Effect of Data Degradation on Motion Re-Identification

Vivek Nair
Computer Science
UC Berkeley
Berkeley, USA
vcn@berkeley.edu

Mark Roman Miller
Computer Science
Illinois Institute of Technology
Chicago, USA
mmiller30@iit.edu

Rui Wang
Carnegie Mellon University
Pittsburgh, USA
ruiwang3@andrew.cmu.edu

Brandon Huang
UC Berkeley
Berkeley, USA
zhaobin@berkeley.edu

Christian Rack
Human-Computer Interaction
University of Würzburg
Würzburg, Germany
christian.rack@uni-wuerzburg.de

Marc Erich Latoschik
Human-Computer Interaction
University of Würzburg
Würzburg, Germany
marc.latoschik@uni-wuerzburg.de

James F. O'Brien
Computer Science
UC Berkeley
Berkeley, USA
job@berkeley.edu

Abstract—The use of virtual and augmented reality devices is increasing, but these sensor-rich devices pose risks to privacy. The ability to track a user's motion and infer the identity or characteristics of the user poses a privacy risk that has received significant attention. Existing deep-network-based defenses against this risk, however, require significant amounts of training data and have not yet been shown to generalize beyond specific applications. In this work, we study the effect of signal degradation on identifiability, specifically through added noise, reduced framerate, reduced precision, and reduced dimensionality of the data. Our experiment shows that state-of-the-art identification attacks still achieve near-perfect accuracy for each of these degradations. This negative result demonstrates the difficulty of anonymizing this motion data and gives some justification to the existing data- and compute-intensive deep-network based methods.

Index Terms—privacy, virtual reality, motion data, signal degradation

I. INTRODUCTION

In recent years, both research, corporate, and consumer interest in virtual reality (VR) has been growing [1]–[3]. The growing adoption of VR devices raises the priority of understanding novel security and privacy threats possible with these devices' use. One class of threats is a privacy risk due to the collection of motion data, a kind of data that is fundamentally necessary for the operation of VR headsets. This data, usually consisting of the 3D position and orientation of the headset itself as well as the hand controllers, can be used to identify users [4]–[7]. On one hand, this behavioral biometric is a benefit to security, as device developers can verify the user of the device continuously and implicitly. On the other hand, our focus is the privacy risk of this motion data, either by malicious applications or by malicious other users in social VR.

Recent work has demonstrated inference [8], [9] and identifiability [4], [10], [11], even in massive pools of people [5]. This has, in turn, motivated defenses [12], [13]. However, these defenses are either very data-intensive, requiring thousands of recordings from hundreds of users, or they do not protect against sample-level data attacks. A simpler approach based on data degradation would be more desirable than a compute- and data-intensive approach, if it were effective.

In this work, we explore the possibility of defending against this identification attack using various methods of reducing the quality of the data. Specifically, we consider:

- *sample-level noise,* the addition of Gaussian noise to each dimension of motion data
- *reduced framerate,* the removal of intervening frames within recorded data by subsampling
- *reduced precision,* rounding the data to the nearest subdivisions of a meter, and
- *reduced dimensionality,* the reduction of the data available to the model to a single dimension.

However, we find identifying signals are robust to these types of degradation. This work demonstrates that the motion necessary to identify an individual is not limited to high-quality data or a controlled environment, but is robust in many situations. This robustness poses a greater challenge to privacy enhancing technologies.

II. RELATED WORK

To set the context for the current work, we first review identification using motion. Then, we review the current space of defenses against identification.

A. Identification using Motion

Because the identification of a user by their motion is both the goal of trusted authentication systems and identification attacks, we interleave these two domains within this review. The distinction we draw in this space is whether the entity with the data (attacker or authenticator) has access to motion data *only*, or whether they have access to some other aspect of the system. To give examples, these other aspects can be

the intent of the mover, e.g., asking the user to perform a specific action; presence within the mover's world, e.g., social engineering attacks like waving and eliciting a wave back; or design of the virtual environment, e.g., eliciting certain actions when playing a virtual game.

