3D Face Modelling, Analysis and Synthesis

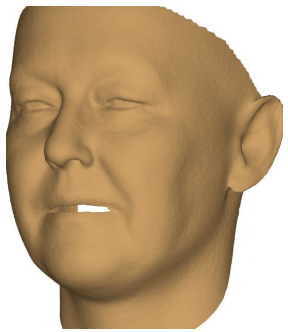
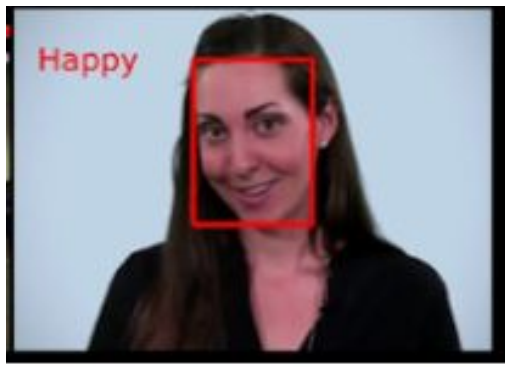
PhD Student: Mohammad Rami Koujan

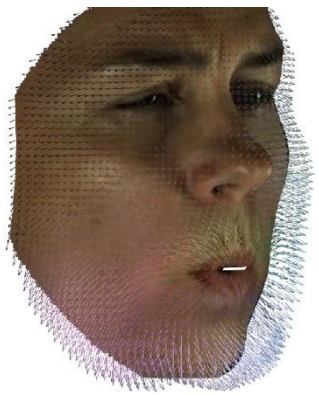
Supervisor: Dr. Anastasios (Tassos) Roussos

College of Engineering, Mathematics and Physical Sciences

University of Exeter

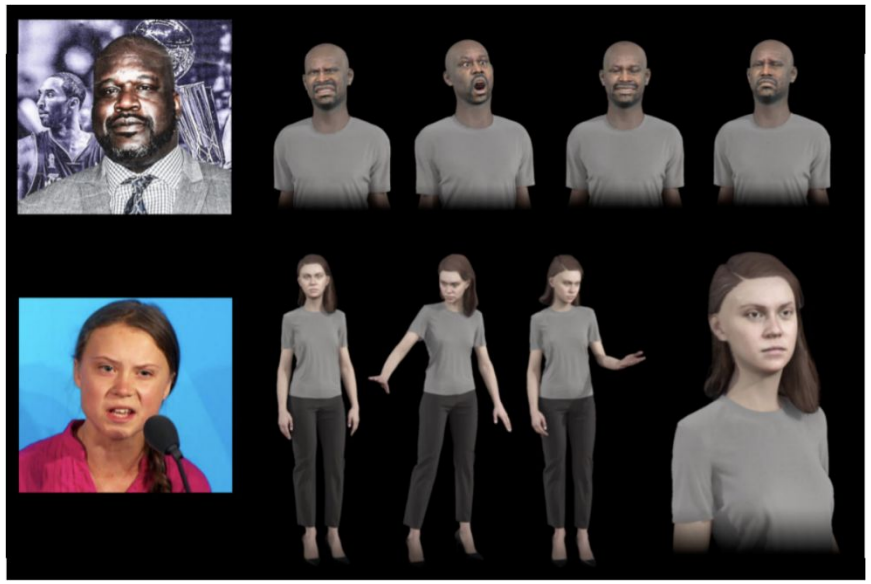
13/07/2021

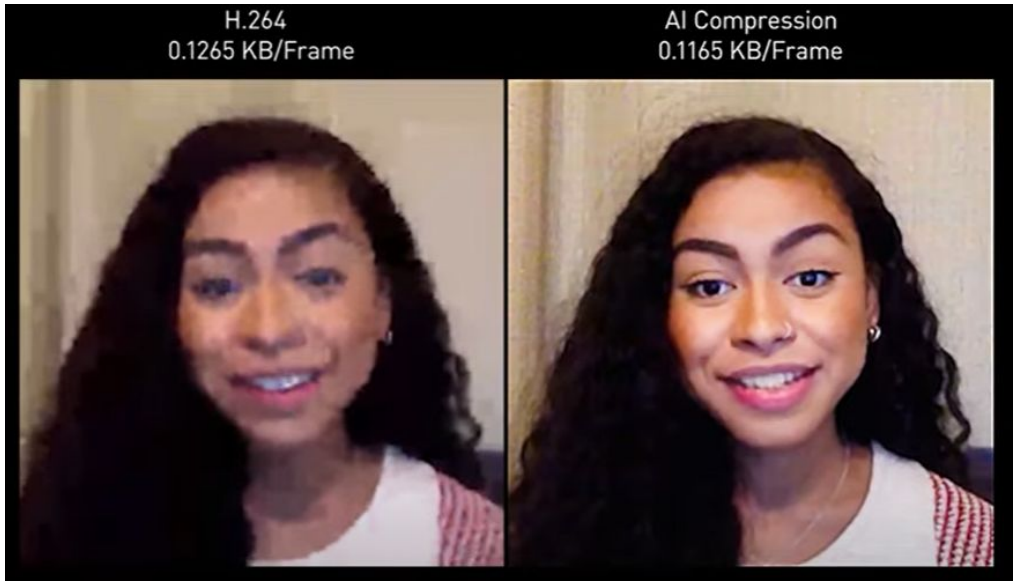









Motivation







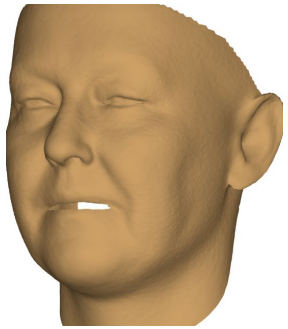
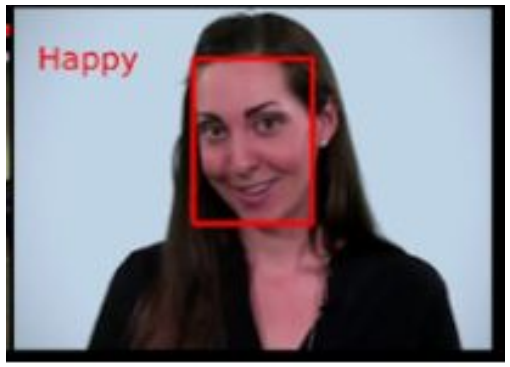


