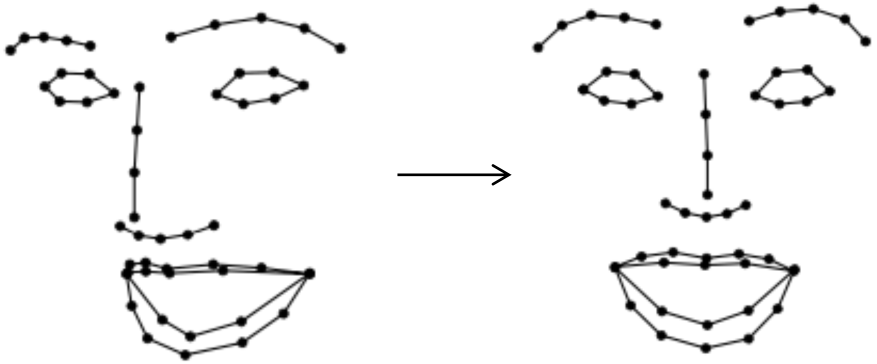
Paper: ARS-06.5

Identity-invariant Facial Landmark Frontalization for Facial Expression Analysis

Vassilios Vonikakis Amazon Web Services Singapore

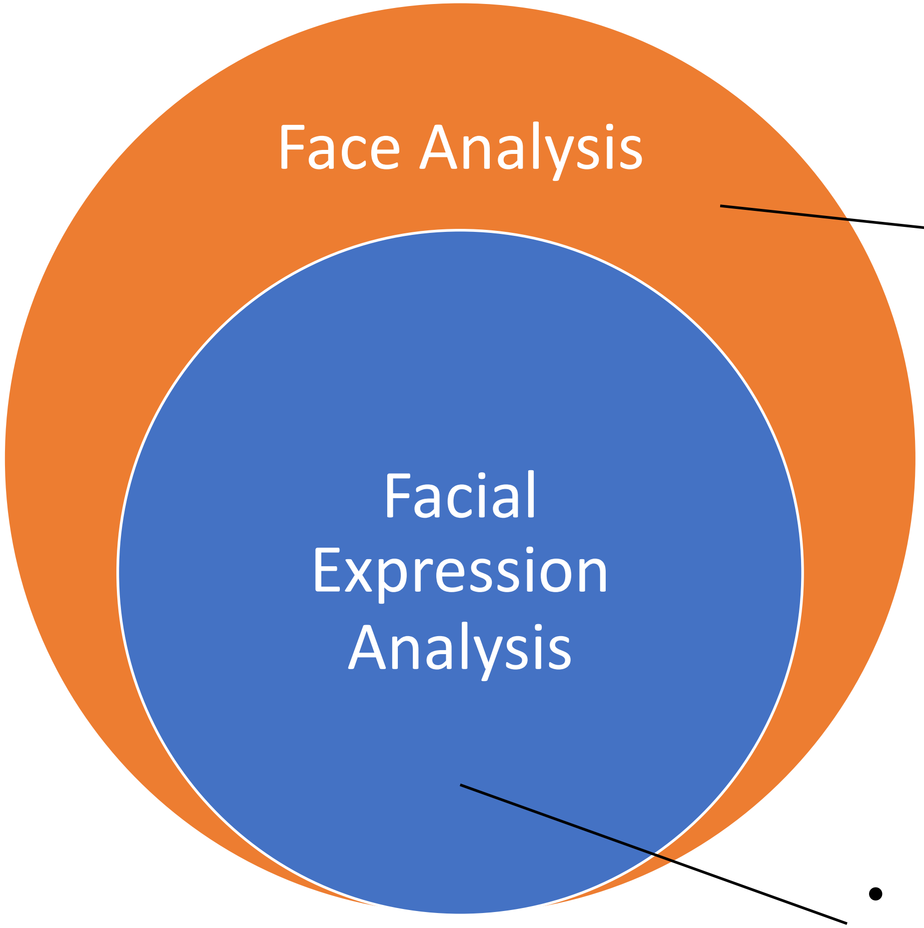
Stefan Winkler National University of Singapore School of Computing



* This work was conducted while both authors were with the Advanced Digital Sciences Center (ADSC), University of Illinois at Urbana-Champaign, Singapore.

