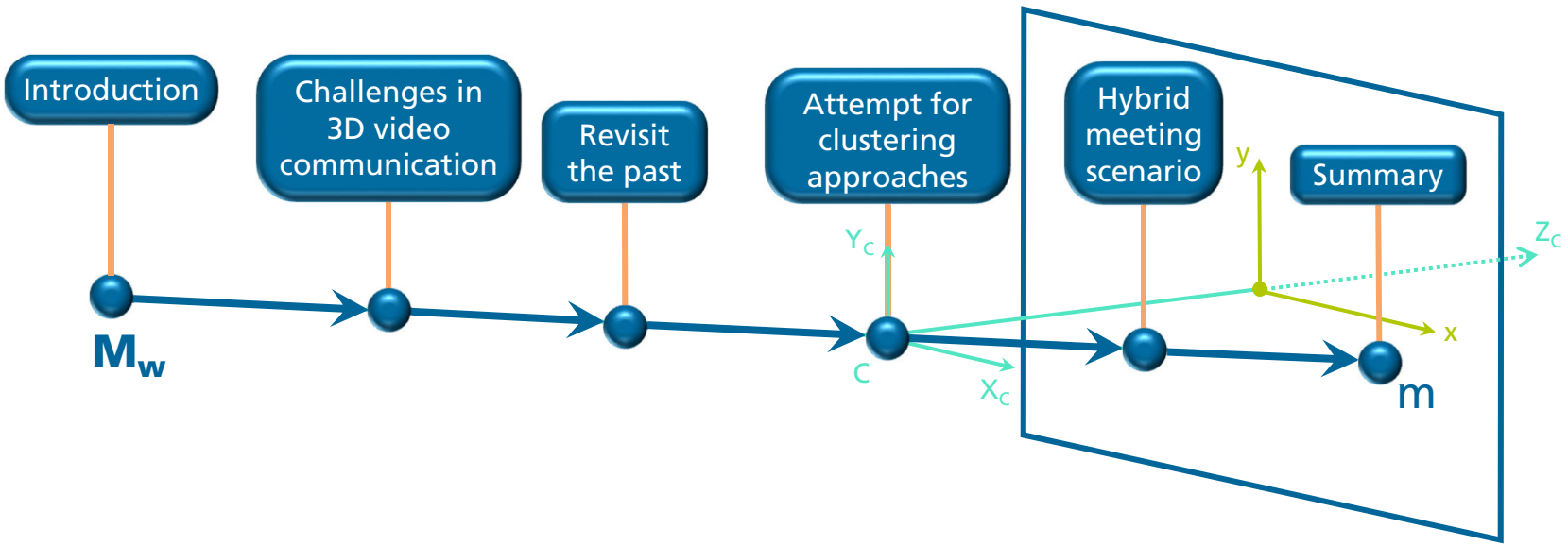
Recent advances in 3D Videocommunication

Dr. Oliver Schreer

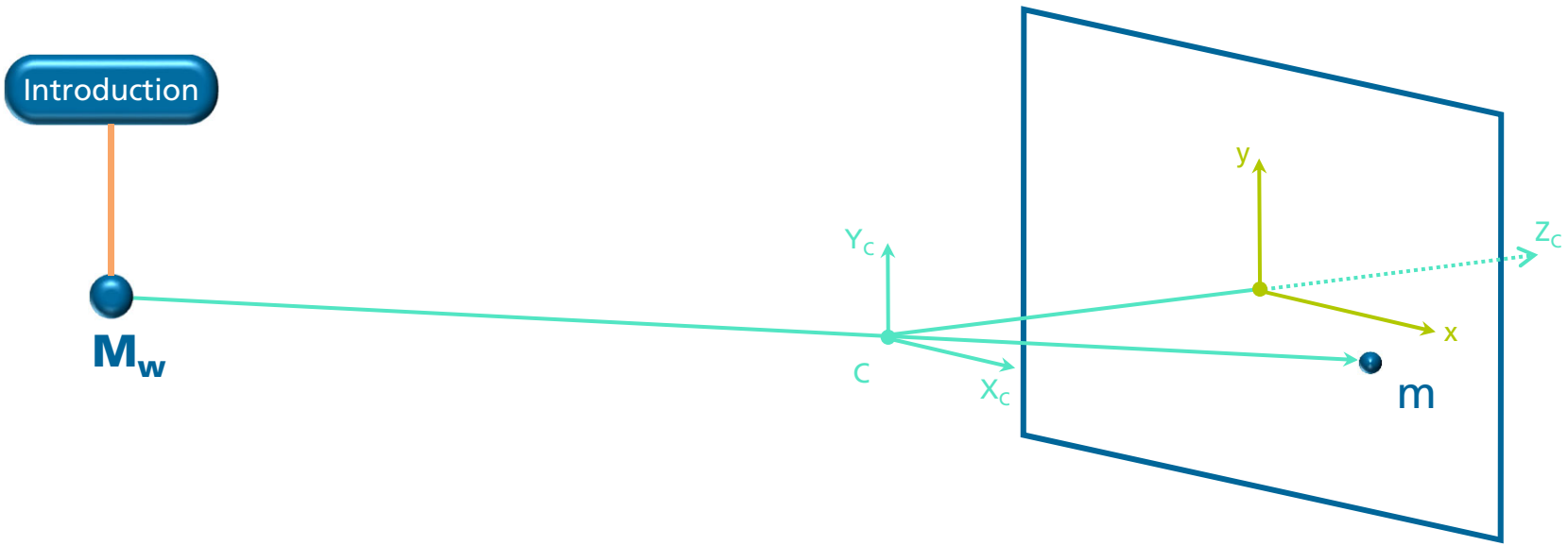
Spring School - Social XR @ CWI, Amsterdam, 7th March 2024

Fraunhofer HHI

A walk along the optical ray



A walk along the optical ray



Short introduction

- Since 1998 – research associate of Fraunhofer Heinrich-Hertz-Institut (HHI)
- 1999 – PhD on “Stereo Image Processing and Navigation in Mobile Robotics”
- Since 2002 – lecturer at TU Berlin on “Stereo Image Processing” and “View Synthesis”
- 2006 – Habilitation degree
- 2015 – Head of research group “Immersive Media & Communication”
- 2 European projects coordinated (FP6 Rushes, H2020 XR4All)
- Participation in 10+ EC projects
- Member of Board of Directors of XR4Europe
- 140+ publications

Vision & Imaging Technologies

Computer Vision & Graphics

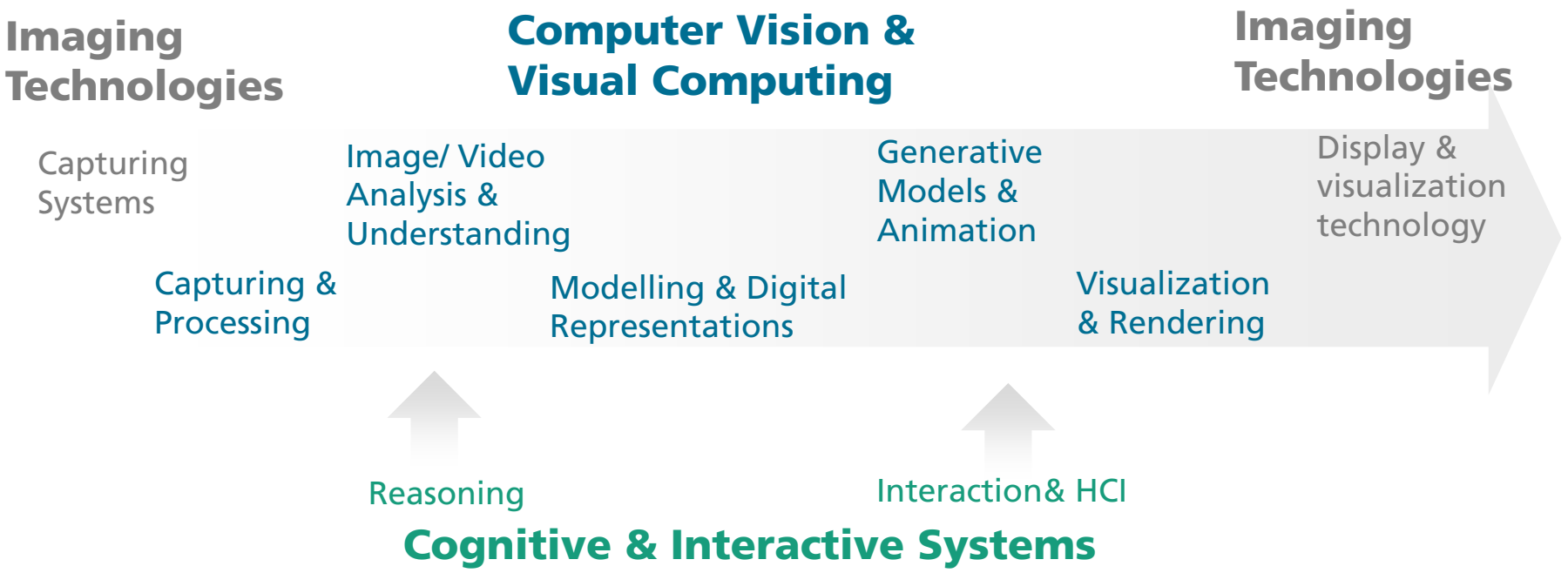
Immersive Media & Communication

Capture & Display Systems

Interactive & Cognitive Systems

Vision & Imaging Technologies

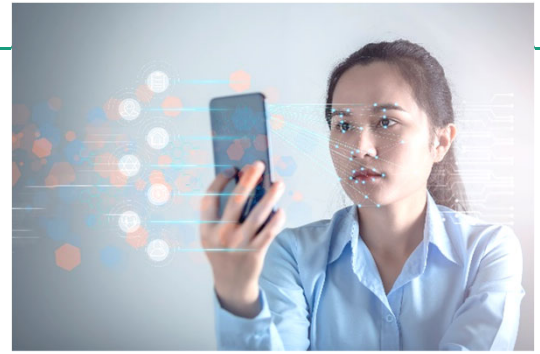
Innovative research along the entire video processing chain



