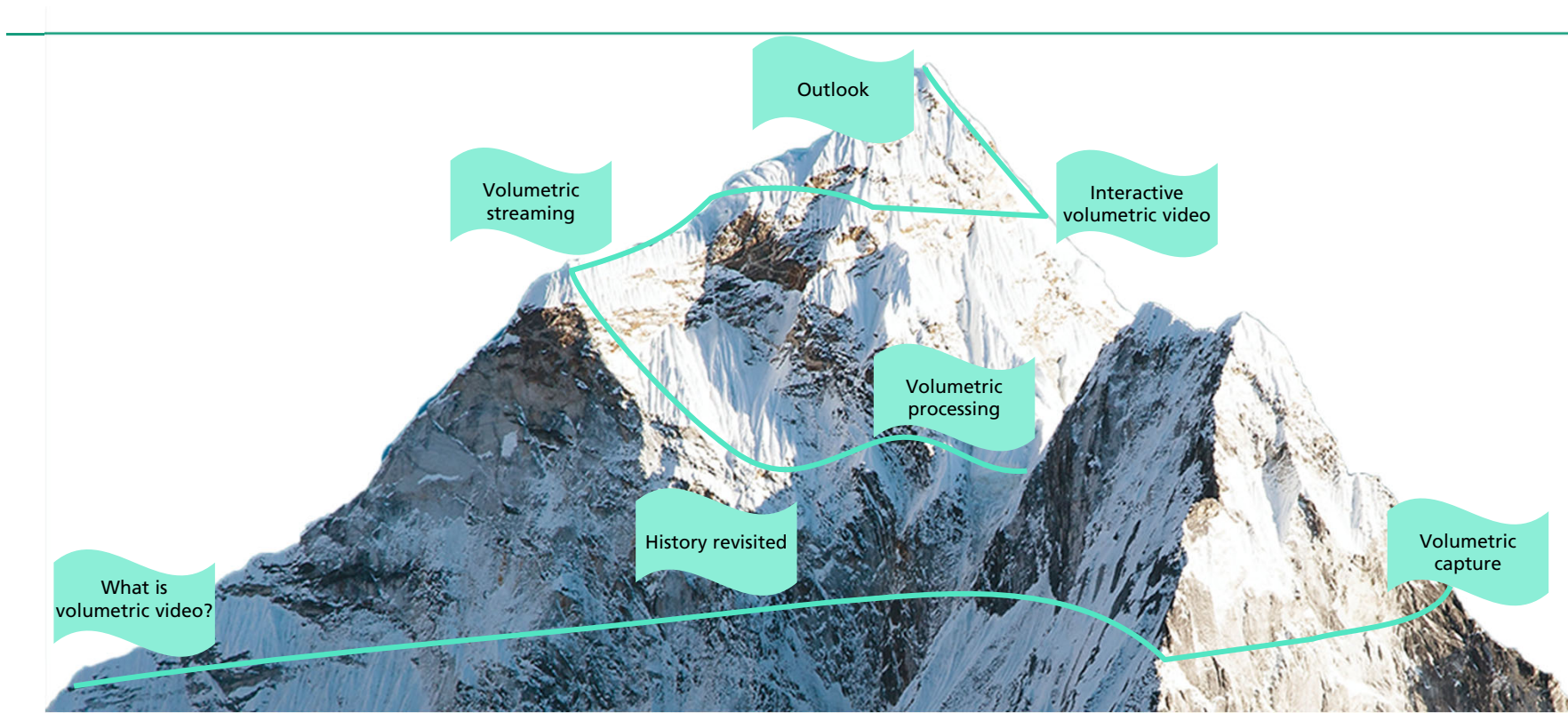

A Journey to Volumetric Video – the Past, the Present and the Future

Dr. Oliver Schreer

Spring School - Social XR @ CWI, Amsterdam, 16th March 2023

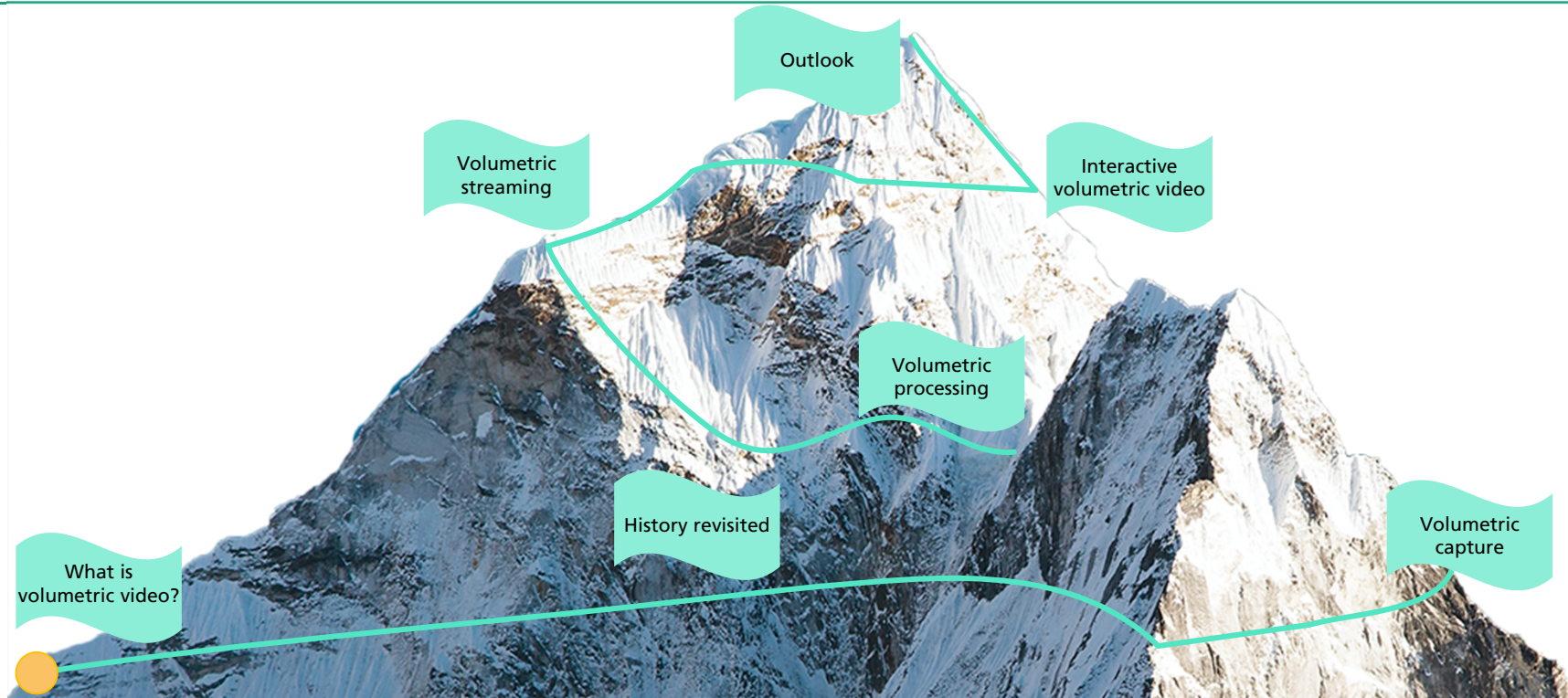


Let's climb the volumetric mountain



Let's climb the volumetric mountain

What is volumetric video?

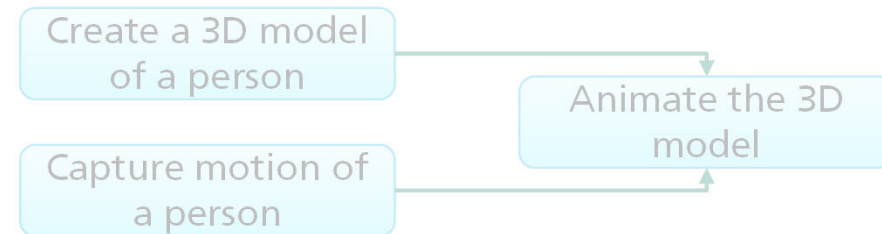


What is volumetric video?

- Persons are currently represented as animated characters in VR
- Limited complexity of motion can be tracked and used for animation
- Facial expressions cannot be captured to full extent
- Reconstruction of dynamic behaviour of clothes is challenging

- Direct capture of any body movement, facial expression and motion of clothes
- Frame-by frame modelling leads to dynamic 3D objects → a sequence of meshes
- 3D models can be viewed from any direction

