A Review of Motion Data Privacy in Virtual Reality

Depeng Xu
Software and Information Systems
UNC Charlotte
Charlotte, USA
depeng.xu@charlotte.edu

Weichao Wang
Software and Information Systems
UNC Charlotte
Charlotte, USA
weichaowang@charlotte.edu

Aidong Lu
Computer Science
UNC Charlotte
Charlotte, USA
aidong.lu@charlotte.edu

***Abstract*—As the metaverse grows with the advances of new technologies, a number of researchers have raised the concern on the privacy of motion data in virtual reality (VR). It is becoming clear that motion data can reveal essential information of people, such as user identification. However, the fundamental problems about what types of motion data, how to process, and on what ranges of VR applications are still underexplored. This work summarizes the work of motion data privacy on these aspects from both the fields of VR and data privacy. Our results demonstrate that researchers from both fields have recognized the importance of the problem, while there are differences due to the focused problems. A variety of VR studies have been used for user identification, and the results are affected by the application types and ranges of involved actions. We also review the biometrics work from related fields including the behaviors of keystrokes and waist as well as data of skeleton, face and fingerprint. At the end, we discuss our findings and suggest future work to protect the privacy of motion data.**

***Index Terms*—Privacy, user identification, authentication, motion, virtual reality, VR**

I. Introduction

The interactions between real people and virtual metaverse rely on the tracking of human motions. Nowadays, all the virtual reality (VR) devices as well as other mobile devices such as smart phones or watches can measure motions at key joints and apply the tracking data to support various interactions. In metaverse, such interactions support users to achieve many actions that we do in the real-life, such as point and walk; and equip users with "super powers" actions like a superhero, such as fly and distant grabbing. They are must-have components of a metaverse.

However, recent work has shown that a variety of private user information can be inferred with the motion data. SoK [1] has provided a comprehensive taxonomy of data attributes, protections, and adversaries in VR, and covered all the layers from software to users, hardware of devices, and network sources. The work pointed out that there are more defensive works than the attacks, but the defensive solutions are far from satisfactory yet. From the angle of the metaverse applications and the underlying technologies, data privacy issues have also been reviewed based on the state-of-the-art solutions including federated learning, differential privacy, homomorphic encryption, and zero-knowledge proofs for different privacy problems [2].

Studies from related fields have also shown that various behavioral patterns can be used for user identification. One example on the large scale is the user behaviors from online services, such as call data records, web browsing histories, and GPS trajectories. Basic statistics of the behavioral patterns like histograms can be used to identify a large portion of users, and therefore serve as unique patterns for users like fingerprints [3]. An example on the small scale is the patterns from just hand movements. The hand patterns with two tasks of 3D pointing and gaming have been used to predict continuous hand trajectory in VR with a regressive model [4]. The results showed high prediction accuracy for the immediate continuous trajectory (from 100ms to 300ms) across all the users and activities, and the trained model could be applied to new users and new activities. In addition, the hand motions recorded from both cameras and VR controllers could be used to detect the keys users entered on a virtual keyboard [5].

This work reviews the privacy leakage of motion data in VR, given the fast growth of metaverse. Specifically we focus on the motion data in VR which only captures the motions of head and two hands, and occasionally enriched with additional data such as gaze or system context information. Our work summarizes the immediate needs of motion data privacy in VR, and the complexity of various factors that may affect the results of user identification. We also discuss the scope of research topics around the core problem of user identification and show the potential of additional privacy issues with the motion data.

The contributions of this work are:

- identify and collect the related work on the topic of user identification with motion data collected with VR devices.
- categorize the research works with the types of VR studies, data engineering, methods of identification, and conclusions.
- review the literature from related fields that use behaviors as biometrics, including skeleton, keystroke motion, fingerprint, wrist and face.
- discuss the problems based on the literature reviews and suggest important factors for the future works.

The following of this paper is organized as follows. We first summarize the related work that are selected for this review and categorize the work from several aspects of data collections, methods, and conclusions. In Section III, we review the behavior privacy problems from several related fields to demonstrate the various works on human behaviors.

We discuss the problems and suggest future works in Section IV, and conclude the paper in Section V.

II. Review of Motion Data for User Identification

We summarize the literature on privacy of motion data in Table I, specifically user identification and authentication. For the data collection, "standard motion" represents the 6DOF motion data, including position (3D) and rotation (3D) of head and two hands, that can be collected with majority VR devices. We review the literature from the following aspects, motion data, VR studies, data processing and feature engineering, classification techniques and results.

A. Survey Methodology

We first used the following keywords to search on the IEEE Xplore and ACM digital libraries.

- privacy, person identification, user identification, authentication
- motion data, skeleton, joint
- virtual reality, VR

We further read through the papers and removed the papers that are not directly related to the review topic. The rest of this section summarizes the remaining 18 papers as the core corp, and we review additional work from related fields in the Section III.

B. Motion Data in VR

To clarify the scope of the topic, motion data privacy in VR, we specify the types of data that are generally collected in VR studies. In majority VR devices, the head and the two hands are monitored with Inertial Measurement Units (IMUs) which combine three types of sensors; an accelerometer, a gyroscope and a magnetometer. The telemetry data is streamed into the storage of the head-mounted displays (HMDs) in real-time.

