IdeaVR: Interaction-edge Detection and Early Authentication Using Native Hand Tracking in Virtual Reality

Haoran Li
Department of Information Science Technology
University of Houston
Houston, Texas, USA
hli80@central.uh.edu

Omar Achkar
Department of Information Science Technology
University of Houston
Houston, Texas, USA
oachkar@central.uh.edu

Xin Fu
Department of Electrical and Computer Engineering
University of Houston
Houston, Texas, USA
xfu8@central.uh.edu

Kyu In Lee
Department of Information Science Technology
University of Houston
Houston, Texas, USA
klee48@central.uh.edu

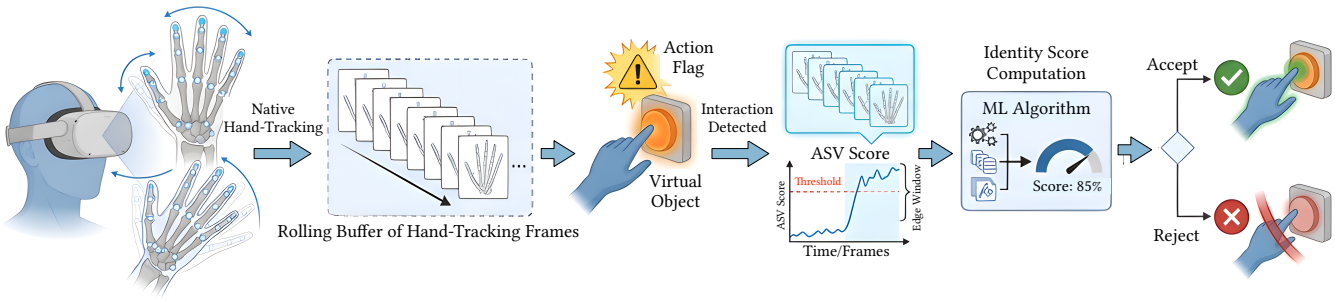


Figure 1: Illustrative runtime authentication scenario. IdeaVR continuously buffers native hand-tracking data. When an object interaction begins, it retrieves the recent buffer, localizes the interaction-edge window, computes an identity score, and accepts or rejects the interaction near the interaction edge.

Abstract

Virtual reality is increasingly used in security-sensitive settings, such as collaborative workspaces, shared asset management, and access-controlled environments, where a single object interaction can immediately alter ownership or application state. Login-time checks cannot protect later interactions, while continuous background authentication does not explicitly align identity decisions with security-relevant events. We present IdeaVR (Interaction-edge Detection and Early Authentication for Virtual Reality), a framework that verifies identity at the edge of each meaningful object interaction before the application treats the action as trusted. IdeaVR continuously buffers headset-native hand-tracking data and, at each interaction flag, uses Aggregated Signal Variance (ASV) onset localization to extract an interaction-edge window for authentication against the enrolled user. Using a 1D CNN, IdeaVR achieves 92.93% TAR and 3.46% FAR on held-out objects across 20 participants and 12 objects under a leave-one-object-per-class-out protocol. Across 12,000 simulated takeover sequences, 96.40% are rejected at the first attempted interaction. Representation ablations show that rich hand-shape features close the gap between interaction-edge and full-interaction windows (92.93% vs. 95.26% TAR), with hot-path latency under 4 ms. In a preliminary zero-shot cross-day evaluation with five participants, IdeaVR achieves 73.46% TAR and 6.56% FAR without recalibration. [1]

CCS Concepts

• **Human-centered computing** → *Human computer interaction (HCI)*; • **Security and privacy** → *Biometrics*; *Usability in security and privacy*.

Keywords

Virtual Reality, Authentication, Behavioral Biometrics

ACM Reference Format:
Haoran Li, Omar Achkar, Xin Fu, and Kyu In Lee. 2026. IdeaVR: Interaction-edge Detection and Early Authentication Using Native Hand Tracking in Virtual Reality. In *32nd ACM Symposium on Virtual Reality Software and Technology (VRST '26), November 16–18, 2026, Sendai, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3822517.3848707

This work is licensed under a Creative Commons Attribution 4.0 International License.
VRST '26, Sendai, Japan
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2811-2/2026/11
https://doi.org/10.1145/3822517.3848707

[1] IdeaVR dataset available at https://osf.io/yrtpg/overview?view_only=d924dadd944f40f4a0102b39ceff8082.

1 Introduction

Virtual reality (VR) systems are increasingly used as interactive computing environments where users manipulate objects, tools, and shared resources through natural hand motion. In these settings, hand motion serves as the primary means through which users express intent and manipulate application state. A grab, hold, move, place, or handover action can immediately affect the virtual environment, transferring object ownership, altering shared workspace state, or triggering irreversible application actions.

This makes authentication timing critical. A user may pass an initial login check, but another person can later gain access to the active headset or session and perform object-centered actions under the legitimate user's identity. If the system authenticates only at login, or issues background decisions at arbitrary intervals, it may identify the takeover only after an object action has already changed application state, too late to reject, delay, or flag the interaction before it produces lasting effects. Authentication should therefore not only ask *who* the user is, but *when* the decision can still protect the next action.

We therefore focus on the *interaction edge*: the short interval immediately before the runtime recognition of a meaningful object action. In common VR interaction frameworks, an object is treated as selected or grabbed only after an interaction condition is satisfied: a select action, grab input, or pickup function being activated [4, 34, 35]. However, the user's intended action begins earlier. Research on grasping shows that finger shaping begins well before contact, with coordinated multi-finger configurations reflecting the upcoming grasp and hand-shape patterns [2, 8, 32]. IdeaVR detects the onset of this coordinated finger closing activity and uses the selected interaction-edge window as the authentication opportunity: it retrieves recent hand motion at the action flag, estimates where task-relevant grasp formation begins, and issues an identity decision before the action is treated as trusted by the application.

Prior work has established that VR motion traces are identity-bearing [12, 13, 28, 30], and that biometric signals should be evaluated beyond single repeated movements or narrow lab settings [14, 26, 31]. However, these systems primarily ask whether users can be identified from motion, rather than when a decision can be issued to protect an upcoming action. Continuous authentication approaches issue decisions across the session but do not align them with security-relevant moments [21, 22]. Motion-forecasting work further shows that authenticating from early partial motion is more challenging than observing complete or near-complete trajectories [11], highlighting the tradeoff between decision timing and signal richness. IdeaVR addresses this tradeoff directly: rather than waiting for a complete interaction trajectory, it authenticates from the interaction-edge window, where the decision can still support action rejection, delay, or flagging.

We present IdeaVR, an interaction-level authentication framework for object-centered VR. IdeaVR continuously maintains recent hand-tracking frames in a rolling buffer and issues one explicit identity check per application-level interaction event. A curl-only Aggregated Signal Variance (ASV) detector estimates where task-relevant grasp formation begins inside the buffer, avoiding dependence on a fixed arbitrary pre-event crop. The downstream authentication model operates on the selected interaction-edge window using a rich hand-motion representation including hand-root pose, fingertip and knuckle geometry, and finger curl dynamics.

