

Evaluation of Segmentation Algorithms for Embodiment Improvement in an XR Application

Jiménez-Moreno, A., et al. (2025, March). Evaluation of Segmentation Algorithms for Embodiment Improvement in an XR Application. In *Proceedings of the 17th International Workshop on IMmersive Mixed and Virtual Environment Systems* (pp. 36-39)

Amaya Jiménez, Silvia Casino,
Elena Conderana, Marta Orduna,
Ester González-Sosa
Pablo Pérez, Álvaro Villegas

The Nokia logo is displayed in white, consisting of the word "NOKIA" in a sans-serif font. It is positioned within a large, stylized circular graphic on the right side of the slide. The graphic features a thick white outer ring and a dark blue inner circle, creating a target-like or tunnel-like effect.

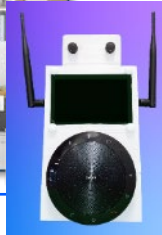
Introduction



Video-Based Self-Avatars (AI)
+
Natural Interaction



The Snowl:
Immersive Telepresence



"Full body video-based self-avatars for mixed reality: from E2E system to user study", in Virtual Reality

Motivation

New challenges

Real use case: *remote training for cooking users*

The student could see 360-degree video and her/his real environment to, for example, identify cooking tools or follow a receipt.

As first step, we decide to focus our research on **training table etiquette**



Oculus meta Quest 2
+ Stereo camera



Motivation

New challenges

Compare the results of different deep-learning based segmentation algorithms to combine real objects and video-based self-avatars:

YOLO-table	YOLO-table + ThunderNet	YOLO-table + EgoHOS
Yolov8: Segment egocentric body, cutlery, plates and glasses	ThunderNet: segment egocentric real body	EgoHOS: segment hands and any object in contact with them

Focus on **inclusivity**:

- People with Intellectual Disabilities (ID) from



Work approach

Research Questions

RQ1: Which algorithm is the best one, objectively and subjectively, to be used for the XR table etiquette application?

RQ2: Do barriers in technology start with evaluation methodologies? Can we adapt the evaluation methodology so the general population and participants with ID can perform the tests similarly?

2D and VR evaluation

2D:

- Evaluation with general population
- Evaluation with users from the Foundation (they are our end users)

VR:

- Evaluation with general population

Experimental evaluation

Test Material

Resolution	FPS	Length
1280x960	50	≈5s

Videos divided into four groups:

