Spatio-Temporally Consistent Face Mesh Reconstruction on Videos
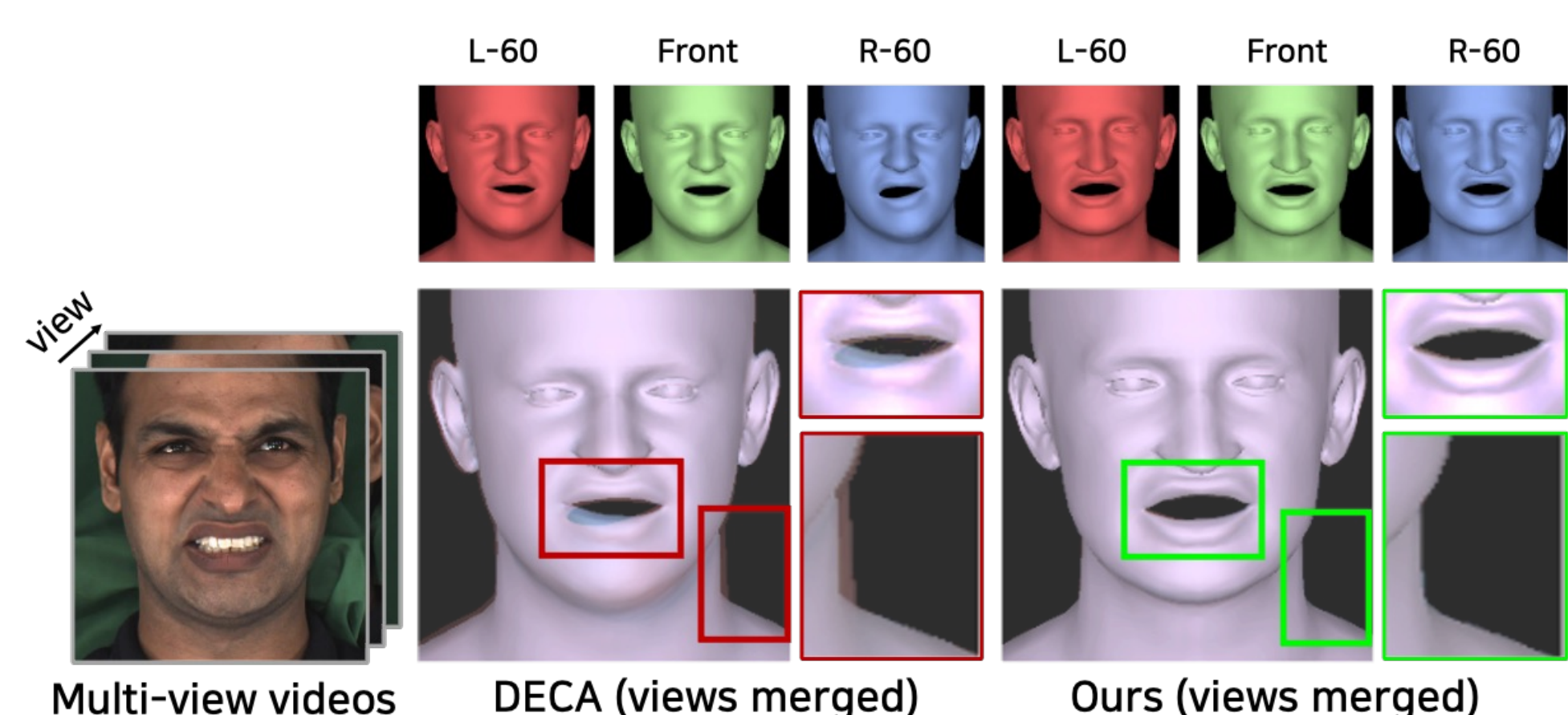
Kim Youwang   Lee Hyun*   Kim Sung-Bin*   Suekyeong Nam   Janghoon Ju   Tae-Hyun Oh