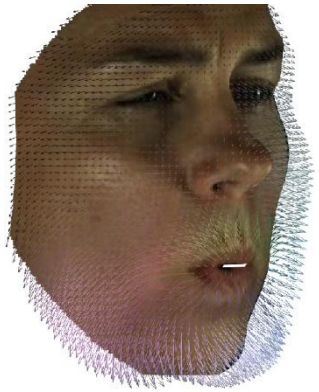
Motivation

- Automatic visual perception of human faces

- Core research question:

  How can we use consumer-grade cameras to build detailed 3D models of human faces, recognise their emotions and even their intentions, synthesise human faces realistically in scenes?






Computer Vision For Faces: Limitations

- Might need active sensors, controlled scenarios or expensive acquisition setups
- Make strong assumptions that might limit the reconstruction quality
- Mostly fail to capture fine-scale motions from monocular input
- Might not run on mobile devices in real time

Overcoming the Limitations

- Research questions:
  - *Is it possible to accurately 3D reconstruct and track human faces from in-the-wild monocular imagery data?*
  - *Can we robustly disentangle the human expression from identity under various challenging acquisition settings?*
  - *Can we capture the fine-scale 3D complex facial motions from only a pair of monocular images?*
  - *Can we reach a photo-realistic and temporally-coherent synthesis performance of human faces indistinguishable from real images?*

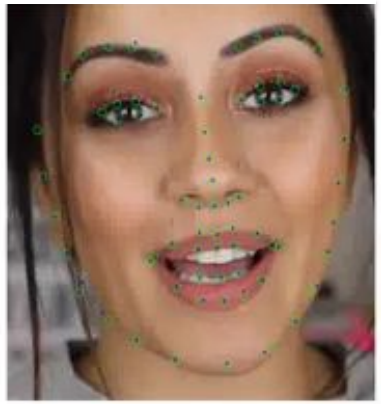
Creating Large-scale Datasets for Training

- FaceVid dataset:
  - **12,000** YouTube **videos**
  - **19M** frames in total
  - **2,500 unique identities**

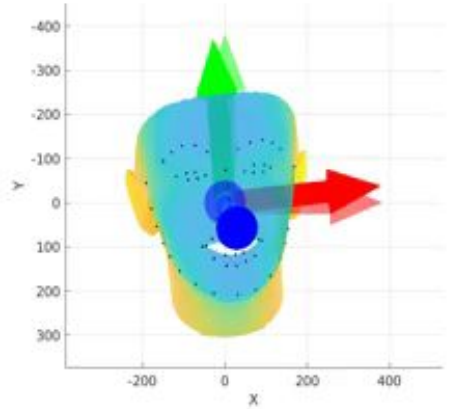
original footage



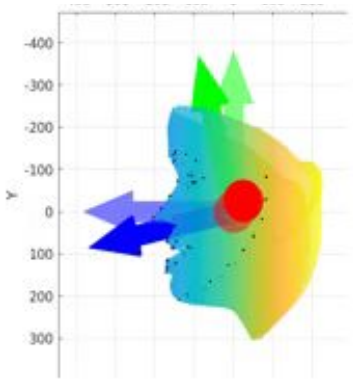
2D landmarks



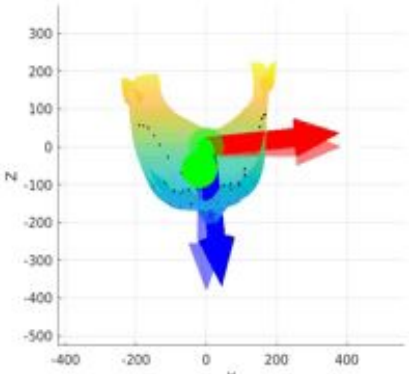
frontal view



side view



top view



Frames from a sample of the 12K videos of Face3DVid dataset

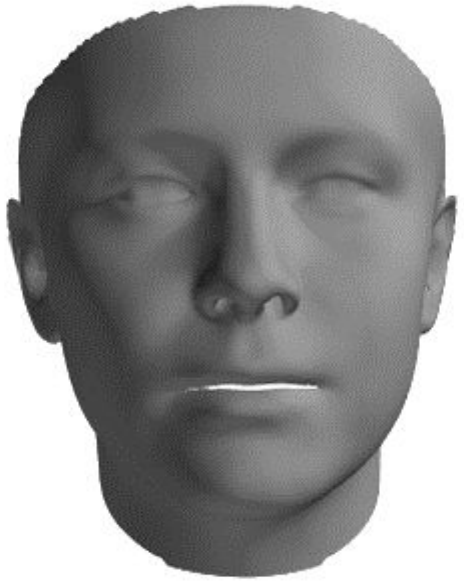
[1] **M.R. Koujan**, A. Roussos and S. Zafeiriou. **DeepFaceFlow: In-the-Wild Dense 3D Facial Motion Estimation**, CVPR'20

