



AHG9/AHG16: Showcase for picture fusion for generative face video SEI message

—

JVET-AH0118

SHIV GEHLOT, GUAN-MING SU, PENG YIN, SEAN MCCARTHY, GARY J. SULLIVAN

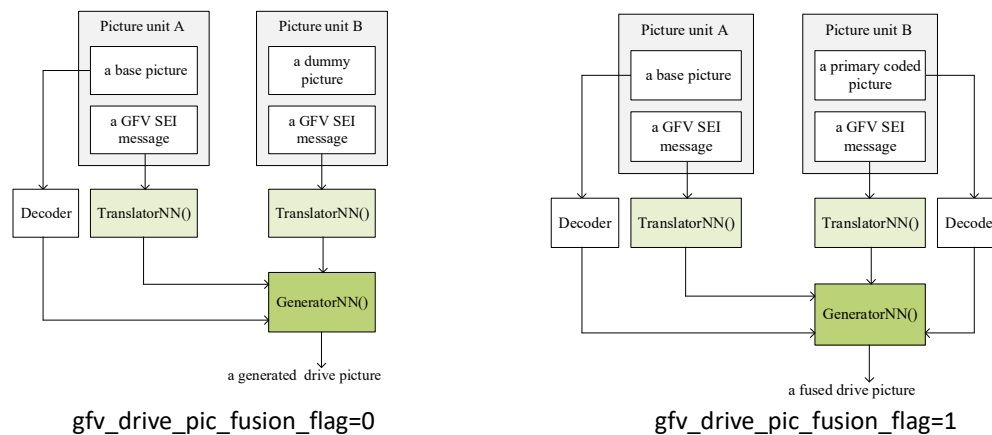
Outline

- Motivation
- Proposal: GFVC with picture fusion
- Showcase
- Conclusions

Motivation



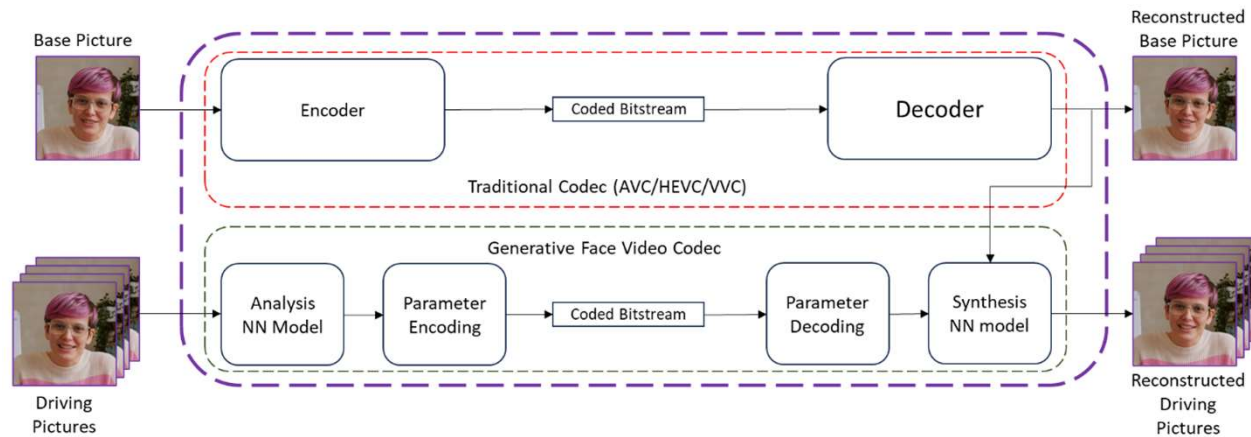
- GFV SEI message now in VSEI TuC (JVET-AG2032)
- Two cases are supported:



- AHG16: GFVC software
 - https://vcgit.hhi.fraunhofer.de/jvet-ahg-gfvc/gfvc_v1/
 - Provides functionality for `gfv_drive_pic_fusion_flag = 0`
- This contribution provides a showcase for `gfv_drive_pic_fusion_flag = 1`

AHG16 GFVC software: GFVC without picture fusion

- Encoder:
 - Base picture is compressed with AVC/HEVC/VVC etc
 - Driving pictures: fed into NN analysis model for key points/compact features extraction, followed by parameter encoding
- Decoder:
 - Base picture: decoded with AVC/HEVC/VVC etc
 - Driving picture: decoded base picture + decoded key points/compact features are fed into NN Synthesis model to generate reconstructed driving picture



Limitation of GFVC without picture fusion

- One observation
 - Entanglement between foreground (face) and background.
 - Background surrounding the face moves with the face.



