Using Siamese Neural Networks to Perform Cross-System Behavioral Authentication in Virtual Reality

Robert Miller* Natasha Kholgade Banerjee† Sean Banerjee‡

Clarkson University
Potsdam, NY

ABSTRACT

In this paper, we provide an approach on using behavioral biometrics to perform cross-system high-assurance authentication of users in virtual reality (VR) environments. VR is currently being explored as a critical tool to ensure seamless delivery of essential services, such as education, healthcare, and personal finance, while enabling users to work from home environments. Due to the sensitive nature of personal data generated, VR applications for essential services need to provide secure access. Traditional PIN or password-based credentials can be breached by malicious impostors, or be handed over by an intended user of a VR system to a confederate to assist the intended user in completing a task, e.g., an exam or a physical therapy routine. Existing approaches that use the behavior of the user in VR as a biometric signature fail when users provide enrollment and use-time data on different VR systems. We use Siamese neural networks to learn a distance function that characterizes the systematic differences between data provided across pairs of dissimilar VR systems. Our approach provides average equal error rates (EERs) ranging from 1.38% to 3.86% for authentication using a benchmark dataset that consists of 41 users performing a ball-throwing task with 3 VR systems—an Oculus Quest, an HTC Vive, and an HTC Vive Cosmos. To compare to prior approaches in VR biometrics, we also obtain average accuracies for the task of identification, where given an input user's trajectory in a use-time VR system, we use Siamese networks to return the user with the top matching trajectory in an enrollment VR system as the label. We report identification results ranging from 87.82% to 98.53% with average improvements of 29.78%±8.58% and 30.78%±3.68% over existing approaches that use generic distance matching and fully convolutional networks on the enrollment dataset respectively.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Security and privacy—Security services—Authentication—Biometrics

1 INTRODUCTION

The consumer space is currently experiencing a transition point where traditionally in-person activities such as education, personal finance, retail, work-from-home, and healthcare are being converted to virtual or hybrid mode, a move that has been accelerated by the COVID-19 pandemic. To enable seamless experience akin to the real-world, virtual reality (VR) is being explored as a critical tool for a diverse range of essential services such as virtual education [11, 27, 32, 57, 70, 75], retail [56, 81], personal finance [8, 77], virtual remote teleoperation and driving [33, 41, 50, 51, 67], and healthcare [7, 13, 18, 39, 42, 52, 69]. Since a large quantity of sensitive user data is likely to be generated, it is critical to ensure that VR applications for essential services are secured against access by unauthorized individuals. A number of approaches have investigated transferring traditional PIN, password, and two-factor authentication systems to VR [5, 6, 19, 21–24, 54, 83]. However, such approaches face vulnerabilities from two perspectives. First, they can be breached by malicious external agents that gain access to credentials of a genuine user. Second, they enable deliberate circumvention by an authorized user providing credentials to a confederate to, for instance, cheat an examination or a therapy session. Deliberate circumvention renders the interventions provided by the VR application ineffective, thereby hindering the efforts of organizations administering the interventions. Recognizing the drawbacks of traditional credentials, a large body of work has emerged on using user behavior in VR—particularly motion trajectories of VR headsets and hand controllers—as a biometric signature [2, 31, 43–47, 49, 53, 55]. Recent approaches have demonstrated high accuracies of upwards of 95% [43, 47, 53]. However, a fundamental limitation of prior work [2, 31, 43–46, 49, 53, 55] is that enrollment data is provided on the same VR system as use-time data.

In this work, we provide an approach to perform cross-VR-system authentication, where enrollment data is provided on one VR system, e.g., an Oculus Quest, and the user interacts with a separate VR system at use-time, e.g., an HTC Vive. Users are likely to interact with one VR system at an organization, e.g., their workplace, their school, and/or their clinic, and have access to an alternate personal VR system in home environments. A user may be asked to provide enrollment data using the organization's VR system while being supervised during enrollment, and may need to access the organization's VR application using their personal VR system. One option may be to recommend re-enrollment with the personal VR system, either by returning to the organization for supervised re-enrollment, or by using remote guidance for re-enrollment without supervision. The former is cumbersome in the long run, and does not maintain seamlessness of authentication with rapid upgrades of VR systems. The latter increases the likelihood of incorrect enrollment, and more critically, lowers security if the user asks a confederate to enroll, thereby enabling permanent circumvention of the system by the confederate. Given these usability and security concerns, it is crucial to provide generalizable cross-system authentication that does not require re-enrollment using a personal VR system.

To date, the work of Miller et al. [47] is the only study that investigates cross-system VR biometrics with data provided on three systems—an HTC Vive, an Oculus Quest, and an HTC Vive Cosmos. They use a generic distance metric to identify users by matching use-time trajectories to enrollment trajectories, yielding cross-system accuracies that are too low for deployment in real-world consumer spaces, with highest average accuracy of 85.12%. Prior same-system learning-based approaches [43, 45, 55] that train solely on enrollment data cannot be directly extended to cross-system authentication, since tracking methods and physical characteristics of the headset and hand controllers, such as mass distribution, size, and aspect ratio, generate system-dependent biases in the data. The biases yield inter-system differences to which same-system methods are agnostic. As we demonstrate with Mathis et al. [43], same-system learning-

*e-mail: romille@clarkson.edu
†e-mail:nbanerje@clarkson.edu
‡e-mail:sbanerje@clarkson.edu

2642-5254/21/$31.00 ©2021 IEEE
DOI 10.1109/VR50410.2021.00035

Study	Classifier	Users	Activities	Features	Acc.
Kupin et al. [31]	Nearest Neighbor	14	Ball-throw	Position of right controller	92.86%
Pfeuffer et al. [55]	Random Forests	22	Point, grab, walk, type	Position, orientation, linear velocity, angular velocity for both controllers and headset	44.44%
Ajit et al. [2]	Perceptron	33	Ball-throw	Position & orientation for both controllers and headset	93.03%
Olade et al. [53]	Nearest Neighbor	25	Grab, rotate, drop	Position & orientation for both controllers and headset + eye position	98.6%
Mathis et al. [43]	FCNs	23	Point at 3D cube	Position & orientation for both controllers	98.91%

Table 1: Summary of related work in same system VR biometrics (FCNs = fully convolutional networks). For all approaches, enrollment and use-time data is provided on the HTC Vive. Acc: accuracy at identifying users from VR behavior.

based methods generate low accuracies for user identification on the use-time system when trained with data from the enrollment system.