No action



Right hand action



Left hand action

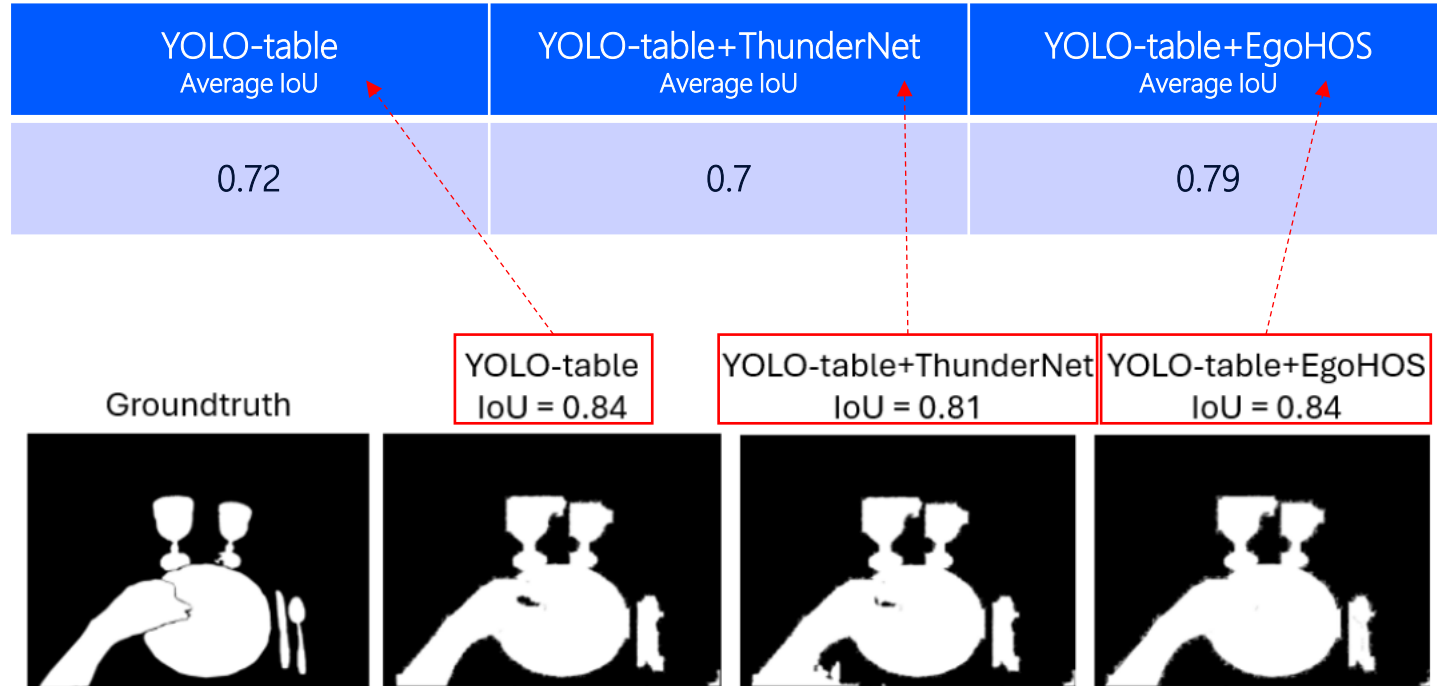


Both-handed action



Experimental evaluation

Objective Evaluation



Subjective assessment

Training session with the “perfect result”

2D – Participants with ID

2D - General population

Perfect result video



Original video



Perfect result



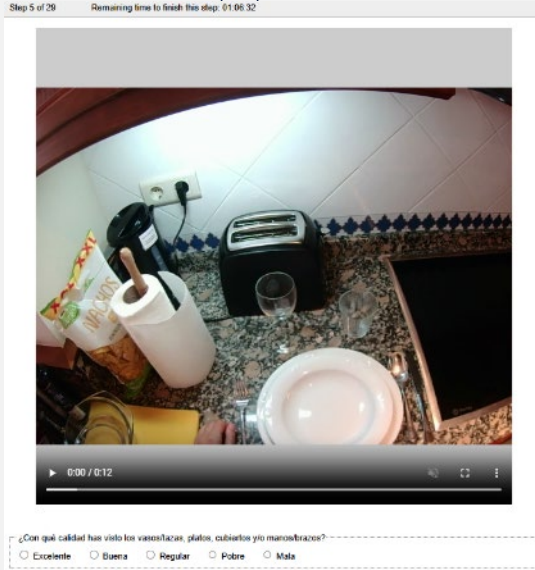
Hands, plates, glasses and cutlery and no other objects are visible

Experimental evaluation

Subjective Evaluation – environment and participants

- Right hand action
- Left hand action
- Both-handed action

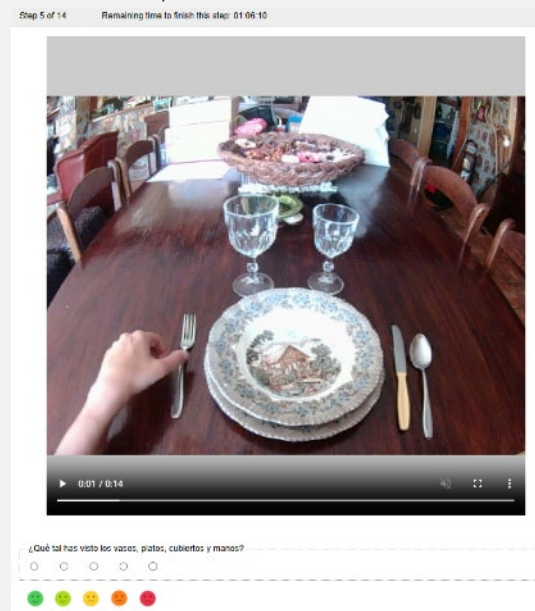
2D - General population



Evaluation of 8 videos with 3 different algorithms
"How good did you see the cutlery, plates, glasses, and the real body?"