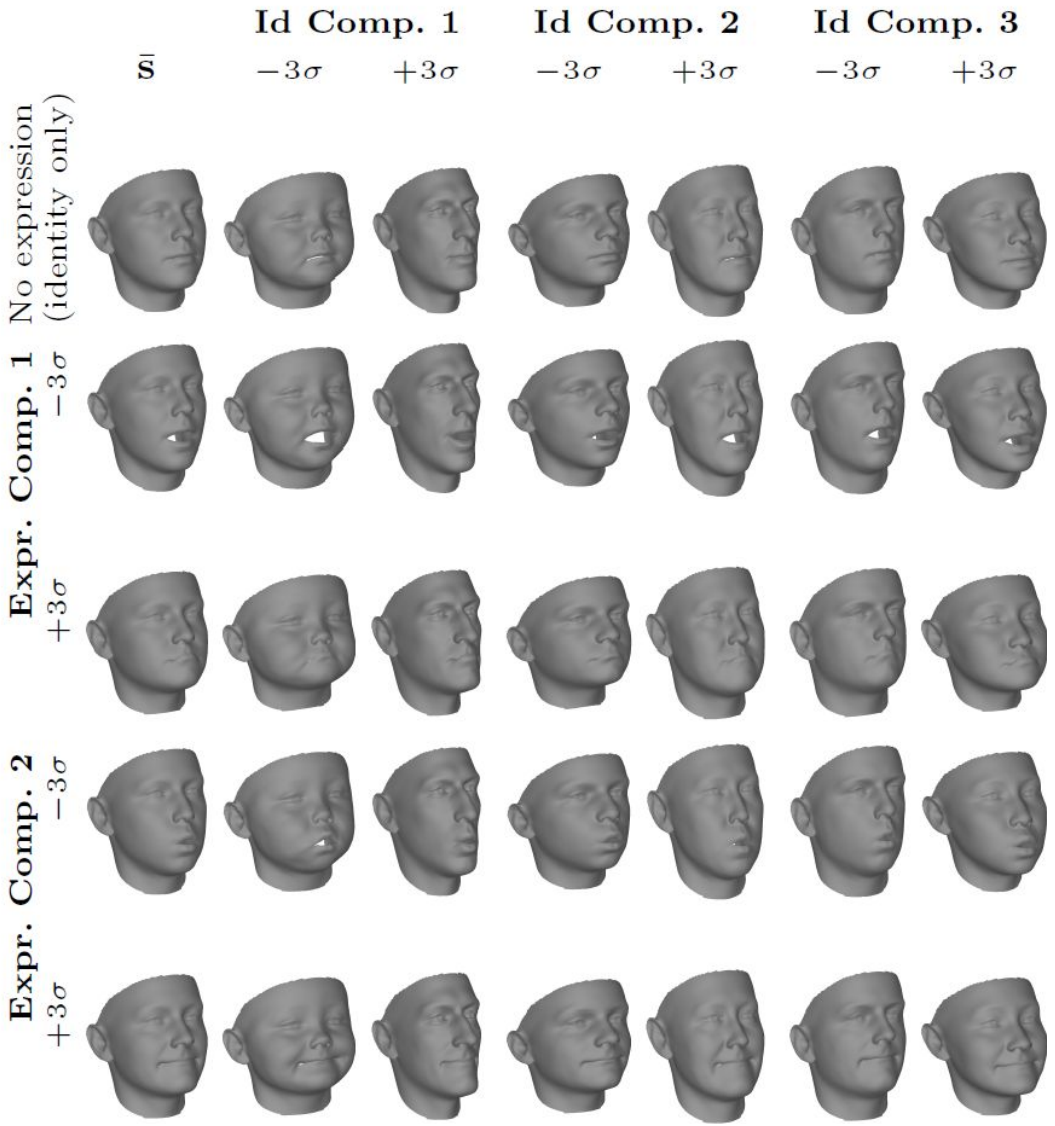
[2] **M.R. Koujan**, L. Alharbawee, G. Giannakakis, N. Pugeault and A. Roussos. **Real-time Facial Expression Recognition ``In The Wild'' by Disentangling 3D Expression from Identity**, FG'20

3DMM of Identity and Expression

- Overall **model of identity & expression** by effectively combining:
  - **identity variation** from LSFM (Booth et al., IJCV 2018) models, with
  - **expression variation** from (Cao et al., IEEE T-VG 2014)

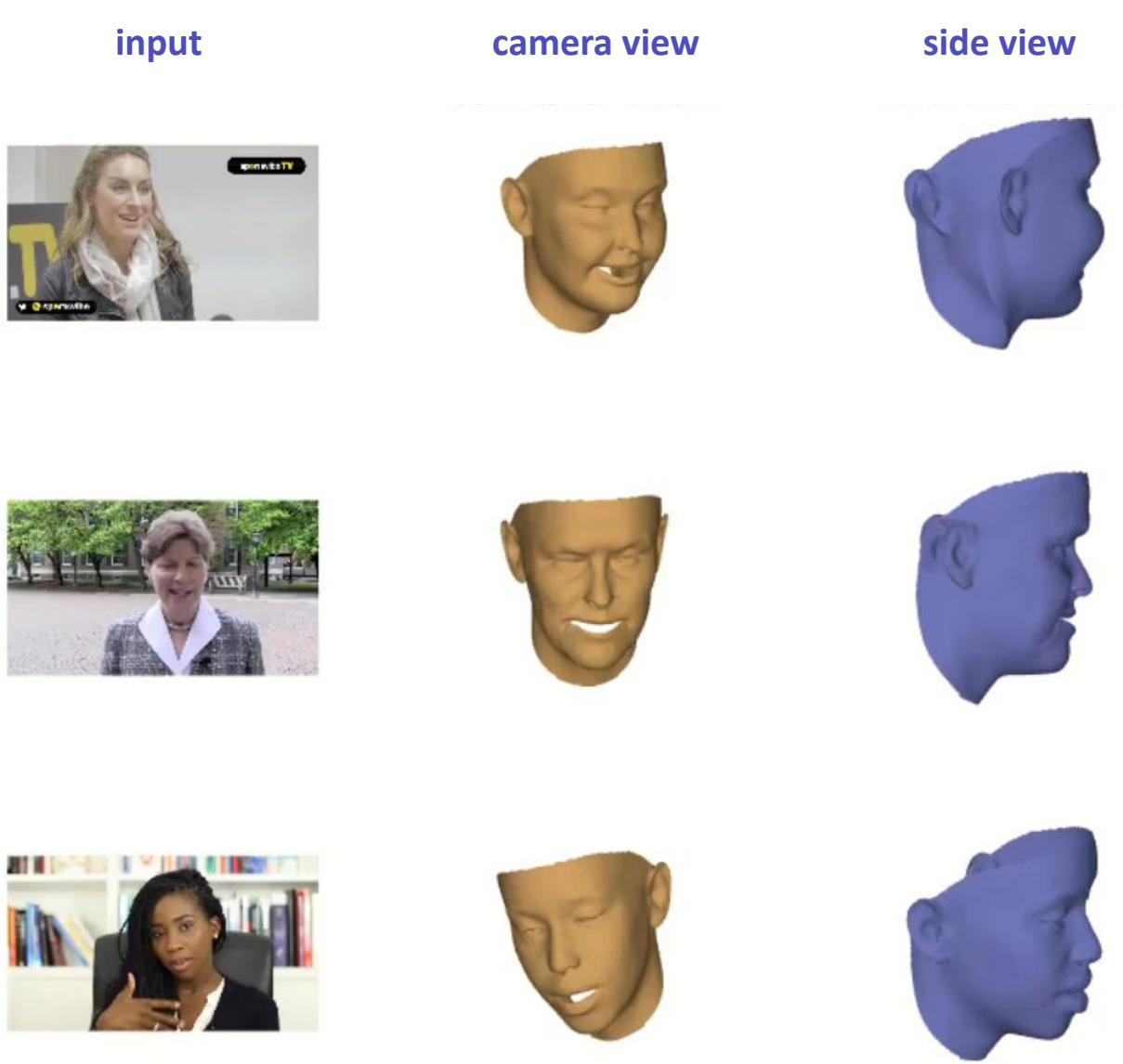
$$\mathbf{S}(\mathbf{p}_{\text{id}}, \mathbf{p}_{\text{exp}}) = \mu + \mathrm{U}_{\text{id}}\,\mathbf{p}_{\text{id}} + \mathrm{U}_{\text{exp}}\,\mathbf{p}_{\text{exp}}$$