Our approach overcomes challenges with prior learning-based methods by using metric learning to characterize systematic variabilities induced by system-specific biases. Since systematic variabilities may be non-linear in nature, we use Siamese neural networks to learn a system-to-system distance metric between headset and hand controller trajectories of an enrollment VR system and a separate use-time VR system. Our work enables authentication by comparing distances against a threshold and identification by returning the user with the minimum distance value. Our approach requires that a training set of users has provided data and ground truth labels on both systems. Test users need only provide ground truth on the enrollment system. Our approach offers two advantages over prior work. First, it is targeted to provide high performance for cross-system authentication, unlike prior learning-based methods [43, 55] where training on one system yields poor accuracy on another system, and unlike methods that use generic distance metrics [2, 31, 47, 53] that fail to capture inter-system variabilities. Second, by virtue of learning a distance metric rather than providing user IDs, our approach is readily generalizable to novel users, unlike the work of Pfeuffer et al. [55] and Mathis et al. [43] where re-training of the neural networks is required every time new users are added to the environment.

We demonstrate results of authentication, i.e., verifying that a user's ID is as claimed, and identification, i.e., recognizing an unknown user's ID, using the dataset of Miller et al. for data provided on the HTC Vive, Oculus Quest, and HTC Vive Cosmos. We demonstrate lowest average equal error rates (EERs) for authentication ranging from 1.38% when the Vive is used at use-time and the Quest at enrollment, to 3.86% when the Cosmos is used at use-time and the Vive at enrollment. We show highest average identification accuracies ranging from 87.82% when the Cosmos is used at use-time and the Vive at enrollment, to 98.53% when the Vive is used at use-time and the Quest at enrollment. We demonstrate an average of 29.78%$\pm$8.58% boost in accuracy over the work of Miller et al. We perform comparisons to the best performing learning-based approach, particularly the work of Mathis et al. [43] who demonstrate highest accuracies on same-system authentication using fully convolutional networks (FCNs). We demonstrate that the FCNs from Mathis et al. provide average accuracies ranging from 50.42% when the Cosmos is used at use-time and the Quest at enrollment, to 72.62% when the Vive is used at use-time and the Quest at enrollment. By learning the distance metric that characterizes inter-system differences between two VR systems, our approach provides a 30.78%$\pm$3.68% boost in accuracy over the work of Mathis et al. We also perform evaluation of the robustness and generalizability of our work, and of the factors influencing authentication accuracy.

2 Related Work

Authentication in VR. Early work in VR authentication has either directly incorporated traditional PIN or password-based techniques into VR environments [19, 22, 23, 54, 83], or extended traditional credentials to include arrangements of VR objects [5, 6, 21, 24]. While PIN/password methods can be used in cross-system authentication, they are rendered unsafe once the unintended user gains access to the PIN/password. Initial work in behavior-based VR authentication [35, 49, 61, 65, 68, 82] has been limited to head motions in Google Glass and Google Cardboard, platforms with limited range of motion due to the absence of hand controllers that are integral to immersive VR. More recently, a growing movement has emerged on performing behavior-based authentication using hand-controller based VR systems, spurred by the proliferation of high-end VR systems and the recognition of the need for VR systems in essential services. Table 1 provides a summary of the prior work in behavior-based VR biometrics when enrollment and use-time data is provided on the same system, particularly the HTC Vive [2, 31, 43, 53, 55]. The approaches have been complemented with real-time implementations of behavior-based authentication [44, 46].

Using existing behavior-based work in VR biometrics without modification is likely to yield low performance, as algorithms based on classifiers [43, 44, 55] are not structured to learn the relationships between the enrollment and use-time VR systems. Methods that use distances between raw trajectories or higher-level features [2, 31, 47, 53] are unlikely to provide high within-user cross-system matches in VR when differences in geometry and weight distribution of devices cause user behavior to be modified across systems. The work of Miller et al. [47] is the first and to date only work to investigate VR biometrics in multiple VR systems. They capture data for 46 users performing the ball-throw from Kupin et al. [31], of which 41 users are right-handed and 5 are left-handed. They perform user identification with the 41 right-handed users using the perceptron proposed in Ajit et al. [2]. Their work uses position, orientation, linear velocity, and angular velocity from the controllers and the headset, and the trigger state of the right controller as features. While within-system accuracies are upwards of 90% and reach 97% for the HTC Vive, cross-system accuracies are significantly lower, with maximum accuracies ranging from around 58.54% to 85.12%.

Table 1 also demonstrates that in general, the number of users spanning VR datasets is low. Collecting data for biometrics in VR environments is challenging since widespread adoption and use of VR devices for mass consumer applications is yet in an embryonic state. Groups performing VR biometric studies conduct lab-based collections of VR interactions, unlike traditional desktop and smartphone applications where data can be collected on personal devices. We use the dataset from the work of Miller et al. [47] as with 46 users each using 3 VR systems, the dataset is the largest of its kind, and has the highest diversity of VR systems, containing systems that use lighthouses and cameras to track device motion.

Gait-based authentication. In using motion trajectories from hand-held and head-mounted devices, our work is related to research in authentication using gait, i.e., the cyclic walking motions of a subject as extracted from accelerometers on smart and wear-

able devices. Gait-based authentication is a well-studied area of research, with a large range of surveys conducted over the past two decades [12, 15, 20, 60, 74]. To the best of our knowledge, the approach of Hoang et al. [26] is the only work that performs cross-sensor gait authentication using accelerometer data. Their work shows an accuracy of 91.33% on identification of 14 subjects by comparing gait features from a Nexus One with features from an LG Optimus G. Data was collected by physically binding the two devices together. The physical binding enables signals for the same walk instance from the two devices to be highly correlated. The work lacks analysis of authentication performance when one device is used in a separate walk instance from the other device, where natural intra-user variability across walk instances is likely to be a confounding factor. The work of Hoang et al. cannot be directly extended to cross-system authentication in VR, as space and usability restrictions prevent headsets or hand controllers belonging to different VR systems from being physically co-mounted on the user. Additionally, trajectories for the large range of tasks that users are likely to perform in VR, e.g., throwing, pointing, swiping, and object-moving may not have characteristic patterns such as repetitions, crests, and troughs generated by walking motions that are used as information sources in Hoang et al.