Fields of Application

Industry & Construction

Medicine



Security

Multimedia

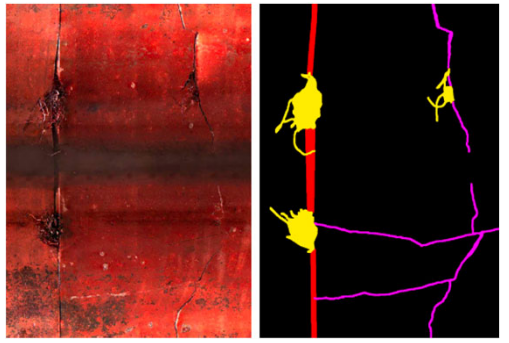
Agriculture & Environment

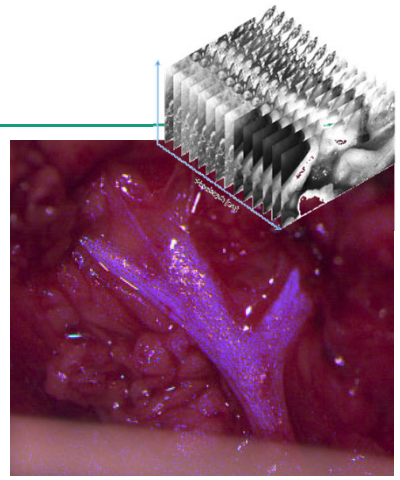
Mobility

Vision & Imaging Technologies

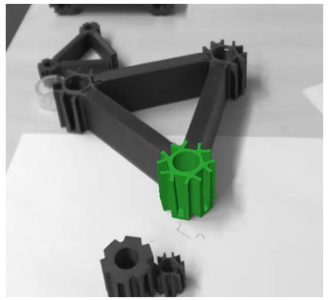
Core Competences

3D Reconstruction



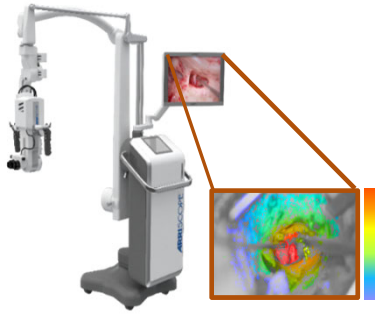


Scene Understanding / Multispectral Imaging



3D Tracking

Analysis and Synthesis of Humans

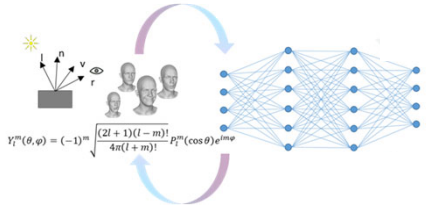


Augmented / Extended Reality

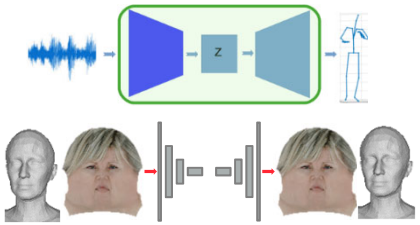
Research Activities

Methodologies

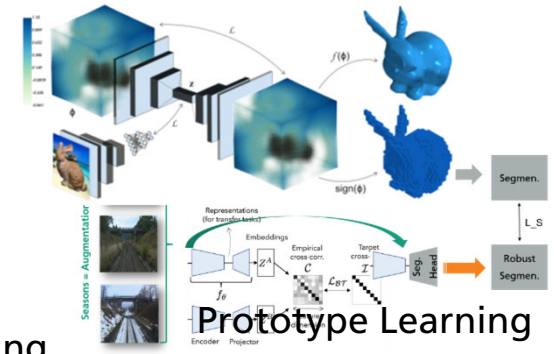
Model-based Deep Learning



Generative Models (GAN, VAE, Diffusion..)

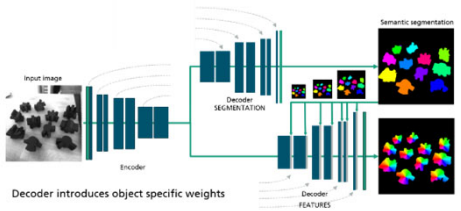


Neural Representations (NeRF, Gaussian Splats)

