Toward Behavioural Analysis of 6-DoF User When Consuming Immersive Media

Silvia Rossi, Irene Viola, Laura Toni and Pablo Cesar

Outline

- Introduction
- User navigation in 3- vs. 6-DoF environment
- User Trajectory Analysis in 6-DoF
- Results
- Conclusion

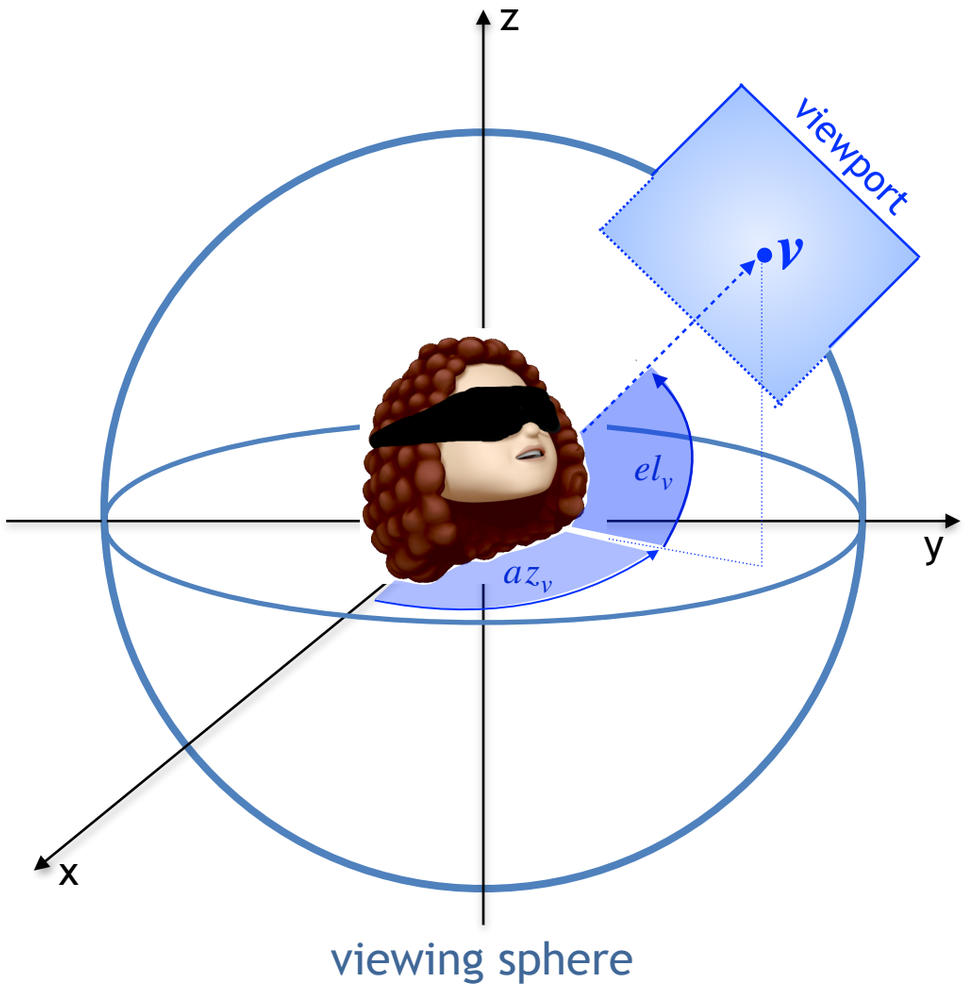
Traditional media consumption

VR: a user-centric era

VR: a user-centric era

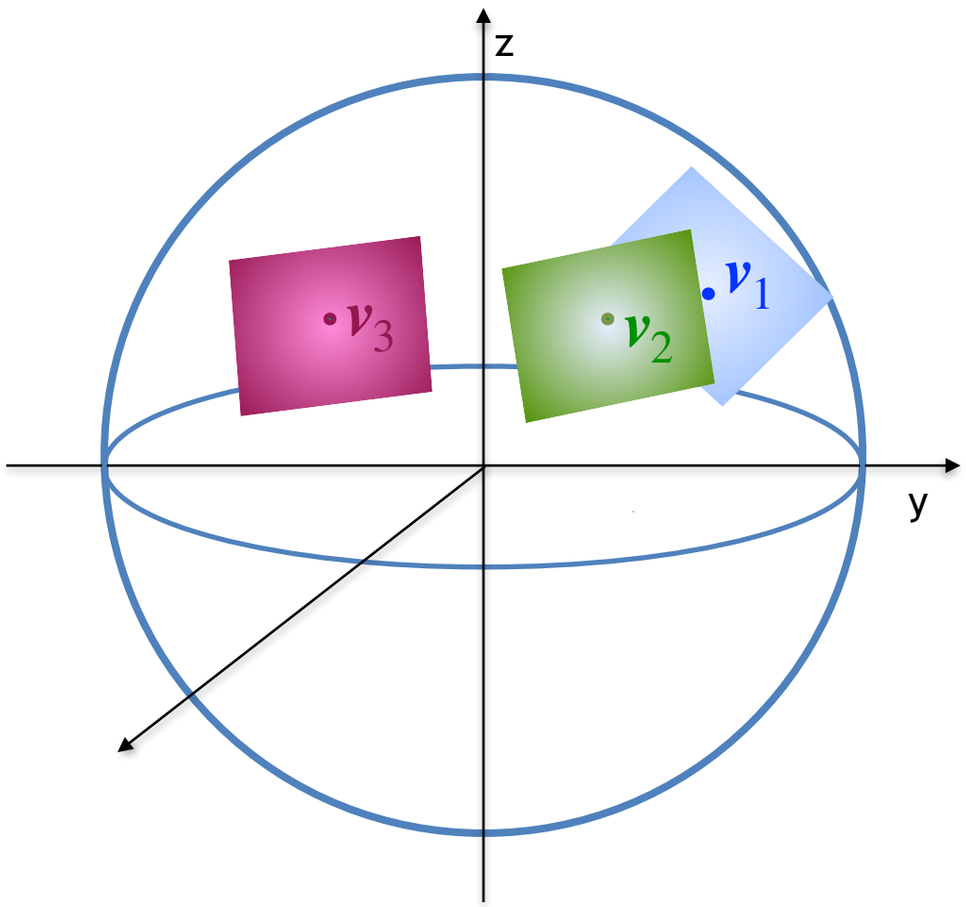
High Quality → Need of understanding user behaviour

User Navigation in 3-DoF



- Omnidirectional video is the de-facto VR format
- The user is virtually positioned at the center of the sphere
- The head is the only “interface” for interactivity
- The media is displayed from an *inward* position

User Navigation in 3-DoF



Distance between **viewport centres** as proxy of viewport overlap [1]

Distance as metric to assess user similarity

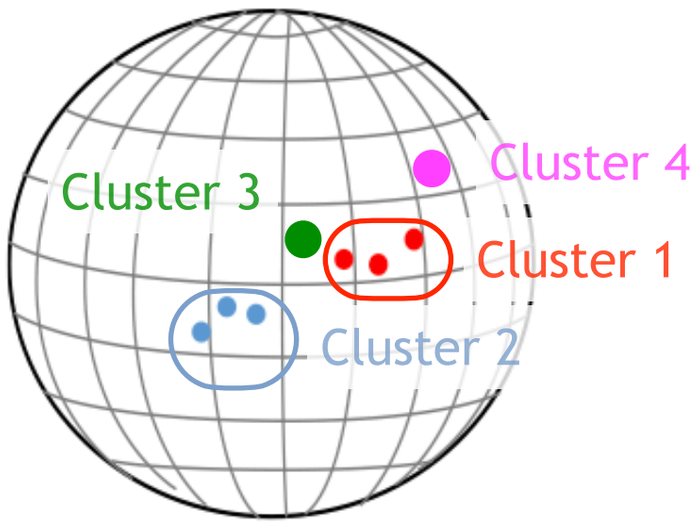
Clique-based clustering to detect user with similar behaviour

✔ The position of viewport center over time is highly informative to identify user behaviour.

[1] S. Rossi, F. De Simone, P. Frossard, and L. Toni. 2019. **Spherical Clustering of Users Navigating 360° Content.** In IEEE International Conference on Acoustics, Speech and Signal Processing.

A clique-based clustering

A **clique-based clustering** has been proposed in [1] to identify similar users that are attending the same portion of the omnidirectional content.