Cross-Domain Biometrics. A large body of work exists in cross-domain biometrics particularly in the domain of matching fingerprints from different contact-based sensors [4, 62, 63], matching contact-based to contactless fingerprints [10, 36–38], iris/periocular biometrics within the same spectrum [29, 58, 64, 79], and cross-spectrum periocular biometrics [3, 28, 59, 66]. Most approaches for fingerprint matching [4, 10, 36, 38, 62, 63] rely on the presence of distinctive features such as minutiae or interest points [40] in a person's fingerprint. These methods cannot be directly transferred to VR behavior, as distinctive reliable regions of a user's trajectory are not readily identifiable *a priori*. Most work in iris/periocular biometrics has similarly used classifiers with hand-crafted features that are specific to 2D images, e.g., histograms of gradients [64, 66], local binary patterns [3, 28, 59, 64], interest points [3, 64], Gabor features [3], Fourier transforms on Laplacian pyramids [58], and ordinal measures [79], fused with features using linear dictionary learning techniques [28]. These features are not immediately adaptable to behavior trajectories with non-linear information such as orientation. The approach of Kandaswamy et al. [29] proposes a multi-source deep transfer learning approach, where the accuracy of a neural network architecture at performing user identification from iris images for a target (use-time) sensor is improved through an iterative approach by transferring weights learnt from multiple sources and fine-tuning through re-training. Their approach requires a prior dataset for a user to re-train the target network, and thereby cannot work when no prior data exists, as in our case where users provide use-time data on a new device for the first time after providing enrollment data on an alternative device. Additionally, since they provide user identities rather than match scores, their approach is not generalizable to new users without network re-training, and their approach cannot be re-targeted to authentication. Lin and Kumar [37] use Siamese networks on hand-crafted features, particularly ridge and minutiae maps, for matching contactless to contact-based fingerprints, and provides best EER of 7.11%. In our work, we avoid using hand-crafted features to prevent biasing, and enable the network to automatically learn relevant features from the input data.

Siamese Neural Networks in Behavior Authentication. Siamese networks have been used for authentication of users interacting with smart devices [9, 14, 16, 80], and in handwriting/signature verification [1, 71, 73, 78]. Prior work in authentication [9, 14, 80] uses Siamese networks to generate distances for a user using the same device at enrollment and use-time, as opposed to different systems as in our work. Work by Fan et al. [16] uses a Myo armband to use EMG data for authentication. While the authors demonstrate the ability to work across multiple devices, the approach requires all users to wear a Myo armband when interacting with each device.

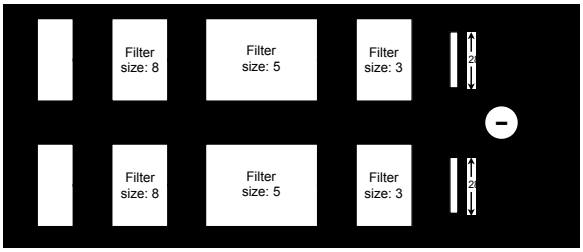

Figure 1: Siamese neural network architecture with FCN limbs. The output of Layer k is generated by performing 1D convolution with the input to the layer using F_k filters, followed by batch normalization and application of the ReLU activation function. The outputs from Layer 3 are pooled into a 128-dimensional feature vector using global average pooling (GAP). The output of the Siamese network is the Euclidean distance between the output vectors of the two limbs.

3 Approach for Cross-System VR Biometrics

Our approach uses Siamese networks to represent a distance function between trajectories from an enrollment and a use-time VR system. As shown in Figure 1, the network consists of limbs that each generate low-dimensional embeddings using headset and hand controller trajectory features from two input instances. The network is trained such that the Euclidean distance between the low-dimensional embeddings from input instances of the same user is low.

Dataset. We use the dataset of Miller et al. [47], the largest VR biometrics dataset in terms of number of users, and the only one with multiple VR systems. The dataset consists of device trajectories recorded for 46 users interacting with three off-the-shelf VR systems, particularly the HTC Vive, Oculus Quest, and HTC Vive Cosmos in that order over a period of several days. Each user uses each system on two separate days, and provides 10 ball-throws on each day. Each throw is recorded for 3 seconds at the frame rate of the corresponding system. Of the 46 users, 41 are right-handed. The Vive uses a lighthouse-based tracking, while the Quest and Cosmos perform tracking using multiple cameras on the headset. We only use the 41 right-handed users to provide comparisons to Miller et al.

Input data preparation. Since the frame rate of the Quest at 75 FPS differs from that of the Vive and Cosmos at 45 FPS, we re-sample the data from the Quest to have the same frame rate as the HTC systems. Given trajectories from the Quest controllers and headset, we re-sample the position and orientation using linear interpolation and spherical linear interpolation respectively. Upon re-sampling, the trajectories from all three VR systems contain 135 trajectory point samples, with right controller position, right controller orientation, left controller position, left controller orientation, headset position, and headset orientation as features. While the dataset of Miller et al. also contains the trigger state of the right controller as a feature, we do not use the trigger state in this work. Prior to use, we normalize the positions of each device's trajectory to have zero mean and unit variance, and we center the positions by subtracting the bounding box center of the trajectory.

Network Architecture. Figure 1 demonstrates the architecture of the Siamese network used in this work. We choose to use FCNs, as they provide the highest accuracy for within-system user identification in Mathis et al. [43], and the second highest accuracy for classification on a variety of time-series tasks [17]. We use a 3-layer FCN for each limb, based on the results from Wang et al. [76]. The input to the first layer consists of a $T \times P$ time series matrix, where T represents the number of time samples within the input trajectory,

E/U	Q1/V1	Q1/V2	Q2/V1	Q2/V2	Q/V	Q1/C1	Q1/C2	Q2/C1	Q2/C2	Q/C	V1/C1	V1/C2	V2/C1	V2/C2	V/C
R,L / P,O	2.93	3.11	1.71	2.37	2.53	***3.24***	***3.24***	3.90	5.16	3.89	5.51	4.95	4.36	5.55	5.09
R,H / P,O	2.48	1.43	1.71	1.80	1.86	3.52	4.53	3.34	***1.77***	3.29	***3.63***	***3.17***	**3.17**	5.46	***3.86***
R,L,H / P,O	**2.22**	***1.38***	***0.68***	***1.28***	**1.39**	3.55	3.91	**2.44**	2.61	***3.13***	4.88	3.49	4.52	5.89	4.70
R,L / P	4.88	3.02	3.17	4.15	3.80	6.21	7.36	7.56	9.02	7.54	5.01	6.56	4.26	6.10	5.48
R,H / P	3.39	5.16	2.68	2.44	3.42	6.77	5.73	4.63	5.85	5.75	4.06	4.45	4.63	**4.63**	4.44
R,L,H / P	2.71	3.75	3.05	2.52	3.01	5.37	6.73	6.45	5.86	6.10	5.95	5.66	5.41	5.25	5.57