Combining Model-based with Model-free Approaches

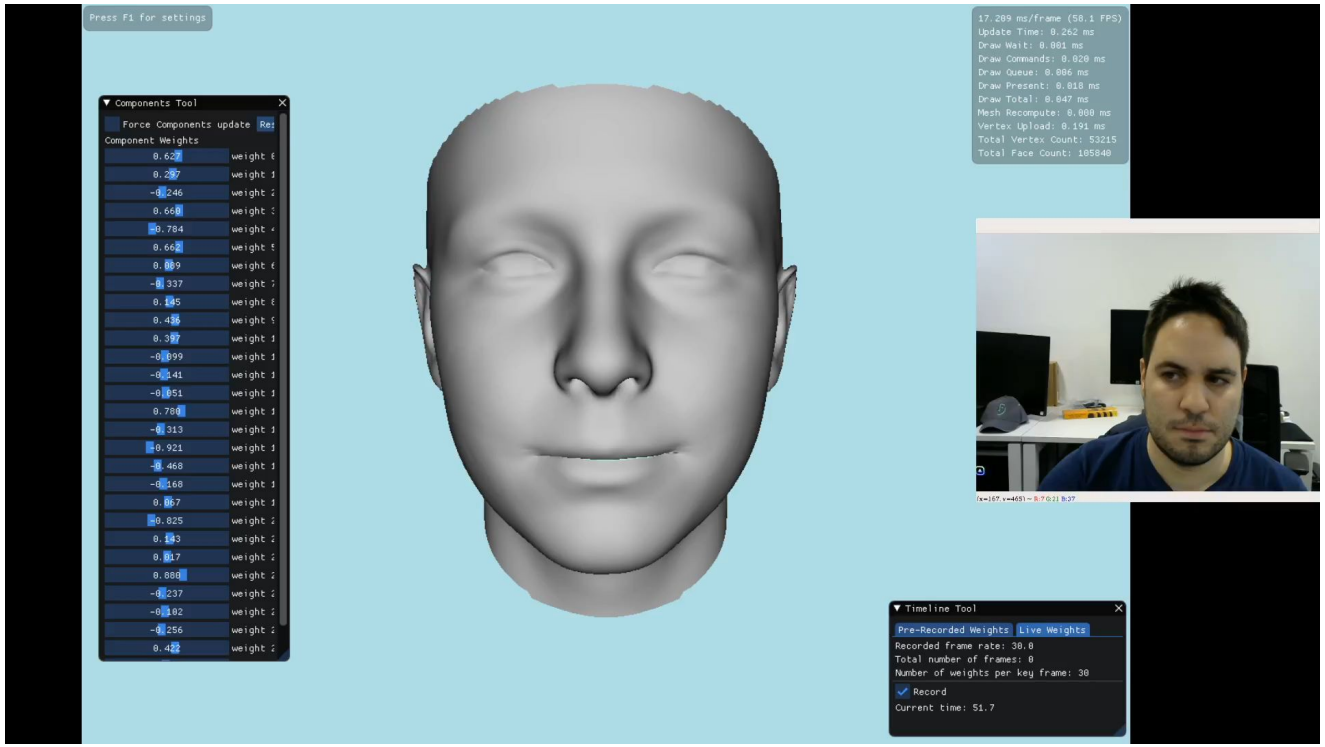
- **Highly-detailed** 4D Face Reconstruction by **combining**:
  - 3D Morphable Modelling, and
  - Dense Nonrigid Structure from Motion



(**Koujan** & Roussos, **Combining Dense Nonrigid Structure from Motion & 3D Morphable Models for Monocular 4D Face Reconstruction**, CVMP 2018)