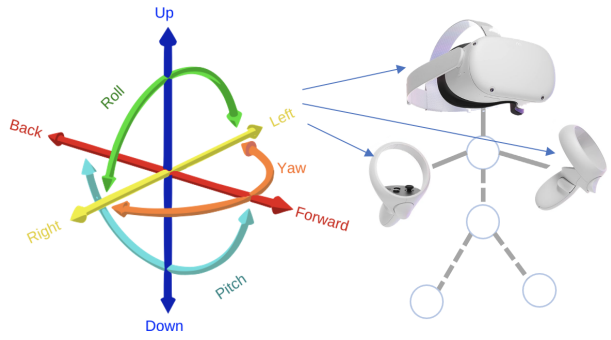

Fig. 1. VR devices collect the motion data that corresponds to the 6 DOF in a 3D space on the three joints, head and two hands.

Specifically, each of the three joints is treated as a rigid body, which means that its shape does not transform over time and its motion can be represented with a solid point. At each joint, multiple 3D metrics can be measured, such as the position and rotation for all the information of 6 DOF.

There are also alternate representations of the 6 DOF motions. Gaming engines such as Unity3D and UnReal provide several additional attributes from the basic position and rotation information. Among these, quaternions are used to represent rotations. A quaternion is a four-tuple of real numbers x,y,z,w, which is a mathematically convenient alternative to the euler angle representation. Up to now, it is still not clear which representation of the rotation is the best for classification models, therefore we include all works that use any of these representations.

Eye tracking data is also available at some devices, such as Quest Pro and Fove. These devices can be much more expensive than the commercial-level VR HMDs. Therefore this review tries to focus on the general VR devices that collect the 6 DOF motion data on the three joints covering head and hands and ignore the work that only uses gaze data.

C. VR Studies

While the privacy of motion data has been well-accepted as an important problem from all related fields of privacy, security and VR, all the research works picked one or two specific VR studies to explore this problem. For example, the game "Beat Saber" has been used in two research groups and several publications in Table I.

We summarize the choices of VR studies and the rationales based on the popularity of the application types. The rationales behind the choices of the VR studies were often explicitly provided in the works, as it is an important factor of the validity and potential impact of the results. The statistics of the study types is shown in Table II, where each study is put under one study type. Due to the overlapping features of some VR studies, such as a bowling game app belonging to both game and sport types, this table may be adjusted if we add these applications to all related types.

1) Game Type: The game type is the most popular choice, since games are still the dominant applications of VR and AR. A rich set of motion data can be collected via game type applications, and generally they require users to perform various tasks by combining all available interaction channels, especially flexible transitions among different buttons and thumbsticks on the VR controllers.

Among all the games, the games with rhythm and well-defined actions are often selected. The top choice is the game "Beat Saber", which has been continuously used by two research groups [6], [7], [24]. The "Beat Saber" has been used to generate the BOXRR-23 dataset which contains 4.7 million motion records from 105,000 Users [26]. Previously, the same group of researchers demonstrated that a large number of real VR users (N=55,541) can be uniquely and reliably identified [7] from this game. In addition, around 50 personal attributes were obtained from the users of the game "Beat Saber", and the results showed that over 40 attributes could be accurately and consistently inferred from VR motion data alone using simple machine learning models [6].

A different data collection method was also deployed without the collaboration with the game companies or related

TABLE I
A SUMMARY OF USER IDENTIFICATION WORK IN VR

ID	VR Study	Data Collection	Classification Methods	Conclusions
[6]	Game "Beat Saber" 1,006 users	Standard motion	RF, CNN, LSTM, and Transformer	40 attributes can be identified with accuracy over 58%.
[7]	Game "Beat Saber", 55,541 users	Standard motion	LightGBM	55,000+ users identified with 94.33% accuracy from 100 seconds of motion.
[8]	Game "Half-Life: Alyx", 71 users	Standard motion + gaze data + physiological data from a subset of 31 users	CNN and GRU	A mean accuracy of 95% for user identification within 2 minutes when trained on the first session and tested on the second.
[9]	Talking with Hands, 34 users	Body, finger, and audio data	RF, Multilayer Perceptron, Fully Recurrent Neural Network, LSTM, GRU	The model with the combination of a long-short term memory architecture and body-relative data correctly identifies any of the 34 subjects with an accuracy of 100% within 150 seconds.
[10]	Searching balls for 25 times, 23 users	Standard motion	Logistic Regression and SVM	The accuracy is around 93% for user authentication.
[11]	A driver dataset with 40 users, and 48 users watching five groups of 18 spherical videos	Head motion	Naive Bayes, PART, Logistic functions, Multilayer Perceptron, LMT	PART and LMT classifiers achieved 99% of accuracy in providing continuous authentication to the user in VR.
[12]	Observation of 360-degree VR video, 511 users	Standard motion	KNN, RF, and GBM	The system identifies 95% of users correctly when trained on less than 5 min of tracking data per person.
[13]	The VR learning study – training the participant how to troubleshoot a surgical robot, 60 users	Standard motion	KNN, RF, and GBM	Personally identify users at accuracy as high as 90.83%.
[14]	An AR everyday application (34 participants) and VR robot teleoperation (35 participants)	Standard motion + gaze data	Logistic Regression, Ridge Classifier, Decision Tree, and RF	Users identified up to 97% F1-score in VR and 80% in AR. Gender and Age inference reached up to 82 and 90% F1-score.
[15]	Picking virtual objects on a plan, 12 users	Standard motion + interaction data	CNN	The highest identification accuracy was 90.92%.
[16]	VR ball-throwing task, 41 users	Standard motion	Siamese neural networks	Identification results ranging from 87.82% to 98.53%.
[17]	41-subject dataset [16], [18], and the 33-subject dataset [19] – ball throwing 10 times on two separate sessions per system.	Standard motion	Siamese network with FCN	The results vary over short, median and long timescales, ranging from no statistically significant relationship to optimal performance for short and long enrollment/input separations by using training sets from users providing long-timescale data.
[20]	An ecologically valid VR training application, 61 users	Standard motion	KNN, RF, GBM	GBM performed the best with an average accuracy of 90.83% from the same session, and the average accuracy of all models were reduced by over 50%.
[21]	Controlled tasks require physical movements, 15 users	Standard motion	Decision Trees, Discriminant Analysis, SVM, Logistic Regression, kNN, Naive Bayes, and Ensemble Classification	User identification's accuracy was 98.6%. Penetration testing with 12 attackers resulted in confidence values ¡50%, although physically similar attackers had higher confidence values.
[22]	Two VR study, bowling and archery, 16 users	Standard motion	LSTM, RNN	An identification accuracy was up to 90% across sessions recorded on different days.
[23]	Stimulus following study which moved a sphere on two elliptical paths, 12 users	Head orientation + gaze	Encoder	100% accuracy on user identification
[24]	game "Beat Saber", 15 users	Standard motion	RF	Overall 86% accuracy for a cross-validated single-session and 71% for a two-session evaluation.
[25]	VR sports and exercises, 24 users	Standard motion	FCN and Inception	90.91% identification accuracy after one week.

