IEEE CONFERENCE ON VIRTUAL REALITY + 3D USER INTERFACES
12-16 MARCH 2022 | VIRTUAL EVENT

Microsoft Research

VirtualCube: An Immersive 3D Video Communication System

Yizhong Zhang[1*] Jiaolong Yang[1*] Zhen Liu[2,1] Ruicheng Wang[3,1]
Guojun Chen[1] Xin Tong[1] Baining Guo[1]

[1] Microsoft Research Asia [2] Nanjing University [3] USTC

([*] Joint first authors with equal contribution)

Background

World Security Council of *S.H.I.E.L.D*
(marvel.fandom.com)

Conference in *Kingsman*
(kingsman.fandom.com)

Jedi Council in *StarWars*
(starwars.fandom.com)

Our Goal

Enabling people in different places to meet with each other as if they were in the same room

- Natural eye contact and high-fidelity facial expressions
- Side conversions and attention switching in multi-person meetings
- Versatile meeting setups



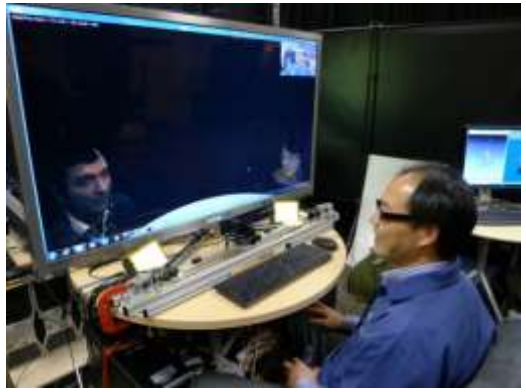
Related Works

[W.-C. Wen et. al, 2000]

[A. Sadagic et. al, 2001]

[H. Baker et. al, 2002]

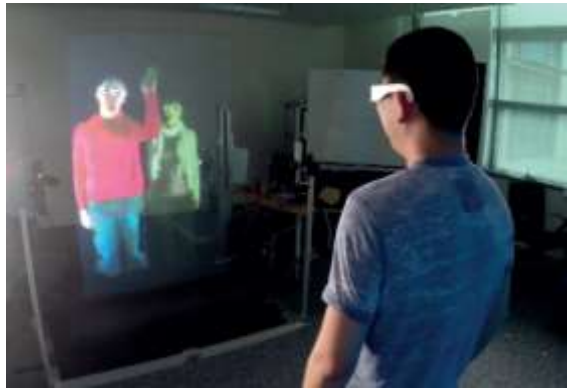
[M. Kuechleret. al, 2006]



[Cha et. al, 2012]

[S. Beck et. al, 2013]

[S. Orts-Escolano et. al, 2016]

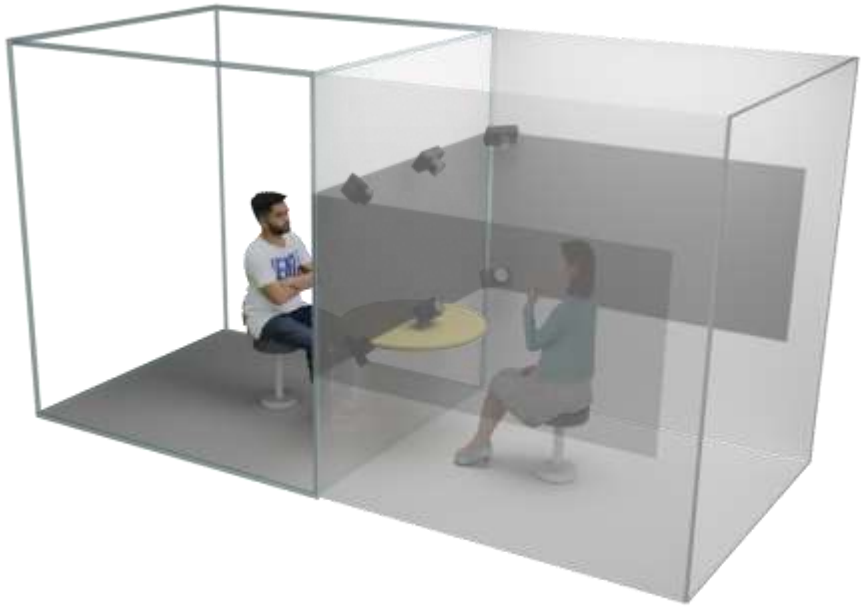


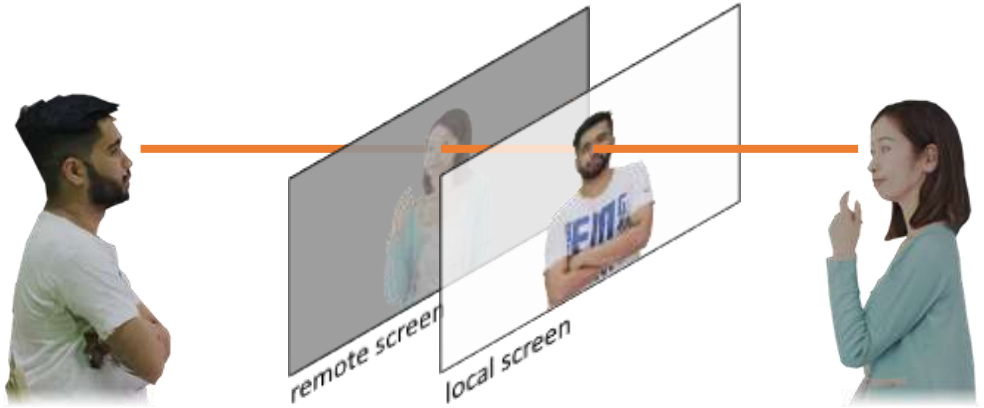
[C. Plüss et. al, 2016]

Key Idea

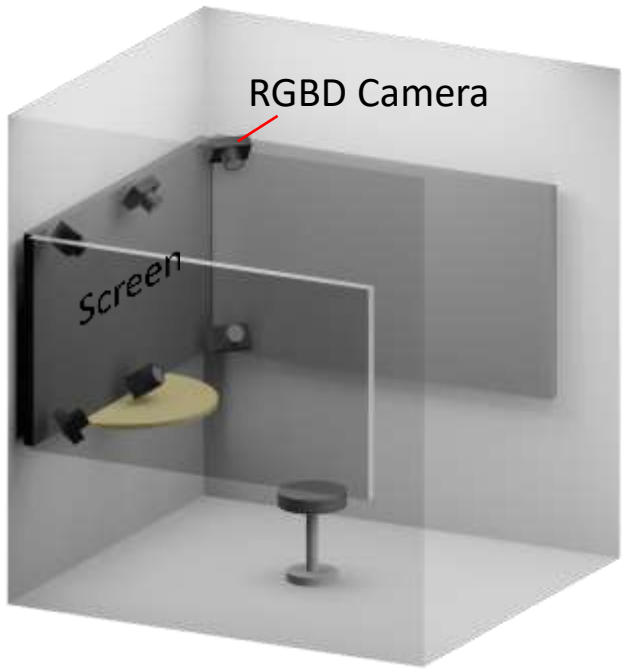
Displaying remote participants as if seeing through the screen

- Mapping all participants in one shared virtual environment
- Life-sized rendering of remote participants from local user's view