Table 2: Average authentication EERs as percentages using 100% of the trajectory points. E = **E**nrollment, U = **U**se-time, Q = Oculus **Q**uest, V = HTC **V**ive, C = HTC Vive **C**osmos, 1 and 2 refer to the day on which the data was captured, R = **R**ight controller, L = **L**eft controller, H = **H**eadset, P = **P**osition, O = **O**rientation. Q/V, Q/C, and V/C represent averages of values in the preceding four columns. EERs in bold are the lowest in their column. Combined bold/italicized EERs are the lowest amongst 100% and 80% of the trajectory points (80% points found in supplementary).

and P represents the number of features at each time sample. The first layer consists of $F_1 = 128$ filters applied using 1-dimensional (1D) convolution with a kernel size of $K_1 = 8$ followed by batch normalization and application of the rectified linear unit (ReLU) activation function to generate an output of size $T \times F_1$. The second layer consists of $F_2 = 256$ 1D convolutional filters with a kernel size of $K_2 = 5$, and generates an output of size $T \times F_2$ after batch normalization and ReLU activation. For the third layer, we use $F_3 = 128$ 1D convolutional filters with a kernel size of $K_3 = 3$ to generate an output of size $T \times F_3$ after batch normalization and ReLU. The output of the third layer is averaged using global average pooling (GAP) to obtain the low-dimensional embedding vector with $F_{\text{out}} = 128$ elements for the corresponding limb. The network output corresponds to the Euclidean distance between the embeddings.

Training. We train the Siamese network using the enrollment and use-time data from a training set of M users. Each user has n trajectories for a total of Mn training trajectories. To train the network, we form M^2n^2 trajectory pairs, where Mn^2 of the pairs come from the same user, while the remaining $M(M-1)n^2$ come from different users. Training involves optimization of the contrastive loss [25]. For the i^{th} pair of time series inputs $\mathbf{X}_1^i$ and $\mathbf{X}_2^i$ and the ground truth output Y^i, the contrastive loss is given as $\left(1-Y^i\right)\left(H_W\left(\mathbf{X}_1^i,\mathbf{X}_2^i\right)\right)^2 + Y^i \max\left(0, m - H_W\left(\mathbf{X}_1^i,\mathbf{X}_2^i\right)\right)^2$, where $H_W\left(\mathbf{X}_1^i,\mathbf{X}_2^i\right) = \left\|G_W\left(\mathbf{X}_1^i\right) - G_W\left(\mathbf{X}_2^i\right)\right\|$ is the Euclidean distance between the low-dimensional embeddings $G_W\left(\mathbf{X}_1^i\right)$ and $G_W\left(\mathbf{X}_2^i\right)$, W represents the weights shared across both limbs, and m represents a margin. Inputs $\mathbf{X}_1^i$ and $\mathbf{X}_2^i$ are fed to the first and second limbs respectively. The contrastive loss [25] has been demonstrated to improve discrimination between dissimilar pairs by forcing them to play a role in optimization of the loss only if their radius is within m. We set m to 1.0. We train using the Adam optimizer [30], with a learning rate $\alpha = 0.001$ and moment parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We use a batch size of 64 during training, and we train for 100 epochs. We use a leave-one-out approach in this work to maximize the quantity of training data available, which yields 41 training folds that we split over 23 machines described in the supplementary material. Each fold takes 50 to 90 minutes to train.

Testing. We perform tests to evaluate the tasks of authentication, i.e., verifying that the user's identity matches their claimed identity with the assumption that the user has provided some known credential, and identification, i.e., returning the identity of an unknown user. For both tasks, we use the trained Siamese network to compute the distance between the use-time trajectory of a test user, and the enrollment trajectories of all users in our dataset. By including all users (i.e., training and test) in the enrollment list, we evaluate the performance of our approach when the input user is compared to a large set of enrollment users. We ensure that neither enrollment nor use-time trajectories for the test user are used during training. To authenticate a test user, we return for each enrollment user the distance to the closest trajectory. We then obtain for a particular choice of threshold, the true accept rate (TAR), i.e., the rate at which the user's claimed identity is correctly accepted, and the false accept rate (FAR), i.e., the rate at which an impostor is incorrectly allowed through. We vary the threshold from 0 to the maximum distance to obtain receiver operating characteristic (ROC) curves, and obtain the EER as being the value of the FAR at which FAR=1-TAR. The EER provides a measure of the strength of a system at correctly rejecting impostors, while keeping false positives low for usability. To assess identification capability, we obtain the user corresponding to the nearest enrollment trajectory as the label for the use-time trajectory, and evaluate accuracies at correctly identifying the user.

4 Results

We demonstrate results for evaluation of our proposed approach using the 41 right-handed users in the Miller et al. [47] dataset by performing leave-one-user-out cross-validation. For the i^{th} cross-validation fold, we leave out all $n = 10$ enrollment and use-time trajectories for the i^{th} user, and train a Siamese network with the enrollment and use-time trajectories of the remaining $M = 40$ users using the training procedure discussed in Section 3. During testing, we obtain the distance for each of the n use-time trajectories of the i^{th} user and the $(M+1)n = 410$ enrollment trajectories for all users, and use the test procedure discussed in Section 3. For authentication, we show best average EERs over all epochs for the neural networks from our work, summarized in Table 2. The table provides EERs by using position (P) and orientation (O) as well as using position only (P) from various combinations of the right controller (R), left controller (L), and headset (H). Column headers provide VR systems compared for enrollment (E) versus use-time (U). The numerals 1 and 2 represent trajectories captured on the first and second day that the system was used. We compare first and second day trajectories for the Vive (V) against the Quest (Q), the Cosmos (C) against the Quest, and the Cosmos against the Vive. Other system pairings are not examined in order to respect temporal order, i.e., to ensure enrollment is done using systems used prior to use-time. In Table 3, we demonstrate comparisons of our accuracy results to the approaches of Miller et al. and Mathis et al. [43] for each evaluation condition. Both methods perform user identification. To replicate the approach of Mathis et al., we train their best performing architecture, i.e., the FCN, to output the user label using the enrollment data of all 41 users. We set up the Mathis et al. FCN to have the same structure as one limb of the Siamese network, with the exception that the Mathis et al. FCN ends with a dense layer consisting of 41 softmax activation outputs ranging from 0 to 1. We retain the output with the maximum value as the recognized user ID. We train the Mathis et al. FCN for 2000 epochs, while keeping the remaining hyperparameters the same as in Section 3. We report averages over