This paper makes the following contributions:

- We introduce interaction-edge authentication for VR, where identity is verified at the early stage of meaningful object actions rather than only at login time or arbitrary intervals.
- We design an action-triggered rolling-buffer pipeline that uses curl-only ASV to localize task-relevant grasp formation onset and produce interaction-edge authentication samples.
- We evaluate IdeaVR in a 20-participant object-controlled study with 12 objects across four Zingg-inspired shape categories [36] under a leave-one-object-per-class-out protocol.
- We show that short interaction-edge hand motion supports interaction-edge authentication on held-out objects, and that local hand geometry and curl dynamics contribute substantially beyond hand-root trajectory alone.

2 Background and Related Work

2.1 Behavioral Biometrics in VR

Virtual reality systems capture rich motion traces as a natural byproduct of interaction: head pose, hand position, finger configuration, and body movement are continuously logged at high frequency simply because the system needs them for rendering. This creates an opportunity: the same data stream that drives the virtual environment may also carry identity-bearing signals that can be used for authentication without additional sensors or explicit user effort.

Prior work has confirmed that VR motion is identity-bearing across a range of signals and tasks. Pfeuffer et al. showed that body motion and body-part relations in VR can identify users [30]. Liebers et al. examined body normalization, gaze behavior, head orientation, and systematic movement signatures as identification signals [12, 15–17]. Nair et al. demonstrated that head and hand motion from a large-scale VR population can uniquely identify users at scale [28]. Jalilov et al. confirmed that hand configuration and finger motion are identity-bearing signals [7]. These findings establish that natural VR behavior carries strong identity information across users and contexts.

Generalization across sessions and tasks remains an open challenge. Rack et al. introduced a large-scale behavioral biometric dataset across two sessions [31], and Liebers et al. studied the stability of behavioral biometrics in a remote field study over eight weeks [14]. Moore et al. showed that identifiability can vary across domain-specific tasks [26], and large-scale telemetry work confirms that motion data is privacy-sensitive at scale [27, 29]. Together these studies show that biometric signals survive variation in session conditions and task context, but that evaluation protocols must test beyond single repeated interactions to establish practical robustness.

IdeaVR builds on this foundation but asks a different system question: not whether users can be identified from VR motion, but when an identity decision can be issued to protect a specific upcoming application action.

2.2 Interaction Preparation and Hand Shape

Before an application recognizes an object action, the user is already engaged in structured motor behavior. Motor control research on grasp preparation shows that finger shaping begins well before contact, with the hand preshaped according to the object's geometry and the intended grip while the arm is still in transport [2, 8]. Object properties such as size, shape, and orientation influence approach trajectory, finger spread, and grip aperture [6]. Finger configurations during grasp preparation are not independent per-finger movements but reflect coordinated postural synergies that are consistent across repeated interactions with similar objects [32]. This structure means that interaction preparation is not arbitrary pre-event motion; it is behaviorally organized, object-directed, and potentially user-specific.

This has direct implications for authentication. If grasp preparation produces consistent, structured hand-shape behavior, it may carry identity information even before the application has recognized the action. Prior work on VR behavioral biometrics, as discussed above, confirms that hand configuration and finger motion are identity-bearing signals in VR. IdeaVR targets this interval specifically: it detects the onset of coordinated finger closing activity and uses the selected window for authentication, operating within the interaction preparation phase before the application treats the object as acted upon.

2.3 Authentication Timing and System Design

The VR authentication literature spans a range of design philosophies. Early work focused on adapting familiar mechanisms (PINs, patterns, and challenge sequences) to the VR context [3, 20], while behavior-based approaches emerged as a way to authenticate users through their natural motion without interrupting the experience [21–25]. Sensing signals have also diversified beyond hand and head motion to include head tremors, vibration, and gaze [1, 5, 10], and broader surveys map the growing design space [33]. Together these systems show that VR/XR devices can support authentication through a range of mechanisms, but they leave open the question of when decisions should be issued relative to security-relevant application events.

That timing question turns out to be subtle. Continuous background methods keep the user from having to do anything explicit, but they issue decisions whenever the system samples, not necessarily when something security-relevant is about to happen. Challenge-based methods can be targeted but interrupt natural use, a cost that matters in practice: Mare et al. found that interruption frequency is one of the primary factors shaping how users experience authentication in daily life [19]. Task-driven approaches [9] narrow the window somewhat, but still do not tie decisions explicitly to application-level object actions.

The stakes of this timing gap become clearer in light of recent motion-forecasting work. Li et al. showed that authenticating from early partial motion is substantially harder than waiting for a complete trajectory [11], which means that issuing a decision early enough to still protect an upcoming action comes at a real signal cost. IdeaVR accepts that tradeoff deliberately. Rather than waiting until the interaction has run its course, it authenticates from the interaction-edge window, near the interaction edge, allowing the application to hold, reject, or flag the associated state update before committing it as trusted.

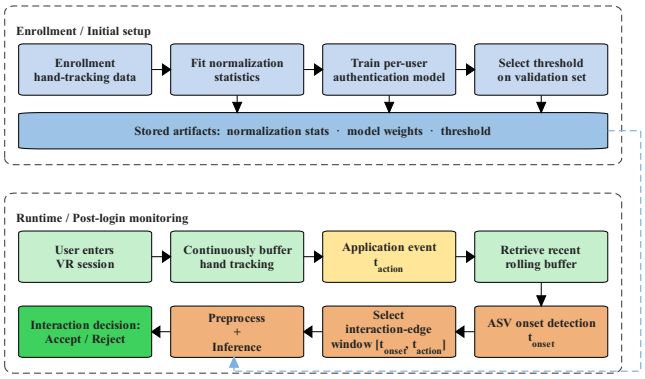


Figure 2: IdeaVR system overview. Enrollment produces stored artifacts (normalization statistics, model weights, and threshold) from hand-tracking data. At runtime, an application-level event at t_{action} triggers buffer retrieval, ASV-guided onset detection to estimate t_{onset}, and authentication over the selected interaction-edge window $[t_{\text{onset}}, t_{\text{action}}]$.

3 Threat Model

IdeaVR addresses session takeover after initial login. We assume a legitimate user has established a VR session, and a second person later gains physical access to the active headset and attempts object-centered actions: grabbing, holding, moving, placing, or otherwise manipulating virtual objects under the legitimate user's identity. Such actions can immediately transfer object ownership, alter shared workspace state, or trigger irreversible application changes. IdeaVR is not a replacement for initial login; it adds interaction-edge protection by checking identity at the early stage of each meaningful interaction.

We assume the attacker does not control the trained model, the hand-tracking pipeline, or the authentication threshold. We evaluate ordinary session takeover, where another person naturally attempts object interactions after gaining access. The reported FAR therefore reflects ordinary impostors and does not represent performance against skilled mimics.