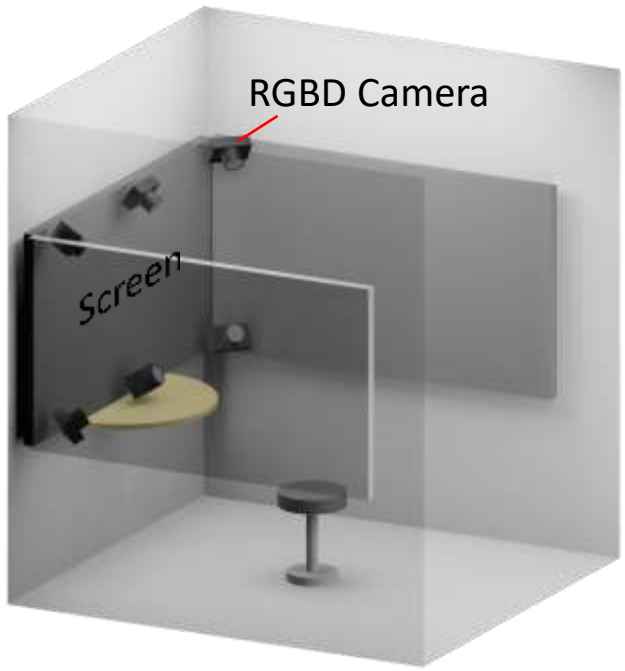
System Hardware



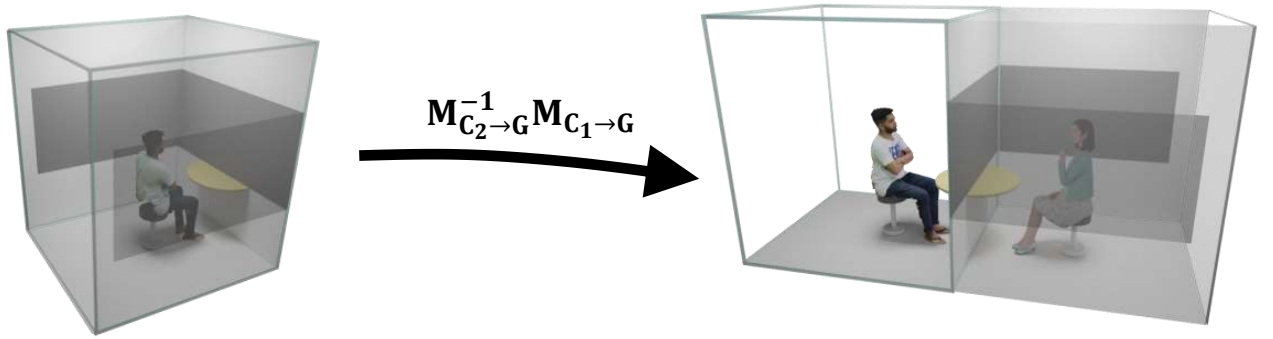
Design prototype

Inside VirtualCube

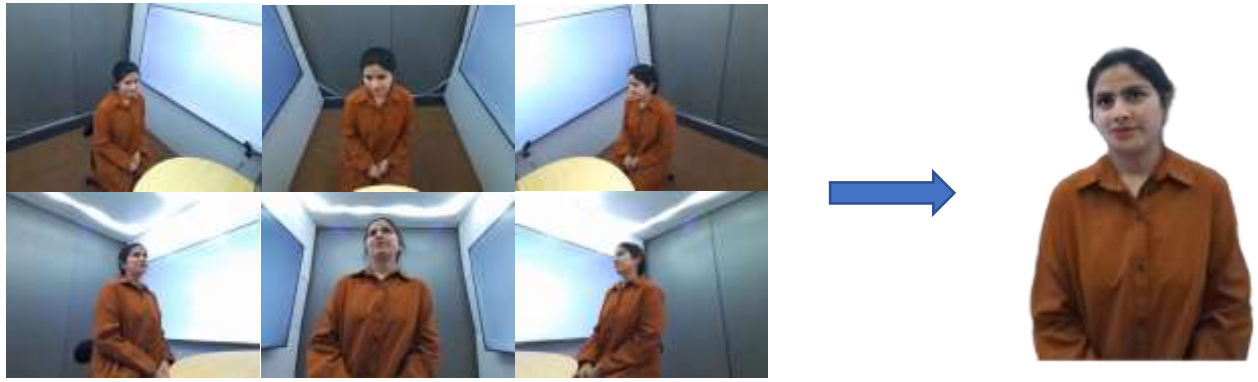
Key Components



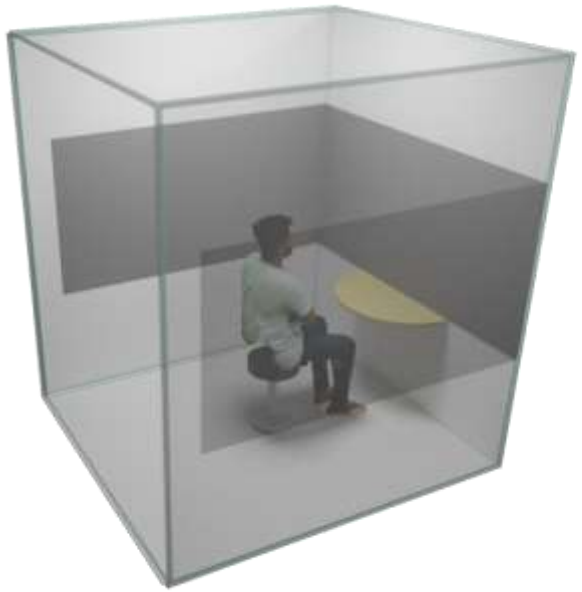
V-Cube Assembly



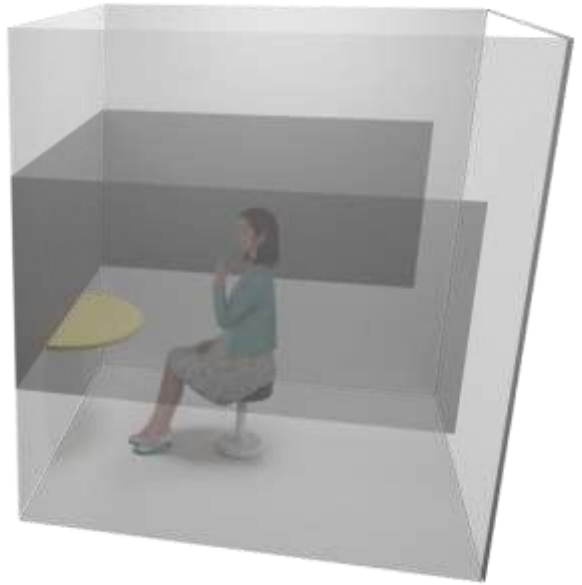
V-Cube View



V-Cube Assembly

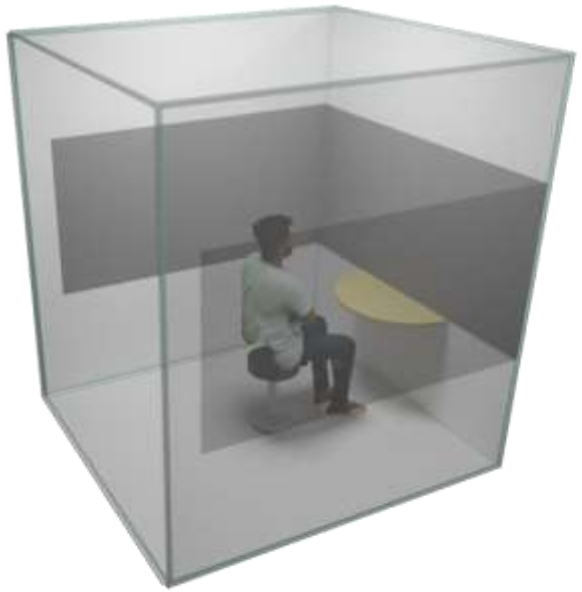


Remote User C1



Local User C2

V-Cube Assembly

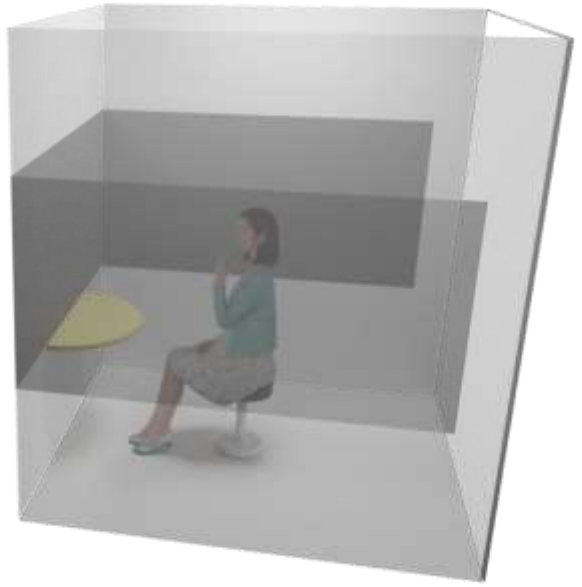


Remote User C1

$\mathbf{M_{C_1 \rightarrow G}}$

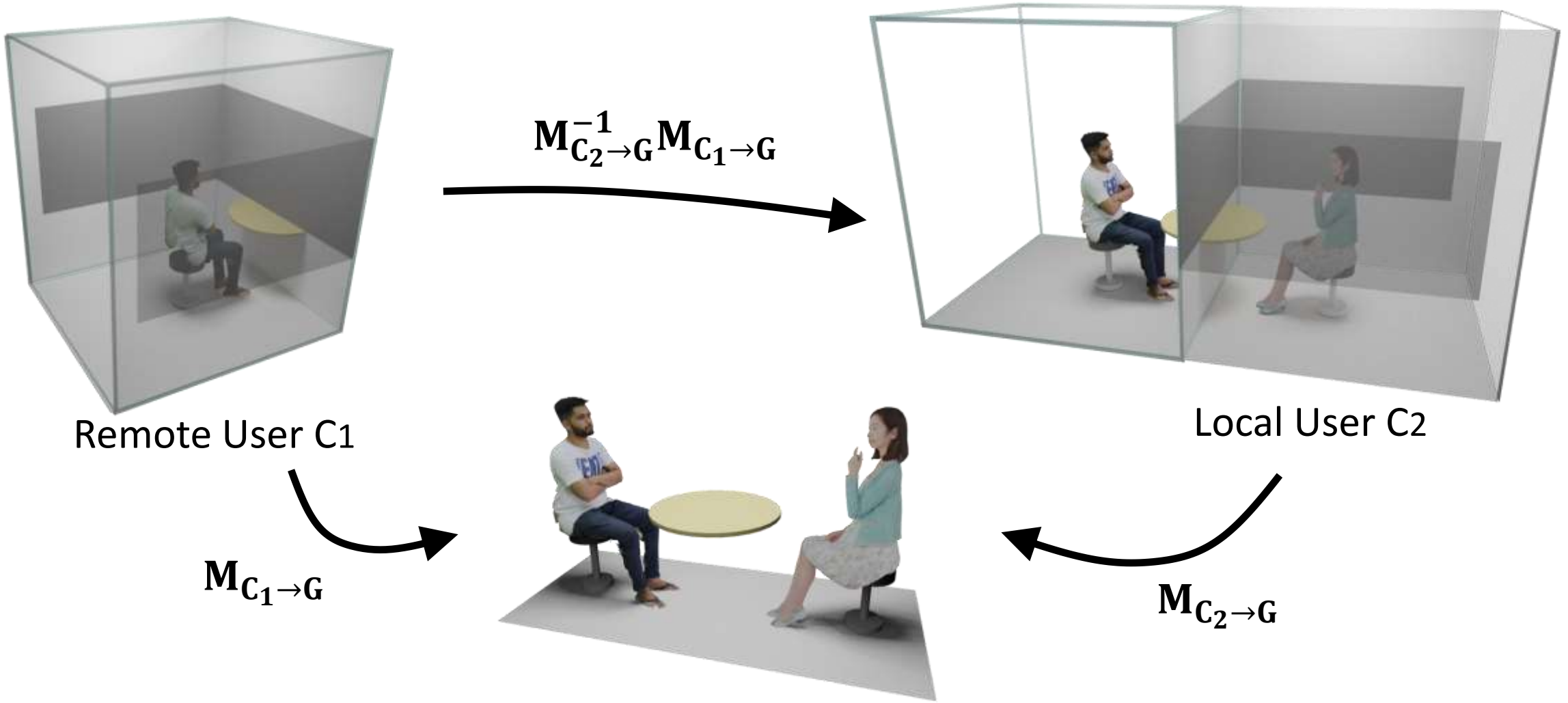
Global Virtual Environment

$\mathbf{M_{C_2 \rightarrow G}}$

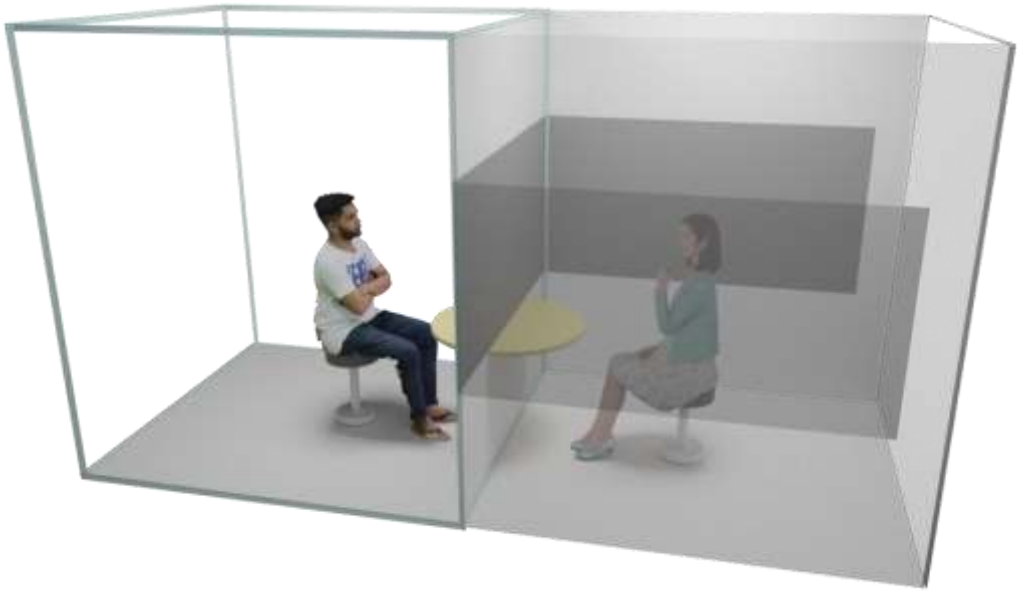


Local User C2

V-Cube Assembly

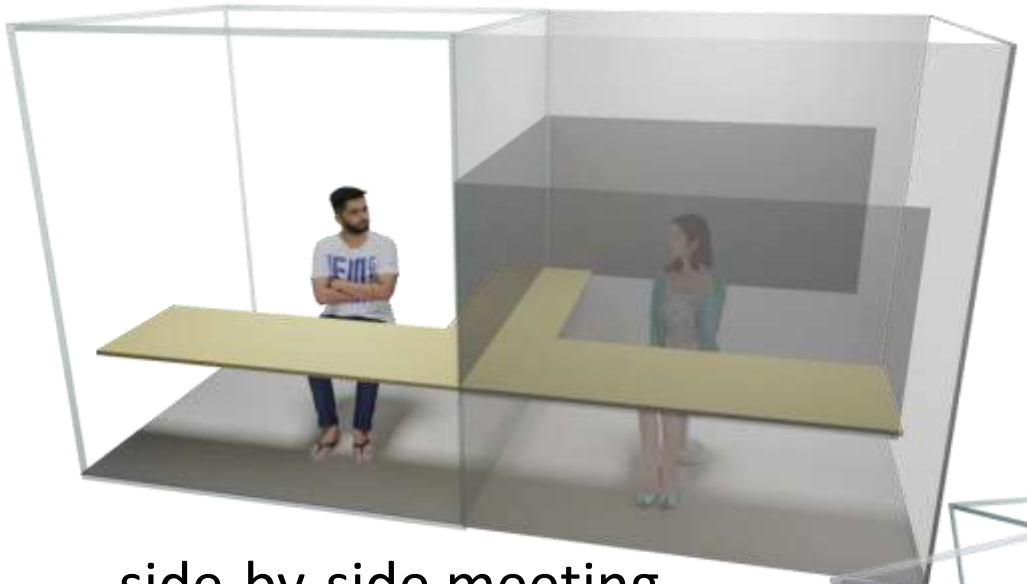


V-Cube Assembly

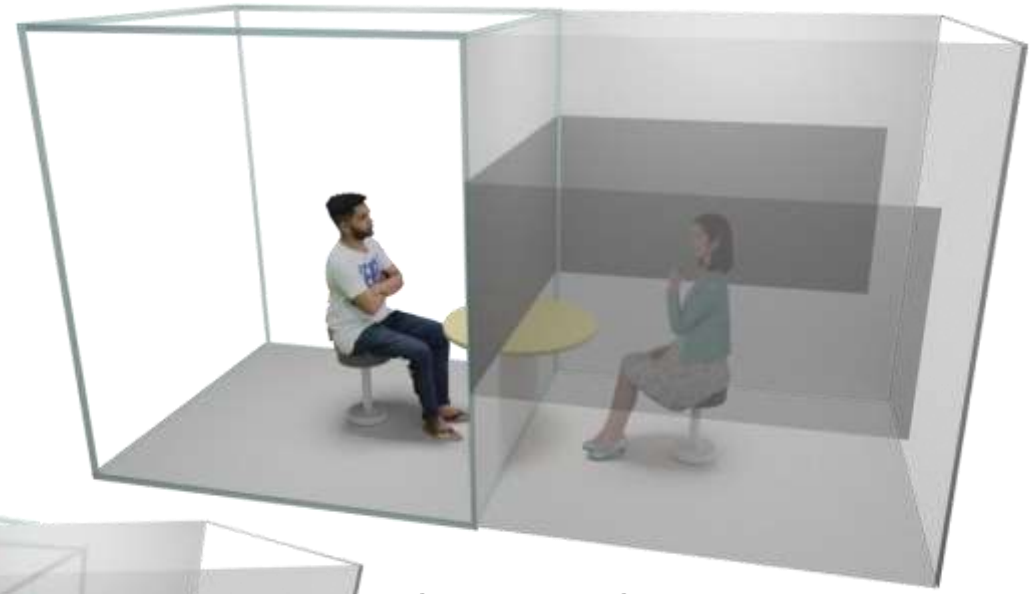


face-to-face meeting

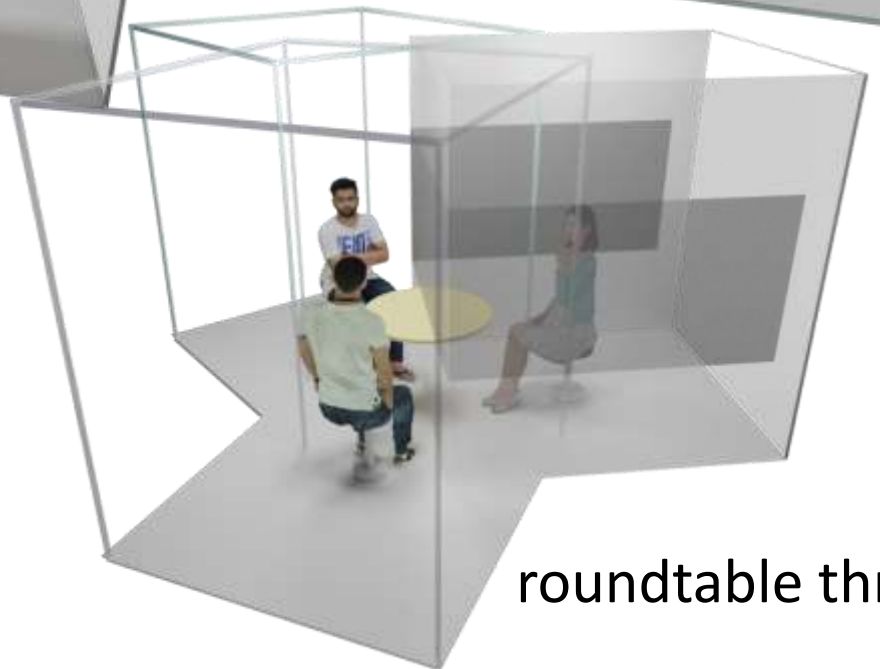
V-Cube Assembly



side-by-side meeting

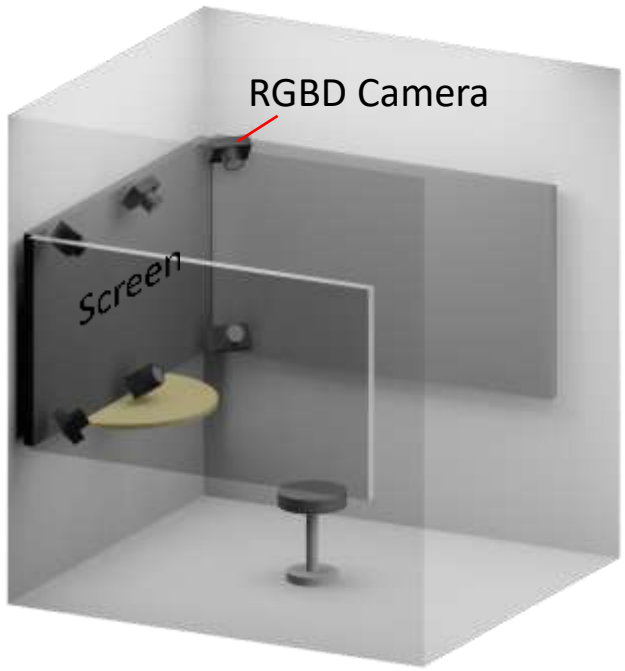


face-to-face meeting

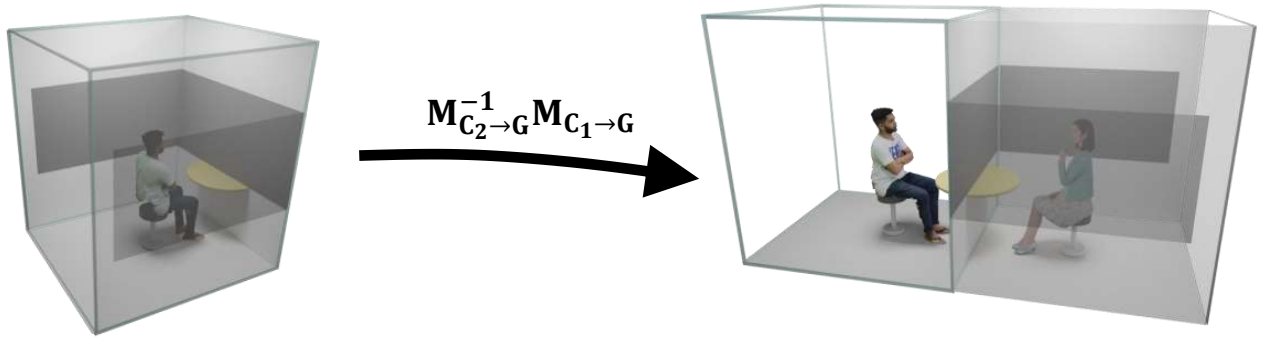


roundtable three-person meeting

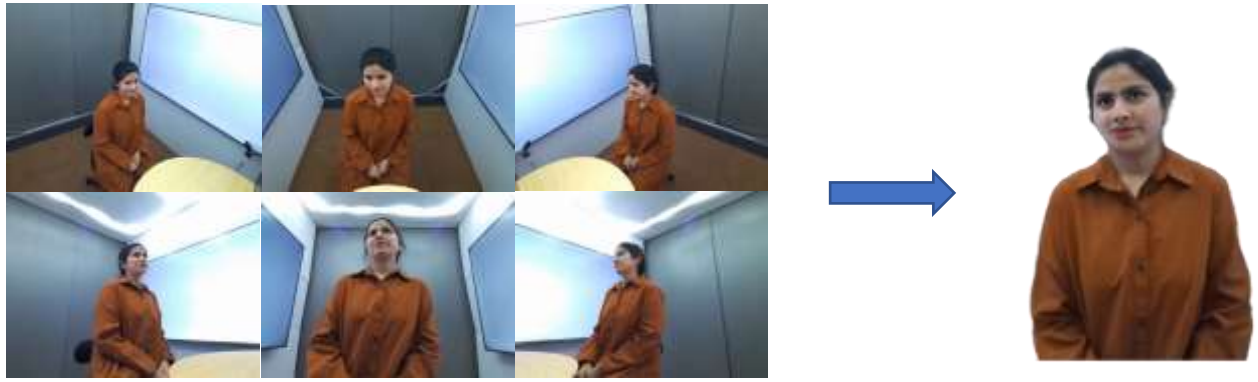
Key Components



V-Cube Assembly



V-Cube View



V-Cube View for 3D Free-View Rendering

Real-time Geometry Proxy

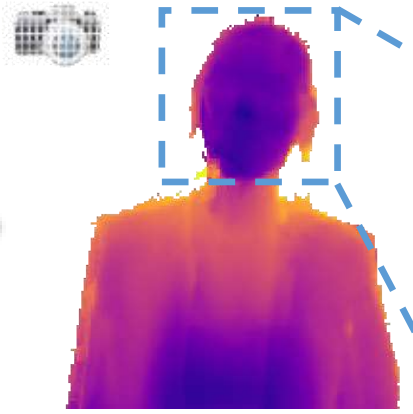
Lumi-Net Rendering

Deep Image Enhancement

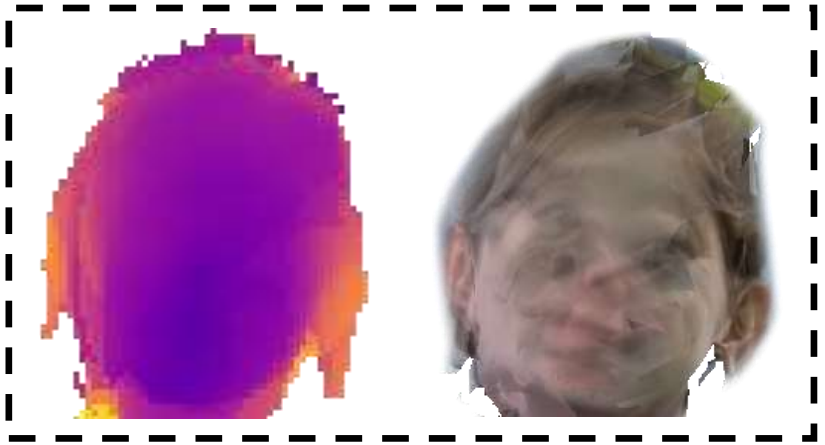
Real-time Geometry Proxy

- Virtual-view depth estimation
  - Fast multiview stereo
  - Used to determine visibility and weight for Lumigraph data blending

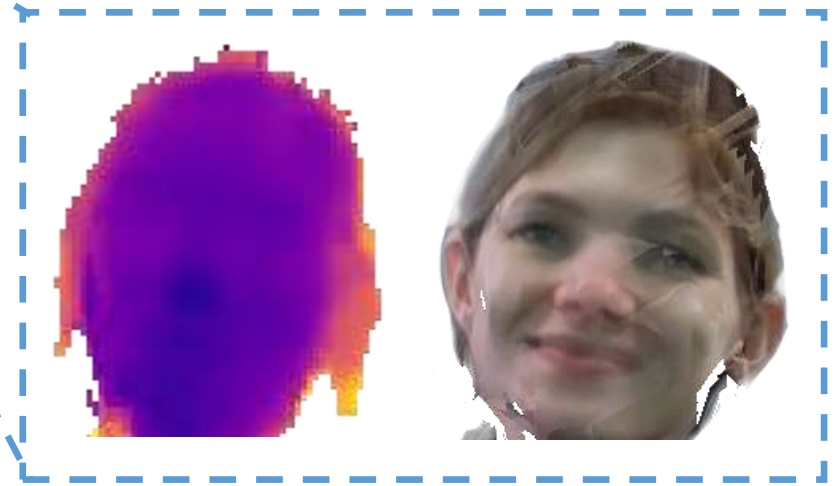
MVS CNN



Estimated Virtual-view Depth

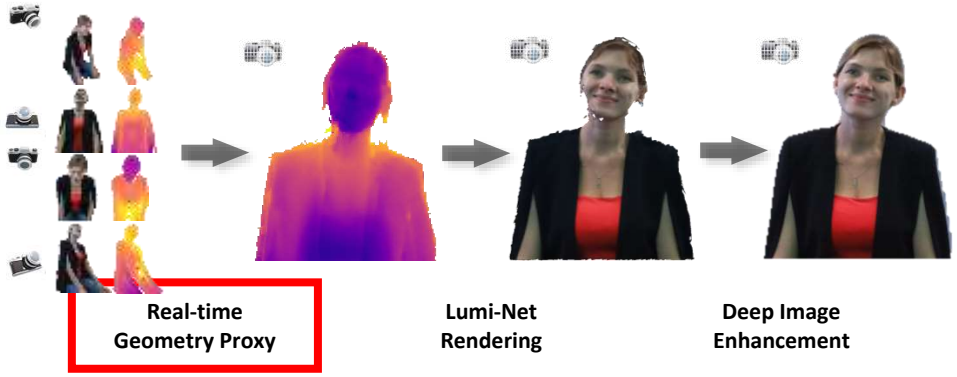


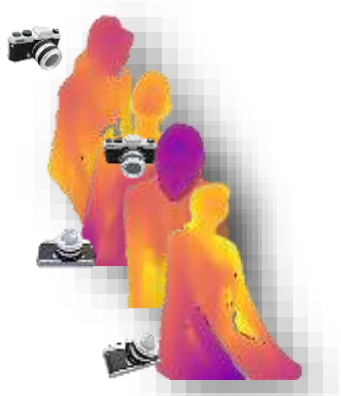
With warped Kinect depth (Texture averaging)



