Egocentric Videoconferencing - Supplemental Document

MOHAMED ELGHARIB*†, Max Planck Institute for Informatics, SIC
MOHIT MENDIRATTA*, Max Planck Institute for Informatics, SIC
JUSTUS THIES and MATTHIAS NIESSNER, Technical University of Munich
HANS-PETER SEIDEL and AYUSH TEWARI, Max Planck Institute for Informatics, SIC
VLADISLAV GOLYANIK and CHRISTIAN THEOBALT, Max Planck Institute for Informatics, SIC

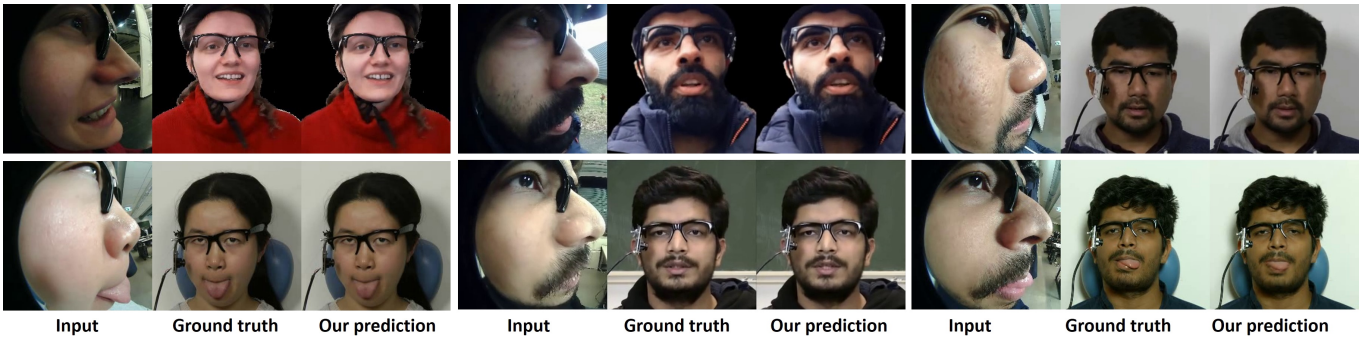


Fig. 1. We present an approach for hands-free videoconferencing. Given the view of an egocentric camera, that is attached to a eye-glasses frame, we predict a frontalised video stream which is common in videoconferencing.

We introduce a method for egocentric videoconferencing that enables hands-free video calls, for instance by people wearing smart glasses or other mixed-reality devices. Videoconferencing portrays valuable non-verbal communication and face expression cues, but usually requires a front-facing camera. Using a frontal camera in a hands-free setting when a person is on the move is impractical. Even holding a mobile phone camera in the front of the face while sitting for a long duration is not convenient. To overcome these issues, we propose a low-cost wearable egocentric camera setup that can be integrated into smart glasses. Our goal is to mimic a classical video call, and therefore, we transform the egocentric perspective of this camera into a front facing video. To this end, we employ a conditional generative adversarial neural network that learns a transition from the highly distorted egocentric views to frontal views common in videoconferencing. Our approach learns to transfer expression details directly from the egocentric view without using a complex intermediate parametric expressions model, as it is used by related face reenactment methods. We successfully handle subtle expressions, not easily captured by parametric blendshape-based solutions, e.g., tongue movement, eye movements, eye blinking, strong expressions and depth varying movements. To get control over the rigid head movements in the target view, we condition the generator on synthetic renderings of a moving neutral face. This allows us to synthesis results at different head poses. Our technique produces temporally smooth video-realistic renderings in real-time using a video-to-video translation network in conjunction with a temporal discriminator. We demonstrate the improved capabilities of our technique by comparing against related state-of-the art approaches.

CCS Concepts: • **Computing methodologies** → **Computer graphics**; *Image manipulation*; *Animation*; *Rendering*.

Additional Key Words and Phrases: Videconferencing, Egocentric, Face Frontalisation, Neural Rendering, Reenactment, Face.