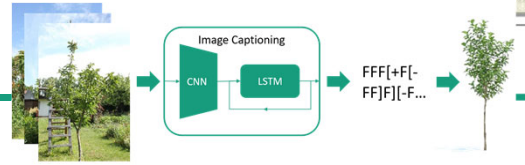


Robustness and Domain Adaptation

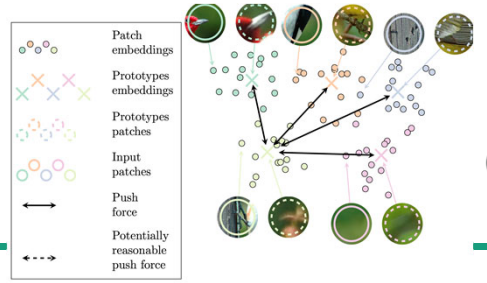
Training with synthetic Data



AI-based Procedural Modeling



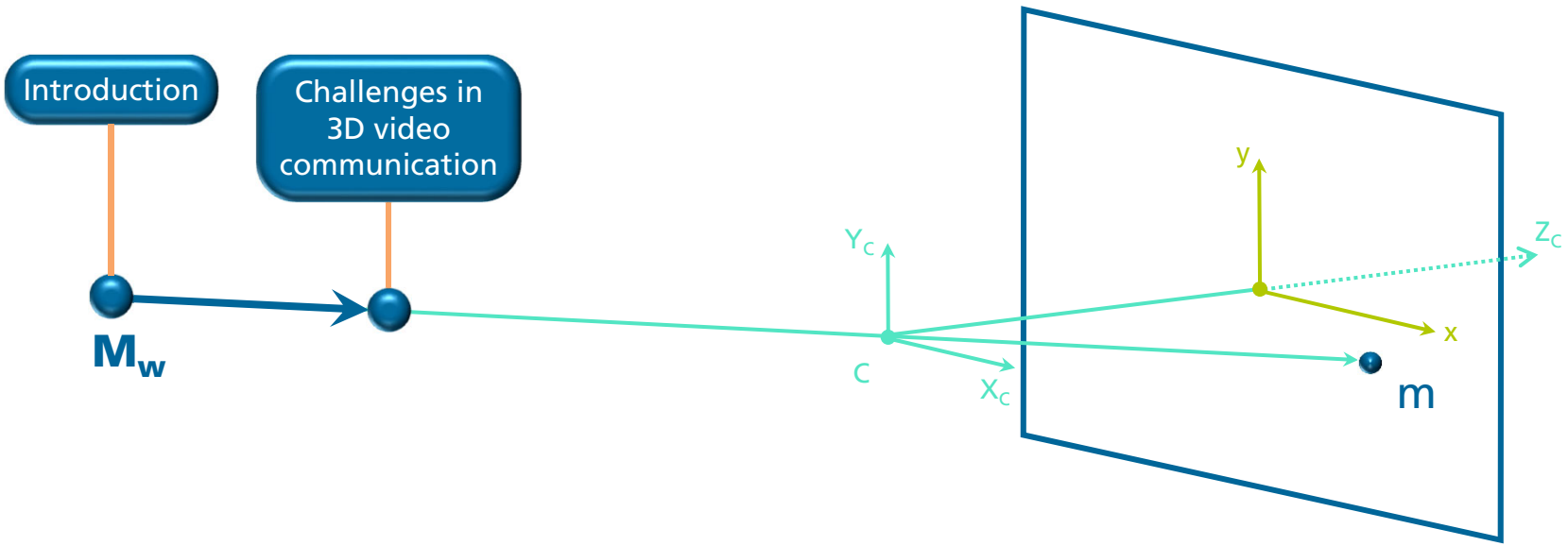
Prototype Learning



Human in the Loop

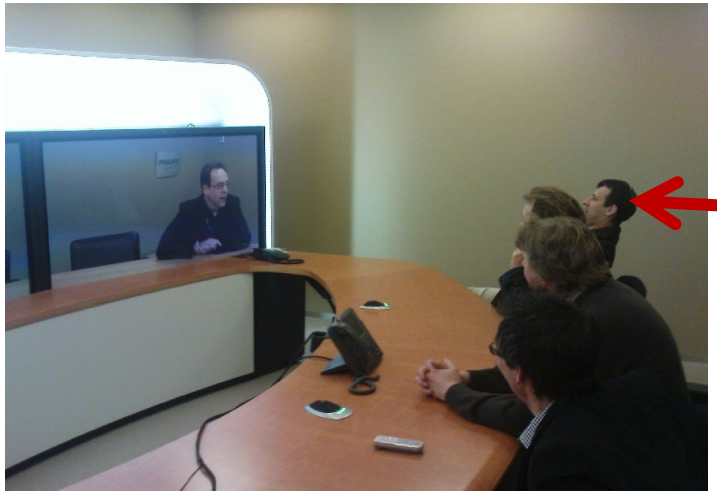
Fraunhofer HHI

A walk along the optical ray

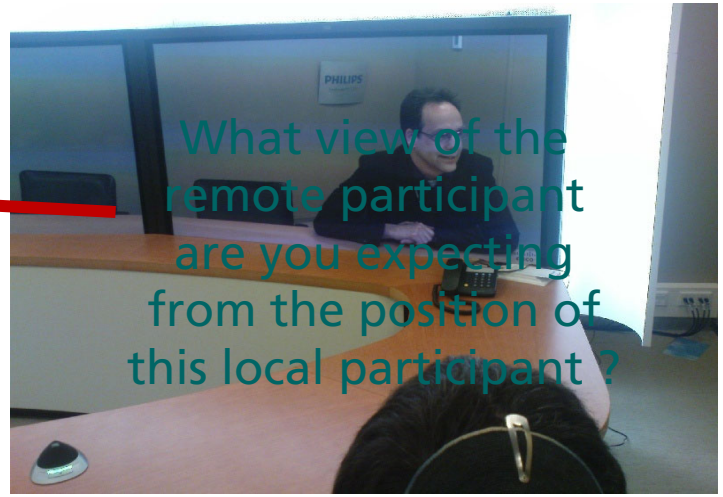


Challenges in 3D video communication

1. Misleading gesture representation in multi-point or Multi-party setups (who is pointing at whom?)

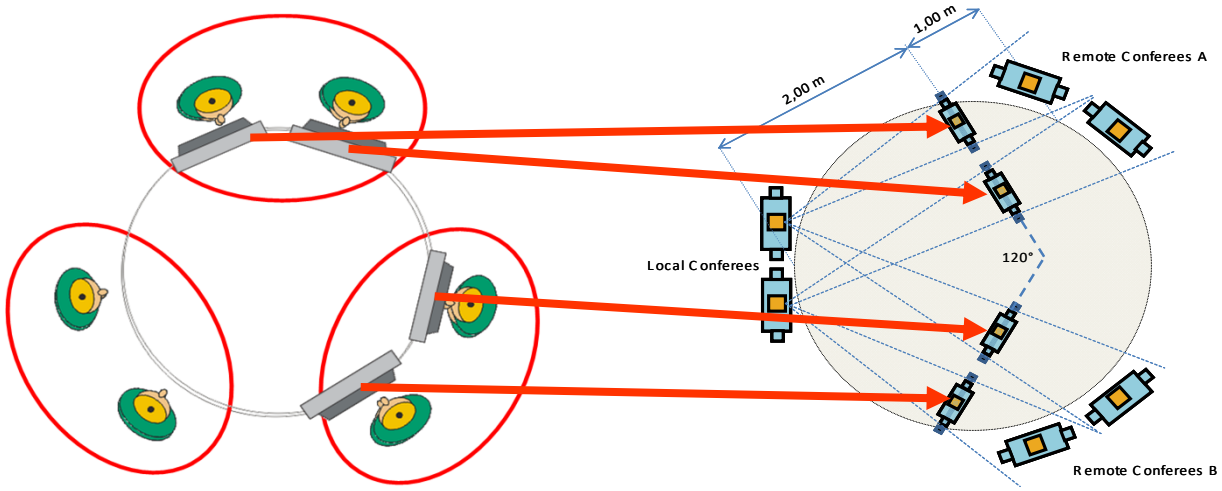


View from left local to right remote



View from right local to right remote

The Principle of the Shared Table

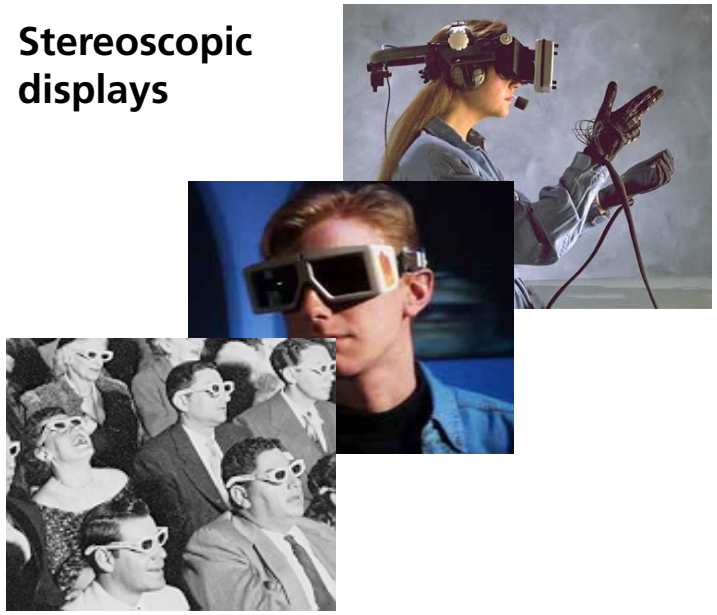


Challenges in 3D video communication

1. Misleading gesture representation in multi-point setups (who is pointing at whom?)
2. No stereoscopic viewing

Which Display Technologies for 3D?

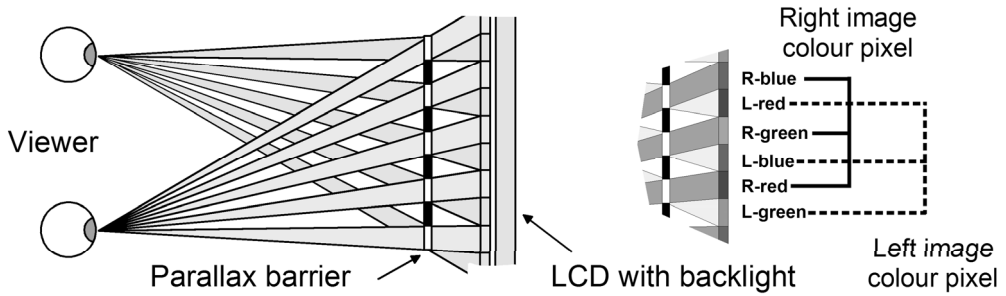
Stereoscopic displays



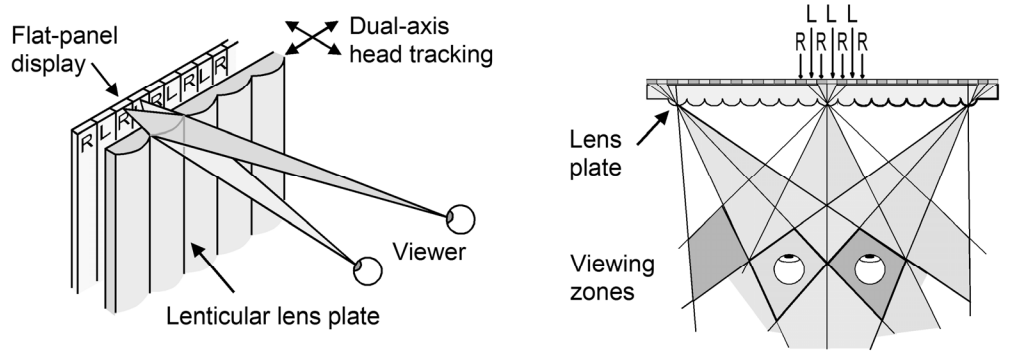
Autostereoscopic displays

The Answer is Given by Our Former Chancellor

Autostereoscopic Display Technology



Parallax barrier
(single user)



Lenticular lens plate
(multi user)

Fraunhofer HHI

Snapshot on Autostereoscopic Display Market

Leia Inc. (former Dimenco)

UHD 32’’

- 8k resolution
- lenticular

Alioscopy 3D

Full HD 21.5” - 55"

- 8 views
- lenticular

UHD 31.5” - 84"

- 16 views
- lenticular

Holografika, Light Field display

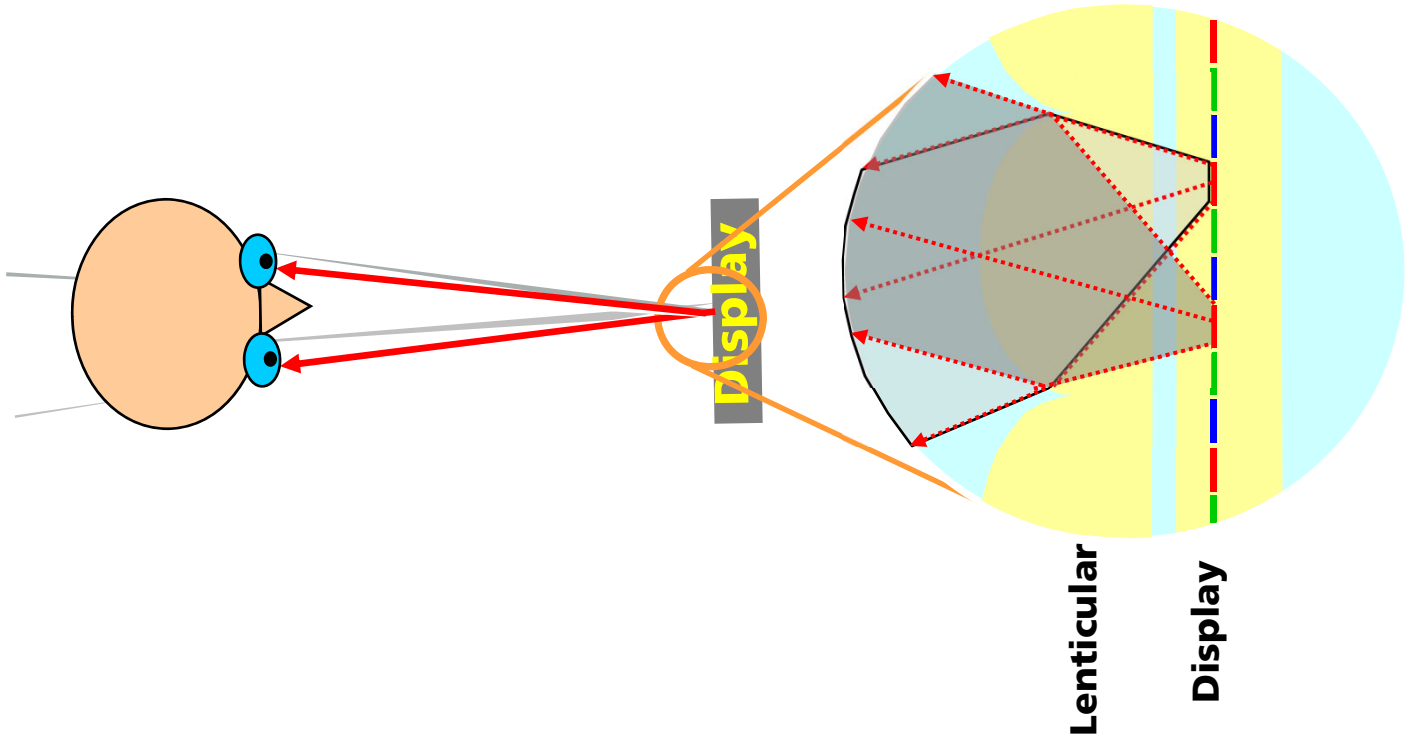
- 30’’ up to 140’’
- 1280 x 768 per view
- Continuous motion parallax
- Still images

Fraunhofer HHI

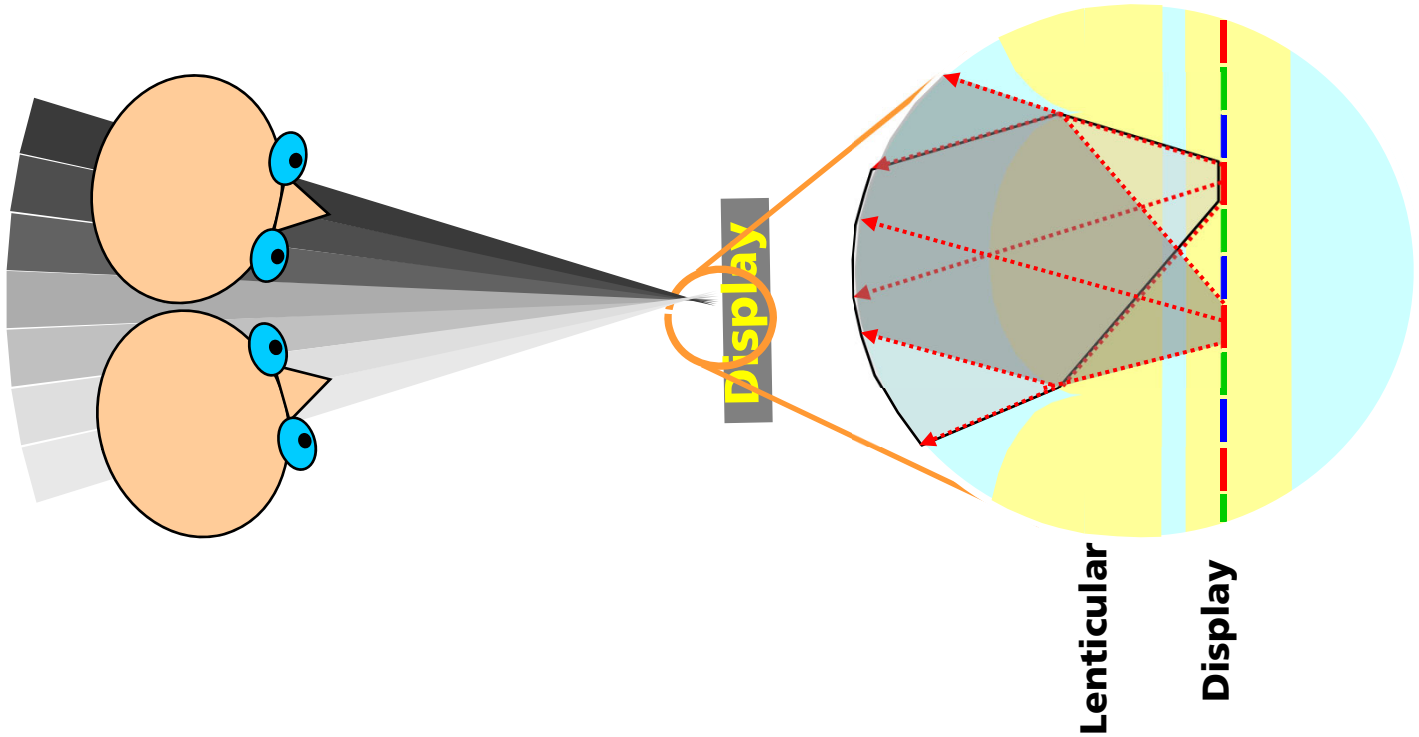
Challenges in 3D video communication

1. Misleading gesture representation in multi-point setups (who is pointing at whom?)
2. No stereoscopic viewing
3. No multi-perspective viewing for multiple users at one site