State of the Art

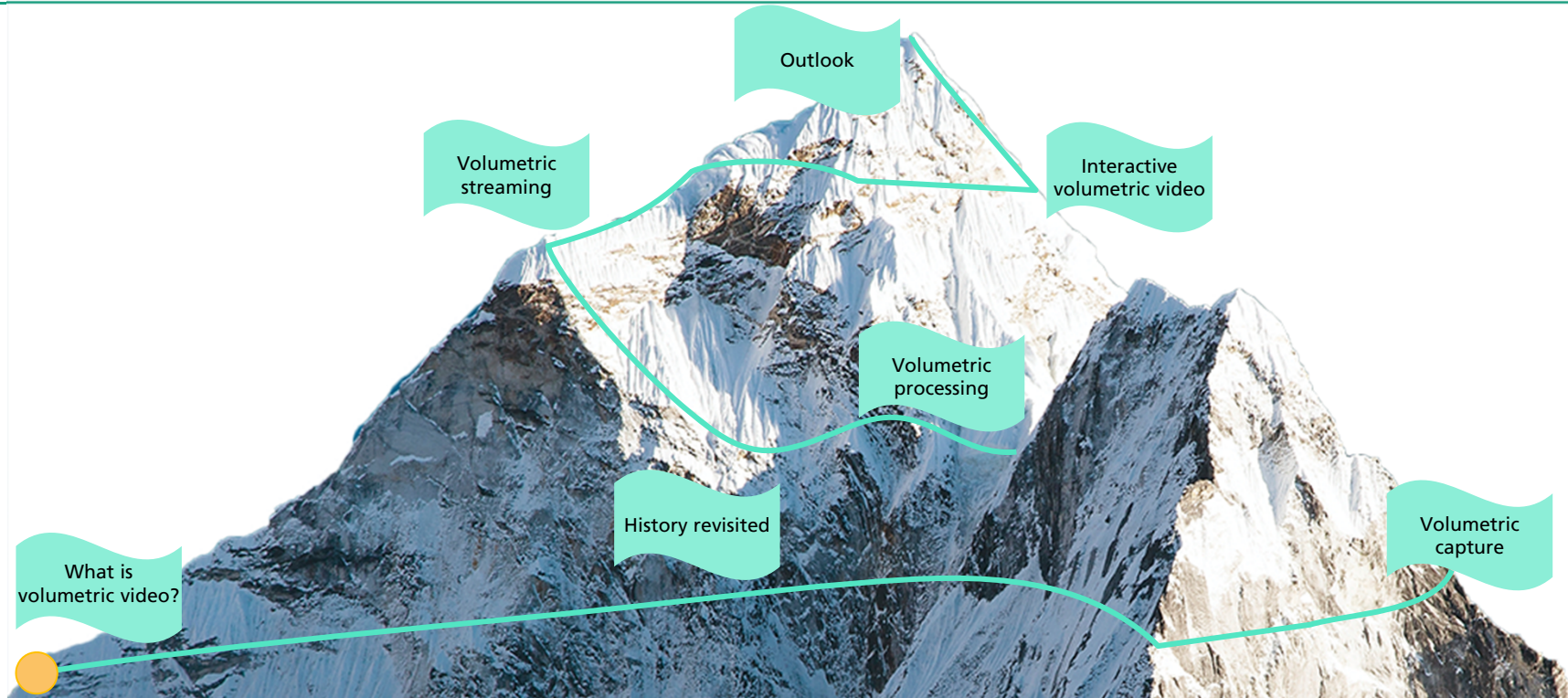


Novel approach

Capture the person and create a 3D model

Let's climb the volumetric mountain

History revisited

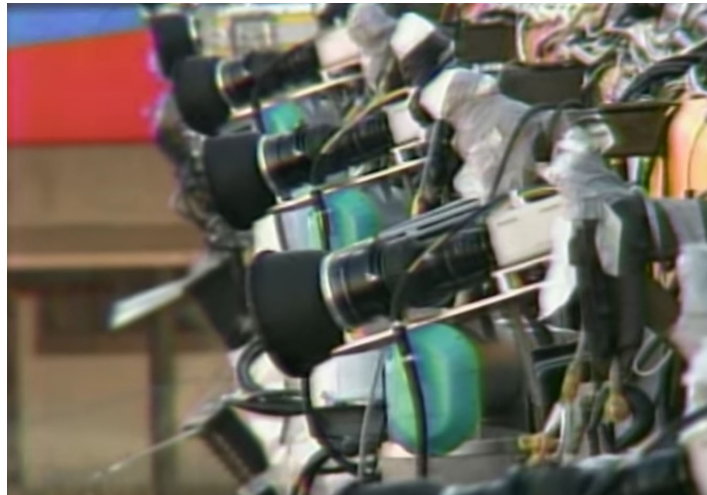


History of Volumetric Video

Term	Virtual View Synthesis	Free View Point TV (FTV)/ Free View Point Video (FVV) / 3D Video			Performance Capture / Volumetric Video / Video grammetry	
	Year	1999	2001	2003	2004	2016
Application	3D Video Communication/ Virtual Eye Contact	Eye Vision, Super Bowl by Takeo Kanade	3DTV, Ray-space capture by Masayuki Tanimoto	3DTV	Training, cultural heritage, sports, music	3D Video Communication, Telepresence
Technology	Intermediate View Interpolation	Hardware-based	Ray-space rendering	Image-based Rendering, Depth Image-based Rendering (DIBR), Multi- View Video Coding (MVC)	Dynamic Mesh Reconstruction	Point Cloud Fusion
Off-/Online	Online	Online	Online	Offline/Online	Offline	Online
# of Cameras	2-3 per participant	30 robotic cameras		up to 100	up to 100	<12 RGBD

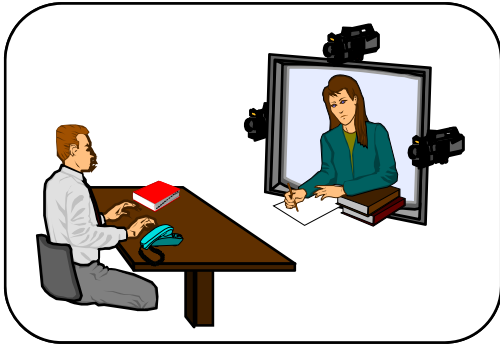
History of Volumetric Video

- 2001 – Hardware based Free Viewpoint Television (FTV) solution
- 30 robotic cameras with 220 degree coverage



Super Bowl 2001, Eye Vision, Takeo Kanade

Revisit the past – virtual eye contact



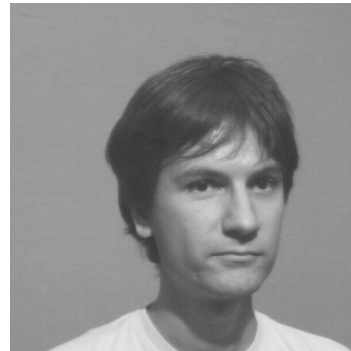
ACTS Panorama
(1995-1998)



global disparity estimator



local disparity estimator



original left



interpolated



original right

Revisit the past – virtual eye contact



IST Project Virtue (2000-2003)



FP7 3DPresence (2008-2011)

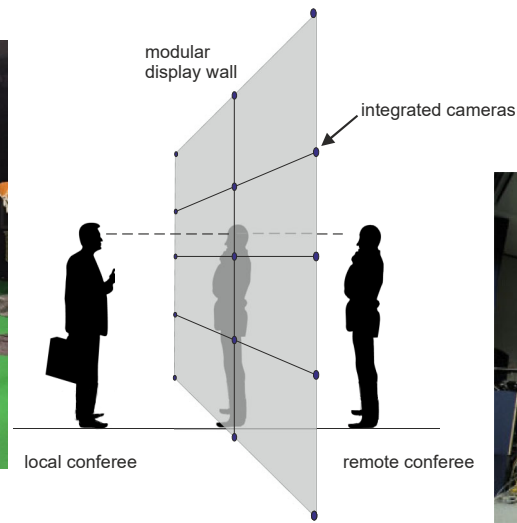
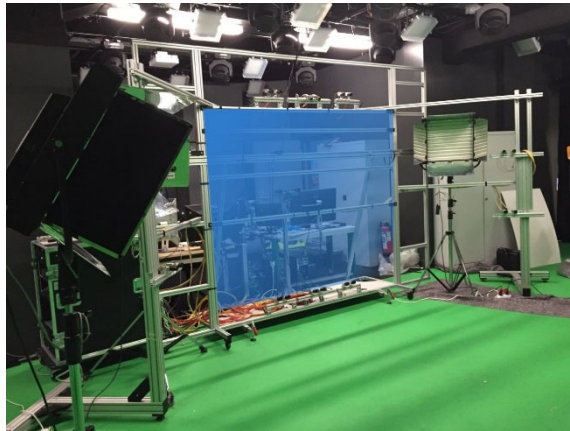


Virtual Eye Contact engine (2010-2011)



TRIFOCAL BASELINE SYSTEM

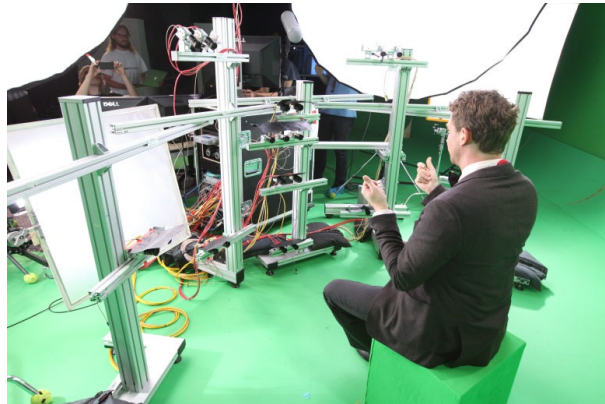
Revisit the past – FVV (free viewpoint video) and FTV (free viewpoint television)



Telepresence wall, 2013,
Fraunhofer HHI

2006, 100 camera system,
Univ. Nagoya, Japan

Revisit the past – Live Performance Capture



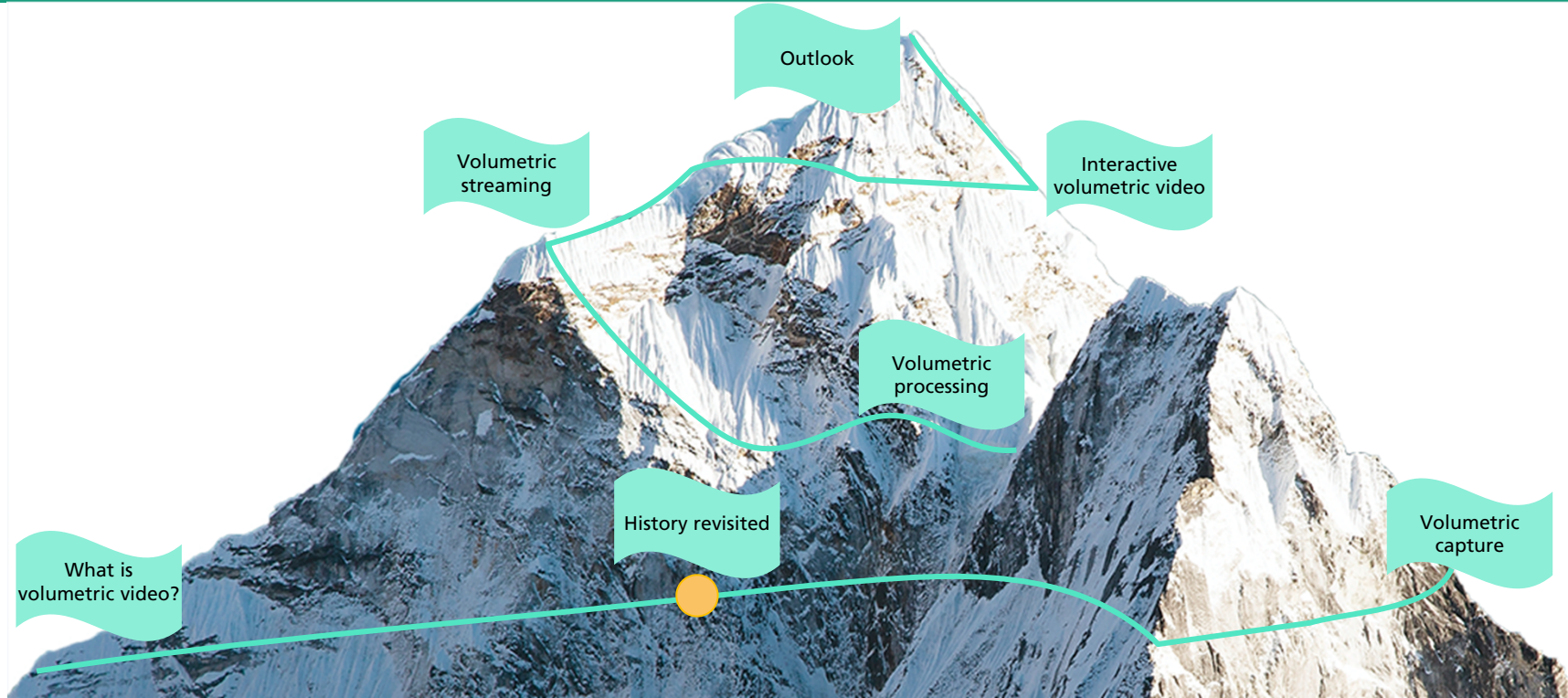
2016, 23 camera system,
Fraunhofer HHI



- Collet A., Chuang M., Sweeney P., Gillett D., Evseev D., Calabrese D., Hoppe H., Kirk A., Sullivan S.. 2015. High-quality streamable free-viewpoint video. ACM Trans. Graph. 34, 4, August 2015.

