De-anonymization Attacks on Metaverse

Yan Meng[1], Yuxia Zhan[1], Jiachun Li[1], Suguo Du[1], Haojin Zhu[1], and Xuemin (Sherman) Shen[2]

[1]Shanghai Jiao Tong University
[2]University of Waterloo

May, 2023

OUTLINE

- Background
- Motivation
- System Design
- Evaluations
- Discussions
- Conclusion

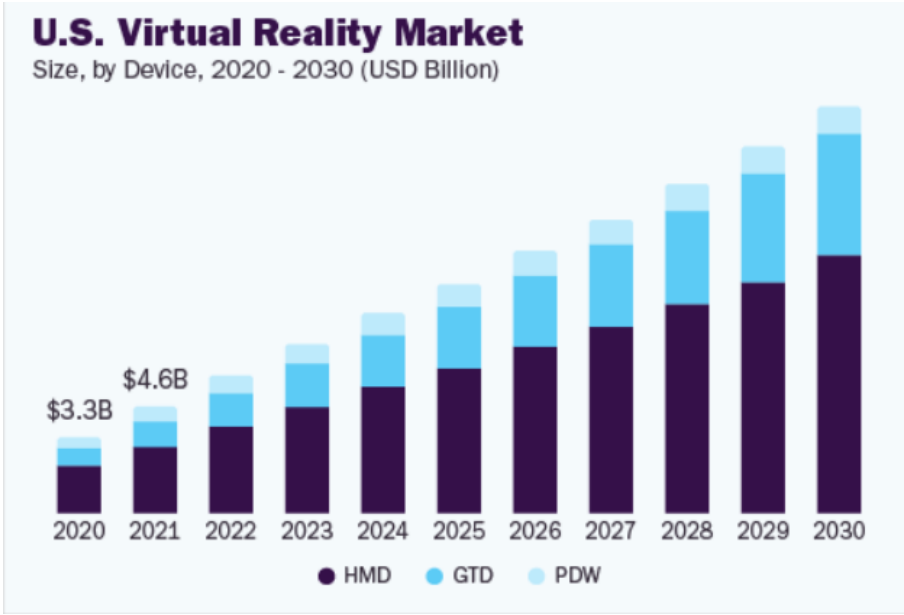
1

Background

VR-driven Metaverse

- Immerse experience & Advanced human-computer interaction
- Undergoing rapid growth of market size

VR scenario[1,2]

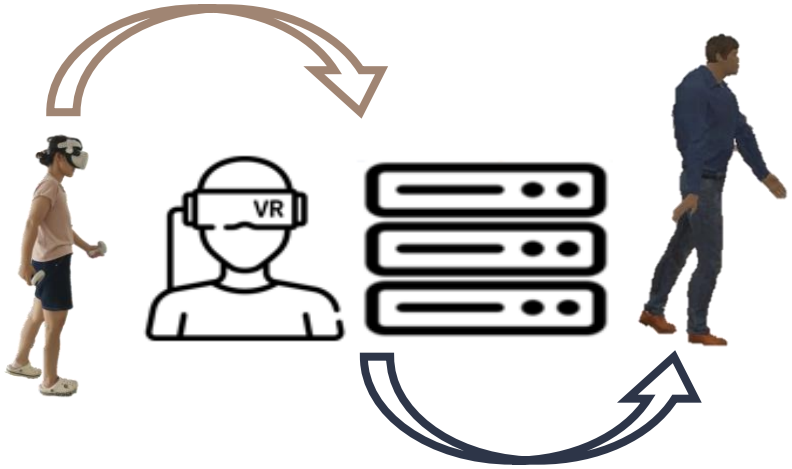


Global Market: $28.4 Billion (2022) [3]
CAGR: 13.8% (2023-2030)

1. https://up.enterdesk.com/edpic/bf/a2/91/bfa2919485d524f1477e06ba82a7e7bb.jpg
2. https://image11.m1905.cn/uploadfile/2018/0323/20180323091601495478.jpg
3. https://www.grandviewresearch.com/industry-analysis/virtual-reality-vr-market

Anonymization in VR

- User's real identity is masked by *avatar*
- Avatar: a digital representation of the user in the virtual world



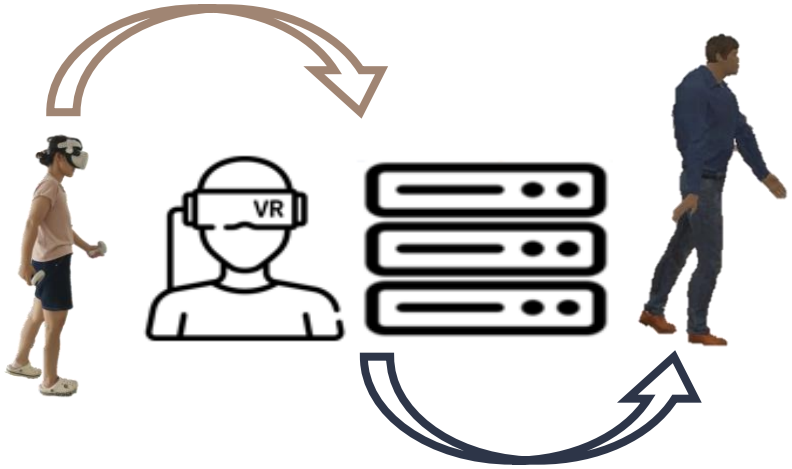
Convert real user to avatar

Various avatars in VR[1,2]

1. https://news.zol.com.cn/764/7647534.html
2. https://www.youtube.com/watch?v=PWLPw4RE9Ig&t=18s

Anonymization in VR

- User's real identity is masked by *avatar*
- Avatar: a digital representation of the user in the virtual world



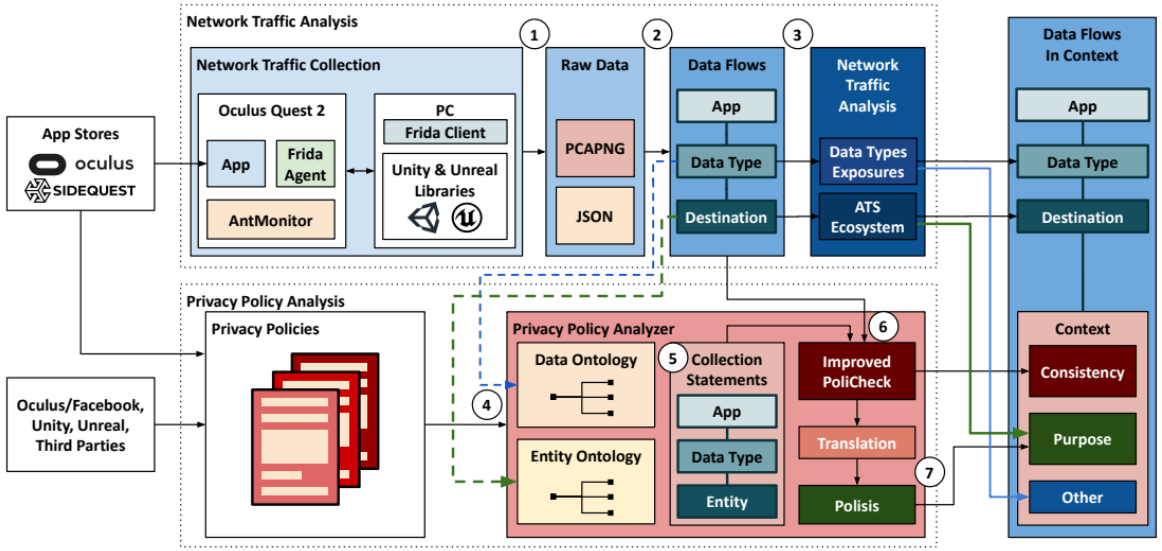
Convert real user to avatar

Various avatars in VR

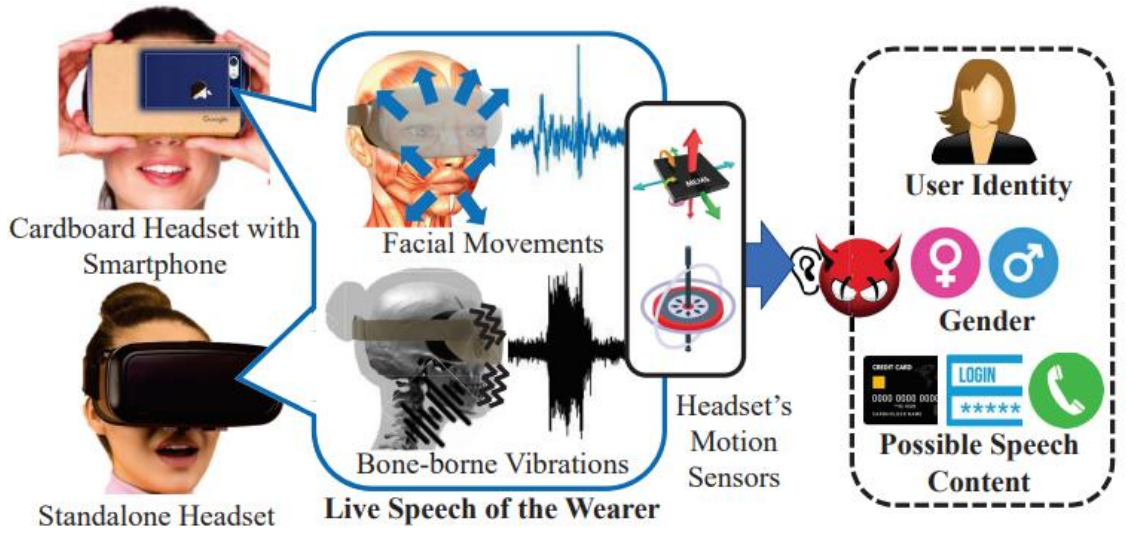
De-anonymization attack: *is there anyway to link user's real identity to his/her avatar?*

