FML: Face Model Learning from Videos

Ayush Tewari[1]   Florian Bernard[1]   Pablo Garrido[2]   Gaurav Bharaj[2]   Mohamed Elgharib[1]   Hans-Peter Seidel[1]   Patrick Pérez[3]   Michael Zollhöfer[4]   Christian Theobalt[1]

[1]Max-Planck-Institute for Informatics   [2]Technicolor   [3]Valeo.ai   [4]Stanford University

1. Goal

- 3D human face model learning from monocular videos
- Joint learning of 3D face reconstruction from monocular images
- Real-time inference at ~200fps

2. Parametric Face Representation

$$\theta =$$

- The parameters define the complete appearance of faces in 3D
- The geometry and reflectance models are learned from videos

3. Background / Challenges

- Construction of current 3D face models:
  - high-quality scans are required
  - not generalizable to in-the-wild faces
- Building a model from 2D data:
  - can capture more general faces
  - ill-posed problem
- Training on videos using multi-frame constraints

4. Pipeline

- Shared identity network to enforce consistent identity parameters across frames
- Siamese network to regress per-frame parameters
- Deformation graph to constrain the problem
- Orthogonality between the learned geometry model and the expression model
- Trained using a self-supervised loss function

Template Mesh   521 vs. 60k vertices

5. Loss Function

$$E(\mathrm{P}) = w_{land}E_{land}(\mathrm{P}) + w_{photo}E_{photo}(\mathrm{P}) + w_{reg}E_{reg}(\mathrm{P})$$

$$\| \quad - \quad \|_2 + \| \quad - \quad \|_{2,1} + \quad + \quad + \text{Stabilization}$$

Landmark Alignment   Photometric Alignment   Laplacian smoothness   Local sparsity

6. Results and Evaluations

Monocular vs. Multi-frame Reconstruction

	Ours				
Train	M = 1	M = 2	M = 4	M = 2	M = 4
Test	M = 1	M = 1	M = 1	M = 2	M = 2
Mean	1.92 mm	1.82 mm	1.76 mm	1.80 mm	**1.74 mm**
SD	0.48 mm	0.45 mm	0.44 mm	0.46 mm	**0.43 mm**

Geometric reconstruction error on the BU-3DFE dataset

Learned Models

Comparison to Tran18

Comparisons to Tewari17, Richardson17, Sela17, Tewari18