Let's climb the volumetric mountain

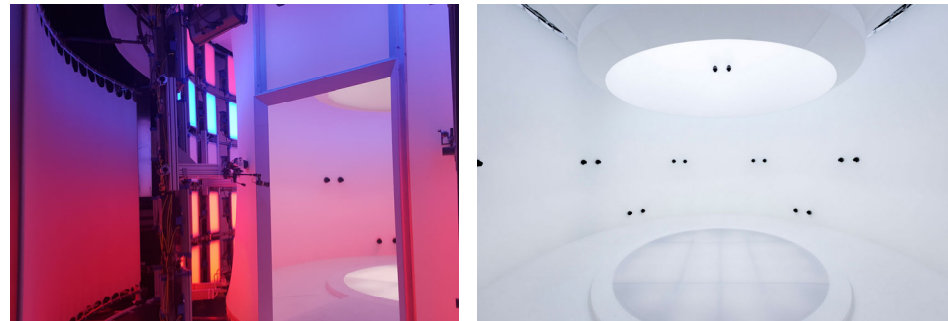
Volumetric capture



Volumetric Video Capture Today



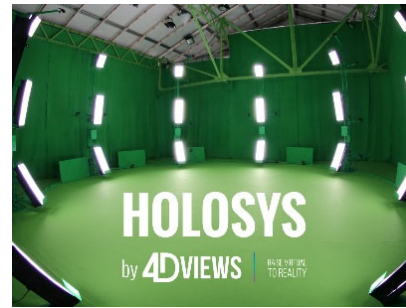
Fraunhofer HHI, Germany



Volucap, Germany



Volograms, Ireland



4DViews – Holosys, France

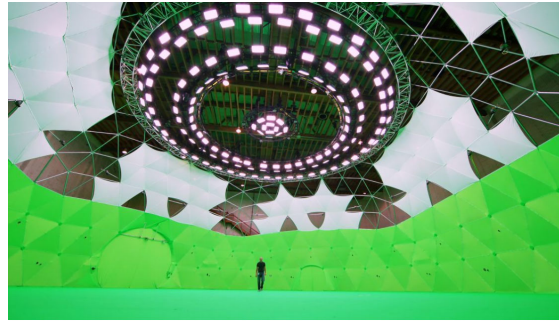


Microsoft Mixed Reality Capture Studio (MS MRC)

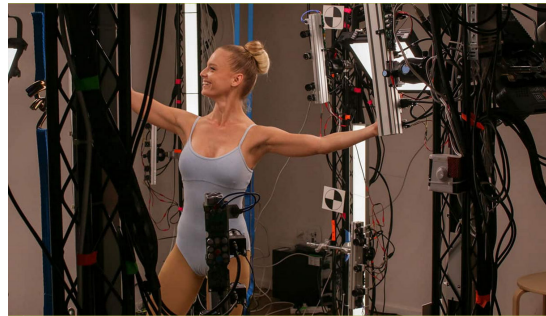
Volumetric Video Capture Today



Polymotion Stage, US



Intel studios, LA, US

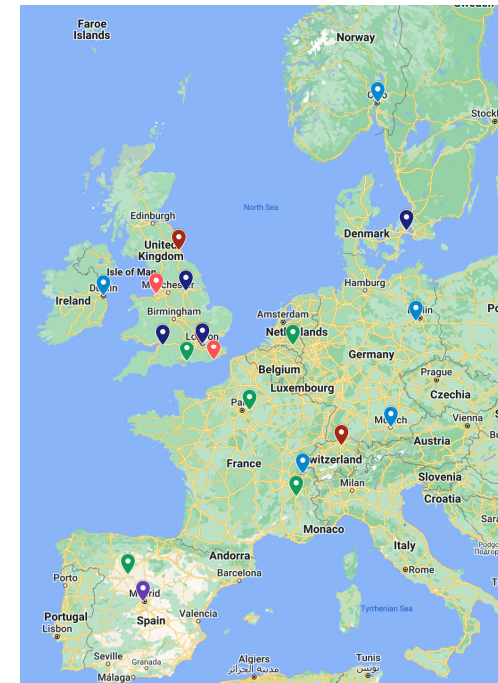


Mantis Vision, Israel

Other studios

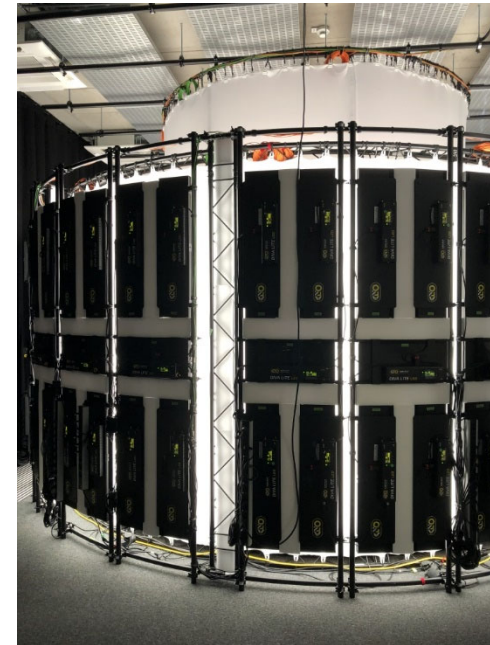
- Dimension Studios (UK)
based on MS MRC
- Metastage (USA)
based on MS MRC
- Jump Studio (Korea)
based on MS MRC
- Korea Immersive Studio (Korea)
- Crescent (Japan)
based on HoloSys
- 4DR Studios (The Netherlands)
based on HoloSys
- CIVIT at Tampere Univ. (FI)