Introduction

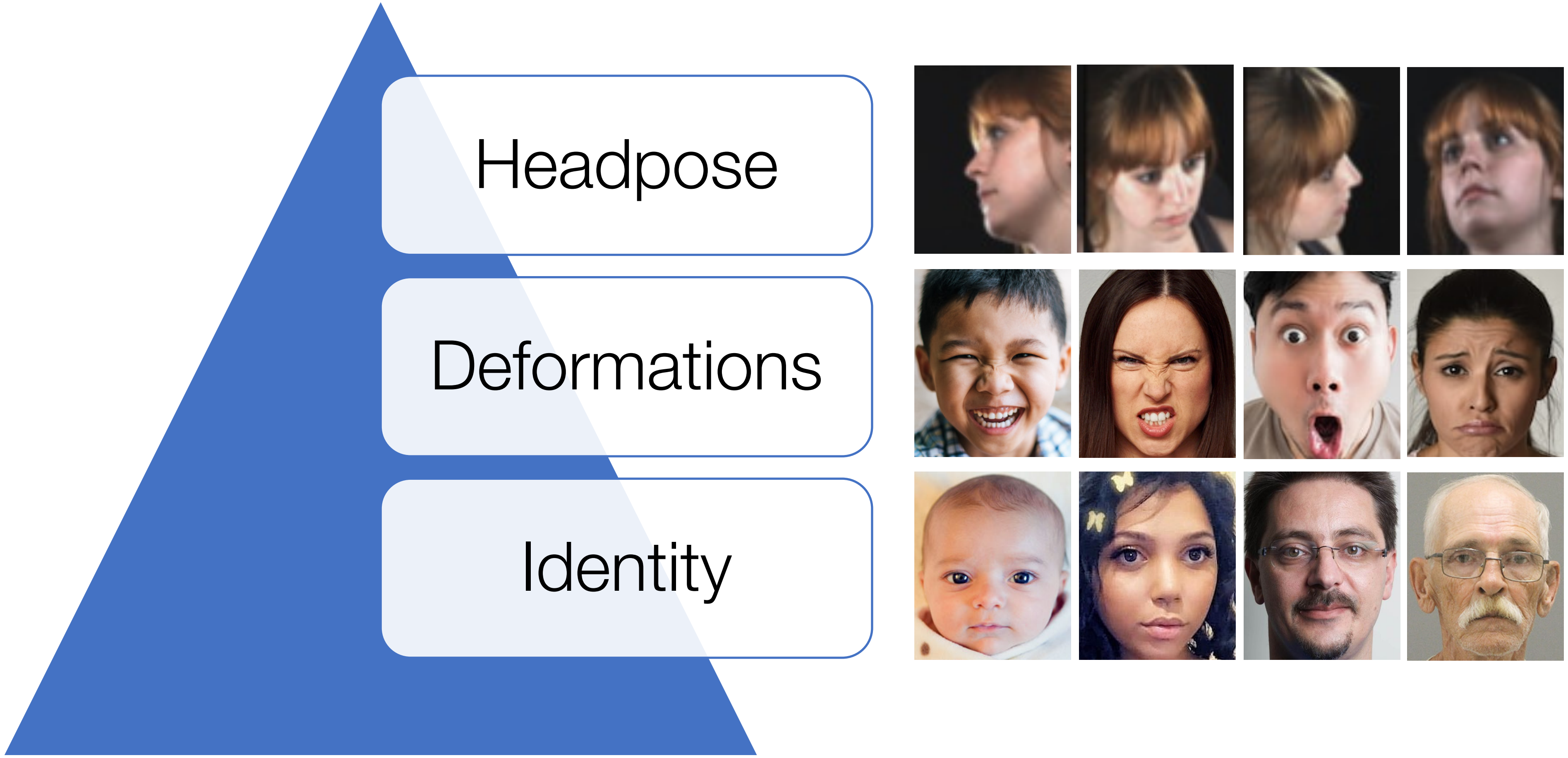
Facial Expression Analysis



- Face recognition/authentication
- Facial demographics (age/gender/race)
- Face transformations (beautifications)
- Facial expression analysis
- Many more…

- Identify the expression that is portrayed on a given face
- (not the identity of the person, or other aspects of it)

3 types of face variability



Eliminating headpose variability

- Headpose invariance
  - Face frontalization
    - Landmark frontalization
      - Frontalization of 2D landmarks
        - Faster and lighter.
        - Potentially better for real-time applications “at the edge”
      - 3D landmark estimation
        - Directly estimates 3D landmarks
        - Promising approach
        - “Heavy” for real-time applications
    - Pixel frontalization
      - 3D reconstruction
      - GANs
        - Operating on pixels
        - Infer the 3D structure of the face and render frontal pixel appearance
        - or use GANs
        - Expensive for real-time applications
  - Include headpose variations in the training set

Difficult to include all combinations of headpose and expressions, especially for difficult to annotate datasets (e.g. Action Units, Arousal/Valence)

Advantages

Proposed method

Operate directly on landmark coordinates

- No additional Facial Expression Analysisture extraction
- Simple computations

Data driven

- Learn directly from data
- Avoid using 3D reconstruction

Fast, light

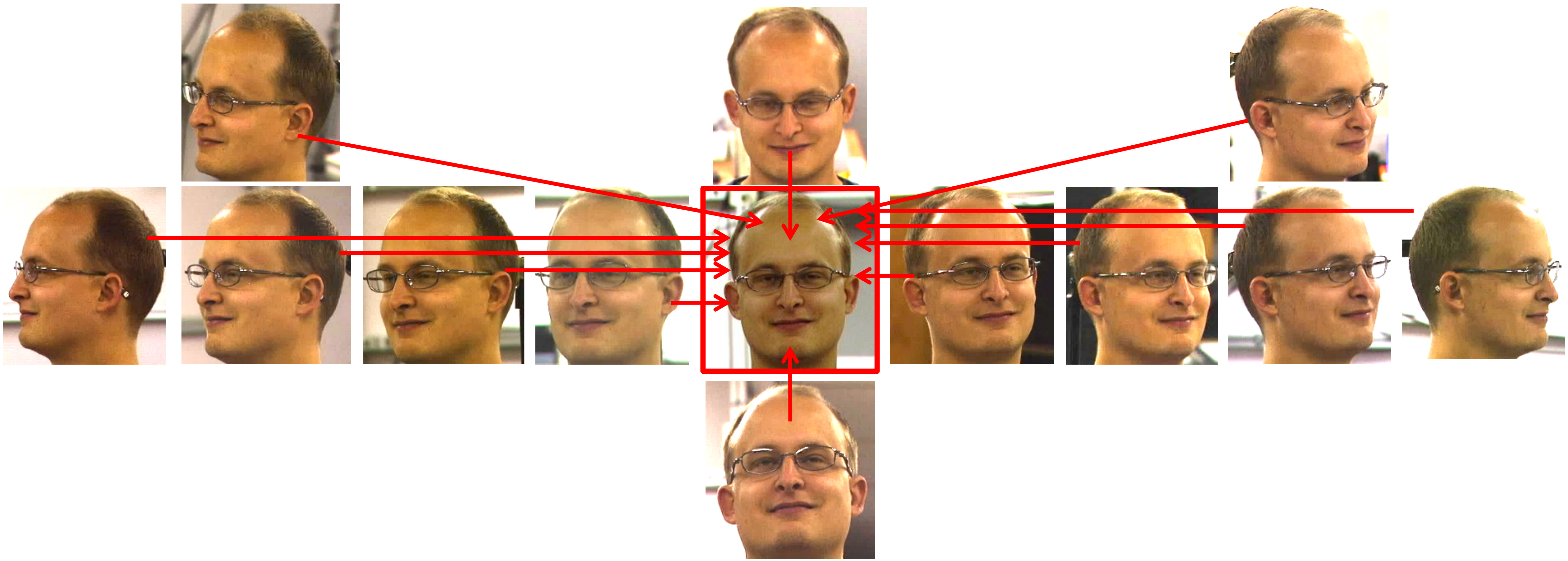
- Minimal computational overhead
- No GPU requirements