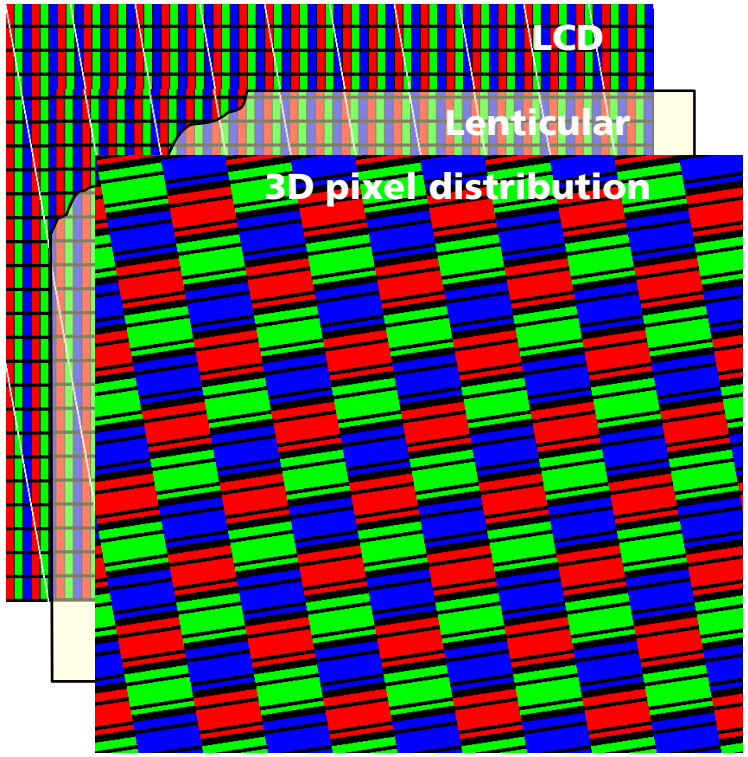
Lenticular Design



Lenticular Design – Multiple Views



Pixel Layout in a 3D View



15 views, slant 1/6

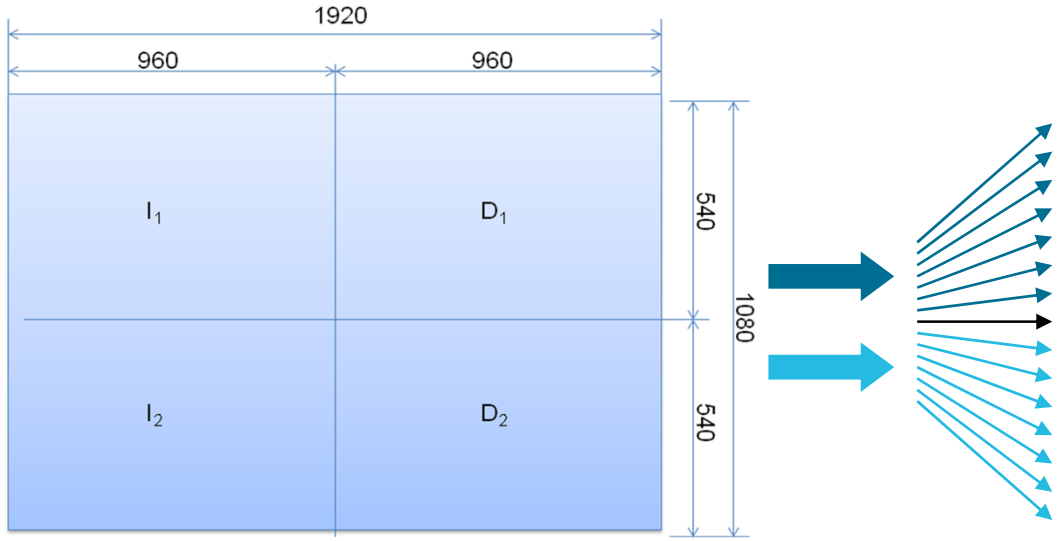
Slant 1/6:

+ Favourable spatial distribution of pixels

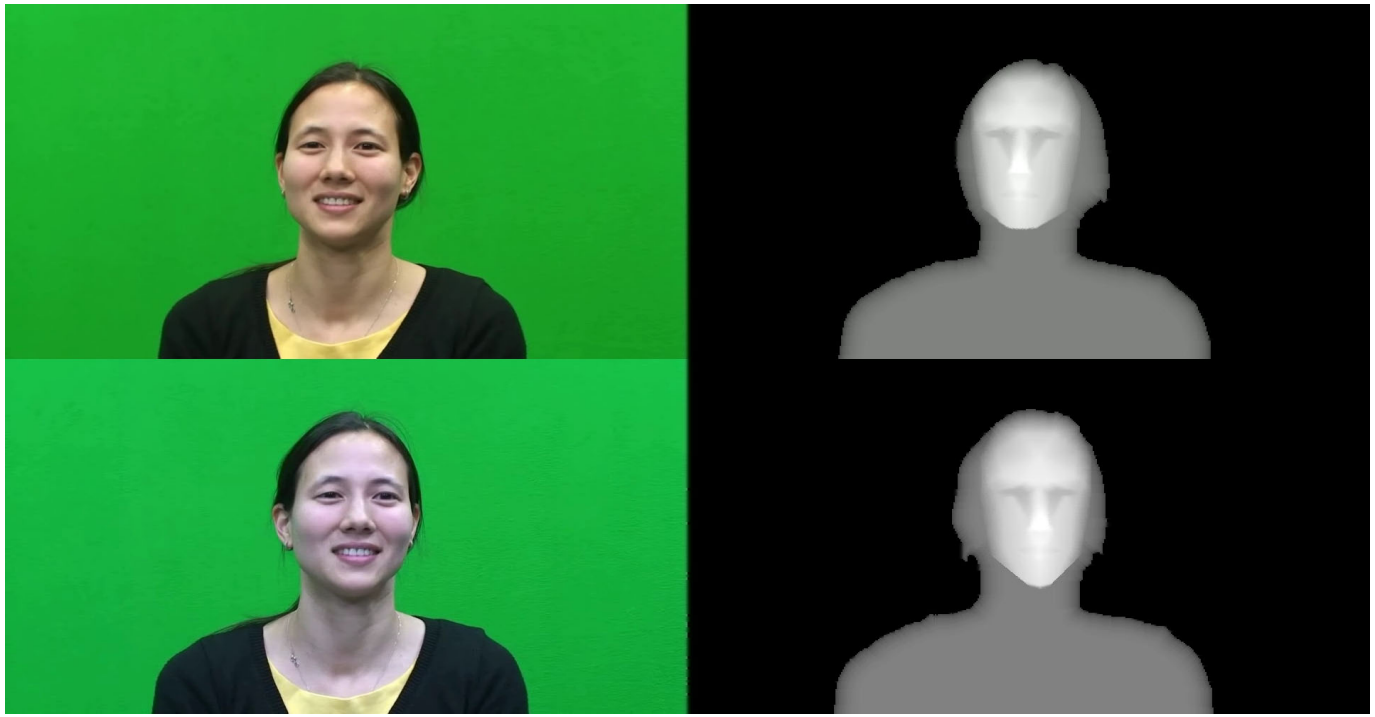
+ Less moiré

Video Format

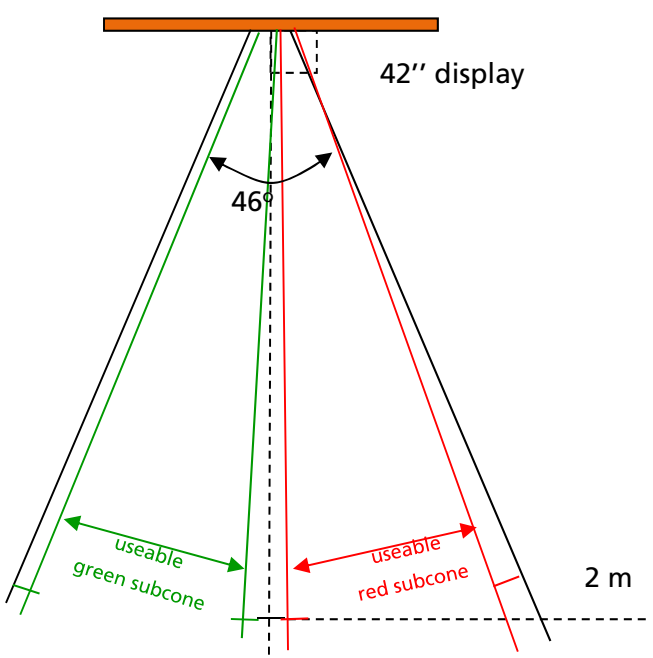
- Input: dual video + depth format → 15 views
- 7 views will be generated from one video + depth
- 2 different perspectives, 7 views each + 1 blank view = 15