Existing De-anonymization Attacks in VR

- **Traffic** analysis solution
- Side-channel attack based on **sensor** information



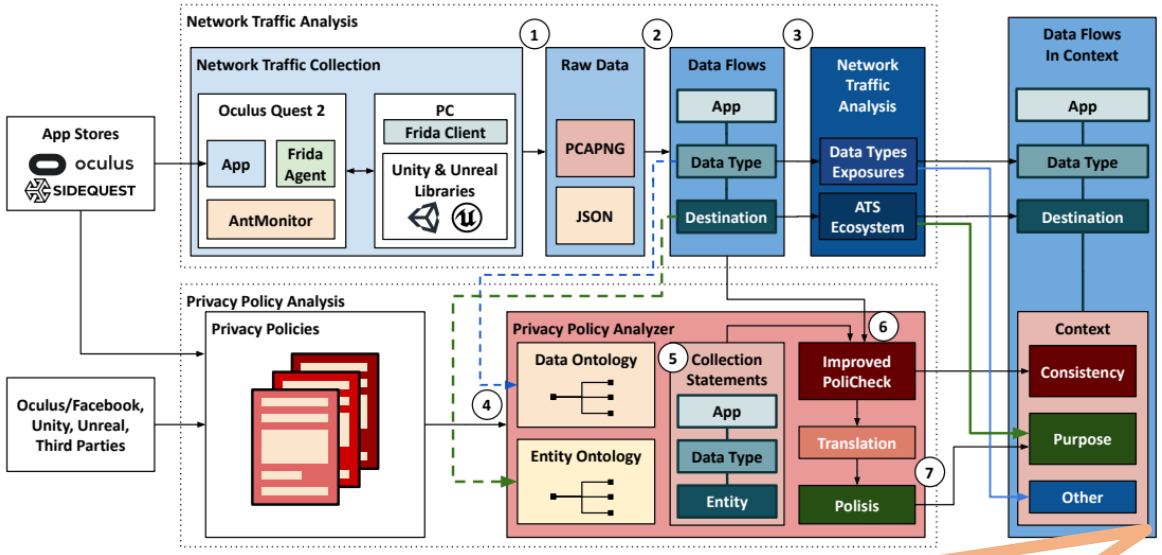
OVRSeen: traffic analysis
(Trimananda et al, USENIX Security'22)

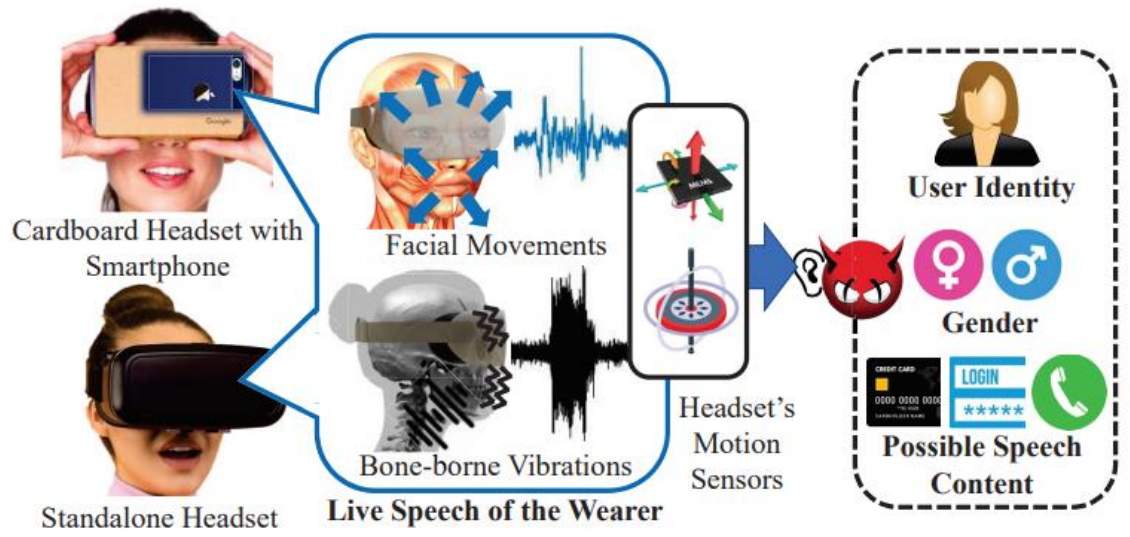


Face-Mic
(Shi et al, MobiCom'22)

Existing De-anonymization Attacks in VR

- **Traffic** analysis solution
- Side-channel attack based on **sensor** information



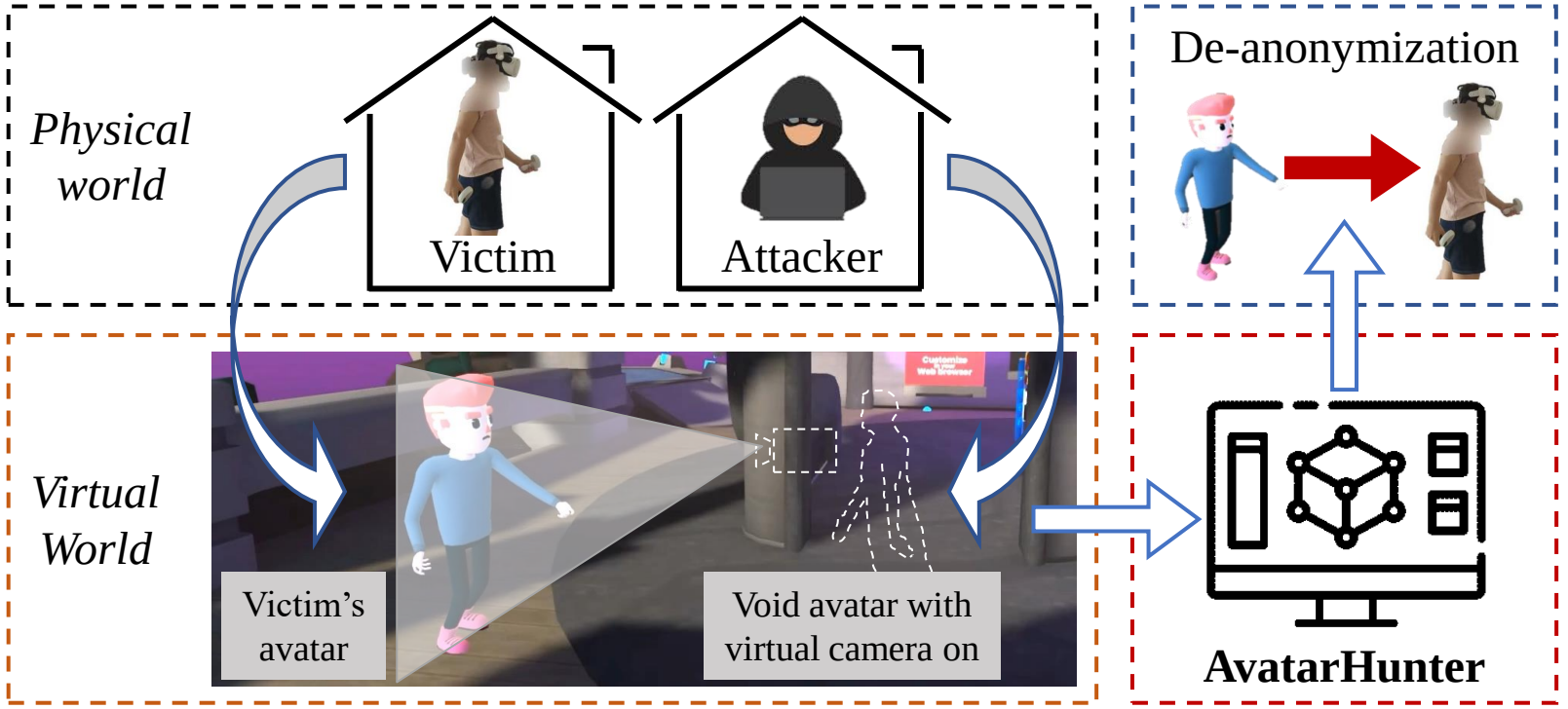


(Trimana … n'21)

Strong attack assumption: *malicious app developers*, *or data collection from VR devices*

Research Goal

- **Is it possible for the external attacker to perform a de-anonymization without requiring any permission?**
- Attack scenario: tracking users in ***avatar-changeable*** VR games or meetings