Applications

- Computation light Facial Expression Analysis systems
- Real-time in edge devices

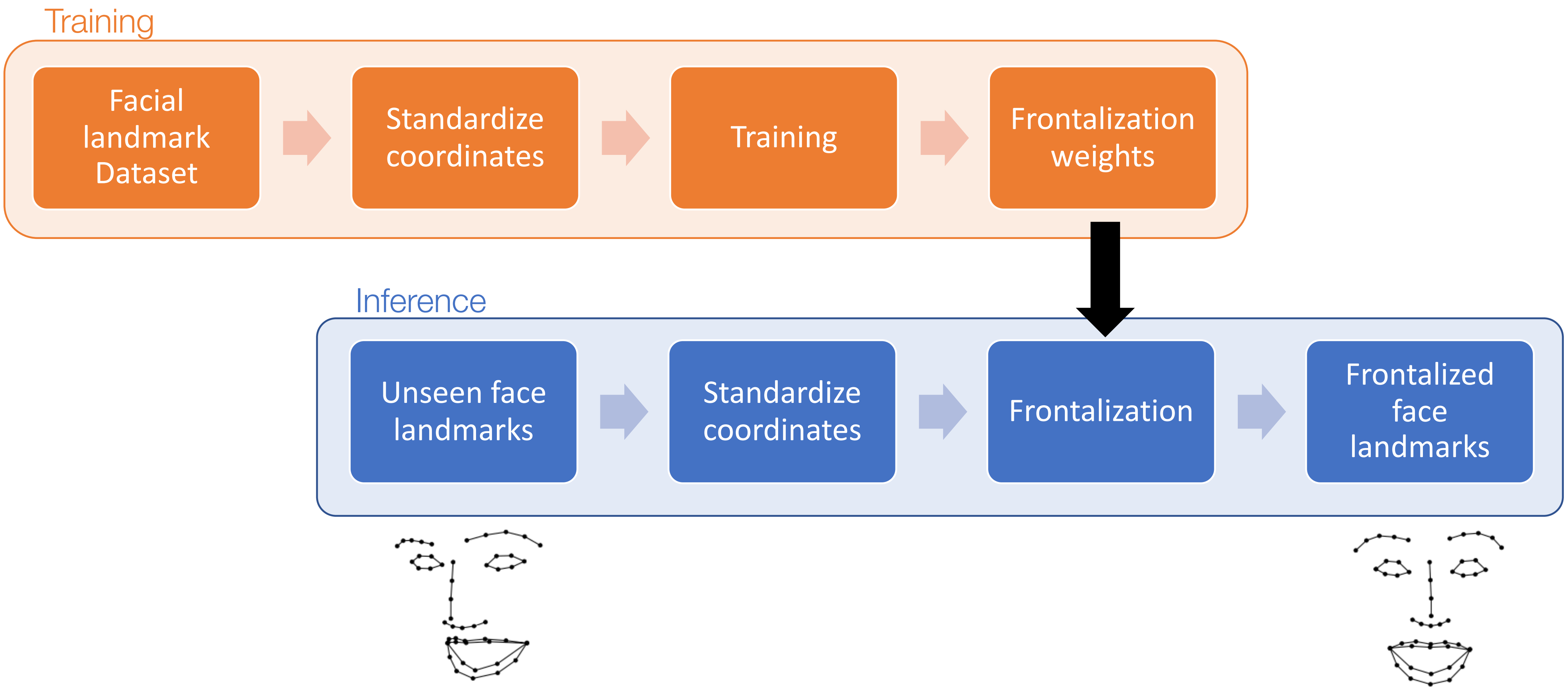
2D Landmark Frontalization

Main idea



- Given a dataset of multiple viewpoints of the same face, and many expressions per subject
- **Learn a mapping** for all facial points to the frontal view, while maintaining expressions

Workflow



Training
Facial landmark Dataset
Standardize coordinates
Training
Frontalization weights
Inference
Unseen face landmarks
Standardize coordinates
Frontalization
Frontalized face landmarks

Coordinate standardization

Face image

Facial landmarks (Xiong & De la Torre 2013)

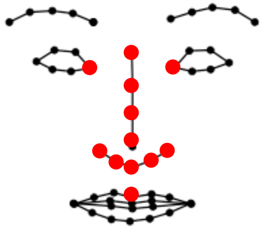
$$\mathbb{R}^{N\times 2}$$

$$\boldsymbol{P} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix}$$

$$\mathbb{R}^{N\times 2}$$

$$\widehat{\boldsymbol{P}} = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_N & y_N \end{bmatrix}$$

Non-isotropic Procrustes (Normalize for Translation, Rotation, Scaling)



$$\mathbb{R}^{2N}$$

$$\boldsymbol{p} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ \vdots \\ x_N \\ y_N \end{bmatrix}$$

Final landmark vector

Frontalization

$$argmin_{x}\|\boldsymbol{Y}-\boldsymbol{AX}\|_2 + \lambda\|\boldsymbol{X}\|_2$$

$$\widehat{\boldsymbol{X}} = (\boldsymbol{A}^T\boldsymbol{A}+\lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{Y}$$