Original

CFTE without picture fusion (QP=22)

Proposal: GFVC with picture fusion

- Basic idea:
 - Segment foreground (face) and background.
 - Fuse generated face with background.



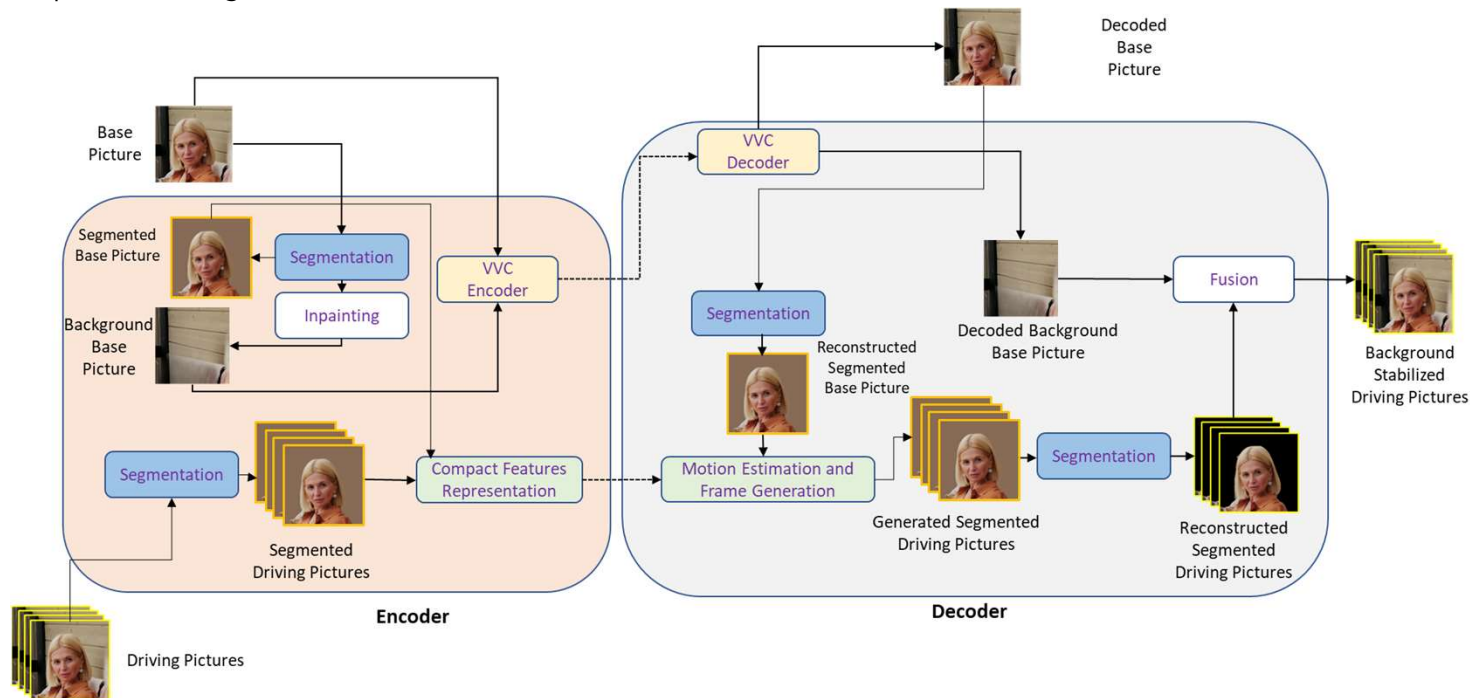
Original

CFTE without picture fusion (QP=22)

CFTE with picture fusion (QP=22)