Notably, while the focus of this work is on privacy, there is an alignment between the present work and the identification of cooperative users. Most of these identification methods are types of *implicit authentication*, authentication based upon actions a user carries out for other reasons [14], and *continuous authentication*, in which users are authenticated several times throughout a session, addressing session hijacking [15]. The same finding - that a particular set of actions and features can be identifying - is a risk to privacy and a benefit to security and usability on these fronts.

In contrast to authentication, we are interested in adversarial identification, where a weaker adversary has access to only the motion data of the target. This kind of attack does not require the trust of the target, access to social interactions with the target [16], or access to the environment the target is in [17].

B. Defenses

In contrast to the attack space, there is less work about defense mechanisms for this data. M. Miller and collaborators [4] reduce the training data streams from 18DOF (head and hands position and rotation) to 3DOF (head rotation only) and reduce accuracy from 95% to 20% on a set of 511. Moore and collaborators [10] reduce accuracy from 89% to 32% on one set of data and 42% to 13% on a second by switching from position-based to velocity-based feature vectors. Nair, Gonzalo, and Song [12] use differential privacy methods on the biometric features they lay out in previous work [17]. Differential privacy methods incorporate a type of noise to each data point within a dataset so that even when the entirety of the dataset is compromised save for one point, the relative likelihood between the data point being the true value and the data point being any other value is bounded above by the privacy parameter. For a formal mathematical definition, see [18]. However, because the protection is only applied to a handful of hand-selected features (height, arm span, etc), the protection leaves many kinds of identifiers unchanged (e.g., degree to which a user looks around a space). There are other types of privacy guarantees, such as k-anonymity and plausible deniability. While to our knowledge these approaches have not been taken on sample-level data, there has been work on protecting eye-tracking data [19].

Leveraging more advanced techniques, Nair and collaborators also have proposed *Deep Motion Masking* [13], which breaks down motion using LSTMs into the variance due to the *action* and variance due to the *user* - in essence, subtracting out the idiosyncrasies of the individual before transmitting motion data. This reduces the identifiability of motions stream while maintaining plausibly-human behavior. However, this work requires significant data and compute power and may not extend to out-of-distribution actions. This motivates us to explore the potential effectiveness of simpler methods to de-identify data, specifically data obfuscation or degradation.

One work that is most similar to the approach here is work by Hanisch and collaborators [20] who investigate the identifiability of gait subject to perturbations, coarsening, removal of data, and normalization. They ultimately conclude that gait anonymization is highly challenging, given their results that the anonymization techniques, for the most part, did not reduce accuracy. In contrast, we investigate Beat Saber, which has a significantly different macro-level structure (see the Data section for more information) and has only recently been discovered to be identifiable motion.

III. METHODS

A. Threat model

It is important to establish the kind of threat under study. In previous work on VR and identifiability, there are two dimensions upon which researchers have categorized threats. First, there is the question of what data is available to the attacker. Nair and collaborators [17] delineate between hardware-level attackers that have access to firmware, client-level attackers that have access to the headset APIs, server-level attackers that have access to the telemetry data sent to the servers and 'unprivileged user' attacker which is another VR system partaking in the same social virtual world. Along this dimension, we focus on the unprivileged user.

The second aspect of space of threat models is the capability of the attacker to influence the behavior of the participant, and the extent to which this can be done. For example, is the attacker designing a virtual world [17], are they another user that is interacting with the target [16], or do they wish not to interact with the target entirely? In our work, we focus on no interaction at all. This may occur because the attacker is working with previously-collected data, does not want to be vulnerable in the virtual world, or has data collected at scale and cannot interact with each target.

Per the framework of Garrido et al. [21], the adversary of interest to us is the "user adversary." This threat actor is selected because it is the least privileged attacker. Therefore, findings based on this work are likely to be applicable to all attacks leveraging VR pose tracking data. It also sets a baseline on threat for all these other conditions. Finally, there are some cases in which this may be the mode of an attacker, e.g., large-scale surveillance where individuals are not queried directly, re-identification attacks where actions are stored for a period of time before being queried, or any other situations in which the attacker does not which to have any direct interaction with the target. Note that this threat model is quite different from the traditional authentication threat model in which a user attempts to gain unauthorized access by posing as another user.