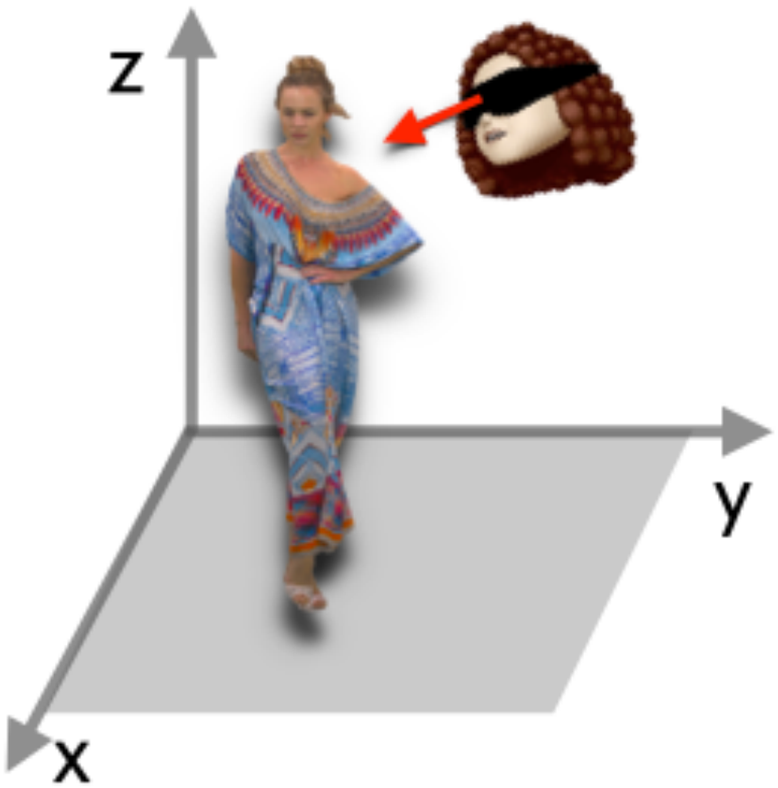
- Users are represented as an unweighted and undirected graph. Nodes are neighbours only if the geodesic distance between user viewport centres is lower than a given threshold.

- In graph theory, a clique is a set of nodes all connected each other.

  → **clusters** are **maximal cliques** detected in the graph.

[1] S. Rossi, F. De Simone, P. Frossard, and L. Toni. 2019. **Spherical Clustering of Users Navigating 360° Content.** In IEEE International Conference on Acoustics, Speech and Signal Processing.

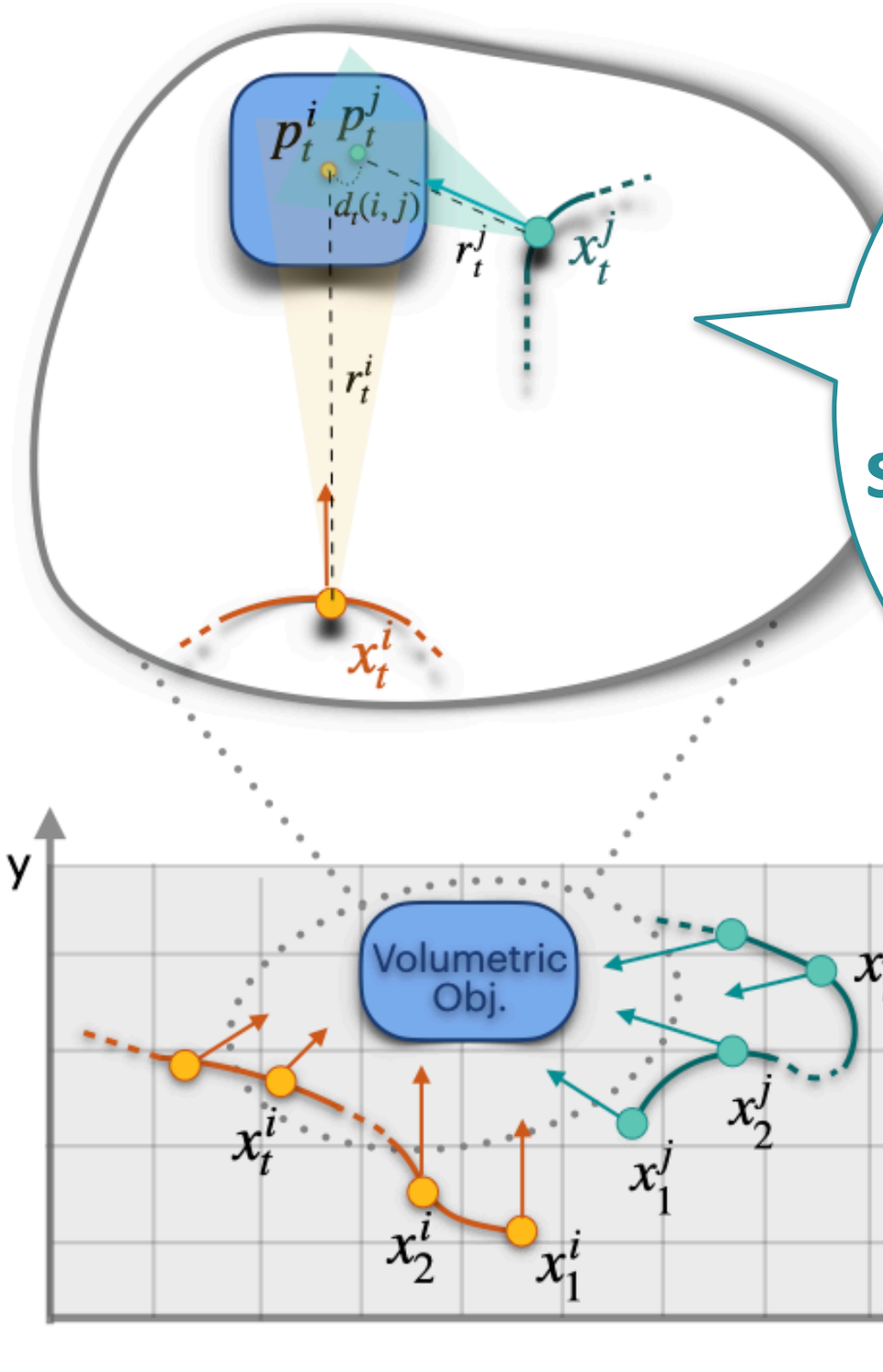
User Navigation in 6-DoF



- Volumetric content (meshes or point clouds)
- The user has now the freedom to move inside the VR space: higher level of interactivity
- The media is displayed from an *outward* position

✗ The position of viewport center over time is not enough to identify user behaviour.