Notes: Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), Long-Short Term Memory (LSTM), k-Nearest-Neighbors (KNN), Random Forests (RF), Gradient Boosting Machines (GBM), Support Vector Machine (SVM), Logistic Model Tree (LMT).

organizations. The data was collected over a period of eight weeks [24], and recorded with a third-party application. The app detected the start and end of a song in Beat Saber and transferred the recorded buffer to the server at the researchers' lab via the HTTP POST request. The transmitted data also included the name of the played song, its difficulty and modifier, the acquired user's score per frame, a timestamp, and a user-specific ID token [24]. The users were encouraged to play multiple times and suggested a specific song at different difficulty levels in the solo game mode. The total data records

TABLE II
SUMMARY OF VR STUDIES AND CONCLUSIONS

Study Type	# of Papers	Selected Result
Game	4	50,000+ users identified with 94.33% accuracy from 100 seconds of motion.
Sport	6	Identification results ranging from 87.82% to 98.53%.
Routine	6	The model with the combination of a LSTM architecture and body-relative data correctly identifies any of the 34 subjects with an accuracy of 100% within 150 seconds.
Training	2	GBM performed the best with an average accuracy of 90.83% from the same session, and the average accuracy of all models were reduced by over 50% from two sessions.

were 375 times during 119 sessions.

Other games have also been used, such as the game "Half-Life: Alyx". The HTC Vive Pro was used to record 71 users of this game for 45 minutes across two separate sessions [8]. The dataset included motion and eye-tracking data, along with physiological data from a subset of 31 users. The game contained navigation for players to walk around or teleport to another location, grabbing of virtual objects with either hand, and interaction methods including "gravity gloves" for remote grabbing and snatch for combat. Overall, the interaction methods often required hand motions and button clicks jointly.

2) Sport Type: Related to the game type, the applications of sport type have often been used, including bowling, ball-throwing tasks, archery, rock climbing, and other types of exercises. Similarly, the popularity of such types among VR applications may play an important role in the selection. VR applications have attracted much attention for sport athletes and rehabilitation of injured or senior people.

For example, a VR study investigated the temporal changes in VR behavior over short, medium, and long timescales. They used the 41-subject dataset [16], [18], which deployed on the HTC Vive, HTC Vive Cosmos, and Oculus Quest; and the 33-subject dataset [19] of a single VR system. The same task, a user throwing a ball at a virtual target 10 times on two separate sessions per system, was used in both datasets. A VR study with a set of typical tasks emphasized joint movements of the head, eyes, arms, wrist, torso and legs, such as grabbing, rotating and dropping [21]. Another study focused on task-driven scenarios and used Oculus Quest [22]. In the bowling task study, users were asked to hit as many pins as possible. In the archery task, users were placed in a forest and asked to shoot an arrow at the bull's eye.

Recently, a study investigated how different types of kinetic signatures influenced user identification in VR [25]. As an example VR exercise app, the game "OhShape" was selected for attractive features similar to "Beat Saber". It required users to adjust their shape to fit through the approaching walls and collect coins and offered four difficulty levels. For VR sport, several games were selected for capturing different types of kinetic movements, such as "Premium Bowling", "Indoor Rock Climbing VR" and "PowerBeatsVR". The key motions such as the throwing motion or when the rocks were hit were captured with the assistance of voice or visual prompts.

3) Routine Type: We categorize several studies in the routine type, as the rationales for choosing these studies are often similar – these actions are commonly used in various VR applications and therefore possess a higher possibility of leaking user identification information. Such applications include talking with hands, observation of 360-degree videos, and controlled tasks with physical movements.

The publicly available dataset "Talking with Hands" [27] used skeleton tracking, finger and audio data collections. Two main tasks were included to create this dataset: free conversation around a given topic and video retelling. The free conversation topics were chosen from a comprehensive set originally designed for casual conversations in English classes, and the video retelling asked users to watch a 5-minute video and then told the story to another user, during which conversations were allowed. The duration of the data ranged from 7 to 20 minutes.

