Combining Dense Nonrigid Structure from Motion and 3D Morphable Models for Monocular 4D Face Reconstruction

Mohammad Rami Koujan[1] Anastasios Roussos[1,2]

Computer Science Department, University of Exeter[1]
Computer Science Department, Imperial College London[2]



December 14, 2018

The 15th ACM SIGGRAPH European Conference on Visual Media Production (CVMP 2018)



Outline

1. Introduction
2. Proposed Framework
3. Experimental Results
   - Qualitative Evaluation
   - Quantitative Evaluation
4. Conclusion and Future Works

Introduction

Our Work

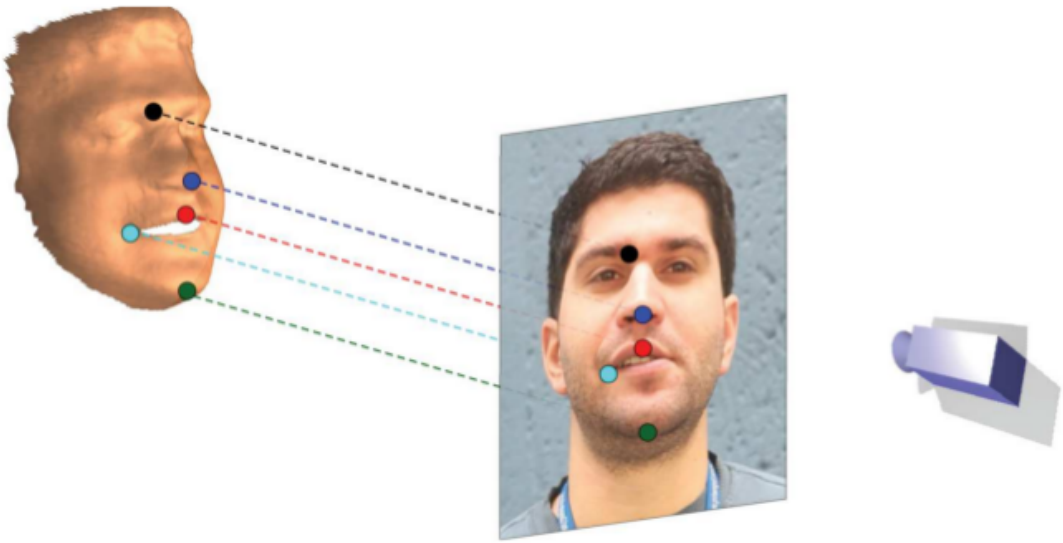
A solution to the 4D face reconstruction problem from monocular in-the-wild videos

(Loading Video...)

Motivation

In-the-wild 4D face Reconstruction

- Deal with recovering the 3D facial geometry from images
- Highly ill-posed problem, especially under unconstrained conditions
- Useful in plenty of applications