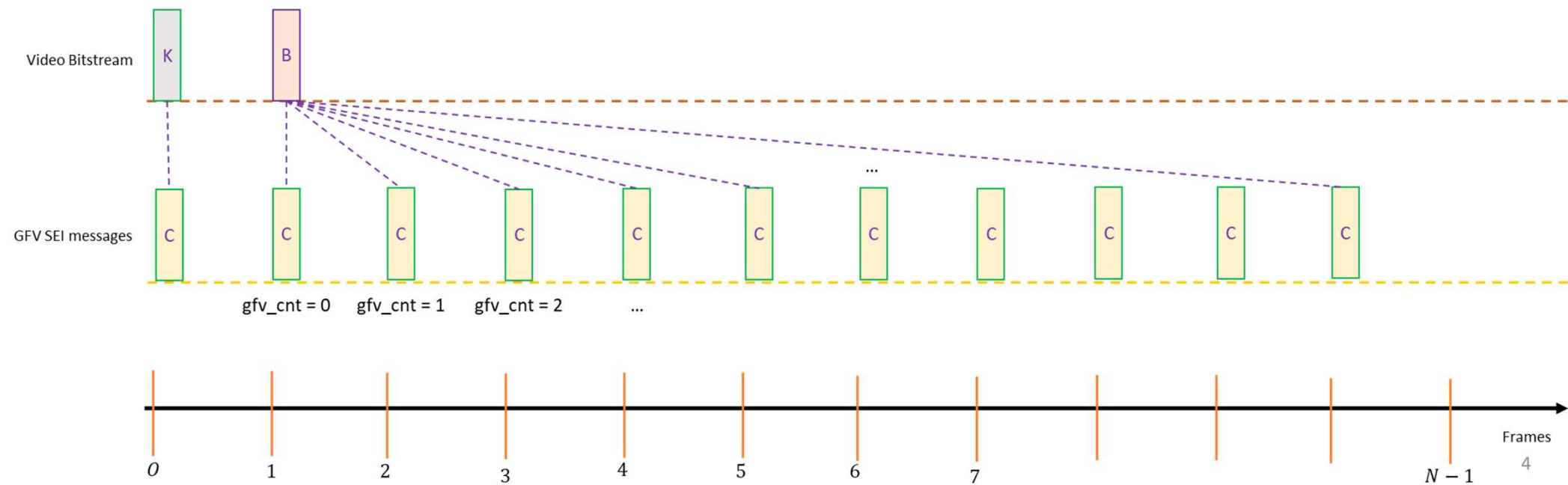
GFVC with picture fusion

- Use picture fusion to mitigate entanglement of face and background
- CFTE is used for GFVC and a static background picture is used as fusion picture
 - Segmentation NN: X. Chen, et al., "Robust Human Matting via Semantic Guidance," ACCV, 2020
 - Inpainting NN: W. Liu, et al., "CoordFill: Efficient High-Resolution Image Inpainting via Parameterized Coordinate Querying", AAAI 2023
- Fusion: alpha blending



Bitstream using static background

- **K**: base picture; **B**: background base picture; **C**: GFV SEI messages
- Bitrate overhead: **B** bitrate (QP22, averages about 6%, overhead is image dependent)



Note: In current GFVC software, **C** is coded using AE.

