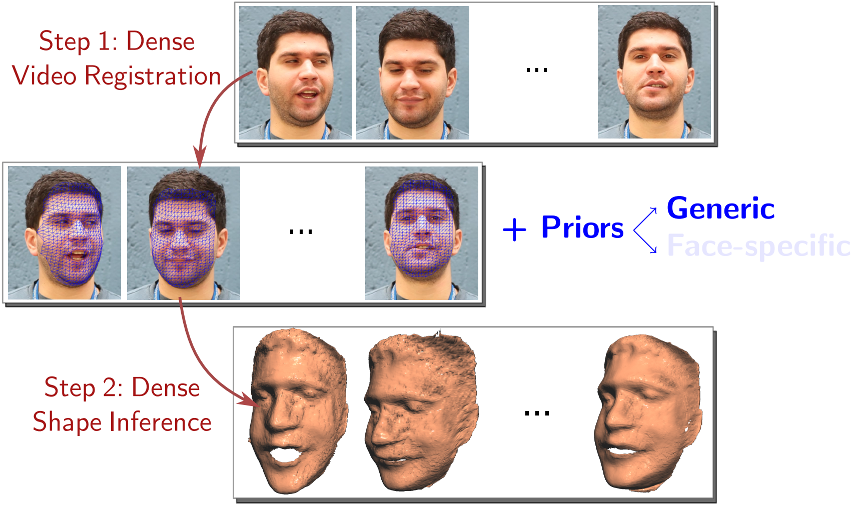
Our typical pipeline



Our typical pipeline



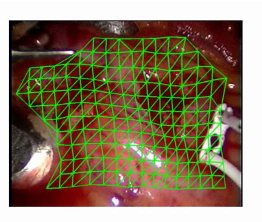
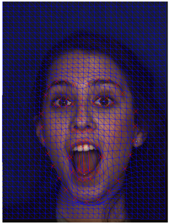
Our pipeline using Generic Priors



(Garg, Roussos, Agapito, A variational approach to video registration with subspace constraints, IJCV'13)
(Garg, Roussos, Agapito, A variational formulation for dense non rigid structure from motion, CVPR'13)

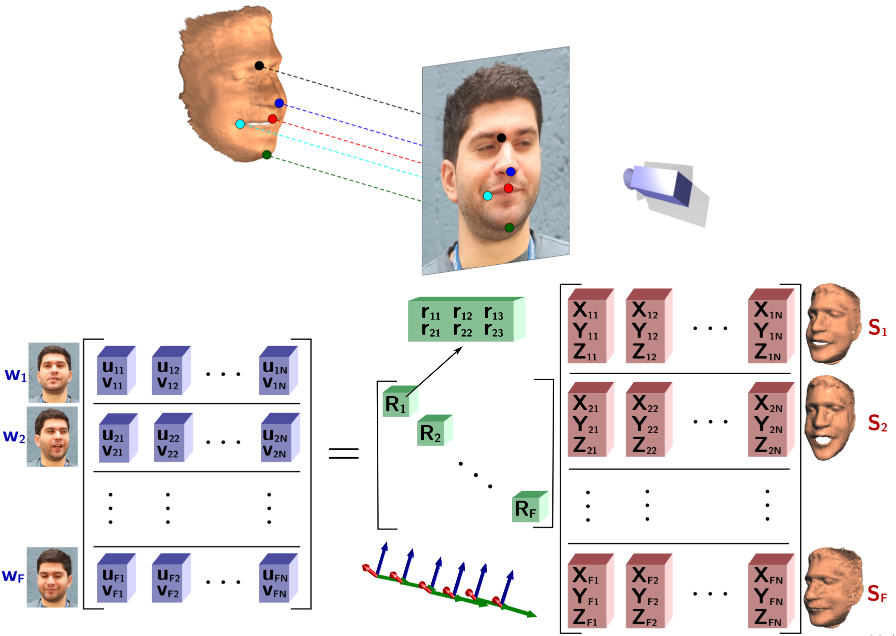
Multi-frame Subspace Flow (MFSF)