B. Data

The data used in this work comes from the Berkeley Open Extended Reality Recordings 2023 (BOXRR-23) dataset [22]. BOXRR-23 consists of 4.7 million motion capture recordings from 105,852 users, derived from "Beat Saber," a popular

virtual reality rhythm game, and "Tilt Brush," a virtual reality drawing application. In this paper, we only use the 500 users with the most recordings from the BOXRR-23 dataset. For each of these users, at least 500 separate recordings are present, with sessions varying in length. After sorting the recordings chronologically, the first 400 recordings per user are used for training, the next 50 are used for validation, and the final 50 are used for testing.

C. Feature Engineering

The registration of a coordinate system is often not amenable to moving, flexible, and diverse human bodies. Over time, different coordinate systems have been developed for specific purposes, such as the anatomical planes (coronal, sagittal, transverse) for medical terminology. For the purposes of our work specifically and of VR more generally, we use a coordinate system that synthesizes the global vertical axis with horizontal axes relative to the headset's forward direction, known as *body-relative coordinates* [11], [23].

To perform this normalization, the forward direction of the head (headset) is projected onto the horizontal plane. The transformation applied to all tracked objects (left hand controller, right hand controller) is the inverse rotation about the vertical axis so that the projected forward direction of the head aligns with the forward direction of the coordinate system. In regards to the question at hand, this would mean the body-space coordinate system is likely to be more effective at separating one's pose from another's than the global coordinate system would be. The use of body-relative coordinates for VR identification models is equivalent to that proposed by Rack et al. [11] and is enabled by the Motion Learning Toolkit [24].

For the features, at each frame processed by the VR device, the position and orientation of the user's left hand, right hand, and head are captured. Three positional coordinates and four orientation coordinates (in quaternion format) are captured for each of the three tracked objects, totaling 21 dimensions captured per frame. After the body-relative transformation is applied, 18 dimensions remain, as the three positional coordinates of the head are eliminated by this transformation. The vertical rotation of the head is also eliminated, but the quaternion representation of the rotation retains use of all four dimensions. The values of interest to us are the first and second derivatives of these 18 values; the result is 36 values per frame describing *body-relative velocity* and *body-relative acceleration*.

Each user's VR device may render frames at a slightly different frequency due to a variety of external factors. To eliminate frame rate as a potential confounding factor, we first resample all motion capture streams to a constant 30 frames per second by using a linear interpolation for positional coordinates and a spherical linear interpolation for orientation quaternions. Each session of a user was then split into 30-second sequences. In summary, an individual sequence has 30 seconds, 30 frames a second, and 36 values per frame; thus, our model has an input shape of (900×36).

D. Model

The model's task is to identify a user based upon their motion. More formally, the model is given a (900×36) sequence as described above. With that sequence, the model attempts to predict the participant who generated that motion, represented as a value of a categorical variable encoded with a one-hot encoding.

The model we have selected is a Long Short-Term Memory (LSTM) model [25], implemented in Python version 3.10.2 using Keras version 2.10.1. The choice of LSTM was to take advantage of the sequential nature of the data. Most hyperparameters for the model were left to the defaults; in particular, the Adam optimizer [26] was used with a learning rate of 0.001. Specifically, we utilize the "LSTM Funnel" architecture described by Nair et al. [13].

The predictions were made per session by taking the entire session of pose tracking data, computing 30-second sequences as described above, and then summing the logarithmic probability of each user reported by the model across all samples. We interpreted this distribution as a probability estimation for the classification of the session as a whole, in line with previous work [4].

The problem type is classification rather than a ranking problem, along the lines of similar work [4], [10], [27]. This method can be contrasted with multiclass AUC, which increases with any improvement in identifiability, not just when a sample is correctly classified.

IV. RESULTS

We study the identifiability of motion data alone through four kinds of methods. As described below, we degraded the quality of the motion data signal in a variety of ways to evaluate what effect, if any, this had on the identification accuracy.