With our estimated depth (Texture averaging)

Real-time Geometry Proxy

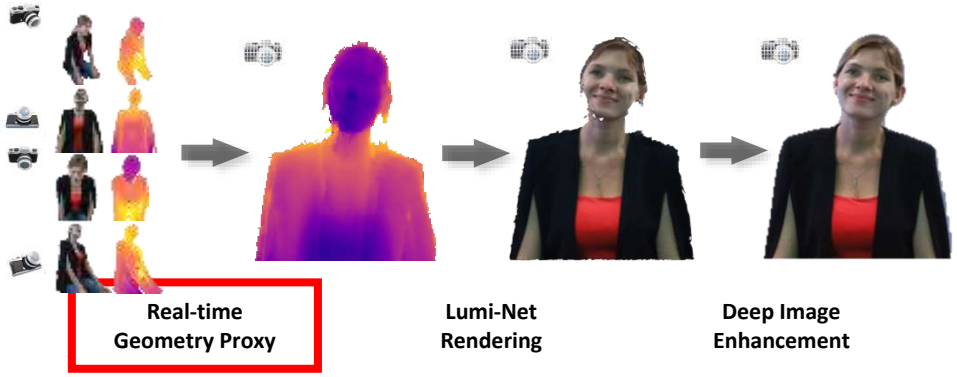




Multiview Depth

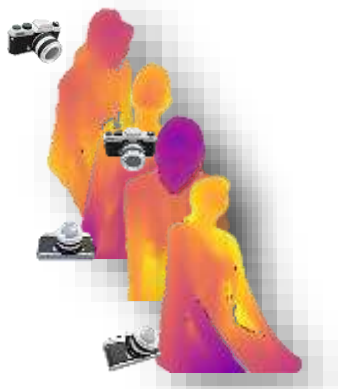
Multiview Color

Real-time Geometry Proxy



Initial Depth

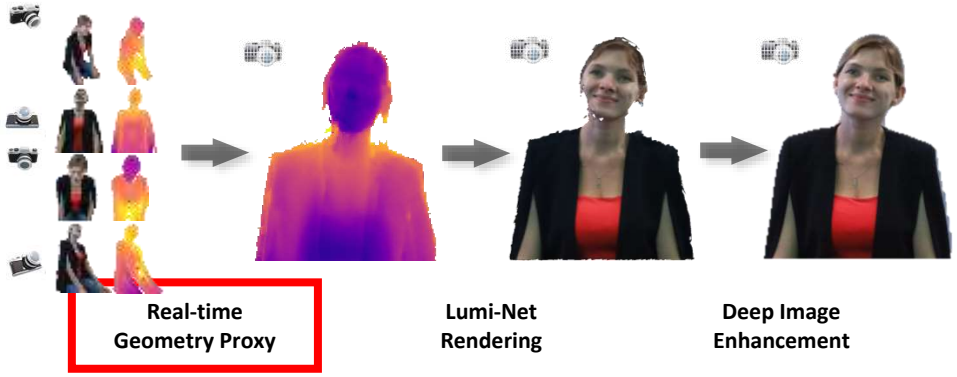
Projection &
Averaging

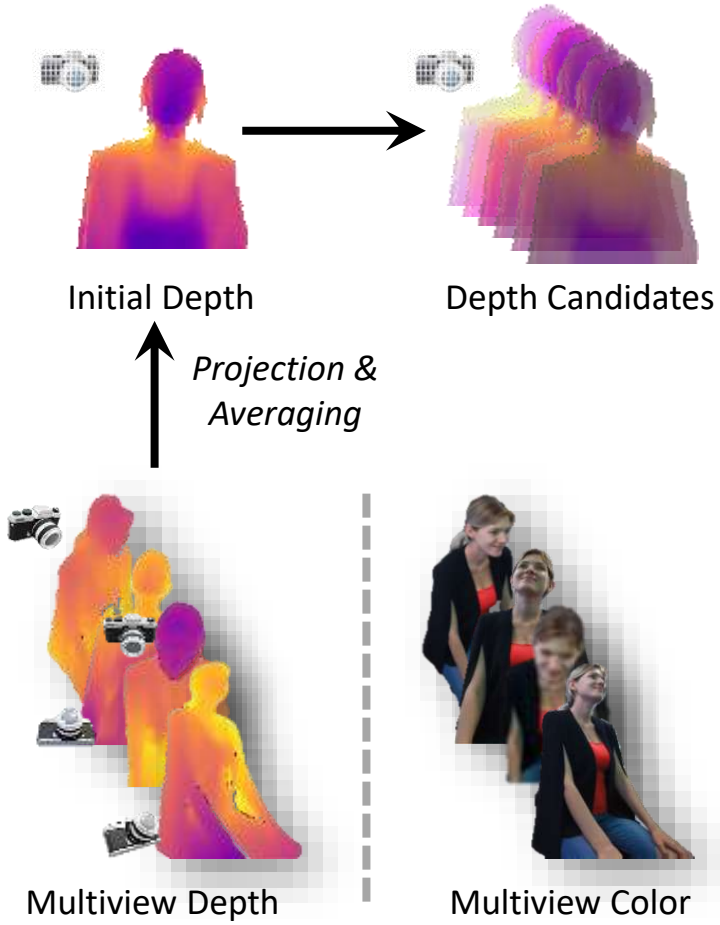


Multiview Depth

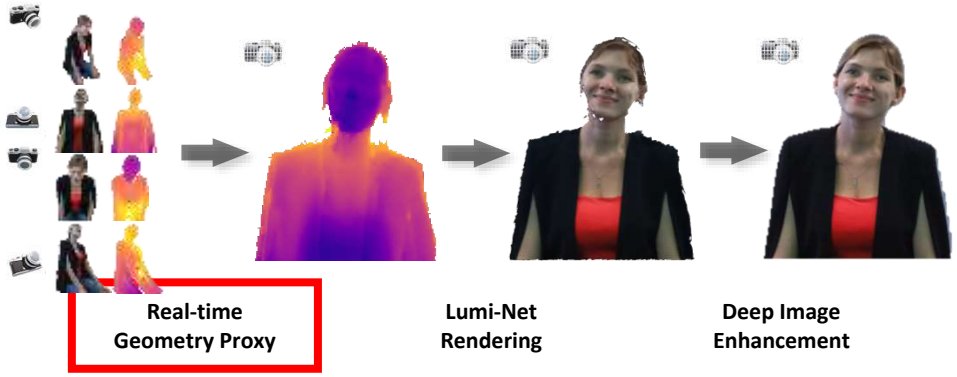
Multiview Color

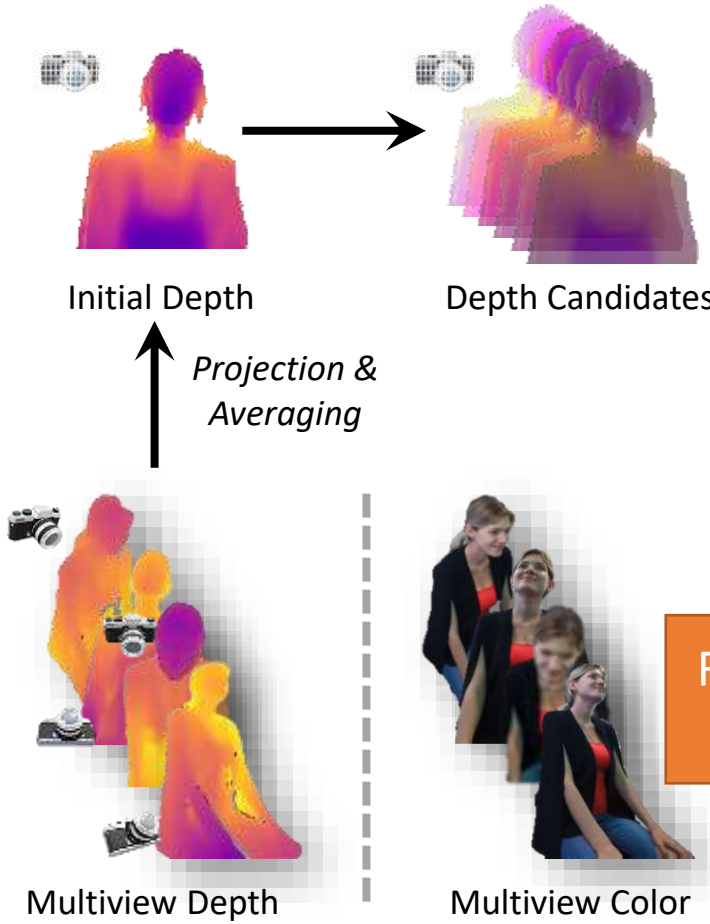
Real-time Geometry Proxy



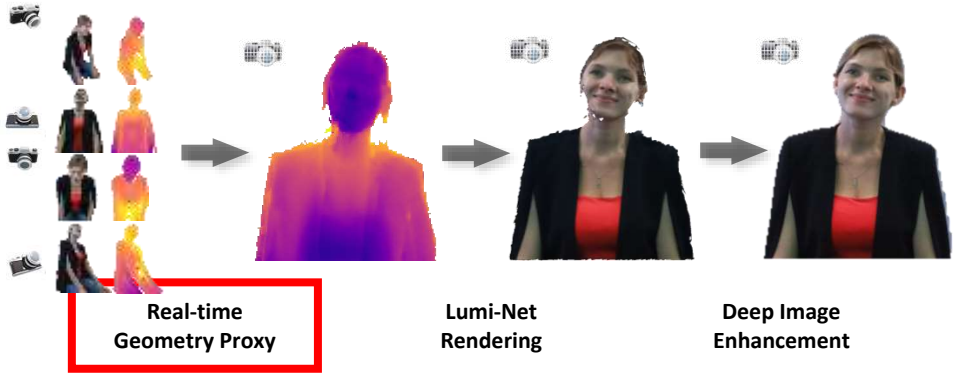


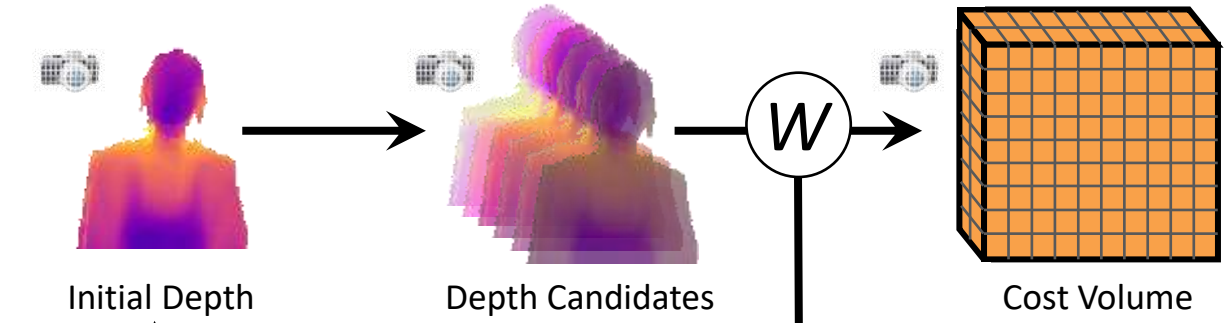
Real-time Geometry Proxy





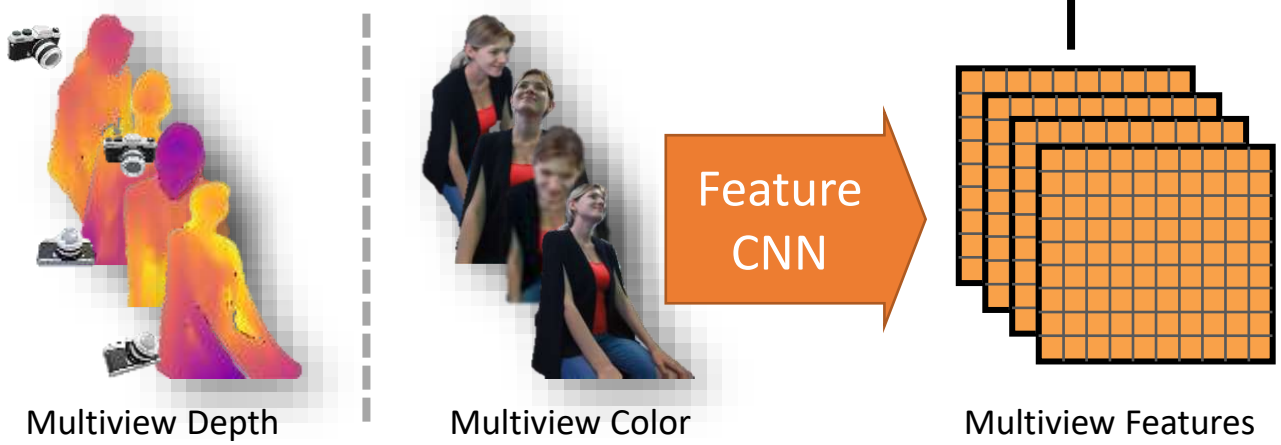
Real-time Geometry Proxy



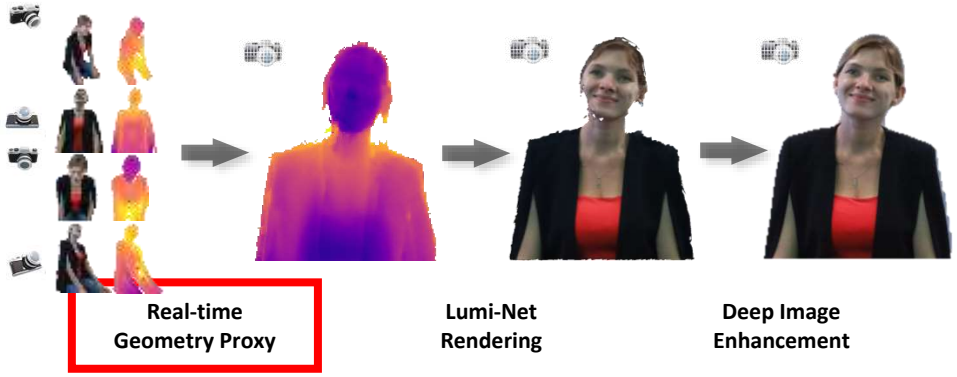


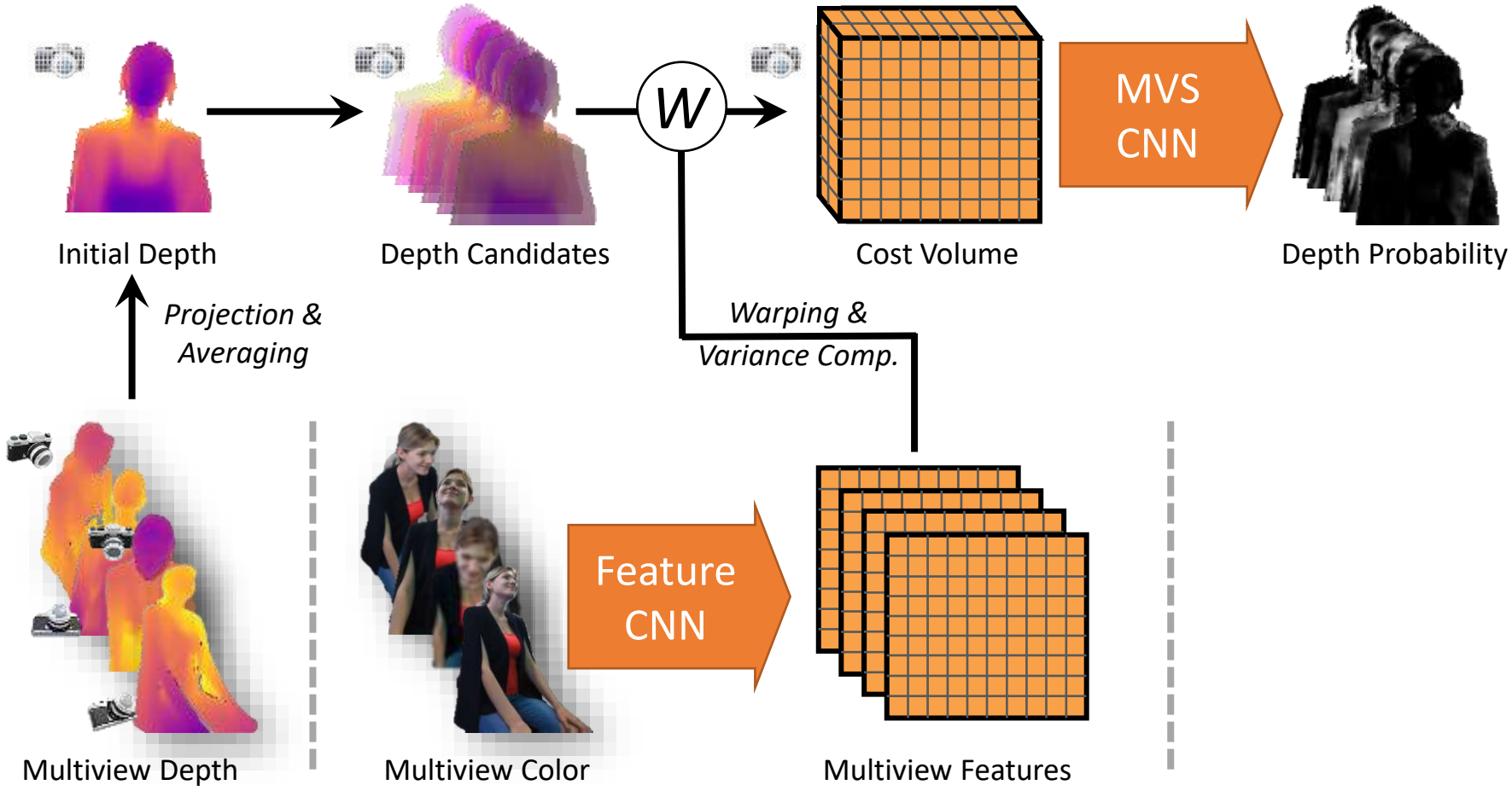
Projection & Averaging

Warping & Variance Comp.