User Navigation in 6-DoF

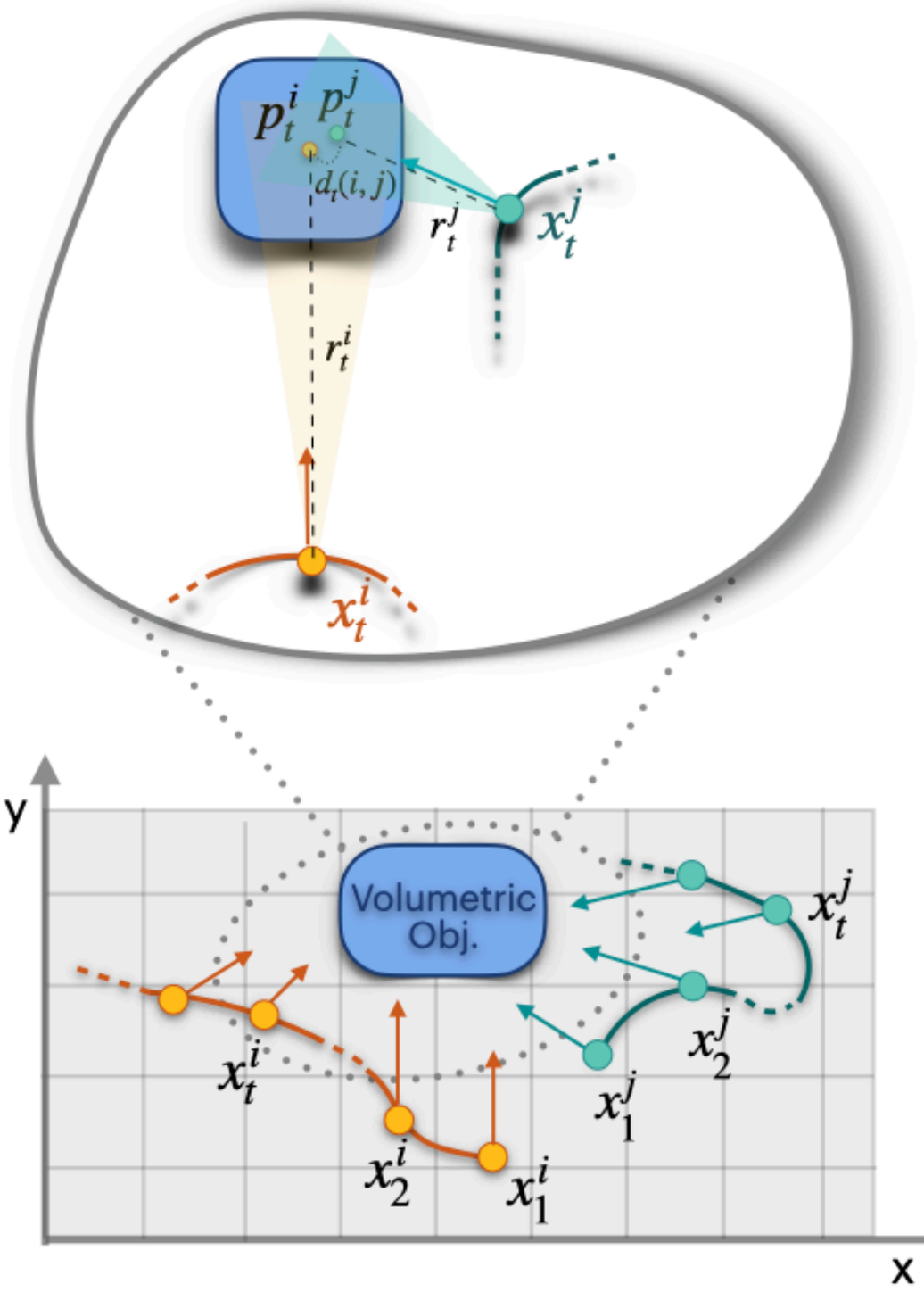


How to assess **user behaviour similarity in 6-DoF?**

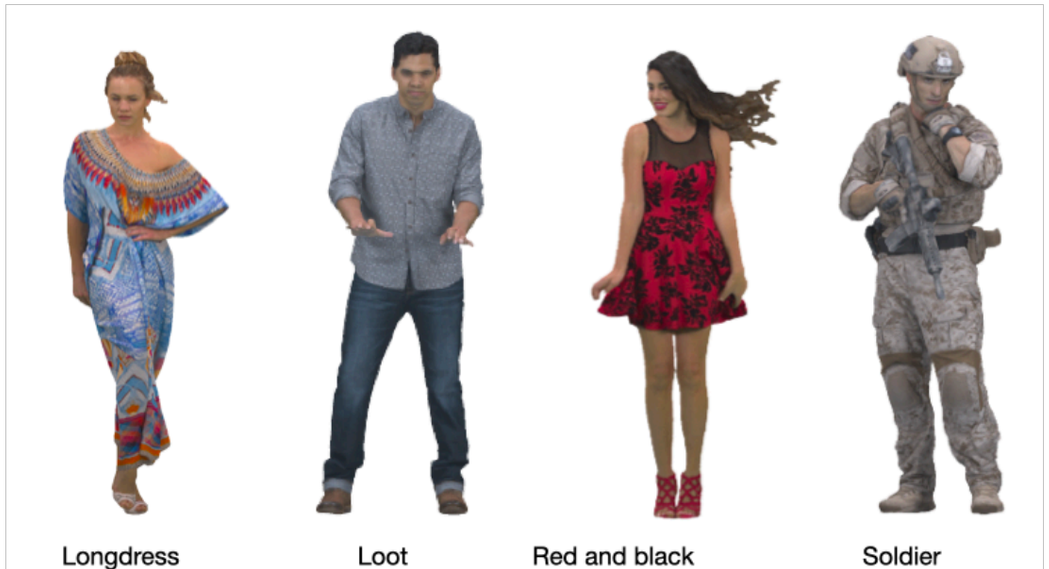
Distance Metrics

To verify if the **overlap ratio** $O_t^{i,j}$ can be substituted with a distance between users, we consider 4 different **distance metrics**:

- $L_x^2 \rightarrow$ euclidean distance between x_t^i, x_t^j user positions in the space
- $L_p^2 \rightarrow$ euclidean distance between p_t^i, p_t^j viewport centres on PC
- $G_p \rightarrow$ geodesic distance between p_t^i, p_t^j viewport centres on PC
- $L_p^1 \rightarrow$ cityblock distance between p_t^i, p_t^j viewport centres on PC



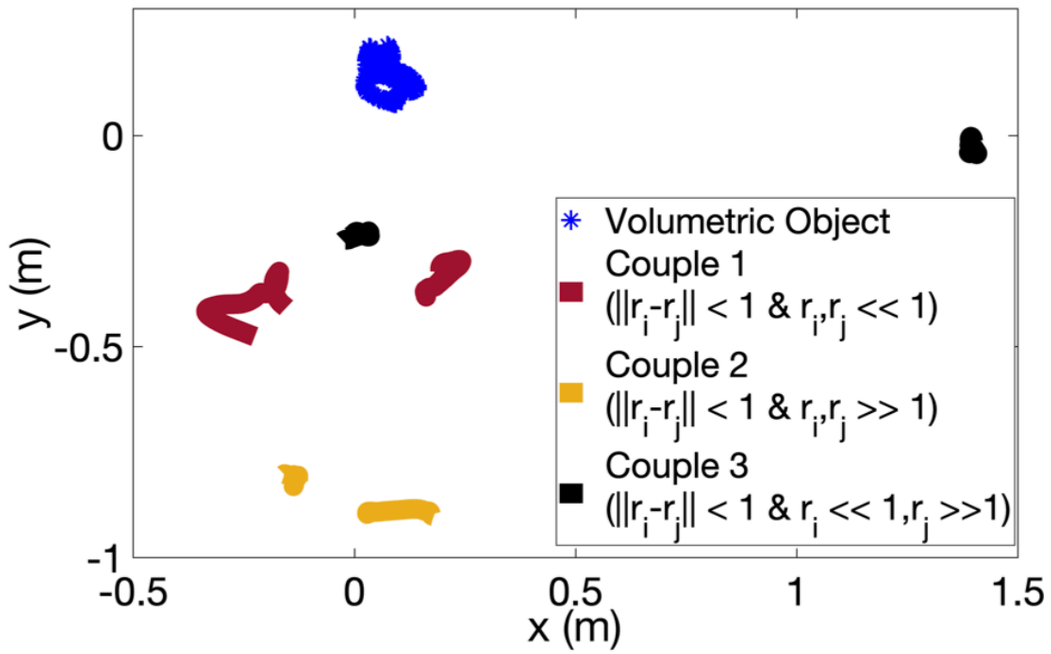
Dataset



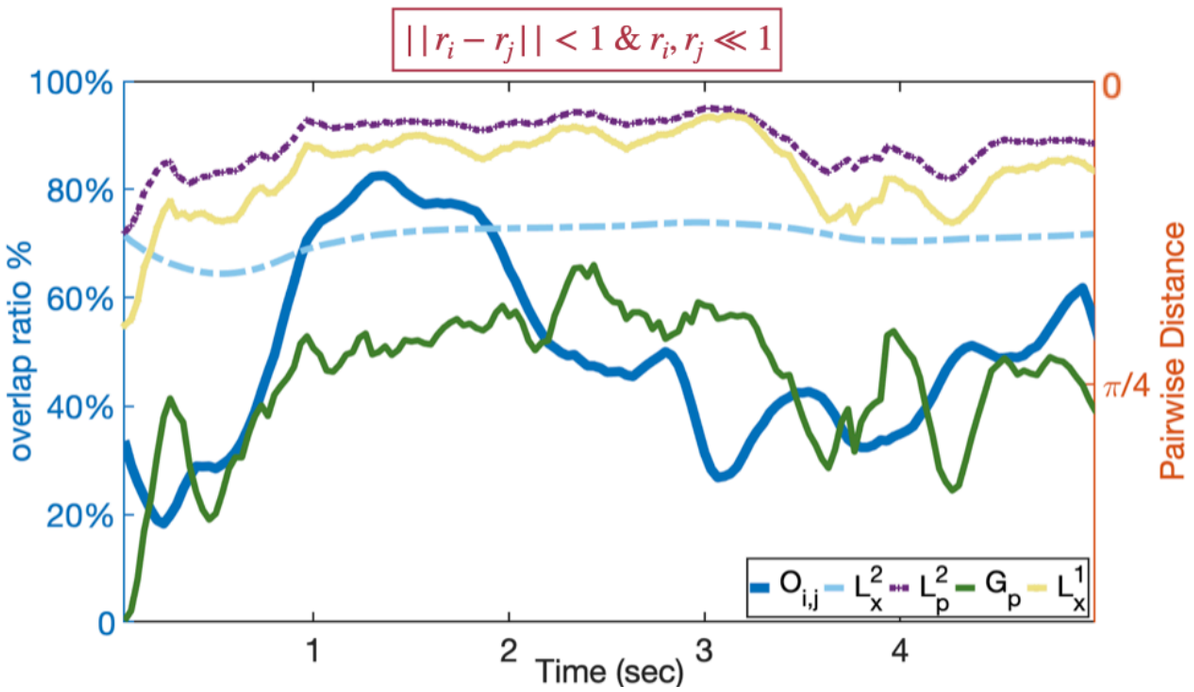
We based our investigations on a publicly available dataset of dynamic Point Clouds (PC) with collected navigation trajectories in 6-DoF [2].