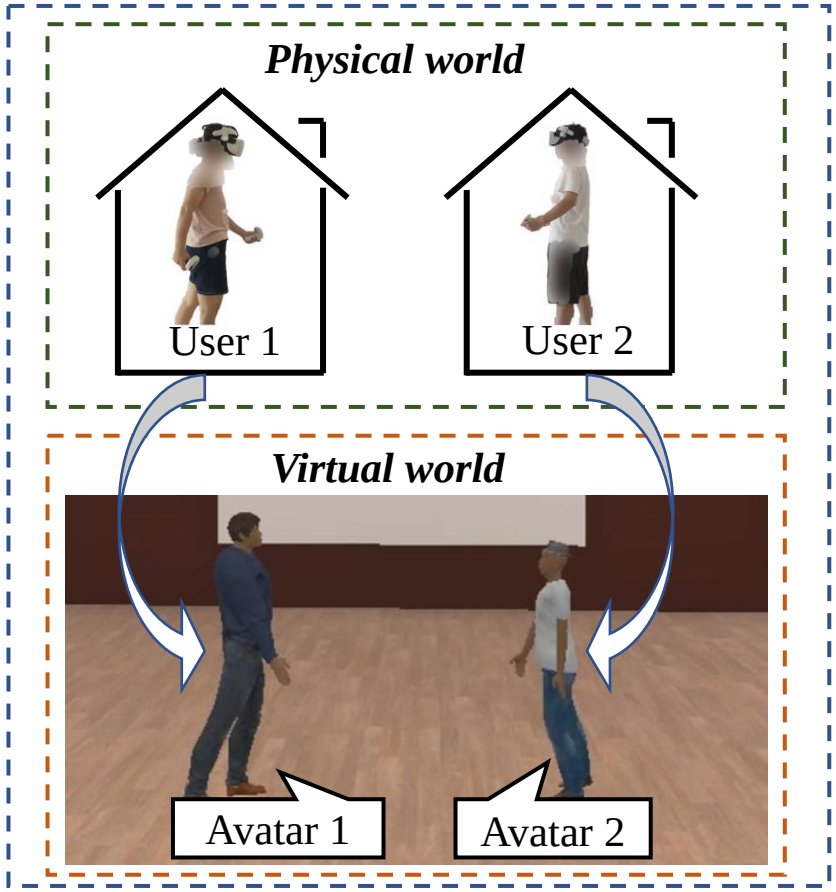


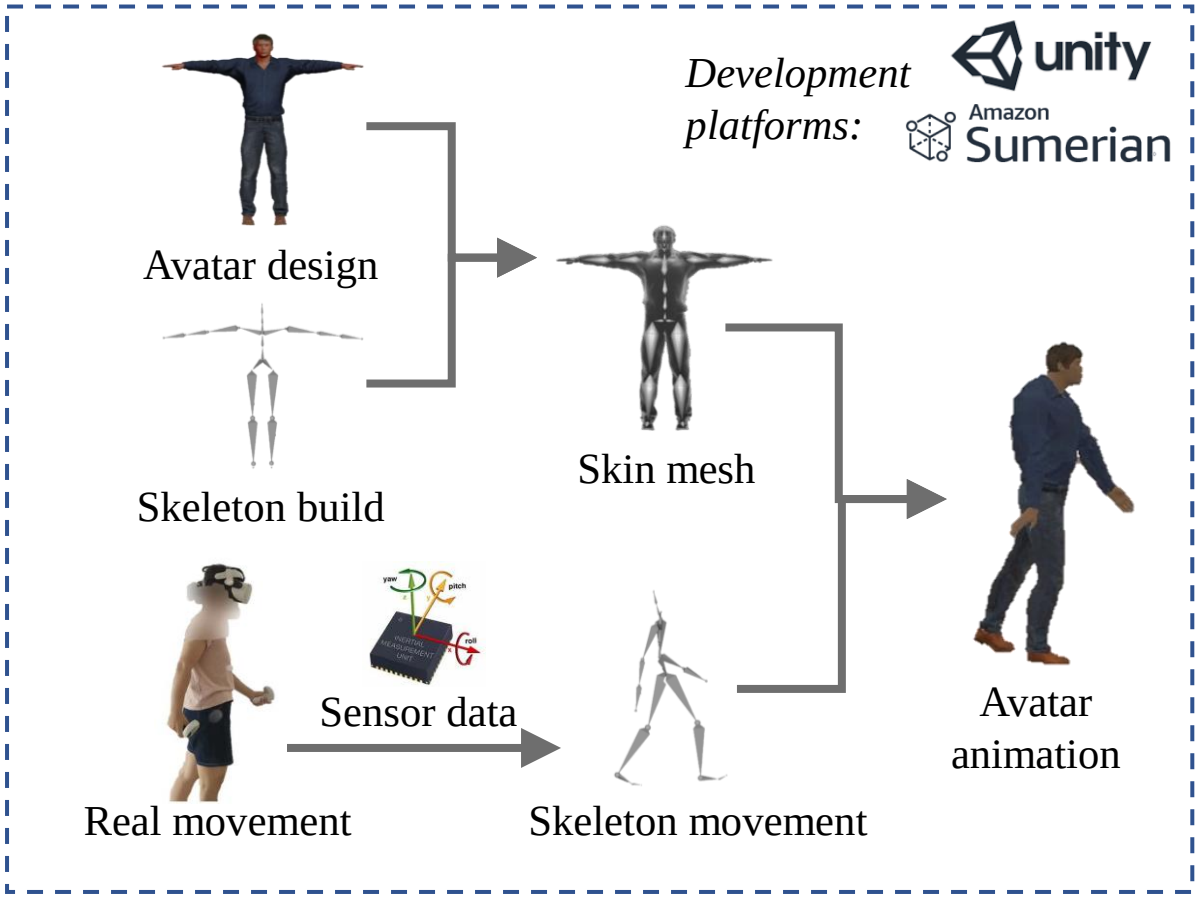
2

Motivation

Basic Insight

- Recall the avatar generation process.

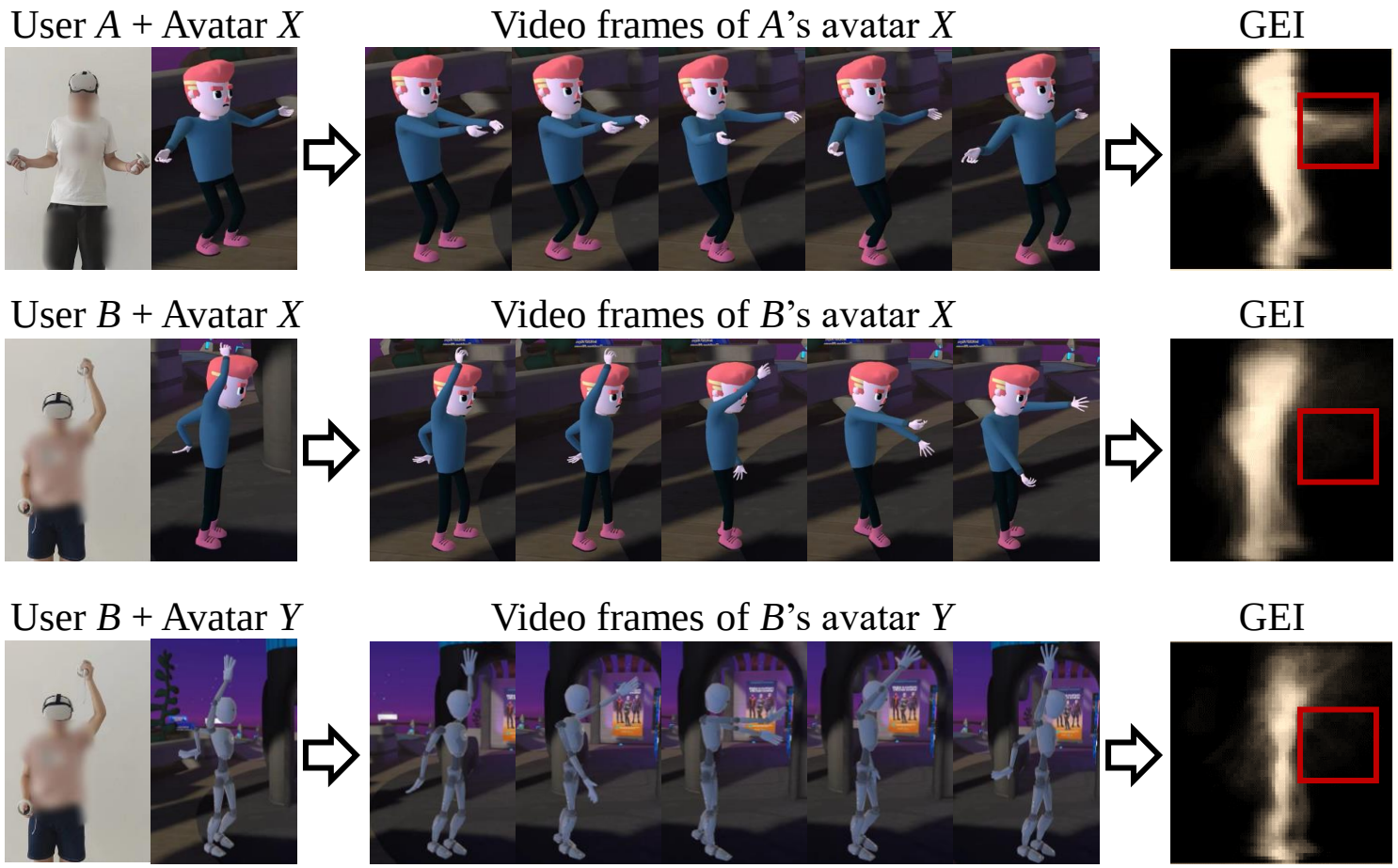




- No matter how avatars are changes, the inherent and unique movement patterns are relatively stable