Deep Learning-based Approaches

Deep Learning for 3D Faces

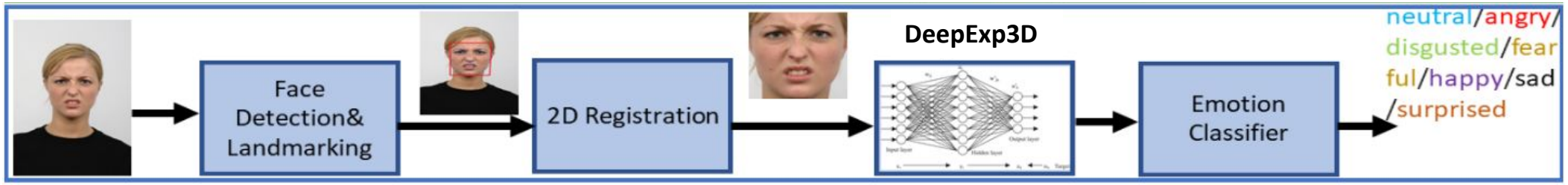




[1] **M.R. Koujan**, A. Roussos and S. Zafeiriou. **DeepFaceFlow: In-the-Wild Dense 3D Facial Motion Estimation**, CVPR'20

[2] **M.R. Koujan**, L. Alharbawee, G. Giannakakis, N. Pugeault and A. Roussos. **Real-time Facial Expression Recognition ``In The Wild'' by Disentangling 3D Expression from Identity**, FG'20

Facial Expression Recognition



- **DeepExp3D**: CNN that **regresses** 3D expression parameters from RGB images **directly**
  - **disentangling** expression parameters from identity and 3D pose



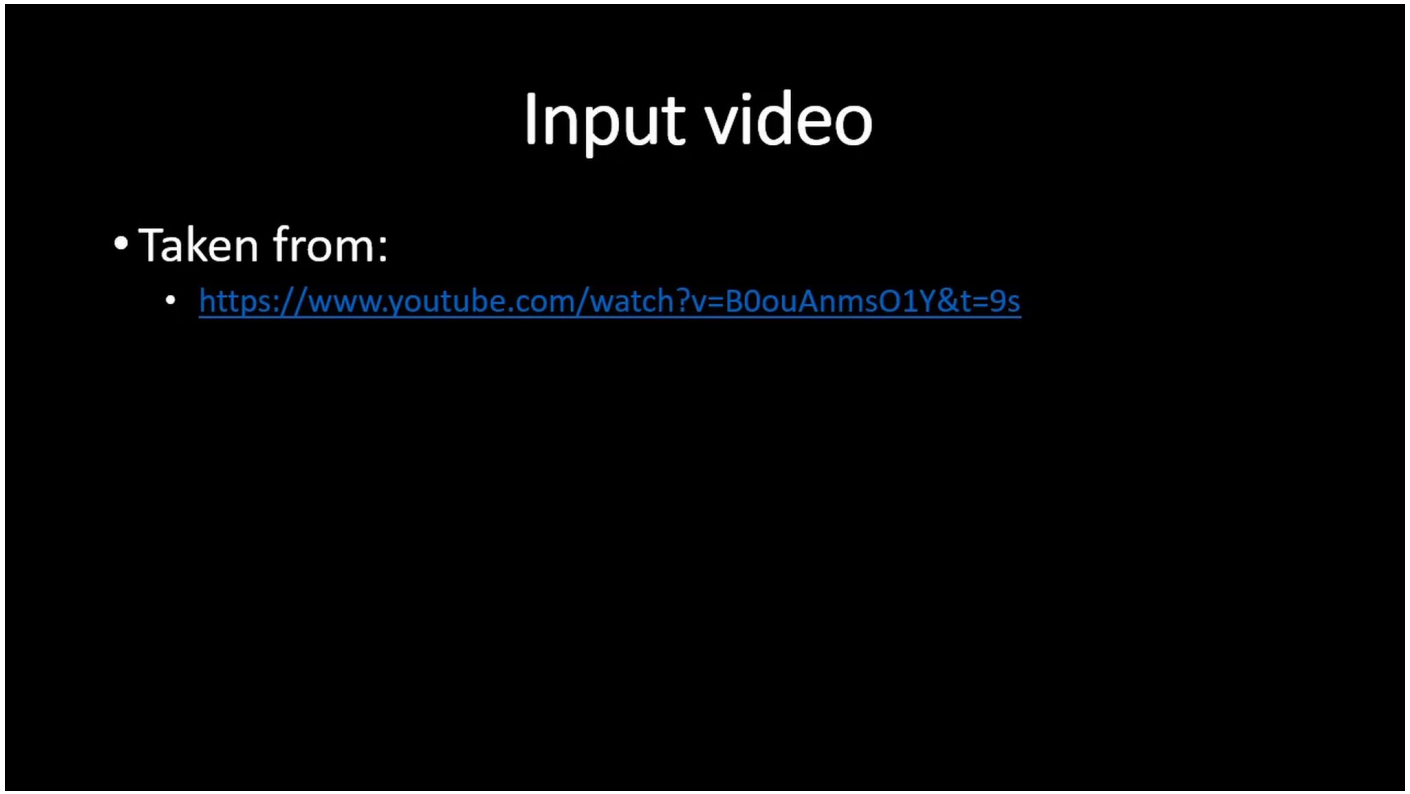
- Recognition output: **emotion classification** from the regressed 3D expression parameters
- Can be used as well for stress recognition

[1] **M.R. Koujan**, L. Alharbawee, G. Giannakakis, N. Pugeault and A. Roussos. **Real-time Facial Expression Recognition ``In The Wild'' by Disentangling 3D Expression from Identity**, FG'20

[2] G. Giannakakis*, **M.R. Koujan***, A. Roussos, and K. Marias. **Automatic stress analysis from facial videos based on deep facial action units recognition**, PPA journal.