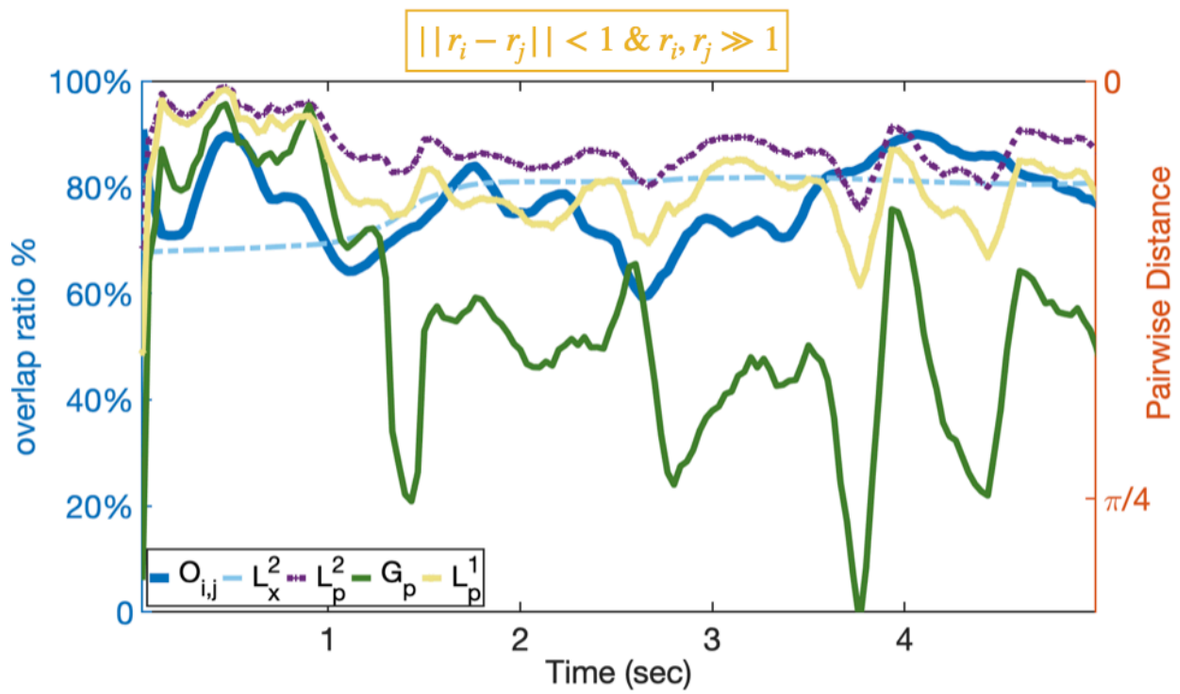
[2] S. Subramanyam, I. Viola, A. Hanjalic, and P. Cesar, "**User Centered Adaptive Streaming of Dynamic Point Clouds with Low Complexity Tiling**", in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.

Distance as proxy for VP overlap?

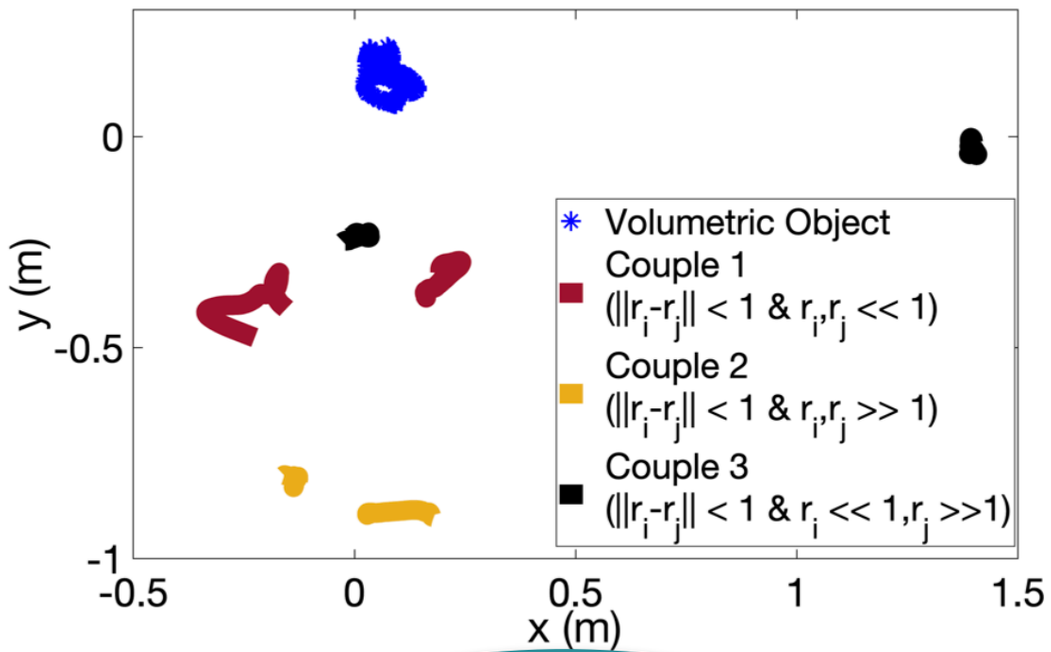


- L_x^2 → euclidean distance between x_t^i, x_t^j user positions in the space
- L_p^2 → euclidean distance between p_t^i, p_t^j viewport centres on PC
- G_p → geodesic distance between p_t^i, p_t^j viewport centres on PC
- L_p^1 → cityblock distance between p_t^i, p_t^j viewport centres on PC