Frontalization weights

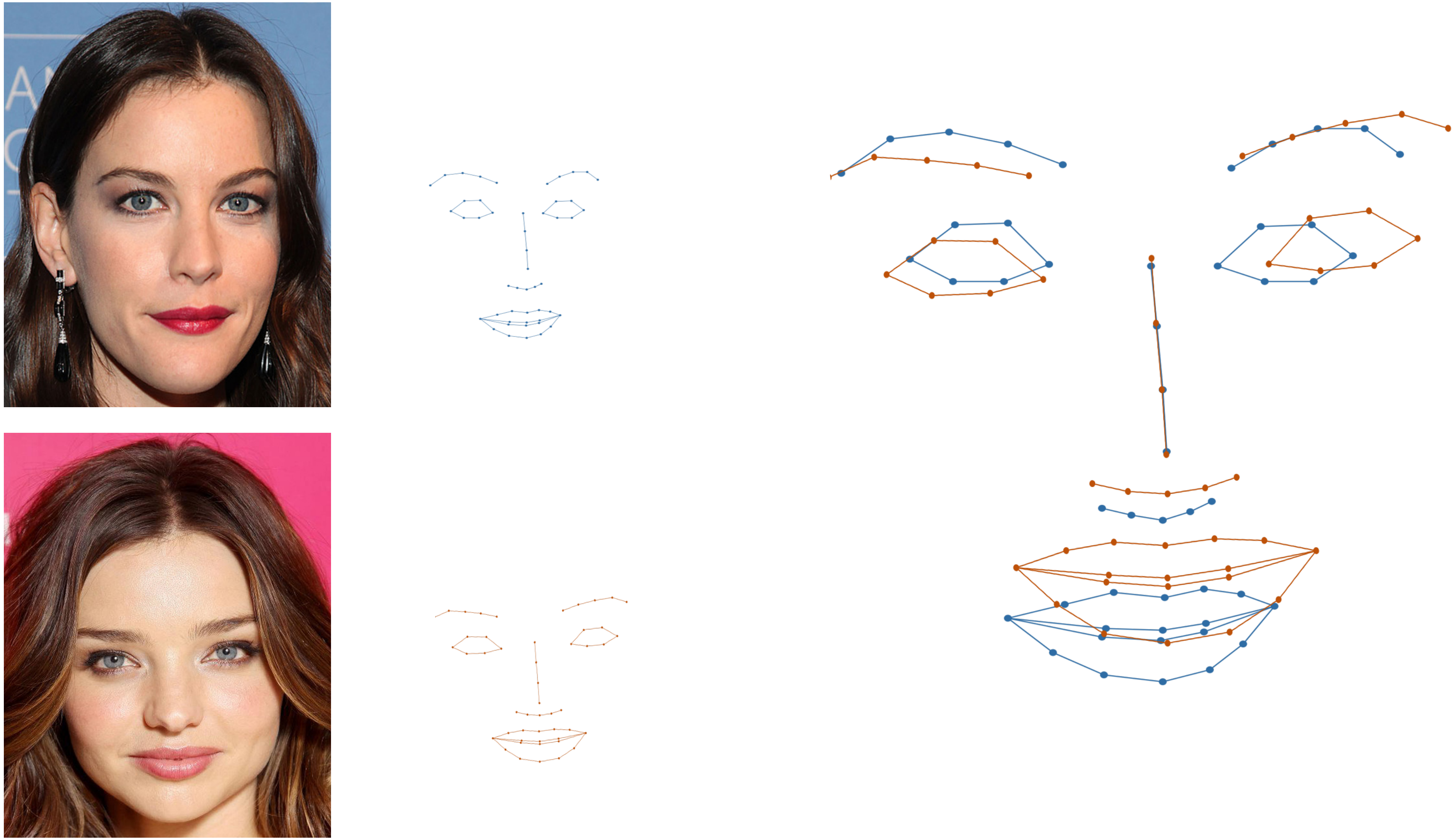
<u>Face dataset</u>

- M subject
- J expressions / subject
- K headposes / expressions / subject

Matrix Y (frontal points)	Matrix A (non-frontal points)	
Person 1, expression 1, frontal Person 1, expression 1, frontal … Person 1, expression 1, frontal	Person 1, expression 1, pose 1 Person 1, expression 1, pose 2 … Person 1, expression 1, pose K	Subject 1 expression1
Person 1, expression 2, frontal Person 1, expression 2, frontal … Person 1, expression 2, frontal	Person 1, expression 2, pose 1 Person 1, expression 2, pose 2 … Person 1, expression 2, pose K	Subject 1 expression2
⋮	⋮	
Person 1, expression J, frontal Person 1, expression J, frontal … Person 1, expression J, frontal	Person 1, expression J, pose 1 Person 1, expression J, pose 2 … Person 1, expression J, pose K	Subject 1 Expression J
⋮	⋮	
Person 2, expression 1, frontal Person 2, expression 1, frontal … Person 2, expression 1, frontal	Person 2, expression 1, pose 1 Person 2, expression 1, pose 2 … Person 2, expression 1, pose K	Subject 2 expression1
Person 2, expression 2, frontal Person 2, expression 2, frontal … Person 2, expression 2, frontal	Person 2, expression 2, pose 1 Person 2, expression 2, pose 2 … Person 2, expression 2, pose K	Subject 2 expression2
⋮	⋮	
Person 2, expression J, frontal Person 2, expression J, frontal … Person 2, expression J, frontal	Person 2, expression J, pose 1 Person 2, expression J, pose 2 … Person 2, expression J, pose K	Subject 2 Expression J
⋮	⋮	