Watching 360-degree videos in VR is also a common application. A dataset was collected by asking users to watch 360-degree videos, each lasts for 20s and randomly selected from a set of 80 videos, and answer questionnaires in VR [12]. In addition, the two user studies, an AR everyday application and VR robot teleoperation, included eleven generic actions (e.g., walking, searching, pointing) with different mental loads [14]. To replace the standard authentication method with password, a ball searching method asked users to look for a ball in the 3D space, and once found continued to search for a new location [10]. An VR study of interacting with virtual objects was performed [15].

4) Training Type: The training applications are becoming popular in VR, since VR is especially suitable for creating realistic environments that are hard or expensive to obtain in real-life education. The users of the training apps, such as students or trainees who are juniors, may be not aware of the privacy leakage as the general population. From the result aspect, the actions in the training apps are more likely to be similar, and could be used for user identification with higher accuracy degrees.

For example, a study trained participants how to troubleshoot a surgical robot [13]. The VR learning study used the HTC Vive and the SteamVR tutorial to train users how to use the Vive. The authors suggested that the VR learning study required lots of motion, and the results indicated that the personal identifiability of user tracking data was likely dependent upon the nature of the underlying VR experience.

Also, an ecologically valid VR training application [20] was performed using the Vive HMD. The study contained a variety of tasks, such as check error message, insert wrench, and use port clutch, which used interactions of walk, look, select, position, rotation or a combined interaction.

D. Data Processing and Feature Engineering

Since the datasets used in these studies are collected with user behaviors in real-life and real-time, the motion data has

to be processed before we can send it to classification models. The process of motion data generally goes through three stages, choices of data representations, data processing and feature engineering and classification models.

1) Data Representations: Just focusing on the motion data, the 6DOF data features can be represented in different ways. So far, all these representations have been used by works listed in table I. Since majority works used world space or did not clarify the space information, we ignore the space category from this summary and focus on the different representations. It is clear that the attributes of position, rotation, and quaternion are the most commonly used. They can still be processed into different features in the following steps.

2) Data Processing: The motion data is often further processed and concatenated as features as the inputs of the classification methods. We summarize several processing operations that are often performed for motion data: absolute or difference data values, normalization, statistical measurements for enrichment, time sequences, and additional records.

a) Absolute or difference values: The absolute values are the records of motion data, and the difference values can measure the changes of records relative to a base value, such as the records from the first time stamp. The absolute values are useful to represent specific locations and status of a motion, especially when they correspond to the virtual objects in the environment. For example, in the "Beat Saber" game, the textual information of the cubes that users interact with motions used the absolute values of both cubes and motions to match each other [7]. The relative positions in relation to the beginning of each kinetic sample was also used, which removed the bias from the starting and absolute positions [25]. The 21 feature vectors were further normalized and aligned the lengths with zero padding.

b) Normalization: The ranges of attributes are different, Euler angles are in [0, 360], quaternions in [-1, 1], and positions can be any real numbers. Therefore, normalization is often used to reduce the undesired weight effect on the classification results. The normalization process can also affect the results, e.g. two types of body normalization methods were tested, arm length and height normalization [22]. Four feature sets were then created with different combinations of motion data and stages of the bowling and archery actions. The choices depend on if it is desirable to apply these attributes, and a study has shown that body normalization in general can increase the identification rates for the Archery and bowling motions [22].

c) Statistical Measurement for Enrichment: Majority works collect statistical measurements from the motion data and use them to enrich the features. For example, the 90 dimensional features came from three categories: direction of head-movement, the magnitude of the head-movement, and movement time which represented the time duration between several head movements [11]. Specifically, they were calculated for each combination of summary statistic (maximum, minimum, median, mean, and standard deviation), body part (head, left hand, right hand), and dimension (x, y, z, yaw, pitch, and roll) [12]. The data processing process divided data into chunks and summarized the features inside each chunk. The results demonstrated that GBM performed the worst around 68.2% accuracy and random forest achieved the best around 95.3% accuracy. Another approach used the principle component analysis (PCA) to cut down the dimensionality with limited sample sizes, and only features produced when 95% of the variance was retained [10].

d) Time sequences: For time sequences, there is often a resampling process to convert all sequences to the same length, and linear interpolation is used to calculate the value of the resampled point based on the values of the left and right nearest points in the original sequence [21]. The data was further transformed with the PCA feature to reduce the data dimensionality before training the models.

e) Additional Collections: When additional data collections are available, such as gaze [14] or psychological data, the combined datasets need to be aligned on the time dimension. For example, the data attributes were collected at different speed, such as 80Hz for eye-tracking, 15Hz or 90Hz for motions with HTC Vive Pro, and 1Hz for heart rate [8]. For motion data, the rotation and position of both hands and the rotation of the head were concatenated as the features. Also the differences of each attribute were used to replace the absolute values, which still kept the 18 features. In [23], the gaze records were processed and normalized with a confidence value about the estimation of the correctness. The feature vector contained 21 attributes with the output of the gaze classifier and head orientation. The lengths of motion data were matched to the gaze data with arithmetic mean function. The features were further centered at the middle of the space to fit into the interval of [0, 1] for both head and eye data, and accumulated with a slicing window with 50% overlap.

3) Feature engineering: It is common that additional features can be computed from the original data and added to the features or used to replace the original processed motion data. Therefore, the features can be in diverse formats depending on the data collection methods and context information based on single or third party apps.

The longest feature vector combined motion features and context features [7]. The motion features from sequences before and after a target event were concatenated with the context features from the VR app. The motion features were absolute values of positions and quaternion rotations, and the statistical measures included min, max, mean, med, stdev for these two attributes among the 3 joints. The context features had 22 types including the position, orientation, type, and color of the block, the angle, speed, location, and accuracy of the cut, and the relative error of the cut in both space and time.