- Robust Subspace Constraints for Video Registration

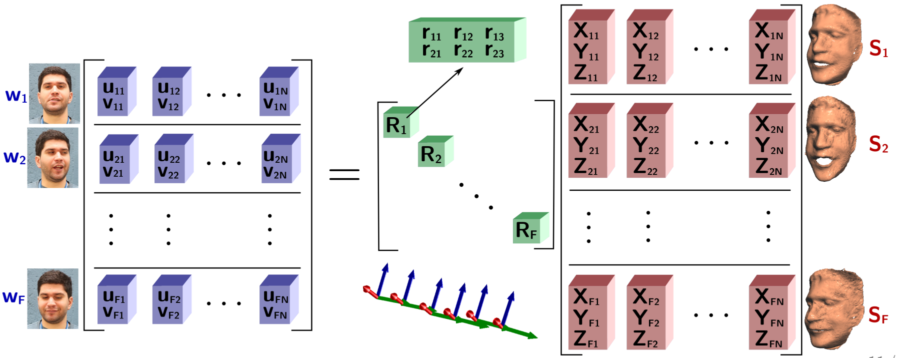
- **The code is now publicly available at**: `https://bitbucket.org/troussos/mfsf`

Orthographic Projection Model

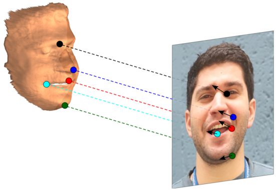


Orthographic Projection Model

$$\mathbf{W} = \mathbf{R}\mathbf{S}$$

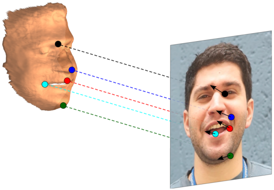


Variational Formulation



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

Variational Formulation

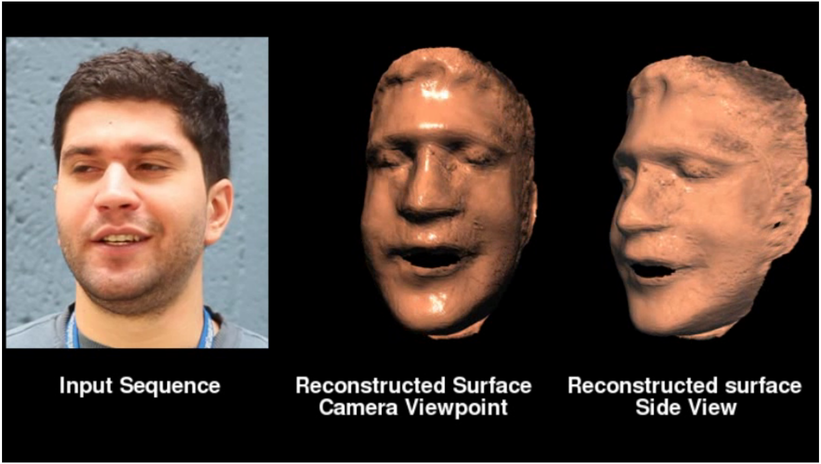


$$\min_{\mathbf{R},\mathbf{S}} \lambda \underbrace{\|\mathbf{W} - \mathbf{R}\mathbf{S}\|^2_{\mathcal{F}}}_{\text{Reprojection error}} + \sum_i \underbrace{TV(S_i)}_{\text{Smoothness prior}} + \tau \underbrace{\|\mathbf{S}\|_*}_{\text{Low rank prior}}$$