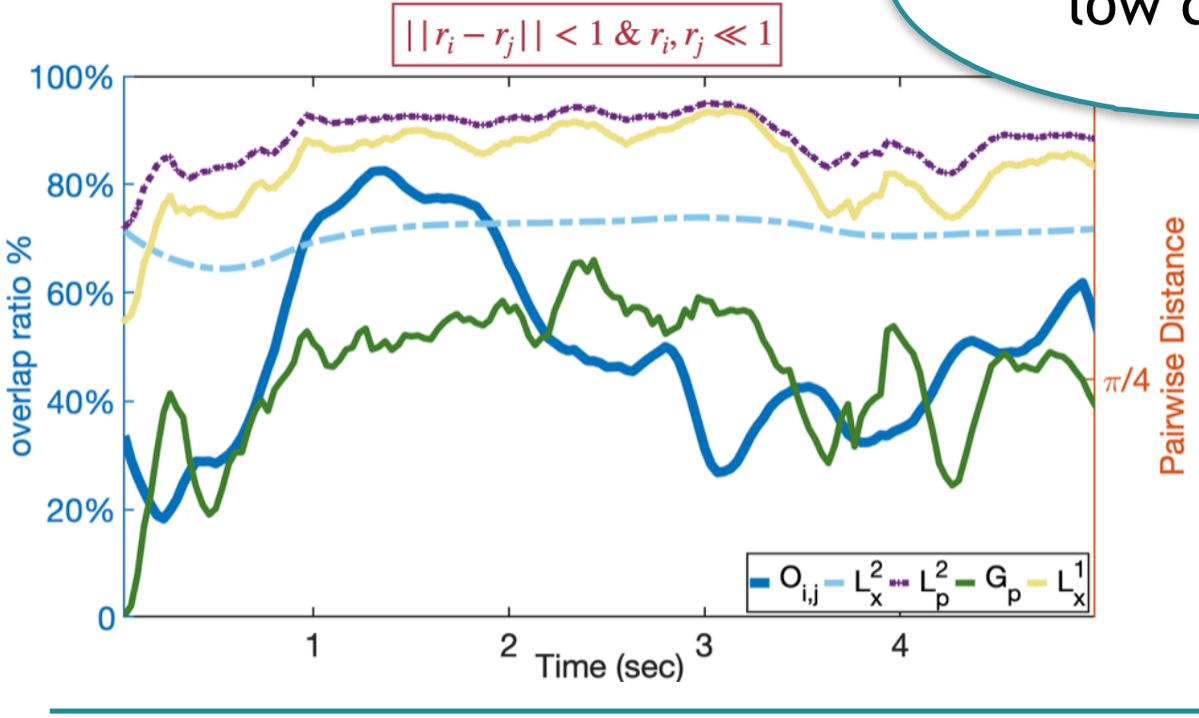


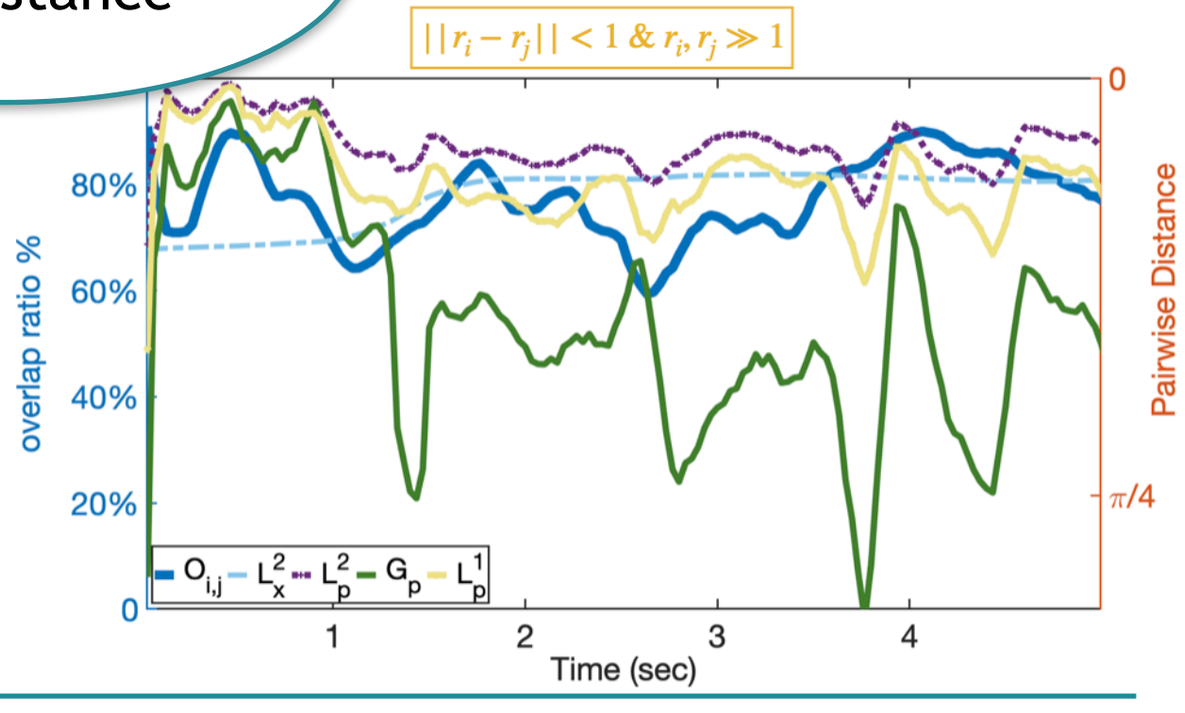
Distance as proxy for VP overlap?



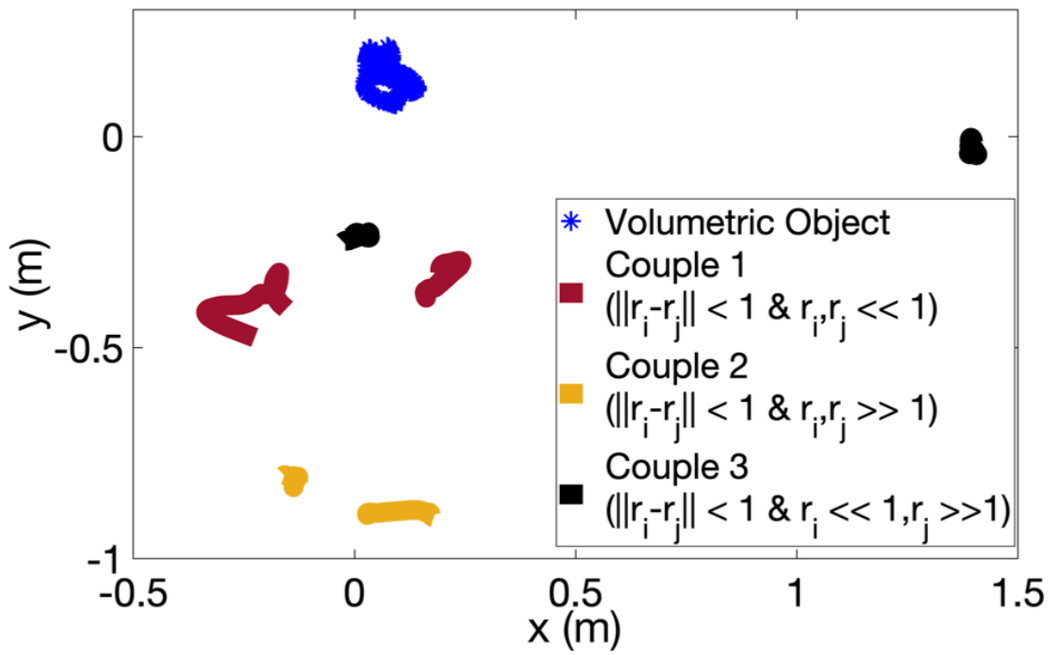
- L_x^2 → euclidean distance between x_t^i, x_t^j user positions in the space
- L_p^2 → euclidean distance between p_t^i, p_t^j viewport centres on PC
- G_p → geodesic distance between p_t^i, p_t^j viewport centres on PC
- L_p^1 → cityblock distance between p_t^i, p_t^j viewport centres on PC