As another example, the motion data collection was standard, but the features combined attributes from both time and frequency domains [10]. The time domain features summarized statistical features, including range, interquartile range, third moment, fourth moment, variance, absolute sum, root mean square, mean, skewness, kurtosis, 25th percentile, and 75th percentile. The frequency domain features included spec-

tral flatness, spectral skewness, spectral kurtosis, spectral centroid, spectral spread, spectral rolloff, spectral entropy, and energy. There were also cross-stream features build on the rotation and position data, including 1-norm, infinity norm, frobenius norm, correlation YX, correlation ZX, correlation ZY, average pitch, average roll, average yaw, standard deviation patch, standard deviation roll, standard deviation yaw. The motion data was first smoothed with a 10th order Butterworth filter and a 5Hz cut-off frequency. The data was then split into 12-second windows with a 50% overlap in between. The time and frequency domain features were computed inside each window.

Another unique work generated two feature vectors, touching conditions with Pc, Dc, Sc, Dh and looking conditions with Dh and Sh. Specifically, Ph Dh, Sh represented the position, direction, speed vectors of the headset; Pc, Dc Sc represented the position, direction, speed vectors of the controller [15]. The data from the starting and ending periods were removed and only the central 20 seconds of the task were used. All the data sequences were later resized to 1024-element using linear interpolation.

E. Classification Techniques

Table I lists the time series classification models in the literature. The following provides several examples of such methods and their results.

It is common that several machine learning (ML) models are used and the best results are selected for the conclusion of a study. For example, several classification models [28] were used [25]. The best user classification results for LSHDEX (butterfly movements for the arms together with sidestep movements) were obtained by Inception [29] with an accuracy of 78.15%, and for LSHDSP (tennis hits) with FCN [30] at 90.91%. The conclusion was that user identification's accuracy was high from low static components of kinetic signatures, and certain movement types elicited more individual behavior than others. Another example used five classification methods in the experiment [11], and the results showed that PART and LMT classifiers achieved 99% of accuracy in providing continuous authentication to the user in VR. Overall, all the five methods achieved over 99% of accuracy for the video watching dataset [31], and performed very differently for the driving dataset [32] (78.3% – 99.6%). Also, in the stimulus following study that used HTC Vive Pro [23], the actual classification was performed through a review on time series classification [28], including Time Convolutional Neural Network (Time-CNN), Multi Layer Perceptron (MLP), Fully Convolutional Neural Network (FCN), Residual Network (ResNet), Encoder, Multi-scale Convolutional Neural Network (MCNN), Time Le-Net (t-LeNet), Multi Channel Deep Convolutional Neural Network (MCDCNN), Time Warping Invariant Echo State Network (TWIESN) and InceptionTime (Inception). The model Encoder [33] achieved the best results – 0.97 and 1.0 for the two stimulus scenarios.

The most complex architecture is designed for accommodating 55,541 users [7]. The model LightGBM was selected after comparing 6 popular classical ML classification models, and a multi-layer hierarchical approach was used to suit the large number of users [7]. Three layers were chosen, each layer contained N classifiers each trained on 1/N of the available classes, and the following layer redistributed the classes from the previous one. The overall classification results were finally determined by taking the highest logarithmic sum of the class probabilities outputs. The results demonstrated that a large number of real VR users could be uniquely and reliably identified [7]. Further, the researchers collected additional data and published the BOXRR-23 dataset which contained 4.7 million motion records from 105,000 Users [26].

Another unique work is to compare motions collected with different VR systems [16]. Based on the VR ball-throwing dataset from [18] with 3 VR systems – an Oculus Quest, an HTC Vive, and an HTC Vive Cosmos, the Siamese neural network was used to learn a system-to-system distance metric between motion data between VR systems. The Siamese neural network has shown good results with other related topics such as handwriting and EEG [16]. Specifically, a Siamese neural network architecture with FCN limbs was used and the identification results ranged from 87.82% to 98.53%.

III. Behavior Privacy from Related Fields

Related to motion data privacy in VR, face and fingerprint have been widely used as biometrics. We review several fields that also use 6DOF or 3DOF data at multiple key joints for user identification or authentication. Most of these fields can use cameras to collect data, such as skeletons, faces and finger motions. The keystroke motions are often collected with VR/AR devices, and can use cameras as well. The wrist motions are collected with smart watches. We provide example works for each related field.

A. Skeletons

Full body skeleton motion data captured by the Microsoft Xbox Kinect v2 has been commonly studied for action recognition, such as the NTU RGB+D dataset [34]. In [35], only using extracted joint distance features from static skeletons and dynamic gait parameters, they could achieve nearly perfect accuracy in identifying persons with only a few frames with simple machine learning classifiers, including k-nearest neighbor (kNN), decision tree, Gaussian Naive Bayesian, neural network with multiplayer perception (MLP), and support vector machine. Another work [36] developed a multi-task deep recurrent neural networks (RNN) to perform action recognition and person identification together. Learning the joint probability of the two could improve the action recognition performance. Research in [37] exploited skeleton data to infer personal attributes like gender of the users. Research in [38] used shift Graph Convolutional Network (GCN) models to accurately classify individuals' identification and gender. Research in [39] used a Siamese network inspired model for re-identifying individuals by correlating their skeleton data from private domains with publicly available datasets. These findings demonstrated high privacy risk in full body

motion data. Recently, several works have studied how to anonymize skeleton data for privacy preservation. Research in [38] created an adversarial training-based frame-by-frame anonymization framework for skeleton action recognition. It modified the skeleton data to confuse a personal ID classifier and a gender classifier while maintaining the performance of an action recognition model. Research in [39] suggested motion retargeting for anonymization. They used deep motion retargeting [40], [41] to get character-agnostic motion data. The spatial structure of the skeleton was transformed, but the essence of the motion pattern remained largely intact, ensuring that the anonymized data was still valuable for downstream applications.