E/U		Q1/V1	Q1/V2	Q2/V1	Q2/V2	Q/V	Q1/C1	Q1/C2	Q2/C1	Q2/C2	Q/C	V1/C1	V1/C2	V2/C1	V2/C2	V/C
R,L / P,O	Ours	94.87	93.65	96.09	97.31	95.48	***90.73***	***93.41***	86.58	84.87	88.90	83.17	87.07	87.56	85.12	85.73
	[43]	73.65	60.73	85.12	70.97	72.62	45.12	46.82	50.48	46.34	47.19	51.95	48.53	53.17	48.53	50.55
	[47]	48.78	45.85	60.24	52.20	60.73	28.05	27.56	25.61	23.71	33.23	57.07	53.66	45.12	44.15	57.62
R,H / P,O	Ours	94.63	98.04	94.39	97.31	96.09	84.39	81.70	91.70	***93.90***	87.92	85.85	86.58	**92.19**	***86.82***	***87.82***
	[43]	57.07	50.24	60.73	56.58	56.16	44.87	34.87	50.00	45.85	43.90	48.53	43.17	57.31	49.26	49.57
	[47]	54.15	48.05	63.66	56.10	47.13	24.39	25.12	30.98	29.27	27.44	64.88	60.73	53.90	49.27	38.23
R, L,H / P,O	Ours	**96.58**	***99.26***	***100.00***	**98.29**	***98.53***	85.36	84.63	**93.17**	92.19	***88.84***	85.85	***87.31***	87.07	85.60	86.46
	[43]	64.14	57.31	70.24	63.65	63.84	48.29	45.12	53.65	54.63	50.42	55.60	47.56	61.70	51.70	54.14
	[47]	55.61	49.02	74.39	63.90	55.49	25.85	29.76	42.93	34.39	33.23	64.63	60.24	55.12	50.49	57.20
R,L / P	Ours	86.34	90.24	92.68	86.58	88.96	77.31	73.41	69.75	68.29	72.19	80.73	76.34	79.51	78.04	78.66
	[43]	62.19	50.73	79.26	67.56	64.94	38.29	43.90	46.34	45.36	43.47	46.82	47.56	49.26	48.04	47.92
	[47]	63.41	47.32	50.24	41.71	50.67	35.61	32.93	30.98	27.07	31.65	45.37	46.34	41.46	35.37	42.14
R,H / P	Ours	94.87	88.04	91.95	93.65	92.74	79.26	79.02	83.65	78.53	80.12	***89.26***	83.65	87.56	82.19	85.67
	[43]	69.26	54.87	78.29	62.92	66.34	39.51	41.95	51.70	46.34	44.88	57.31	53.41	64.39	52.92	57.01
	[47]	63.17	49.02	42.44	33.90	47.13	30.98	30.49	22.93	22.20	26.65	48.29	47.80	28.54	28.29	38.23
R, L,H / P	Ours	93.90	87.56	92.43	92.68	91.64	81.21	79.26	73.65	80.24	78.59	81.21	78.29	80.00	81.70	80.30
	[43]	69.02	58.78	80.97	67.80	69.14	47.07	48.53	53.17	50.97	49.94	55.12	51.46	61.70	53.41	55.42
	[47]	60.98	52.20	53.17	43.41	52.44	36.59	34.88	31.95	29.27	33.17	50.24	48.54	36.83	38.29	43.48
Best	[47]	68.05	59.51	82.68	71.46	70.43	41.95	41.95	50.73	44.63	44.82	66.10	62.44	63.41	58.54	62.62

Table 3: Average identification accuracies as percentages using 100% of the trajectory points. Same conventions used in Table 2 apply here, except that highest accuracies are bolded/italicized.

all $n = 10$ trajectories for all 41 users using our approach and the methods of Miller et al. and Mathis et al. We show the best average accuracy over all epochs for our work and Mathis et al. The last row of Table 3 provides maximum accuracies obtained by Miller et al. upon examining 2^{13} feature combinations.

Evaluation of results. For user authentication, our approach provides lowest EERs when Quest trajectories are used at enrollment and Vive trajectories at use-time, with an average across all day combinations of 1.39% when all features and trajectory points are used, as shown in Table 2. EERs for comparing Cosmos use-time trajectories against Quest enrollment and Vive enrollment trajectories are higher, with lowest across-day averages obtained at 3.13% and 3.86% respectively. We discuss potential reasons for differences in performance between system pairings in Section 5. Similar trends are observed for user identification, as shown in Table 3, where the highest average accuracy across all days is 98.53% for comparing Quest enrollment to Vive use-time when all features are used, and average accuracies across all captures for Cosmos use-time trajectories are lower at 88.84% and 87.86% when compared to Quest and Vive respectively as enrollment. ROC curves for authentication and confusion matrices for identification are shown in the supplementary material. Prior approaches that perform matching using nearest-neighbor distances with the ball-throwing motion [2, 31, 47] demonstrate higher accuracies when the later portions of the trajectory corresponding to the non-directed hand return are removed, since non-directed motions may demonstrate high variability and may have limited use as a behavior signature. In the supplementary, we provide EERs and accuracies for evaluating Siamese networks when 80% of the trajectory points are retained. Using 80% of the points with all devices and features yields average authentication EERs of 1.38%, 3.33%, and 4.73%, and identification accuracies of 84.93%, 86.95%, and 97.68% for Vive-enrollment-Cosmos-use, Quest-enrollment-Cosmos-use, and Quest-enrollment-Vive-use respectively. The comparable performance using 80% and 100% of the points indicate that the Siamese network is successful at automatically eliminating unnecessary portions of the trajectory. The worst performance—with highest average EERs of 3.80%, 7.54%, and 5.57%, and lowest average accuracies of 88.96%, 72.19%, and 78.65%— is obtained when solely position features are used, and when the head is eliminated for all except EER of Vive-enrollment-Cosmos-use, indicating that extremity orientations and head motions during an intentional action contribute as a signature. As a baseline, we apply our approach to within-system authentication with position and orientation. We obtain lowest EERs for Quest, Vive, and Cosmos of 0.73%, 0.99% and 1.57% respectively, and highest accuracies of 99.75%, 99.51%, and 98.04% respectively. Best results are obtained when all devices are used for Quest, right controller and headset for Vive, and right and left controller for Cosmos.