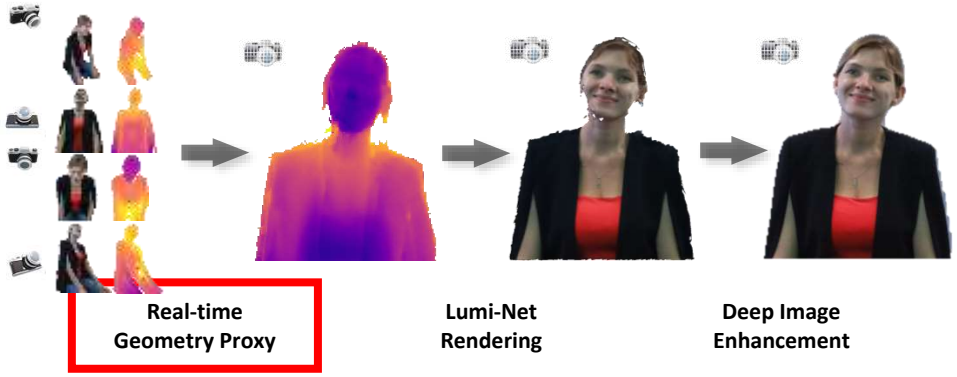


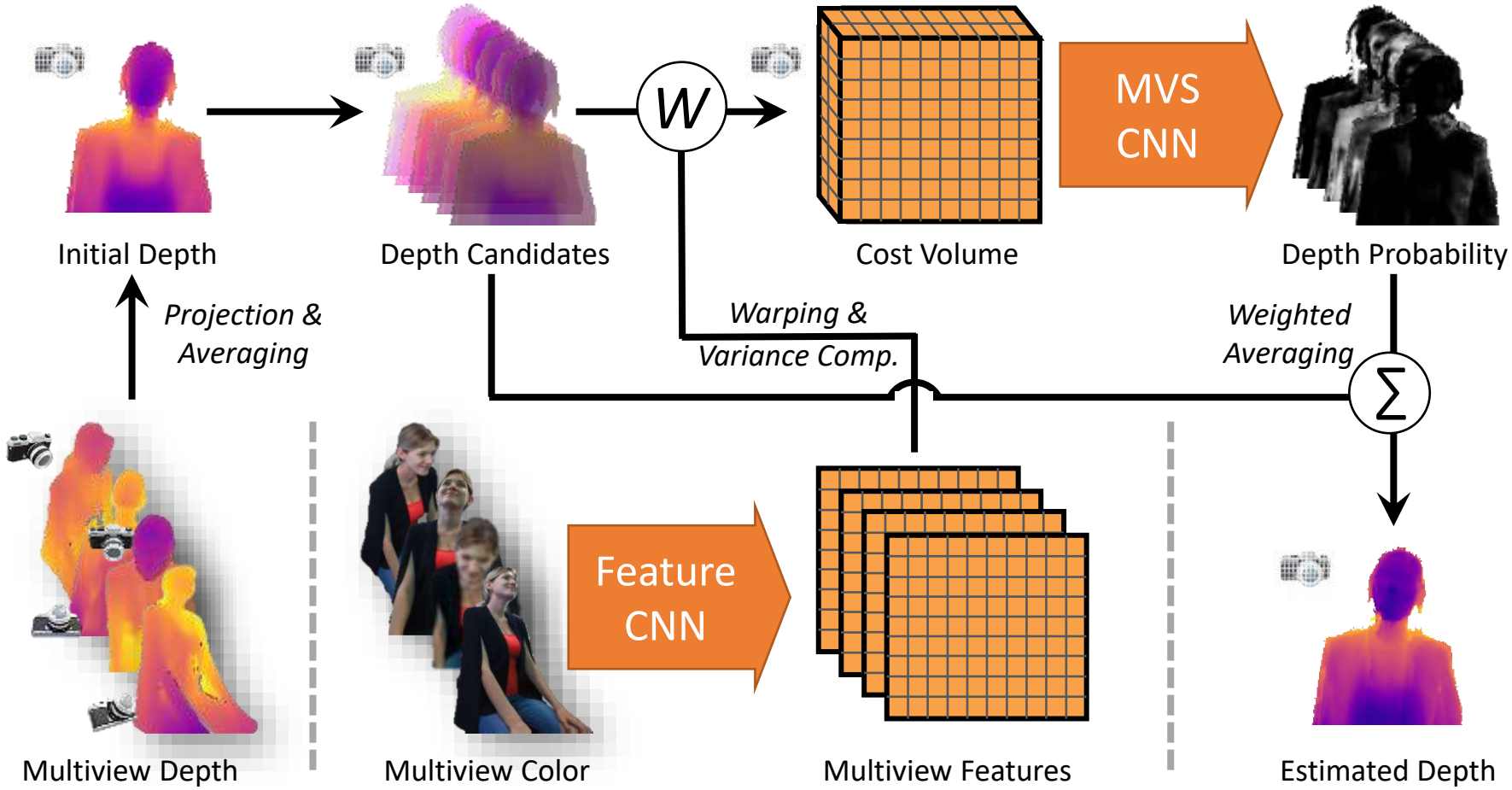
Real-time Geometry Proxy





Real-time Geometry Proxy





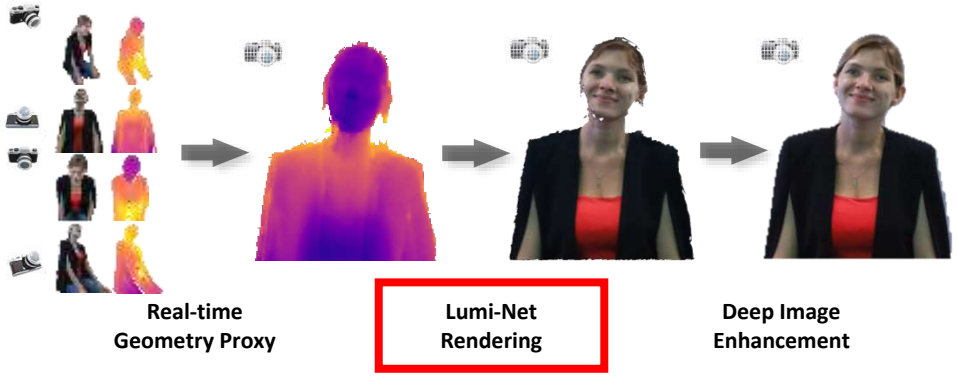
V-Cube View for 3D Free-View Rendering

Real-time Geometry Proxy

Lumi-Net Rendering

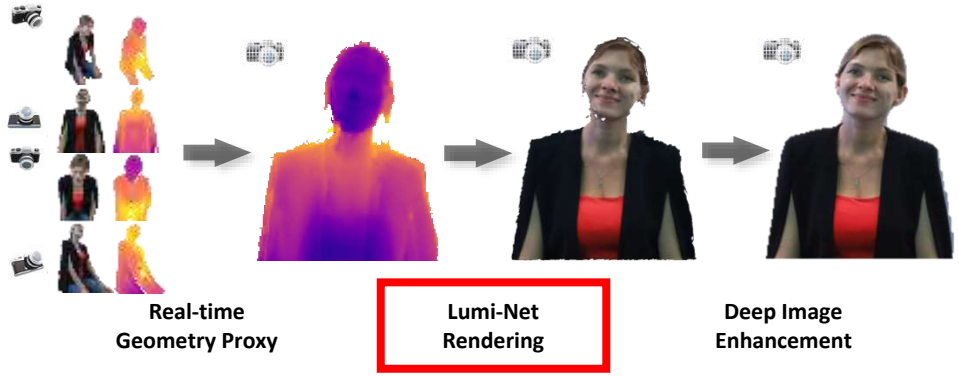
Deep Image Enhancement

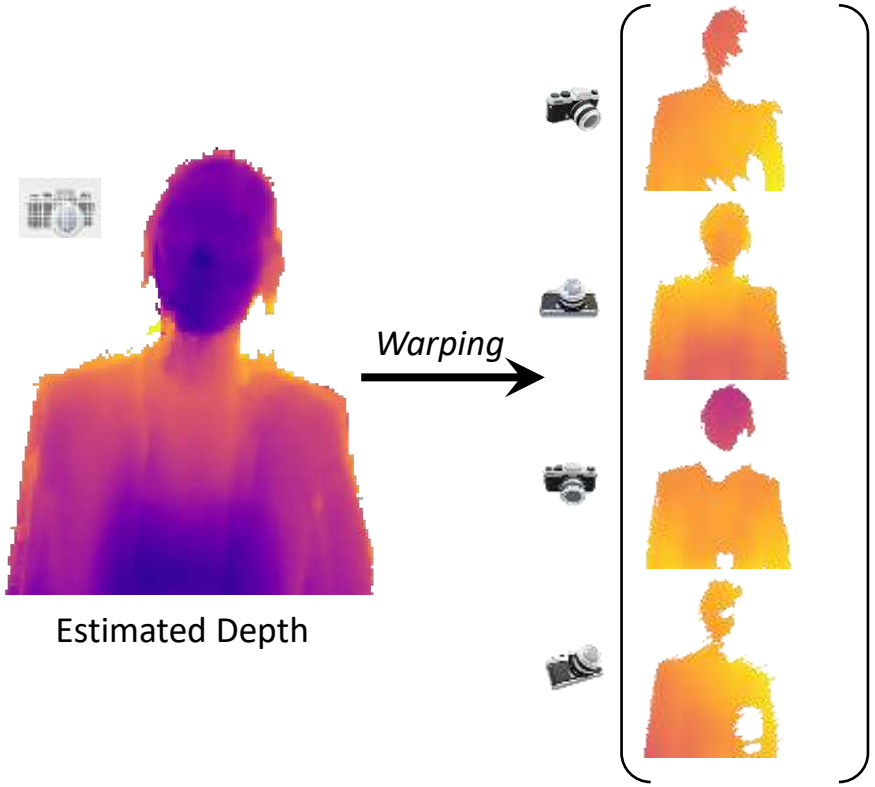
Lumi-Net rendering



Estimated Depth

Lumi-Net rendering

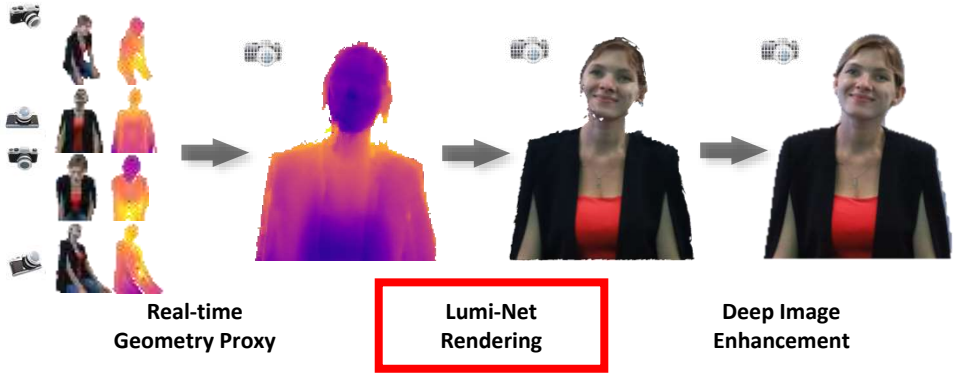


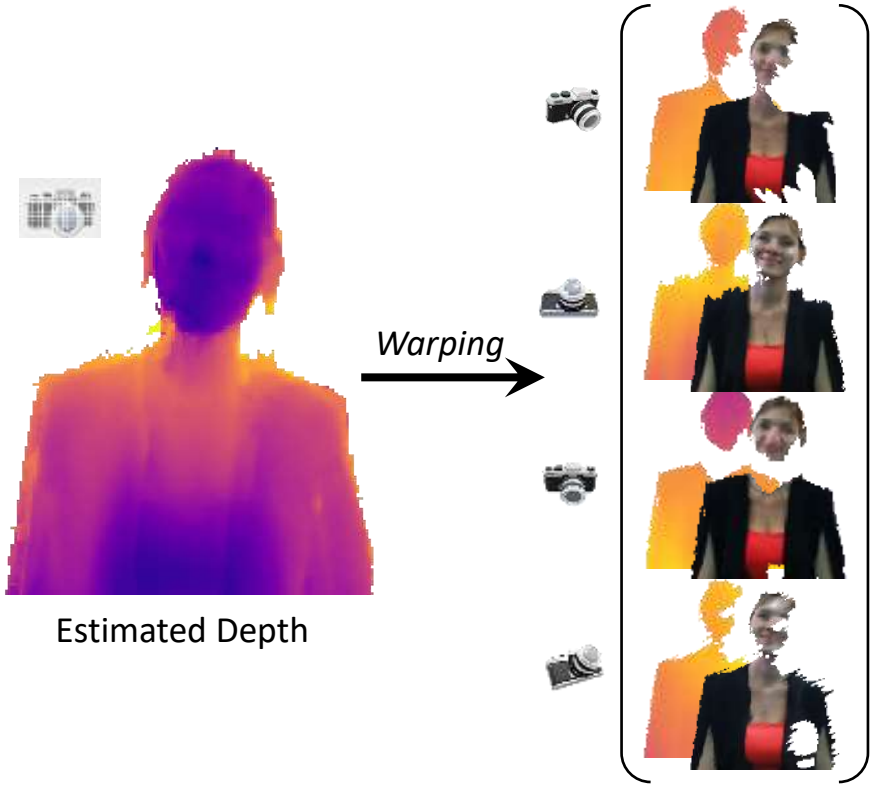


Estimated Depth

Warped Depth, Color, etc.

Lumi-Net rendering

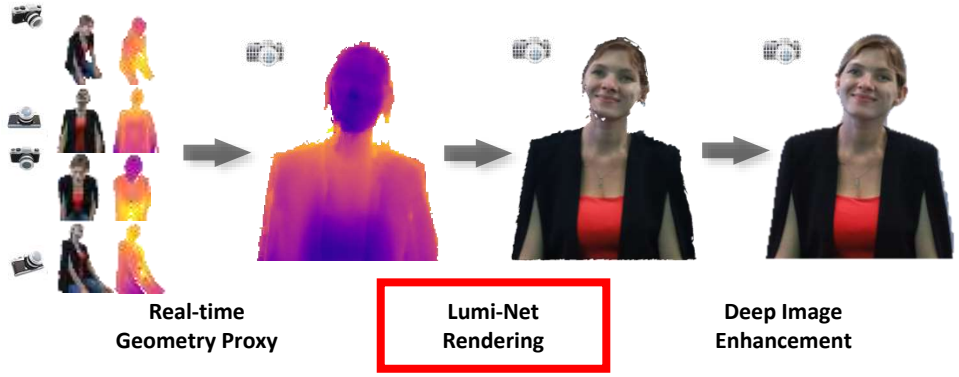


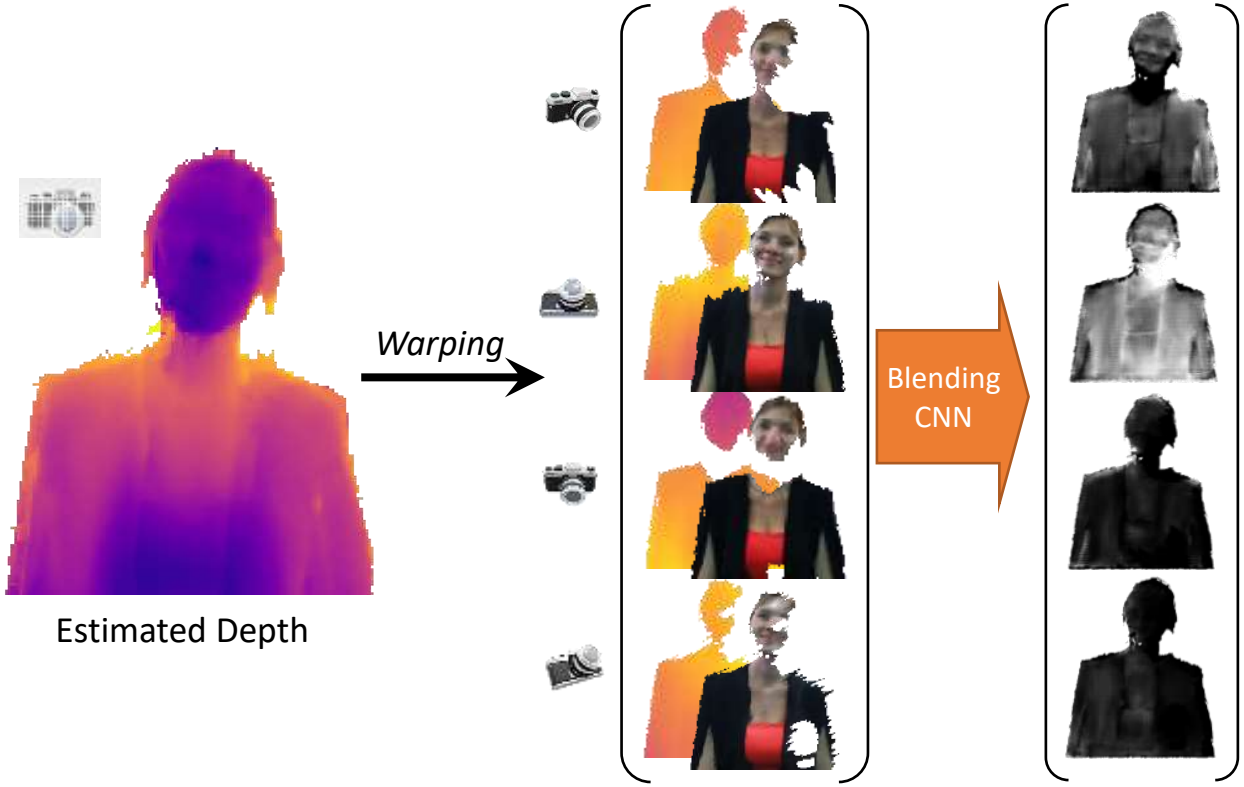


Estimated Depth

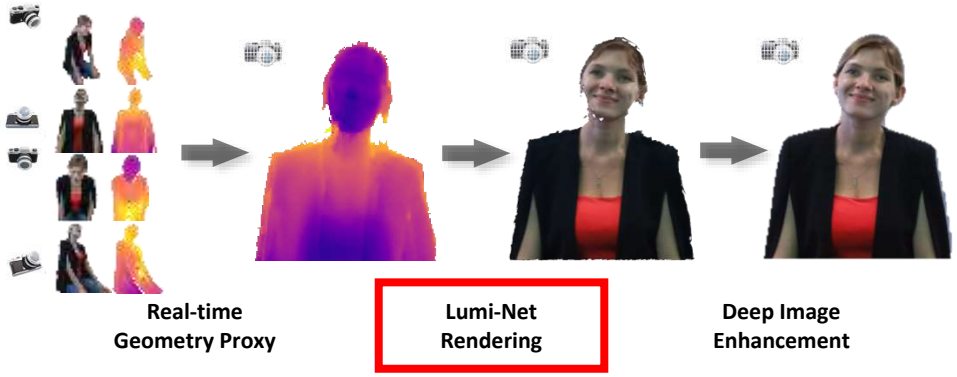
Warped Depth, Color, etc.