Motivation

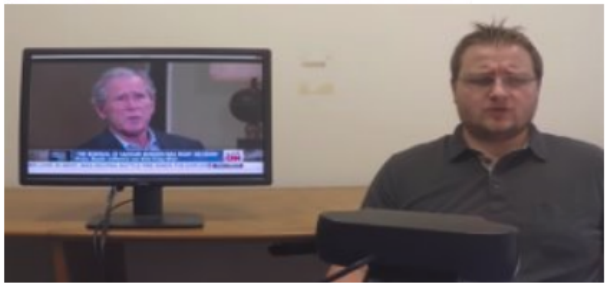
Applications



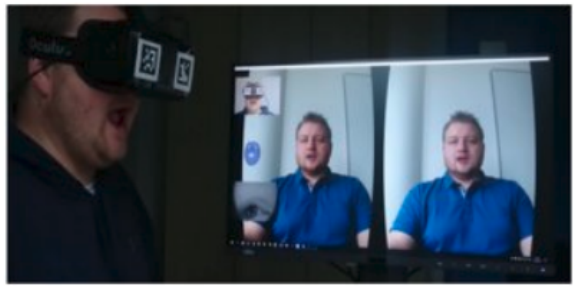
CAO C., el al. (2014). ACM Graph



ICHIM A. E., et al. (2015). ACM Graph



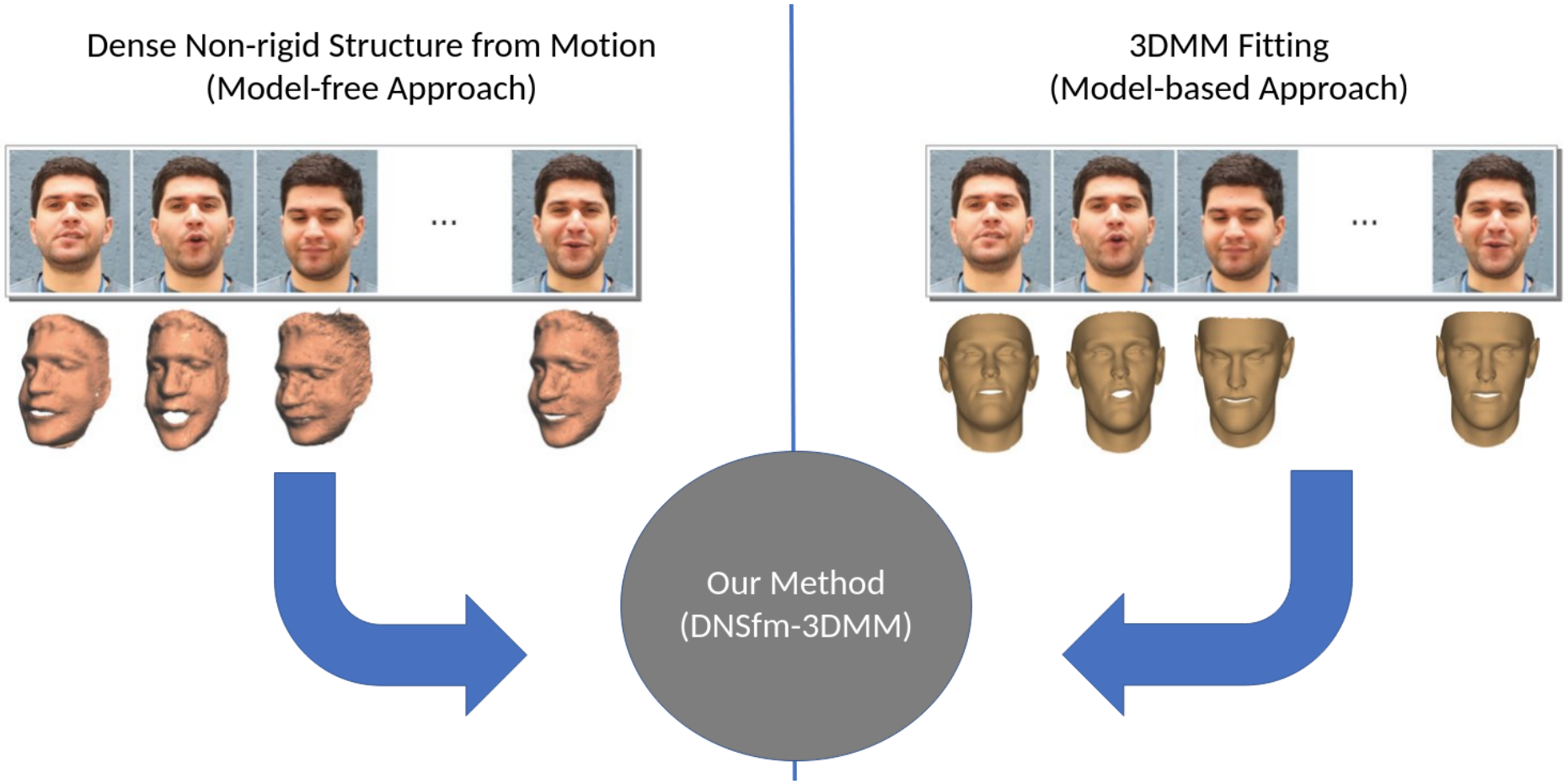
Thies et al. CVPR 2016



Thies et al. CVPR 2016

(M. Zollhofer, et al. EUROGRAPHICS 2018 K.)

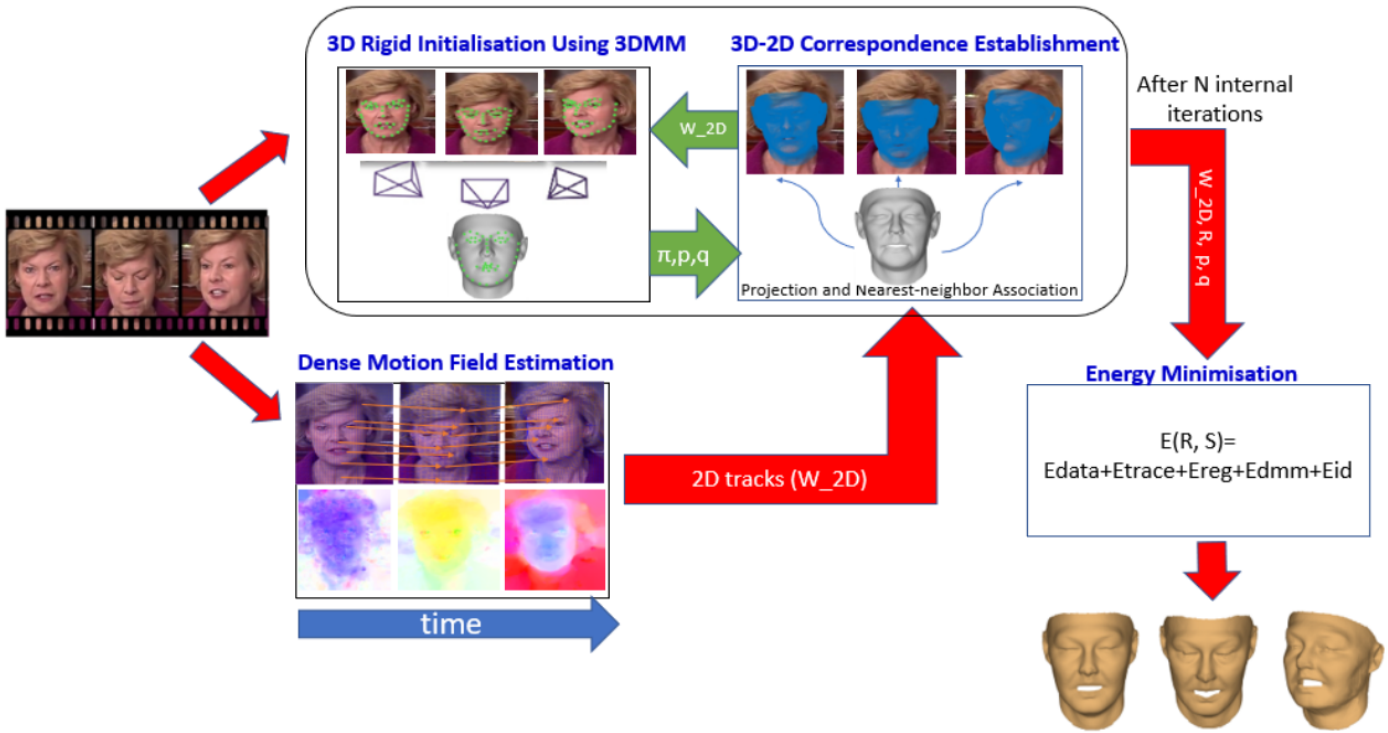
Proposed Framework



(Garg, Roussos, Agapito, IEEE CVPR 2013), (J. Booth, et al. PAMI 2018)