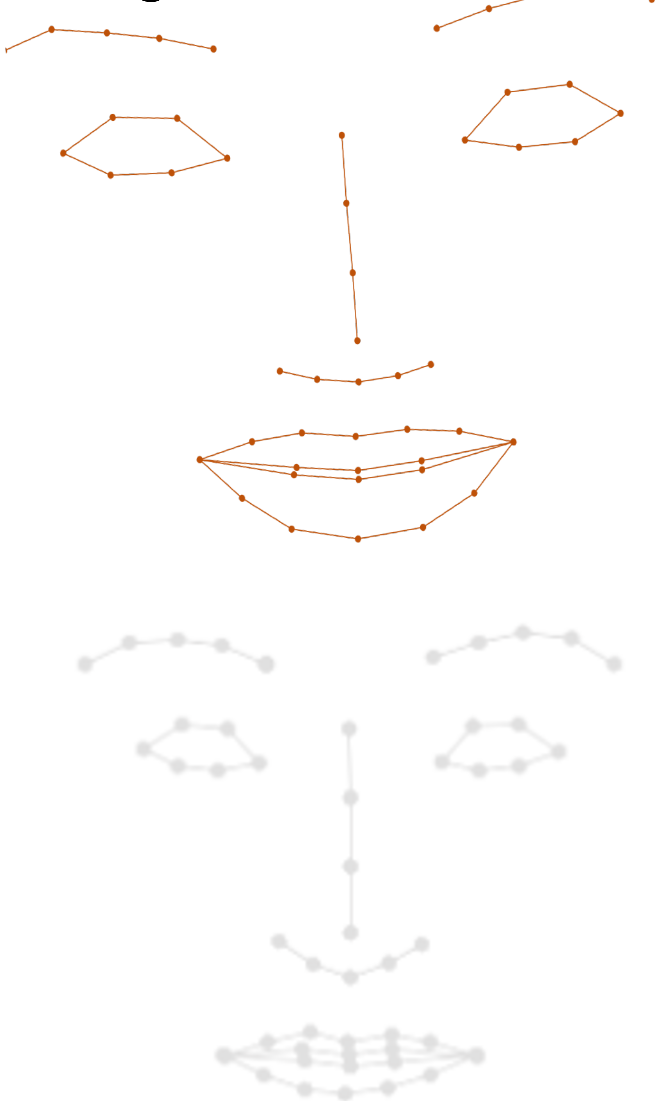
Subject M

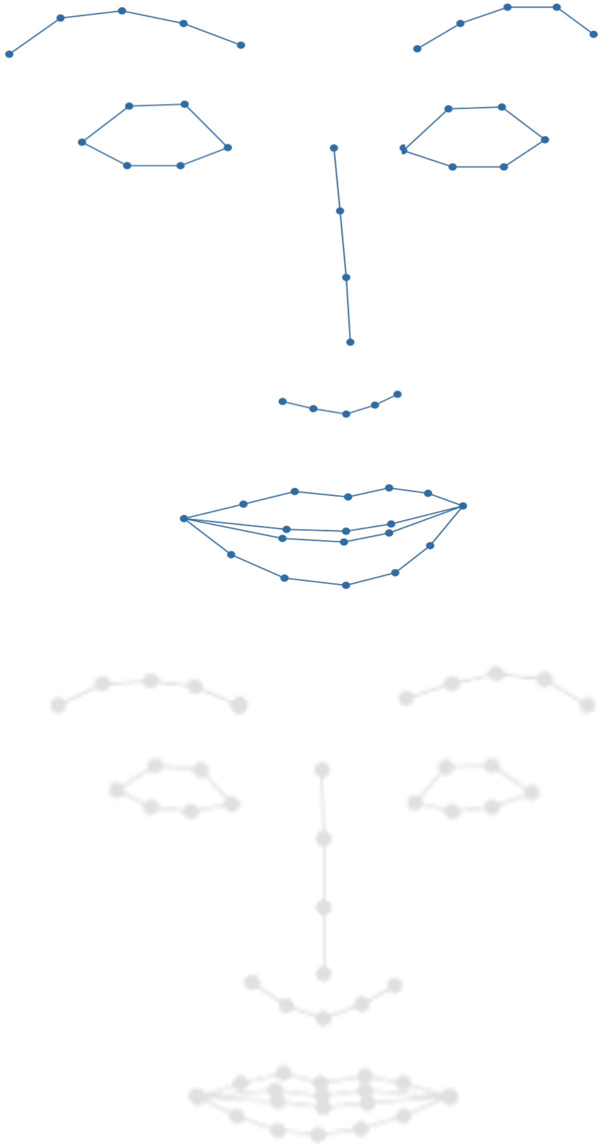
Identity invariance



- Although facial landmarks remove a big part of a face’s visual appearance, there still identity information
- Each person has different relative positions for the eyes, mouth, nose, eyebrows etc.

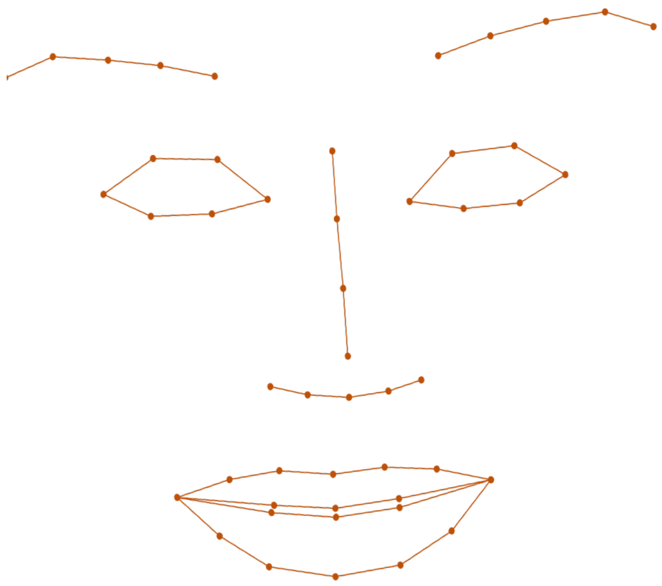
Identity invariance

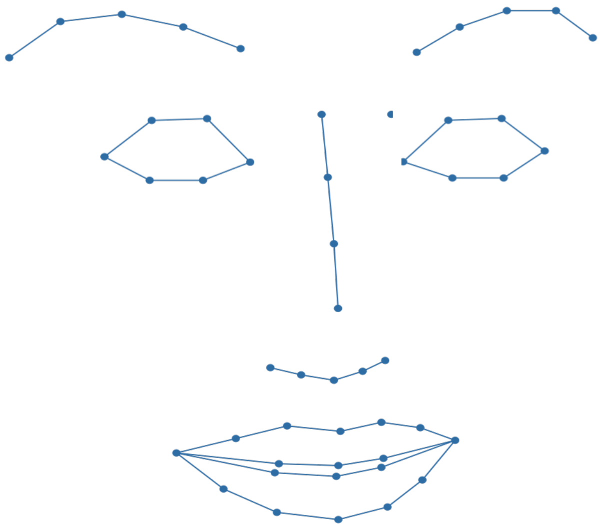




- We reposition the facial parts to a standardized position
- The shape of the parts does not change (expression remains the same). Only their relative position on the face.