4 System Overview

4.1 Interaction-Edge Authentication

IdeaVR operates in two phases: enrollment and runtime authentication. During enrollment, the system collects interaction-edge samples from the legitimate user using the same sensing and preprocessing pipeline used at runtime. These samples are used to fit preprocessing parameters, train the authentication model, and select an operating threshold on validation data. After enrollment, all parameters and the threshold are fixed for runtime inference.

At runtime, IdeaVR continuously maintains recent hand-tracking frames in a rolling buffer but does not issue frame-by-frame authentication decisions. Instead, each application-level interaction event triggers one explicit identity check. When the event occurs, IdeaVR retrieves the recent buffer, estimates where task-relevant

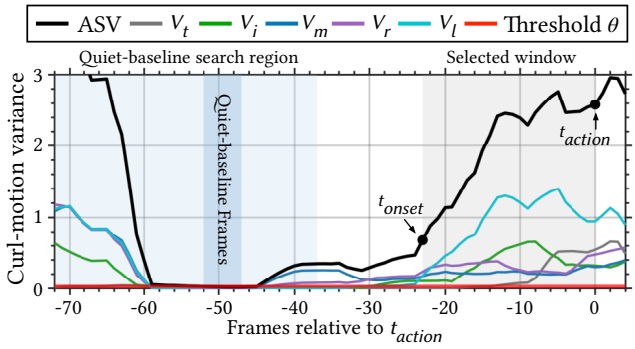


Figure 3: Representative trial showing the quiet-baseline region, ASV signal, detected onset t_{onset}, action time t_{action}, and the resulting selected window.

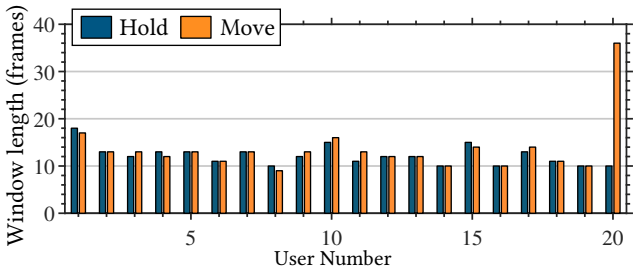


Figure 4: ASV-selected window lengths across users for HOLD and MOVE interactions; bars show the median window length.

interaction preparation begins, constructs an interaction-edge sample, and checks whether the current actor matches the enrolled user before the action is treated as trusted by the application. This design provides continuous protection across the active session while aligning each decision with a moment that is about to affect application state. Figure 2 illustrates the full pipeline.

4.2 Rolling Buffer and Event Trigger

The buffer is not the authentication window itself; it defines the maximum search horizon from which ASV can recover the interaction preparation that led to the current event. The horizon is a tunable system parameter; a deployment can set it based on expected interaction style, memory budget, and latency constraints. In our current implementation, the buffer contains 72 frames.

Each buffered frame contains hand-root position expressed relative to the headset frame, fingertip and knuckle positions expressed relative to the hand root, and per-finger curl values. This preserves both global approach motion and local hand-shape behavior while reducing dependence on absolute world position. Model-specific z-score normalization is applied after window selection, using statistics estimated from training data and held fixed for validation, test, and runtime inference.

When an interaction event occurs, IdeaVR treats the first flagged frame as t_{action}, the runtime recognition point for the object action. The system retrieves the preceding buffer and passes it to the ASV detector, which estimates t_{onset}, the beginning of task-relevant grasp formation. The end of the selected window is always fixed at t_{action}; only the start varies based on the detected onset.

4.3 ASV-Guided Window Selection

Prior work shows that grasp-closure timing varies with object and movement conditions [18], suggesting that a fixed interaction-edge crop may not consistently capture the same phase of grasp formation. To address this, IdeaVR uses Aggregated Signal Variance (ASV) to approximately localize grasp onset within the rolling buffer and adapt the authentication window to each interaction. ASV is not the biometric classifier; it estimates t_{onset} to obtain a behaviorally aligned window. The resulting ASV-selected window lengths in Figure 4 vary across users and interaction intents, providing empirical evidence that the useful interaction-edge duration is not uniform across interactions.

For each buffer, IdeaVR computes an ASV curve from the five finger-curl channels. Let $c_t \in \mathbb{R}^F$ be the finger-curl vector at frame t, where $F = 5$, and let $\Delta c_t = c_t - c_{t-1}$. The ASV signal is

$$\text{ASV}_t = \sum_{j=1}^{F} \text{Var}\big(\Delta c_{t-w+1}^{(j)}, \ldots, \Delta c_t^{(j)}\big), \tag{1}$$

where w is the rolling variance window length. In our implementation, $w = 5$ frames, and the resulting ASV curve is smoothed using an 11-frame rolling mean.

IdeaVR estimates an activity threshold from quiet baseline frames in the early part of the buffer, excluding frames that show coordinated finger closing or elevated activity from prior-task residual motion. The threshold is

$$\theta = \text{median}(\mathcal{B}) + k \cdot \text{IQR}(\mathcal{B}), \tag{2}$$

where $\mathcal{B}$ is the quiet baseline set, and $Q_{25}(\mathcal{B})$ and $Q_{75}(\mathcal{B})$ denote its 25th and 75th percentiles, respectively, with $\text{IQR}(\mathcal{B}) = Q_{75}(\mathcal{B}) - Q_{25}(\mathcal{B})$. This estimator reduces sensitivity to isolated noisy frames, while additional threshold floors prevent near-zero baselines from triggering on noise or gradual curl drift.

In the current implementation, a candidate onset must also show positive curl change across at least half of the fingers over a 3-frame window, with a 4-frame lookahead used to verify continued closing activity. When multiple candidate grasp onsets occur within the buffer, IdeaVR selects the onset closest to t_{action} when it is separated from earlier motion by a quiet gap; otherwise, the candidates are treated as part of one continuous grasp-formation episode and the earliest onset is used. If no valid onset is found, t_{onset} is set to the start of the 72-frame buffer preceding t_{action}, ensuring that every interaction event produces an authentication sample. Figure 3 illustrates the ASV-guided selection process in a representative trial, while Figure 4 shows the distribution of selected window lengths across users and interaction types.

4.4 Scoring and Decision

After the interaction-edge window is selected, IdeaVR converts the frame slice into the model-ready representation used during enrollment and runtime inference. The ASV detector uses only curl channels for window selection, but the authentication model

operates on the full input bundle: hand-root pose, fingertip coordinates, knuckle coordinates, and finger curl values, allowing richer hand-motion information to inform the identity decision.

The selected window may vary in length depending on the detected onset. Variable-length interaction-edge windows are handled according to the sequence model. The CNN uses a fixed 36-frame tail crop nearest to the interaction edge, whereas the Transformer preserves the full detected window, up to the 72-frame buffer limit; shorter sequences are post-padded with zeros, with padding masks used to exclude padded frames from attention. Classical baselines resample each detected window to 36 frames before descriptor extraction.