B. Keystroke Motions

The problem of keystroke snooping has also raised increasing interests [42]. The attack to recognize the user's virtual typing is not necessarily easy, and it is built on a sequence of operations, including 3D keystroke position estimation, 3D cursor position estimation, keyboard alignment, and finally KNN classification. The results demonstrated that the attack can recognize the user's virtual typing with over 89.7% accuracy. Similarly, the keyboard input from HoloLens has been studied [43]. The data collection included motion data, gaze, and system information. With 25 users and 750 inference trials of passwords consisting of 4–8 lowercase English letters, results achieved a top-5 accuracy of 93%.

An interesting work is to provide a 2D video of VR users and allow an attacker to mimic the behaviors of victims for password-based authentication [44]. The results demonstrated that an attacker can match their 3D enrollment trajectories with the motion trajectories extracted from the 2D video to defeat behavior-based VR security. The results vary dramatically across different conditions.

C. Finger Motions

The finger movements have also been used to identify users [45]. The uni- and bimanual finger behavior from 16 users were gathered from the interaction with eight different universal interface elements, such as buttons and sliders, in VR (Quest 2) /AR (HoloLens2). The positions and rotational Euler coordinates were recorded for a location, and each hand contained 26 locations. The mean, min, max, and standard deviation further aggregate values within each time window. The Random Forest classifier achieved 95% accuracy of user identification across sessions. The feature set, F10, consisted of the head and all virtual bones of the index finger and thumb, including the fingertip, while the head position.y played an important role as it suggested the height of a user and it made a difference in the classification process.

D. Wrist Motions

Related to motion data in VR, wrist-wearables such as smartwatches and fitness bands have also been studied from the privacy aspect. While efforts have shown that it is possible to infer certain behaviors with a high success probability [46], other works concluded that consumer-grade wrist-wearables were difficult to inference handwriting due to unique and/or inconsistent behaviors [47].

E. Faces

Facial features and expressions have been studied as a biometric for avatar authentication [48]. A number of approaches have shown that face recognition can achieve an accuracy from low 60% to near perfect results. Multimodal biometric approaches are common nowadays, including forms of multi-sensor, multi-algorithm, multi-instance, and multi-model which often combine face data with fingerprint and voice.

IV. Discussions

We discuss several aspects of user privacy with motion data in VR and suggest possible future work directions.

A. User Identification with VR Studies

The results of user identification from our reviews vary dramatically. Just take the accuracy of classification models as an example, the accuracy can be around 90% for only 20 users or 97% for over 55,000 users. The purpose of the table II is for understanding the differences in the results. The game and routine types generally achieved higher accuracy values than the sport and training types. While the data collection and classification methods may all affect the results, zooming into the detailed data processing and feature engineering components reveals that the choices of motions and combinations of motions can be directly related to the classification results. The studies in our sport and training types contain more diverse motions without a fixed order, compared to the studies in the game and sport types. For example, each song of the "Beat Saber" contains a specific sequence of cubes and obstacles at the exact same time stamps, and a bowling action can be segmented to the prepare, throw and end stages. However, training a user to complete tasks with combinations of different interactions without clear sequences, or everyday exercises that involve a set of actions may both create challenges for the classification tasks. Therefore, we should expect different accuracy values based on the detailed motions in each study.

B. Applications

On the positive aspect, motion data has been used for authentication for VR systems. For example, Dice Palette supported 3D freehand interaction in user registration and login, and thereby improving security by avoiding bystanders from guessing gestures [49]. Also, FMHash (i.e., Finger Motion Hash) generated a compact binary hash code from a piece of in-air-handwriting of an ID string, and used hash searching to achieve convenient sign-in and sign-up with in-air-handwriting gesture ID on mobile and wearable systems [50].

From the current results, we suggest that certain VR applications should pay special attention to the privacy issue when they satisfy two criteria: 1) highly repetitive or rhythmic actions are involved and 2) users are students or seniors

that need special attention to privacy protection. Therefore, application fields such as education, training, and rehabilitation should enforce privacy protection policy and adopt defensive solutions.

C. Research Topics

Beyond the core problem of user identification, the latest works have also shown that motion data can be used to explore additional sensitive information. The VR sport and exercise study [25] showed that motion data from different kinetic actions could influence user identification results, and in addition revealed the sport capability or even health condition of users. The VR Game "Beat Saber" was used to show that 40 attributes from 1,006 users could be classified with good accuracy degrees, including age, race, rich or poor, and even the fabrics of the clothes that users wear while playing the game [6]. In addition, a large-scale example with 55,541 users has shown that the number of classes is not a problem in real-life applications [7]. In summary, we expect three branches of research directions for this field.

- New methods of data collection, engineering, classification that can improve user identification across different applications, devices, time, user, etc.
- New methods to explore sensitive information that can be revealed by the motion data with the enrichment of additional data collections.
- New methods to protect the privacy of users according to the applications and user preferences.