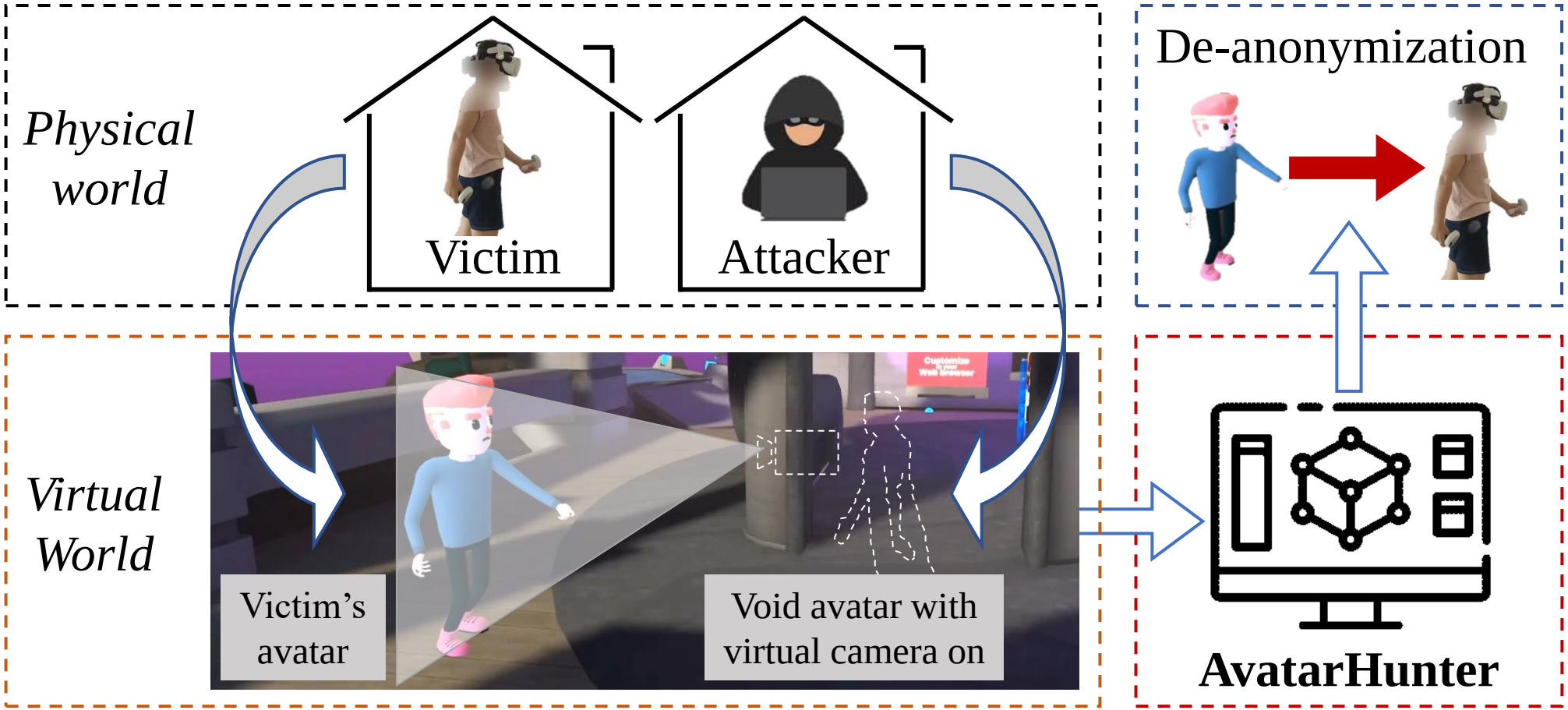
Validation of Insight

- Two users, same avatars → Different patterns → Different features
- Different avatars → Stable features



Attack model

- Our proposed attack system: AvatarHunter

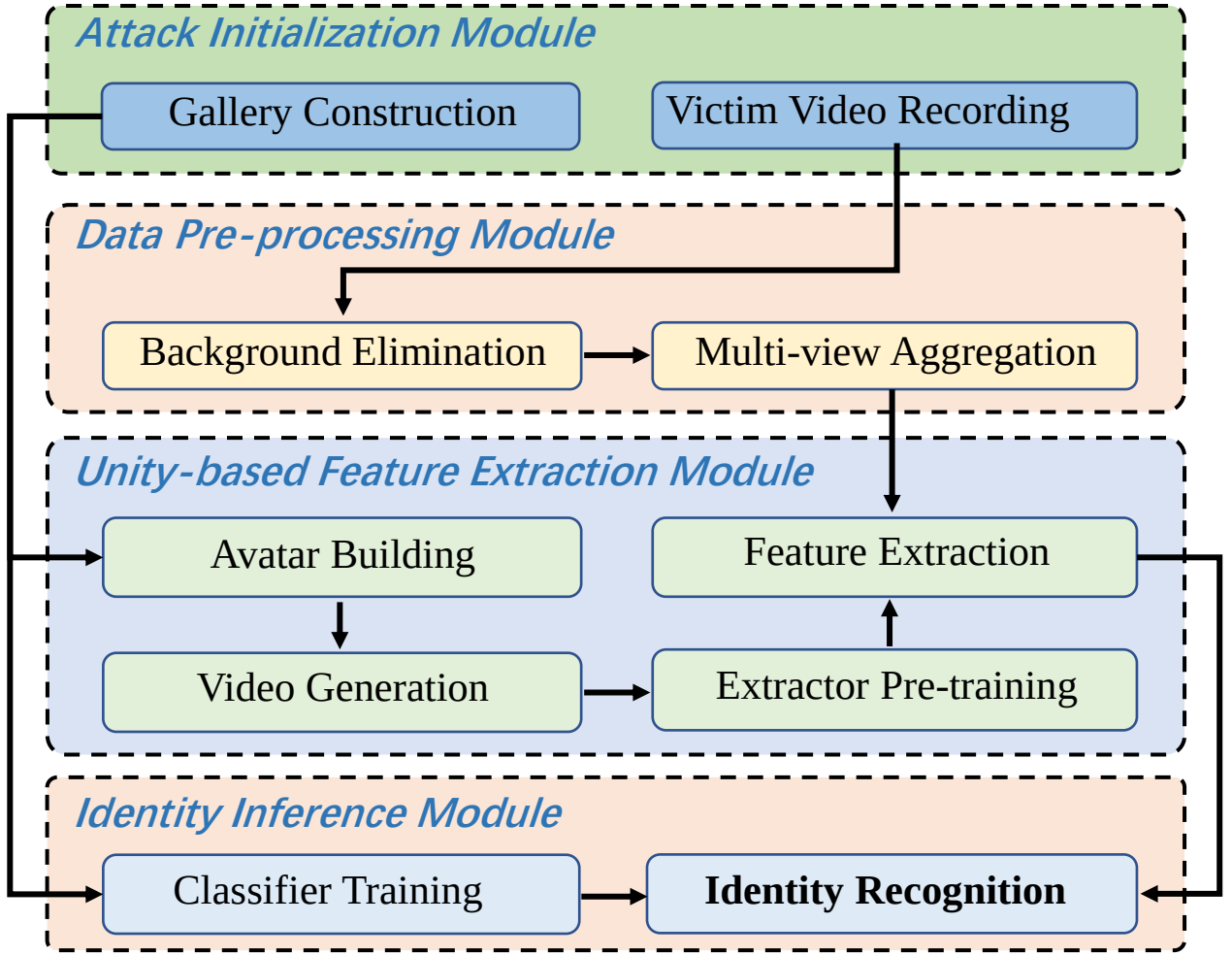


3

System Design

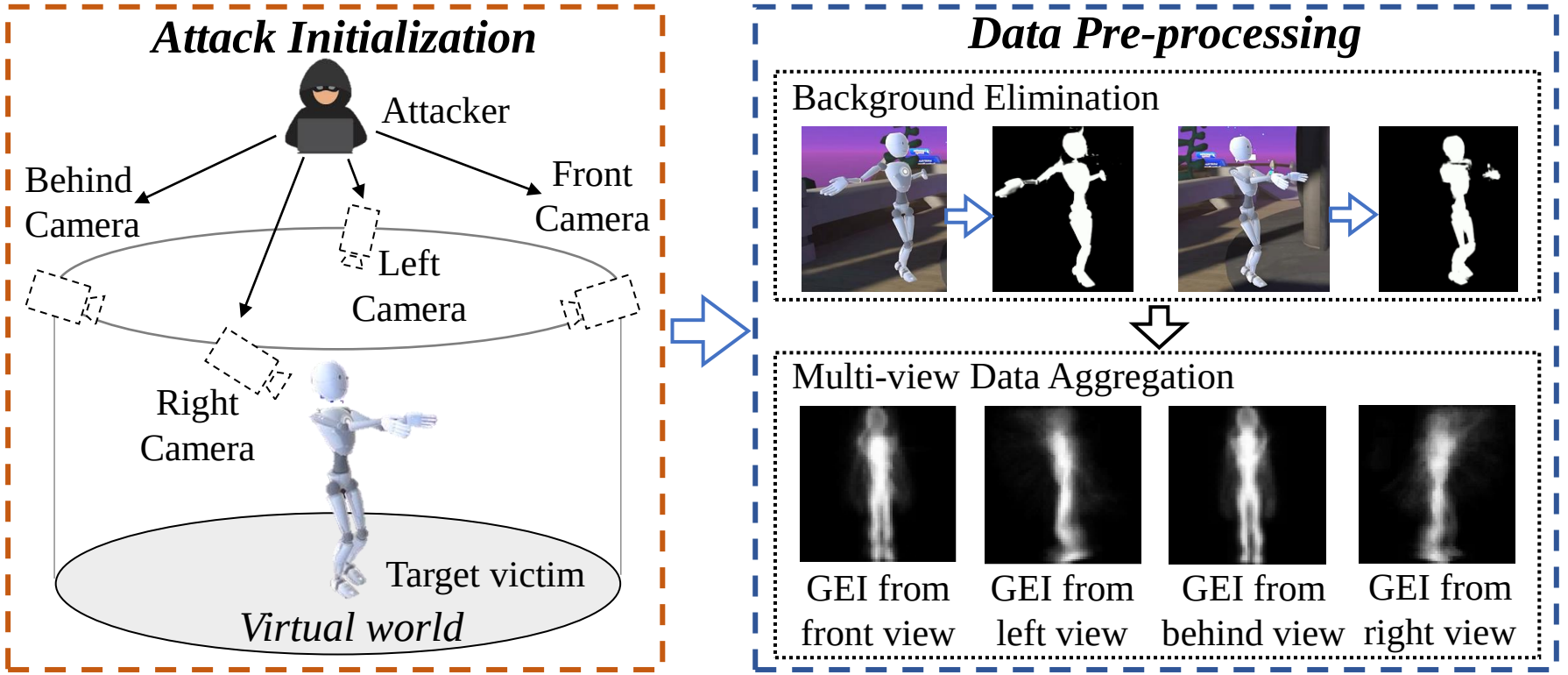
System Overflow of AvatarHunter

- AvatarHunter consists of four modules:
  - Attack Initialization
  - Data Pre-processing
  - Feature Extraction
  - Identity Inference