High overlap, low distance

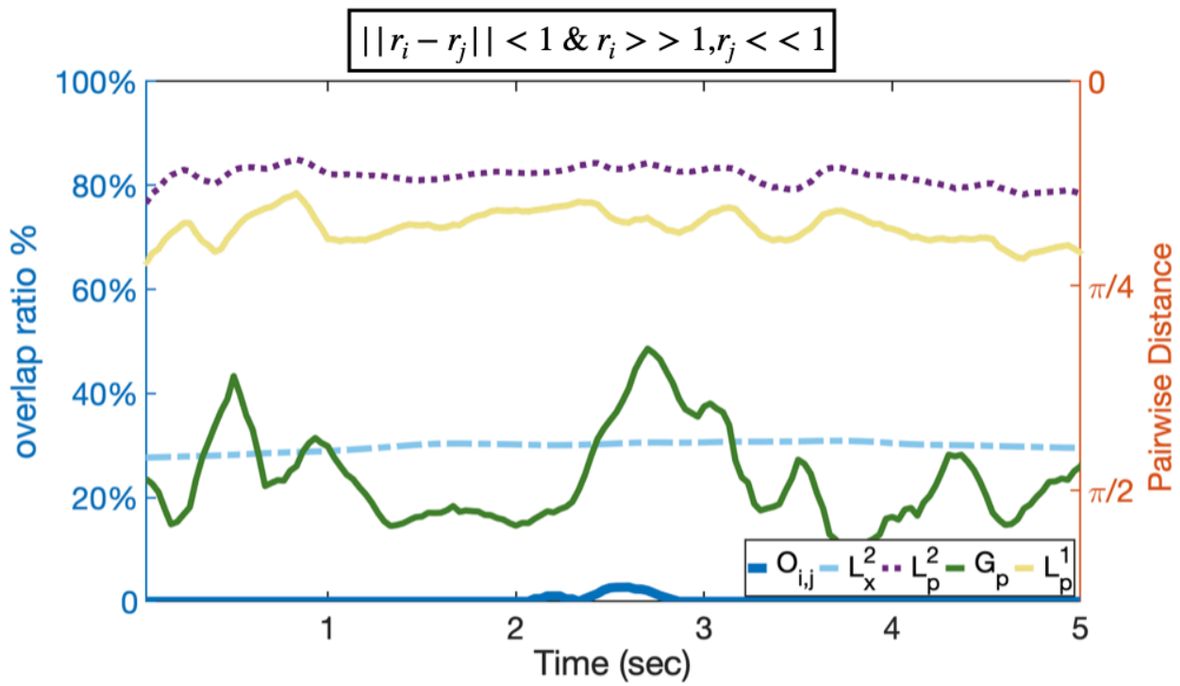




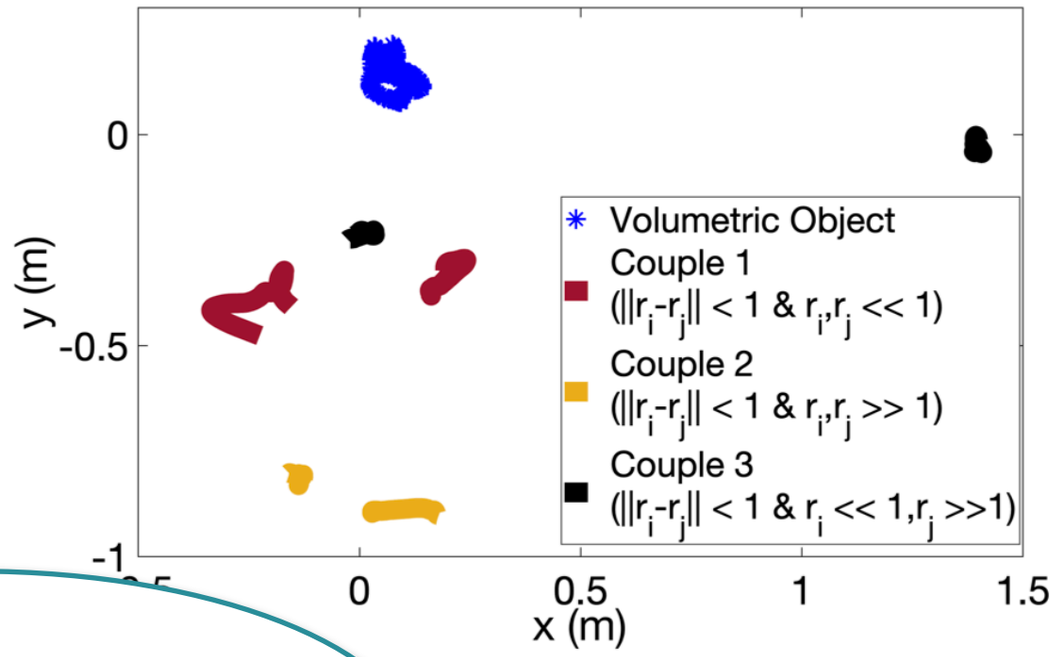
Distance as proxy for VP overlap?



- L_x^2 → euclidean distance between x_t^i, x_t^j user positions in the space
- L_p^2 → euclidean distance between p_t^i, p_t^j viewport centres on PC
- G_p → geodesic distance between p_t^i, p_t^j viewport centres on PC
- L_p^1 → cityblock distance between p_t^i, p_t^j viewport centres on PC

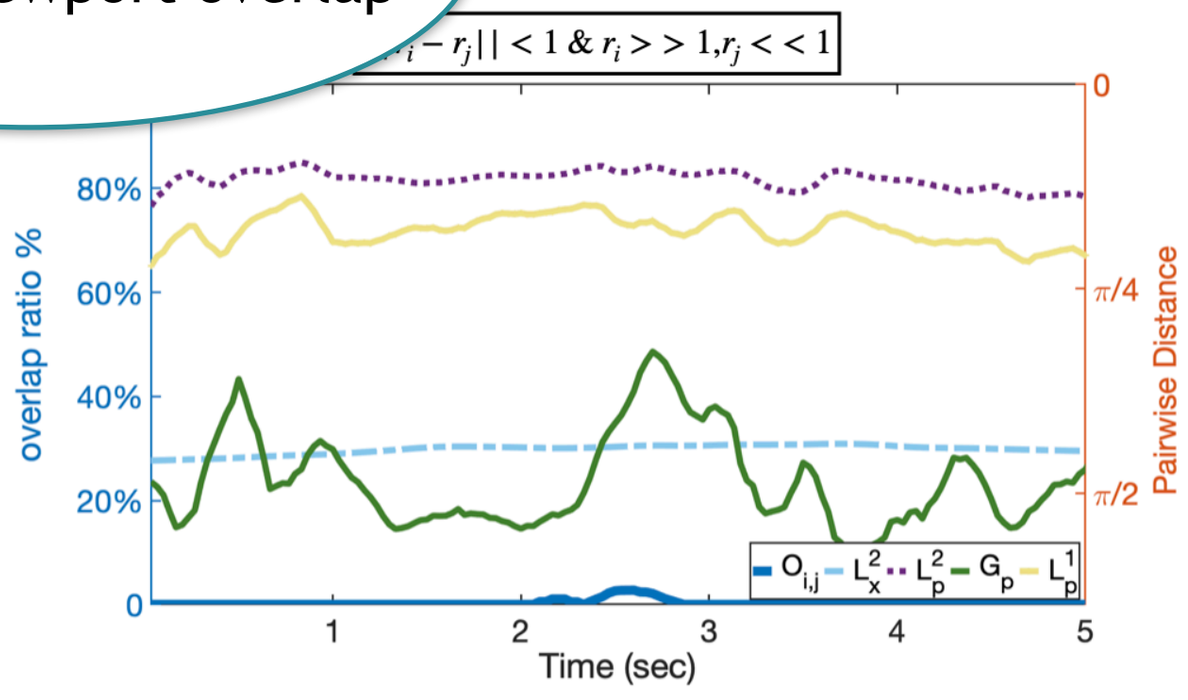


Distance as proxy for VP overlap?



- L_x^2 → euclidean distance between x_t^i, x_t^j user positions in the space
- L_p^2 → euclidean distance between p_t^i, p_t^j viewport centres on PC
- G_p → geodesic distance between p_t^i, p_t^j viewport centres on PC
- L_p^1 → cityblock distance between p_t^i, p_t^j viewport centres on PC