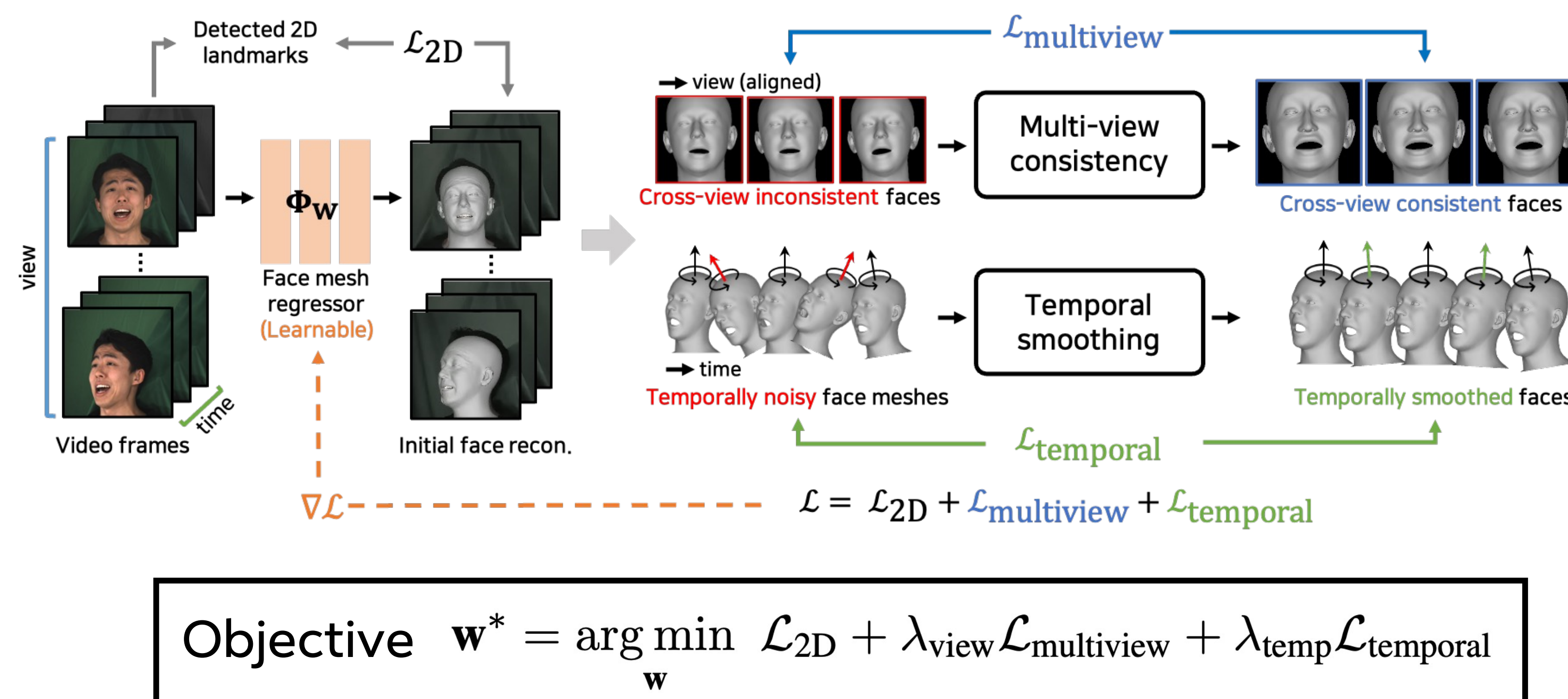POSTECH AMi Algorithmic Machine Intelligence Lab   KRAFTON

Overview

- The goal of 3D face mesh reconstruction:

  2D images, videos → accurate & consistent 3D faces.

- Multi-view videos: severely occluded faces

 → most neural models fail to recon. accurate faces.