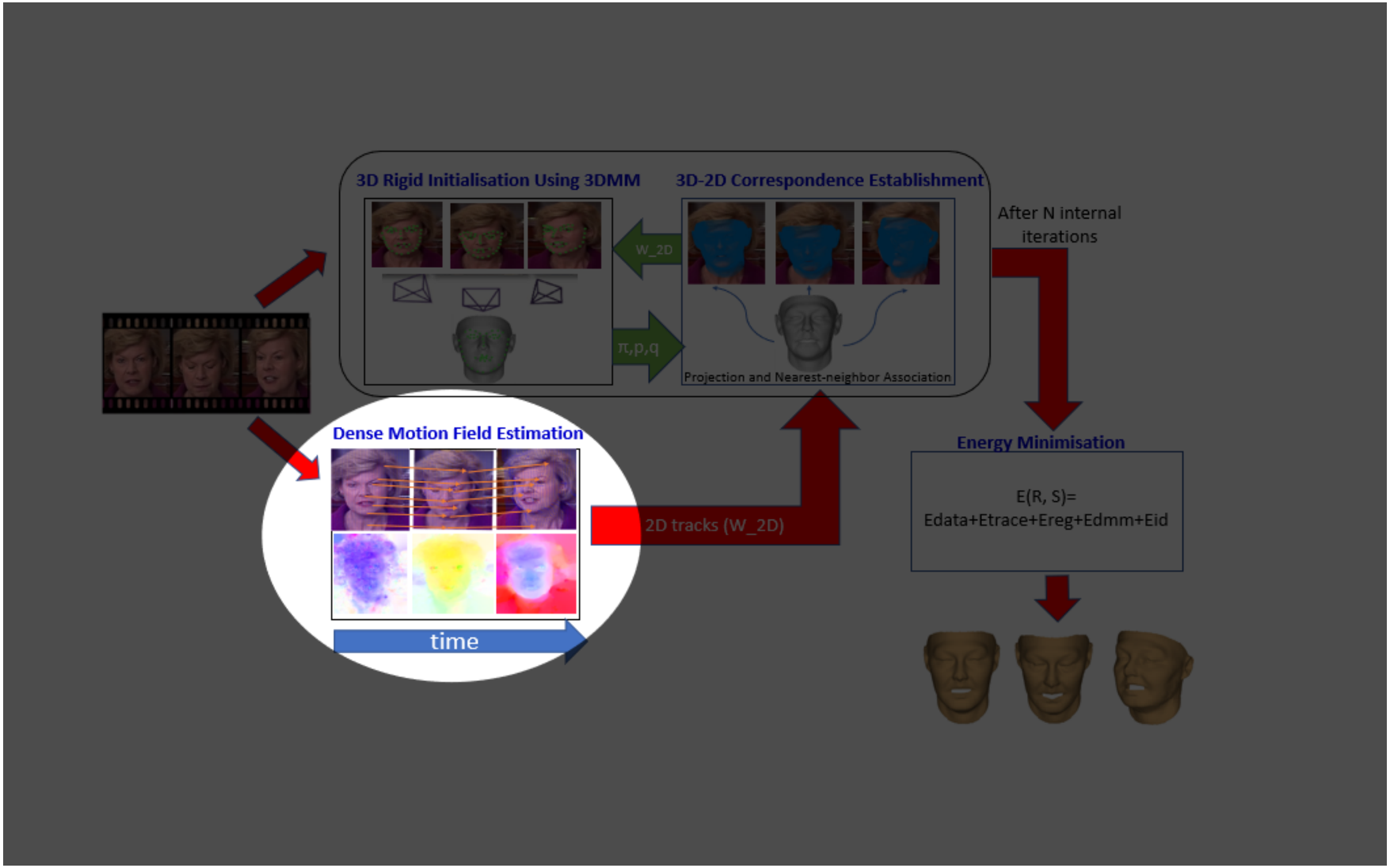
Proposed Framework

- Combination of two schemes:
  1. Dense non-rigid structure from motion
  2. 3D Morphable Models (3DMM): Large Scale Facial Model (LSFM), learnt from 10K face scans [J. Booth, et al. IJCV 2017]



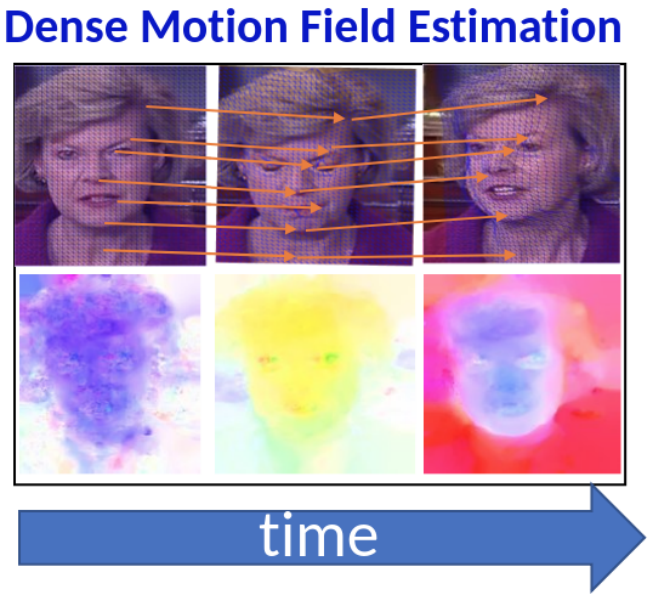
Our proposed method marked by the different adopted stages

