

DeepFaceFlow: In-the-wild Dense 3D Facial Motion Estimation

Mohammad Rami Koujan[1,4], Anastasios Roussos[1,3,4], Stefanos Zafeiriou[2,4]

University of Exeter[1], Imperial College London[2], FORTH-ICS[3], Facesoft.io[4]



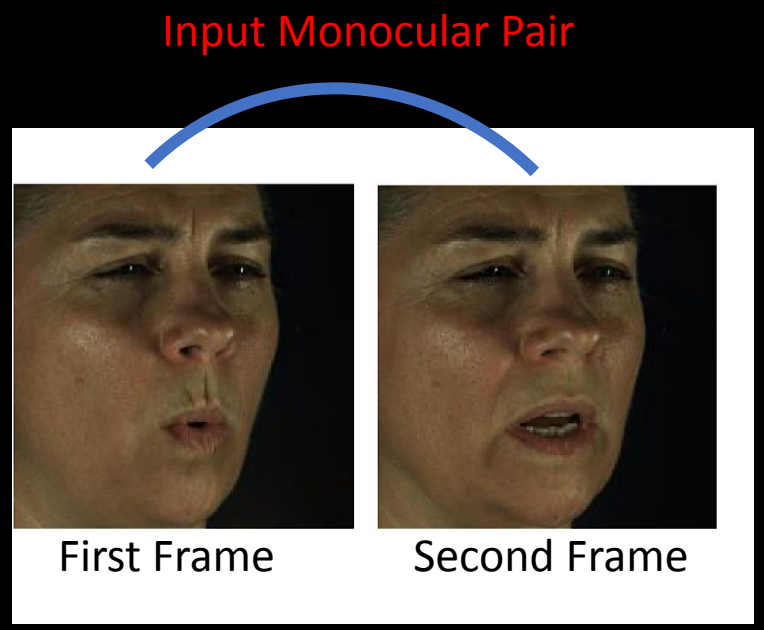






Overview

Overview

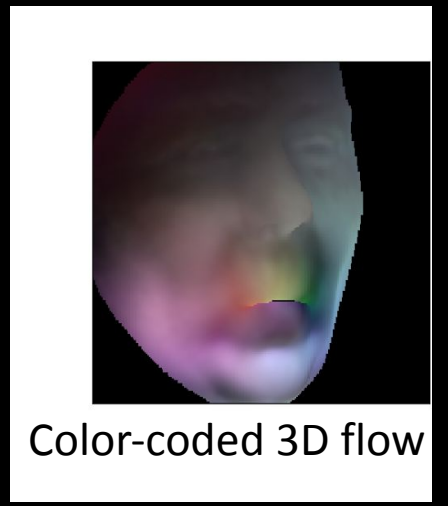


Overview

Input Monocular Pair



Output image



Overview

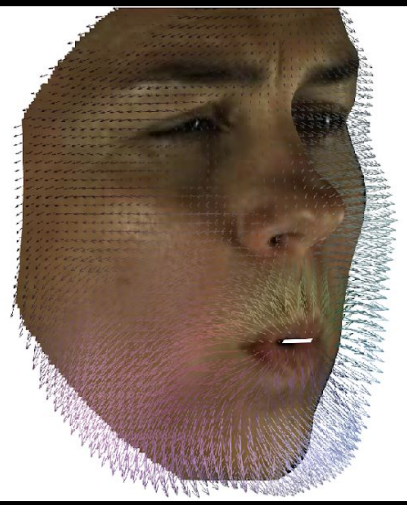