Volumetric Video Capture Today

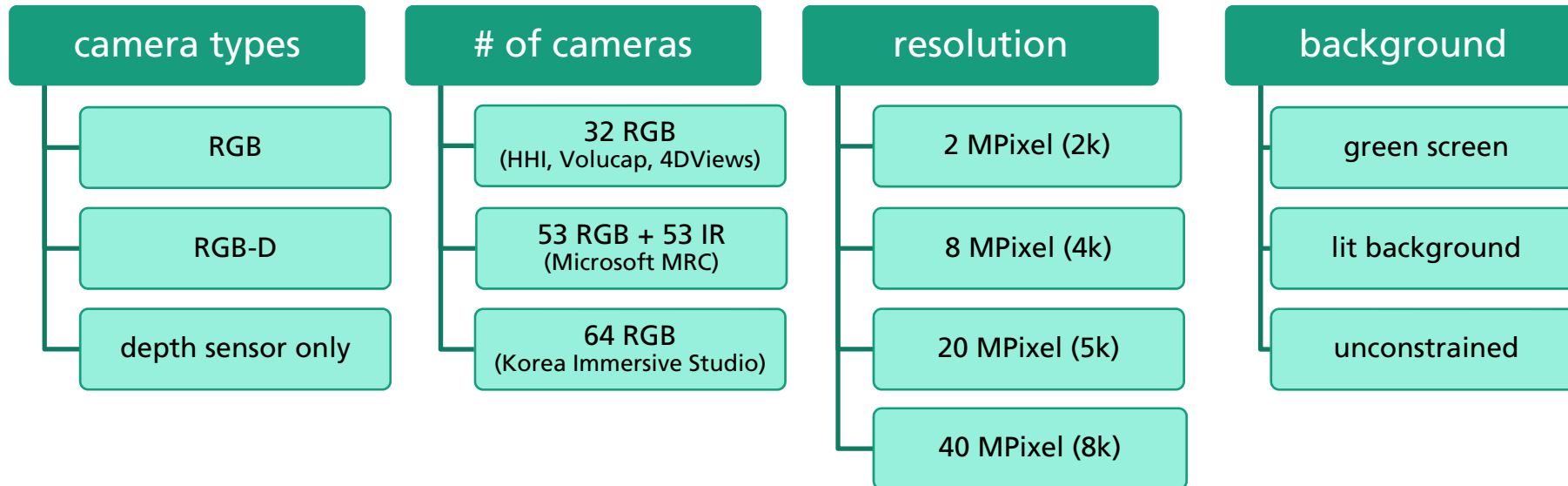


<https://arcturus.studio/blog/2022/04/12/volumetric-capture-studio-near-you/>

New studio @ Fraunhofer HHI

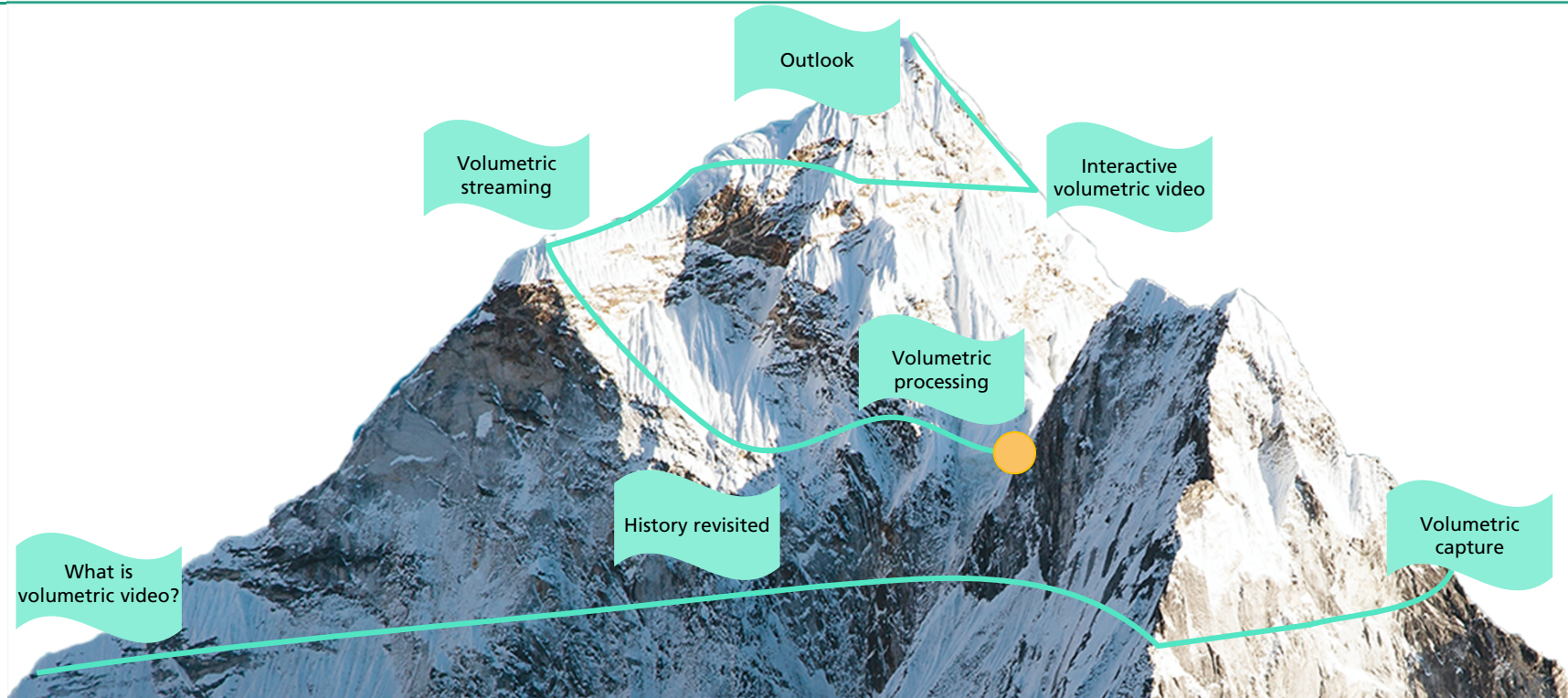


Volumetric capture

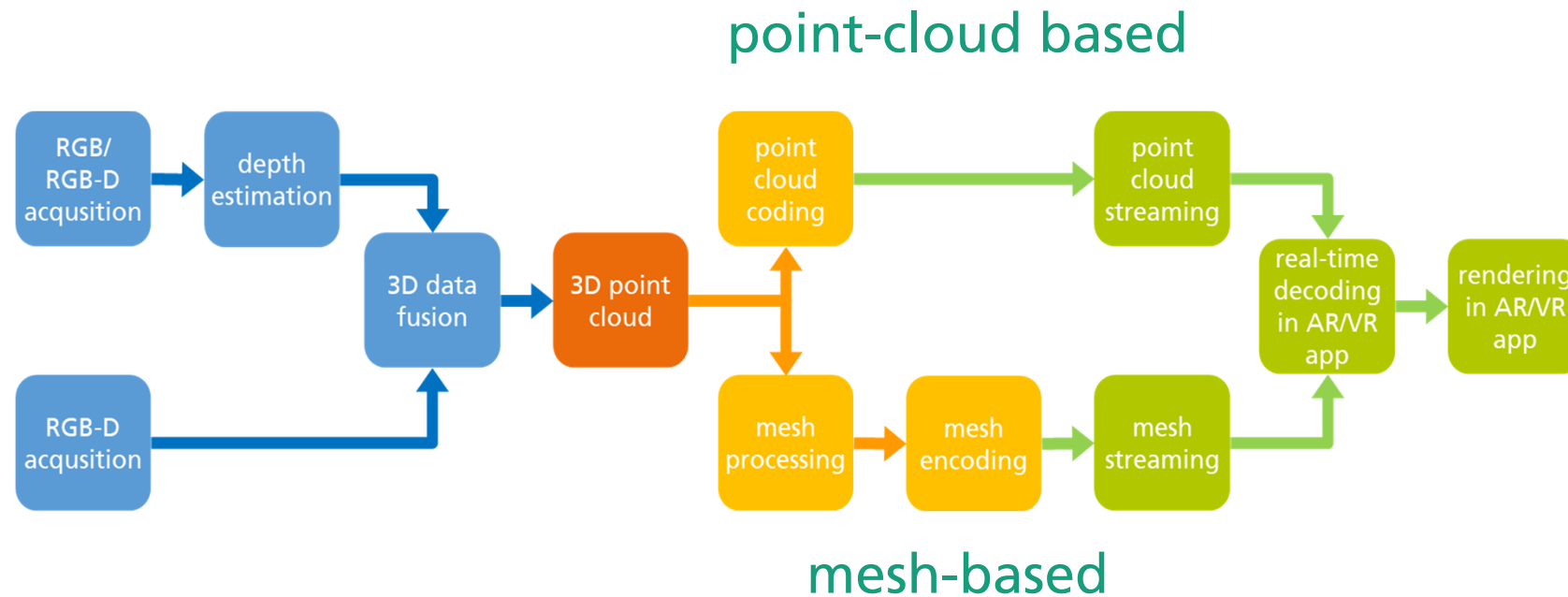


Let's climb the volumetric mountain

Volumetric processing

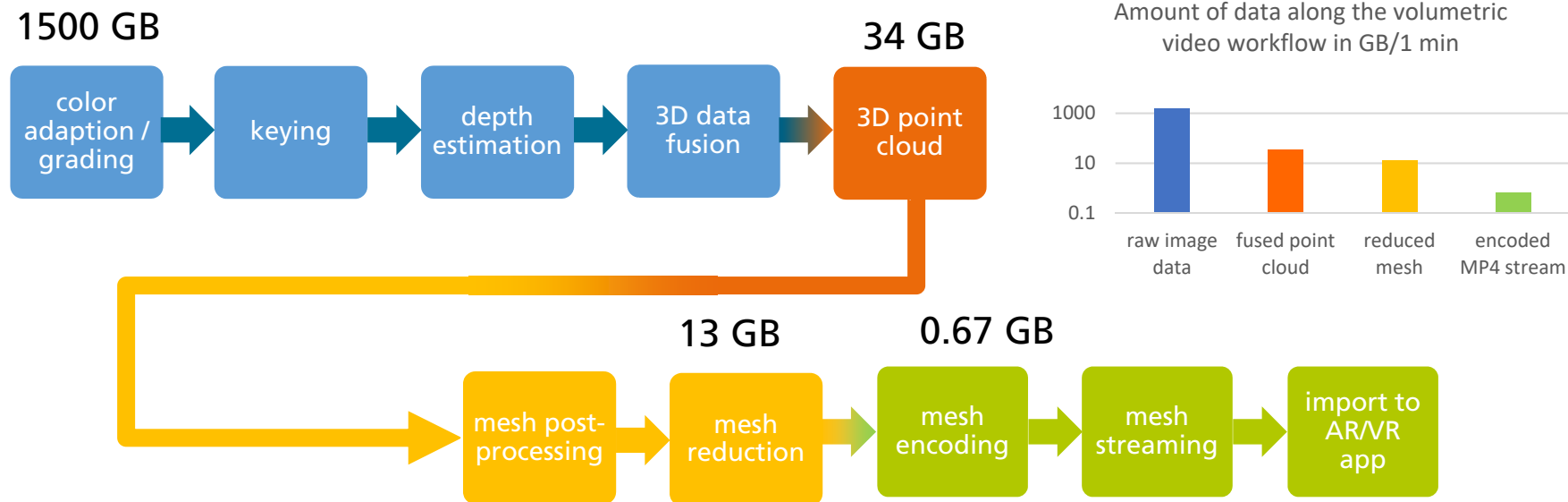


Volumetric Video Production and Workflow



Volumetric Video Production and Workflow

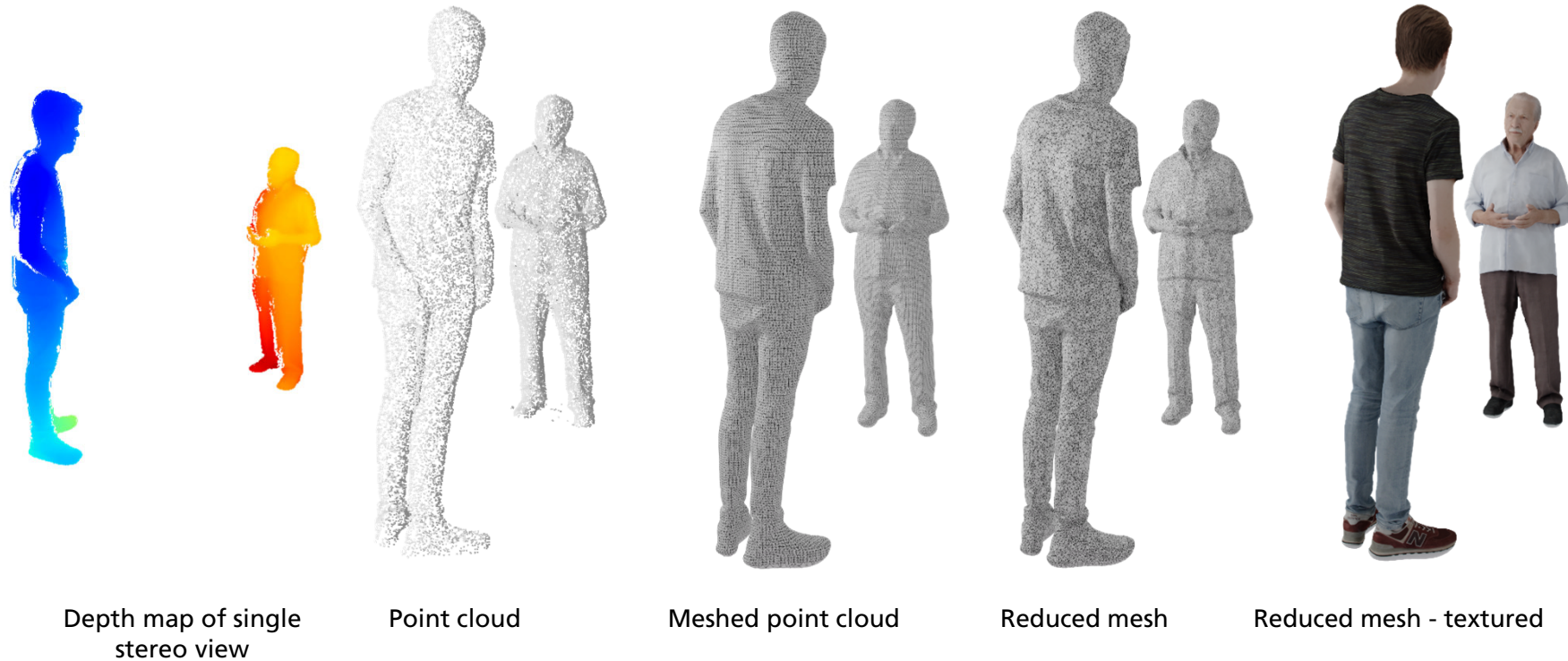
Interactive volumetric video requires mesh-based workflow



■ O. Schreer et al., Advanced Volumetric Capture and Processing, SMPTE Motion Imaging Journal, Vol. 128, June 2019

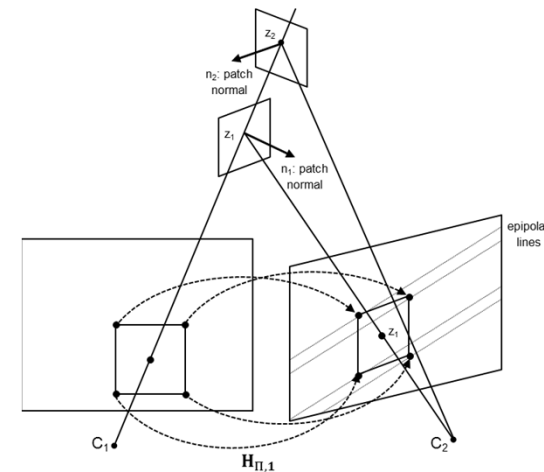
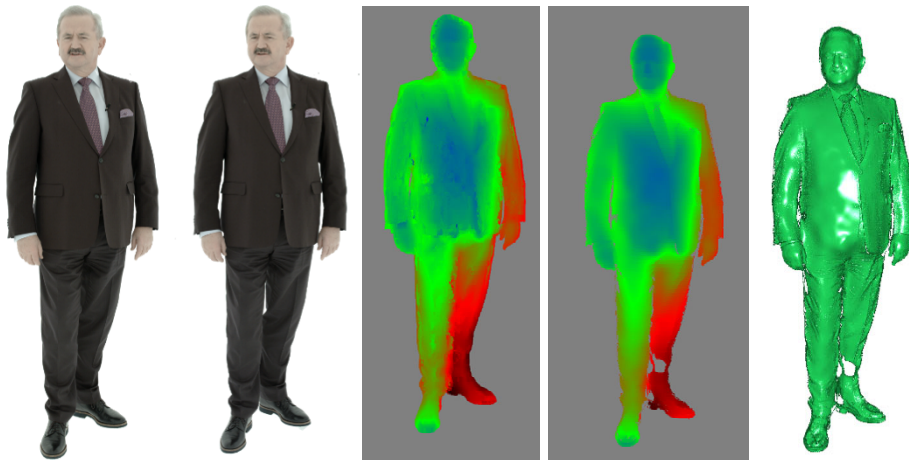
Volumetric Video Production and Workflow