Proposed Framework



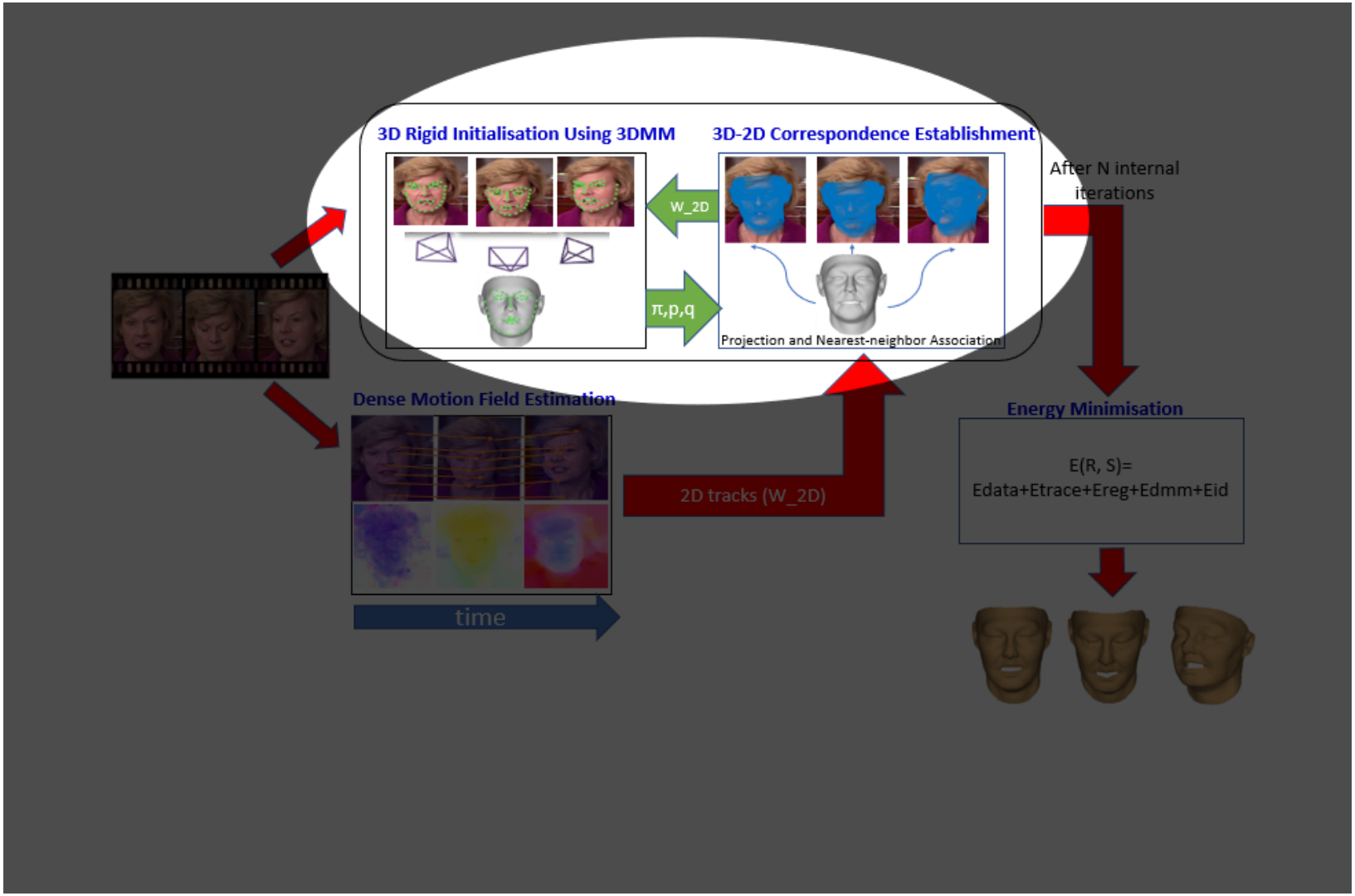
Dense Non-rigid Structure from Motion

- Select a reference frame from the input sequence
- Compute the dense optical flow from this frame to all others
- Variational approach: brightness consistency, low-rank motion basis and spatial regularization terms



(Garg, Roussos, Agapito, IJCV 2013)

Proposed Framework



3D Rigid Initialisation Using 3DMM

Procedure

1. Parse the input video into a sequence of frames
2. Extract 68 facial landmarks
3. Compute the camera matrix using the landmarks and the mean face of the 3DMM
4. Estimate the 3DMM identity and expression parameters utilizing the extracted landmarks and computed camera matrix

Advantages of incorporating 3DMM

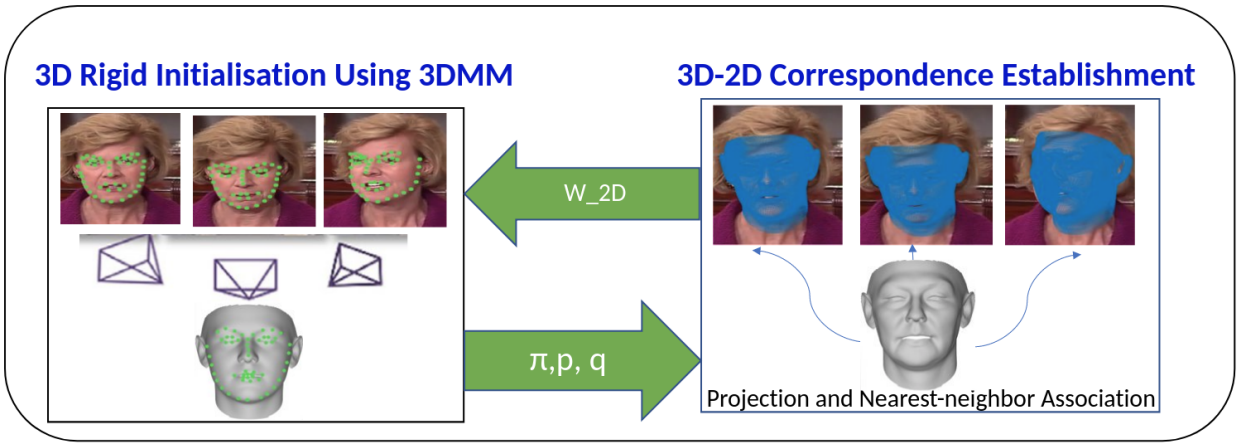
- Deal with the lack of rotation in the input video
- More robust against occlusions and noisy landmarks