IdeaVR is model-agnostic after the interaction-edge sample has been prepared. The deployed model outputs an authentication score s for the enrolled user, and IdeaVR compares this score against a validation-selected threshold τ:

$$d = \begin{cases} \text{accept}, & s \geq \tau, \\ \text{reject}, & s < \tau. \end{cases} \tag{3}$$

The threshold is selected on validation data to maximize TAR subject to a target FAR of at most 5%, then held fixed for test and runtime inference. Scores below the threshold are flagged as suspicious; the application can then reject the action, freeze or delay the object update, request re-authentication, or log the event for later mitigation. Response policy is application-defined and outside the scope of this paper.

5 Experiment Design and Dataset

5.1 Study Overview

The experiment instantiates the session-takeover setting described in Section 3. The security-relevant event is not initial login but a later object-centered action performed inside an already authenticated VR session. We therefore collect natural hand motion around concrete object interactions rather than explicit login gestures or challenge-response tasks.

Each participant provides legitimate interaction-edge behavior through normal object manipulation. Each flagged interaction is treated as one candidate authentication event. Section 6 describes how these samples are partitioned for training, validation, and testing.

The study includes 20 participants, 12 virtual objects, and two interaction intents: hold and move. The objects are organized into four Zingg-inspired shape categories, with three objects per category, exposing IdeaVR to object-induced variation in approach motion, finger spread, and pre-contact hand shaping while keeping the interaction context controlled.

5.2 Apparatus and Logging

The study used a Meta Quest 3S headset with native hand tracking. The VR application was built in Godot 4.3 with Godot XR Tools and required no controllers, eye tracking, external cameras, or wearable sensors. The application logged hand-tracking frames and interaction flags at 72 Hz.

Each frame contains the hand-tracking channels described in Section 4.2. Each interaction flag marks the beginning of a meaningful hold or move interaction and is used to form an ASV-selected interaction-edge authentication sample. For each flagged interaction, IdeaVR reconstructs a bounded interaction-edge buffer before the first flagged frame.

Model training, evaluation, and latency measurements were performed on a desktop workstation with an Intel Core Ultra 9 285K CPU, 64 GB RAM, and an NVIDIA GeForce RTX 5080 GPU with 16 GB VRAM, running Windows 11 Enterprise with a Linux-based Python/PyTorch environment. Latency results estimate computation cost on this workstation rather than full on-headset deployment overhead.

5.3 Participants and Ethics

The study includes 20 participants recruited from the university community. All participants were right-handed, and the evaluation used right-hand tracking streams. Using one hand keeps the study controlled and avoids mixing handedness effects with identity recognition. The IdeaVR framework is not tied to a particular hand. Demographic variables beyond handedness were not collected.

The study was conducted under institutional review board approval. Participants were informed of the study purpose and the type of data collected before data collection began. No personal identifying information was collected as part of the dataset. Each participant was assigned an internal identifier with no mapping to real identity.

5.4 Object and Interaction Design

The study uses 12 virtual objects organized into four Zingg-inspired shape categories, as shown in Figure 5. Zingg-style classification describes coarse three-dimensional form using relationships among long, intermediate, and short object axes. In our VR setting, these categories are useful because object geometry can affect approach trajectory, finger spread, and pre-contact hand shaping. The object groups are: Equant (Ball, Cube, Soda), Prolate (Banana, Bottle, Screwdriver), Oblate (Burger, Cheese Slice, Plate), and Bladed (Knife, Spatula, Wrench).

Each object is used with two interaction intents, hold and move, as illustrated in Figure 6. In the hold condition, the participant reaches for the object, holds it for 2 seconds, and releases it, forming a short object-engagement task that emphasizes approach and early stabilization. In the move condition, the participant picks up the object and moves it through a designated path to a target location, producing more sustained hand motion and stronger trajectory constraints. Together, hold and move allow us to study interaction-edge authentication across two controlled action contexts with different motor demands.

5.5 Procedure and Dataset Construction

Participants remained seated throughout and completed object-centered interaction trials inside the VR application. Object appearances were shuffled across trials to reduce the chance that models would learn predictable sequence patterns rather than user-specific interaction preparation behavior. Each participant was scheduled to complete 20 repetitions of each object–intention combination, corresponding to 480 planned interactions per participant and 9,600 planned interactions across the 20-participant study. After data collection, 9,474 interactions were retained, corresponding to a 98.69%

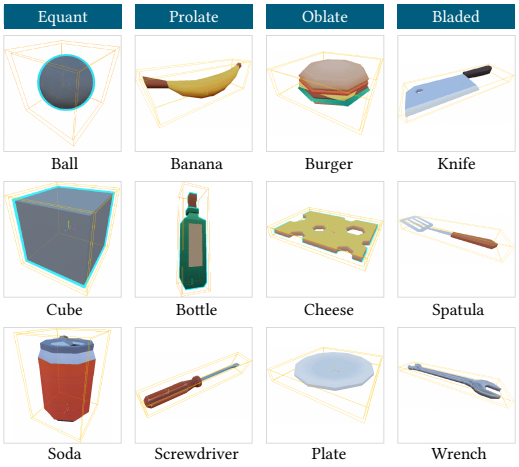
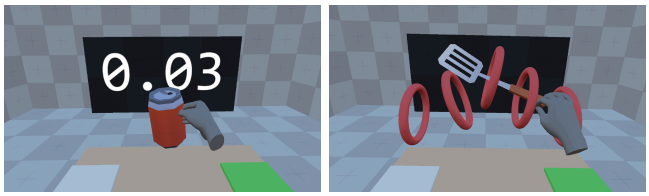


Figure 5: Object design with Zingg-inspired groups.



(a) Hold interaction (b) Move interaction

Figure 6: Hold and move interaction experiment.

Table 1: Leave-one-object-per-class-out fold assignment.

Fold	Equant	Prolate	Oblate	Bladed
Fold 1	Ball	Banana	Burger	Knife
Fold 2	Cube	Bottle	Cheese	Spatula
Fold 3	Soda	Screwdriver	Plate	Wrench

completion rate; a complete collection session lasted approximately 40 minutes.

The resulting dataset consists of continuous hand-tracking recordings across all trials, with interaction flags logged at the moment each object-centered hold or move action is recognized by the application. Each session captures the full hand-tracking stream alongside the flagged interaction events, providing the raw material from which authentication samples are later extracted. Section 6 describes how these samples are partitioned for training, validation, and testing.

6 Authentication Model and Evaluation Protocol

6.1 Authentication Samples and Representation

Each application-flagged object action produces one authentication sample, drawn from a rolling buffer of 72 frames (1 s at 72 Hz). The interaction flag defines t_{action}, and IdeaVR uses ASV to estimate t_{onset} from the preceding buffer. The resulting interaction-edge interval $[t_{\text{onset}}, t_{\text{action}}]$ is the model input for that authentication attempt (see Figure 2). A sample is indexed by participant, collection trial, logged file, and task label.