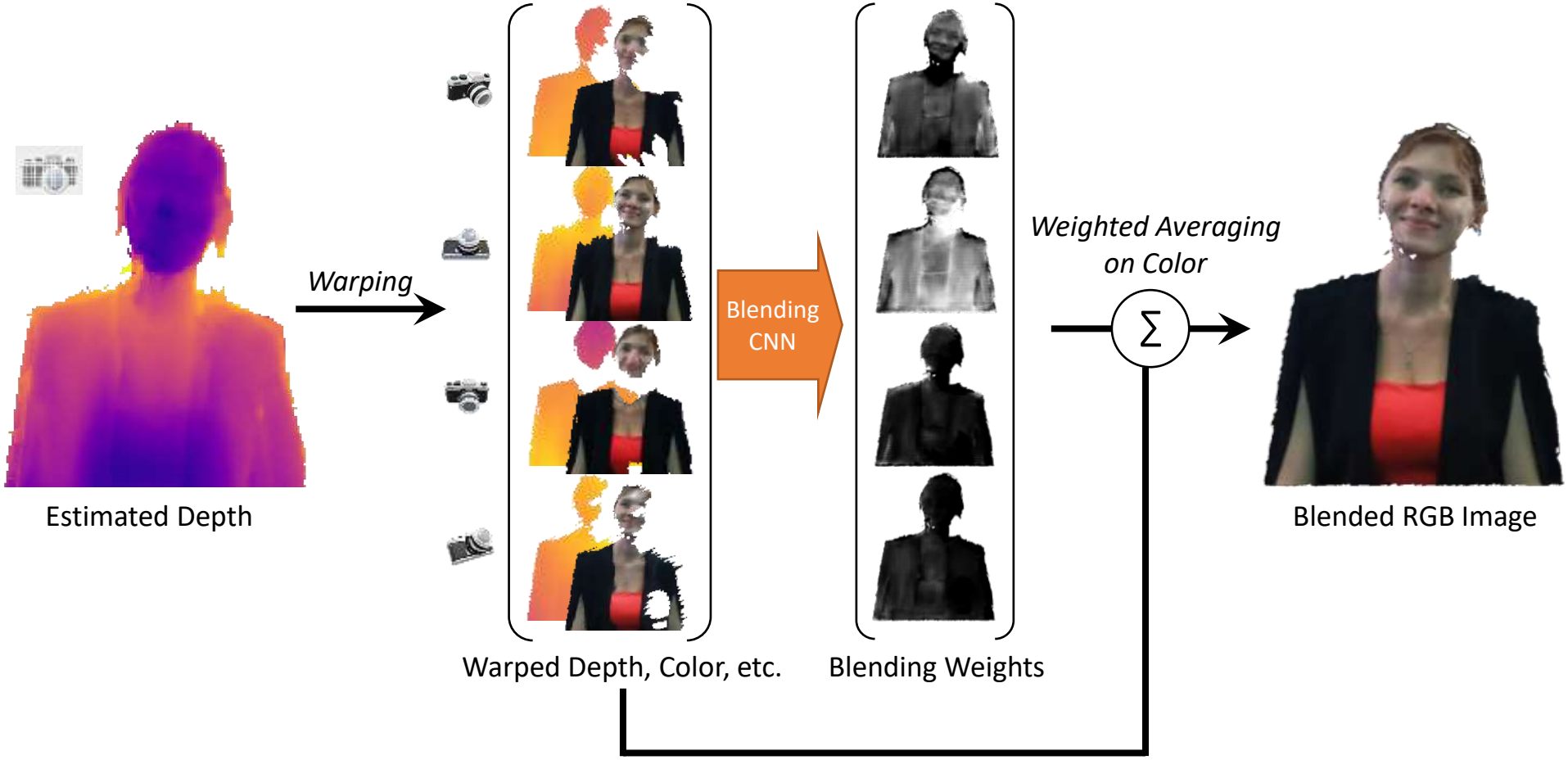
Lumi-Net rendering





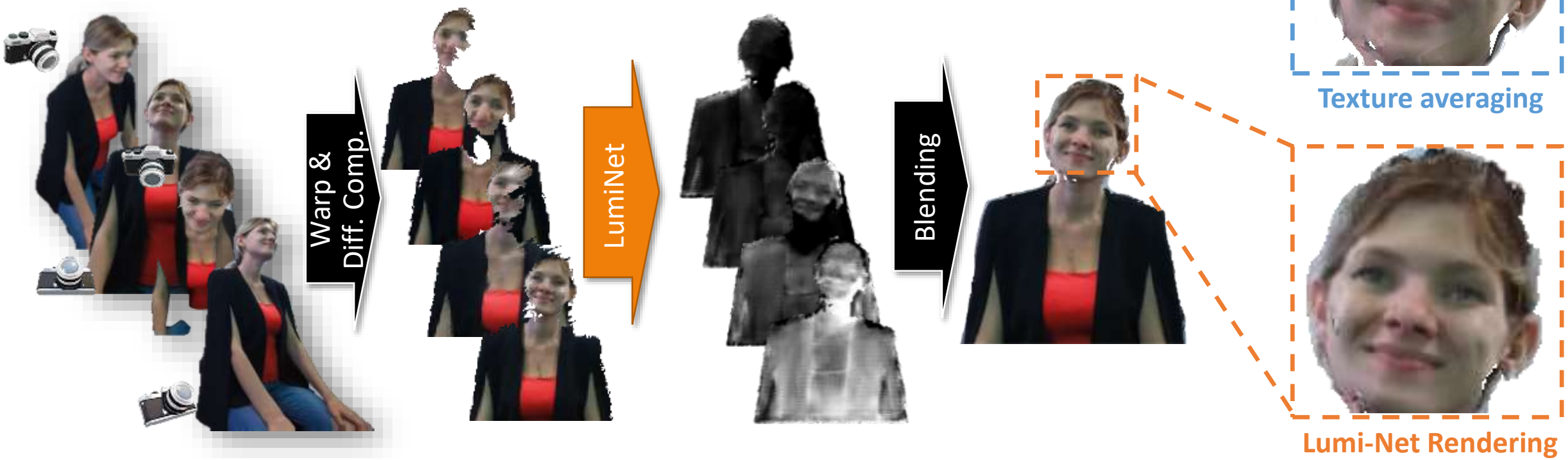
Lumi-Net rendering





Lumi-Net rendering

- Make the best use of sampled lightfield data
  - Minimal Angle Deviation Principle
  - Resolution Sensitivity Principle
- End-to-end neural network training



V-Cube View for 3D Free-View Rendering

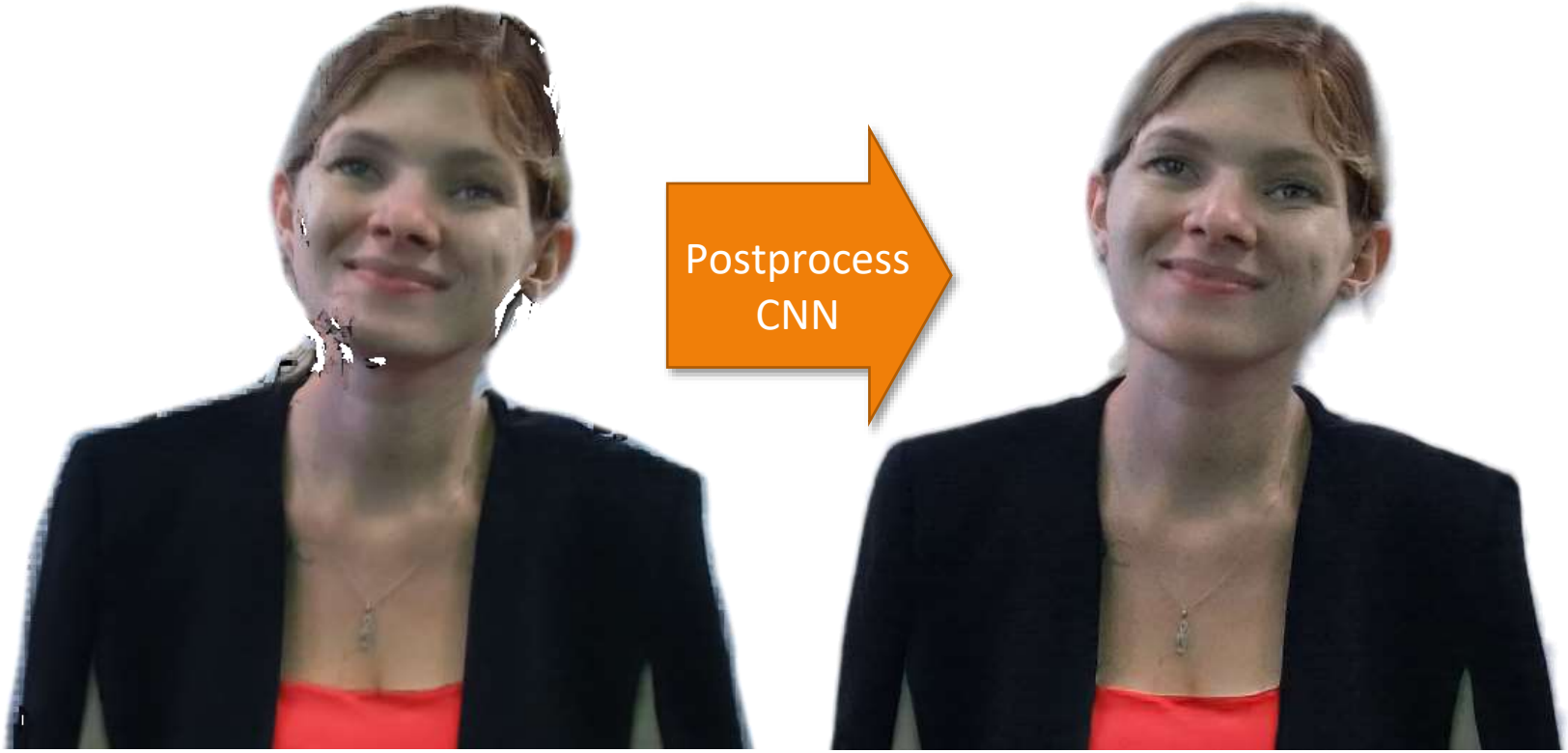
Real-time Geometry Proxy

Lumi-Net Rendering

Deep Image Enhancement

Deep Image Enhancement

- Postprocessing CNN
  - Alpha prediction, noise removal, etc.
  - Perception-based optimization; adversarial learning for sharpening



Lumi-Net Rendering

Enhanced (Final)

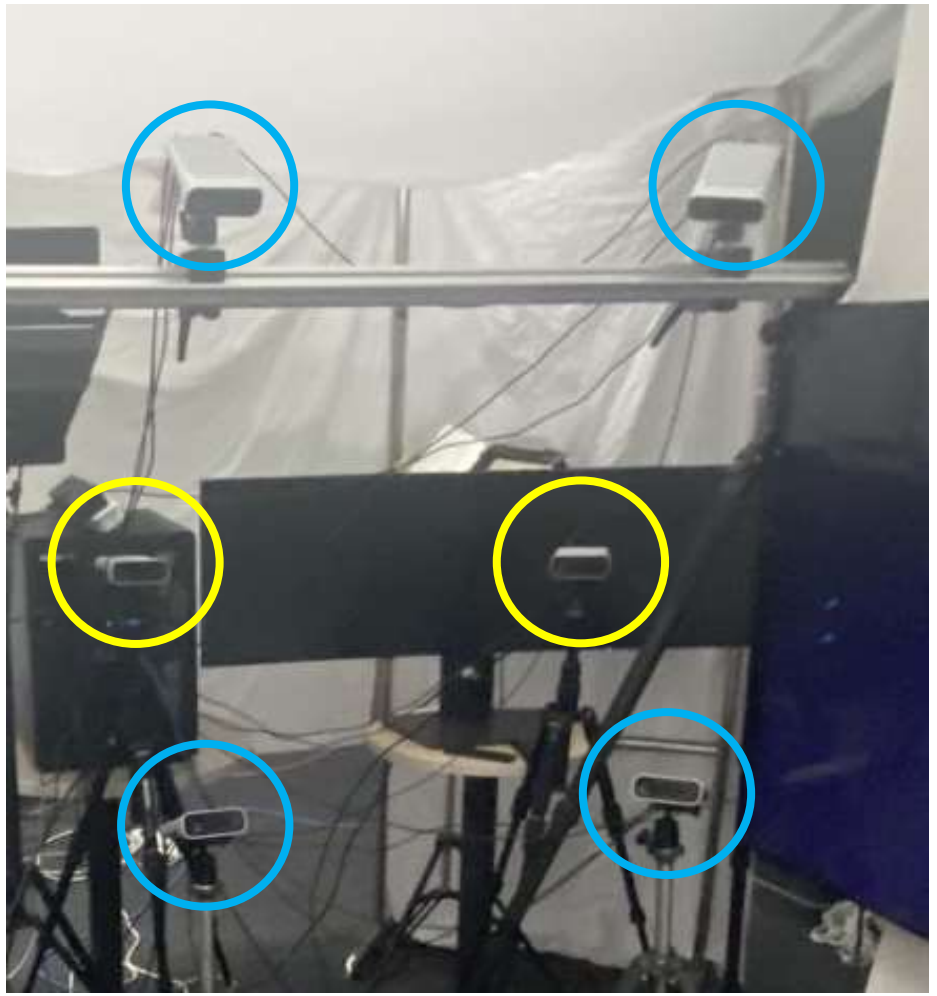
Deep Image Enhancement

- Postprocessing CNN
  - Alpha prediction, noise removal, etc.
  - Perception-based optimization; adversarial learning for sharpening