$\Rightarrow$ frequently encountered scenarios in in-the-wild videos

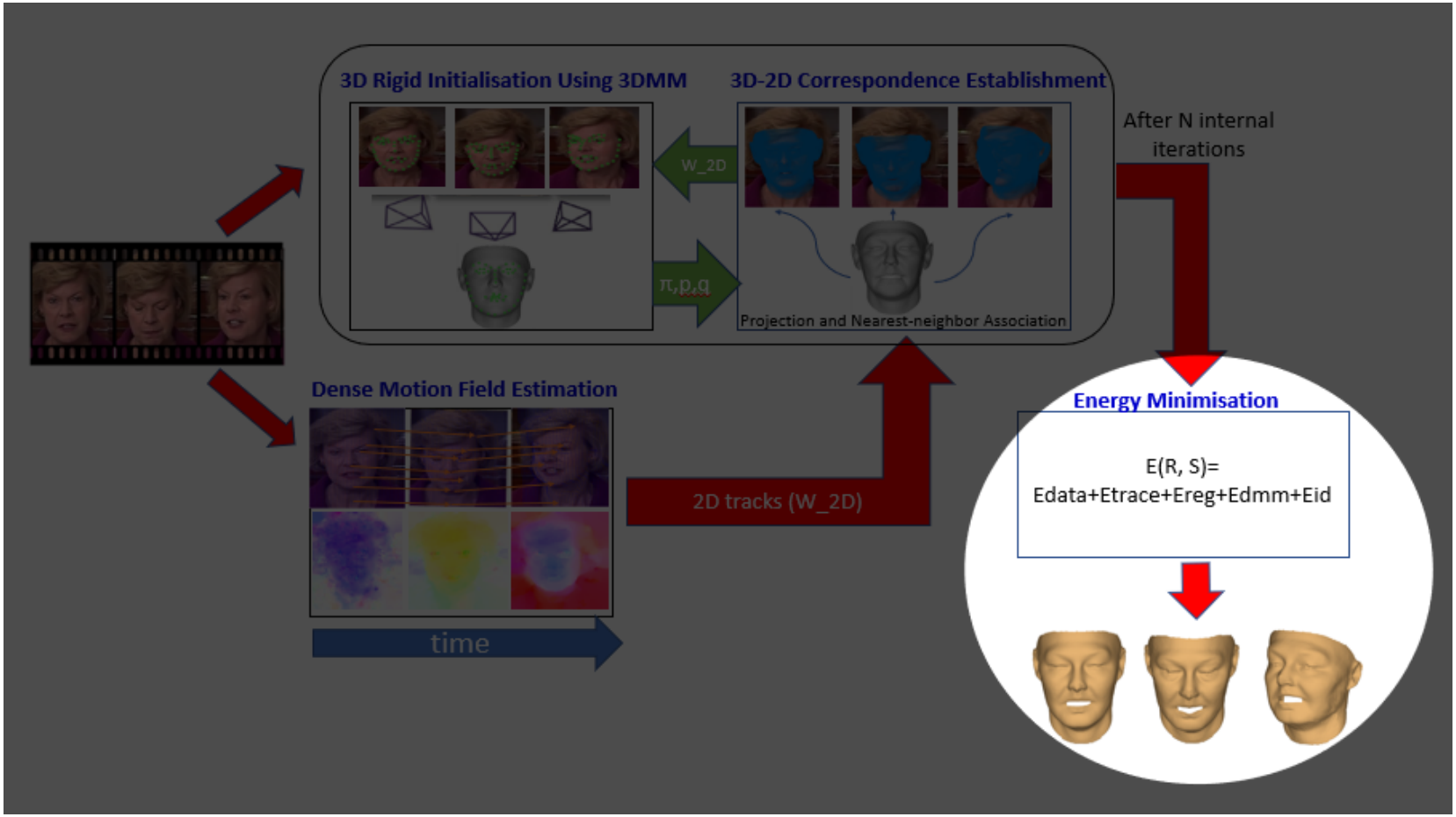
3D-2D correspondence Establishment

Aim

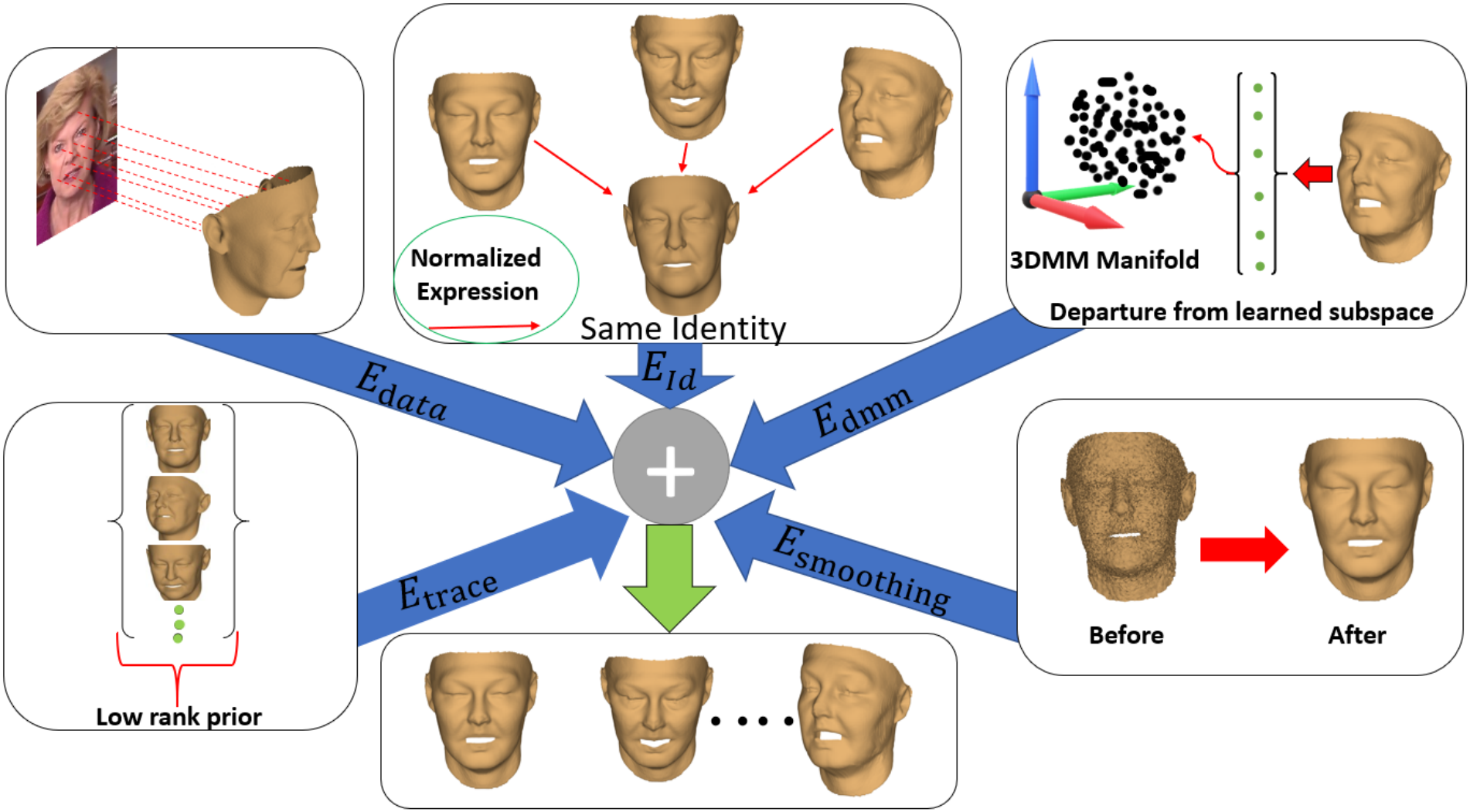
- Given the following:
  1. Dense 2D tracks extracted by the multi-frame subspace flow
  2. Rigid 3D face initialisation and camera matrix
- Match each 3D point on rigid initialisation to its 2D point track by projection and nearest-neighbor association
- Recompute the rigid initialisation and refine the correspondence iteratively



Proposed Framework



Energy Formulation



Energy Formulation

Energy Minimisation