For each enrolled user, samples from that user are labeled genuine, while samples from all other participants are labeled impostor. This one-vs-rest formulation matches the runtime setting: the system checks whether the current actor matches the enrolled identity associated with the active session. The decision is made once per object action, not once per arbitrary time clip.

IdeaVR represents each interaction-edge sample as a sequence of hand-tracking channels. Coordinate normalization follows the scheme described in Section 4.2.

We evaluate four input bundles of increasing richness: hand position only (3 total channels), hand pose (6 total channels), hand pose with local joint geometry (36 total channels), and the full hand geometry representation with finger curl (41 total channels).

6.2 Models

We evaluate both sequence-based and feature-based authentication models. Sequence models receive normalized interaction-edge windows and learn temporal representations directly. We evaluate a 1D CNN and a Transformer as representative sequence models. The CNN uses three temporal convolution stages followed by squeeze-and-excitation and temporal-attention pooling, while the Transformer uses two encoder blocks with $d_{\text{model}} = 64$, four attention heads, and a 256-dimensional feed-forward layer. Both neural models are trained using binary cross-entropy with logits and AdamW for up to 25 epochs with batch size 128. Feature-based models summarize each window into engineered descriptors and are evaluated with Random Forest (RF) and one-class SVM (OCSVM) as baselines. For RF and OCSVM, each input channel is summarized by its mean, standard deviation, minimum, maximum, and median. The RF uses 300 trees with minimum leaf size 2 and square-root feature subsampling. The OCSVM uses an RBF kernel, with $\nu \in \{0.01, 0.05, 0.1, 0.2\}$ and $\gamma \in \{\text{scale}, \text{auto}, 0.001, 0.01, 0.1\}$ selected on validation data. The purpose of this comparison is to test whether interaction-edge windows provide useful authentication evidence across model families. Training and threshold selection are performed separately for each enrolled user.

6.3 Leave One Object Per Class Out Folds

We evaluate IdeaVR under an object-holdout protocol that approximates deployment conditions where enrollment cannot cover every object a user may later manipulate. As shown in Table 1, we use a three-fold leave-one-object-per-class-out design. The object groups follow the Zingg-inspired categories shown in Figure 5. In each fold, one object from each shape group is withheld from training and validation and used only for testing. The remaining objects are used for training and validation. Across the three folds, every object is withheld exactly once.

Both hold and move variants are included on the train and test sides of each fold. By holding out one object per Zingg category, the protocol tests whether interaction-edge authentication generalizes to unseen objects when the shape category is represented in enrollment.

6.4 Performance Metrics and Operating Point Selection

Test performance is reported using true acceptance rate (TAR), false acceptance rate (FAR), and equal error rate (EER). In the one-vs-rest formulation, a true positive (TP) is a genuine sample correctly accepted, a false negative (FN) is a genuine sample incorrectly rejected, a false positive (FP) is an impostor sample incorrectly accepted, and a true negative (TN) is an impostor sample correctly rejected:

$$\mathrm{TAR} = \frac{TP}{TP + FN}, \quad \mathrm{FAR} = \frac{FP}{FP + TN}. \tag{4}$$

EER denotes the operating point at which FAR equals the false rejection rate (1 − TAR). Validation splits are grouped by collection trial so that samples from the same trial are not split across training and validation, reducing leakage from repeated within-trial behavior. Authentication thresholds are selected on validation data to maximize TAR subject to a target FAR of at most 5%, then held fixed for test and runtime inference. TAR and FAR are reported at this operating point. EER is computed separately for each object fold and averaged across folds; it is not used for threshold selection.

6.5 Ablations, Takeover Simulation, and Timing

We include two ablations to clarify what supports the interaction-edge decision. First, within the same ASV-selected interaction-edge window, we compare four input representations: global position only, global pose (position + rotation), global pose with local fingertip and knuckle geometry, and full hand geometry including finger curl. This tests whether local hand-shape behavior contributes beyond global hand motion. Second, we compare interaction-edge and full-interaction windows under two input conditions, global position only and full hand geometry, to isolate the timing cost of early authentication. This tests whether rich hand-shape features can close the performance gap between the interaction-edge window and the full interaction window.

For takeover evaluation, we simulate interaction-level session takeover using held-out impostor interactions. For each enrolled user, impostor actions are randomly ordered to form takeover sequences, and we measure how many attempted interactions occur before the first rejection. This reports detection in interaction units rather than wall-clock time, matching IdeaVR's interaction-edge decision model.

We also report decision latency for the runtime path: buffer retrieval, ASV onset detection, sequence preparation, model inference, and threshold comparison. Timing is measured offline from logged samples on the desktop workstation described in Section 5.2 and estimates the computational cost of the decision path after an interaction event occurs.

7 Results

7.1 Authentication Performance

Table 2 summarizes authentication performance under the leave-one-object-per-class-out protocol. Thresholds are selected on validation data to maximize TAR subject to a target FAR of at most 5%, then held fixed on held-out test objects. The sequence models provide the strongest operating tradeoff. The CNN reaches 92.93% TAR while keeping FAR at 3.46%, below the validation target, with an EER of 5.52 ± 0.53%. This relatively low balanced-error point suggests that the interaction-edge signal can support a useful trade-off between security, by limiting false acceptance, and usability, by avoiding excessive rejection of legitimate users. The Transformer follows the same pattern with 89.81% TAR, 4.19% FAR, and 6.76 ± 0.32% EER, suggesting that the interaction-edge window contains identity-specific temporal structure that is useful across sequence-model architectures.

Table 2: Authentication performance across three object folds. TAR, FAR, and EER are averaged across folds; Hold and Move report the full CNN by interaction intent.

Model	Input	Test TAR	Test FAR	EER
RF	Engineered features	92.36 ± 1.03%	10.92 ± 0.75%	8.90 ± 0.16%
OCSVM	Engineered features	63.69 ± 4.37%	4.66 ± 0.82%	27.41 ± 3.04%
CNN	Sequence	**92.93 ± 0.22%**	**3.46 ± 0.29%**	**5.52 ± 0.53%**
– Hold	Sequence	92.60 ± 0.63%	3.66 ± 0.34%	5.93 ± 0.51%
– Move	Sequence	93.27 ± 0.18%	3.26 ± 0.28%	5.10 ± 0.78%
Transformer	Sequence	89.81 ± 1.00%	4.19 ± 0.37%	6.76 ± 0.32%

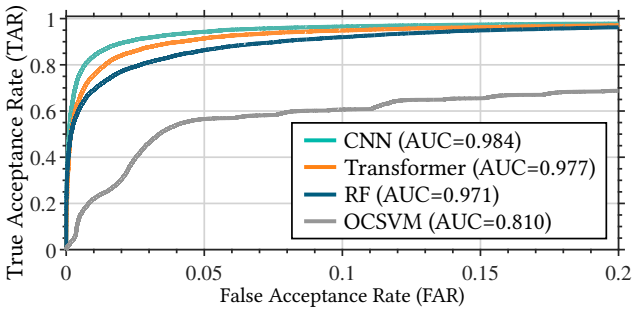


Figure 7: ROC curves for all models on the same-day object-holdout evaluation. Scores are concatenated across all users and three object folds before computing each curve. AUC (area under the curve) summarizes discrimination across all thresholds; higher is better.