Our Algorithm

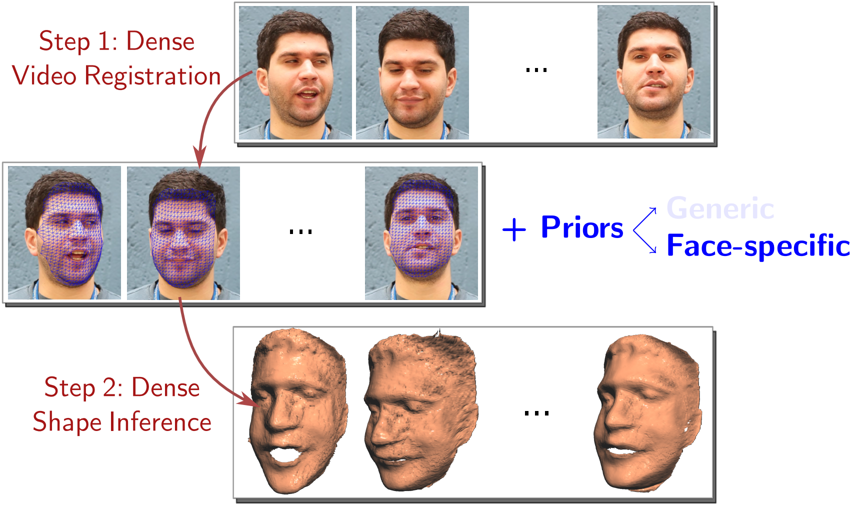
- **Initialize R** and **S** using rigid factorisation.
- Minimize energy via **alternation**:
  - Step 1: Rotation estimation.
  - Step 2: Shape estimation.
- Efficient and highly **parallelizable** algorithm → **GPU-friendly**

Result on a Real Sequence



But what about In-the-wild Videos?

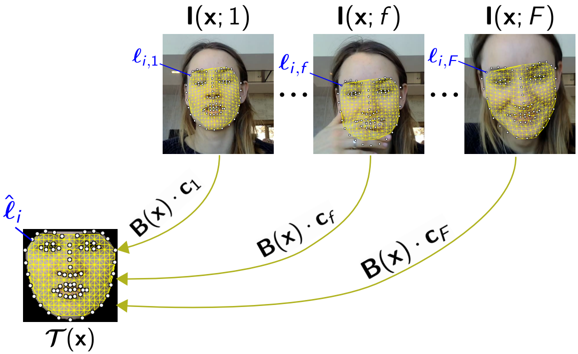
Our pipeline using Face-specific Priors



(Snape, Roussos, Panagakis, Zafeiriou, “Face Flow”, ICCV’15)

Presentation Outline

Face Flow Energy Formulation



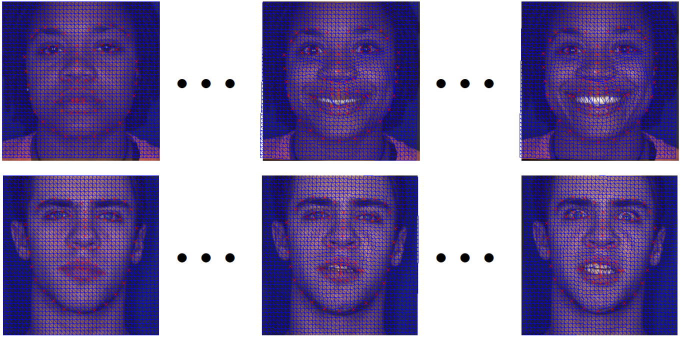
$$\mathrm{C} = [\mathbf{c}_1 \cdots \mathbf{c}_f \cdots \mathbf{c}_F] = \begin{bmatrix} \mathrm{C}_s \\ \mathrm{C}_{nr} \end{bmatrix} \begin{array}{l} \} \text{ similarity (first 4 rows)} \\ \} \text{ non-rigid deformations} \end{array}$$

Minimise w.r.t. C:

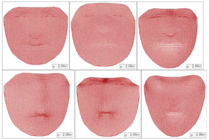
$$\underbrace{\sum_{f=1}^{F} \int_{M} \|\mathcal{T}(\mathbf{x}) - \mathbf{I}\left(\mathbf{B}(\mathbf{x}) \cdot \mathbf{c}_f \,;\, f\right)\|^2 \,\mathrm{d}\mathbf{x}}_{\textit{image data term}} + \beta \underbrace{\sum_{f=1}^{F} \sum_{i=1}^{L} \left\| \mathbf{B}(\hat{\ell}_i) \cdot \mathbf{c}_f - \ell_{i,f} \right\|^2}_{\textit{landmarks term}}, \text{ s.t. } \underbrace{\operatorname{rank}(\mathrm{C}_{nr}) \leq \lambda}_{\textit{temporal coherency}}$$