20 participants (9 women and 11 men) with an average age of 36.05 years

2D - Participants with ID

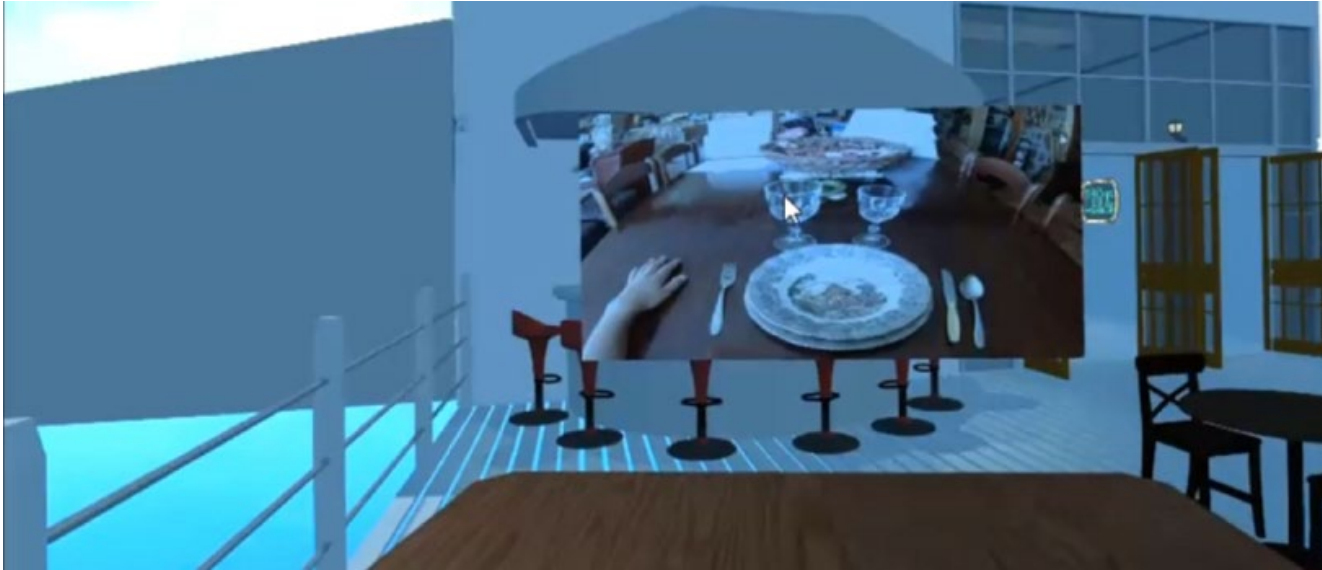


Evaluation of 3 videos with 3 different algorithms
"How good did you see the cutlery, plates, glasses, and the real body?"

19 participants (9 men and 2 women considered) with an average age of 45.82 years

Experimental evaluation

Subjective Evaluation – environment and participants

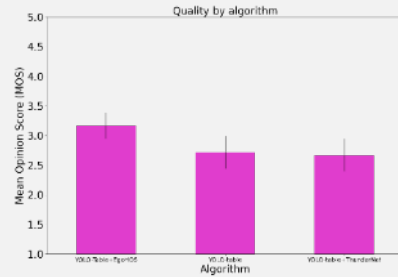
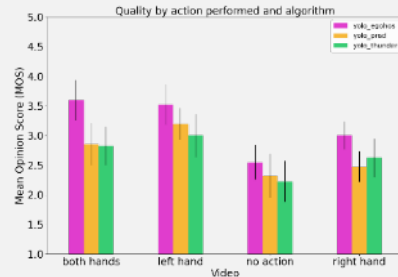


16 participants (7 women and 9 men) with an average age of 43.56 years

Experimental evaluation

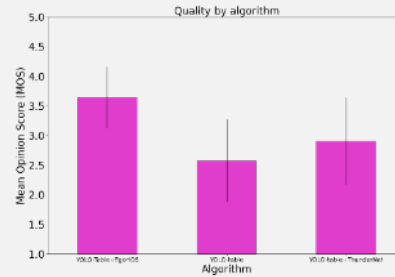
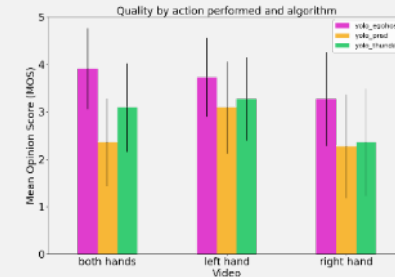
Subjective Evaluation – results