Full End-to-End Automatic Processing Chain



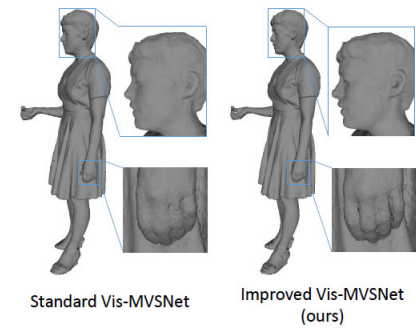
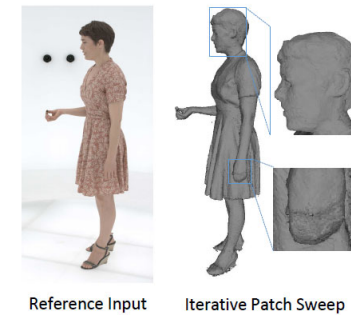
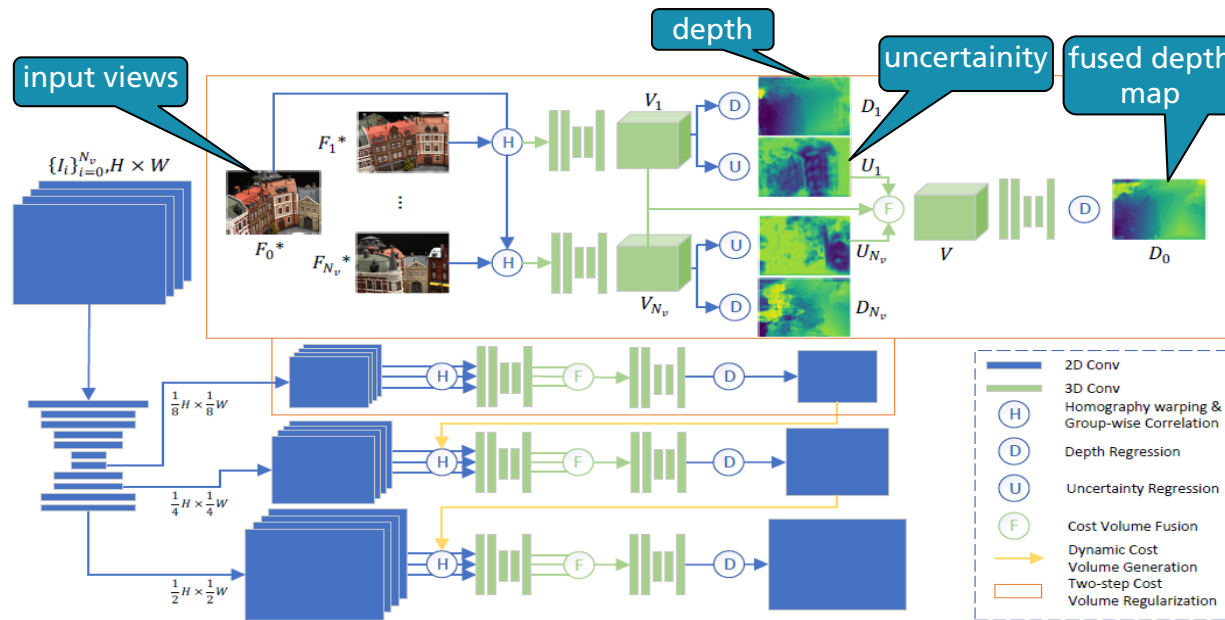
Stereo-based Depth Estimation (Surface-Region Matching)

- Spatial candidates and a statistically guided update for comparison (no search range)
- Spatial 3D patches are projected from $L \rightarrow R$ and $R \rightarrow L$ 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)

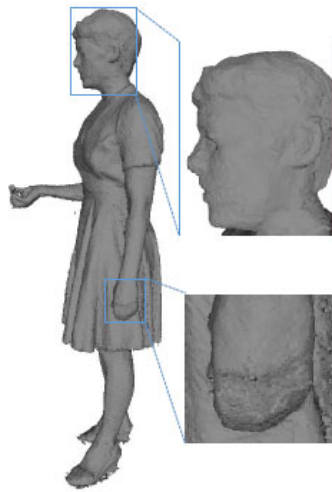


■ J. Zhang, Y. Yao, S. Li, Z. Luo, T. Fang. Visibility-aware multi-view stereo network, BMVC , 2020.

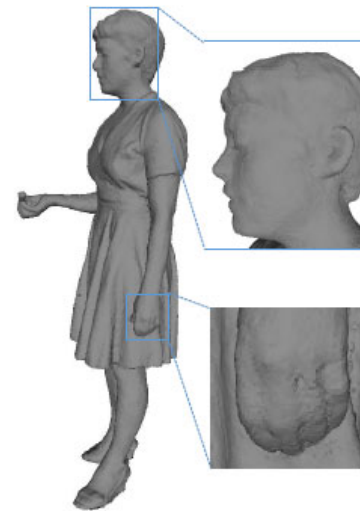
Learning-based Depth Estimation (Vis-MVS)



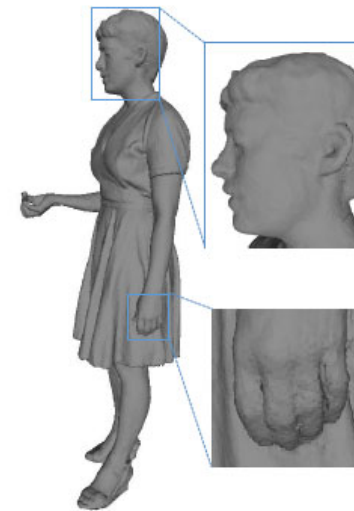
Reference Input



Iterative Patch Sweep



Standard Vis-MVSNet

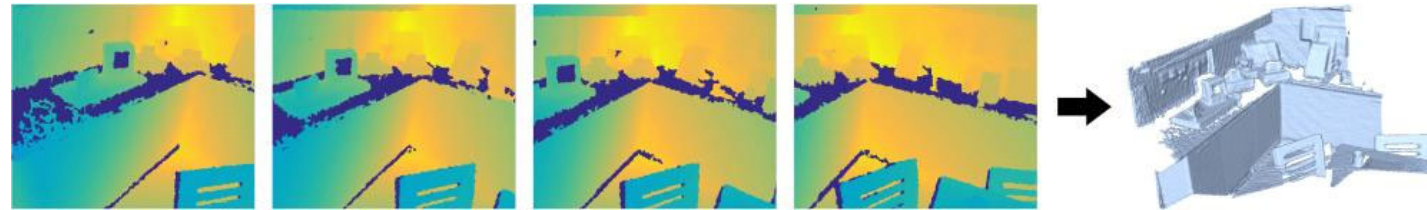


Improved Vis-MVSNet
(ours)

- J. Zhang, Y. Yao, S. Li, Z. Luo, T. Fang. Visibility-aware multi-view stereo network, BMVC , 2020.

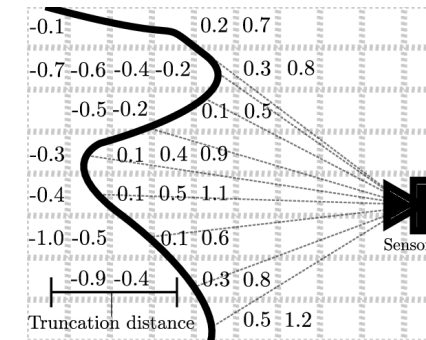
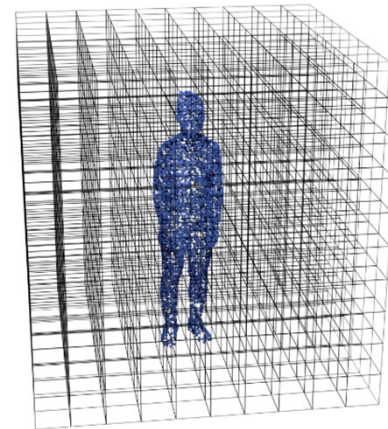
Point cloud fusion

Fuse multiple depth maps into a single coherent point cloud



Truncated Signed Distance Function (TSDF)

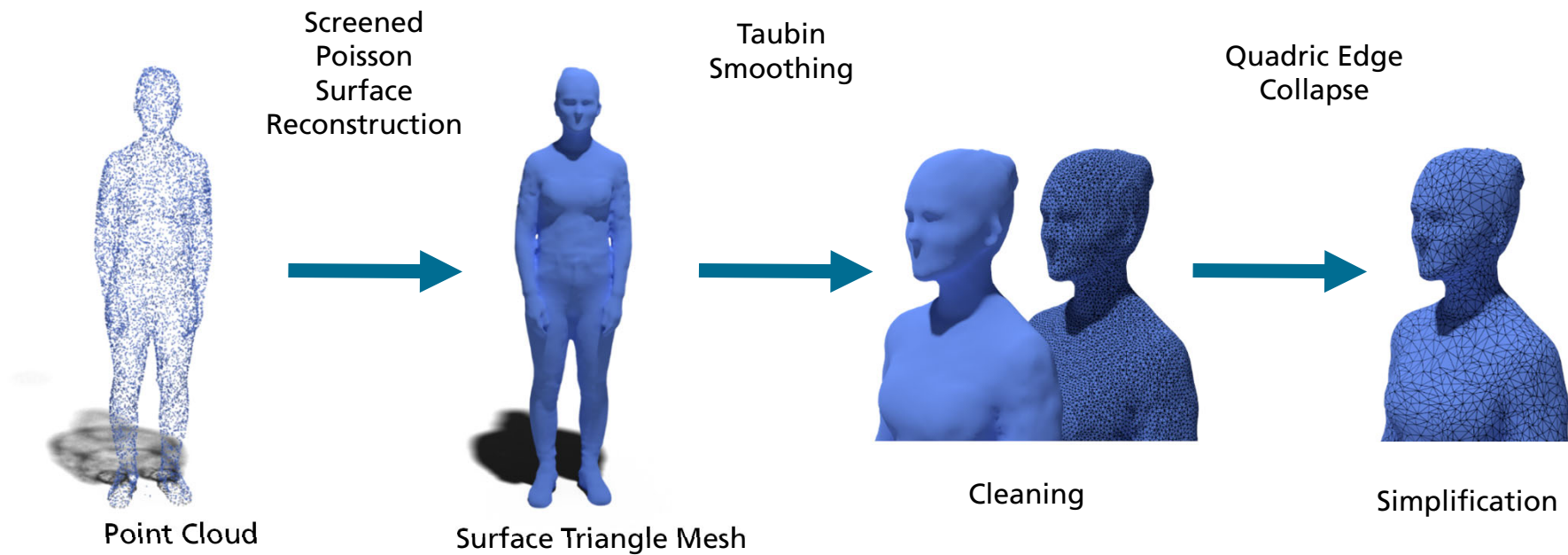
- Define voxel resolution
- Calculate signed distance to the nearest range surface
- Introduce weight depending on viewing and surface
- Extract isosurface with $D(x) = 0$



- A. Zeng, Sh. Song, M. Niessner, M. Fisher, J. Xiao, Th. Funkhouser, (2017). 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions, CVPR, 2017.
- Th. Whelan, M. Kaess, H. Johannsson, M. Fallon, J. Leonard, J. McDonald (2014). Real-time large-scale dense RGB-D SLAM with volumetric fusion. The International Journal of Robotics Research. 34. 598-626. 10.1177/0278364914551008.