Input Monocular Pair



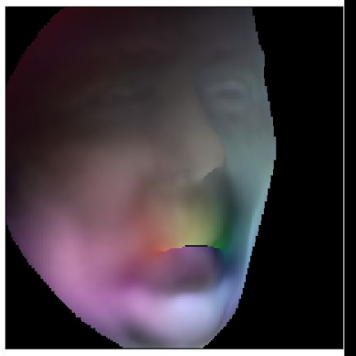
First Frame

Second Frame

3D Shape of first frame
+ Estimated 3D Motion Vectors

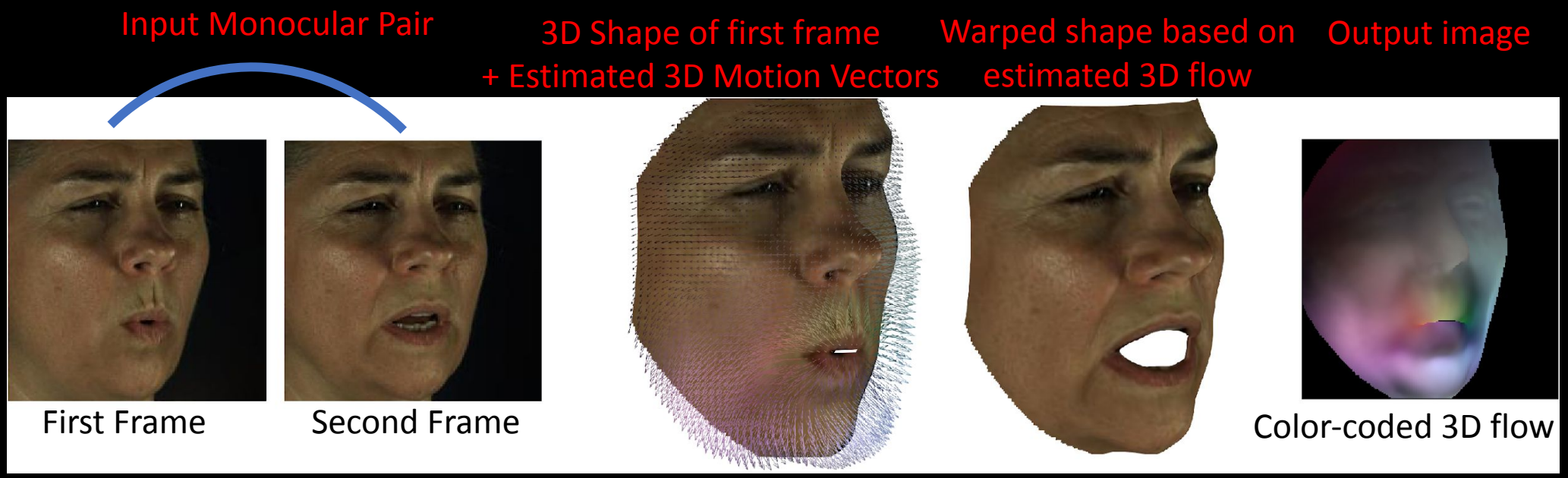


Output image

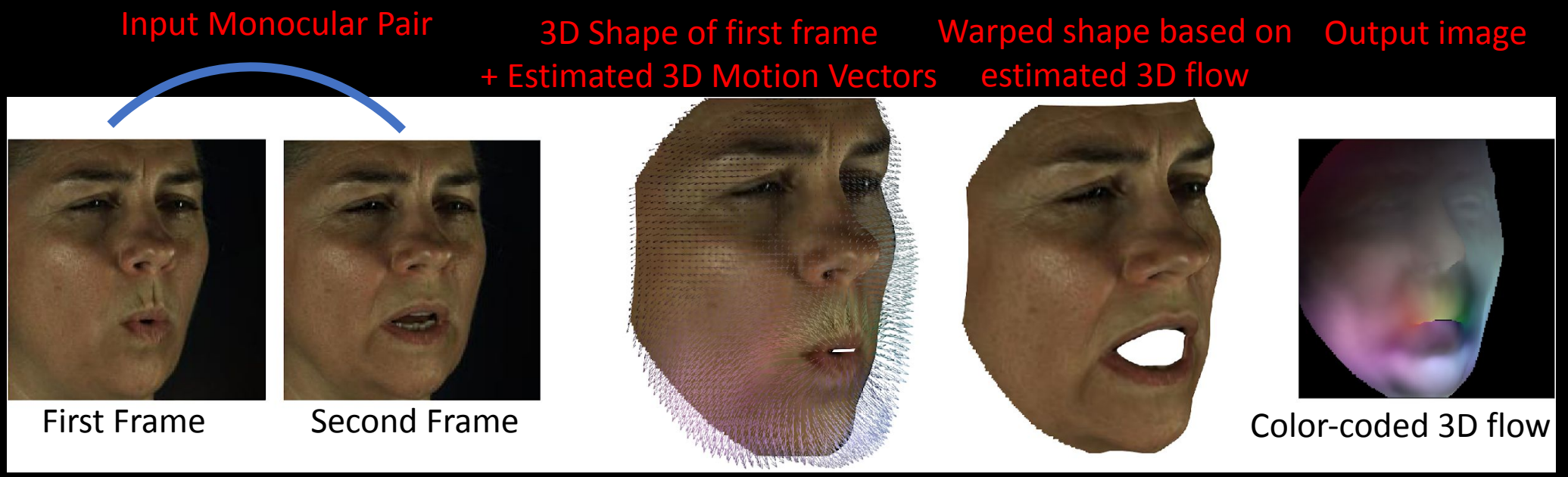


Color-coded 3D flow

Overview


Input Monocular Pair
3D Shape of first frame
+ Estimated 3D Motion Vectors
Warped shape based on
estimated 3D flow
Output image
First Frame
Second Frame
Color-coded 3D flow