A. Added Noise

First, we attempted to thwart the identification models by introducing random noise to each dimension of the motion telemetry stream. Specifically, we added zero-centered Gaussian noise with increasingly large standard deviations (σ) ranging from 0.1 to 5.0. This range represents a spread of values that interpolate between small, perceptible changes (10cm) to unrealistically large (5m). The results of this process are shown in Figure 1. We found that a per-user identification accuracy of 100% can be achieved with noise as high as $\sigma = 2.0$, and with over 90% accuracy even when $\sigma = 5.0$.

B. Reduced FPS

Next, we attempted to reduce the frame rate from the baseline of 30 FPS to as low as 1 FPS. The method of reduction was direct subsampling (rather than interpolation). Interestingly, downsampling motion data to 15 FPS and 10 FPS resulted in almost no reduction in per-sample or per-user accuracy. Further reductions to 5 FPS, 3 FPS, and 1 FPS resulted in degraded per-sample accuracy, but 100% per-user accuracy was still achieved at just 1 FPS, as shown in Figure 2.
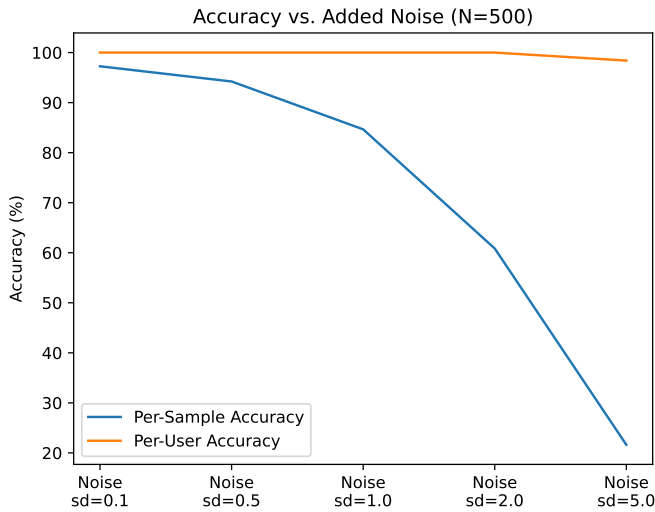
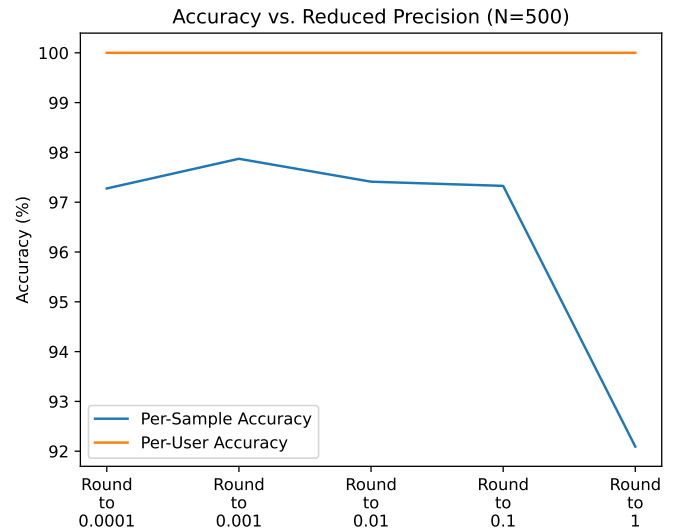Fig. 1. Cross-session identification accuracy with increasing Gaussian noise added to the telemetry signal

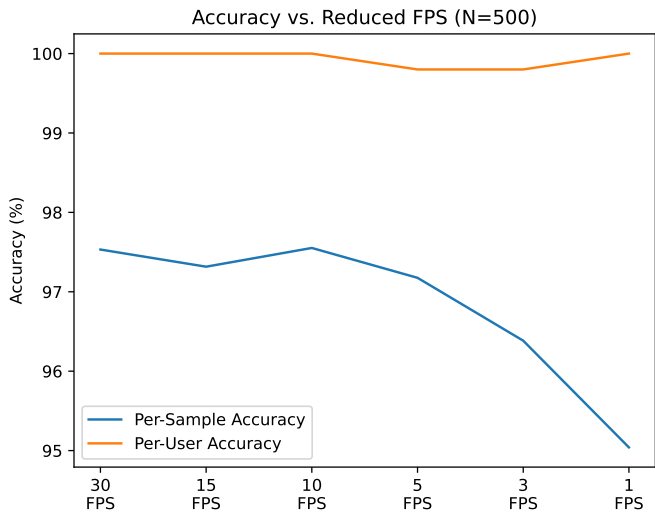Fig. 3. VR identification accuracy with rounded signal

