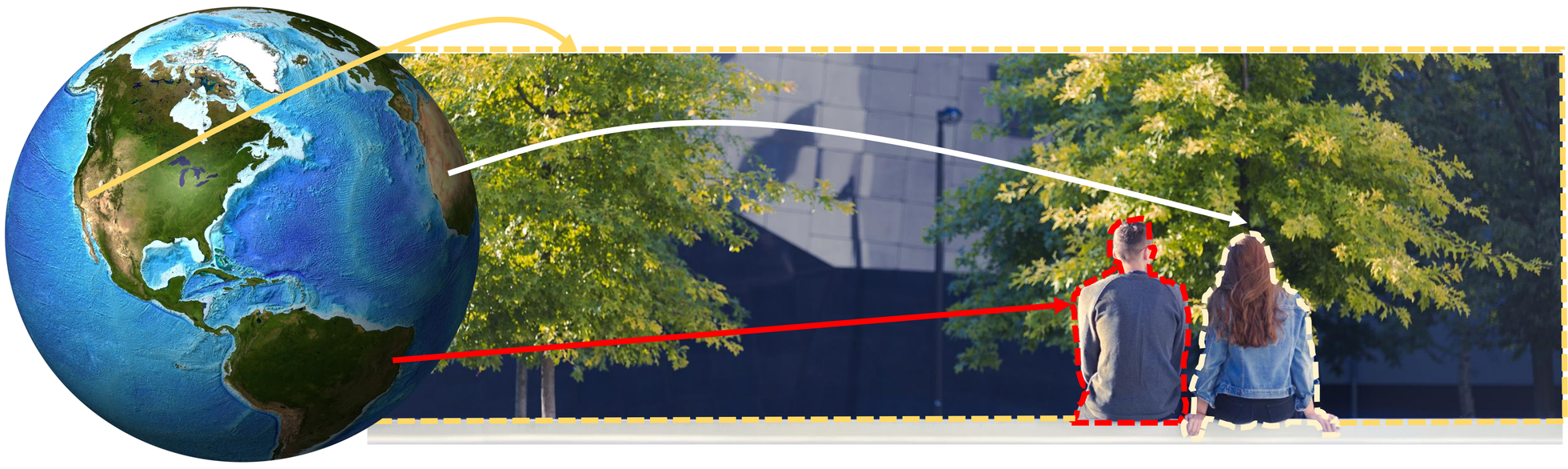
Context-Aware Remote User Placement for Immersive 360 Communications

A. Jiménez-Moreno, B. Rodriguez-Arbol, E. González-Sosa, P. Pérez, A. Villegas

Nokia

ABSTRACT. We use immersive technology (360 video and avatars) for personal communications, to provide a more natural experience among participants, on-site or remote, and to facilitate the interaction with the environment. To elicit sense of presence for people physically situated in another place we use avatars. In this scenario we propose an automatic placement of the avatars based on Artificial Intelligence in real time, calculating the positions according to the real environment, avoiding occlusions with other avatars and relevant parts of the scene. The result is a more natural communication experience for all peers.

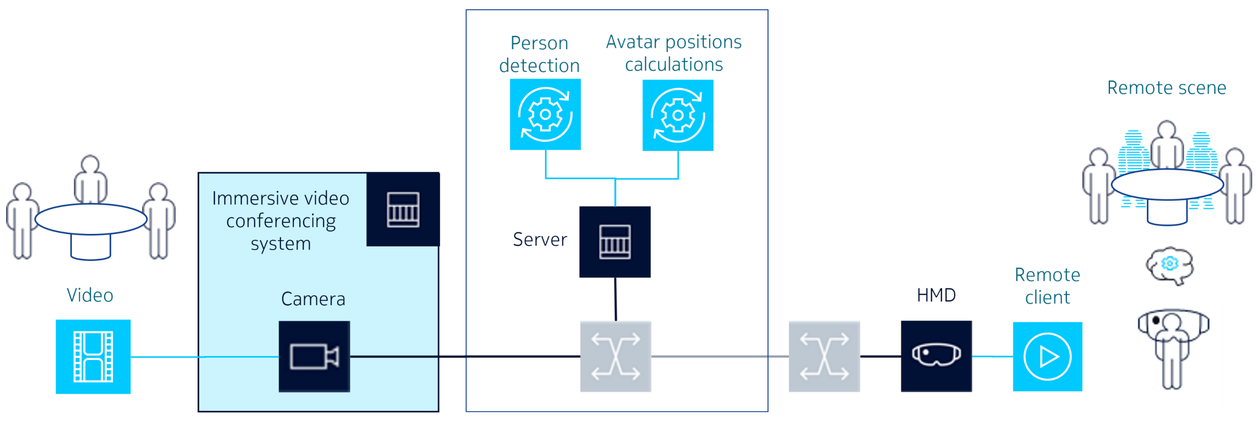


Main Components

- Scene analysis in 360 video
- Artificial Intelligence to recognize areas of interest for avatar positioning
- Offloading computationally expensive algorithms
- Real time processing
- Unity client combining all this information in the immersive scene