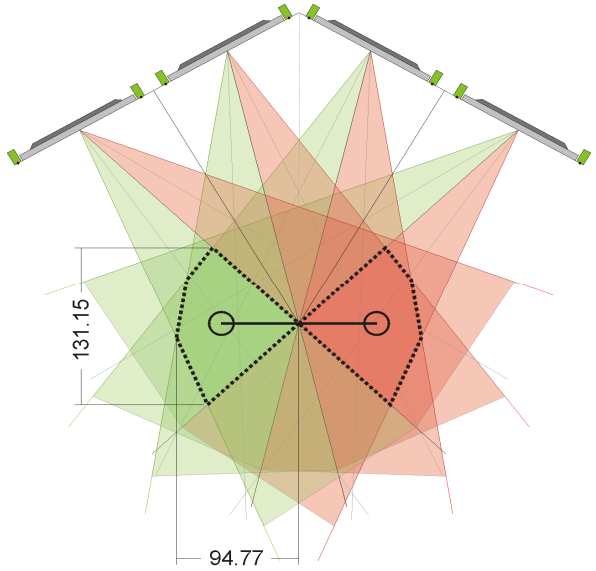


Example of Dual Video + Depth Format



Arrangement of 4 Multi-Perspective Displays





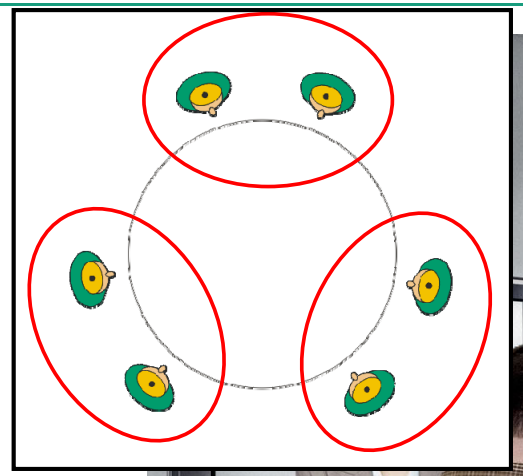
Arrangement with 4 multi-perspective displays

Example

Challenges in 3D video communication

1. Misleading gesture representation in multi-point setups (who is pointing at whom?)
2. No stereoscopic viewing
3. No multi-perspective viewing for multiple users at one site
4. Lack of eye contact

Problem of Eye Contact



Problem

Displacement angle

Solution

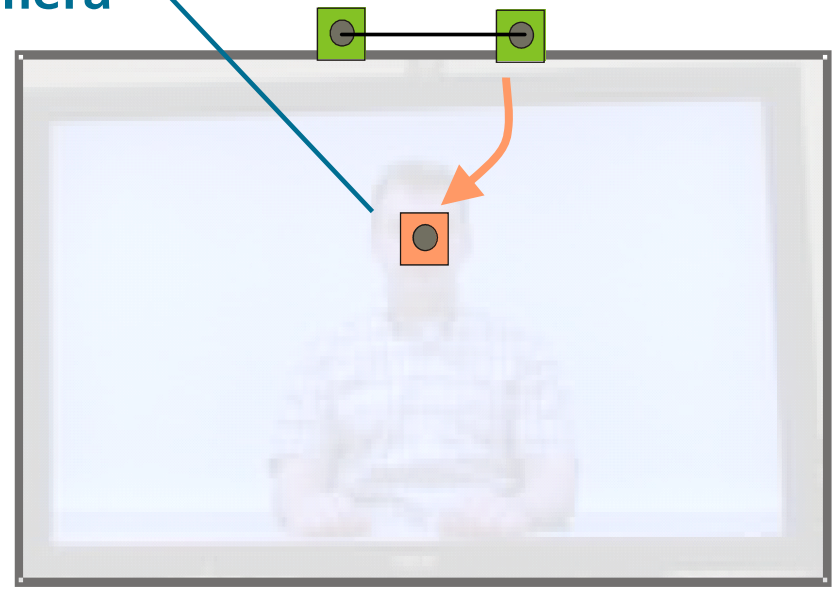
Virtual camera

Realization

- Disparity/depth estimation
- Virtual rotation of conferee

Novel view rendering

Video + depth for virtual camera



Summary

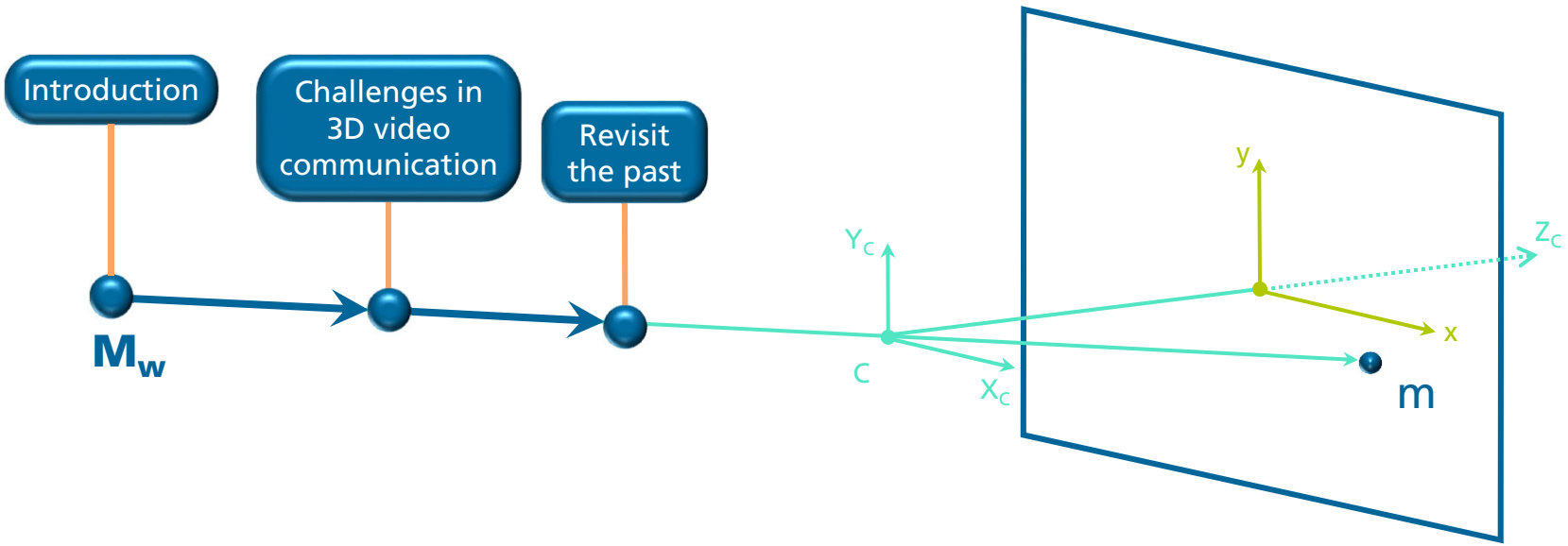
- Gesture representation in multi-point setups (who points at whom?)

  → shared table principle
- Stereoscopic viewing → autostereoscopic displays
- Multi-perspective viewing for multiple users → novel multi-perspective 3D display
- Lack of eye contact → virtual view rendering

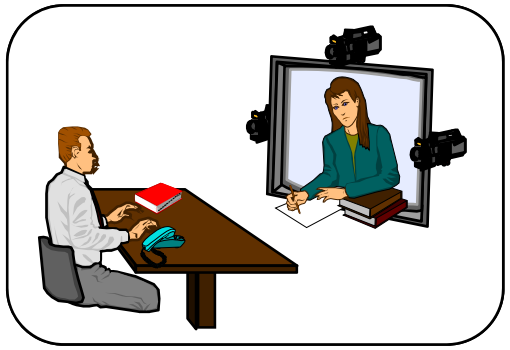
Real-time 3D video processing is required for

- ➢ **support of multi-view multi-perspective 3D displays**
- ➢ **provision of eye contact by virtual view rendering**

A walk along the optical ray



Revisit the past – virtual eye contact



ACTS Panorama
(1995-1998)

global disparity estimator

local disparity estimator

original left

interpolated

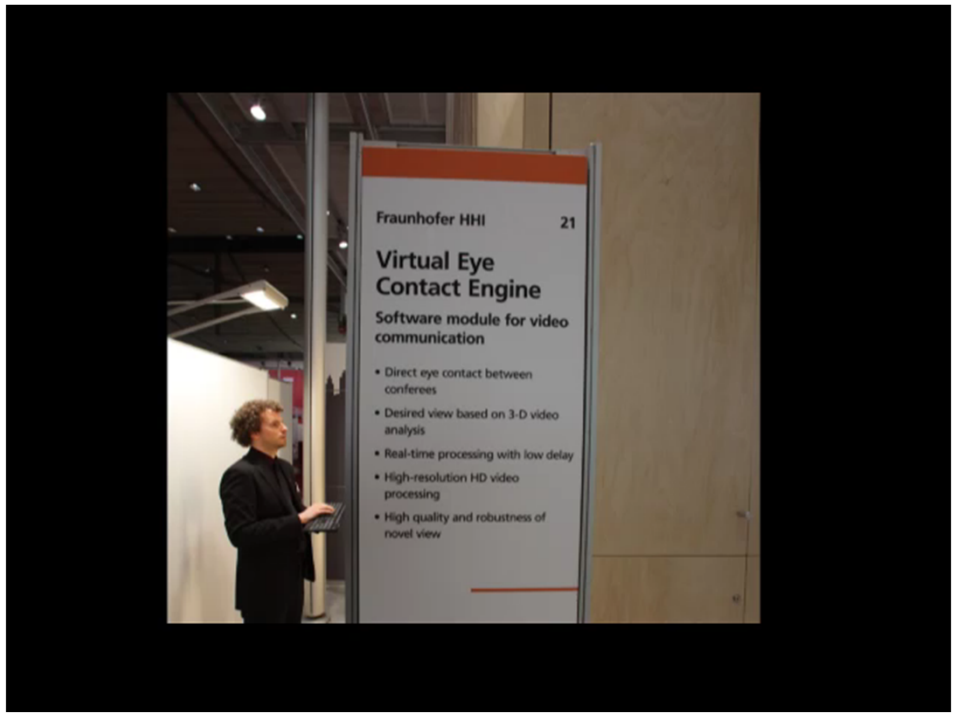
original right

Revisit the past – virtual eye contact

IST Project Virtue (´00-´03)

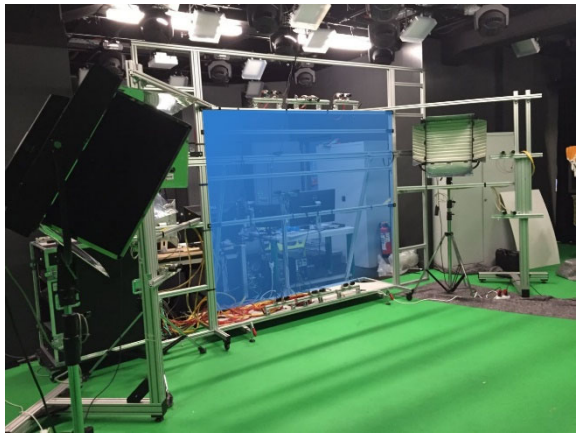
FP7 3DPresence (´08-´11)

Virtual Eye Contact Engine (´10-´11)