Fig. 2. VR identification accuracy with reduced FPS

C. Reduced Precision

We also attempted to reduce the precision of VR motion data in the spatial (rather than temporal) dimensions. We did so by rounding all values in the VR motion data to the nearest 0.0001, 0.001, 0.01, 0.1, and 1 meters. The results, shown in Figure 3, show minimal impact on accuracy for all degrees of rounding between 0.0001 and 0.1. Rounding all dimensions to the nearest full meter did have a significant impact on per-sample accuracy, but still allowed 100% per-user accuracy to be achieved.

D. Reduced Dimensions

Finally, we attempted to reduce the dimensionality of the motion data. We began by eliminating the dimensions associated with the users' heads, leaving only their hands. Next, we eliminated all positional dimensions, leaving only the

rotations of the users' hands. Further, we eliminated the users' right hands, leaving only their left hand rotations. Finally, we eliminated the i, j, and k quaternion elements, leaving only the w coordinate of the quaternion, which corresponds to rotational magnitude. The results of these reductions are shown in Figure 4; each dimensionality reduction accompanied a corresponding drop in per-sample accuracy, with 100% per-user accuracy still being observed.
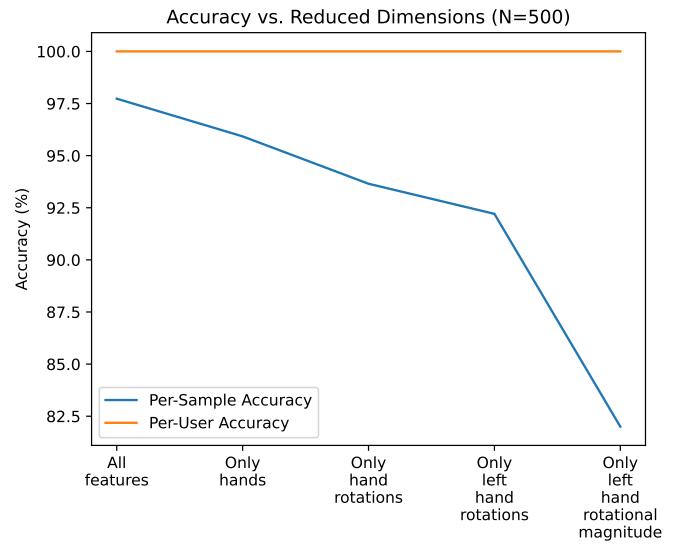

Fig. 4. VR identification accuracy with reduced dimensions

V. DISCUSSION

A. Summary of Results

We performed several methods of signal degradation on VR motion data. In each case, whether Gaussian noise is added to the data, the framerate is reduced, the values are rounded,

or only one dimension is tracked, identification is still very high, leading to an accuracy of 100% at the session level. This illustrates that given enough data, it does not take ideal conditions for motion data to be identifying. Rather, based upon these results, it would seem a user is more likely to run into user experience issues through the obfuscation than to successfully anonymize the data through signal degradation. For example, negotiating personal space in a social VR setting would be very difficult to do if other avatar positions are only updated once per second.

B. Implications for Privacy

First, the degraded signal results further dispel the myth that motion-based identification in VR requires high-quality data or a controlled laboratory environment. In fact, we show that a low-quality device or network will still be good enough to identify VR because the signal - whatever is identifying within the motion - is very robust. In short, you don't need many conditions for identification to be feasible, just a lot of data.

Considering virtual reality motion identification as a whole, there are several steps to take. First, developers should protect this data with standard practices for personally identifying data [1]. When this data needs to be shared with others, it can be helpful to reduce the time span available, minimize variation in activities, or modify data to produce security guarantees like k-anonymity, plausible deniability, or differential privacy [12], [19]. Furthermore, there are developments in law that need to be made to clarify the legal status of this data based on its risks to privacy [28].