V. Conclusions and Future Work

This review summarizes the latest work on the privacy of motion data in VR. We cover the 18 papers searched from the ACM/IEEE digital libraries with several related keywords and a number of biometrics works from related fields. While the number of the literature is small, both the fields of VR and data privacy have recognized the problem. In addition, a variety of VR studies have been explored, covering several types of applications. The privacy of motion data has also expanded from the problem of user identification to the recognition of more detailed human behaviors, across devices, across time, and across users. It is also clear that we are still at the early stage of this topic, and there are many open problems from the aspects of study design, data collection, data processing, feature engineering, classification models, and defensive solutions for protecting the user's privacy.

Acknowledgment

This work was supported in part by NSF grant 1840080.

References

[1] G. Munilla Garrido, V. Nair, and D. Song, "Sok: Data privacy in virtual reality," *Proceedings on Privacy Enhancing Technologies*, vol. 2024, no. 1, p. 21–40, Jan. 2024. [Online]. Available: http://dx.doi.org/10.56553/popets-2024-0003

[2] C. Chen, Y. Li, Z. Wu, C. Mai, Y. Liu, Y. Hu, J. Kang, and Z. Zheng, "Privacy computing meets metaverse: Necessity, taxonomy and challenges," *Ad Hoc Networks*, vol. 158, p. 103457, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870524000684

[3] F. M. Naini, J. Unnikrishnan, P. Thiran, and M. Vetterli, "Where you are is who you are: User identification by matching statistics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 358–372, 2016.

[4] N. M. Gamage, D. Ishtaweera, M. Weigel, and A. Withana, "So predictable! continuous 3d hand trajectory prediction in virtual reality," in *The 34th Annual ACM Symposium on User Interface Software and Technology*, ser. UIST '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 332–343. [Online]. Available: https://doi.org/10.1145/3472749.3474753

[5] Z. Ling, Z. Li, C. Chen, J. Luo, W. Yu, and X. Fu, "I know what you enter on gear vr," in *2019 IEEE Conference on Communications and Network Security (CNS)*, 2019, pp. 241–249.

[6] V. Nair, C. Rack, W. Guo, R. Wang, S. Li, B. Huang, A. Cull, J. F. O'Brien, M. Latoschik, L. Rosenberg, and D. Song, "Inferring private personal attributes of virtual reality users from head and hand motion data," 2023.

[7] 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," in *32nd USENIX Security Symposium (USENIX Security 23)*. Anaheim, CA: USENIX Association, Aug. 2023, pp. 895–910. [Online]. Available: https://www.usenix.org/conference/usenixsecurity23/presentation/nair-identification

[8] C. Rack, T. Fernando, M. Yalcin, A. Hotho, and M. E. Latoschik, "Who is alyx? a new behavioral biometric dataset for user identification in xr," *Frontiers in Virtual Reality*, vol. 4, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frvir.2023.1272234

[9] 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.

[10] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead, "Unsure how to authenticate on your vr headset? come on, use your head!" in *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, ser. IWSPA '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 23–30. [Online]. Available: https://doi.org/10.1145/3180445.3180450

[11] M. Sivasamy, V. Sastry, and N. Gopalan, "Vrcauth: Continuous authentication of users in virtual reality environment using head-movement," in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, pp. 518–523.

[12] M. MR, H. F, J. H, L. JA, and B. JN, "Personal identifiability of user tracking data during observation of 360-degree vr video," *Scientific Reports*, vol. 10, 2020.

[13] A. G. Moore, R. P. McMahan, H. Dong, and N. Ruozzi, "Personal identifiability of user tracking data during vr training," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 556–557.

[14] P. P. Tricomi, F. Nenna, L. Pajola, M. Conti, and L. Gamberini, "You can't hide behind your headset: User profiling in augmented and virtual reality," *IEEE Access*, vol. 11, pp. 9859–9875, 2023.

[15] D.-M. Pham, "Human identification using neural network-based classification of periodic behaviors in virtual reality," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 657–658.

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

[17] ——, "Temporal effects in motion behavior for virtual reality (vr) biometrics," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 563–572.

[18] ——, "Within-system and cross-system behavior-based biometric authentication in virtual reality," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 311–316.

[19] A. Ajit, N. K. Banerjee, and S. Banerjee, "Combining pairwise feature matches from device trajectories for biometric authentication in virtual reality environments," in *IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2019.

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

[21] I. Olade, C. Fleming, and H.-N. Liang, "Biomove: Biometric user identification from human kinesiological movements for virtual reality systems," *Sensors*, vol. 20, p. 2944, 05 2020.

[22] J. Liebers, M. Abdelaziz, L. Mecke, A. Saad, J. Auda, U. Gruenefeld, F. Alt, and S. Schneegass, "Understanding user identification in virtual reality through behavioral biometrics and the effect of body normalization," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ser. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445528

[23] J. Liebers, P. Horn, C. Burschik, U. Gruenefeld, and S. Schneegass, "Using gaze behavior and head orientation for implicit identification in virtual reality," in *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3489849.3489880

[24] J. Liebers, C. Burschik, U. Gruenefeld, and S. Schneegass, "Exploring the stability of behavioral biometrics in virtual reality in a remote field study: Towards implicit and continuous user identification through body movements," in *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: https://doi.org/10.1145/3611659.3615696

[25] J. Liebers, P. Laskowski, F. Rademaker, L. Sabel, J. Hoppen, U. Gruenefeld, and S. Schneegass, "Kinetic signatures: A systematic investigation of movement-based user identification in virtual reality," in *In Proceedings of the CHI Conference on Human Factors in Computing Systems*, ser. CHI '24, 2024.