Showcase: Static Background

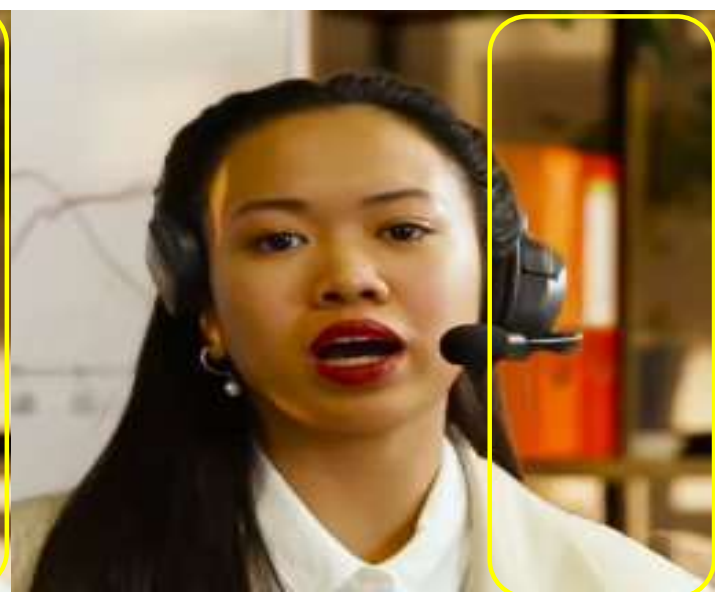
C012



Original

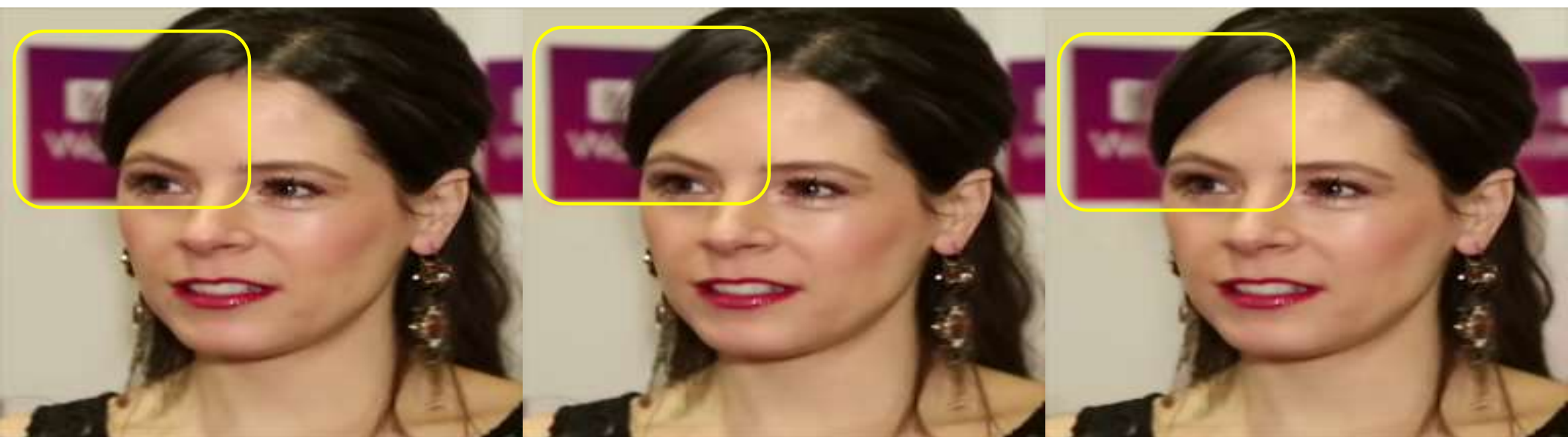


CFTE without picture fusion (QP=22):10.4kbps



CFTE with picture fusion (QP=22):11.3kbps

V001



Original

CFTE without picture fusion (QP=22):5.5kbps

CFTE with picture fusion (QP=22):5.9kbps

V002



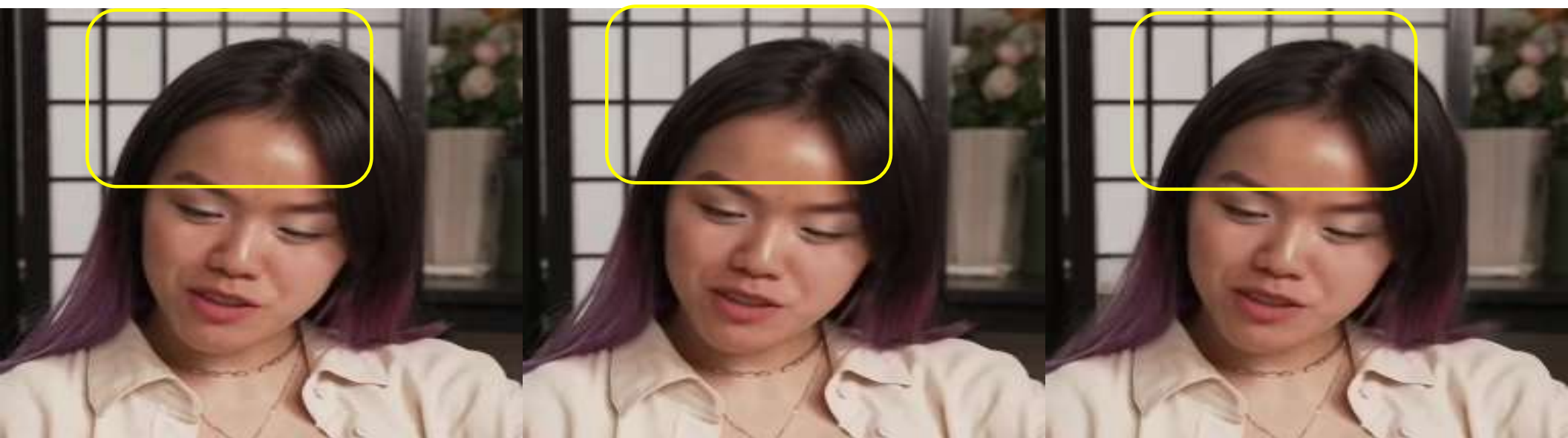
Original

CFTE without picture fusion (QP=22):6.4kbps

CFTE with picture fusion (QP=22): 6.8kbps

CFVQA (16, 17, 18): challenge clips
Dynamic Background

C018

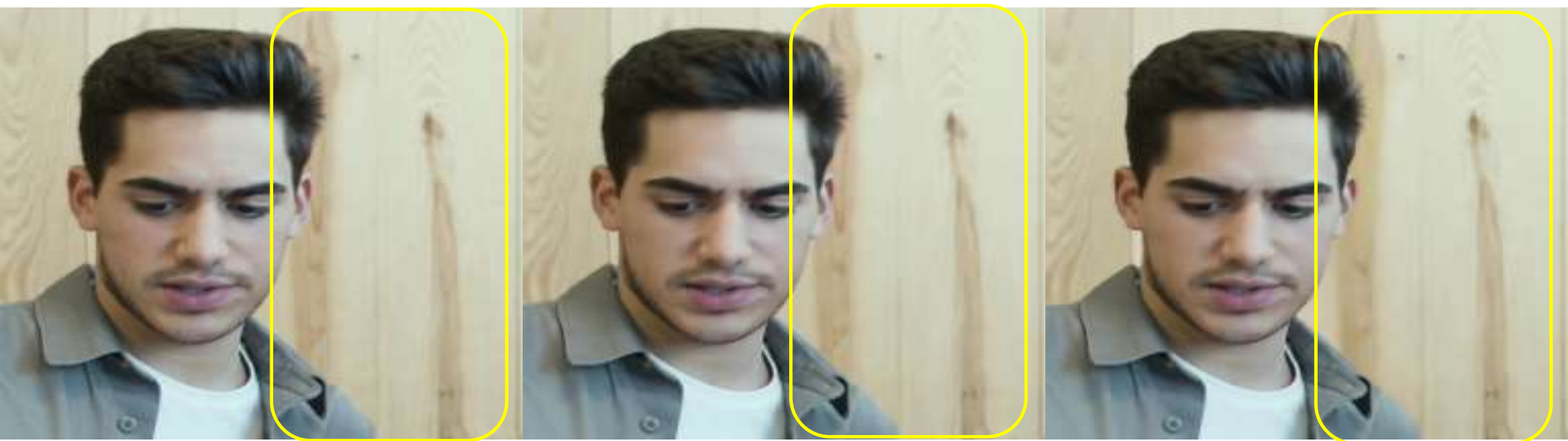


Original