Multi-view videos   DECA (views merged)   Ours (views merged)

STREAM optimization

$$\mathcal{L} = \mathcal{L}_{2D} + \mathcal{L}_{multiview} + \mathcal{L}_{temporal}$$

Objective $\mathbf{w}^* = \arg\min_{\mathbf{w}} \ \mathcal{L}_{2D} + \lambda_{view}\mathcal{L}_{multiview} + \lambda_{temp}\mathcal{L}_{temporal}$
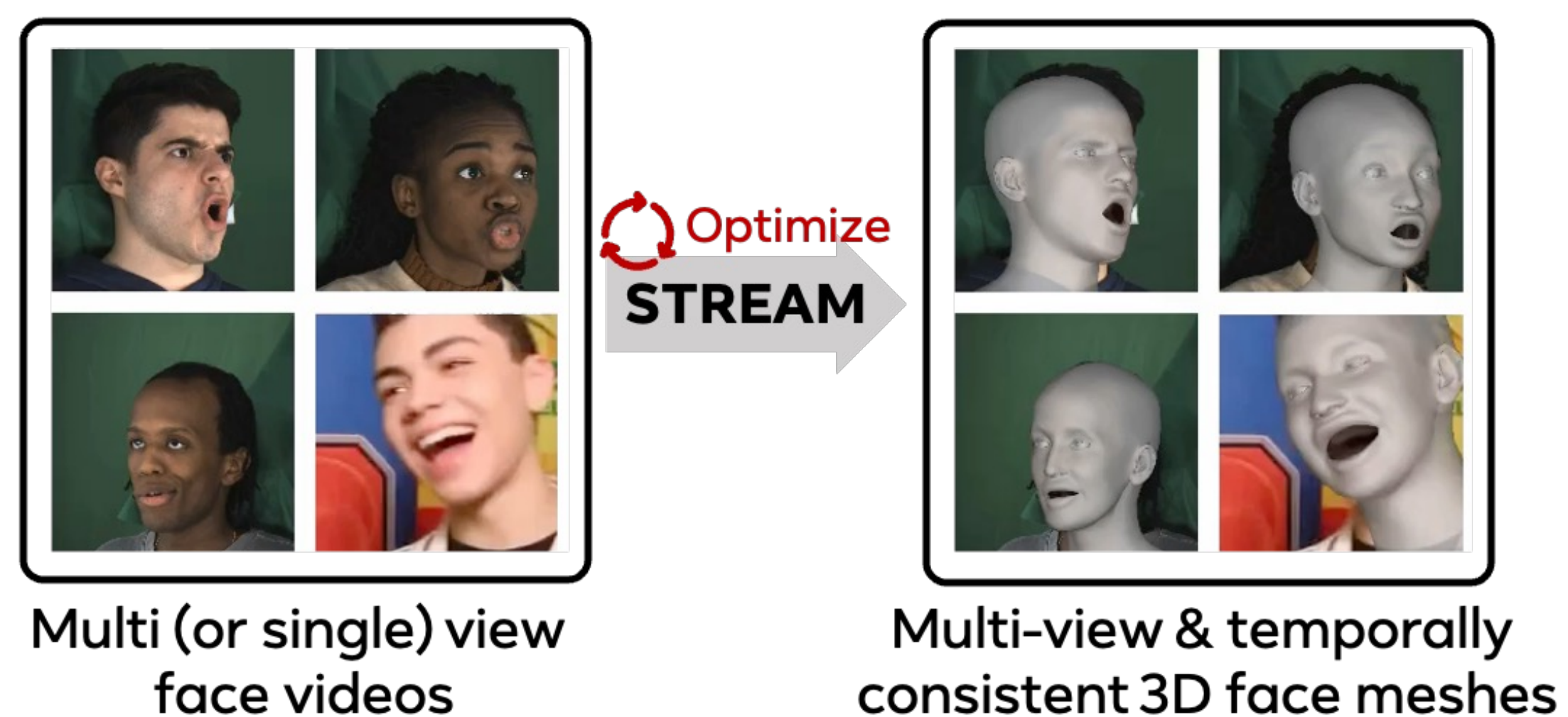
Multi-view bootstrapping