Overview



We propose a framework for the high-fidelity dense 3D flow estimation between a pair of monocular facial images

Proposed Framework

Proposed Framework

- Input: only a pair of monocular RGB images

Proposed Framework

- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow

Proposed Framework

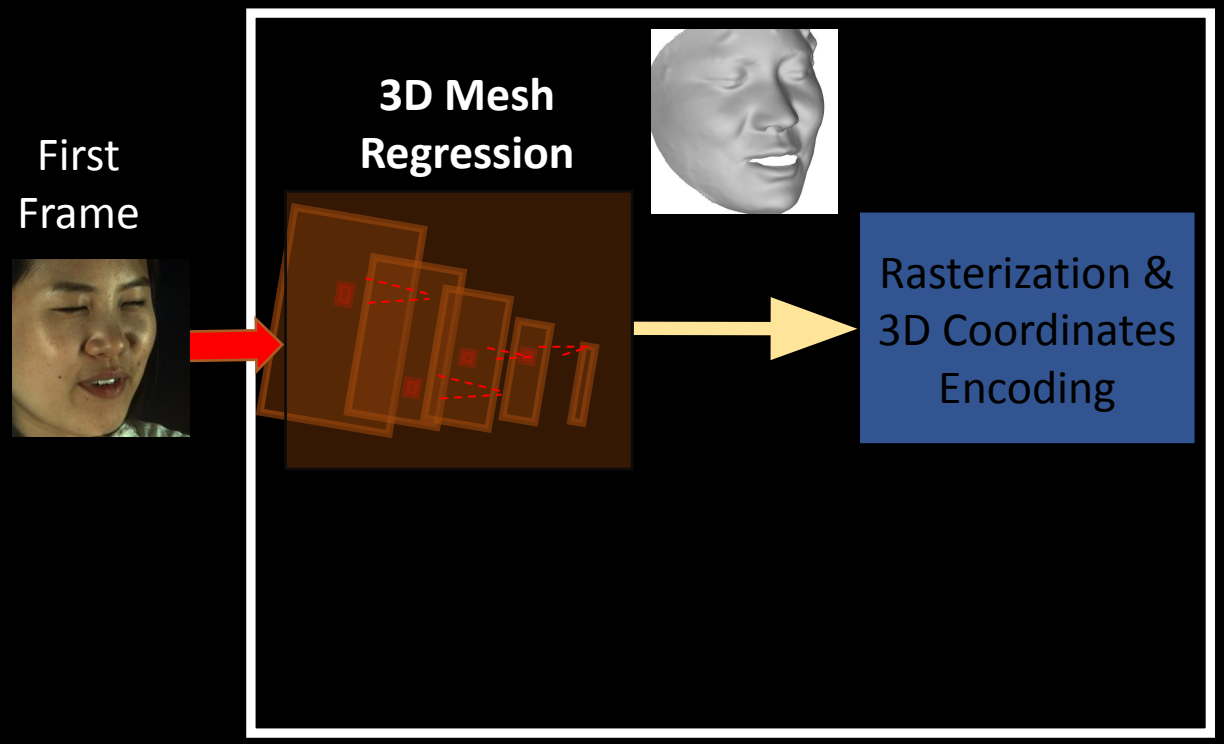
- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow

First Frame



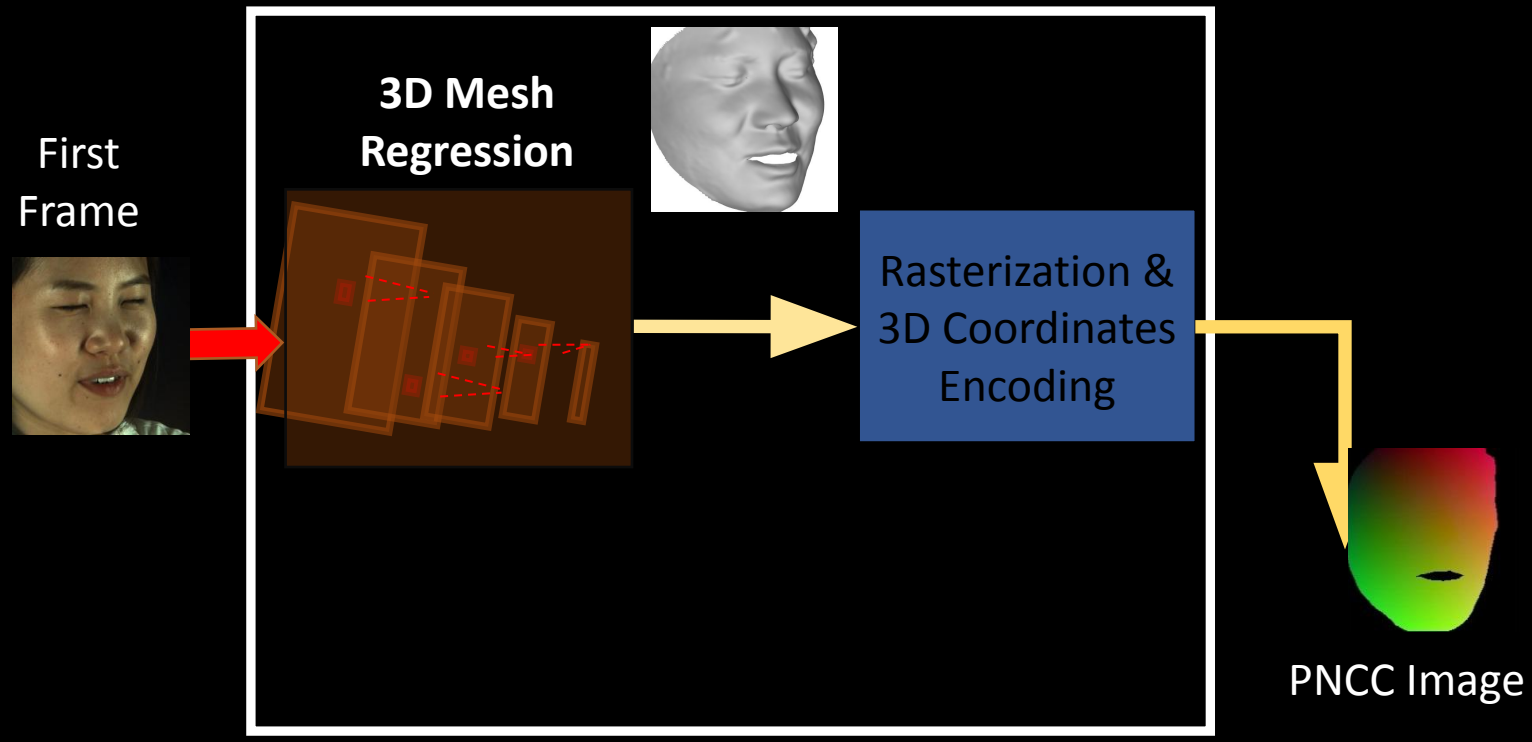
Proposed Framework

- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow



Proposed Framework

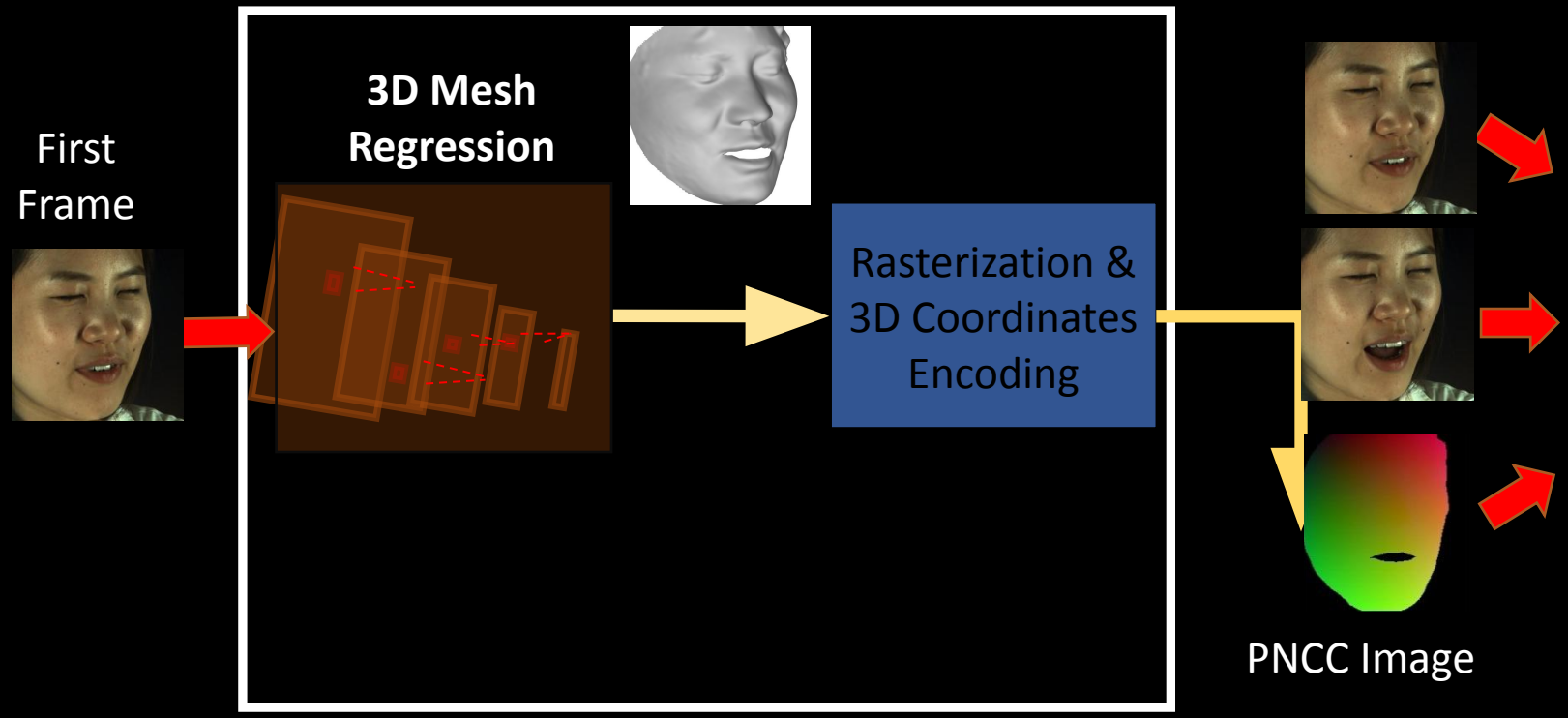
- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow



First stage: 3DMeshReg works as an initialisation for the 3D facial shape in the first frame. This estimation is rasterized in the next step and encoded in an RGB image, termed as Projected Normalised Coordinates Code (PNCC) storing the x-y-z coordinates of each corresponding visible 3D point.