Comparison to Mathis et al. [43]. Our identification results demonstrate an overall improvement of 30.78%±3.68% over the approach of Mathis et al., with the average taken over accuracies for all system pairings, feature sets, and trajectory points. Their approach shows highest average accuracies of 73.96%, 54.57%, and 57.80% for Quest-enrollment-Vive-use, Quest-enrollment-Cosmos-use, and Vive-enrollment-Cosmos-use when 80% of the trajectory points are used. Quest-Cosmos and Vive-Cosmos accuracies are highest when using all devices and features, while Quest-Vive accuracies are highest when the headset is left out. The low performance of Mathis et al. is explained by their approach being agnostic to systematic differences between two systems. Since the training data does not contain information from the use-time system, the networks are trained to be biased toward the enrollment system. On average, Mathis et al. show lower accuracies for the same feature set when all points are used in comparison to when 80% of the points are

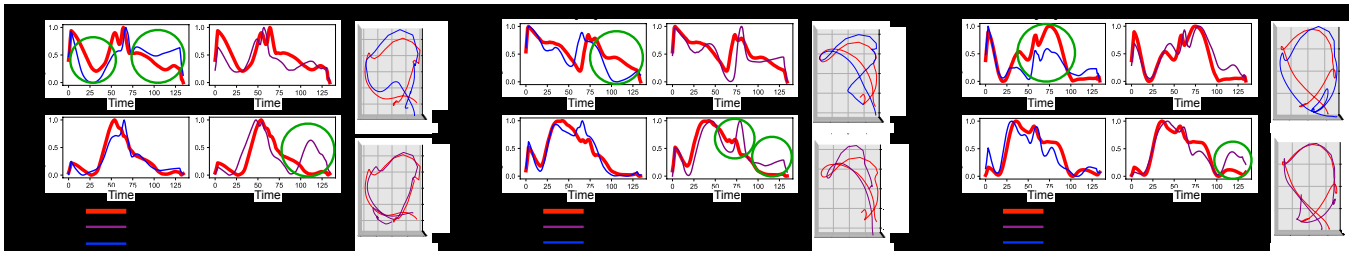

Figure 2: Comparison of results from our work against Mathis et al. [43]. Each block shows average activation maps for the closest trajectory for the user returned by Mathis et al. in the first column and the best enrollment trajectory returned using our approach in the second column, together with 3D plots in the third column demonstrating use-time trajectories in red, and matched enrollment trajectories returned using our approach in purple and Mathis et al. in blue. The green circles demonstrate that the network trained using the Mathis et al. approach yields activations that are farther for the correct user, and the Siamese network yields activations that are farther for the incorrect user and closer for the correct user.

used, indicating that the Mathis et al. FCN is likely unsuccessful at recognizing that non-directed hand motions during the return phase may differ for hand controllers of different systems. The Siamese networks used in our work place less emphasis on the latter portions of the trajectories, where the physical characteristics of the device may induce irregularities in non-directed return motions.

Figure 2 provides a visualization of the results obtained using the FCN from Mathis et al. against our approach for 3 users, with one user shown per block. For all figures, we show visual results for the position of the right controller due to the ease of viewing the large-scale motion of the right controller, however, the analysis applies to position and orientation features from both controllers and the headset. In each user's block, the top two activation plots represent average activations generated using the first layer of the FCN limb of our Siamese network, while the bottom two plots represent average activations generated from the first layer of the Mathis et al. FCN. We generate the average activation by averaging the outputs of the ReLU function over all filters given the trajectory features as input. The first layer activations, though not representative of the comprehensive influence of the network, provide a reasonable visualization of the capability of the earlier part of the network in generating system-independent representations. The approach of Mathis et al. returns the user ID rather than a trajectory match. To facilitate visualization, given a use-time trajectory's activation in red, we show the closest matching activation for an enrollment trajectory from the user identified by Mathis et al. in blue in the first column of each block. The closest match is generated by taking the sum-squared distance between the use-time activation and all $n = 10$ enrollment activations for the detected user, and returning the enrollment activation with the lowest distance. In the second column of each block, we show the activation for the trajectory of the best matching user as returned by our approach in purple compared to the use-time activation in red. The third column of each block shows 3D plots for the use-time trajectory in red, and the enrollment trajectories for Mathis et al. for the closest activation in blue and the closest user using our Siamese network in purple. The figure demonstrates that Siamese activations for the correct user are closer than the incorrect user, while Mathis et al. activations for the correct user are farther. The green circles represent regions where the activation maps deviate from each other, resulting in an incorrect match using the approach of Mathis et al., due to its being agnostic to systematic deviations between two VR systems. We leverage the Siamese architecture to learn systematic deviations, thereby ensuring that activations from the same user demonstrate similarity.

Comparison to Miller et al. [47]. Our identification results demonstrate an overall improvement of 29.79%±8.58% averaged over system pairings when accuracies using our best performing features are compared against accuracies from best performing features from Miller et al. shown in the last row of Tables 3. The

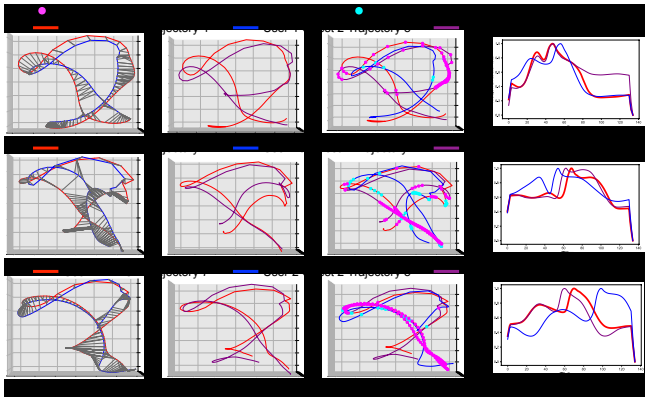

Figure 3: Comparison of results from our work against Miller et al. [47] shown using right hand trajectories. (a) Miller et al. yield an incorrect user match in blue to the use-time trajectory in red. (b) Our Siamese network returns the correct match in purple (c) by weighting portions of the trajectory with similar activations in pink. Similar activations for the incorrect match in cyan are fewer. (d) Average activations plotted against time, demonstrating similarity of shape between use-time and enrollment trajectories for the same user.