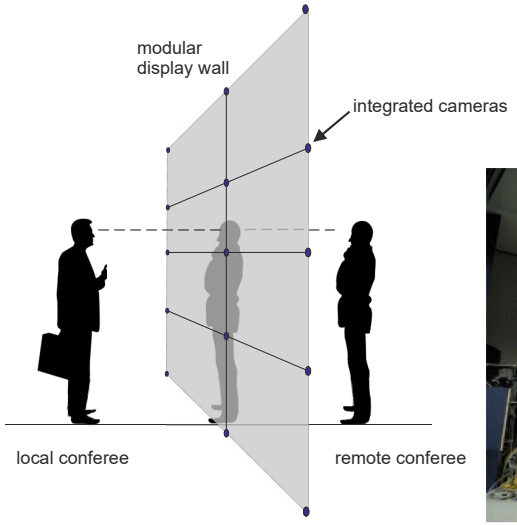


Revisit the past – virtual eye contact

Revisit the past – FVV (free viewpoint video) and FTV (free viewpont television)

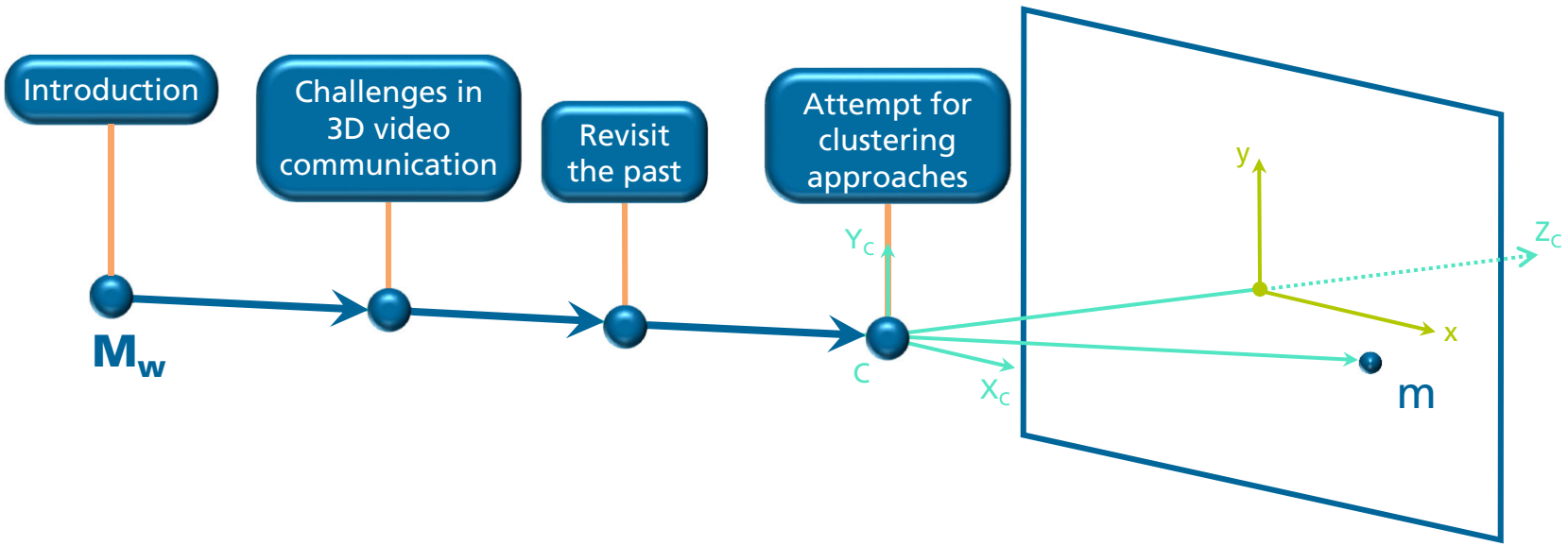


Telepresence wall, 2013,
Fraunhofer HHI

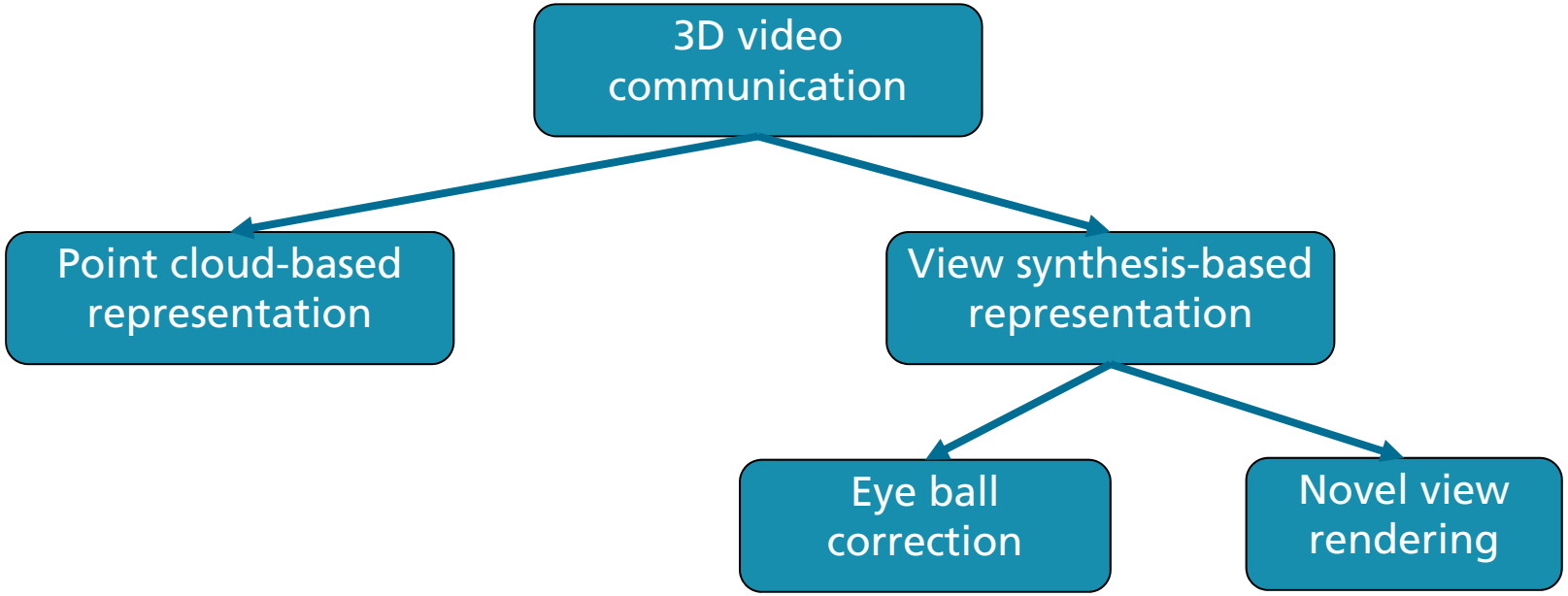


2006, 100 camera system,
Univ. Nagoya, Japan

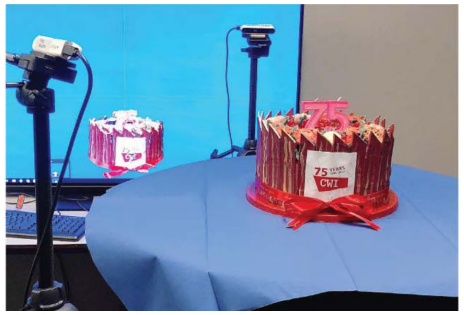
A walk along the optical ray



Attempt for classification

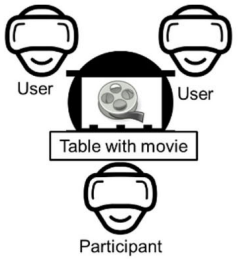


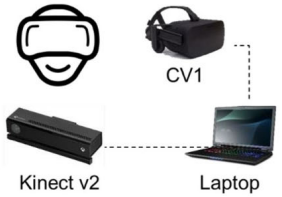
Point-cloud based representations



- Real-time Social XR system by CWI[1]

- TogetherVR by TNO[2]

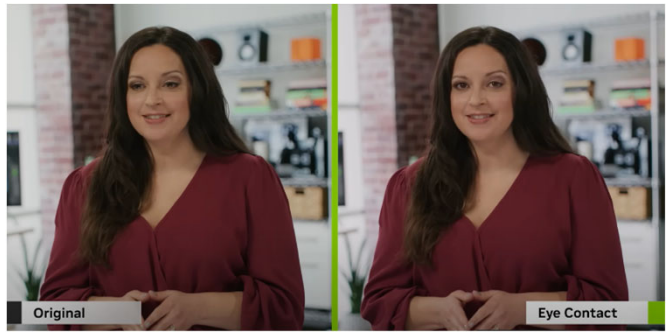




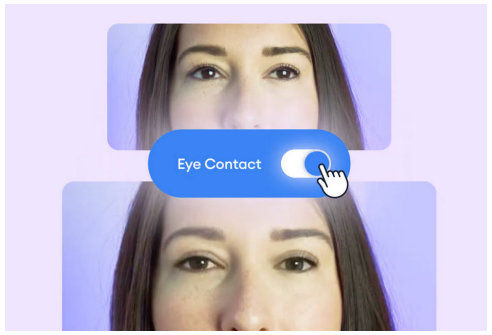
[1] Viola, J. Jansen, S. Subramanyam, I. Reimat and P. Cesar, "VR2Gather: A Collaborative, Social Virtual Reality System for Adaptive, Multiparty Real-Time Communication," in *IEEE MultiMedia*, vol. 30, no. 2, pp. 48-59, April-June 2023, doi: 10.1109/MMUL.2023.3263943.