Attack Initialization & Pre-processing

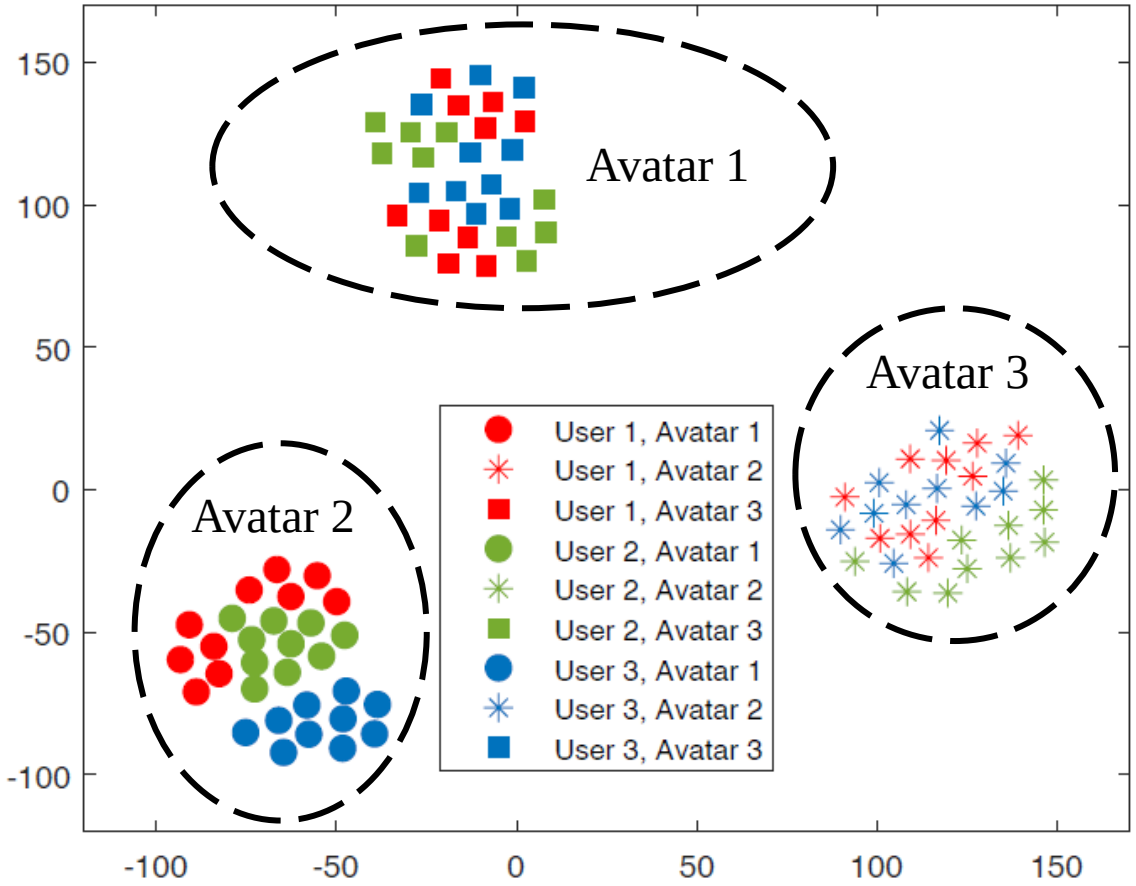
- Multiple-void camera
- Preprocessing
  - Background elimination
  - Silhouette aggregation



Unity-based Feature Extraction

Attack Initialization → Data Pre-processing → **Feature Extraction** → Identity Inference

- Firstly, directly using existing gait recognition solution (i.e., GaitSet [1] in this study)
- √ Appearance signature
- ✘ Movement signature

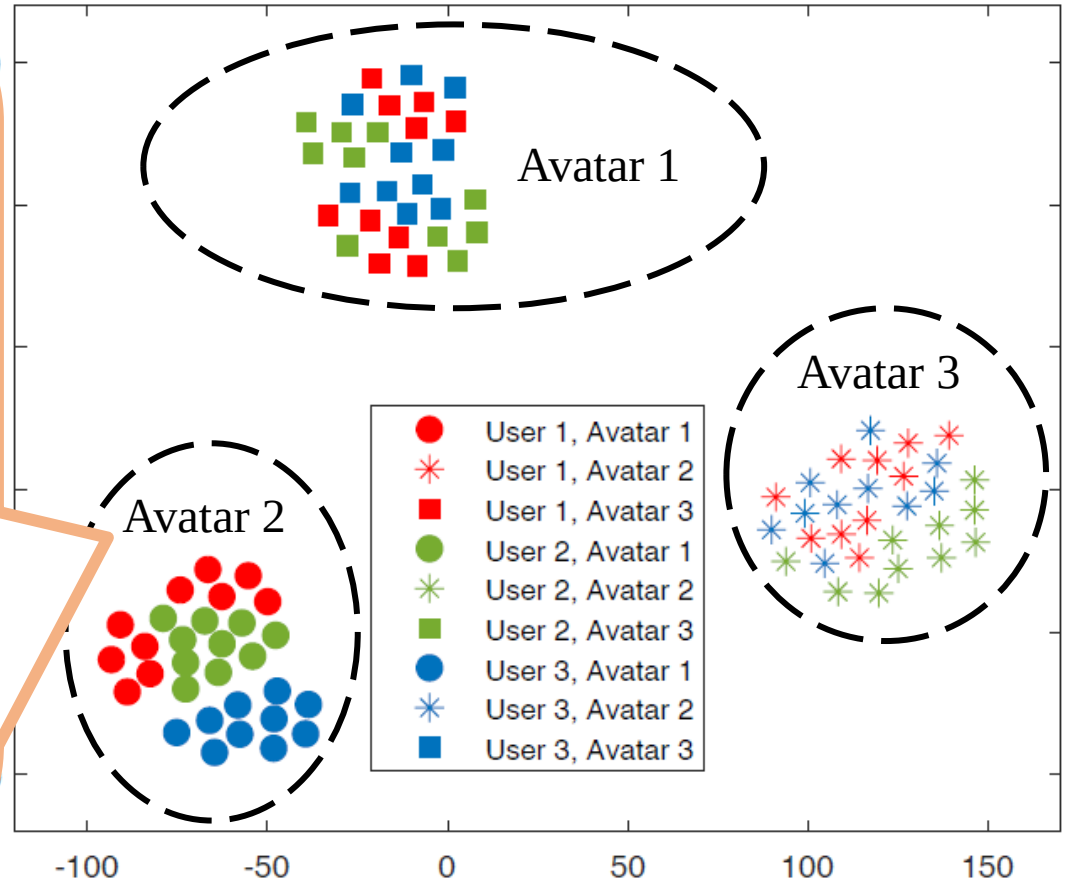


[1] Chao et al., "Gaitset: Cross-view gait recognition through utilizing gait as a deep set," IEEE TPAMI, vol. 44, no. 7, pp. 3467–3478, 2022.

Unity-based Feature Extraction

Attack Initialization | Data Pre-processing | **Feature Extraction** | Identity Inference

1. Reason: Existing feature extractors are trained using the videos from physical world.
2. *Not suitable for Avatar-changeable scenario.*
3. ***Needs to re-construct the feature extractor.***