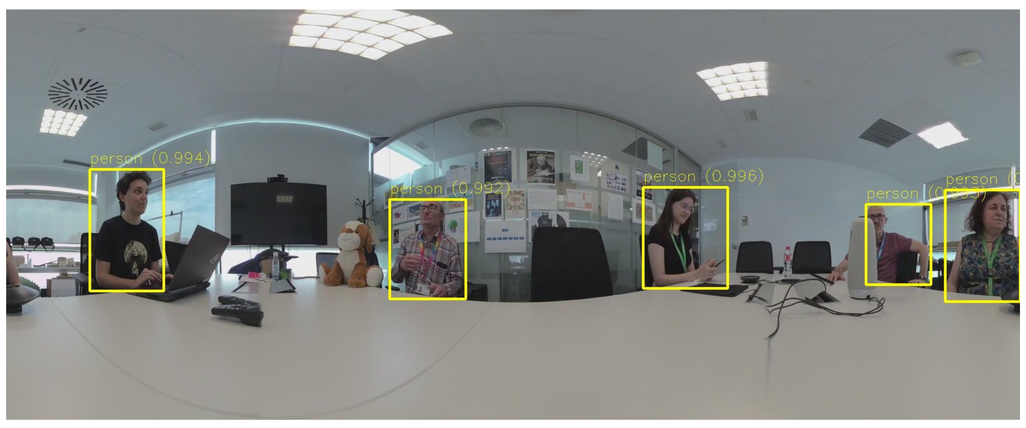
Architecture

- Based on the immersive video conferencing system in [1]
  - Video encoded with H.264 or H.265 and transmitted over UDP with RTP headers
  - Audio managed by Mumble library
  - Connectivity via Ethernet, WiFi, or mobile network (4G or 5G)

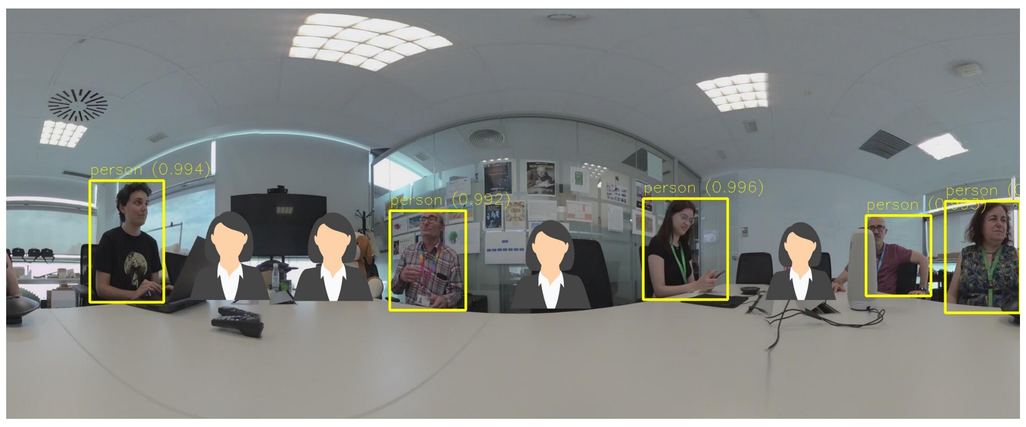


AI for scene analysis

- Person detection with Faster R-CNN [2] implemented with the library Detectron2 [3]
  - Avg. Intersection over Union of 65%.



- Calculation of available positions in the local scene
  - Between and above the calculated bounding boxes



Experimental evaluation

- Ten-minute meeting with four local users and two remote users
- 8Mbps video with 4K resolution encoded with H.264

- Our method improved the immersiveness as the avatars were correctly placed and steady, generating a comfortable scene
- More realistic avatars and a more complete scene analysis are considered for future work

[1] R. Kachach et al., "A Multi-Peer, Low Cost Immersive Communication System for Pandemic Times," 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, Portugal, 2021, pp. 691-692, doi: 10.1109/VRW52623.2021.00228

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., vol. 28. Curran Associates, Inc., 2015.

[3] Y. Wu et al., "Detectron2", 2019 https://github.com/facebookresearch/detectron2

IEEE AIxVR

Los Angeles, USA, Jan. 17-19 2024

Work supported by Project COMODIA, from 2021 call for R&D Projects in Artificial Intelligence and other digital Technologies and their integration in value chain (C005/21-ED). Exp. Number: 2021/C005/00144164

GOBIERNO DE ESPAÑA · VICEPRESIDENCIA PRIMERA DEL GOBIERNO · MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL · SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL · red.es · Plan de Recuperación, Transformación y Resiliencia · Funded by the European Union NextGenerationEU

NOKIA eXtended Reality Lab