Distance metrics fail in capturing viewport overlap

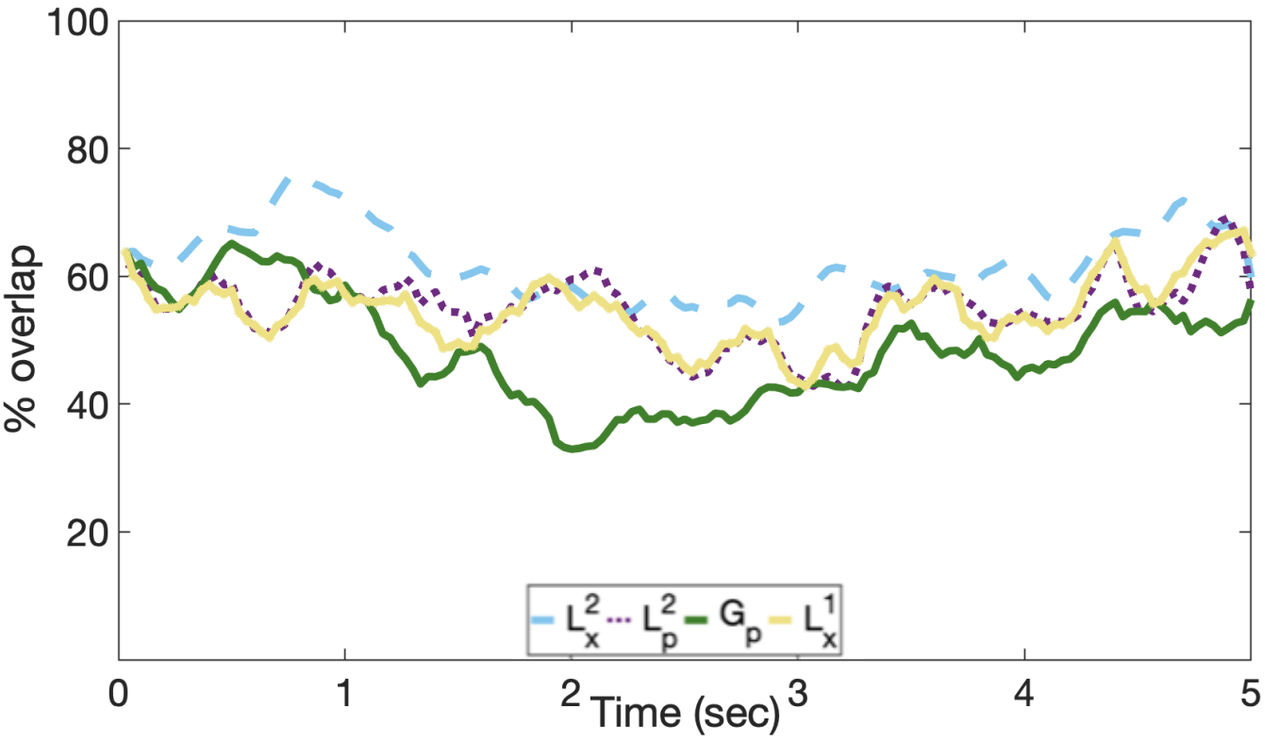


Clustering results

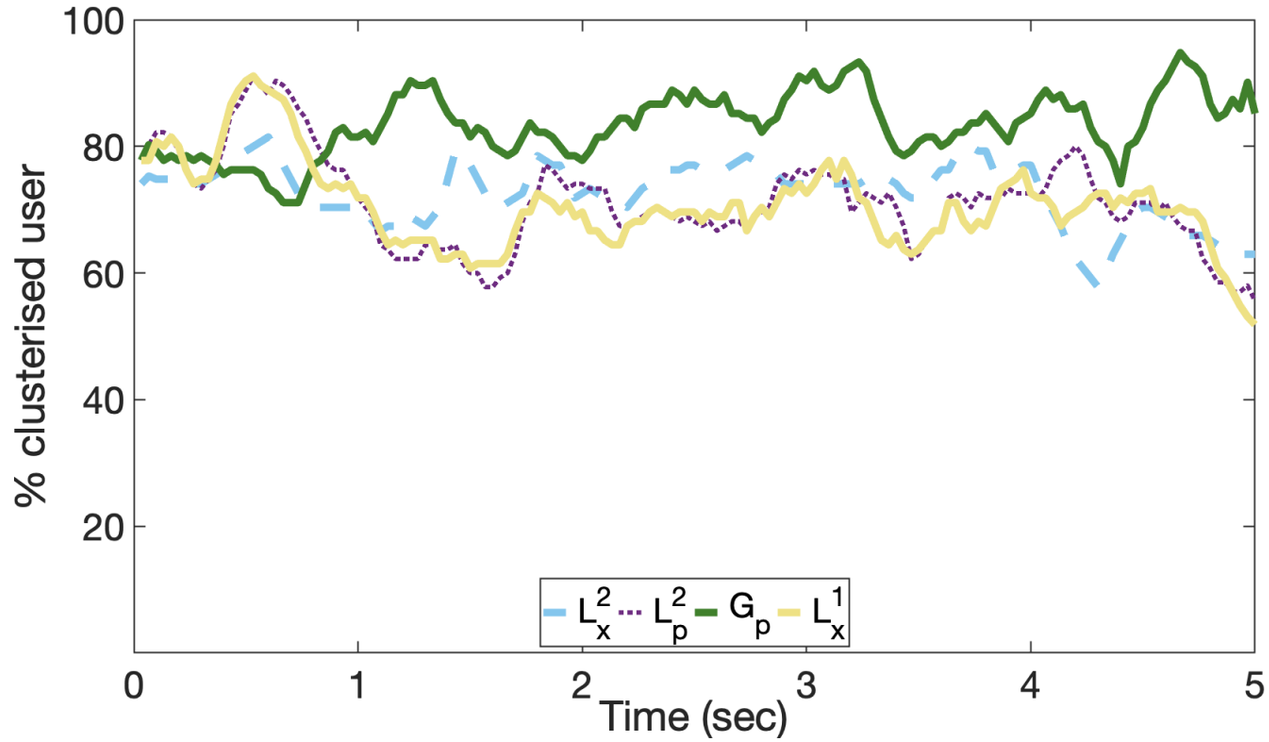

Red and black

We now applied a **clique-based clustering** proposed in [1] to identify similar users that are attending the same portion of the omnidirectional content.

Mean Overlap Over Time*



Mean N. Clustered Users Over Time*



- L_x^2 → euclidean distance between x_t^i, x_t^j user positions in the space
- L_p^2 → euclidean distance between p_t^i, p_t^j viewport centres on PC
- G_p → geodesic distance between p_t^i, p_t^j viewport centres on PC
- L_p^1 → cityblock distance between p_t^i, p_t^j viewport centres on PC

*** clusters > 2 users**

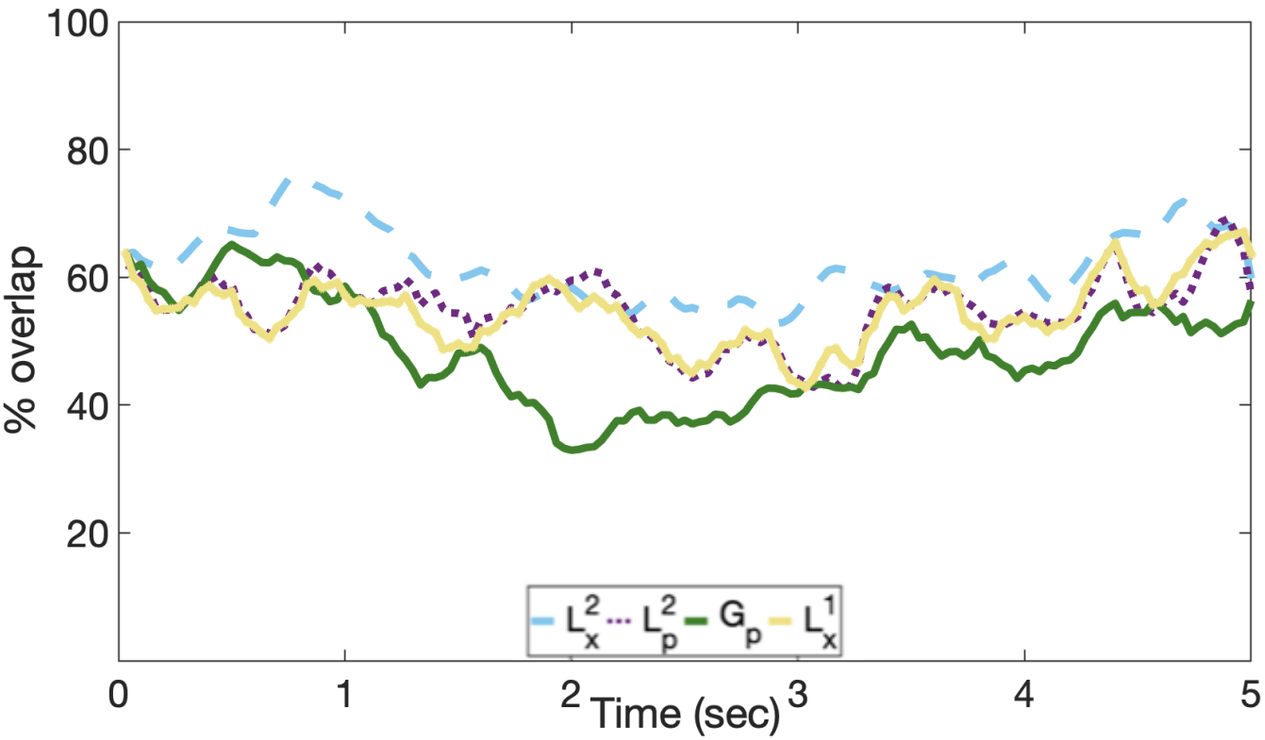
[1] S. Rossi, F. De Simone, P. Frossard, and L. Toni. 2019. **Spherical Clustering of Users Navigating 360° Content.** In IEEE International Conference on Acoustics, Speech and Signal Processing.