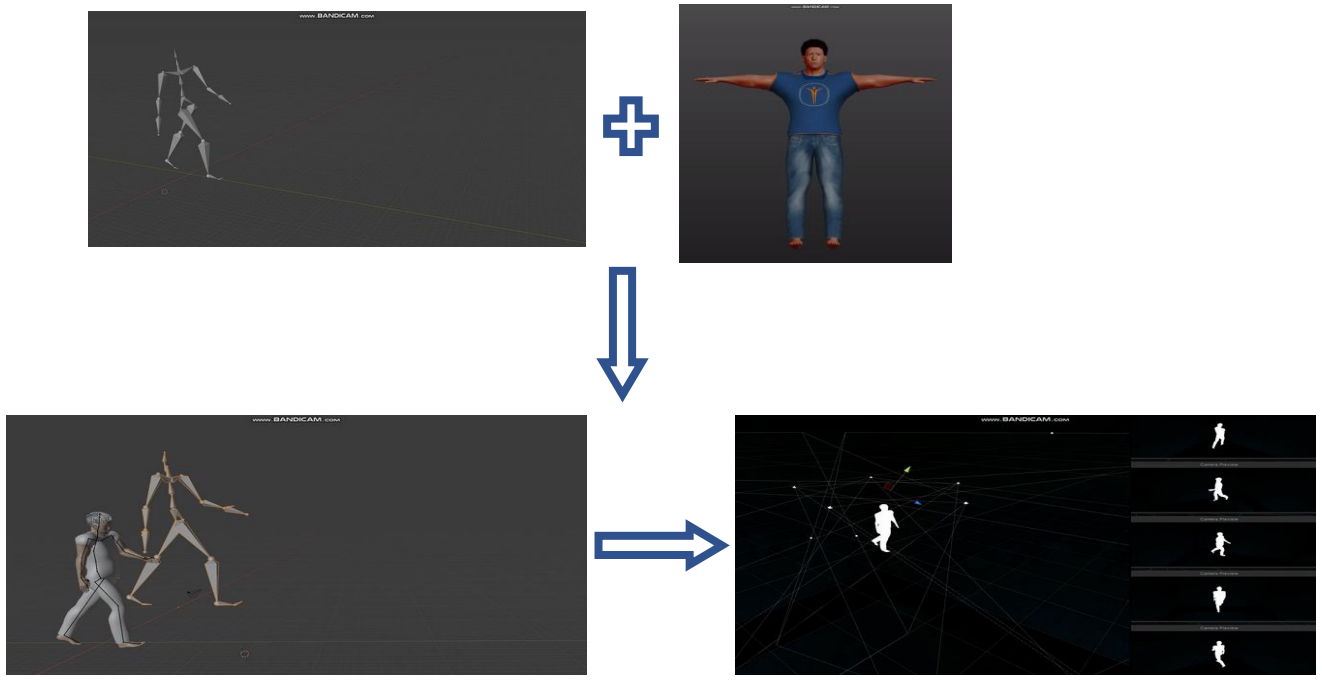


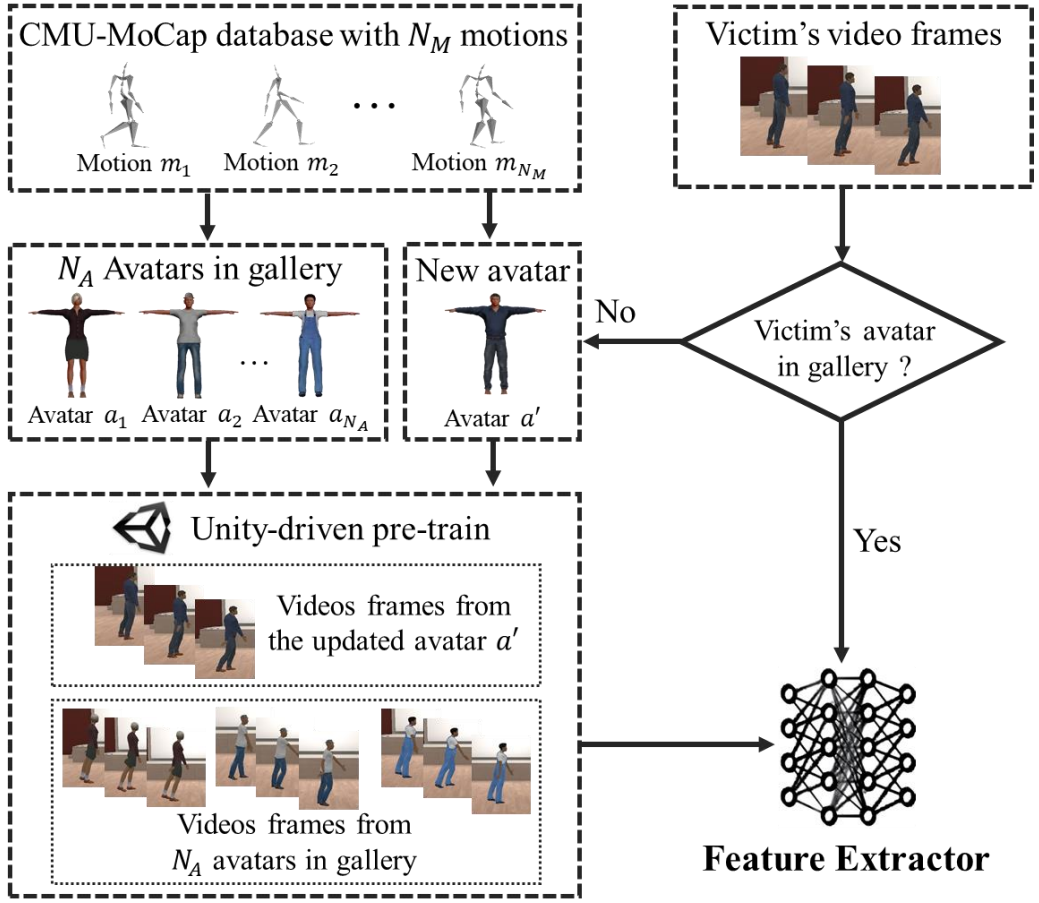
T-SNE of existing feature extractor

Unity-based Feature Extraction

Attack Initialization → Data Pre-processing → **Feature Extraction** → Identity Inference

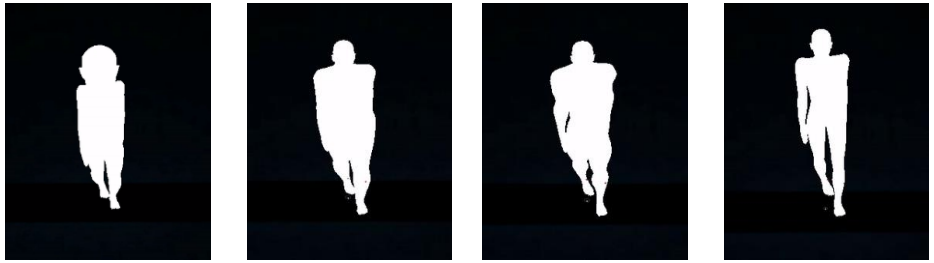
- Unity-based solution



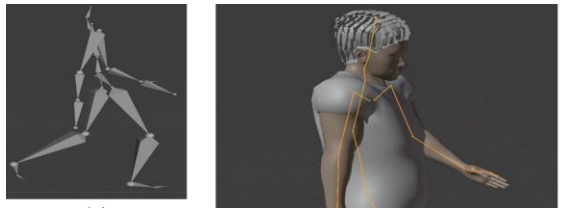


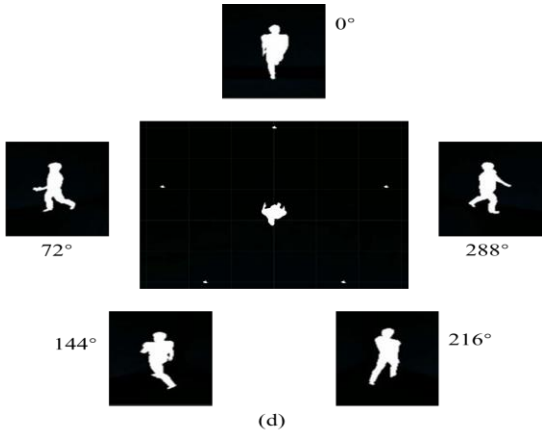
Unity-based Feature Extraction

- Unity-based solution
- Various avatars



Child Fat Strong Tall

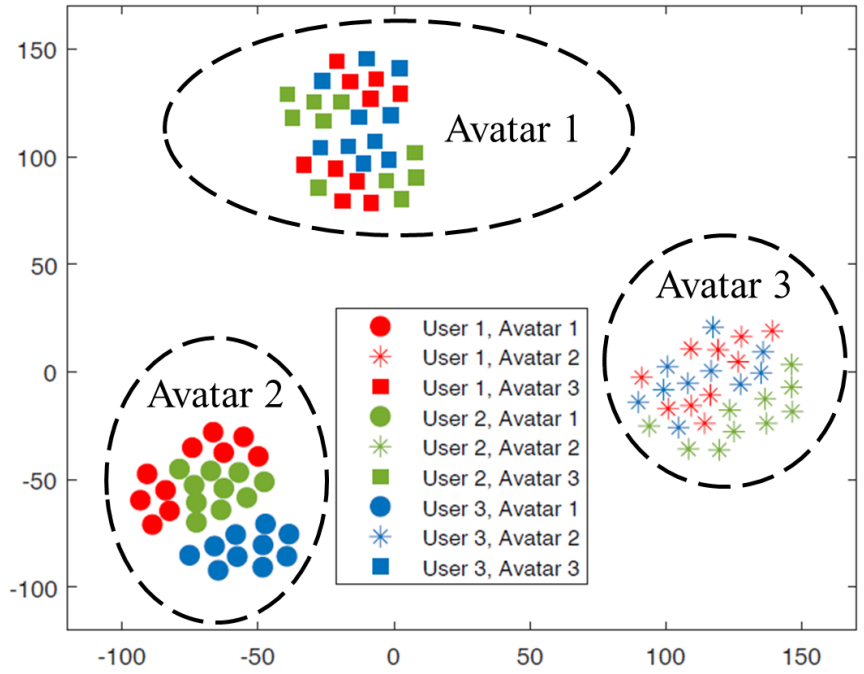




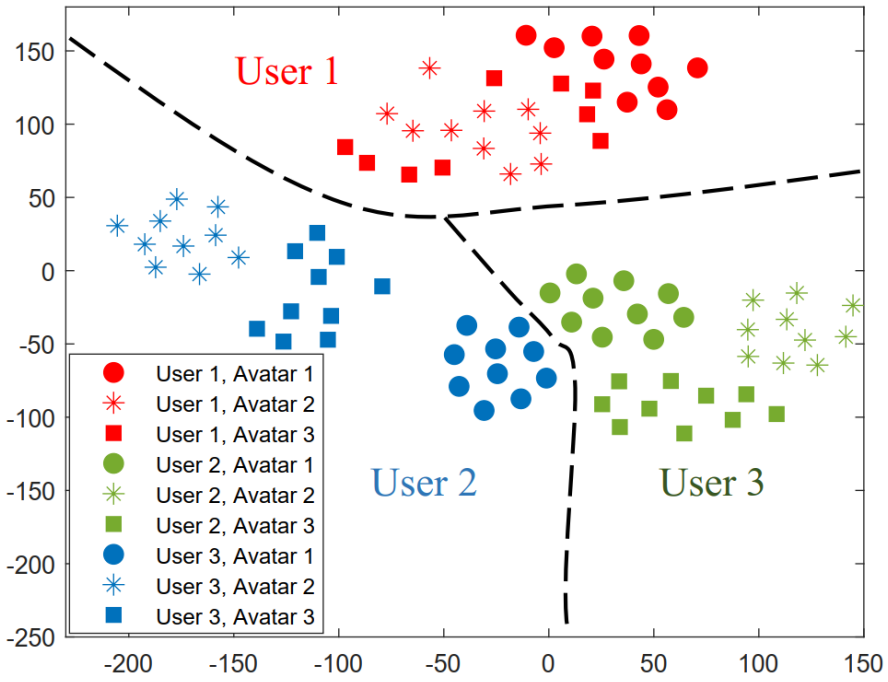
Unity-based Feature Extraction

Attack Initialization | Data Pre-processing | **Feature Extraction** | Identity Inference