- Vertex level merging of per-view mesh predictions:
  - Higher weight on visible vertices → occlusion robust.
 - Serves as a pseudo-supervision during optimization.
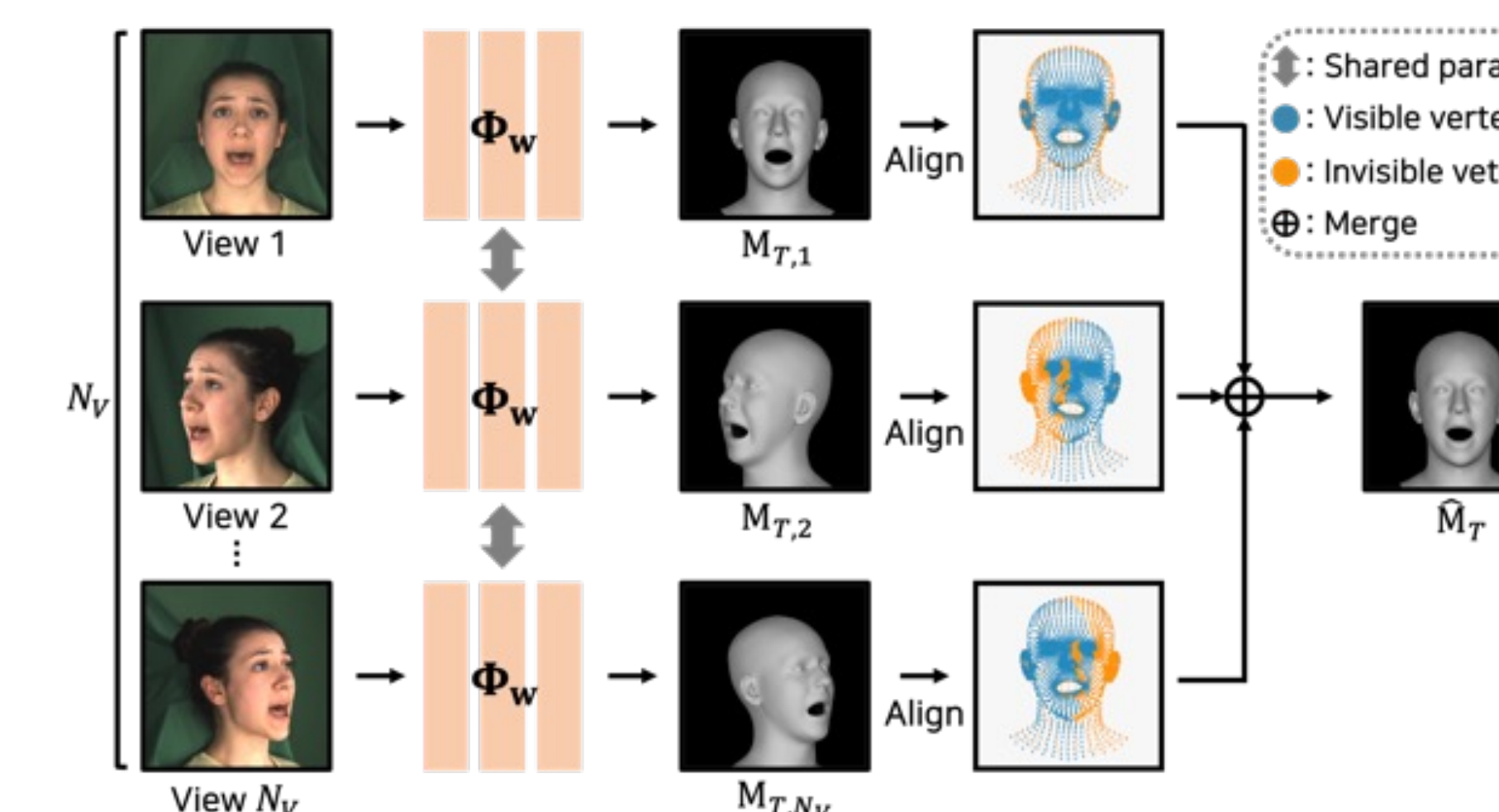
Contributions

- **STREAM**: a **3DMM optimization** method to obtain
  - Multi-view consistent facial geometries.
  - Temporally smooth facial motion.

Multi (or single) view face videos   Multi-view & temporally consistent 3D face meshes

- Design **spatio-temporal bootstrapping** methods.
- Propose **evaluation metrics** for 4D face meshes.