Identity invariance





Frontalization & Identity standardization

Identity standardization function

$$\boldsymbol{p_s} = f(\boldsymbol{p}, \boldsymbol{p_t})$$

Identity standardized face points

Original face points

Template face points

$$argmin_x \|\boldsymbol{Y} - \boldsymbol{AX}\|_2 + \lambda \|\boldsymbol{X}\|_2$$

$$\widehat{\boldsymbol{X}} = (\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I})^{-1}\boldsymbol{A}^T\boldsymbol{Y}$$

<u>Face dataset</u>

- M subject
- J expressions / subject
- K headposes / expressions / subject

Matrix Y (frontal points)

Identity standardized

Matrix A (non-frontal points)

Matrix Y	Matrix A	
Person 1, expression 1, frontal Person 1, expression 1, frontal … Person 1, expression 1, frontal	Person 1, expression 1, pose 1 Person 1, expression 1, pose 2 … Person 1, expression 1, pose K	Person1 expression1
Person 1, expression 2, frontal Person 1, expression 2, frontal … Person 1, expression 2, frontal	Person 1, expression 2, pose 1 Person 1, expression 2, pose 2 … Person 1, expression 2, pose K	Person1 expression2
⋮		
Person 1, expression J, frontal Person 1, expression J, frontal … Person 1, expression J, frontal	Person 1, expression J, pose 1 Person 1, expression J, pose 2 … Person 1, expression J, pose K	Person1 Expression J
⋮		
Person 2, expression 1, frontal Person 2, expression 1, frontal … Person 2, expression 1, frontal	Person 2, expression 1, pose 1 Person 2, expression 1, pose 2 … Person 2, expression 1, pose K	Person2 expression1
Person 2, expression 2, frontal Person 2, expression 2, frontal … Person 2, expression 2, frontal	Person 2, expression 2, pose 1 Person 2, expression 2, pose 2 … Person 2, expression 2, pose K	Person2 expression2
⋮		
Person 2, expression J, frontal Person 2, expression J, frontal … Person 2, expression J, frontal	Person 2, expression J, pose 1 Person 2, expression J, pose 2 … Person 2, expression J, pose K	Person2 Expression J
⋮		

Person M

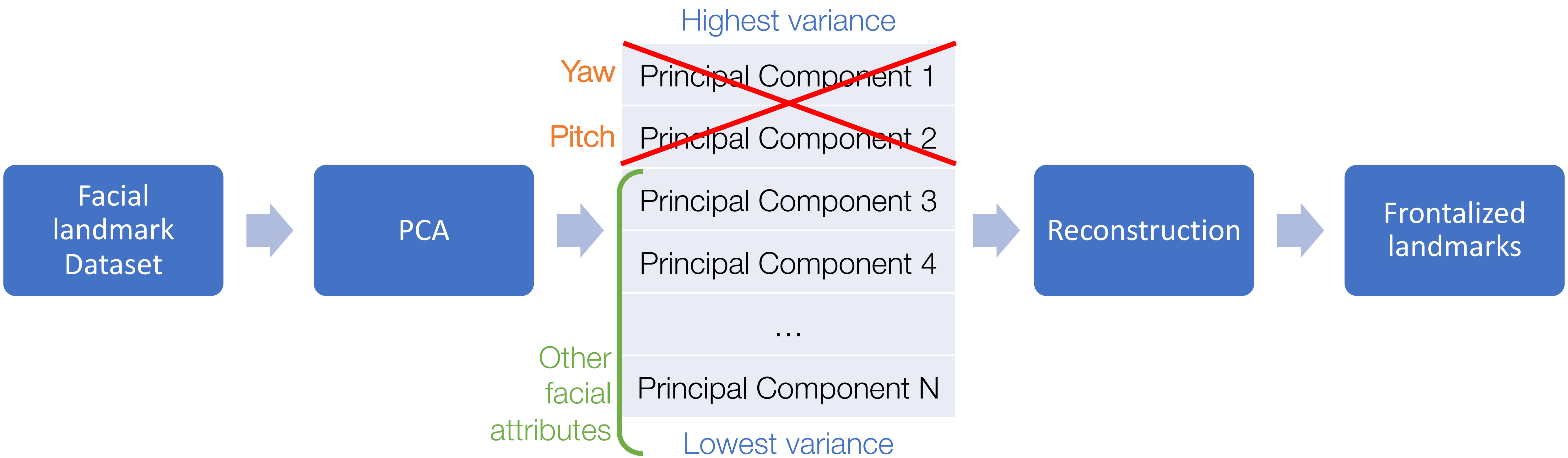
Experimental results

Datasets

Dataset	Pitch	Yaw	
Radboud [17]	–	$0°, \pm 45°, \pm 90°$	Training datasets
Karolinska [18]	–	$0°, \pm 45°, \pm 90°$	
CAS-PEAL [19]	$\pm 15°$	$0°, \pm 22°, \pm 45°, \pm 67°, \pm 90°$	
PIE [20]	$\pm 15°$	$0°, \pm 22°, \pm 45°, \pm 67°, \pm 90°$	
Multi-PIE [21]	$-30°$	$0°, \pm 15°, \pm 30°, \pm 45°, \pm 60°, \pm 75°, \pm 90°$	Testing datasets