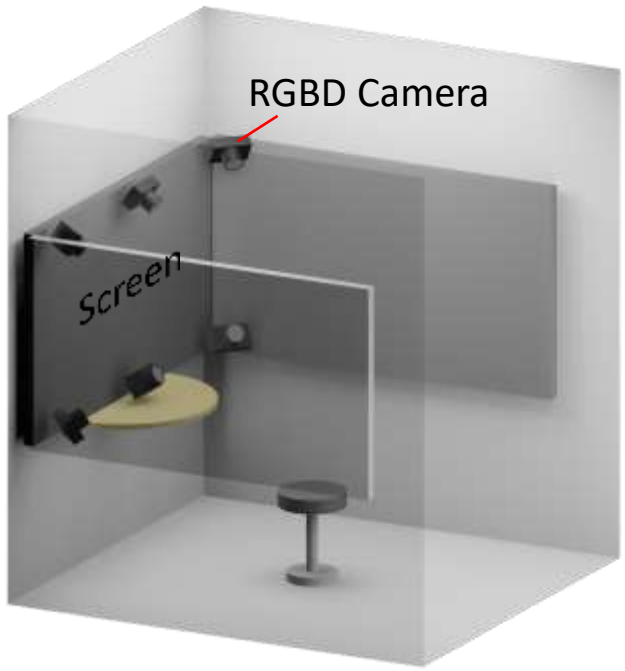
Training Data

Training Data

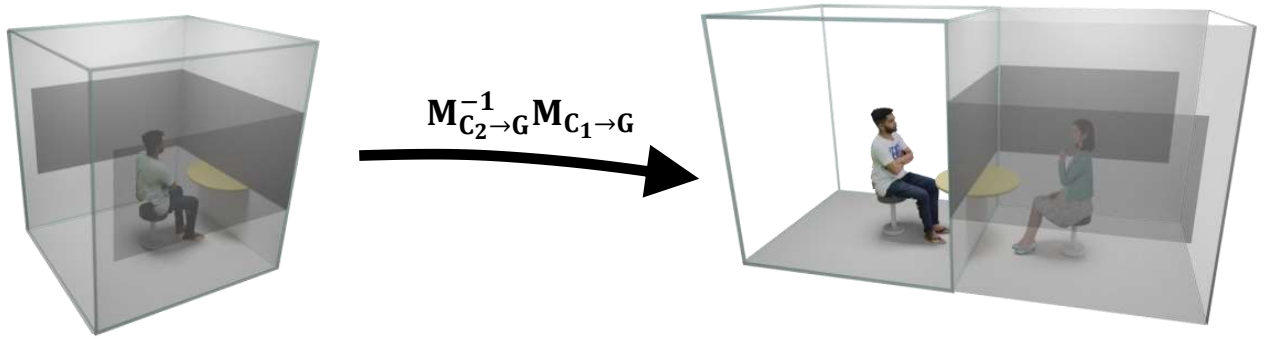
Training Data



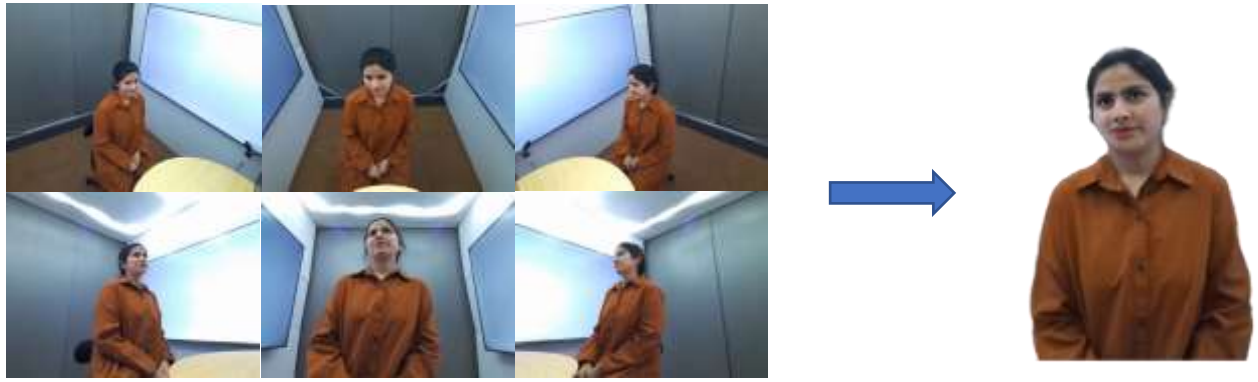
Key Components



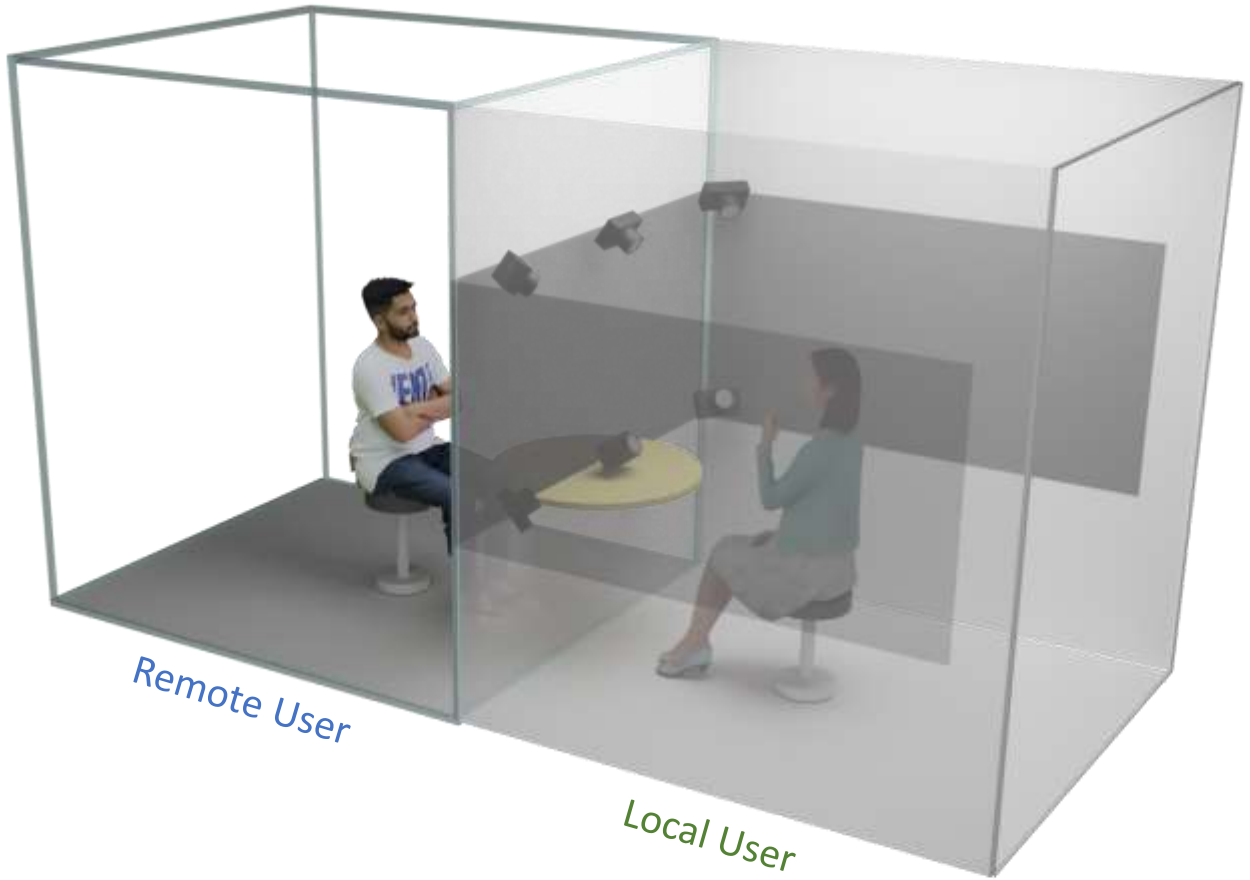
V-Cube Assembly



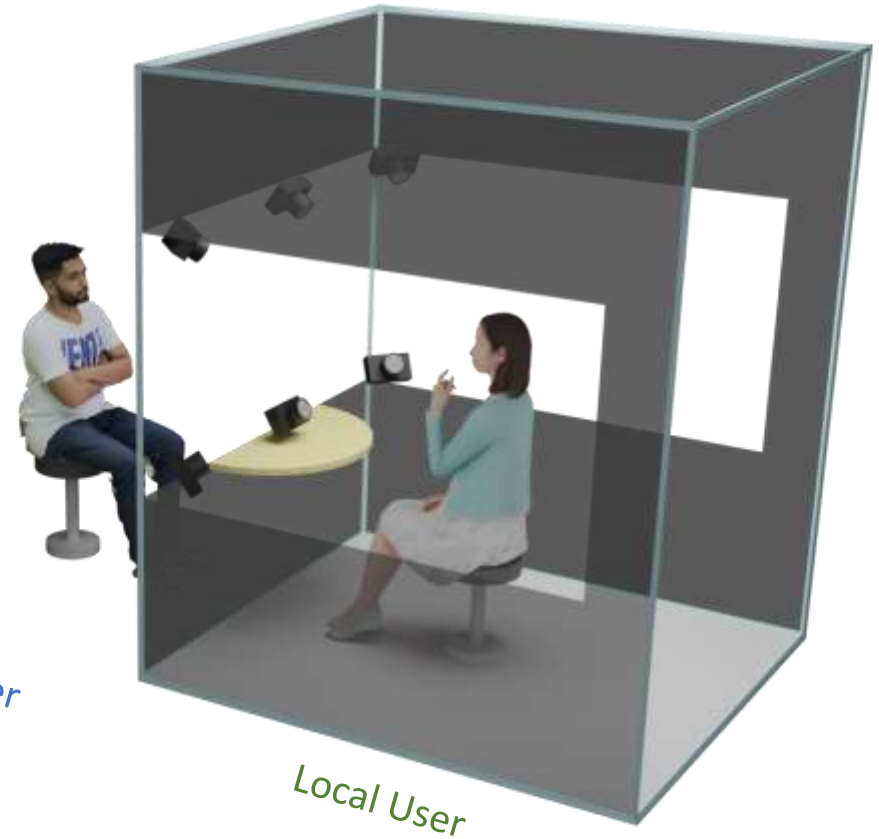
V-Cube View



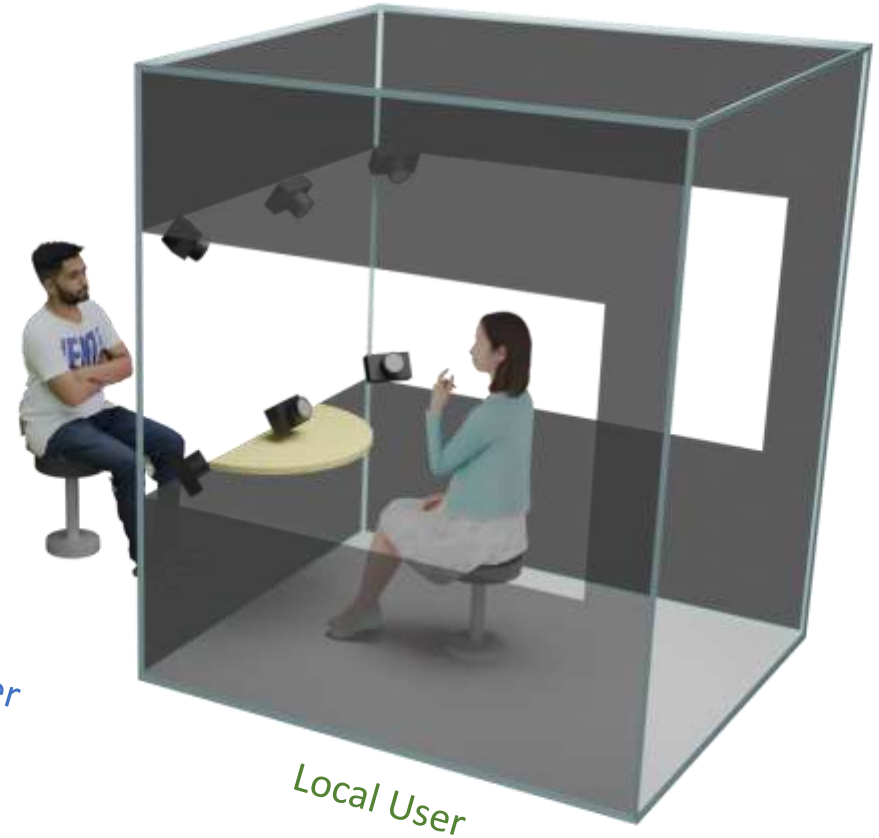
System Workflow



System Workflow



System Workflow



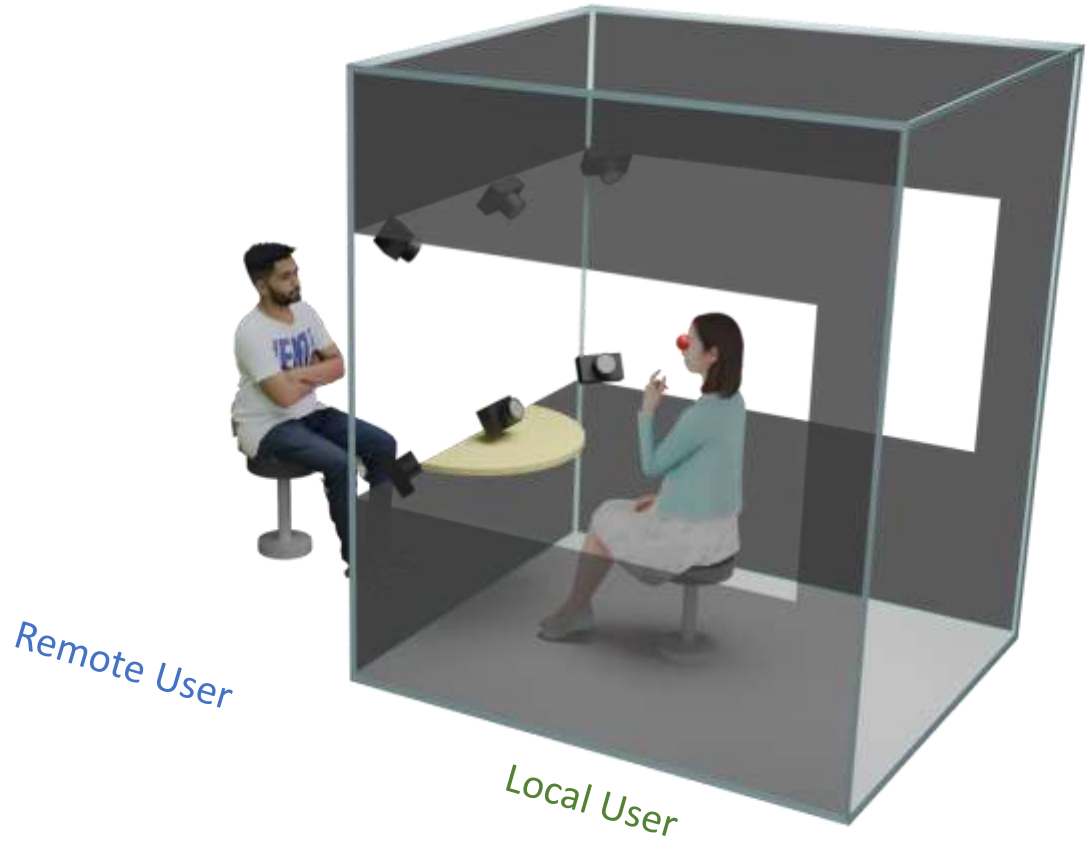
Remote User (Sender)

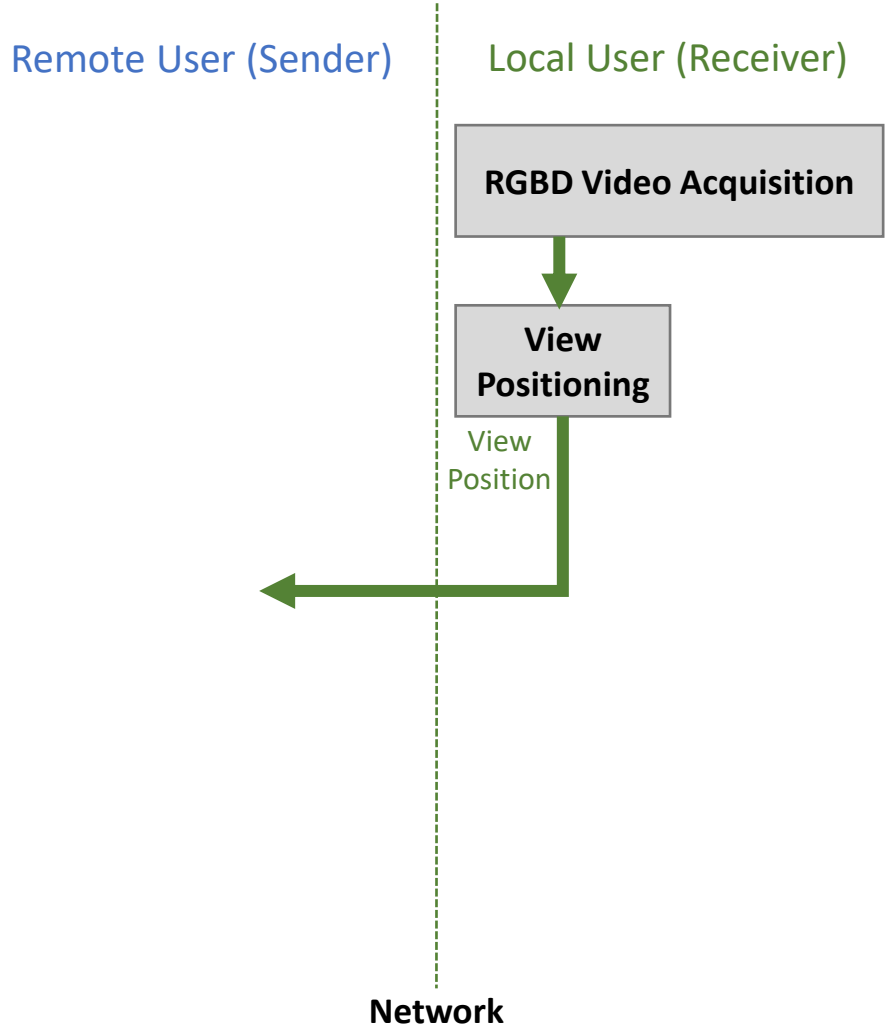
Local User (Receiver)