approach of Miller et al. demonstrates higher performance when the nearest neighbor distance calculation used in their work uses 80% of the trajectory points and disregards the latter 20% of the trajectory involving non-directed motions. Their approach provides best average accuracies of between 51.28% and 72.74% using 80% of the points, and between 44.82% and 70.43% using 100% of the points. Figure 3 provides a visual demonstration of the trajectory matches returned by the approach of Miller et al. [47] compared against our matches for three randomly selected users. For a use-time trajectory shown in red, Figure 3(a) demonstrates that the closest matching enrollment trajectory by the approach of Miller et al. shown in blue yields an incorrect user match. Figure 3(b) shows that the Siamese network in our approach returns the closest trajectory belonging to the correct user, shown in purple. As shown in Figure 3(c), the Siamese network weights portions of the trajectory of higher relevance, as shown by the points in pink that represent similar average activation outputs generated from the first neural network layer. We detect similar activations as those that are greater than a threshold of 0.5, and that have an absolute difference within a tolerance of 0.03. The plots demonstrate that similar high activations occur earlier in the trajectory, during the directed motions performed in the lift, poise, and thrust phase of the ball-throwing action, whereas the hand-return portions of the trajectory are weighted lower by the Siamese network. Points in cyan on the incorrectly matched trajectory from

E/U	Q1/V1	Q1/V2	Q2/V1	Q2/V2	Q1/C1	Q1/C2	Q2/C1	Q2/C2	V1/C1	V1/C2	V2/C1	V2/C2
R,L / P,O	92.68	89.15	86.78	90.68	***88.10***	***89.76***	82.20	78.27	80.12	81.05	81.20	80.37
R,H / P,O	89.93	89.66	88.44	88.68	79.10	73.78	84.93	87.93	83.22	84.15	***83.76***	79.93
R,L,H / P,O	***92.80***	***93.34***	***94.71***	***94.54***	84.46	82.73	***89.22***	***88.68***	***84.49***	***85.63***	83.63	***80.73***
R,L / P	76.61	77.66	79.41	69.49	69.71	67.56	61.02	57.90	65.98	62.85	62.59	63.88
R,H / P	84.76	75.83	85.20	75.29	68.56	72.22	77.15	73.68	77.20	74.51	73.90	76.37
R,L,H / P	86.61	77.73	84.56	81.07	75.61	71.20	66.59	72.05	74.66	69.63	69.56	72.63

Table 4: Rate of occurrence of correct user in top 10 enrollment trajectories as a percentage for 100% the trajectory points. The same conventions used in Table 3 apply here, except that highest occurrence rates are bolded/italicized.

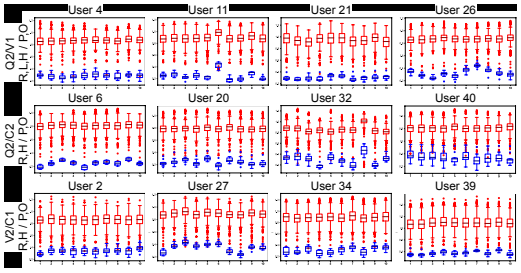

Figure 4: Box plots of outputs from the Siamese network. The vertical axis represents the output value. The horizontal axis represents the use-time trajectory for each user. Blue plots correspond to enrollment trajectories for the same user, while red plots correspond to those of different users. The plots show randomly chosen users under the highest-performing system pairings and feature sets.

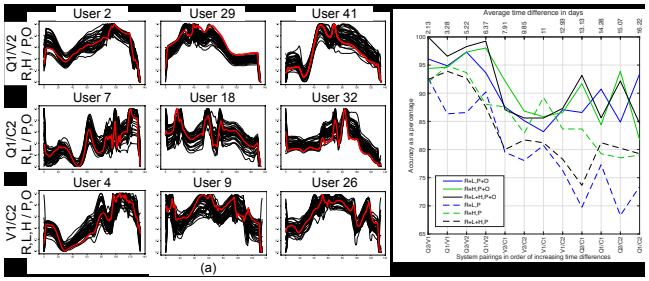

Figure 5: (a) Plots of activation outputs from first layer of Siamese network for use-time trajectories along the vertical axis against time along the horizontal axis. Red and black plots represent maps when the trajectory is used in the test and training set respectively. Plots are generated for randomly chosen users for various system pairings and feature sets. The plots demonstrate that individual networks learn and generate similar information for the same user. (b) Plot of accuracies for various device/feature combinations against increasing time differences within system pairings.

Miller et al. show fewer similar activations to the use-time trajectory under the Siamese network. Figure 3(d) shows average activations plotted against time. The use-time trajectory activations match those from their own enrollment trajectory more closely than those from the incorrect user returned by Miller et al.

Evaluation of robustness. In Table 4, we evaluate the robustness of our approach by obtaining how often the correct user was identified within the top 10 enrollment trajectories when 100% of the trajectory points are used; results for using 80% trajectory points are shown in the supplementary. As shown by Table 4, on average more than 80% (i.e., 8 or higher) of the top 10 trajectories belong to the correct user using our approach. Quest/Vive occurrence rates exceed 90%. Figure 4 shows box-plots of distances for each use-time trajectory to enrollment trajectories belonging to the same user and different users in blue and red respectively. The plots are shown for randomly chosen test users for the best performing system pairing and feature set. The plots demonstrate a clear separation of distances for each test user. Separation is observed even when distances are large, e.g., in the case of Users 11, 32, and 27. The separations support the high rate of occurrence of the correct user in the Top 10 enrollment trajectories in Table 4. The results demonstrate that the Siamese network is trained to learn a repeatable pattern of similarity.