Learning the Deformation Basis

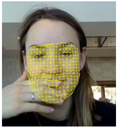
- Data-driven MFSF on BU4D videos, using landmarks constraint:



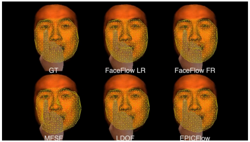
- Computed face deformation basis $\mathbf{B}(\mathbf{x})$:



- Real sequence:



- Benchmark sequences:

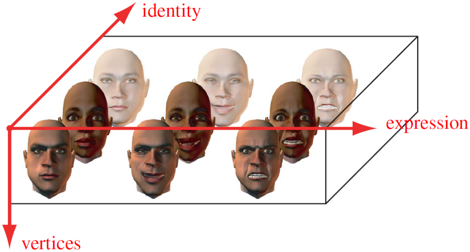


	Original		Illum.		Ilum.+Occ.	
	RMSE	AE95	RMSE	AE95	RMSE	AE95
Face Flow Low-Rank	2.95	5.52	**3.56**	**6.63**	**4.48**	**8.47**
Face Flow Full-Rank	3.24	6.01	3.76	7.02	5.83	11.50
MFSF	1.73	3.20	6.33	13.68	8.25	17.30
LDOF	**1.56**	**2.79**	4.84	9.98	6.54	11.44
EPICFlow	1.66	3.25	4.02	9.61	5.15	11.61
SIFTFlow	2.65	5.15	4.89	11.81	11.82	23.05

- LDOF: Large Displacement Optical Flow (Brox and Malik, T-PAMI 2011)
- EPICFlow: Edge-Preserving Interpolation of Correspondences for Optical Flow (Revaud et al., CVPR 2015)
- SIFTFlow: Sift flow (Liu et al, T-PAMI 2011)

Presentation Outline

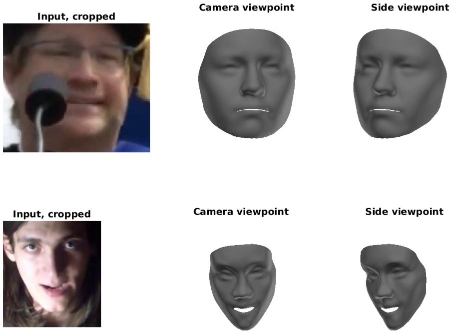
- Bilinear face model:



$$\mathbf{f} \approx \bar{\mathbf{f}} + \mathcal{M} \times_2 \mathbf{w}_2^T \times_3 \mathbf{w}_3^T$$

(Vlasic, Brand, Pfister, Popovic, “Face transfer with multilinear models”, SIGGRAPH’06)
(Bolkart, Wuhrer, “A Groupwise Multilinear Correspondence Optimization for 3D Faces”, ICCV’15)

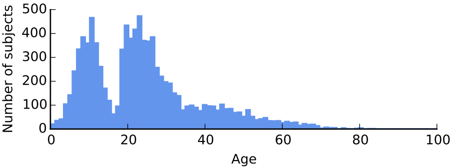
Results (Ongoing Work)



Presentation Outline

Large-Scale Facial Modelling (LSFM)

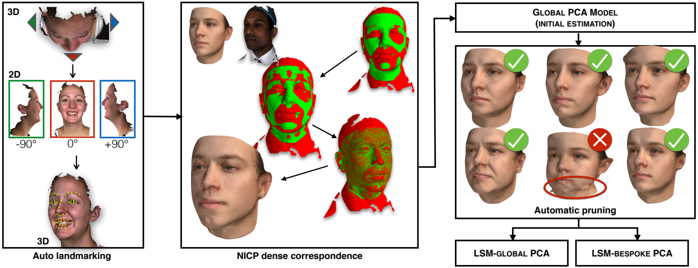
Synthetic faces generated by our **LSM** model