RGBD Video Acquisition

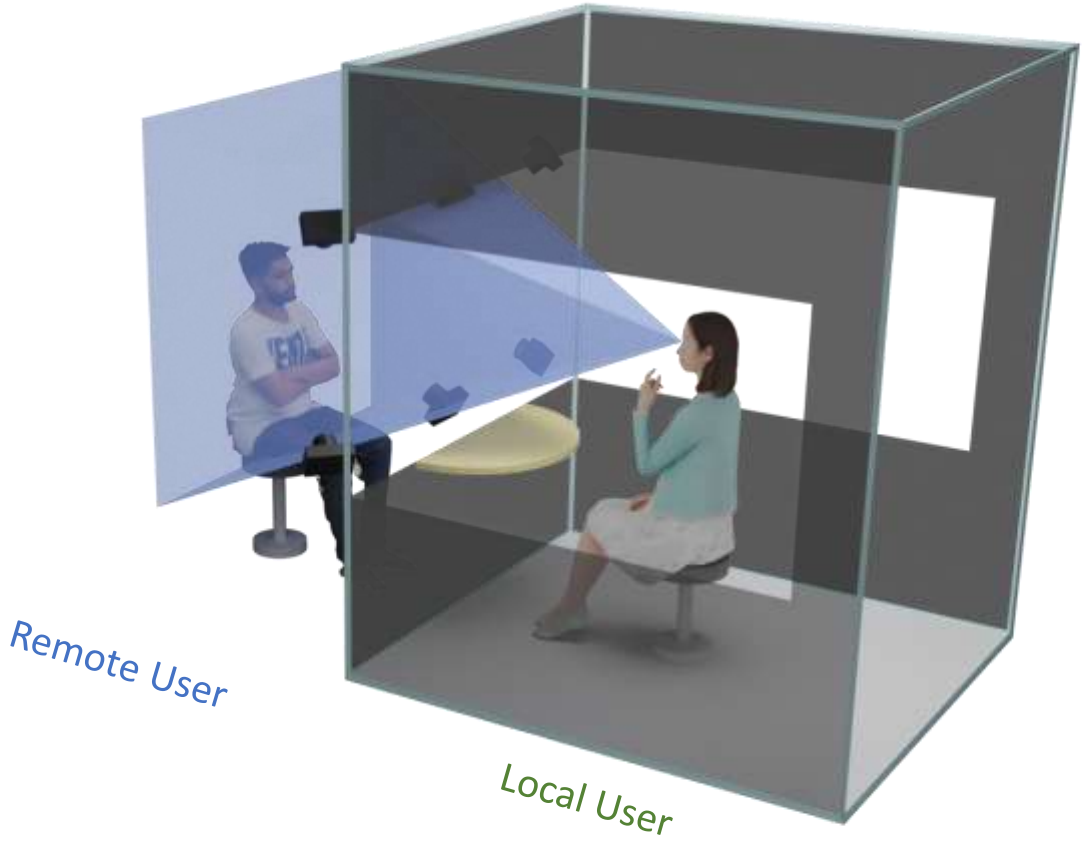
Network

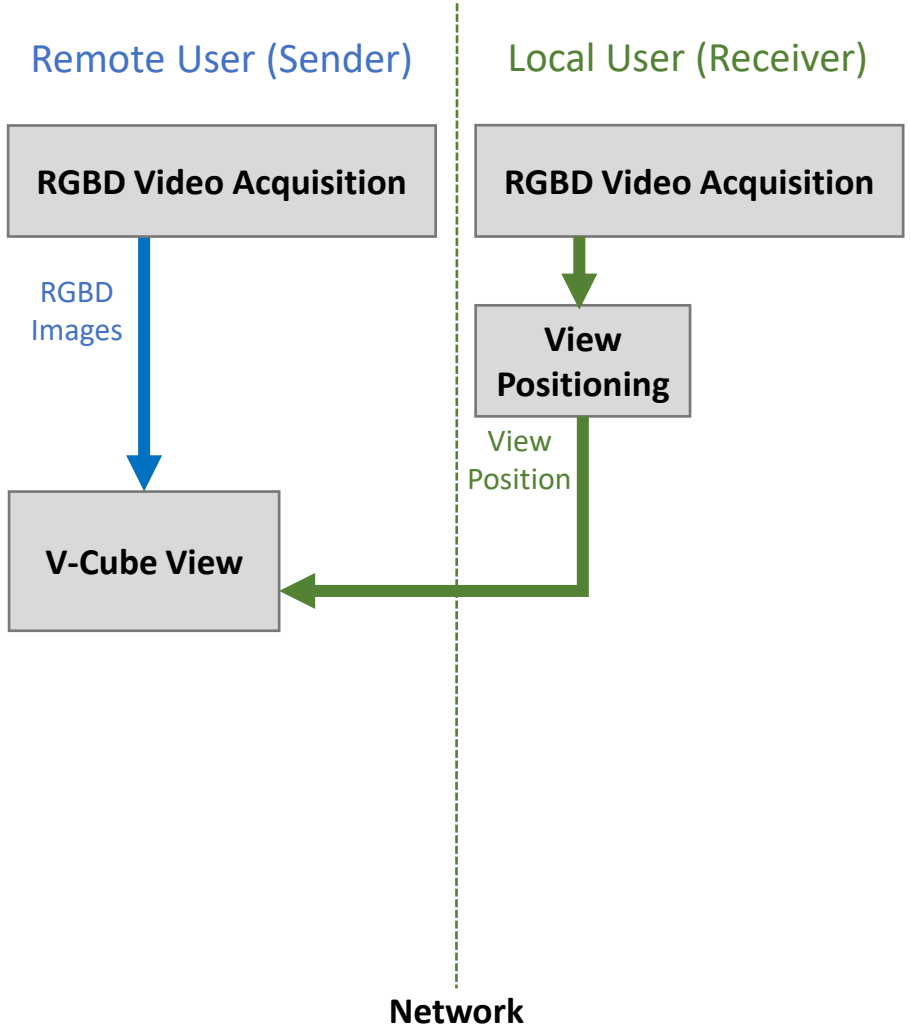
System Workflow



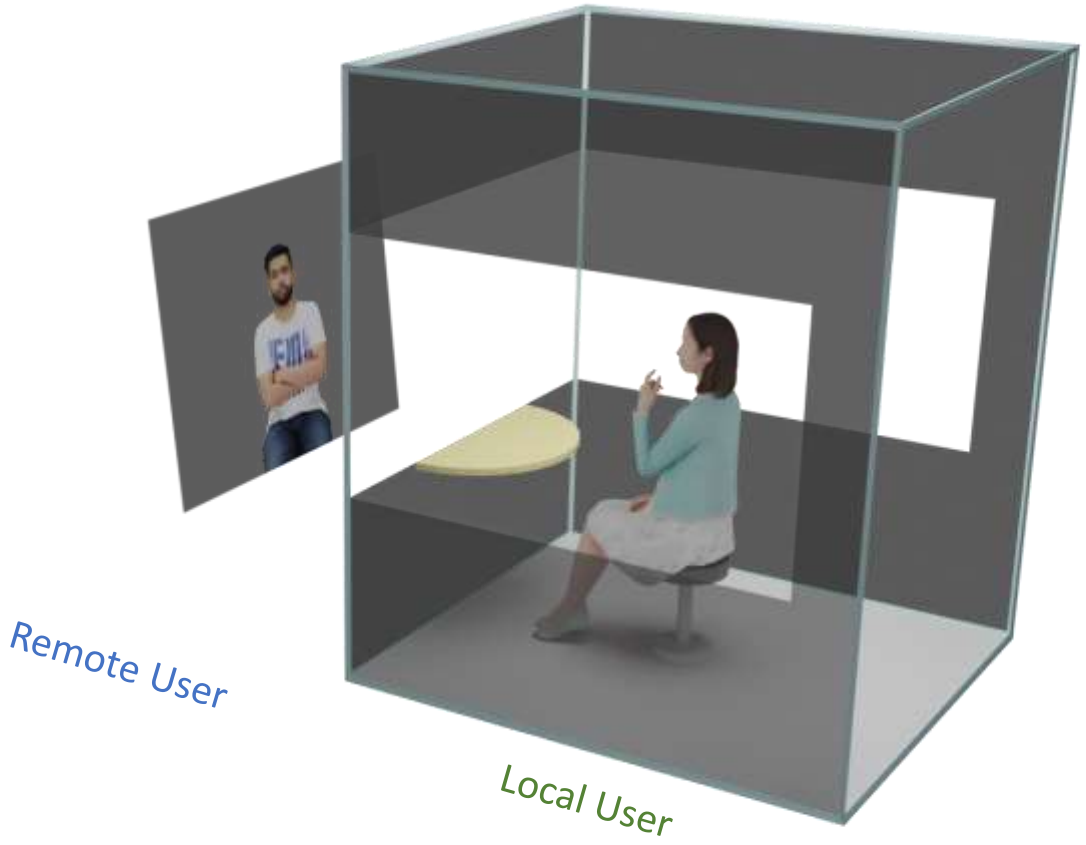


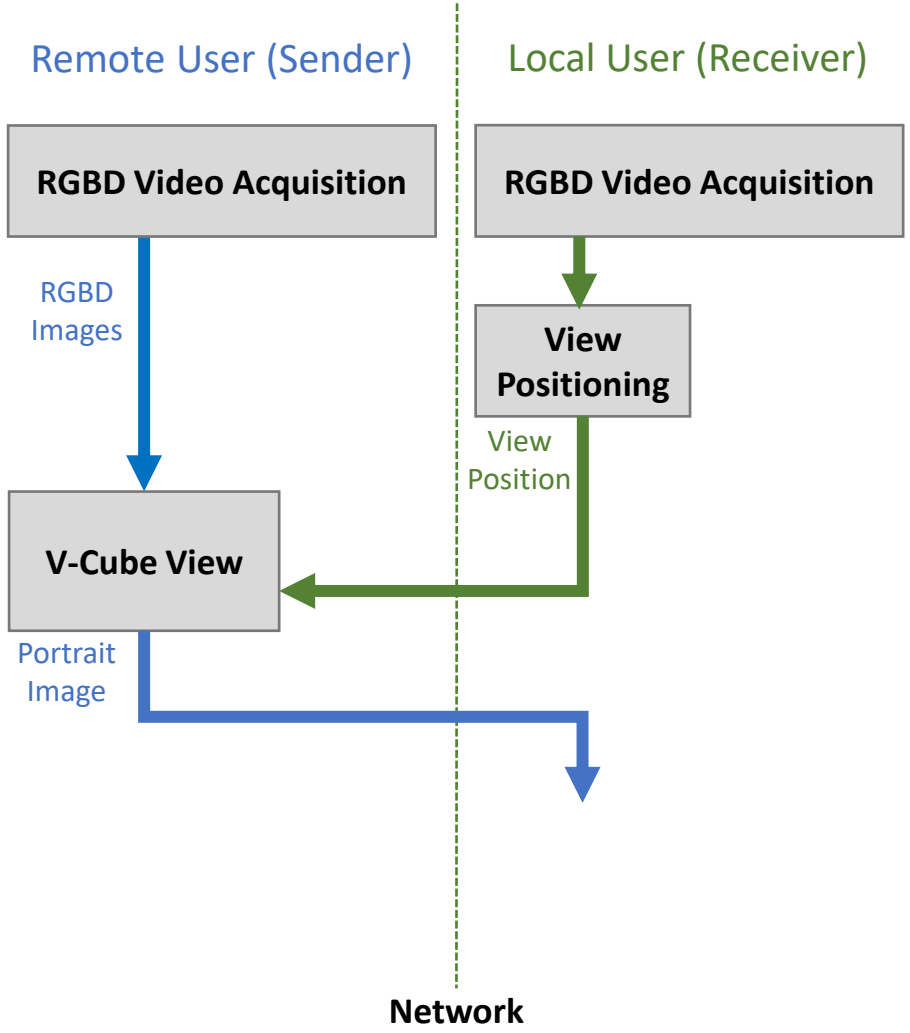
System Workflow



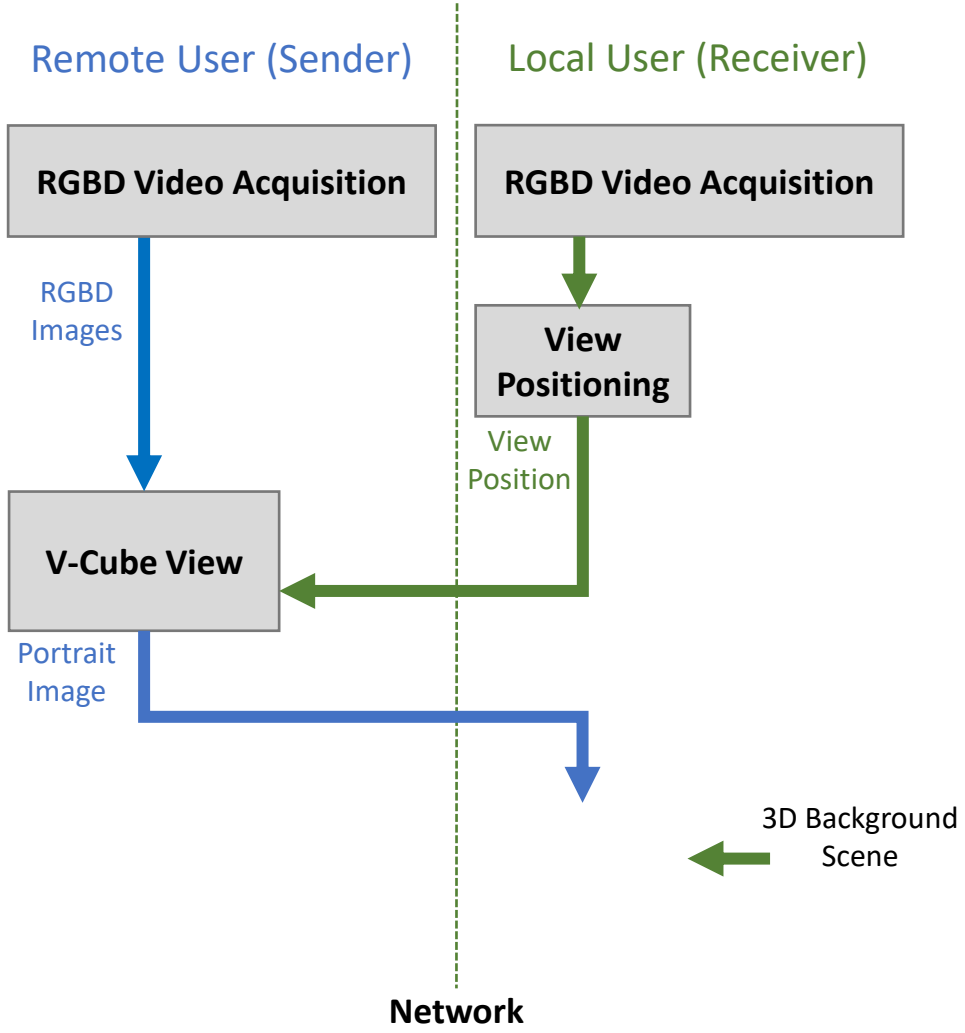


System Workflow

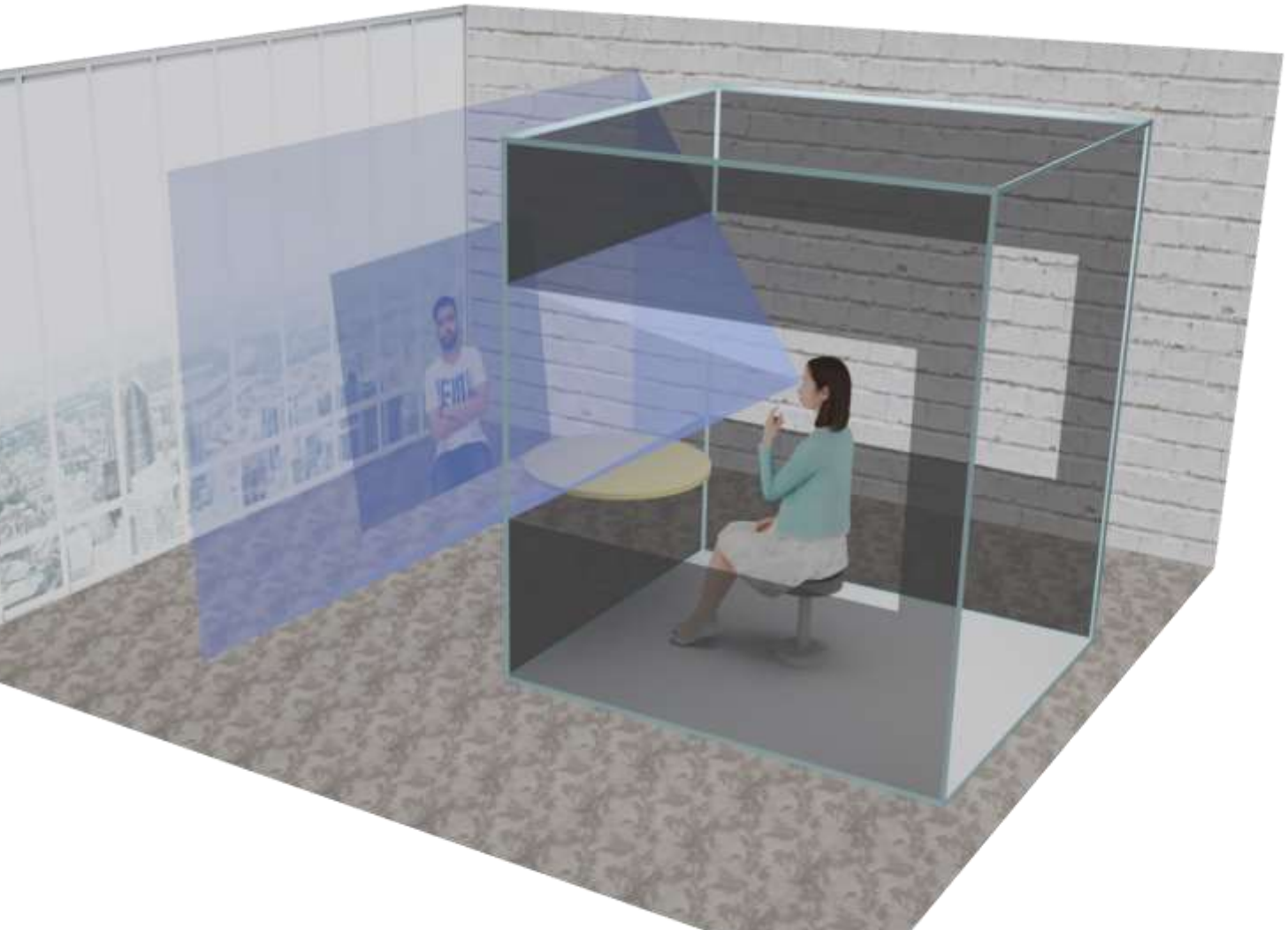


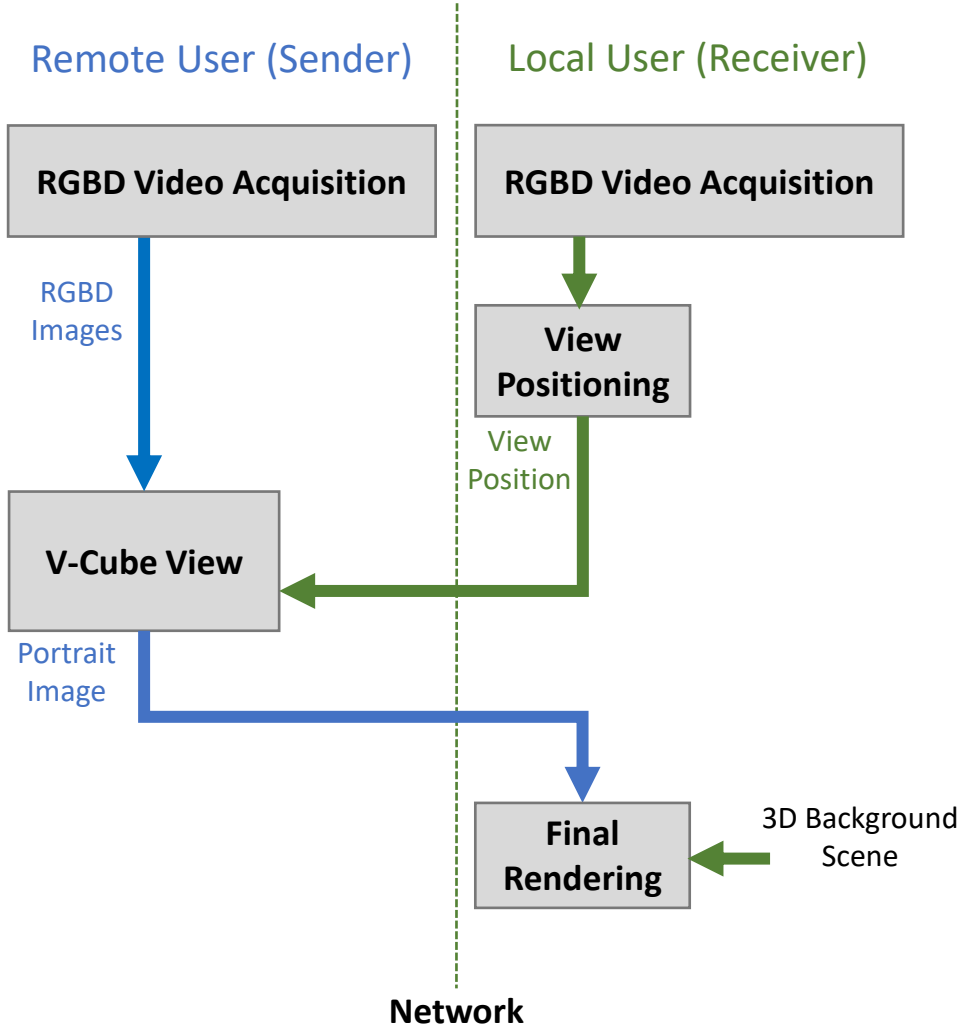


System Workflow

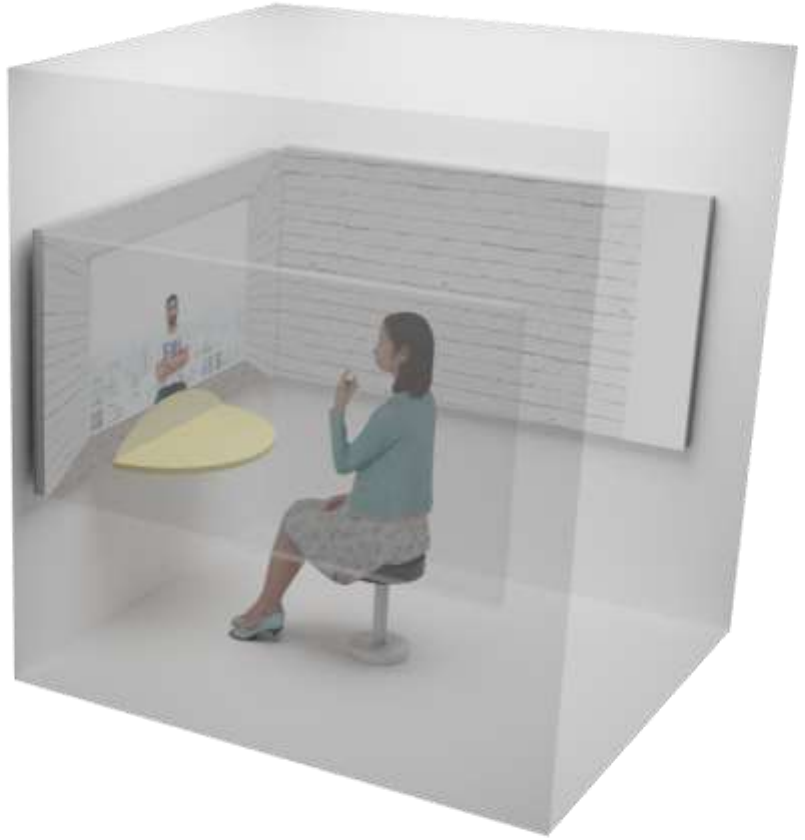


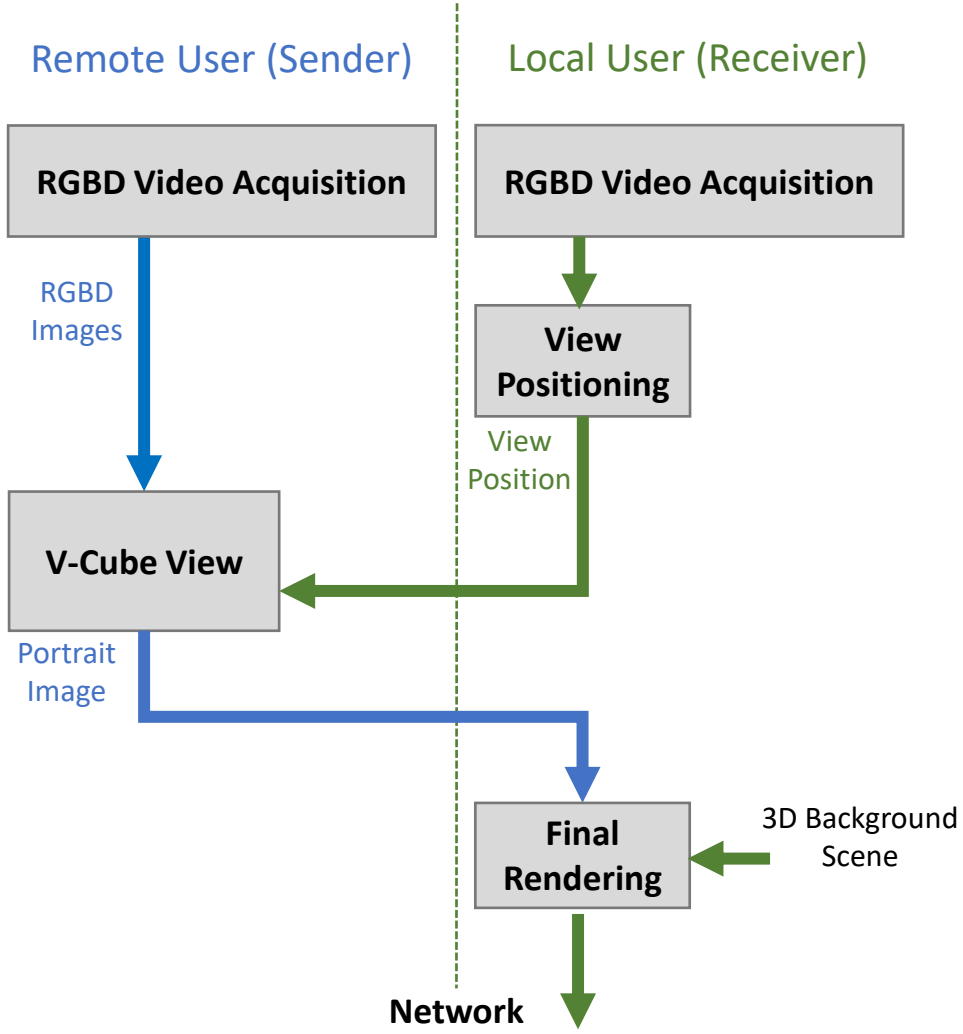
System Workflow



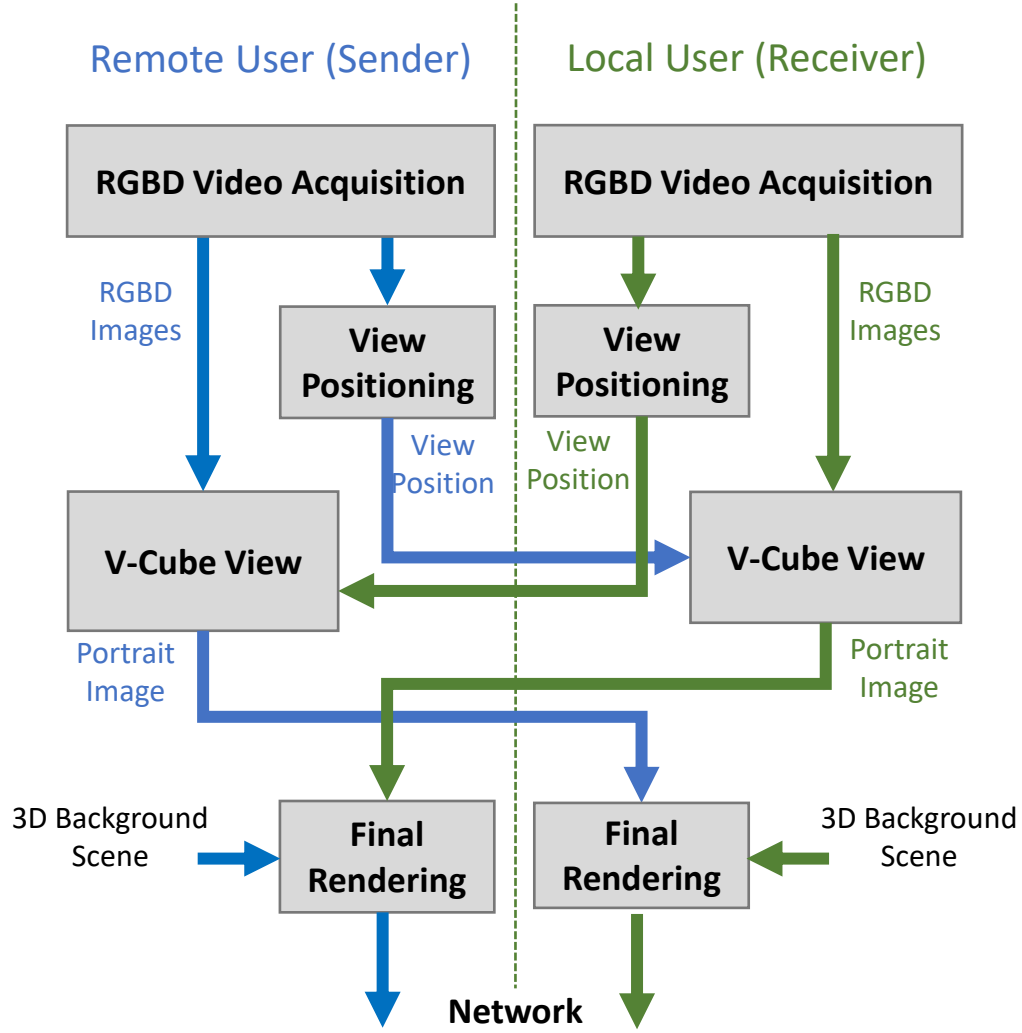


System Workflow

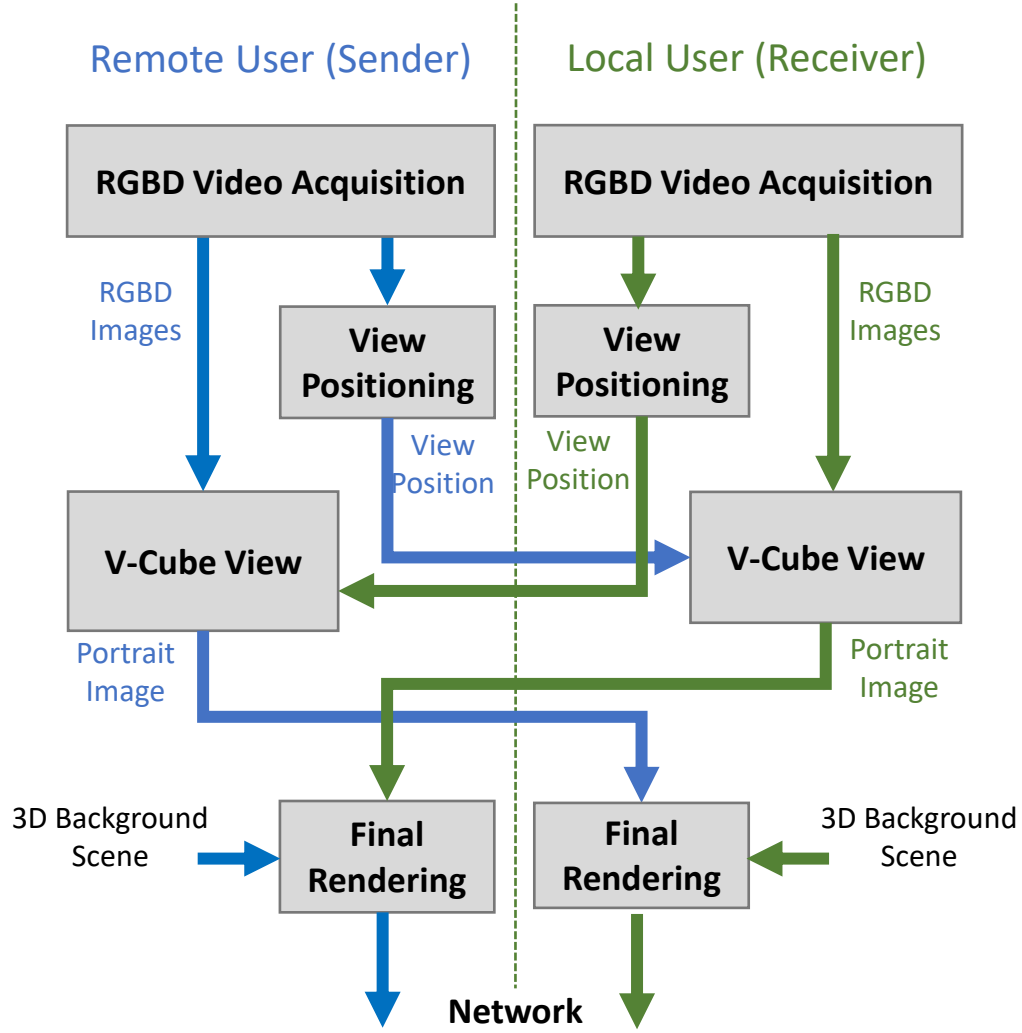




System Workflow



System Workflow

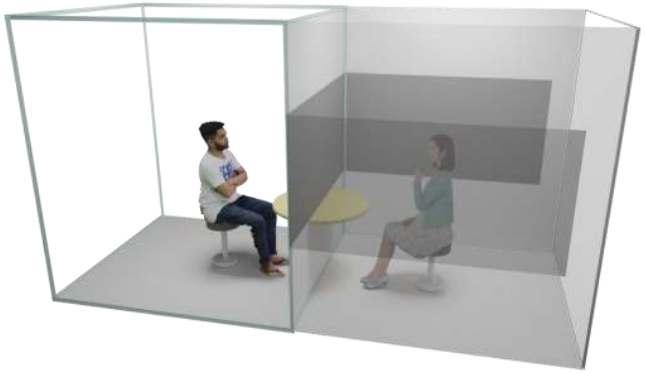


Implementation Details

- Hardware Specification
  - Core i9-10980XE CPU
  - 64GB RAM
  - Nvidia GeForce RTX 3090 ×2 for V-Cube View
  - Nvidia GeForce RTX 2080 for display

- System Performance
  - End-to-end delay: ~300ms
  - Frame rate: 23fps one GPU

Demo I: one-to-one meeting

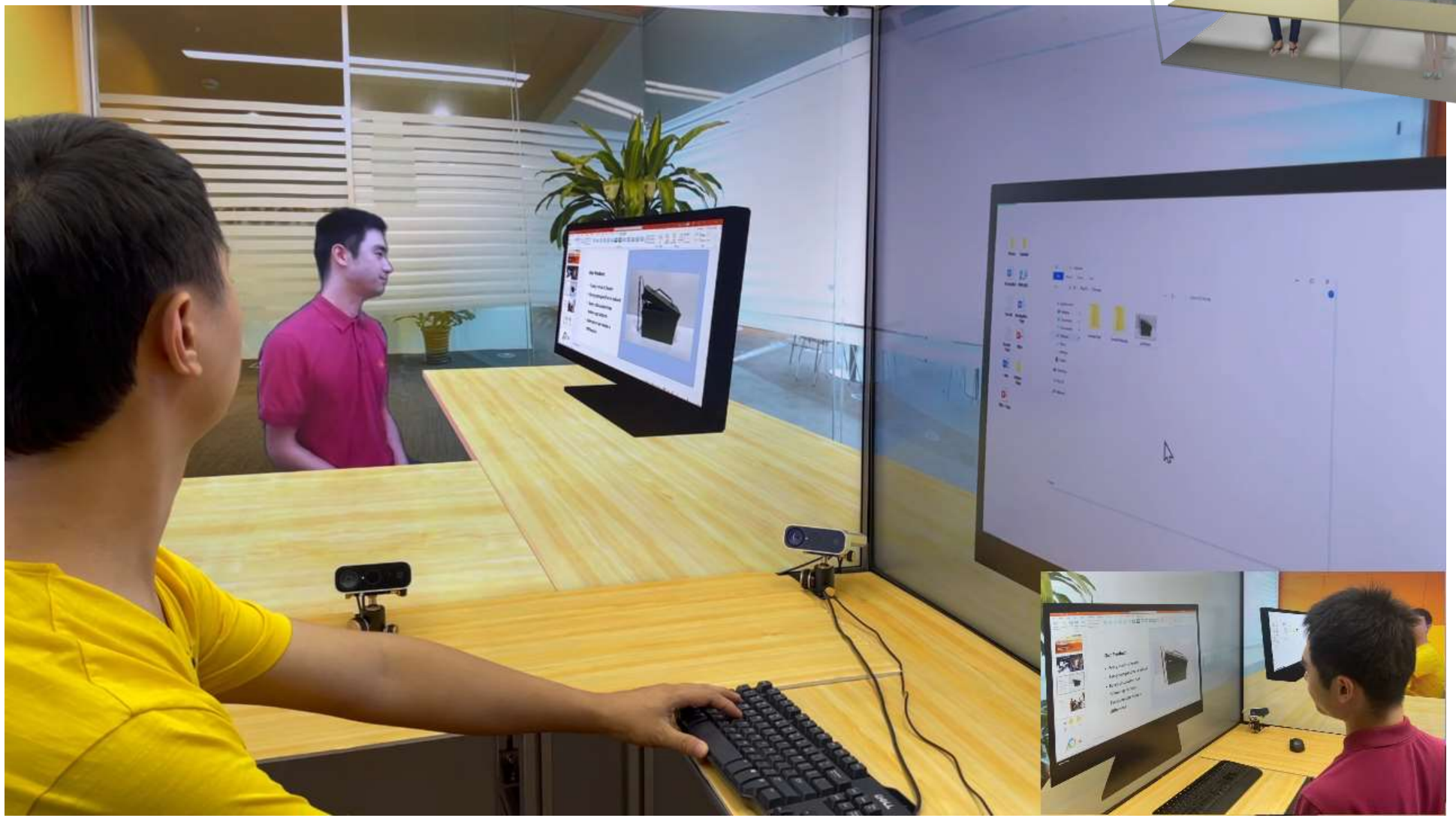