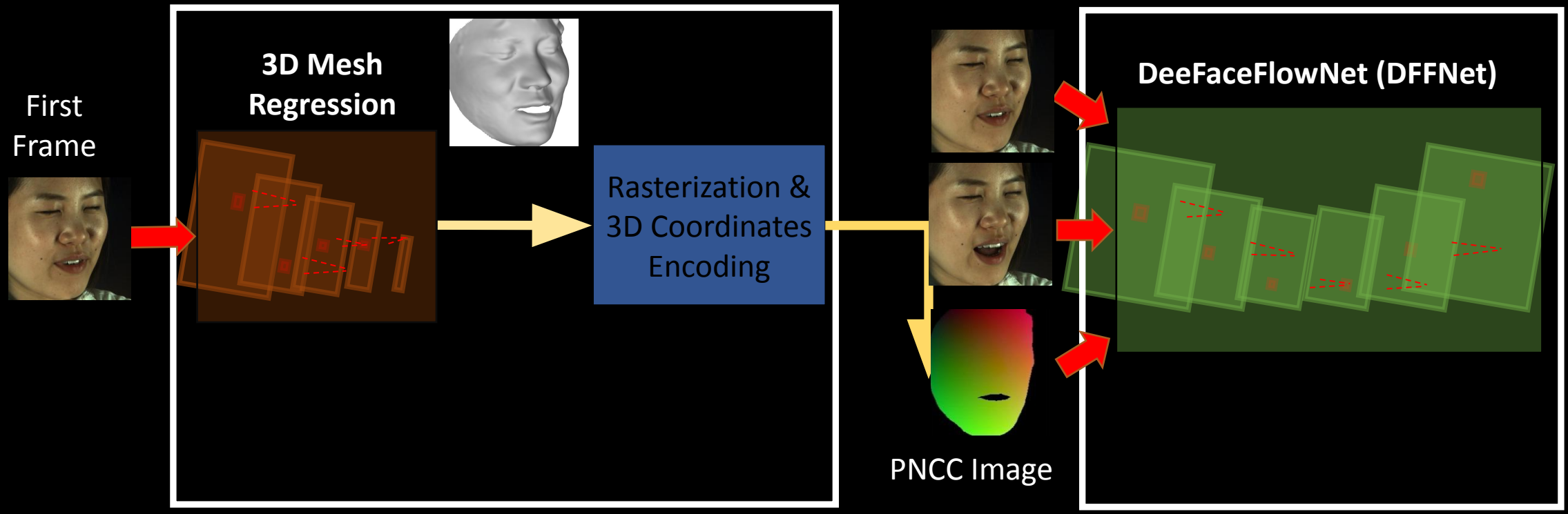
Proposed Framework

- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow



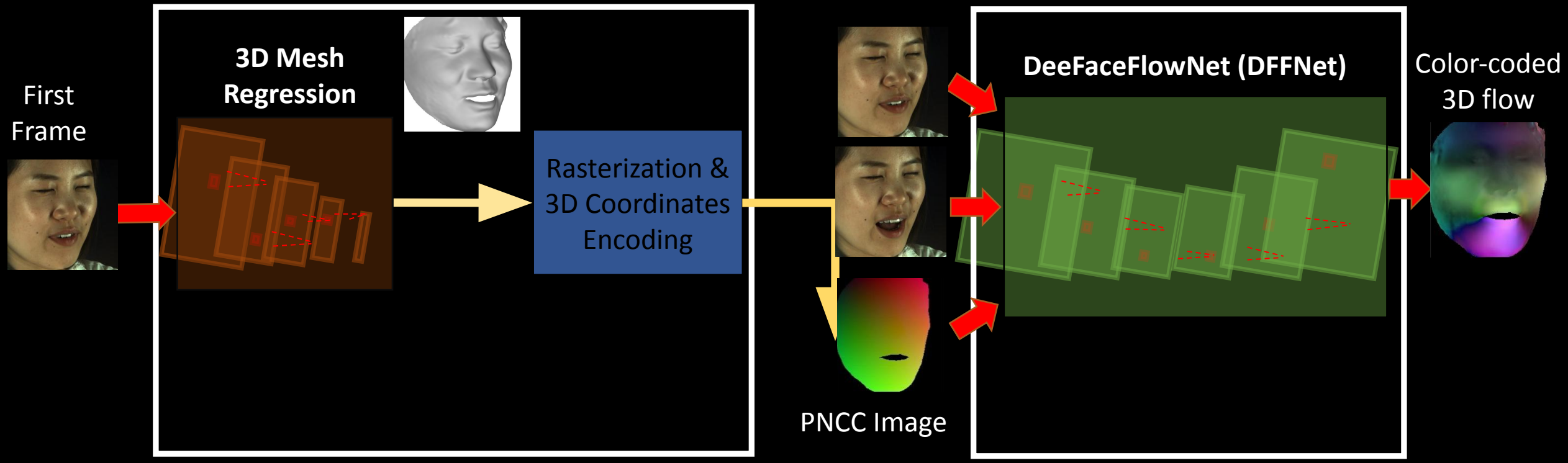
Proposed Framework

- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow



Proposed Framework

- Input: only a pair of monocular RGB images
- Output: 3-channel image encoding the per-pixel 3D flow



Given the pair of images as well as the PNCC, the second stage (DeepFaceFlowNet) estimates the 3D flow using a deep fully-convolutional network (DeepFaceFlowNet).

Application-oriented Results

1- 3D Facial Reconstruction and Tracking Based on Estimated 3D Flow





Application-oriented Results

2- Video-to-video Translation Based on Our Estimated 3D flow and Head2head[1] Framework

[1] M.R. Koujan*, M. Doukas*, A. Roussos, S. Zafeiriou. “Head2Head: Video-based Neural Head Synthesis.” 2020 15th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2020)



Source video



PNCC
(used by both methods)



Reenactment by
FNet2

Reenactment by
Our Method

[FNet2]: Ilg, Eddy, et al. "Flownet 2.0: Evolution of optical flow estimation with deep networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017

For more details and results, please check the project webpage:
https://github.com/mrkoujan/DeepFaceFlow