- High-resolution 3D statistical model
- Automatically built from of $\sim$**10,000** 3D scans
- **Largest-scale** Morphable Model ever constructed
- Model **publicly available**

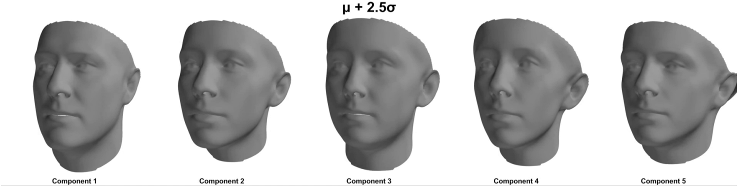
(Booth, Roussos, Zafeiriou, Ponniah, Dunaway, "A 3D Morphable Model learnt from 10,000 faces", CVPR'16)

Automatic Pipeline for Constructing our LSFM models

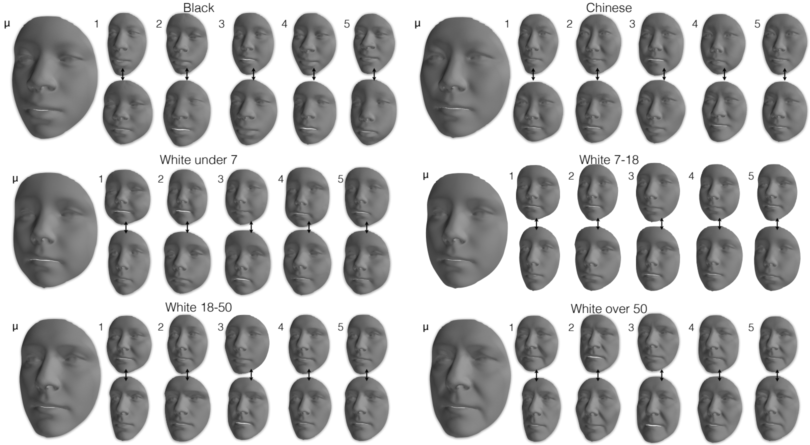


- Fully automatic
- State-of-the-art image localisation on synthetic views
- Natively 3D approach to dense mesh correspondence
- Building **global model** but also **models tailored by age/gender/ethnicity**

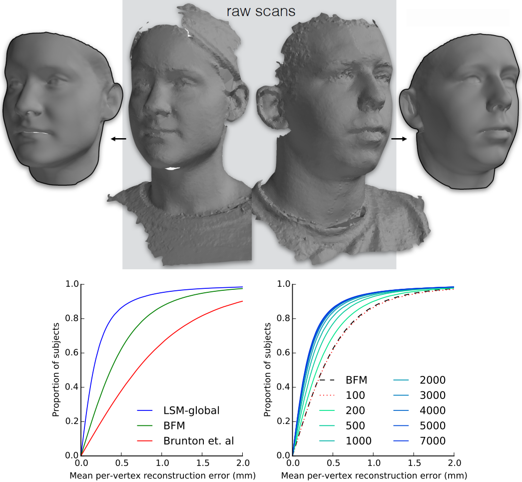
Global LSFM Model



Bespoke LSFM Models



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)

Presentation Outline

Conclusions

- **Robust** algorithms for detailed 3D reconstruction from **in-the-wild** face videos
  - dense variational methods
  - robust penalisers and low-rank matrix priors
  - efficient optimisation approaches
  - appropriate shape priors
  - state-of-the-art facial landmark localisation

Conclusions

- **Robust** algorithms for detailed 3D reconstruction from **in-the-wild** face videos
  - dense variational methods
  - robust penalisers and low-rank matrix priors
  - efficient optimisation approaches
  - appropriate shape priors
  - state-of-the-art facial landmark localisation
- **Dense 3D face modelling** with **unprecedented quality**
  - large-scale datasets are extremely valuable
  - fully-automated construction pipeline
  - far more diverse than existing models