- Alternate between the estimation of the motion matrix $\mathbf{R}$ and the shape matrix $\mathbf{S}$, keeping the other unalterable
- Minimising w.r.t $\mathbf{R}$ while fixing $\mathbf{S}$ depends on $(E_{data}) \rightarrow$ Levenberg-Marquardt
- Minimising w.r.t $\mathbf{S}$ while fixing $\mathbf{R} \rightarrow$ primal-dual algorithm

Experimental Results

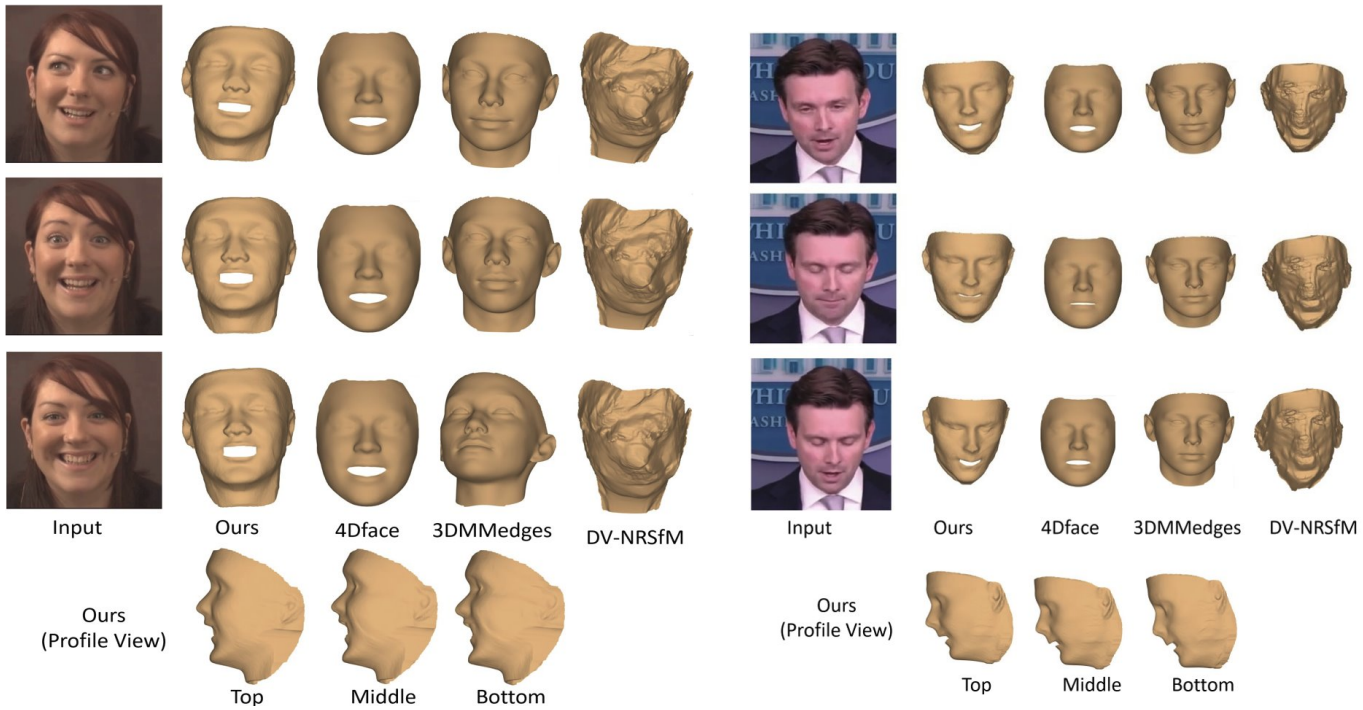
Methods we compare against

- 3DMMedges: A. Bas, et al. ACCV 2016 (3DMM-based approach)
- 4Dface: P. Huber, et al. ACM SIGGRAPH 2016 (3DMM- and media-based super resolution)
- DV-NRSfM: R. Garg, et al. CVPR 2013 (dense variational reconstruction)