- Performance Improvement
- From appearance feature to movement signature



GaitSet



Unity-based feature extractor

Identity Inference

- Classification model:
  - Pre-collected Gallery vs Test Datasets
  - Random Forest based method

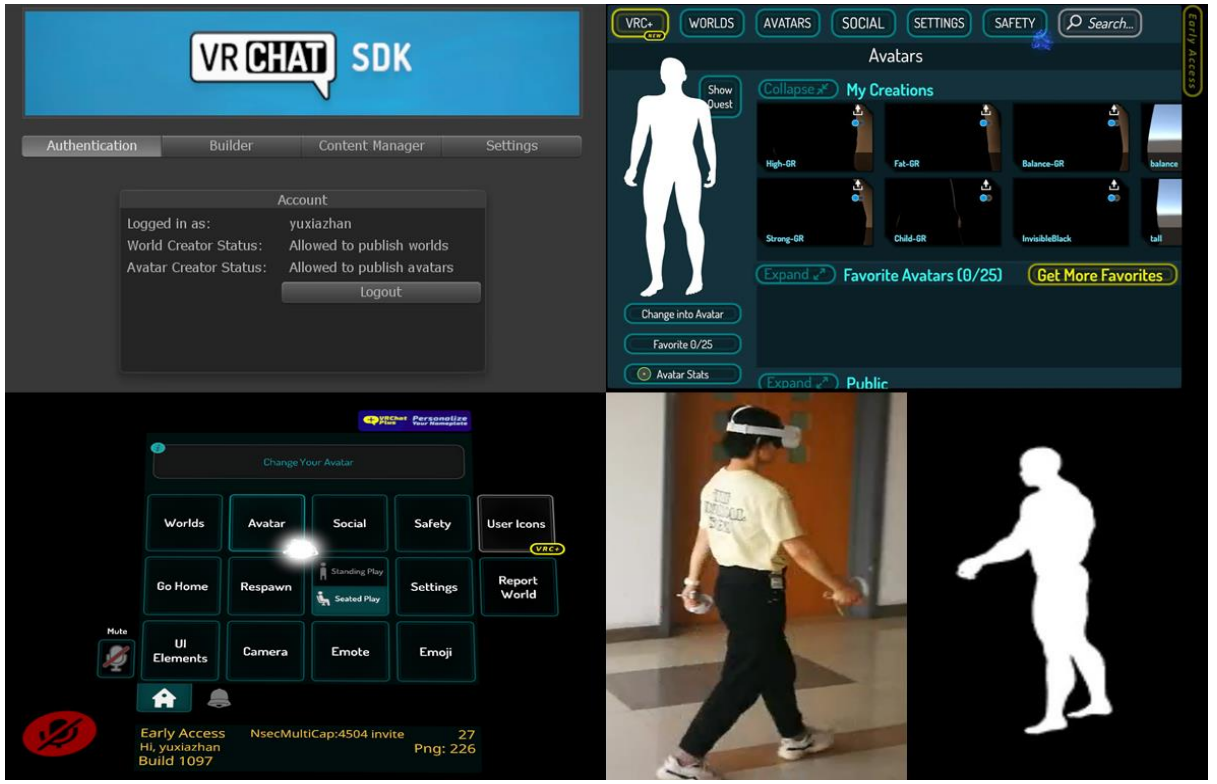
4

Evaluations

Dataset Construction

- Experimental Platform
  - VR device: Meta Oculus Quest 2
  - VR applications: VRChat[1]

Meta Quest 2

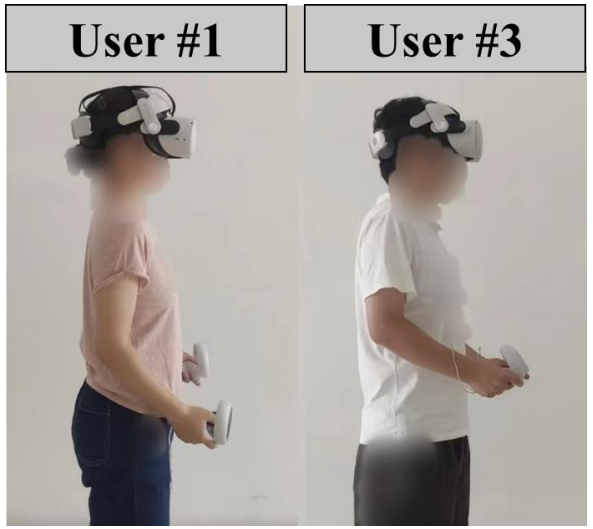


VRChat

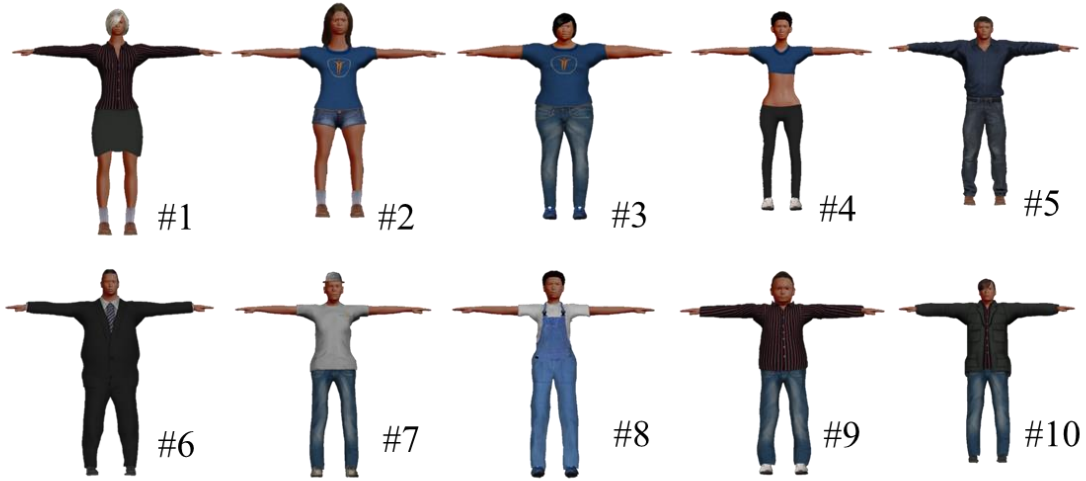
1. https://hello.vrchat.com/

Dataset Construction

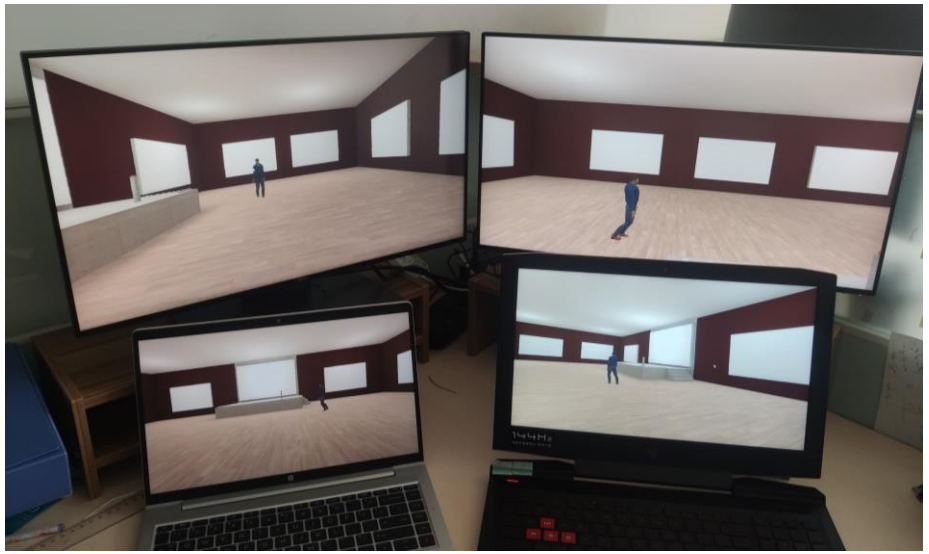
- 10 Users, 10 Avatars, 4 cameras
- Total 1000 trials (video clips)
- Gallery: Testing = 3:7