- Detected faces with OpenCVs Haar Cascade
- Detected facial landmarks with SDM (Xiong & De la Torre 2013)
- Total number of valid detected faces: 87K (after mirroring 174K)

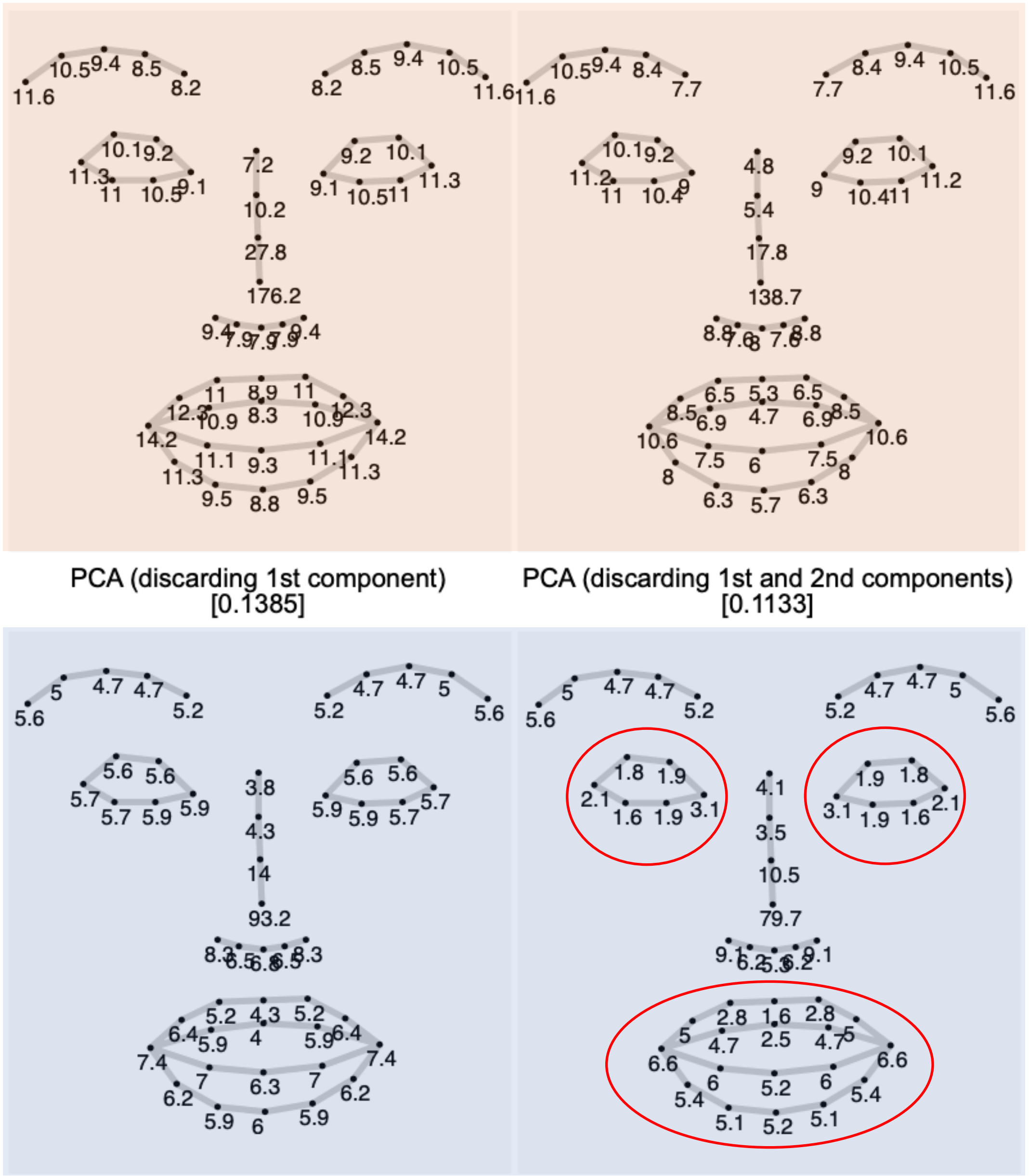
PCA-based frontalization



M.F.Valstar, E.Sanchez-Lozano, J.F.Cohn, L.A.Jeni, J. M. Girard, Z. Zhang, L. Yin, and M. Pantic, "FERA 2017 – addressing head pose in the third facial expression recognition and analysis challenge," in Proc. 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG), 2017, pp. 839–847.

Experimental results

Method	Mean normalized LS residual *
PCA (-1st component)	0.1385
PCA (-1st & 2nd component)	0.1133
Proposed	0.0775
Proposed + Identity standardization	**0.0602**



PCA (discarding 1st component) [0.1385]