2D - General population



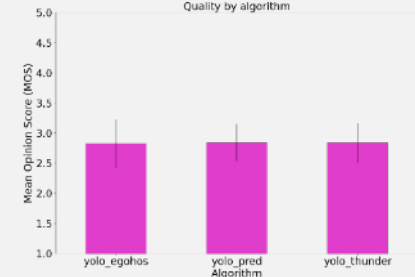
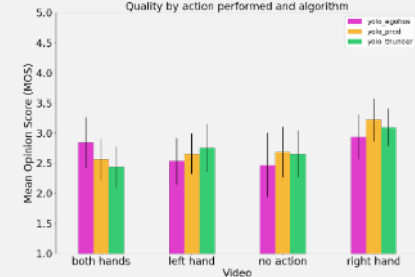
YOLO-table+EgoHOS is the preferred algorithm

2D- Participants with ID



YOLO-table+EgoHOS is the preferred algorithm

VR – General population



There is no preferred algorithm

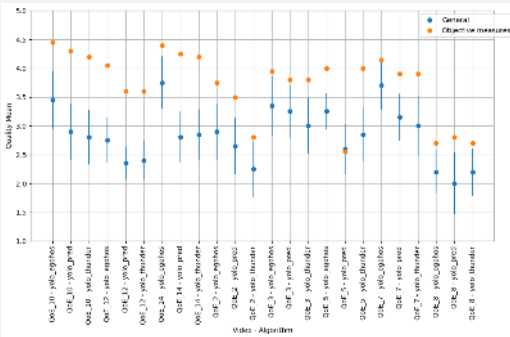
Experimental evaluation

Objective vs Subjective – results

Pearson's correlation

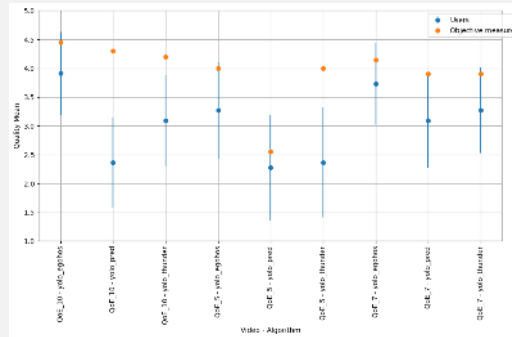
Objective– Subjective General Population	Objective– Subjective Population from the Foundation	Objective– Subjective VR population
0.99	0.87	-0.704

2D - General population



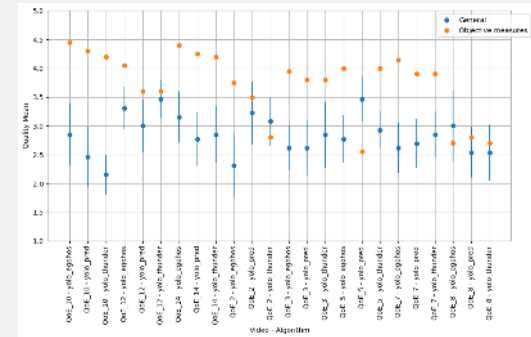
Subjective scores follow the same trend as objective measures

2D – Participants with ID



Subjective scores follow the same trend as objective measures

VR – General population



Differences between objective and subjective trends

Conclusions

Problems and general conclusions

Issues encountered:

- Limited literature on subjective evaluation of segmentation algorithms, especially in VR
- For the participants, **2D evaluation** is more difficult to understand (the direct application is not visible) but it is easier to evaluate
- In **VR evaluation**, the environment is more realistic, but other factors come into play, such as latency, depth perception, and other environmental conditions

General conclusions:

- Good **adaptation** for participants with ID, as subjective evaluations in 2D from both the general population and Foundation participants show similar results to the objective evaluation
- **YOLO-table+EgoHOS** is the preferred algorithm in the 2D evaluation
- Does it make sense to compare 2D evaluation with the evaluation in VR if there are many more factors that affect?
- More VR evaluations would be needed to draw clearer conclusions

Jiménez-Moreno, A., et al. (2025, March). Evaluation of Segmentation Algorithms for Embodiment Improvement in an XR Application. In *Proceedings of the 17th International Workshop on IMmersive Mixed and Virtual Environment Systems* (pp. 36-39)

Thank you!

Silvia Casino,
silvia.casino_colom.ext@nokia.com

This work has been funded by the Spanish Government program UNICO I+D 5G-6G 2022, project INCLUVERSO 5G (TSI-064200-2022-009), within the framework of the Recovery, Transformation and Resilience Plan