Existing Video Conferencing

Our VirtualCube

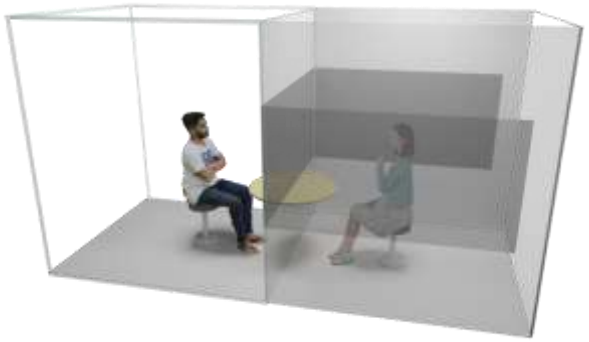
Demo II: group meeting

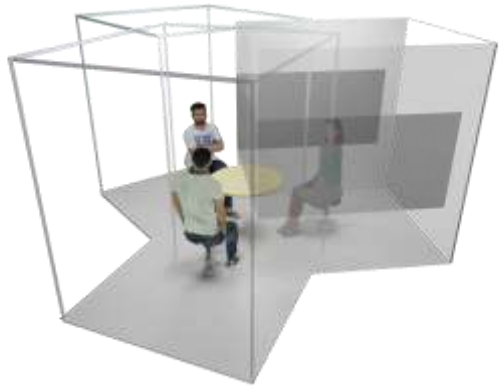
Demo III: working side-by-side

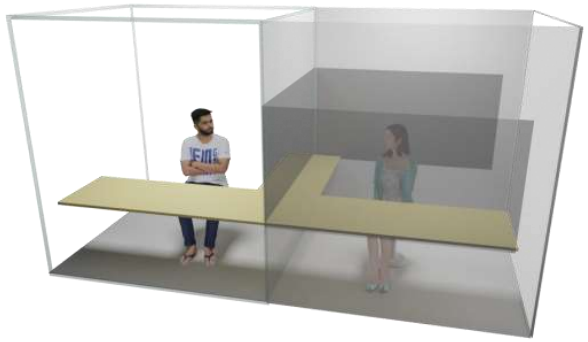


Conclusion

- Standardized and simplified
  - all using off-the-shelf hardware
- Real-time high-quality rendering
  - achieved by v-cube view algorithm
- Versatile modeling
  - achieved by v-cube assembly







IEEE CONFERENCE ON VIRTUAL REALITY + 3D USER INTERFACES
IEEE VR 2022
12-16 MARCH 2022 | VIRTUAL EVENT

Microsoft Research

Thanks

https://www.microsoft.com/en-us/research/project/virtualcube/

Acknowledgments

- Sicheng Xu for network training
- Yuxiao Guo for neural network conversion with NNFusion
- Steve Lin for paper proofreading
- Our users for their valuable feedback
- Anonymous reviewers for their helpful suggestions
- 3D human models in illustrative figures are authored by Perspectiv, ddd and rusttm88 from sketchfab.com under CC BY 4.0 License.