The CNN is also consistent across held-out object folds and interaction intents. Across folds, TAR remains within 92.75–93.24% and FAR within 3.07–3.77%, indicating that performance is not driven by a particular held-out object combination. Hold and Move likewise show similar operating behavior, with EERs of 5.93 ± 0.51% and 5.10 ± 0.78%, respectively. Despite differences in intents and ASV-selected window lengths, this consistency suggests that the interaction-edge identity signal generalizes across both unseen objects and distinct manipulation contexts.

The classical baselines illustrate why the operating point matters. The RF baseline reaches 92.36% TAR but its FAR rises to 10.92%, above the intended false-acceptance region, making it less suitable for interaction-edge authentication despite its high genuine acceptance. OCSVM keeps FAR near target at 4.66% but TAR drops to 63.69% with high fold variance (±4.37%), indicating both excessive genuine rejection and sensitivity to the held-out object set. The CNN therefore gives the best overall balance. Receiver operating characteristic (ROC) curves across all models are shown in Figure 7.

Table 3: Ablation of input representation and window source across three object folds. TAR, FAR, and EER are averaged across folds. Panel A varies input features; Panel B compares interaction-edge and full-interaction windows.

Configuration	TAR	FAR	EER
Panel A: Input representation (Interaction-edge window)			
Hand position only (3D)	53.87%	4.46%	17.99%
Hand pose (6D)	73.82%	4.43%	12.57%
Hand pose + local joint geometry (36D)	80.57%	4.18%	9.96%
Full hand geometry (41D)	92.93%	3.46%	5.52%
Panel B: Window source			
Hand position only			
Interaction-edge window	53.87%	4.46%	17.99%
Full-interaction window	76.04%	4.36%	11.19%
Full hand geometry			
Interaction-edge window	92.93%	3.46%	5.52%
Full-interaction window	95.26%	2.85%	4.10%

7.2 Input Representation and Window Source Ablation

Table 3 presents two complementary ablations that together motivate the IdeaVR design: which features to use, and when to authenticate.

The interaction-edge window is the basis for both ablations. ASV localized a valid onset in 99.2% of interactions, indicating that coordinated finger closing activity can be consistently detected by the proposed selector in our setting. Using only this window, we progressively enrich the input representation from global hand position alone to the full hand geometry. Global hand position only reaches 53.87% TAR, which is insufficient for reliable authentication. Adding rotation improves TAR to 73.82%, and further adding local fingertip and knuckle geometry raises it to 80.57%. The largest single gain comes from including finger curl dynamics, bringing TAR to 92.93%, an improvement of 39 percentage points over position alone at comparable FAR. This suggests that IdeaVR does not simply learn where the hand moves, but how the hand begins to shape itself for the upcoming interaction.

We then compare the interaction-edge window against the full-interaction window (flagged action frames after t_{action}), both padded to the same fixed-length input, under two input conditions. With global hand position only, moving to the full interaction window improves TAR substantially (53.87% $\rightarrow$ 76.04%), showing that global trajectory alone is weak at the pre-grasp edge. With the full hand geometry, however, the same shift yields only a marginal improvement (92.93% $\rightarrow$ 95.26%, Δ2.33pp). This near-closure of the timing gap is the central result: rich hand-shape features substantially reduce the accuracy cost of early authentication. IdeaVR can authenticate before the object action takes effect without overly sacrificing accuracy, precisely the property required for interaction-edge protection.

7.3 Interaction-Level Takeover and Decision Latency

Table 4 reports takeover simulation results using held-out impostor interactions. For each enrolled user, impostor samples are randomly ordered into a takeover sequence and evaluated until the first rejection or a 10-interaction horizon is reached. Across folds, the first-action reject rate is 96.40% and the within-two-actions reject rate is 99.81%. The median number of interactions before the first rejection is 1, and the mean is 1.038, indicating that nearly all simulated takeovers are detected within the first two attempted object actions.

Table 4: Interaction-level takeover simulation using held-out impostor actions. Detected rate reports detection within a 10-interaction horizon.

Metric	Fold 1	Fold 2	Fold 3	Mean
Detected rate	100.00%	100.00%	100.00%	100.00%
First-action reject rate	96.28%	95.93%	97.00%	96.40%
Within-2-actions reject rate	99.75%	99.70%	99.98%	99.81%
Median interactions to reject	1	1	1	1
Mean interactions to reject	1.040	1.044	1.030	1.038

The per-decision hot-path latency averages 3.95 ms across all three object folds on the evaluation workstation described in Section 5.2, with ASV onset detection (1.07 ms) and CNN inference (1.79 ms) accounting for the majority of cost; one-time session-load and normalization costs are excluded. At 72 Hz, a single frame spans approximately 13.9 ms. The measured 3.95 ms decision latency remains within one frame interval on the evaluation workstation; end-to-end on-headset latency remains to be evaluated.

7.4 Preliminary Zero-Shot Cross-Day Stability

Five participants returned two weeks after the initial study. Models, normalization parameters, and thresholds learned from the full 20-participant Day 1 dataset were applied directly to Day 2 samples without recalibration. The cross-day evaluation achieved 73.46% TAR and 6.56% FAR (fold range 6.15%–7.31%), compared to 92.93% TAR and 3.46% FAR same-day, a 19.47 percentage point TAR reduction and 3.10 percentage point FAR increase. These differences are descriptive rather than matched estimates given the asymmetry in cohort size ($N = 5$ cross-day vs. $N = 20$ same-day), but the result shows that Day 1 decision boundaries retain useful identity information across sessions without recalibration. At the same time, the observed degradation indicates meaningful temporal drift and should not be interpreted as deployment-ready cross-session robustness without user-specific recalibration or adaptation.

For context, Liebers et al. [14] conducted a longitudinal study spanning up to eight weeks. For an approximately two-week gap between their first and third sessions, they reported weighted-F1 decreases of 18 percentage points for $N = 8$ and 23 percentage points for $N = 15$ using full-body game motion. Table 5 places these results alongside IdeaVR as contextual evidence of the magnitude of temporal drift in VR behavioral biometrics. The values are not directly comparable because the studies use different metrics, authentication formulations, motion signals, participant cohorts, and tasks. IdeaVR's 19.47-percentage-point difference should therefore be interpreted only as a descriptive change relative to its main same-day result.

Table 5: Contextual cross-session comparison over an approximately two-week interval.