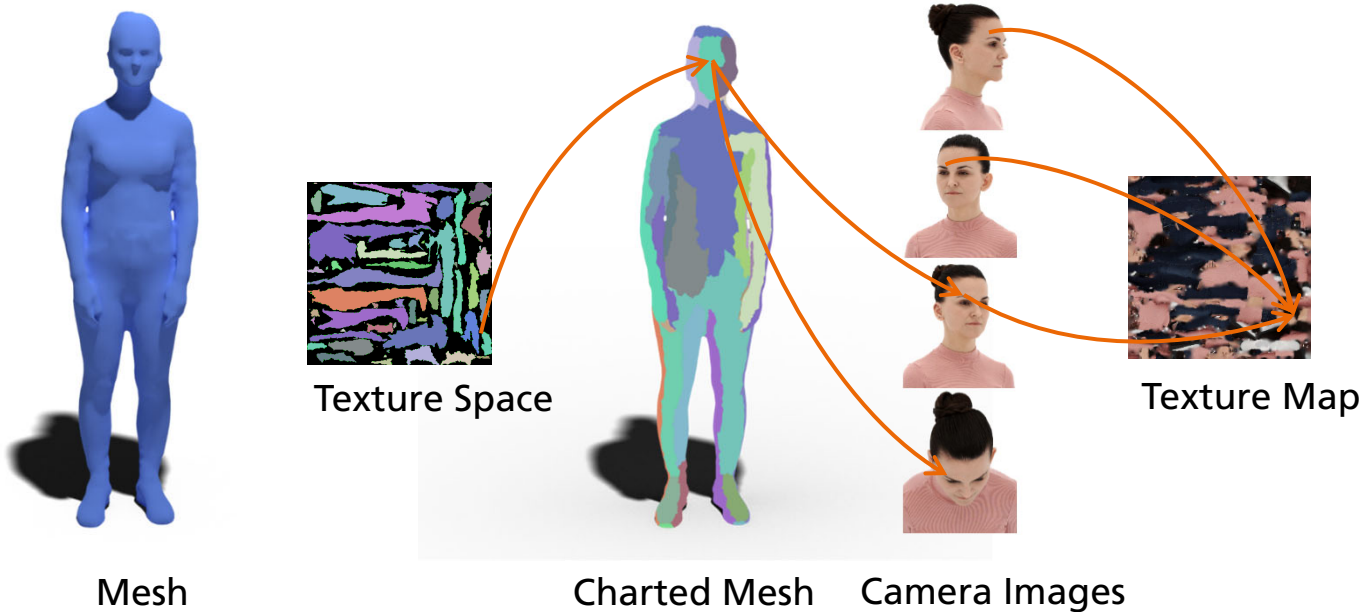
....

Meshing & Mesh Post-processing

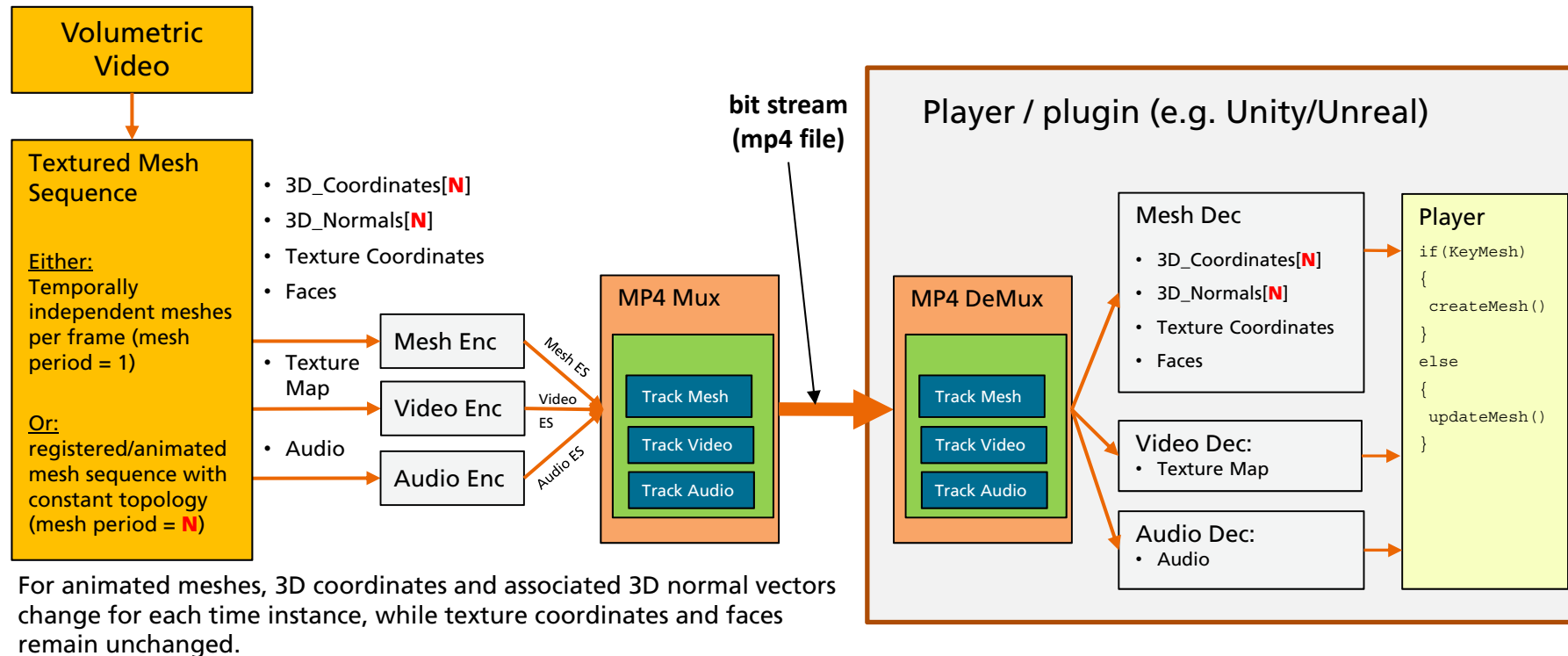


- Kazhdan M., Hoppe H., (2013), "Screened poisson surface reconstruction", ACM Trans. Graph. 32, 3 (July 2013).
- Taubin, G., (1995), "Curve and surface smoothing without shrinkage," Proceedings of IEEE Int. Conf. on Computer Vision, 1995, pp. 852-857, doi: 10.1109/ICCV.1995.466848.
- Garland, M., Heckbert, P. S., (1997), "Surface simplification using quadric error metrics.", in Proc. of the 24th annual Conf. on Computer graphics and interactive techniques.
- Catmull, E. E., 1974, "A Subdivision Algorithm for Computer Display of Curved Surfaces".

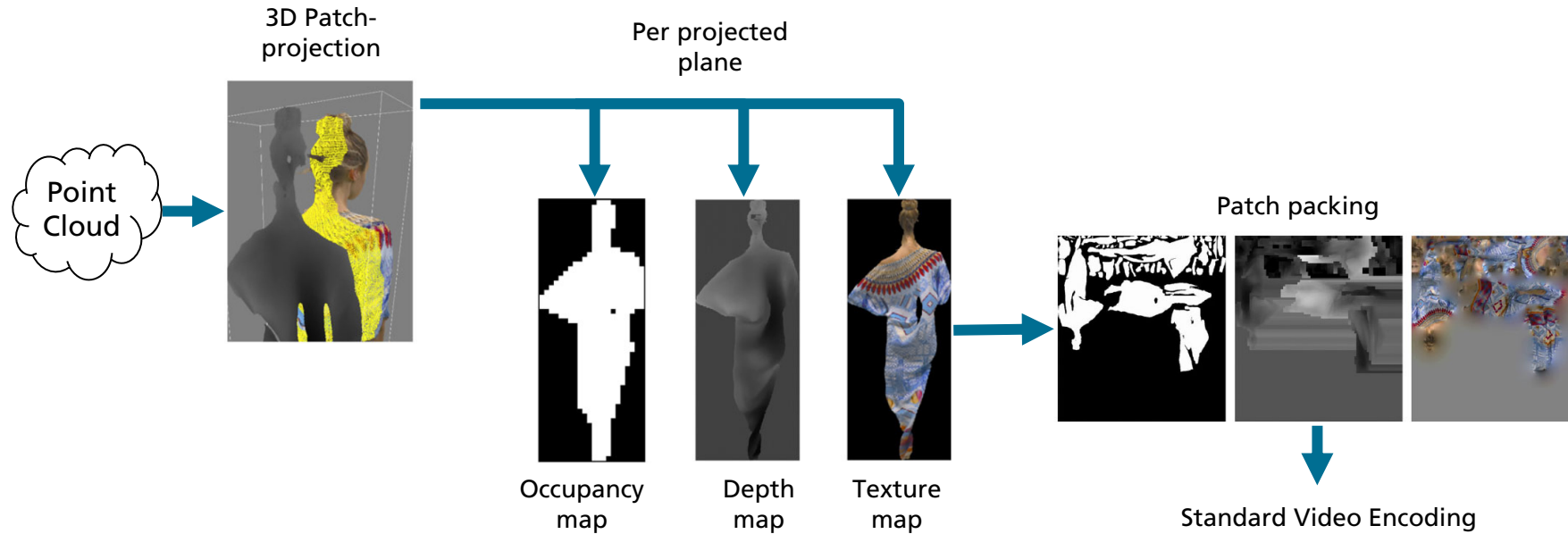
Charting and texturing



Mesh Coding and Transmission



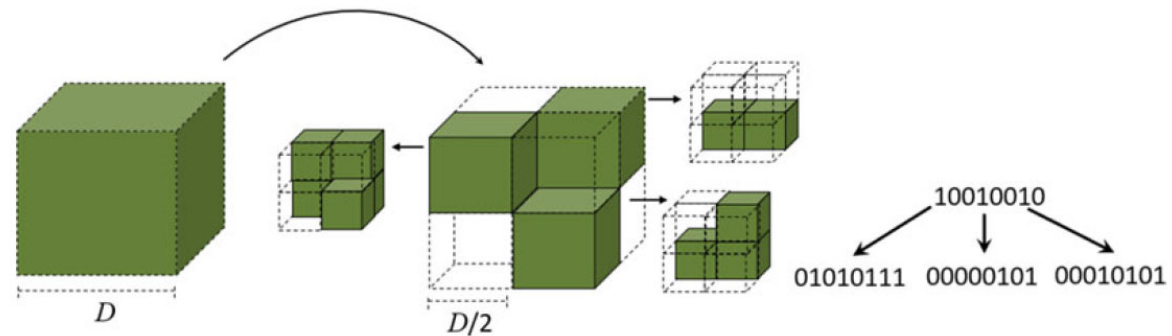
Video-based Point Cloud Coding (V-PCC)



- Graziosi, D., Nakagami, O., Kuma, S., Zaghetto, A., Suzuki, T., & Tabatabai, A. (2020). An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC). *APSIPA Transactions on Signal and Information Processing*, 9, E13. doi:10.1017/ATSIP.2020.12

Geometry-based Point Cloud Coding (G-PCC)

- Point cloud can be represented by an octree or trisoup structure
- Lossless entropy encoding → reduction by 1/10
- Encoding by
 - region-adaptive hierarchical transform
 - Predicting transform
 - Lifting transform