[26] 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 105,000 xr users," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–8, 2024.

[27] G. Lee, Z. Deng, S. Ma, T. Shiratori, S. Srinivasa, and Y. Sheikh, "Talking with hands 16.2m: A large-scale dataset of synchronized body-finger motion and audio for conversational motion analysis and synthesis," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 763–772.

[28] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019.

[29] H. Ismail Fawaz, B. Lucas, G. Forestier, C. Pelletier, D. F. Schmidt, J. Weber, G. I. Webb, L. Idoumghar, P.-A. Muller, and F. Petitjean, "Inceptiontime: Finding alexnet for time series classification," *Data Mining and Knowledge Discovery*, vol. 34, no. 6, p. 1936–1962, Sep. 2020. [Online]. Available: http://dx.doi.org/10.1007/s10618-020-00710-y

[30] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 1578–1585.

[31] "A dataset for exploring user behaviors in spherical video streaming," https://wuchlei-thu.github.io.

[32] "Virtual reality driving simulator dataset," https://www.kaggle.com/sasanj/virtual-reality-driving-simulator-datasetdataset.csv.

[33] J. Serrà, S. Pascual, and A. Karatzoglou, "Towards a universal neural network encoder for time series," 2018.

[34] J. Liu, A. Shahroudy, M. Perez, G. Wang, L. Duan, and A. C. Kot, "NTU RGB+D 120: A large-scale benchmark for 3d human activity understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 10, pp. 2684–2701, 2020.

[35] W. Zhao, S. Yang, T. Qiu, and X. Luo, "Person identification based on static features extracted from kinect skeleton data," in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021, pp. 2444–2449.

[36] H. Wang and L. Wang, "Learning content and style: Joint action recognition and person identification from human skeletons," *Pattern Recogn.*, vol. 81, no. C, p. 23–35, sep 2018. [Online]. Available: https://doi.org/10.1016/j.patcog.2018.03.030

[37] A. M. Glandon, J. Zalameda, and K. M. Iftekharuddin, "Transfer learning using infrared and optical full motion video data for gender classification," in *Infrared Technology and Applications XLIX*, G. F. Fulop, D. Z. Ting, and L. L. Zheng, Eds., vol. 12534, International Society for Optics and Photonics. SPIE, 2023, p. 1253418. [Online]. Available: https://doi.org/10.1117/12.2663972

[38] S. Moon, M. Kim, Z. Qin, Y. Liu, and D. Kim, "Anonymization for skeleton action recognition." AAAI Press, 2023. [Online]. Available: https://doi.org/10.1609/aaai.v37i12.26754

[39] T. Carr, A. Lu, and D. Xu, "Linkage attack on skeleton-based motion visualization," in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, ser. CIKM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 3758–3762. [Online]. Available: https://doi.org/10.1145/3583780.3615263

[40] K. Aberman, R. Wu, D. Lischinski, B. Chen, and D. Cohen-Or, "Learning character-agnostic motion for motion retargeting in 2d," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 75:1–75:14, 2019.

[41] K. Aberman, P. Li, D. Lischinski, O. Sorkine-Hornung, D. Cohen-Or, and B. Chen, "Skeleton-aware networks for deep motion retargeting," *ACM Trans. Graph.*, vol. 39, no. 4, p. 62, 2020.

[42] Y. Wu, C. Shi, T. Zhang, P. Walker, J. Liu, N. Saxena, and Y. Chen, "Privacy leakage via unrestricted motion-position sensors in the age of virtual reality: A study of snooping typed input on virtual keyboards," in *2023 IEEE Symposium on Security and Privacy (SP)*, 2023, pp. 3382–3398.

[43] S. Luo, X. Hu, and Z. Yan, "Holologger: Keystroke inference on mixed reality head mounted displays," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2022, pp. 445–454.

[44] R. Miller, N. K. Banerjee, and S. Banerjee, "Using external video to attack behavior-based security mechanisms in virtual reality (vr)," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2022, pp. 684–685.

[45] U. G. Jonathan Liebers, Sascha Brockel and S. Schneegass, "Identifying users by their hand tracking data in augmented and virtual reality," *International Journal of Human–Computer Interaction*, vol. 40, no. 2, pp. 409–424, 2024. [Online]. Available: https://doi.org/10.1080/10447318.2022.2120845

[46] T. Yan, Y. Lu, and N. Zhang, "Privacy disclosure from wearable devices," in *Proceedings of the 2015 Workshop on Privacy-Aware Mobile Computing*, ser. PAMCO '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 13–18. [Online]. Available: https://doi.org/10.1145/2757302.2757306

[47] R. Wijewickrama, A. Maiti, and M. Jadliwala, "dewristified: handwriting inference using wrist-based motion sensors revisited," in *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, ser. WiSec '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 49–59. [Online]. Available: https://doi.org/10.1145/3317549.3319722

[48] M. L. Gavrilova and R. V. Yampolskiy, "Applying biometric principles to avatar recognition," in *2010 International Conference on Cyberworlds*, 2010, pp. 179–186.

[49] X. Su, C. Cao, X. Xia, L. Chen, and B. Che, "Dice palette: Vr authentication based on freehand 3d interaction," in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2023, pp. 939–940.

[50] D. Lu, D. Huang, and A. Rai, "Fmhash: Deep hashing of in-air-handwriting for user identification," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–7.