System	Metric	Same-session	Cross-session	Change
Liebers et al. [14]	Weighted F1	83% / 86%	65% / 63%	−18 / −23 pp
IdeaVR	TAR	92.93%	73.46%	−19.47 pp

8 Discussion

8.1 Signal Source and Interaction-Edge Timing

The results support the core premise of IdeaVR: the interaction-edge provides enough information for effective identity checking using only short interaction-edge hand motion. The CNN maintains usable genuine acceptance while keeping false acceptance below the target operating region under the held-out-object protocol, and the takeover simulation shows that most impostor attempts are rejected at the first interaction.

The representation ablation clarifies where this signal comes from. Hand-root position alone performs poorly, but performance improves substantially as hand orientation, local finger geometry, and curl dynamics are added. This argues against the interpretation that IdeaVR is simply learning scene-relative reach location or a repeated movement path. The signal appears to come from how the hand begins to shape itself for the upcoming interaction, a combination of global approach behavior and local grasp preparation that is user-specific even before the application has recognized the action.

The window source comparison reinforces why the interaction-edge timing matters. Global hand-root position becomes more informative when the model observes the full interaction window, but that later trajectory is only available after the object action has already started. With rich hand-shape features, however, the gap between interaction-edge and full-interaction windows nearly closes (92.93% vs. 95.26% TAR), showing that IdeaVR can issue a decision near the interaction edge, before the application commits the action as trusted, with only a 2.33 percentage point reduction in TAR.

8.2 Role of ASV and Cross-Day Generalization

ASV should be interpreted as a lightweight window-selection mechanism rather than the main biometric classifier. Its role is to localize grasp formation within the rolling buffer using curl activity, without depending on absolute hand location. Figure 4 shows that ASV-selected window lengths vary across users and interaction types: most users cluster around short windows, but some exhibit substantially longer interaction-edge episodes, particularly for move interactions. This variation motivates adaptive window selection rather than assuming the same pre-event duration for every interaction.

In zero-shot cross-day transfer, TAR decreases by 19.47 percentage points while FAR increases to 6.56%, with no Day 2 recalibration or adaptation. This indicates meaningful temporal drift and shows that the same-day decision boundary does not transfer unchanged across sessions.

8.3 Limitations and Future Work

This study has several limitations. The cross-day evaluation is preliminary given the five-participant subset and single gap interval. The observed degradation indicates that practical deployment may require periodic or adaptive recalibration as user behavior and tracking conditions drift over time. Full longitudinal evaluation with adaptive recalibration remains important future work.

The current evaluation does not include skilled mimic attackers. Targeted imitation may increase false acceptance and degrade authentication performance. Future work should therefore evaluate informed mimic attacks and compare whether interaction-edge motion is more or less difficult to imitate than full action trajectories, in connection with prior work on behavioral biometric vulnerabilities [11].

An uninterrupted interaction provides only one authentication opportunity at its initial edge, while release, re-engagement, or object switching creates additional edges. Longer continuous interactions may require complementary authentication using cues such as non-dominant-hand or head motion.

Finally, IdeaVR was evaluated with right-hand interactions in a controlled seated setting and a 12-object set to isolate native hand-tracking authentication. These constraints limit generalization to varied postures, handedness, object geometries, occlusion, and tracking conditions. The framework is not inherently tied to one hand; future work should examine these factors and full on-headset deployment with live mitigation policies.

9 Conclusion

We presented IdeaVR, an interaction-edge authentication framework for VR that issues one identity decision per object interaction from continuous headset-native hand tracking. Rather than checking identity only at login or issuing background estimates at arbitrary intervals, IdeaVR buffers recent hand motion and authenticates near the interaction edge, before the application commits the action as trusted.

Through an object-controlled study with 20 participants and 12 virtual objects across four Zingg-inspired shape categories, a curl-only ASV detector localizes task-relevant grasp formation within the rolling buffer, and a 1D CNN authenticates the selected interaction-edge window with 92.93% TAR and 3.46% FAR on held-out objects. Representation ablations show that local hand geometry and curl dynamics contribute substantially beyond hand-root trajectory alone. A takeover simulation shows that 96.40% of impostor interactions are rejected at the first attempted action.

These findings support a broader design principle for VR security: authentication should not only ask *who* the user is, but *when* the decision can still protect the next action.

Acknowledgments

All study procedures were reviewed and approved by the University of Houston Institutional Review Board under ID STUDY00005734. This work was supported by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00463802).

References

[1] Arman Bhalla, Ivo Sluganovic, Klaudia Krawiecka, and Ivan Martinovic. 2021. MoveAR: Continuous Biometric Authentication for Augmented Reality Headsets. In *Proceedings of the 7th ACM on Cyber-Physical System Security Workshop (CPSS '21)*. Association for Computing Machinery, New York, NY, USA, 41–52. doi:10.1145/3457339.3457983

[2] Umberto Castiello. 2005. The Neuroscience of Grasping. *Nature Reviews Neuroscience* 6, 9 (2005), 726–736. doi:10.1038/nrn1744

[3] Ceenu George, Mohamed Khamis, Emanuel von Zezschwitz, Henri Schmidt, Marinus Burger, Florian Alt, and Heinrich Hussmann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *Proceedings 2017 Workshop on Usable Security (USEC '17)*. Internet Society, San Diego, CA, USA, 1–12. doi:10.14722/usec.2017.23028

[4] Godot XR Tools. [n. d.]. Godot XR Tools: Pickup Function. https://godotvr.github.io/godot-xr-tools/docs/pickup/. Accessed: 2026-06-08.

[5] Zhixiang He, Fengyuan Ran, Jing Chen, Yangyang Gu, Kun He, Ruiying Du, Ju Jia, and Cong Wu. 2025. HT-Auth: Secure VR Headset Authentication via Subtle Head Tremors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 3, Article 85 (Sept. 2025), 26 pages. doi:10.1145/3749980

[6] Lorna S. Jakobson and Melvyn A. Goodale. 1991. Factors affecting higher-order movement planning: a kinematic analysis of human prehension. *Experimental Brain Research* 86, 1 (1991), 199–208. doi:10.1007/BF00231054

[7] Amin Jalilov, Eunyong Cheon, and Ian Oakley. 2026. Leveraging Biometric-Rich Hand Gestures for Head-Mounted Display Authentication. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery.