[2] Simon N. B. Gunkel, Hans M. Stokking, Martin J. Prins, Nanda van der Stap, Frank B. ter Haar, and Omar A. Niamut. 2018. Virtual reality conferencing: multi-user immersive VR experiences on the web. Proc. of the 9th ACM Multimedia Systems Conference (MMSys '18)

Creating eye contact through manipulating eyes



NVIDIA Maxine Eye Contact



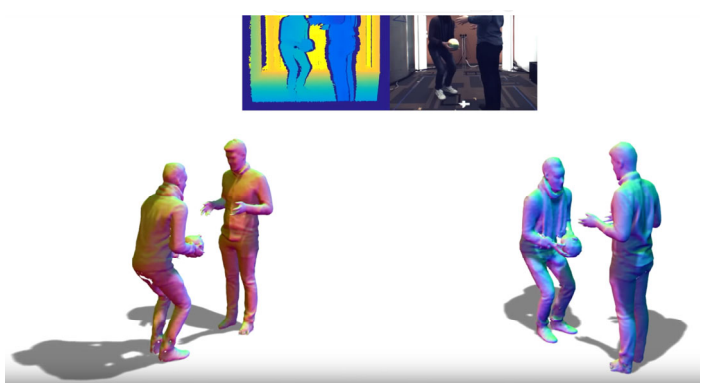
VEED.IO

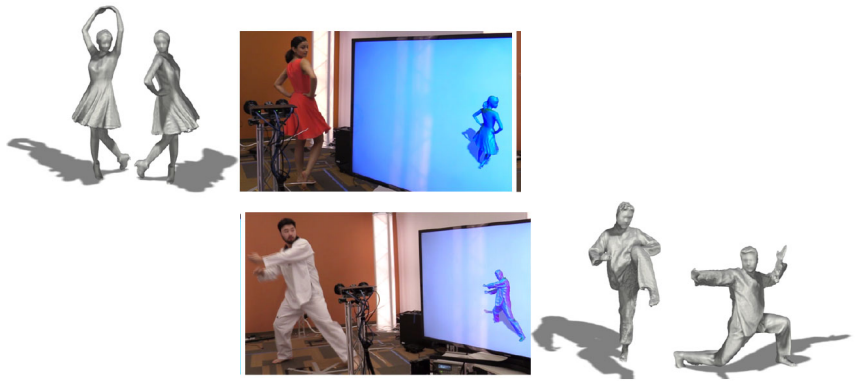
Isikdogan et. al. 2020, WACV

… and many more such as Microsoft eye contact correction for Surface Pro X and Windows 11

- AI-based algorithm for gaze estimation and redirection
- Provides natural eye contact
- Solution for one-to-one communication (Zoom, …)
- … but NO solution for one-to-many (who is looking at whom)

Real-time performance capture – Fusion4D



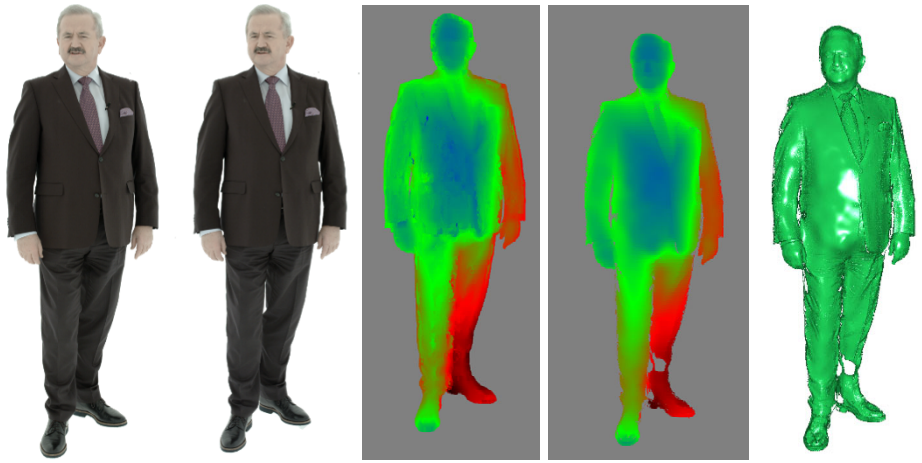


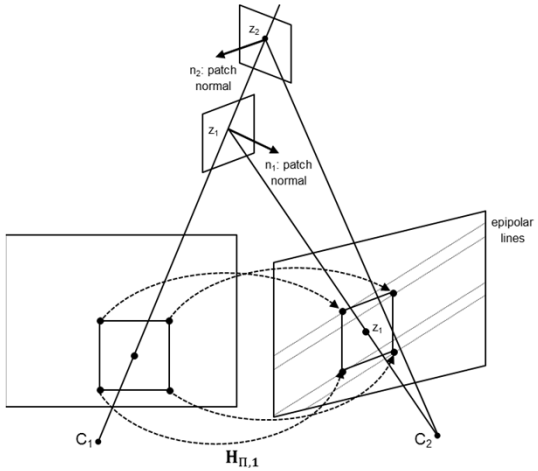
- 24 cameras (1 Mpixel res) in trinoccular arrangement
- Depth estimation using the PatchMatch Stereo algorithm [Bleyer et al. 2011]
- 8 depth maps fused to mesh

M. Dou, S. Khamis, Y. Degtyarev, Ph. Davidson, S.R. Fanello, A. Kowdle, S. Orts-Escolano, Ch. Rhemann, D. Kim, J. Taylor, P. Kohli, V. Tankovich, S. Izadi. 2016. Fusion4D: real-time performance capture of challenging scenes. ACM Trans. Graph. 35, 4, Article 114 (July 2016)

Stereo-based Depth Estimation (Surface-Region Matching)

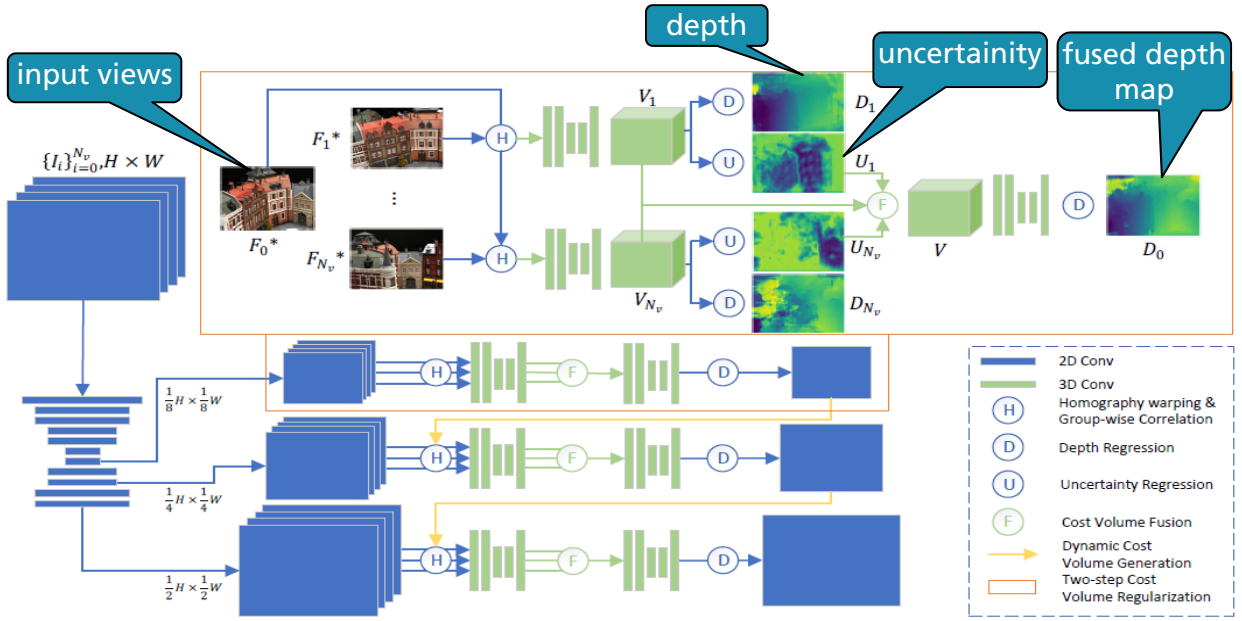
- Spatial candidates and a statistically guided update for comparison (no search range)
- Spatial 3D patches are projected from L→R and R→L followed by consistency check
- Iterative structure serves parallel processing on GPU

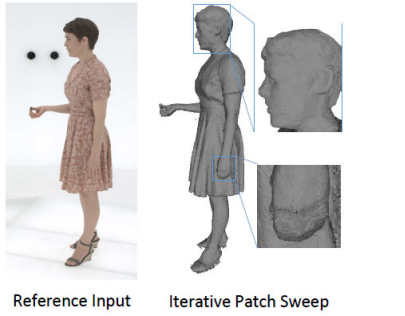


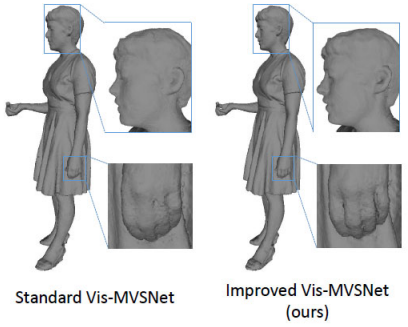


- O. Schreer, I. Feldmann, S. Renault, M. Zepp, P. Eisert, P. Kauff, Capture and 3D Video Processing of Volumetric Video, IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, September 2019.

Learning-based Depth Estimation (Vis-MVS)