Experimental Results

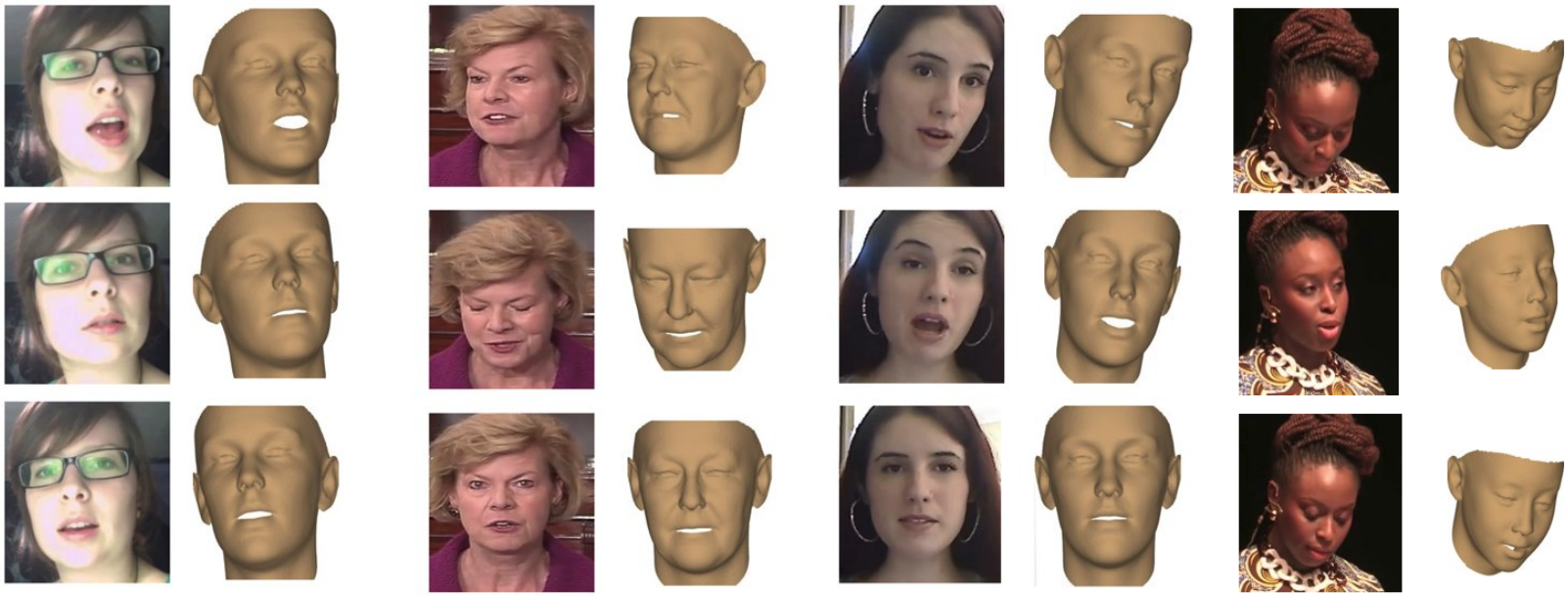
Qualitative Evaluation

- In-the-wild videos from the 300VW dataset were used.



Experimental Results

Qualitative Evaluation

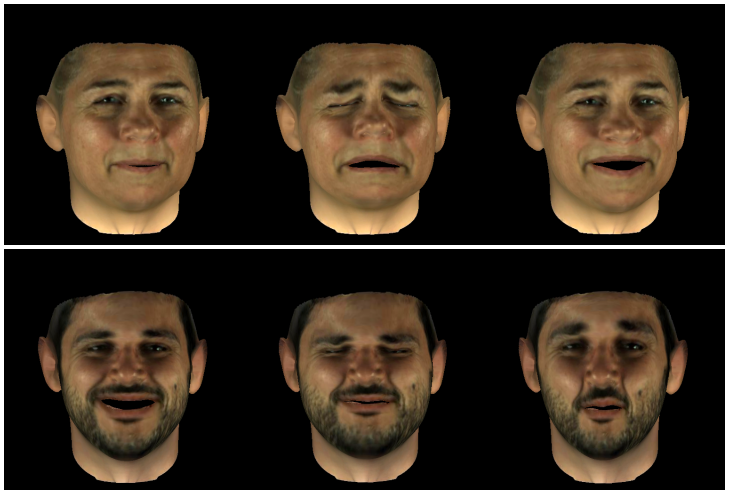


Reconstruction results of our method

Experimental Results

Quantitative Evaluation

- Two synthetic videos were generated with the help of a high-resolution DI4DTM face scanner
- Each video consists of 440 frames exhibiting various natural expressions and head pose variations