[8] Marc Jeannerod. 1984. The Timing of Natural Prehension Movements. *Journal of Motor Behavior* 16, 3 (1984), 235–254. doi:10.1080/00222895.1984.10735319

[9] Alexander Kupin, Benjamin Moeller, Yijun Jiang, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Task-Driven Biometric Authentication of Users in Virtual Reality (VR) Environments. In *MultiMedia Modeling (Lecture Notes in Computer Science, Vol. 11295)*. Springer, 55–67. doi:10.1007/978-3-030-05710-7_5

[10] Feng Li, Jiayi Zhao, Huan Yang, Dongxiao Yu, Yuanfeng Zhou, and Yiran Shen. 2024. VibHead: An Authentication Scheme for Smart Headsets through Vibration. *ACM Transactions on Sensor Networks* 20, 4, Article 91 (May 2024), 21 pages. doi:10.1145/3614432

[11] Mingjun Li, Ashutosh Shivakumar, Natasha Kholgade Banerjee, and Sean Banerjee. 2025. Motion Forecasting Attacks on Behavioral Biometric Authentication Systems in Virtual Reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '25)*. Association for Computing Machinery. doi:10.1145/3756884.3766052

[12] Jonathan Liebers, Mark Abdelaziz, Lukas Mecke, Alia Saad, Jonas Auda, Uwe Gruenefeld, Florian Alt, and Stefan Schneegass. 2021. 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 (CHI '21)*. Association for Computing Machinery. doi:10.1145/3411764.3445528

[13] Jonathan Liebers, Sascha Brockel, Uwe Gruenefeld, and Stefan Schneegass. 2024. Identifying Users by Their Hand Tracking Data in Augmented and Virtual Reality. *International Journal of Human–Computer Interaction* 40, 2 (2024), 409–424. doi:10.1080/10447318.2022.2120845

[14] Jonathan Liebers, Christian Burschik, Uwe Gruenefeld, and Stefan Schneegass. 2023. 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 (VRST '23)*. Association for Computing Machinery. doi:10.1145/3611659.3615696

[15] Jonathan Liebers, Patrick Horn, Christian Burschik, Uwe Gruenefeld, and Stefan Schneegass. 2021. 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 (VRST '21)*. Association for Computing Machinery, New York, NY, USA, Article 22, 9 pages. doi:10.1145/3489849.3489880

[16] Jonathan Liebers, Patrick Laskowski, Florian Rademaker, Leon Sabel, Jordan Hoppen, Uwe Gruenefeld, and Stefan Schneegass. 2024. Kinetic Signatures: A Systematic Investigation of Movement-Based User Identification in Virtual Reality. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, 1–19. doi:10.1145/3613904.3642471

[17] Jonathan Liebers, Frieder Sykora, Niklas Pfützenreuter, Uwe Gruenefeld, David Goedicke, and Stefan Schneegass. 2025. User Identification in Virtual Reality through Behavioral Biometrics and the Influence of Colocated Interactions. In *Proceedings of the 31st ACM Symposium on Virtual Reality Software and Technology (VRST '25)*. Association for Computing Machinery, New York, NY, USA, Article 83, 13 pages. doi:10.1145/3756884.3765978

[18] Madhur Mangalam, Mathew Yarossi, Mateusz P. Furmanek, and Eugene Tunik. 2021. Control of Aperture Closure During Reach-to-Grasp Movements in Immersive Haptic-Free Virtual Reality. *Experimental Brain Research* 239, 5 (2021).

[19] Shrirang Mare, Mary Baker, and Jeremy Gummeson. 2016. A Study of Authentication in Daily Life. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, 189–206.

[20] Florian Mathis, John Williamson, Kami Vaniea, and Mohamed Khamis. 2020. RubikAuth: Fast and Secure Authentication in Virtual Reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*. Association for Computing Machinery, 1–9. doi:10.1145/3334480.3382827

[21] Robert Miller, Ashwin Ajit, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Realtime Behavior-Based Continual Authentication of Users in Virtual Reality Environments. In *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR '19)*. IEEE, 253–2531. doi:10.1109/AIVR46125.2019.00058

[22] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2020. 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 '20)*. IEEE, 311–316. doi:10.1109/VRW50115.2020.00070

[23] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2021. Using Siamese Neural Networks to Perform Cross-System Behavioral Authentication in Virtual Reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (IEEE VR '21)*. IEEE, 140–149. doi:10.1109/VR50410.2021.00035

[24] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2022. Combining Real-World Constraints on User Behavior with Deep Neural Networks for Virtual Reality (VR) Biometrics. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR '22)*. IEEE, 409–418. doi:10.1109/VR51125.2022.00060

[25] Robert Miller, Natasha Kholgade Banerjee, and Sean Banerjee. 2022. Temporal Effects in Motion Behavior for Virtual Reality (VR) Biometrics. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR '22)*. IEEE, 563–572. doi:10.1109/VR51125.2022.00076

[26] Alec G. Moore, Tiffany D. Do, Nicholas Ruozzi, and Ryan P. McMahan. 2023. Identifying Virtual Reality Users Across Domain-Specific Tasks: A Systematic Investigation of Tracked Features for Assembly. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR '23)*. IEEE, 396–404. doi:10.1109/ISMAR59233.2023.00054

[27] Gonzalo Munilla Garrido, Vivek Nair, and Dawn Song. 2024. SoK: Data Privacy in Virtual Reality. *Proceedings on Privacy Enhancing Technologies* 2024, 1 (2024), 21–40. doi:10.56553/popets-2024-0003

[28] Vivek Nair, Wenbo Guo, Justus Mattern, Rui Wang, James F. O'Brien, Louis Rosenberg, and Dawn Song. 2023. Unique Identification of 50,000+ Virtual Reality Users from Head & Hand Motion Data. In *32nd USENIX Security Symposium (USENIX Security '23)*. USENIX Association, 895–910.

[29] Vivek Nair, Wenbo Guo, Rui Wang, James F. O'Brien, Louis Rosenberg, and Dawn Song. 2024. 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* 30, 5 (May 2024), 2239–2246. doi:10.1109/TVCG.2024.3372087

[30] Ken Pfeuffer, Matthias J. Geiger, Sarah Prange, Lukas Mecke, Daniel Buschek, and Florian Alt. 2019. Behavioural Biometrics in VR: Identifying People from Body Motion and Relations in Virtual Reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300340

[31] Christian Rack, Tamara Fernando, Murat Yalcin, Andreas Hotho, and Marc Erich Latoschik. 2023. Who Is Alyx? A New Behavioral Biometric Dataset for User Identification in XR. *Frontiers in Virtual Reality* 4 (2023). doi:10.3389/frvir.2023.1272234

[32] Marco Santello, Martha Flanders, and John F. Soechting. 1998. Postural Hand Synergies for Tool Use. *Journal of Neuroscience* 18, 23 (1998), 10105–10115. doi:10.1523/JNEUROSCI.18-23-10105.1998

[33] Sophie Stephenson, Bijeeta Pal, Stephen Fan, Earlence Fernandes, Yuhang Zhao, and Rahul Chatterjee. 2022. SoK: Authentication in Augmented and Virtual Reality. In *2022 IEEE Symposium on Security and Privacy (IEEE S&P '22)*. IEEE, 267–284. doi:10.1109/SP46214.2022.9833742

[34] Unity Technologies. [n. d.]. XR Interaction Toolkit: Interaction Overview. https://docs.unity.cn/Packages/com.unity.xr.interaction.toolkit@3.2/manual/architecture.html. Accessed: 2026-06-08.

[35] Unity Technologies. [n. d.]. XR Interaction Toolkit: XR Grab Interactable. https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@2.0/manual/xr-grab-interactable.html. Accessed: 2026-06-08.

[36] Theodor Zingg. 1935. *Beitrag zur Schotteranalyse*. Ph. D. Dissertation. ETH Zurich.