Facial Expression Recognition

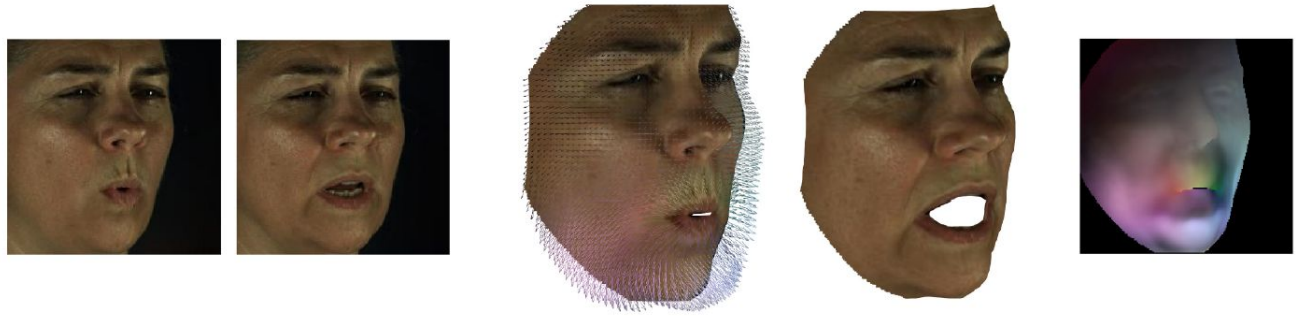
- application on **videos**:
  - applied on **each frame** independently
  - **real-time performance**: 50 frames per second

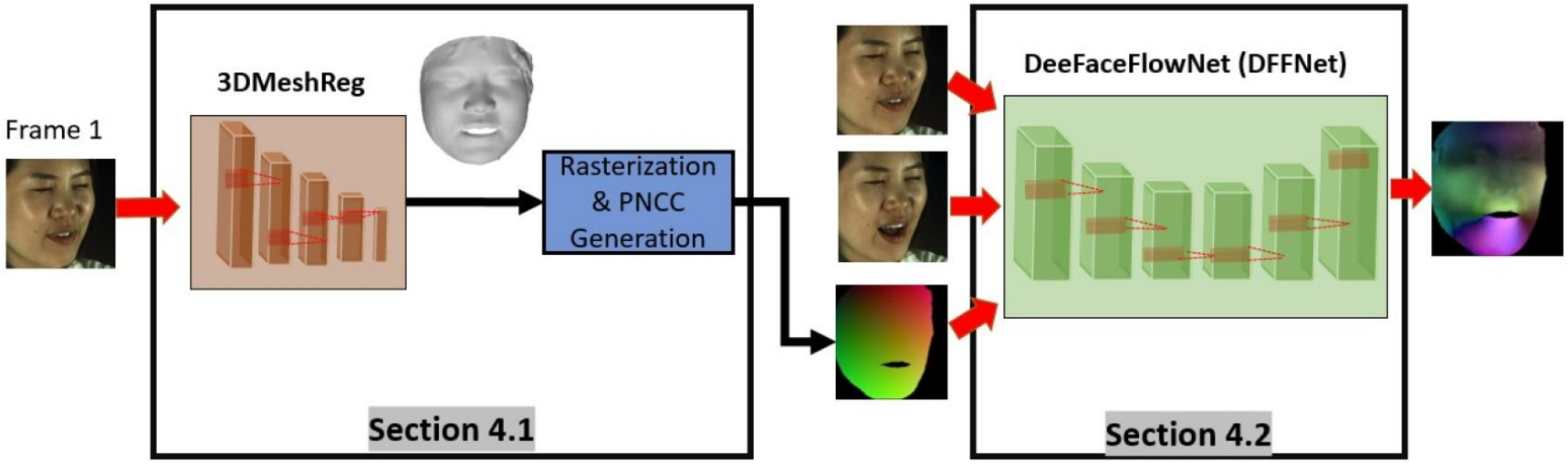


M.R. Koujan, L. Alharbawee, G. Giannakakis, N. Pugeault and A. Roussos. **Real-time Facial Expression Recognition ``In The Wild'' by Disentangling 3D Expression from Identity**, FG'20

Highly-detailed Facial Motion Estimation

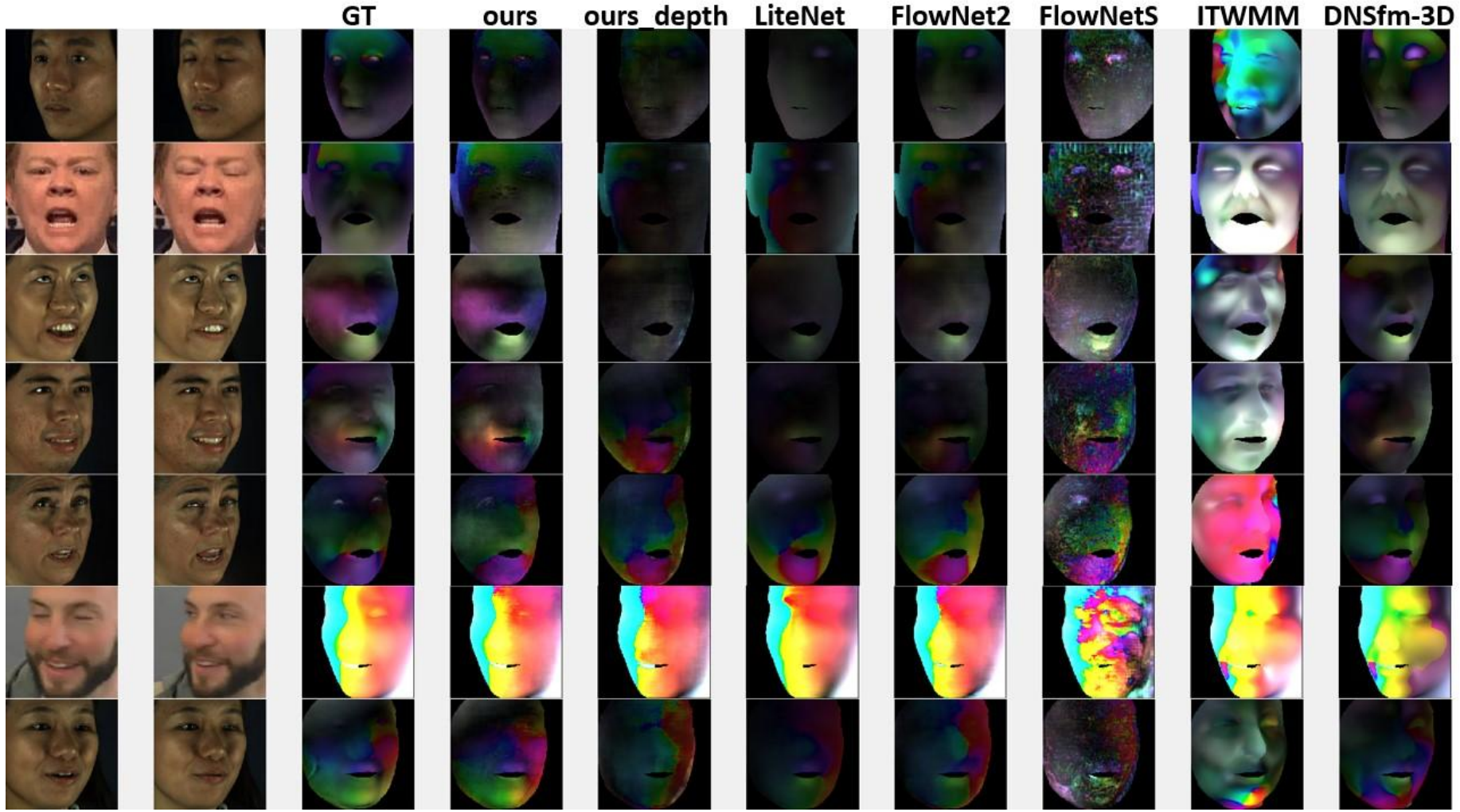
- Deep learning **3D optical flow** for in-the-wild facial videos