Courtesy of D. Graziosi et al, 2020

Volumetric video for VR



Lana, 2019, HHI and Volucap

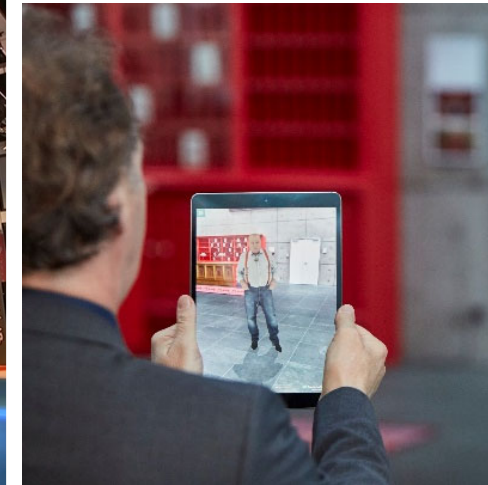


Tiger Tale, 2018, HHI and Trotzkind

Volumetric video for AR



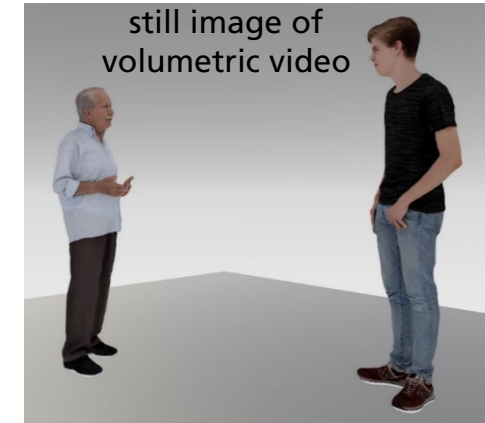
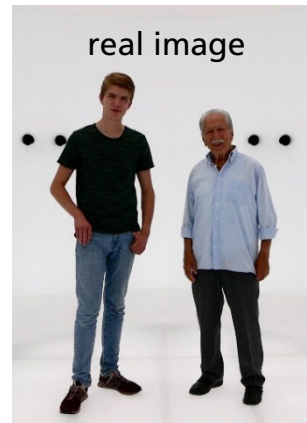
Linda Zervakis, german TV anchor woman, Volucap



Uli Wegner, boxing legend, 2018, HHI and Inflections

VR Experience “Ernst Grube – the legacy”

- Volumetric video is the relevant technology to preserve memories from the past
- Responsible concept design for story telling of dark periods of history
- Interview with Ernst Grube, one of the last German survivors of the Holocaust.
- The complete VR production “ERNST GRUBE – THE LEGACY” consists of five episodes lasting about 8-12 minutes each.
- Worldwide longest VR Experience using volumetric video



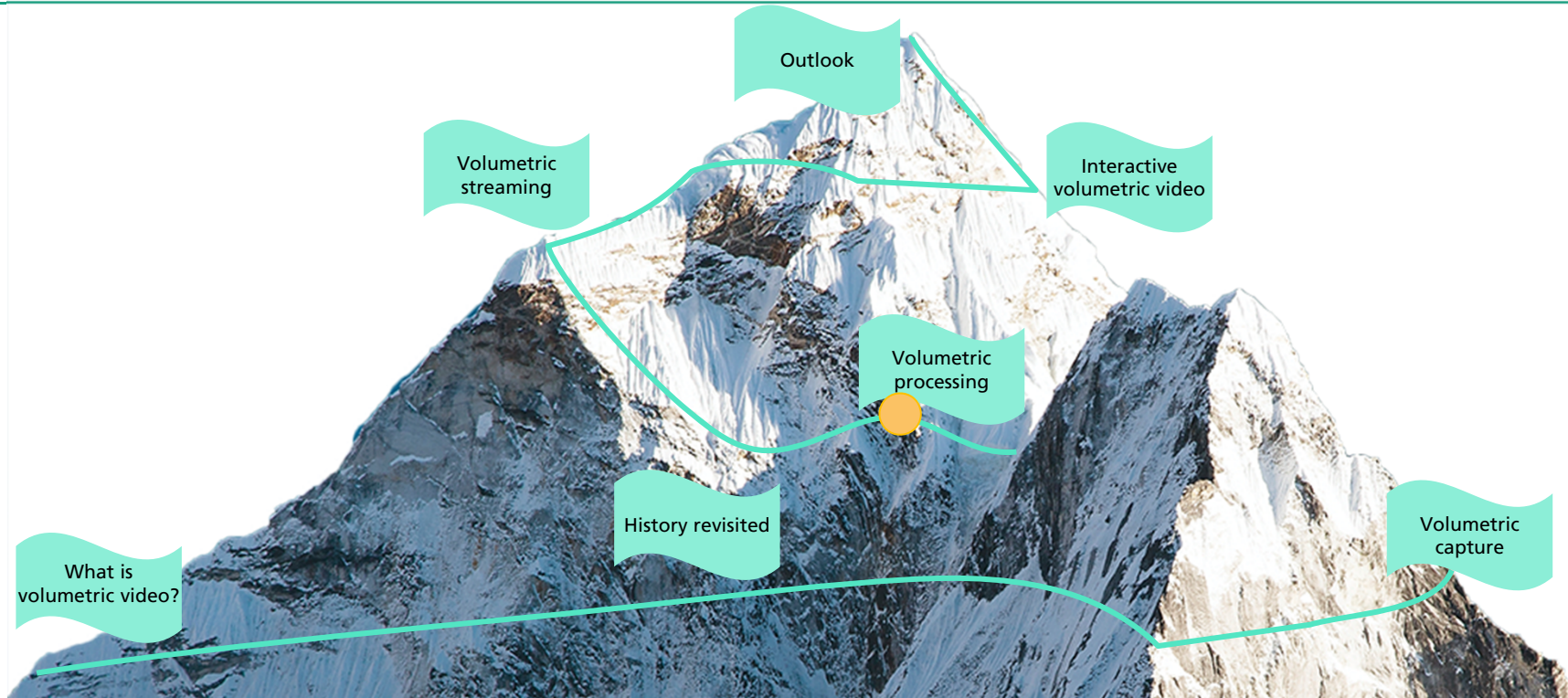
Preserving memories of the past



Ernst Grube - The Legacy

Let's climb the volumetric mountain

Volumetric streaming



Volumetric Video Streaming for Mixed Reality



■ Problem

- Complex rendering

■ Solution

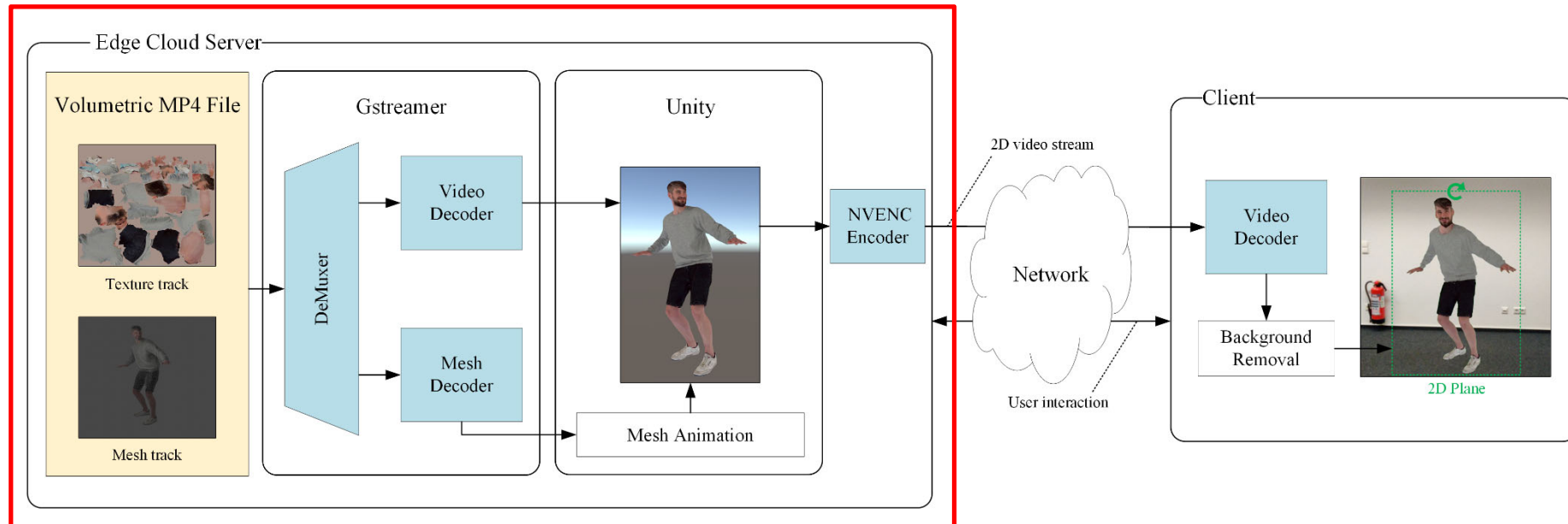
- Offload rendering to cloud server
- Transmit 2D video stream based on user's head pose

■ Drawback

- Requires ultra low latency system and network

Streaming of Interactive Volumetric Video

Split Rendering for Mixed Reality



- S. Gül et al. "Interactive Volumetric Video from the Cloud", IBC, September 2020.
- S. Gül et al., Cloud Rendering-based Volumetric Video Streaming System for Mixed Reality Services, ACM MMSys 2020

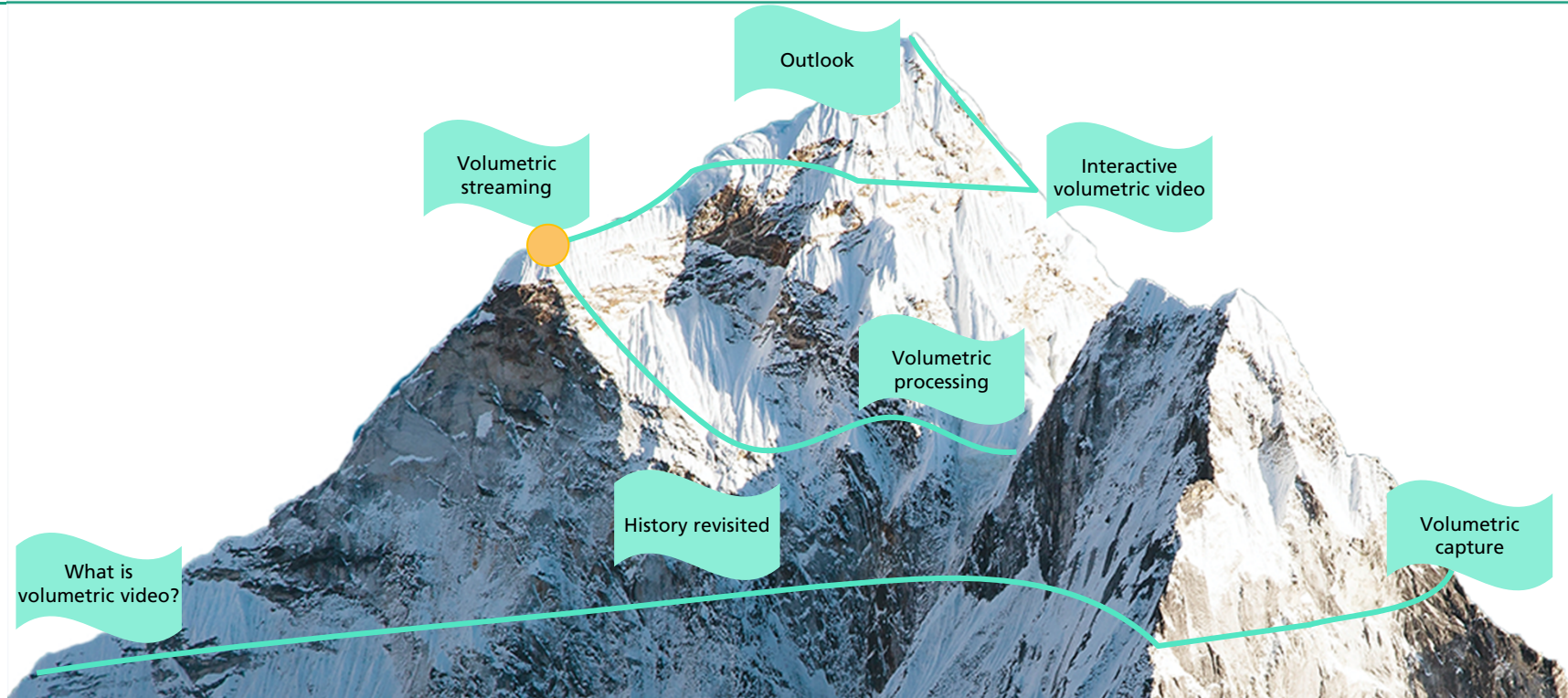
Demo: Split rendering for Mixed Reality



■ J. Son et al. , "Split Rendering for Mixed Reality: Interactive Volumetric Video in Action" SIGGRAPH Asia 2020 XR, December 2020