C. Limitations and Future Work

Some limitations of this work are that the manipulations are not applied together, e.g., there is no combination of added noise and reduced FPS. Additionally, this model is only trained on Beat Saber data, and while this application was not specially selected for its identifiability, it remains to be seen how this identifiability and potential defense extends beyond a single application type.

While some work [12] weakens the relationship between real-world biometrics like height and arm length from a user's virtual avatar, it may be plausible, given a user's preferences, to disconnect those two entirely. Care must be taken in this approach, though, as normalization may make other idiosyncrasies more prominent [29]. Another approach to avoid this tradeoff is to use transformed social interaction [30] so that gestures that might otherwise be identifiable can come from another recorded value but still be communicative.

VI. CONCLUSION

In this work, we test the effectiveness of signal degradation against state-of-the-art re-identification methods. Despite employing various degradation methods such as adding noise, reducing framerate, precision, or dimensionality, we found that identification accuracy remained remarkably high, almost always achieving 100% at the session level. This result underscores the robustness of re-identification attacks based upon motion data. If simpler privacy protection methods are effective, they need to extend beyond these kinds of signal degradation; if simpler methods are not available, then the evidence can show the necessity of more complex machine learning to protect privacy. We hope this negative result can further define the boundaries between and aid future research in protecting privacy in VR.

REFERENCES

[1] Y. Wang, Z. Su, N. Zhang, R. Xing, D. Liu, T. H. Luan, and X. Shen, "A survey on metaverse: Fundamentals, security, and privacy," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 319–352, 2023.

[2] B. Falchuk, S. Loeb, and R. Neff, "The Social Metaverse: Battle for Privacy," *IEEE Technology and Society Magazine*, vol. 37, no. 2, pp. 52–61, Jun. 2018, conference Name: IEEE Technology and Society Magazine.

[3] R. Di Pietro and S. Cresci, "Metaverse: Security and Privacy Issues," in *2021 Third IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, Dec. 2021, pp. 281–288.

[4] M. R. Miller, F. Herrera, H. Jun, J. A. Landay, and J. N. Bailenson, "Personal identifiability of user tracking data during observation of 360-degree VR video," *Scientific Reports*, vol. 10, no. 1, pp. 17 404–17 413, 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-74486-y

[5] V. Nair, W. Guo, J. Mattern, R. Wang, J. F. O'Brien, L. Rosenberg, and D. Song, "Unique Identification of 50,000+ Virtual Reality Users from Head & Hand Motion Data," Feb. 2023, arXiv:2302.08927 [cs]. [Online]. Available: http://arxiv.org/abs/2302.08927

[6] R. Miller, N. K. Banerjee, and S. Banerjee, "Using siamese neural networks to perform cross-system behavioral authentication in virtual reality," *Proceedings - 2021 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2021*, pp. 140–149, 2021.

[7] ——, "Temporal Effects in Motion Behavior for Virtual Reality (VR) Biometrics," *Proceedings - 2022 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2022*, pp. 563–572, 2022.

[8] M. R. Miller, E. Han, C. DeVeaux, E. Jones, R. Chen, and J. N. Bailenson, "A Large-Scale Study of Personal Identifiability of Virtual Reality Motion Over Time," Mar. 2023, arXiv:2303.01430 [cs]. [Online]. Available: http://arxiv.org/abs/2303.01430

[9] V. Nair, C. Rack, W. Guo, R. Wang, S. Li, B. Huang, A. Cull, J. F. O'Brien, L. Rosenberg, and D. Song, "Inferring Private Personal Attributes of Virtual Reality Users from Head and Hand Motion Data," Jun. 2023, arXiv:2305.19198 [cs]. [Online]. Available: http://arxiv.org/abs/2305.19198

[10] A. G. Moore, R. P. McMahan, H. Dong, and N. Ruozzi, "Personal identifiability and obfuscation of user tracking data from VR training sessions," *Proceedings - 2021 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2021*, pp. 221–228, 2021.