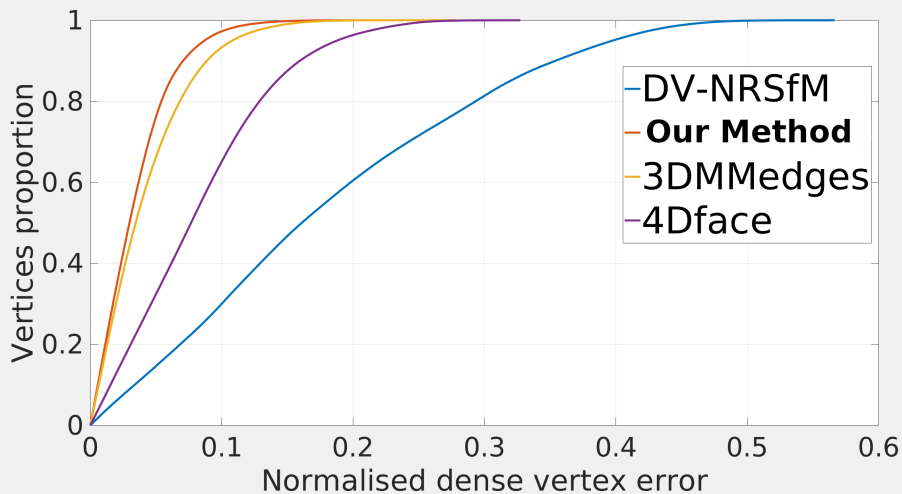


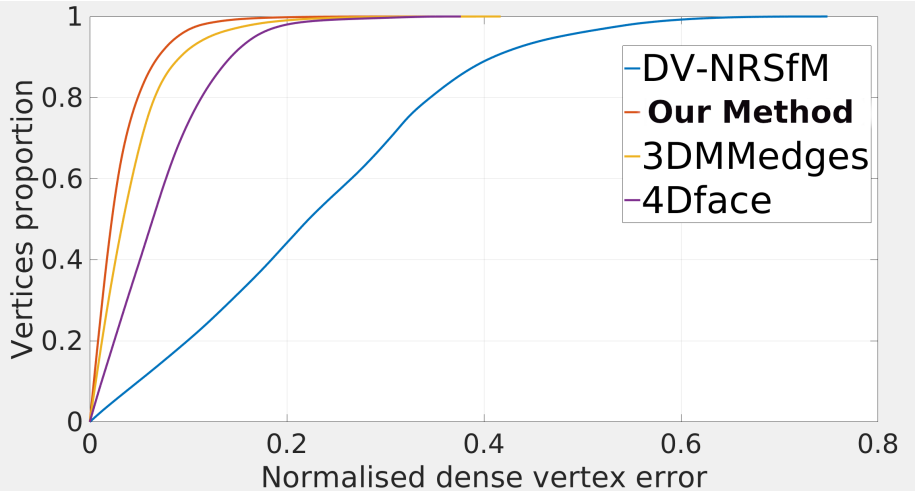
Three frames from the two synthetic videos we generate in this paper

Experimental Results

Quantitative Evaluation

- Comparison against some State-of-the-art methods
- Average per-vertex error was computed as a metric



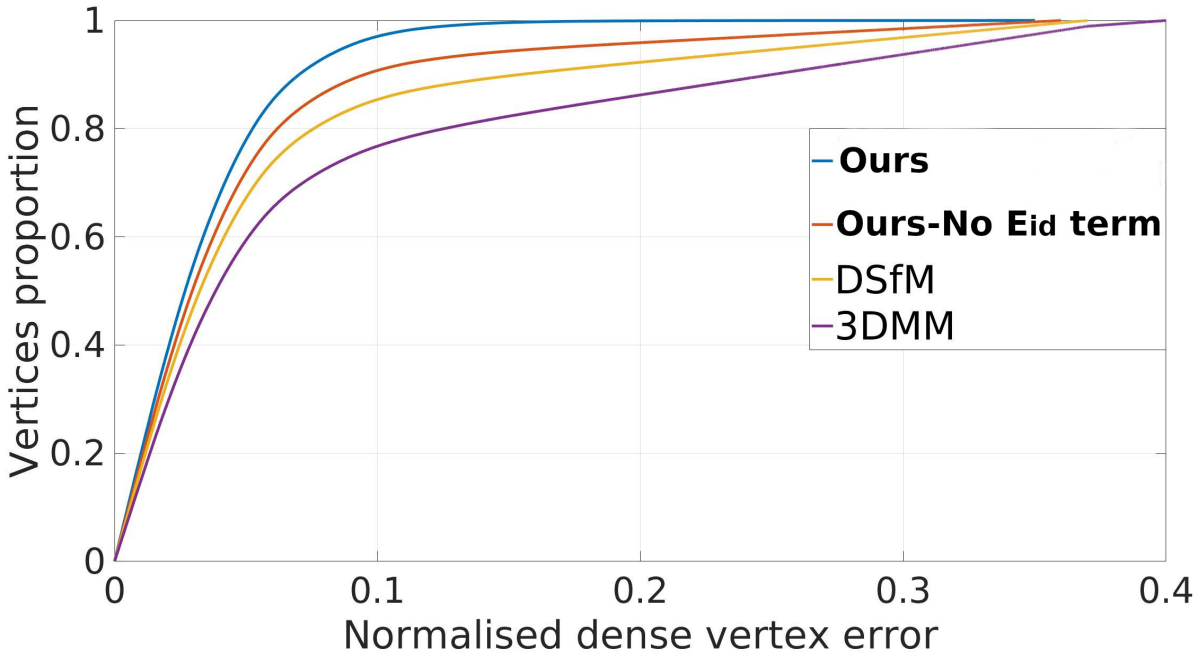


Average per-vertex error. Methods: DSfM-3DMM (ours), DV-NRSfM (R. Garg. et. al. CVPR 2013), 3DMMedges (A. Bas, et. al. ACCV 2016), 4Dface (P.Huber, et.al ACM SIGGRAPH 2016)

Experimental Results

Effect of E_{id} and E_{dmm} Energy Terms

- 1000 synthetic frames were generated
- Average per-vertex discrepancy was used again as a metric



Quantitative evaluation of E_{id} and E_{dmm} on 1000 synthetic frames

Conclusion

- A solution for the problem of 4D face reconstruction and tracking from monocular videos has been proposed
- Combine both Dense Nonrigid Structure from Motion (DNSfM) and 3D Morphable Models (3DMM)
- Quantitative and qualitative evaluations have been conducted showing the potential of the proposed method
- Effect of 3DMM subspace soft constraints was evaluated and shown to offer a rewarding combination

Future Work

- GPU implementation of our framework to expedite the required computations
- Using our method for generating pseudo ground truth for large-scale dataset of videos

Thank You!

(Loading Video...)