Multi-view consistent 3D faces

Input   DECA   STREAM

Down   Front   L-30   L-60   R-30   R-60   Top
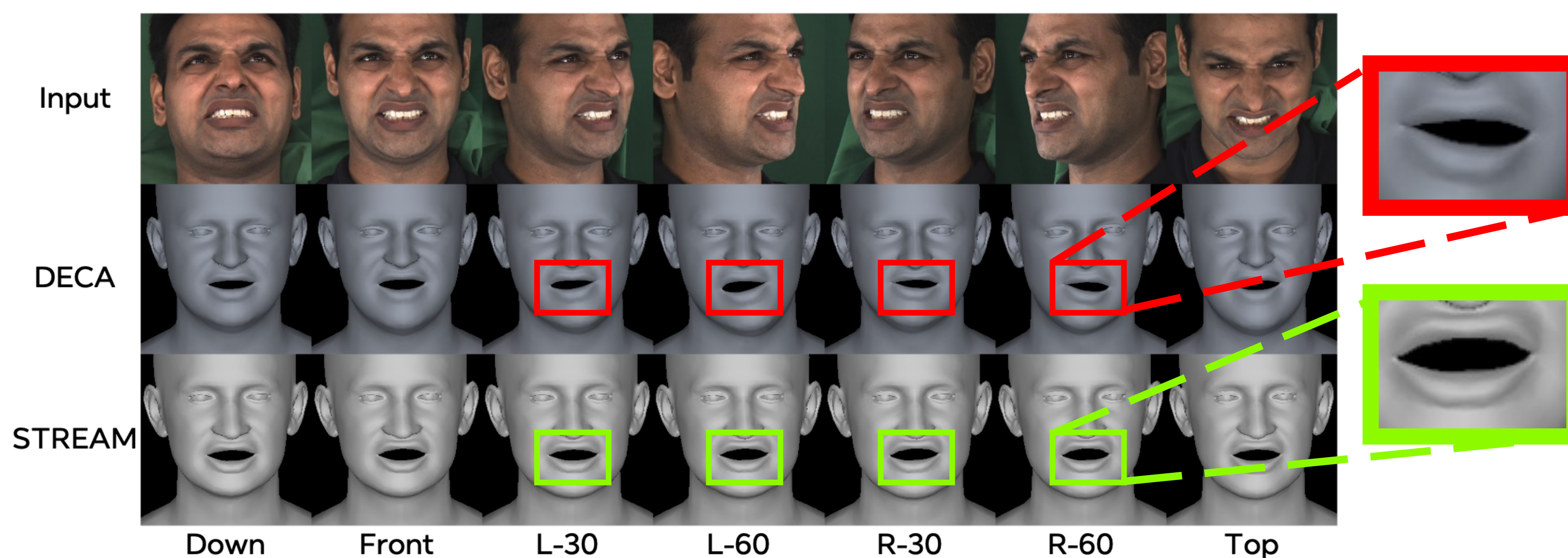
STREAM optim. achieves

1. Multi-view consistent face geometry.
2. Occlusion robust 3D reconstruction.

Cross-view Vertex Distances

DECA

	Down	Front	L-30	L-60	R-30	R-60	Top
Down	0.000						
Front	0.135	0.000					
L-30	0.181	0.165	0.000				
L-60	0.248	0.251	0.127	0.000			
R-30	0.178	0.151	0.198	0.260	0.000		
R-60	0.216	0.229	0.235	0.251	0.138	0.000	
Top	0.226	0.170	0.239	0.319	0.209	0.273	0.000
	Down	Front	L-30	L-60	R-30	R-60	Top

STREAM (Ours)

	Down	Front	L-30	L-60	R-30	R-60	Top
Down	0.000						
Front	0.079	0.000					
L-30	0.107	0.087	0.000				
L-60	0.122	0.112	0.063	0.000			
R-30	0.108	0.086	0.103	0.116	0.000		
R-60	0.123	0.110	0.113	0.111	0.065	0.000	
Top	0.114	0.081	0.108	0.126	0.108	0.125	0.000
	Down	Front	L-30	L-60	R-30	R-60	Top

Qualitative results: 3D faces on videos

Take-home message

- **STREAM**: a multi-view and temporal aware optimization for

  3D face reconstruction on videos.

- **Breakthrough**: simple bootstrapping for 3D face meshes.

- **Application**: dataset of accurate 3D faces for internet videos.