10 Users



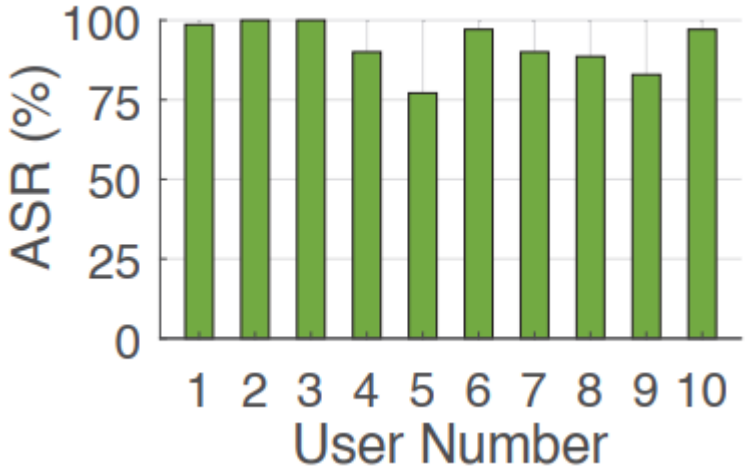
10 Avatars



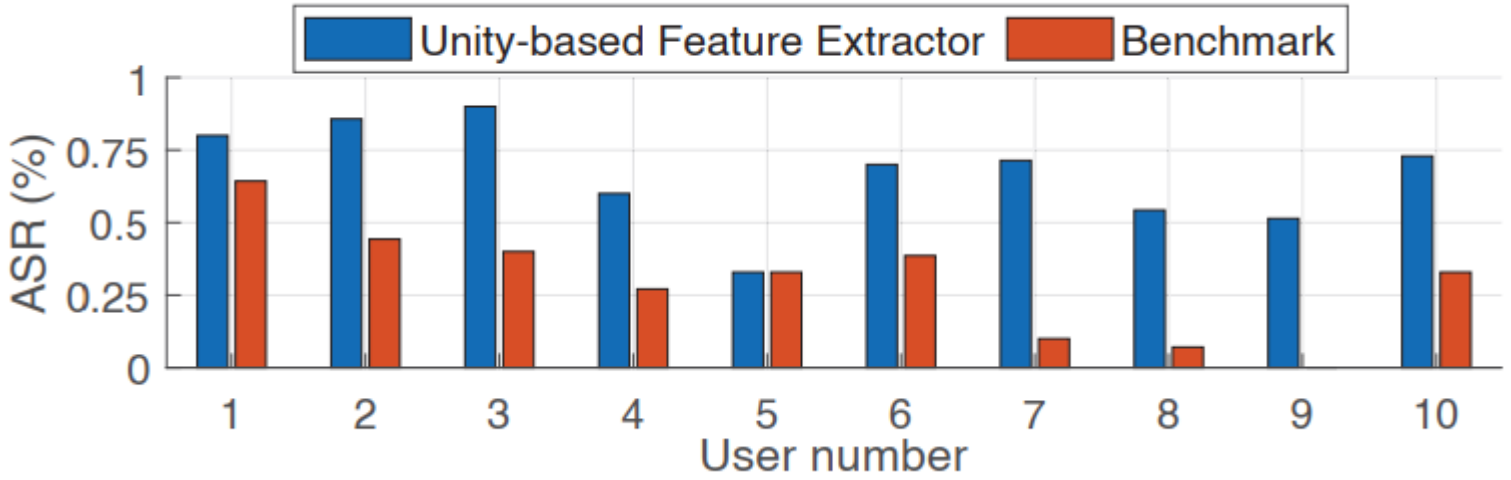
Views from 4 cameras

Overall Performance

- Closed-world avatar setting: 92.1%
- Open-world avatar setting: 66.9%
- AvatarHunter (66.9%) vs BenchMark (29.7%)



Closed-world scenario



Open-world scenario

Impact of Various Factors

- Length of Recording Video

Frame length	10	20	30	60	90
ASR in scenario-1 (%)	87.0	91.6	92.1	93.0	93.4
ASR in scenario-2 (%)	63.6	65.3	66.9	65.9	65.0

- Gallery Size

Gallery size	10%	30%	50%	70%	90%
ASR in scenario-1 (%)	78.8	92.1	94.6	94.0	98.0
ASR in scenario-2 (%)	63.0	66.9	64.0	62.7	66.0

- Camera Number

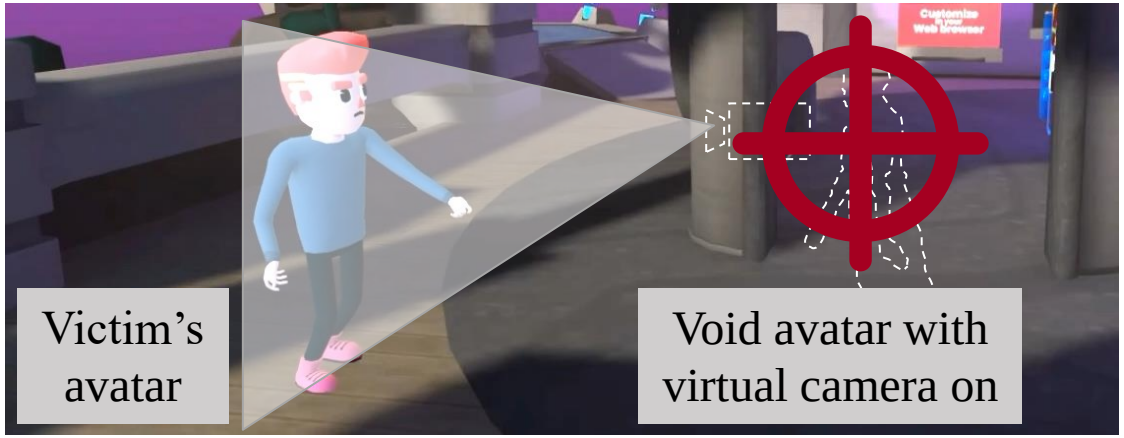
Camera combination	F	FB	FR	FRL	FBRL
ASR in scenario-1 (%)	72.9	90.1	91.9	88.4	92.1
ASR in scenario-2 (%)	51.3	62.3	68.6	66.4	66.9

5

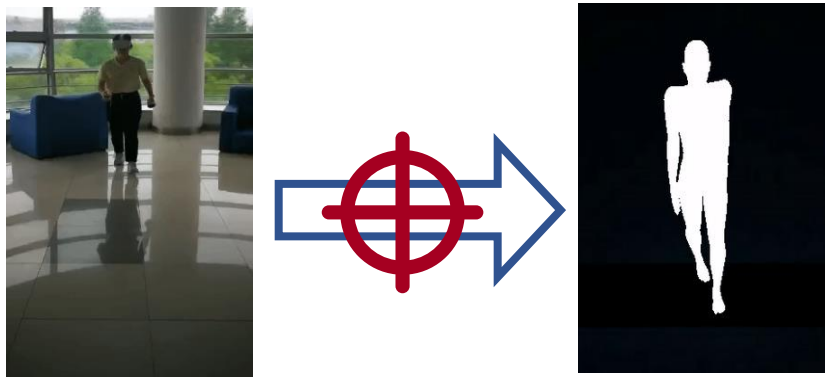
Discussions

Countermeasures

- Detecting Suspicious Users in VR



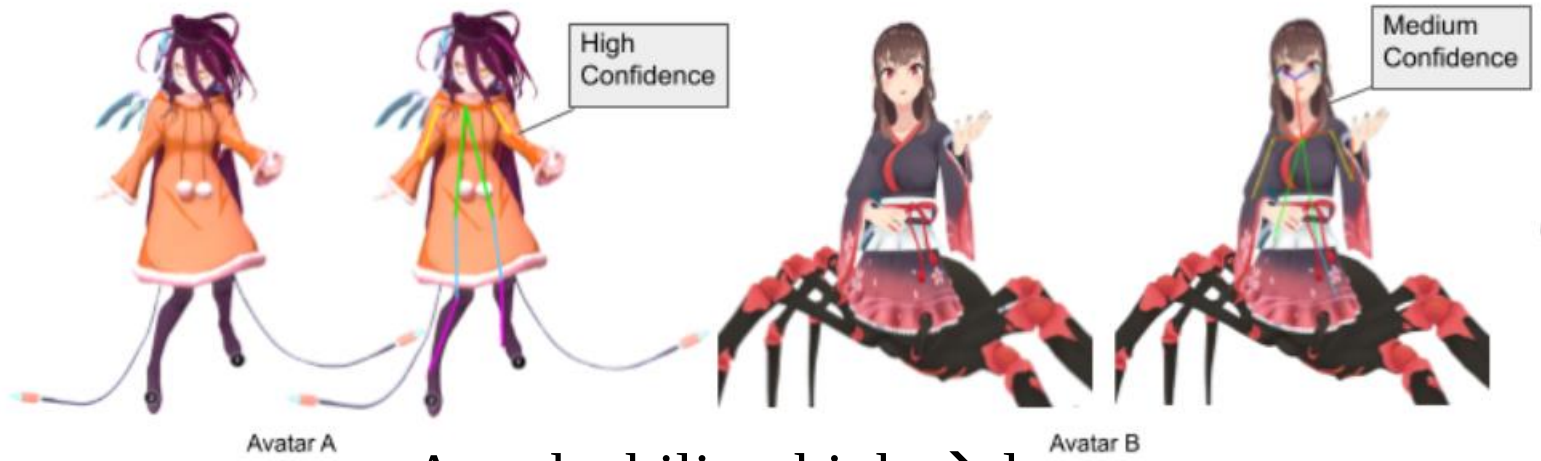
- Adding Noises during Avatar Generation



- Restricted access control (invite uses they trust)

Limitations

- Not Suitable for Non-humanoid Avatars



Attack ability: high → low

- Pre-collecting victim's information in VR

6

Conclusions

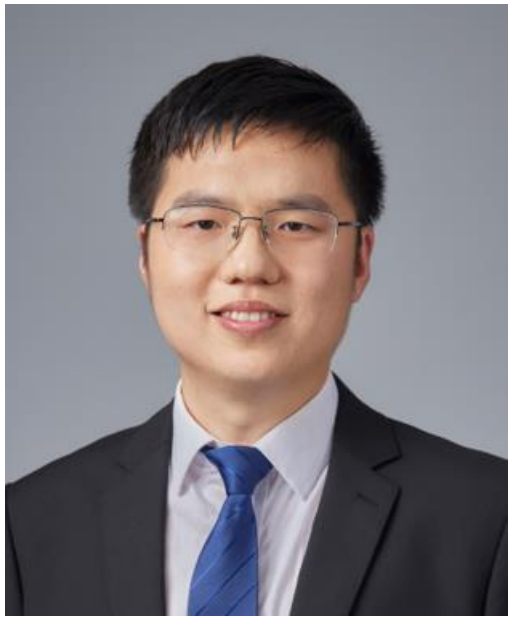
Conclusion

- Summary:
  - Propose AvatarHunter, a non-intrusive and user-unconscious de-anonymization attack in VR scenarios.
  - Leverage Unity-based feature extractor to characterize the victim's movement signature.
  - Proved to be effective and robust to various factors.

- Future Work
  - More universal attack scenarios.
  - Practical and useful countermeasures.

Conclusion

Thank you!



Yan Meng
Assistant Professor
Email: yan_meng@sjtu.edu.cn