Highly-detailed Facial Motion Estimation

- **Qualitative results**



Full Head Re-enactment

- we transfer **facial expressions** along with **head pose** and **eye gaze**
- nearly **real-time** speeds
- using **3D-based representation** & **Generative Adversarial Networks** (GANs)



M.R. Koujan*, M. Doukas*, A. Roussos and S. Zafeiriou. **Head2Head: Video--based Neural Head Synthesis**, FG 2020
M.R. Koujan*, M.Doukas*, A.Roussos, V. Sharamanska. **Head2Head++: Deep Facial Attributes Re-targeting**, TBIOM journal 2020

Full Head Re-enactment

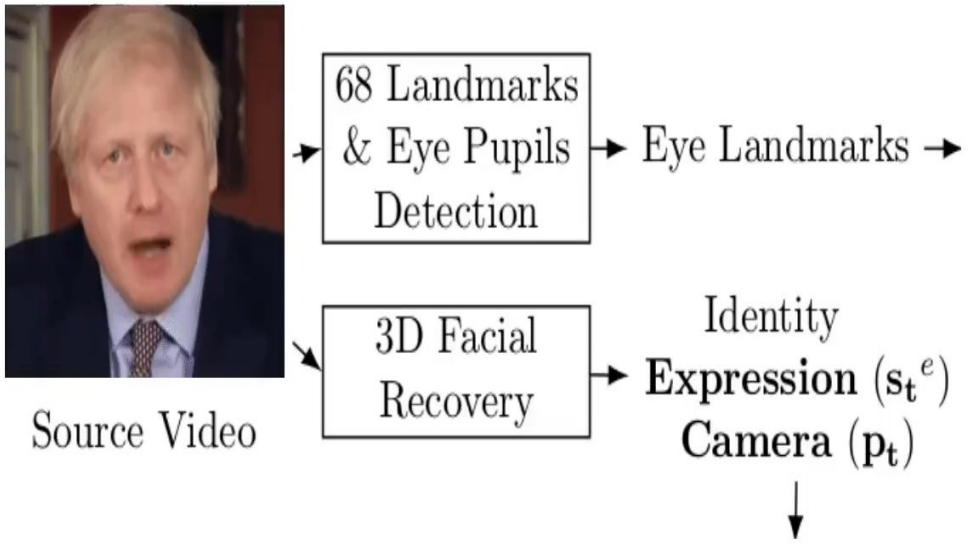
- applications:
  - tele-presence
  - performance capture for visual effects
  - actor dubbing
  - video games
  - virtual tutors

Turnbull to Biden



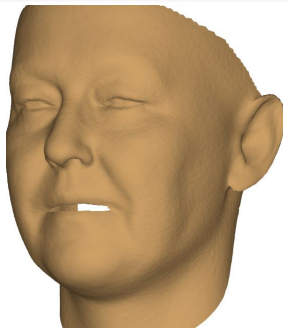
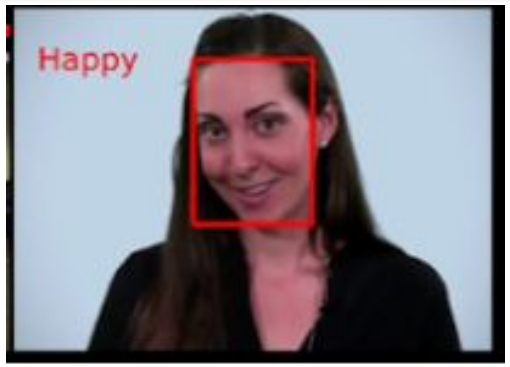
M.R. Koujan, M. Doukas, A. Roussos and S. Zafeiriou. **Head2Head: Video--based Neural Head Synthesis**, FG 2020