Clustering results

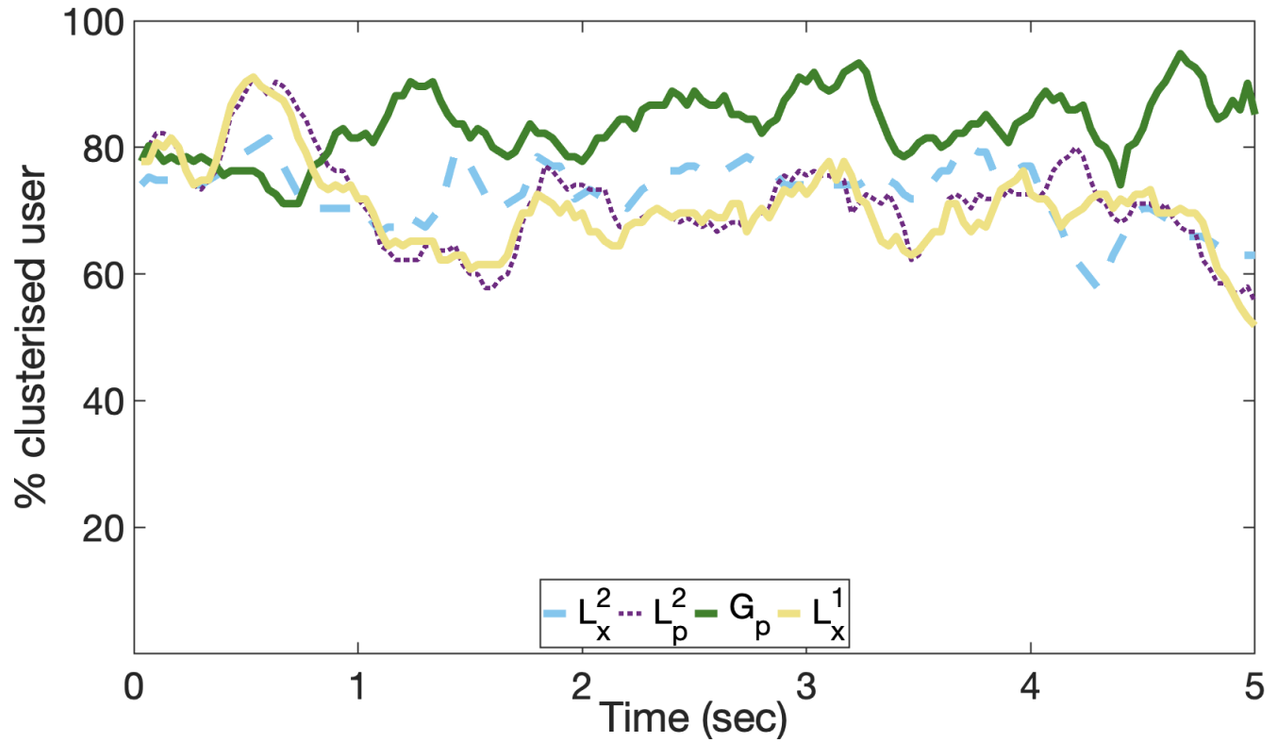

Red and black

We now applied a **clique-based clustering** proposed in [1] to identify similar users that are attending the same portion of the omnidirectional content.

Mean Overlap Over Time*



Mean N. Clustered Users Over Time*



Clusters based on:

- G_p → able to identify consistent groups of users
- L_x^2 → perform better in terms of viewport overlap

*** clusters > 2 users**

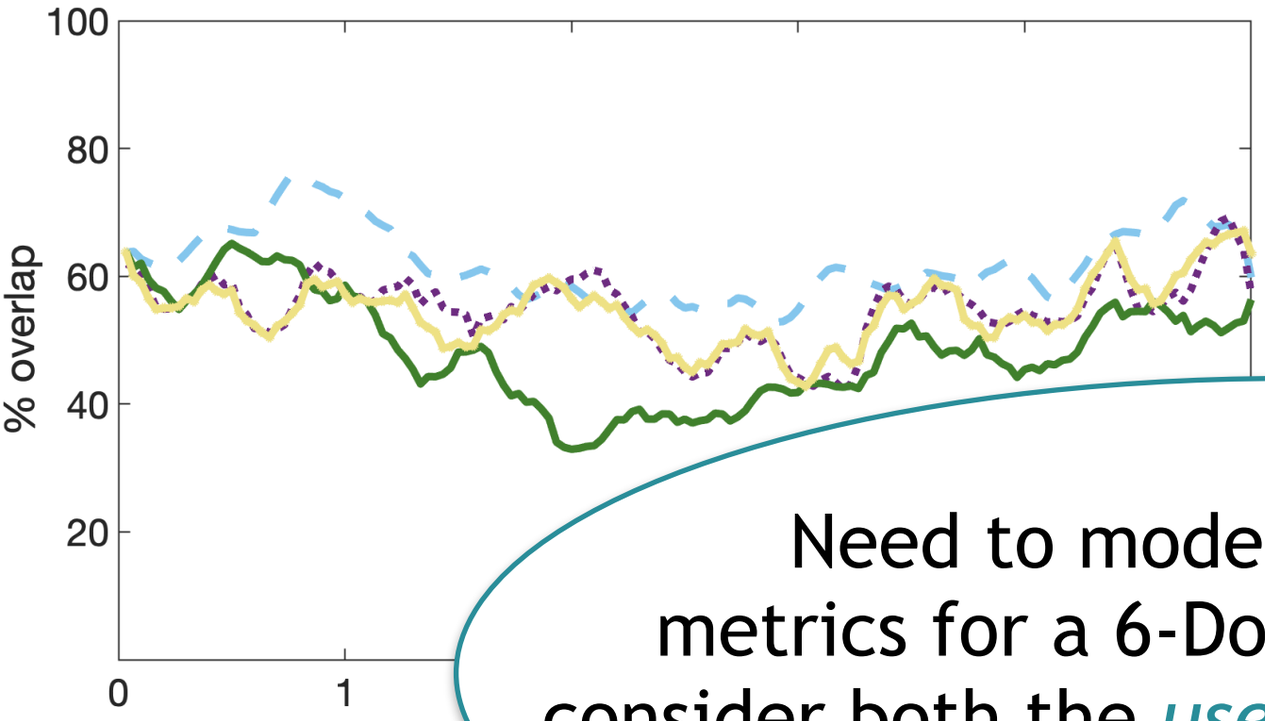
[1] S. Rossi, F. De Simone, P. Frossard, and L. Toni. 2019. **Spherical Clustering of Users Navigating 360° Content.** In IEEE International Conference on Acoustics, Speech and Signal Processing.

Clustering results

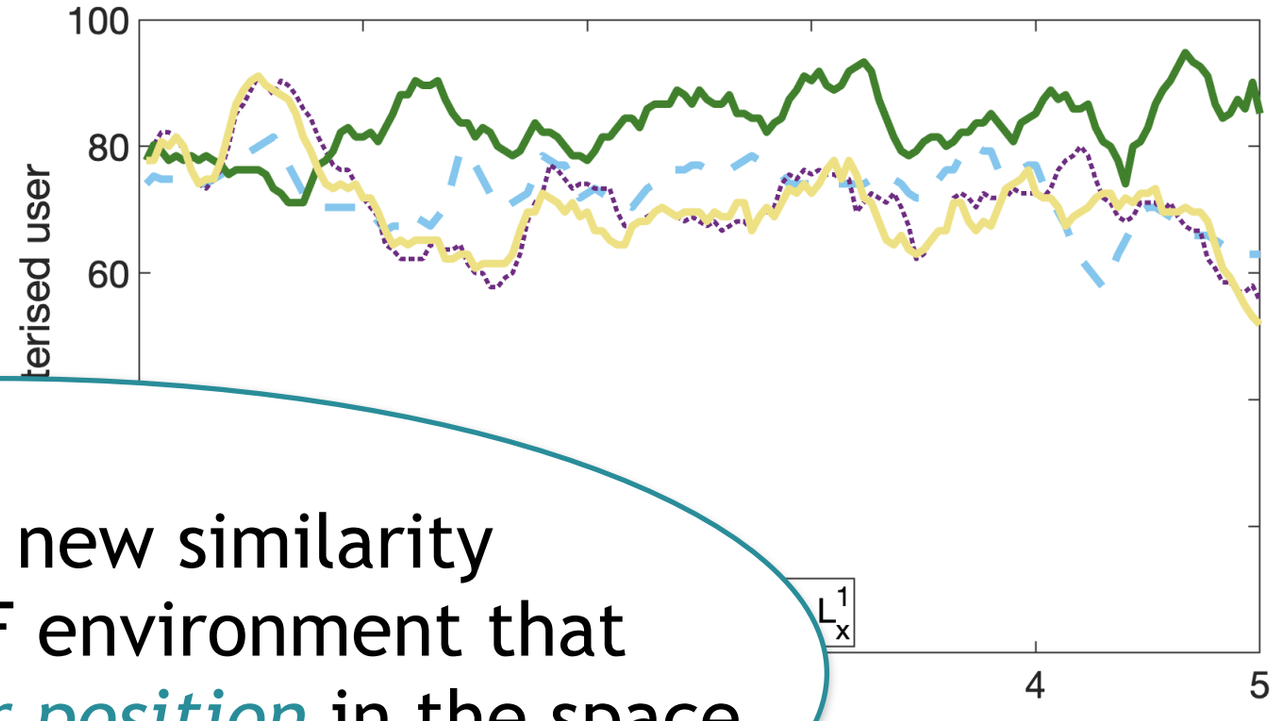

Red and black

We now applied a **clique-based clustering** proposed in [1] to identify similar users that are attending the same portion of the omnidirectional content.

Mean Overlap Over Time*



Mean N. Clustered Users Over Time*



Need to model new similarity metrics for a 6-DoF environment that consider both the *user position* in the space and *viewing direction*.

G_p ... of users

L_x^2 → perform better in terms of viewport overlap

*** clusters > 2 users**

[1] S. Rossi, F. De Simone, P. Frossard, and L. Toni. 2019. **Spherical Clustering of Users Navigating 360° Content.** In IEEE International Conference on Acoustics, Speech and Signal Processing.

Conclusion

We have shown key differences of interactivity in 3- and 6-DoF VR space:

- The definition of navigation trajectory is different
- Current metrics fail in capturing similarity among users

➡ We are currently working on metrics that better model the user behaviour in 6-DoF and exploring different datasets.

Thank you for your attention

Questions?