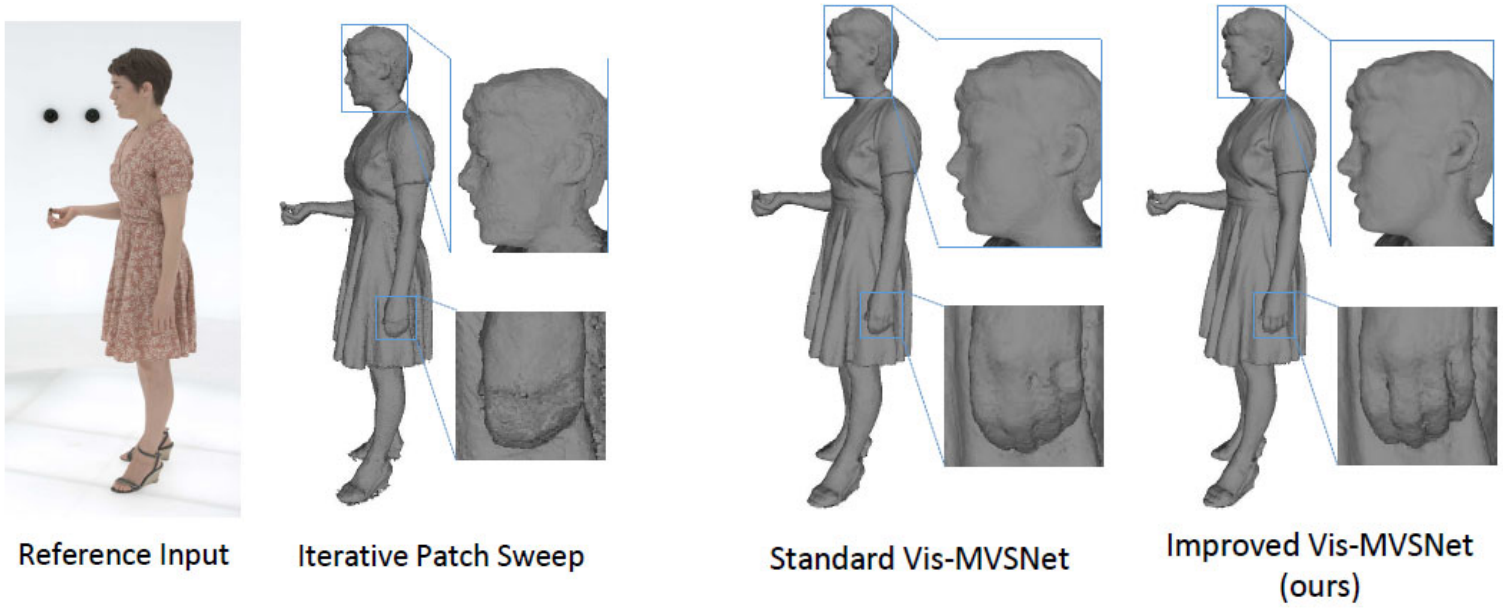




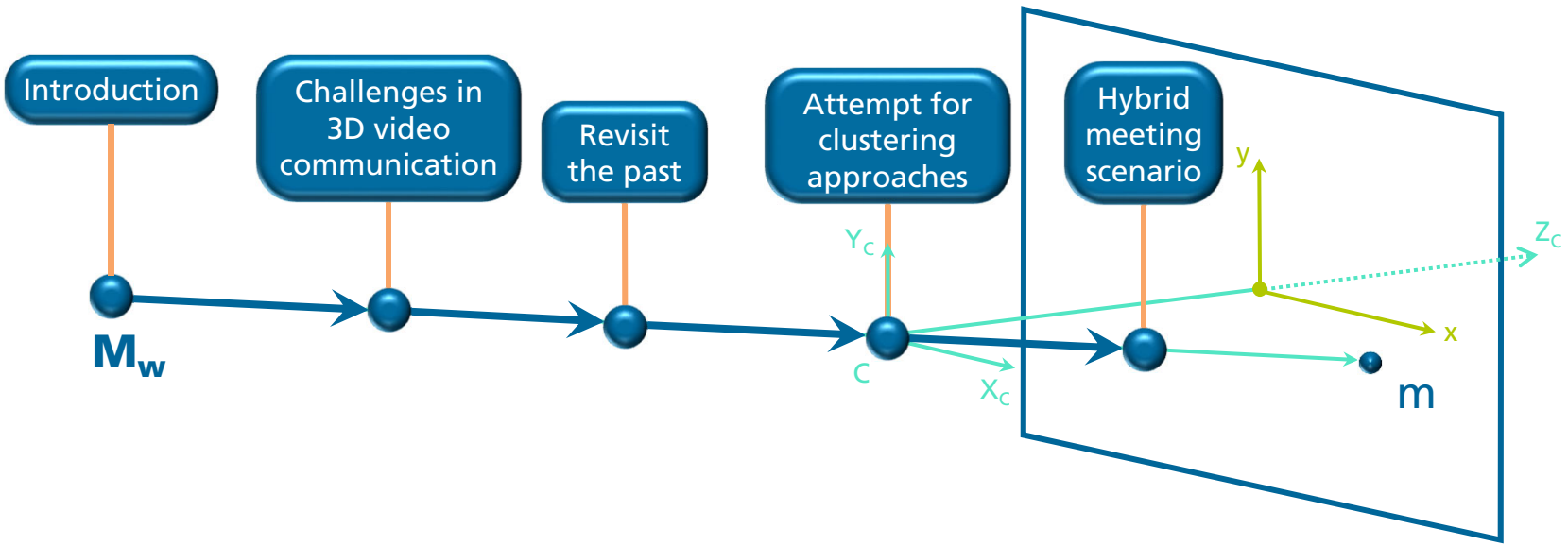


- D. Chen, M. Worchel, I. Feldmann, O. Schreer, P. Eisert, Accurate human body reconstruction for volumetric video, 2021 Int. Conf. on 3D Immersion, Brussels, Belgium, December 2021, **Best paper award**

Learning-based Depth Estimation (Vis-MVS)



A walk along the optical ray



Hybrid meeting scenario

- Two real persons meet a remote person

Challenges

- Eye contact
- Different perspectives for both real person
- High-quality rendering

Hybrid meeting scenario

The worlds first two view display

Features

- View separation at two distinct regions in space
- Moving area in a 50cm circle
- 4k panel allows rendering of 2 views in HD resolution

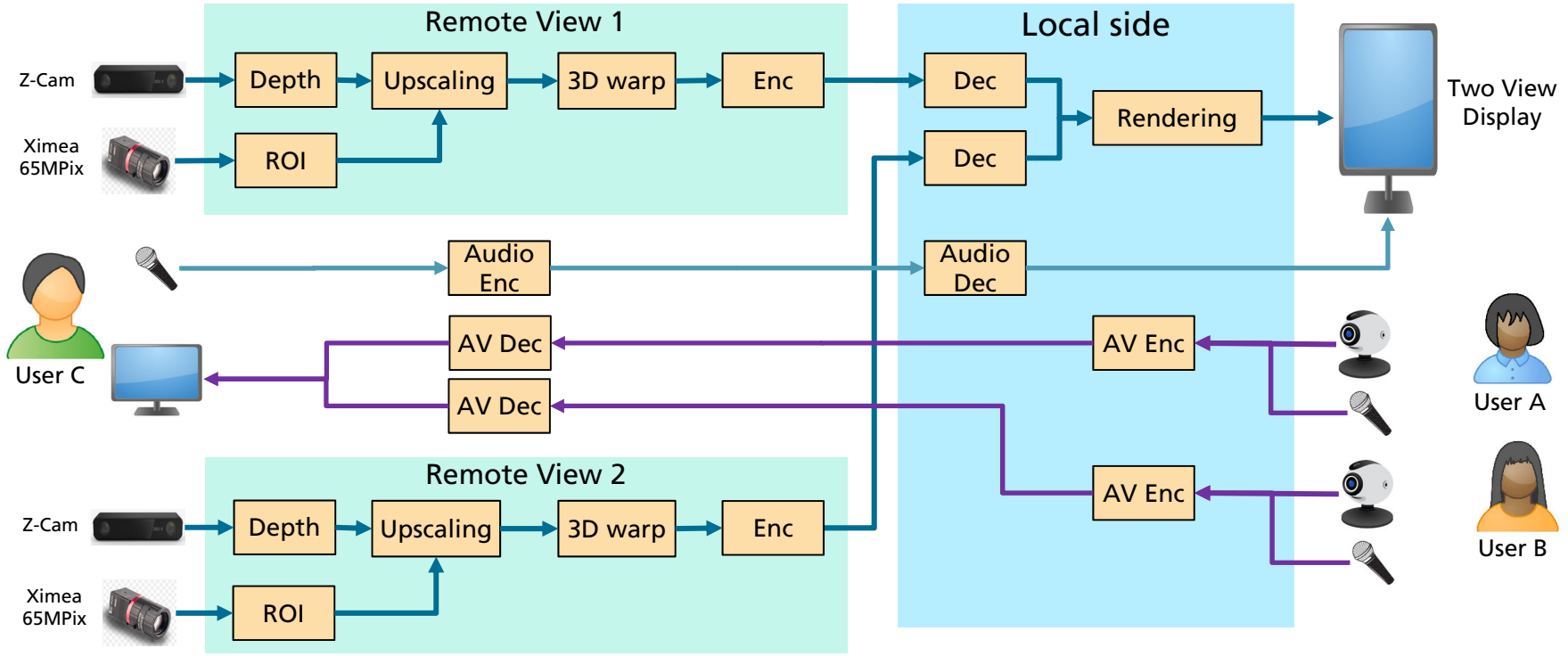
Hybrid meeting scenario

Remote 3D Setup

Features

- Low-res 3D camera
- High-res 65 MPix camera
- Fusion of low-res depth and high-res texture
- Real-time @12 FPS
- Delay below 50 msec.

3D Multi-View Workflow



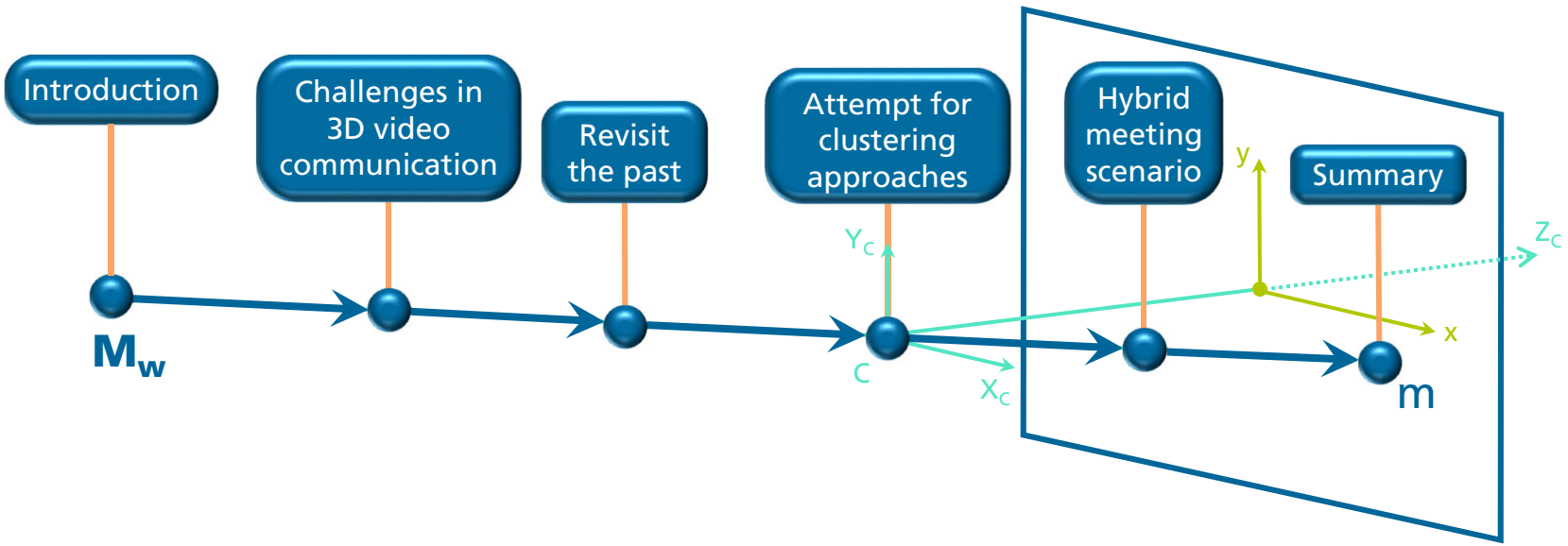
Hybrid meeting scenario

Results

Real-time low delay 3D processing

Boundary processing using Codeformer

A walk along the optical ray



Fraunhofer HHI

Summary

- True immersive video communication is a relevant topic since two decades
- Pandemic has led to a new rise of the topic
- Various challenges for immersive video communication require different solutions
- Eye contact for point-to-point communication is solved
- Real-time multi-party immersive 3D video communication still a challenge
- Deep Learning supports high-quality novel view rendering

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut, HHI

WE PUT SCIENCE INTO ACTION.

Contact:

Dr. Oliver Schreer
oliver.schreer@hhi.fraunhofer.de

Einsteinufer 37
10587 Berlin

Many thanks to my colleagues, contributing to this talk:
Decai Chen, Ingo Feldmann, Marcus Zepp