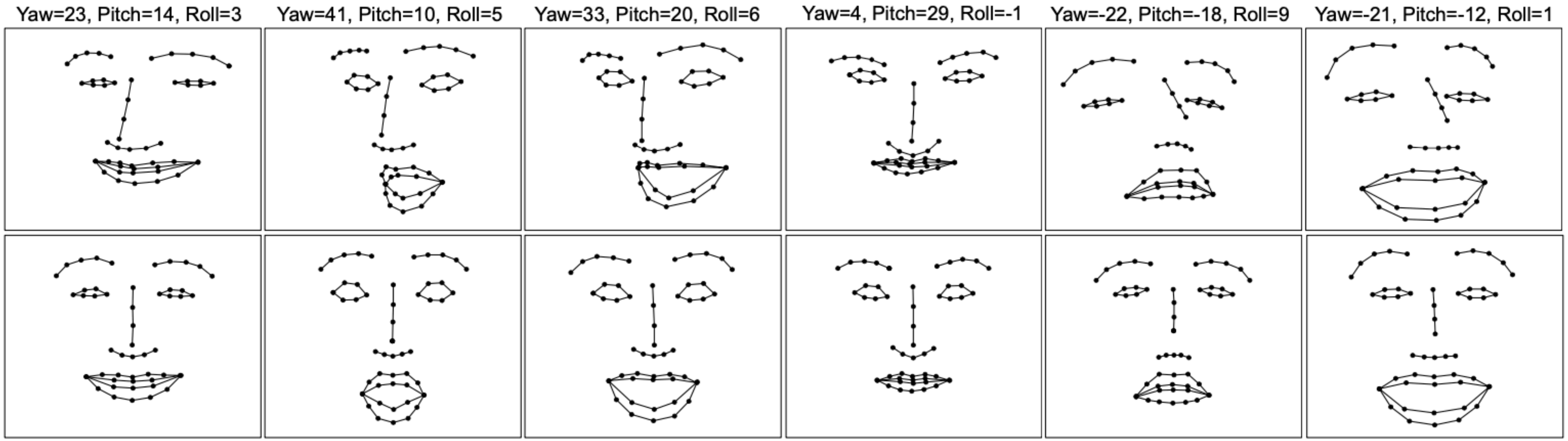
PCA (discarding 1st and 2nd components) [0.1133]

Proposed (without identity standardization) [0.0775]

Proposed (with identity standardization) [0.0602]

* same metric used in face alignment e.g. Xiong & De la Torre 2013

Qualitative results



- Good performance for YAW, at least within [0, ±60º]
- Performance for PITCH not as good (fewer training examples)
- Expression preserved in the frontalized face

Realtime frontalization

https://www.youtube.com/watch?v=5FimHoNv7Dg&feature=emb_logo

Conclusions

Operate directly on landmark coordinates

- No additional Facial Expression Analysisture extraction
- Simple computations

Data driven

- Learn directly from data
- Avoid using 3D reconstruction

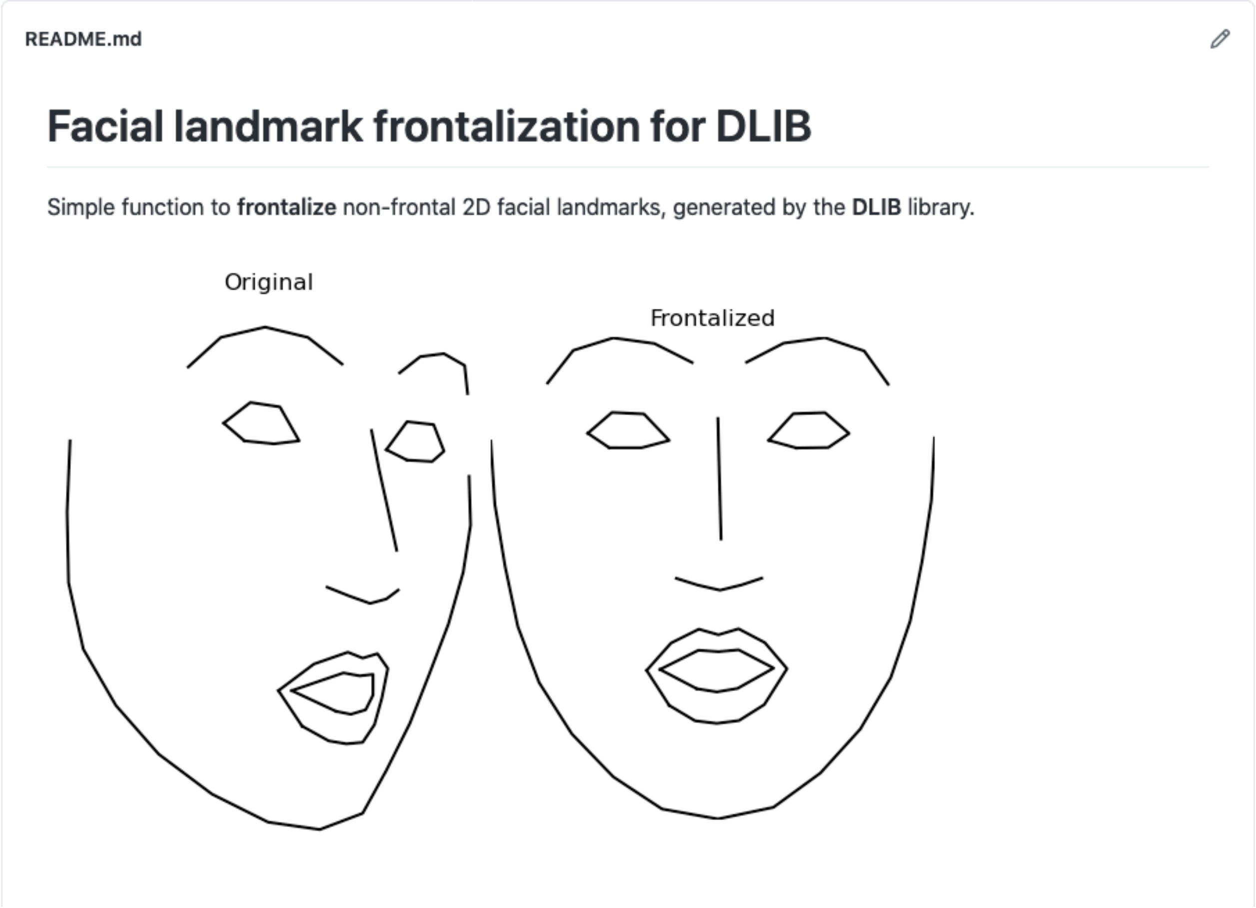
Fast, light

- Minimal computational overhead
- No GPU requirements

Applications

- Computation light Facial Expression Analysis systems
- Real-time in edge devices

Landmark frontalization for DLIB



- Free Python implementation of the code, for facial landmarks extracted by DLIB library
- https://github.com/bbonik/facial-landmark-frontalization

Thank you