[11] C. Rack, K. Kobs, T. Fernando, A. Hotho, and M. E. Latoschik, "Versatile user identification in extended reality using pretrained similarity-learning," Feb. 2023, arXiv:2302.07517 [cs]. [Online]. Available: http://arxiv.org/abs/2302.07517

[12] V. Nair, G. M. Garrido, and D. Song, "Going Incognito in the Metaverse," *ArXiv*, 2022. [Online]. Available: http://arxiv.org/abs/2208.05604

[13] V. Nair, W. Guo, J. F. O'Brien, L. Rosenberg, and D. Song, "Deep motion masking for secure, usable, and scalable real-time anonymization of virtual reality motion data," 2023.

[14] M. Jakobsson, E. Shi, P. Golle, R. Chow *et al.*, "Implicit authentication for mobile devices," in *Proceedings of the 4th USENIX conference on Hot topics in security*, vol. 1. USENIX Association, 2009, pp. 25–27.

[15] I. Traoré and A. A. E. Ahmed, "Introduction to continuous authentication," in *Continuous Authentication Using Biometrics*. IGI Global, 2012, pp. 1–22. [Online]. Available: https://doi.org/10.4018/978-1-61350-129-0.ch001

[16] B. Falk, Y. Meng, Y. Zhan, and H. Zhu, "POSTER: ReAvatar: Virtual Reality De-anonymization Attack through Correlating Movement Signatures," *Proceedings of the ACM Conference on Computer and Communications Security*, pp. 2405–2407, 2021.

[17] V. Nair, G. M. Garrido, D. Song, and J. F. O'Brien, "Exploring the privacy risks of adversarial VR game design," *Proc. Priv. Enhancing Technol.*, vol. 2023, no. 4, pp. 238–256, 2023. [Online]. Available: https://doi.org/10.56553/popets-2023-0108

[18] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3, pp. 211–407, 2013. [Online]. Available: http://www.nowpublishers.com/articles/foundations-and-trends-in-theoretical-computer-science/TCS-042

[19] B. David-John, K. Butler, and E. Jain, "Privacy-preserving datasets of eye-tracking samples with applications in XR," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2774–2784, 2023, conference Name: IEEE Transactions on Visualization and Computer Graphics.

[20] S. Hanisch, E. Muschter, A. Hatzipanayioti, S.-C. Li, and T. Strufe, "Understanding Person Identification Through Gait," *Proceedings on Privacy Enhancing Technologies*, vol. 2023, no. 1, pp. 177–189, Jan. 2023. [Online]. Available: https://petsymposium.org/popets/2023/popets-2023-0011.php

[21] G. M. Garrido, V. Nair, and D. Song, "Sok: Data privacy in virtual reality," 2023.

[22] V. Nair, W. Guo, R. Wang, J. F. O'Brien, L. Rosenberg, and D. Song, "Berkeley open extended reality recordings 2023 (boxrr-23): 4.7 million motion capture recordings from 105,852 extended reality device users," 2023.

[23] C. Rack, A. Hotho, and M. E. Latoschik, "Comparison of data encodings and machine learning architectures for user identification on arbitrary motion sequences," in *2022 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2022, pp. 11–19.

[24] Christian Rack, Lukas Schach, and Marc E. Latoschik, "Motion Learning Toolbox," 2023. [Online]. Available: https://github.com/cschell/Motion-Learning-Toolbox

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[27] I. Olade, C. Fleming, and H. N. Liang, "Biomove: Biometric user identification from human kinesiological movements for virtual reality systems," *Sensors (Switzerland)*, vol. 20, no. 10, pp. 1–19, 2020.

[28] B. Heller, "Watching androids dream of electric sheep: Immersive technology, biometric psychography, and the law," *Vanderbilt Journal of Entertainment and Technology Law*, vol. 23, 2020.

[29] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Grunefeld, F. Alt, and S. Schneegass, "Understanding user identification in virtual reality through behavioral biometrics and the efect of body normalization," *Conference on Human Factors in Computing Systems - Proceedings*, 2021, ISBN: 9781450380966.

[30] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk, "Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments," *Presence: Teleoperators and Virtual Environments*, vol. 13, no. 4, pp. 428–441, 2004.