CFTE without picture fusion (QP=22):9.6kbps

CFTE with picture fusion (QP=22):11.3kbps

C016

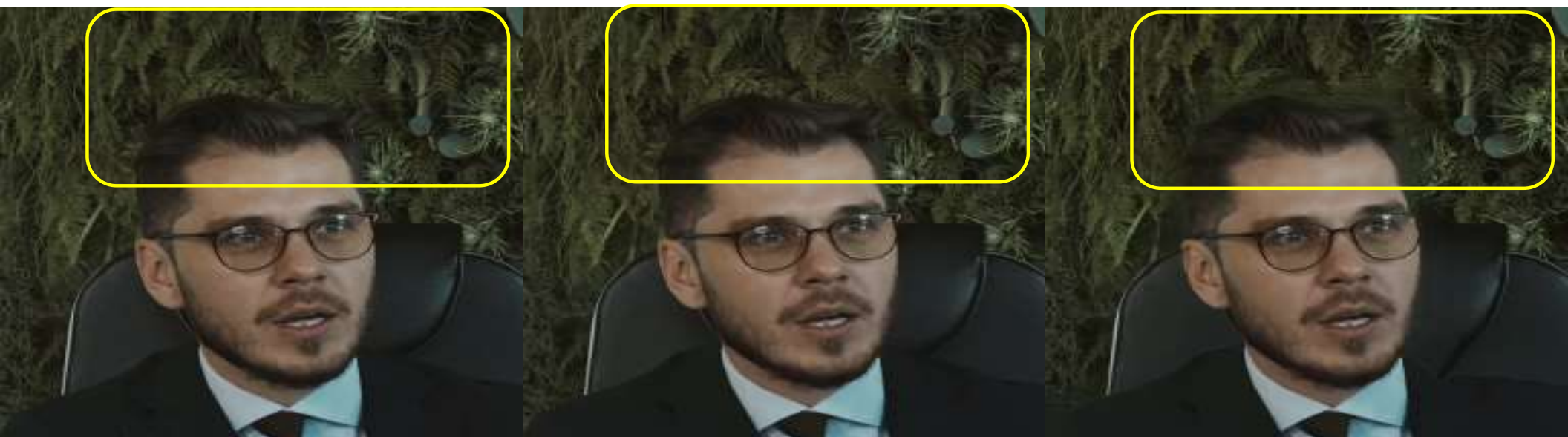


Original

CFTE without picture fusion (QP=22): 8.4kbps

CFTE with picture fusion (QP=22): 8.7kbps

C017



Original

CFTE without picture fusion (QP=22): 17.6kbps

CFTE with picture fusion (QP=22): 18.3kbps

Conclusion

- A showcase for picture fusion for GFV SEI.
 - Illustrate the mitigation of background warping with face
 - Limited to static background
- Propose to further improve picture fusion technology
 - Improve segmentation, inpainting, and fusion module
 - Handle dynamic background