ACM Reference Format:
Mohamed Elgharib, Mohit Mendiratta, Justus Thies, Matthias Nießner, Hans-Peter Seidel, Ayush Tewari, Vladislav Golyanik, and Christian Theobalt. 2020. Egocentric Videoconferencing - Supplemental Document. *ACM Trans. Graph.* 39, 6, Article 268 (December 2020), 2 pages. https://doi.org/10.1145/3414685.3417808

*Indicates equal contribution
†Contact us through: elgharib@mpi-inf.mpg.de

Authors' addresses: Mohamed Elgharib, elgharib@mpi-inf.mpg.de, Max Planck Institute for Informatics, SIC; Mohit Mendiratta, mmendira@mpi-inf.mpg.de, Max Planck Institute for Informatics, SIC; Justus Thies, justus.thies@tum.de; Matthias Nießner, niessner@tum.de, Technical University of Munich; Hans-Peter Seidel, hpseidel@mpi-sb.mpg.de; Ayush Tewari, atewari@mpi-inf.mpg.de, Max Planck Institute for Informatics, SIC; Vladislav Golyanik, golyanik@mpi-inf.mpg.de; Christian Theobalt, theobalt@mpi-inf.mpg.de, Max Planck Institute for Informatics, SIC.

0730-0301/2020/12-ART268
https://doi.org/10.1145/3414685.3417808

A APPENDIX

Tab. 1–2 lists the sequences used in our experiments. For each sequence, we indicate the total number of frames. We use 7,500 frames for training our technique, 2,500 frames for validation and the rest for testing.

Table 1. List of sequences used in our experiments.

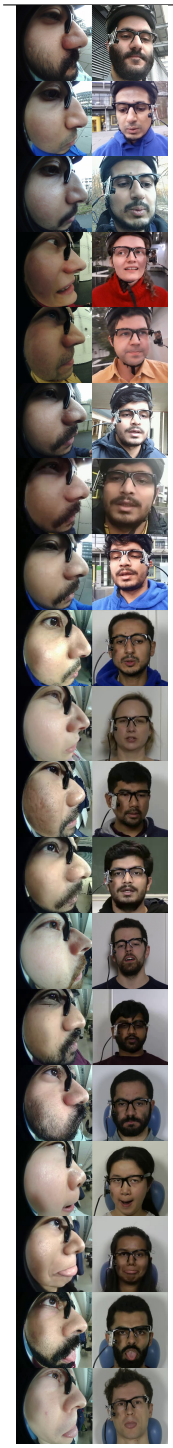
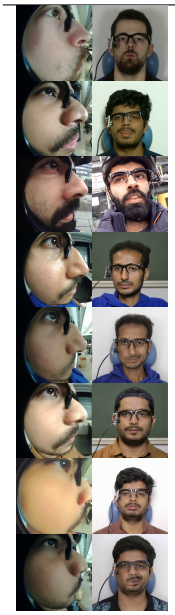
Image	Name	Type	Number of frames
	ID1V1.	outdoor dynamic	13500
	ID2V1.	outdoor dynamic	5700
	ID2V2.	outdoor dynamic	13640
	ID3.	indoor dynamic	13050
	ID4.	indoor dynamic	13450
	ID5V1.	outdoor dynamic	12600
	ID5V2.	outdoor dynamic	14750
	ID5V3.	outdoor dynamic	5796
	ID2V3.	sitting scenario	14500
	ID6.	sitting scenario	13800
	ID7.	sitting scenario	14730
	ID5V4.	sitting scenario	14700
	ID8V1.	sitting scenario	13500
	ID9.	sitting scenario	14260
	ID1V2.	stress expressions	15600
	ID10.	stress expressions	14500
	ID11.	stress expressions	14600
	ID12V1.	stress expressions	14740
	ID13.	stress expressions	14500

Table 2. List of sequences used in our experiments.

Image	Name	Type	Number of frames
	ID8V2.	stress expressions	13880
	ID5V5.	stress expressions	14400
	ID12V2.	outdoor dynamic	17300
	ID2V4.	sitting scenario	16000
	ID2V5.	stress expressions	16000
	ID5V6.	sitting scenario	13000
	ID5V7.	stress expressions	13000
	ID5V8.	stress expressions	13000