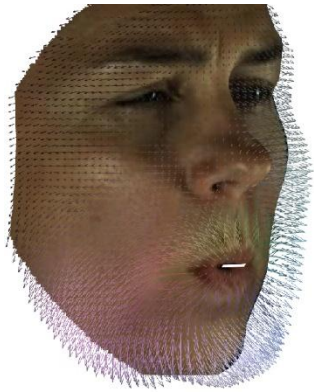
Full Head Re-enactment Pipeline



Conclusions

Conclusions

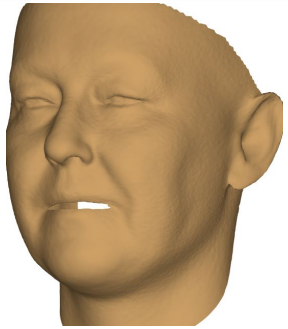
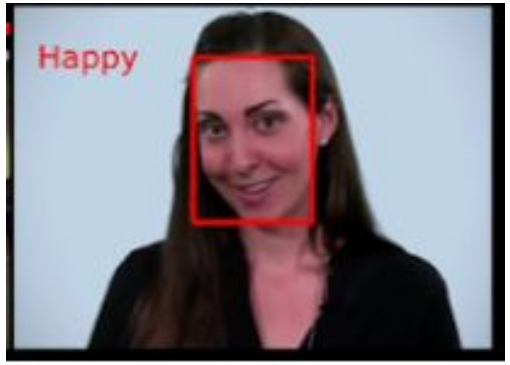




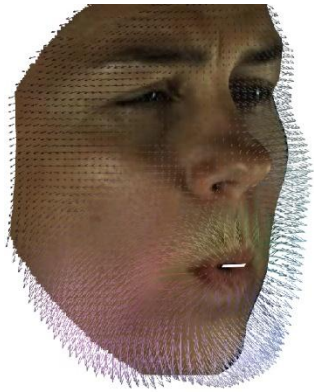





- **Pioneering methodologies** for 3D face shape modelling, analysis and synthesis
- **Collection and annotation** of a large-scale dataset of in-the-wild facial performances
- Using **monocular input only**, our methods yield **state-of-the-art results** on estimating:
  - dense dynamic 3D shape and motion
  - facial expressions
  - realistic re-enactment

Conclusions

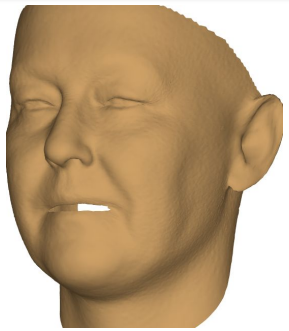
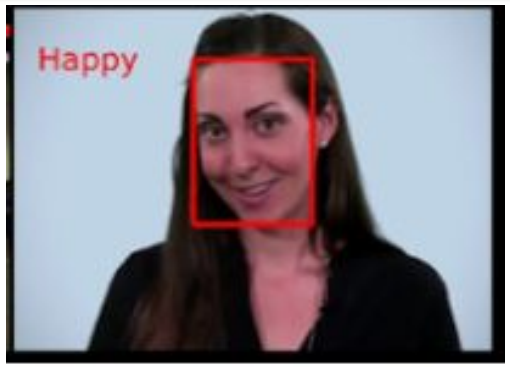




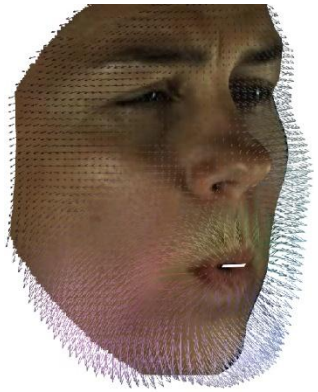


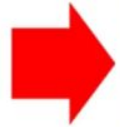



- **Key components:**
  - Large-scale datasets of in-the-wild facial videos
  - highly-detailed and realistic **shape priors**
  - robust **model fitting**
  - latest advances of **deep learning** for Computer Vision

Future Works




- Extend the 4D face reconstruction pipeline to textures as well
- Include the temporal aspect for FER in videos
- Estimate 3D flow for occluded 3D vertices not visible in the reference frame
- Deploy a Full-head 3DMM of face, eyes, ears and tongue for synthesis

Publications

- **M.R. Koujan**, and A. Roussos. "**Combining dense nonrigid structure from motion and 3d morphable models for monocular 4d face reconstruction**." Proceedings of the 15th ACM SIGGRAPH European Conference on Visual Media Production. 2018.
- **M.R. Koujan**, N. Dochev, A. Roussos, "**Real-Time Monocular 4D Face Reconstruc- tion using the Large Scale Facial Models**". Demo paper at the ACM SIGGRAPH European Conference on Visual Media Production (CVMP 2018), London, UK, De- cember 2018
- **M.R. Koujan**, L. Alharbawee, G. Giannakakis, N. Pugeault and A. Roussos, "**Real- Time Facial Expression Recognition "In The Wild" by Disentangling 3D Expression from Identity**," in 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (FG), Buenos Aires, AR, 2020 pp. 539-546
- **M.R. Koujan***, M. Doukas*, A. Roussos and S. Zafeiriou, "**Head2Head: Video-Based Neural Head Synthesis**," in 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (FG), Buenos Aires, AR, 2020 pp. 319-326
- G. Giannakakis, **M.R. Koujan**, A. Roussos and K. Marias, "**Automatic Stress De- tection Evaluating Models of Facial Action Units**," in 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (FG), Buenos Aires, AR, 2020 pp. 817-822
- **M.R. Koujan**, M. Doukas, A. Roussos and S. Zafeiriou, "**ReenactNet: Real-Time Full Head Reenactment**," in 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) (FG), Buenos Aires, AR, 2020 pp. 327-327
- **M.R. Koujan**, A. Roussos, and S. Zafeiriou. "**DeepFaceFlow: In-the-wild Dense 3D Facial Motion Estimation**." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR 2020
- **M.R. koujan***, M. Doukas*, V. Sharmanska, A. Roussos and S. Zafeiriou, "**Head2Head++: Deep Facial Attributes Re-Targeting**," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 1, pp. 31–43, 2021
- **M.R. Koujan*** , G. Giannakakis*, A. Roussos and K. Marias, "**Automatic stress analysis from facial videos based on deep facial action units recognition**." Pattern Analysis and Applications (2021): 1-20, 2021

Thank You!