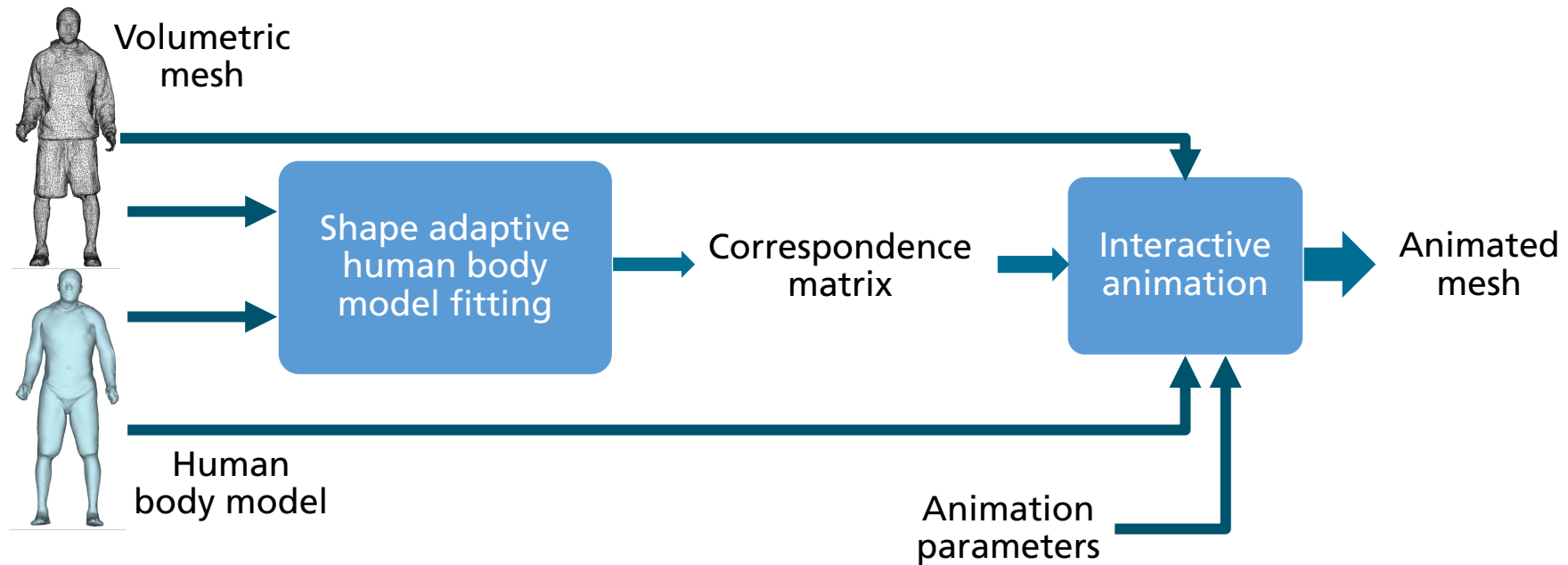
Let's climb the volumetric mountain

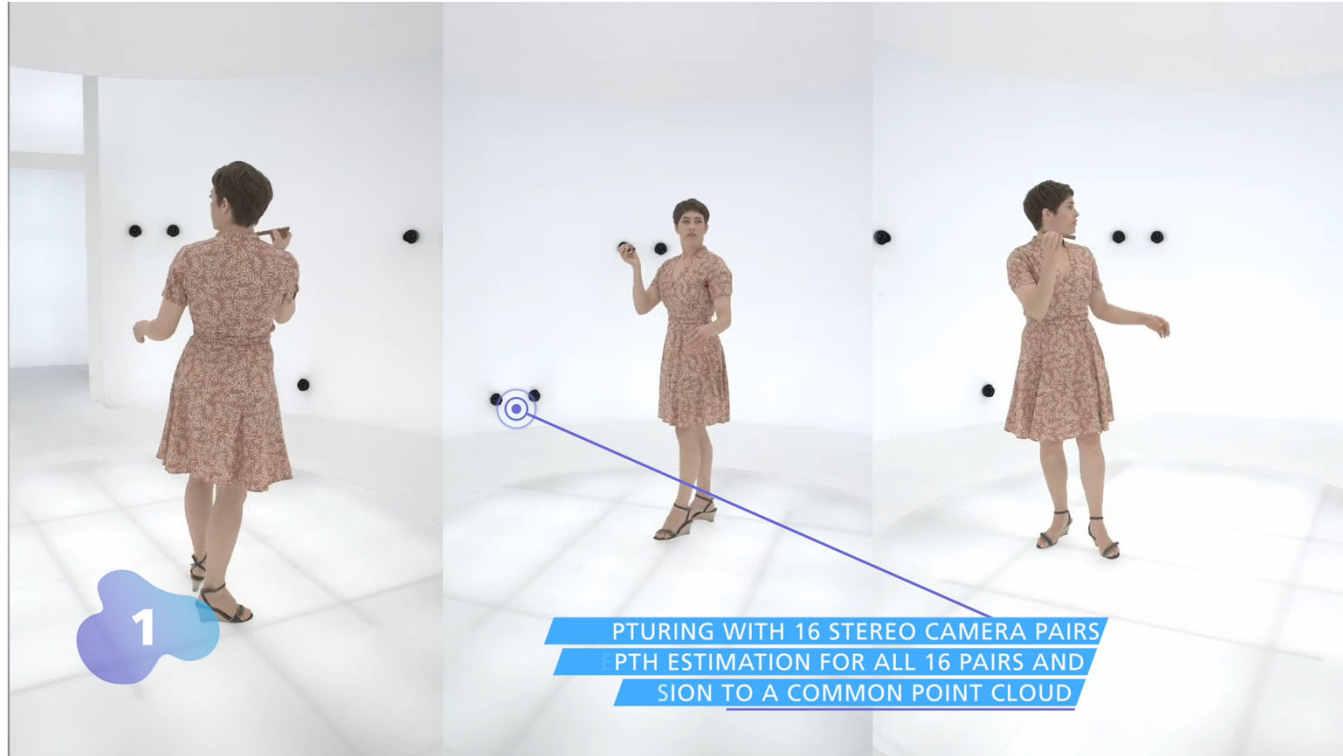
Interactive volumetric video



Interactive Animation of Volumetric Video

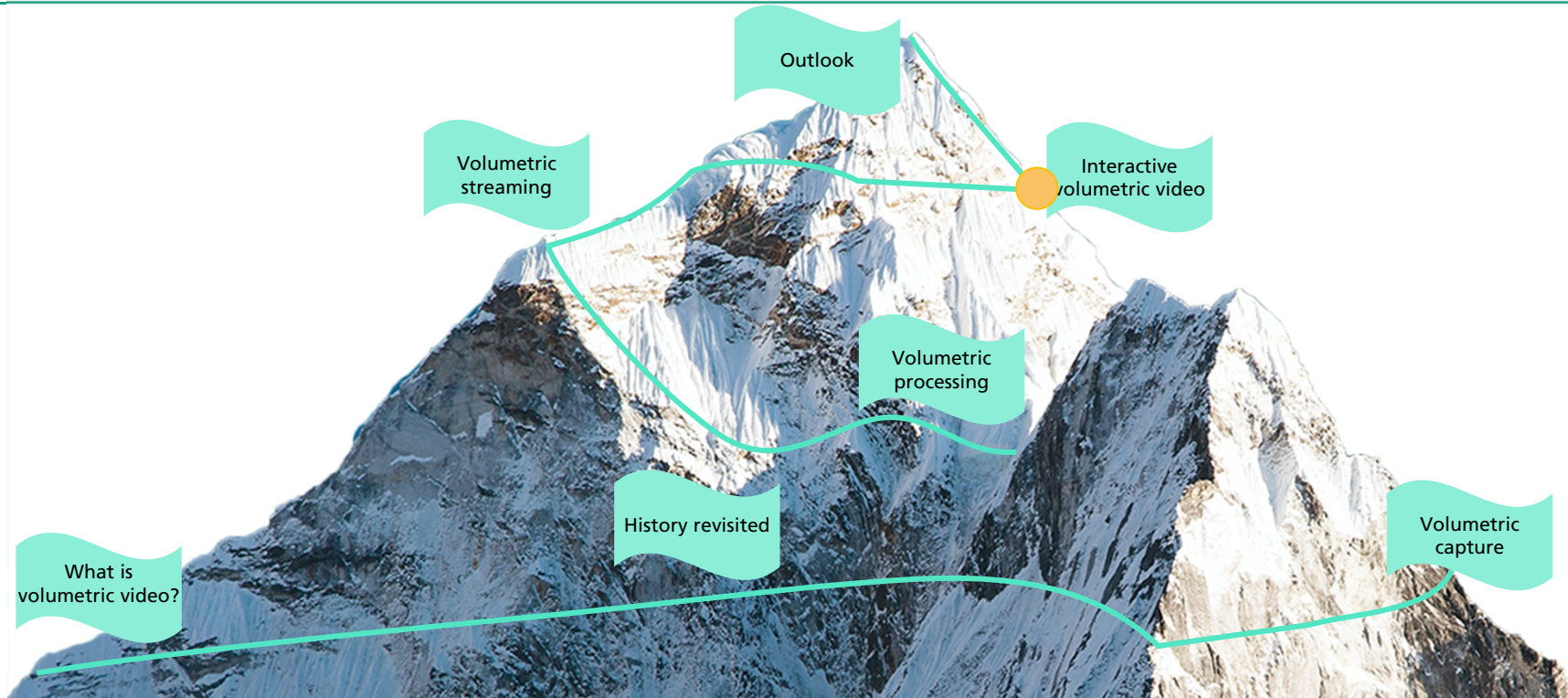
Workflow





Let's climb the volumetric mountain

Outlook



Outlook

- What are the major challenges?
 - Mobile/portable and cheaper volumetric capture
 - High-quality reconstruction with fewer cameras, or even just one?
 - Exploit AI in various processing steps or even the complete reconstruction?
 - 3D reconstruction of complex structures eg. hairs, moving clothes, reflective/transparent surfaces → NeRF
 - Estimation of material properties for high-quality rendering
 - High-quality for real-time telepresence, not just fused point clouds from RGBD

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute, 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:

Anna Hilsmann, Decai Chen, Rodrigo Diaz, Peter Eisert, Ingo Feldmann, Serhan Gül, Cornelius Hellge, Son Jangwoo, Wieland Morgenstern, Sylvain Renault, Markus Worchel, Marcus Zepp