Evaluation of generalizability using leave-one-out. One concern that may arise is whether the leave-one-out approach leveraged in this work to overcome data deficiency can be used to evaluate generalizability of the method in generating high accuracy when more users may be added to the test set. To assess this concern, we visualize the average activation at the first layer against time in Figure 5(a) for use-time trajectories of several users for various system pairings and feature sets. The plot in red represents the average activation when the use-time trajectory is part of the test set for the network trained when the corresponding user is the left-out user. The plots in black represent average activations when the use-time trajectory is used to train the remaining 40 out of 41 networks when the user forms a part of the training set for those networks. The black plots demonstrate that different networks learn similar information using input from the same user. The red plots demonstrate that a network trained by leaving out the user generates information similar to the output if the user were used during training, indicating that with 40 training users, the network is generalizable to novel users.

5 Discussion

Our approach shows lower average EERs and higher average accuracies for comparing Vive trajectories against the Quest, while EERs are higher and correspondingly accuracies are lower when Cosmos trajectories are compared against the Vive and the Quest. A number of factors play a role in the variability in accuracy across system pairings, such as the time duration between data collection which can influence evolution of long-term user behavior, clothing and accessories of the person, the type of tracking method used—lighthouse versus camera— and the influence of mass distribution of the system components on the motion of the user. The time span in days between successive captures is 1.15 ± 0.76 between the two Quest captures, 2.13 ± 2.26 between the second Quest and first Vive captures, 3.09 ± 2.03 between the two Vive captures, 7.91 ± 8.80 between the second Vive and first Cosmos captures, and 1.93 ± 1.54 between the two Cosmos captures. The time spans induce Quest and Vive trajectories be closest in time, followed by Vive and Cosmos trajectories. Quest and Cosmos trajectories are the furthest apart. To evaluate the influence of time, in Figure 5(b) we show plots of the

accuracies for the Siamese network from for 100% trajectory points against the average day difference over users for that system pairing. When position only is used as a feature, the accuracy drops as the time between the system pairings grows, indicating that users may change their positional placement over time. When orientation is included as a feature, the performance drops when the Cosmos is used at use-time, however, the enrollment system plays a lesser role. Variability in tracking appears to have a limited influence on accuracy, since using the lighthouse-based Vive at use-time and the camera-based Quest at enrollment provides high accuracy, whereas the accuracy of the camera-based Cosmos does not consistently drop over time whether the Vive or the Quest is used during enrollment. The accuracy drop may be explained by differences in influence of the physical characteristics of the VR system devices on the user's motion. During user grasp of the smaller 3-inch handle of the Cosmos controller, the shift in center of mass toward the heavier infrared (IR) emitter ring may influence limb motion. For both the Quest and the Vive controllers, the center of mass is more likely to be closer to the user's hand. While the Vive's handle at 5.75 inches is longer than the Quest's handle at 3.5 inches, the Quest's IR ring is thinner than the Cosmos and the Vive. For users with use-time Cosmos trajectories showing low accuracy, the Cosmos's emitter ring may weight down the user's hand further back during the poise, and may restrict thrust during forward motion. The drop for Cosmos, shown in the lower same-system performance, may also be explained by inaccuracy of inside-out tracking in localizing rapid motions or tracking extremities that leave the cameras' fields of view.

The use of behavior to distinguish genuine users from impostors comes with ethical concerns as the underlying behavior of users can be maliciously used for marketing or user monitoring in decentralized environments when user behavior data is processed remotely. Similar ethical concerns have been raised in smartphone-based authentication and work by Murmuria et al. [48] localizes the processing on the device to ensure no personally identifiable information is available to external applications. In our work, once the Siamese networks have been trained they can be deployed on a user's device to ensure security is maintained in a local environment. The use of deep learning as performed in our work enables rapid deployment on user devices for local processing, given that despite resource-intensive training phases, trained networks can be deployed on resource-constrained devices to run in real-time, as has been demonstrated for small form-factor computing such as the Raspberry Pi [34, 72]. For instance, work by Miller et al. [46] demonstrates a real-time VR system for authenticating users that can operate fully offline by passing user trajectories to a pre-built classifier. During testing, our approach provides interactive performance at 80 milliseconds to authenticate an input use-time trajectory by running the FCN on trajectory using a GTX 1080Ti GPU, and computing distances to the enrollment trajectory FCN outputs using a single core of an AMD Ryzen 2700X 3.7GHz CPU. While the distance computations scale with the number of enrolled users, computation of the F_{out}-dimensional enrollment limb (i.e., Limb 1) outputs of the Siamese network can be done offline, necessitating a single run of the use-time limb (i.e., Limb 2) and computation of independent Euclidean distances that can be run in parallel on modern multi-core hardware. For severely resource-constrained devices, future work can investigate the possibility of matching an input user to enrolled users within a group, where group-level similarities are identified by clustering trajectories and/or intermediate limb outputs. Once the Siamese networks have been trained using enrollment/use-time pairs from a sufficient set of training users to achieve high authentication performance, for new test users the Siamese networks need only be used to perform computation of distance value against their enrollment data, with minimal need for re-training. In future work, we will evaluate performance trends with varying user numbers, to identify asymptotic behavior of authentication and influence of intermittent re-training on resource use during training and deployment.

While Mathis et al. [45] argue that the knowledge-based approach remains the minimally invasive method for authentication, knowledge-based methods fail to address the challenge of a user deliberately handing over their credentials to an ally, e.g., in the case of a student attempting to circumvent a VR examination. In such a scenario, a single-point-of-entry check of the user's identity using the de-facto password/PIN is insufficient. Thus, as we consider future applications of VR, the user session must be continuously monitored to ensure the legitimate user remains in control. Miller et al. [46] use VR behavioral biometrics to continuously maintain the security of a user's session by verifying whether a user's behavior matches past behavior. In their demonstration, a confederate takes over a user session, but is immediately identified as an impostor through their behavior. Once trained, our Siamese network can be deployed for continuous session monitoring whereby each user trajectory is passed to the network for testing, and the user session is locked if the claimed user identity does not match the active user.

There is a substantial scope for future research emanating from our findings. Work is needed on disambiguating factors influencing the trajectories of the various devices comprising the VR system by, for instance, performing studies on the same day, or keeping the time difference between system pairs constant, i.e., no more than a day apart, with randomization so that systems are not presented to users in the same order. Detailed investigation of the influence of tracking mechanisms is essential, by comparing camera-based systems such as the Oculus Quest, the Oculus Quest 2, and the HTC Vive Cosmos against each other, comparing lighthouse-based systems such as the HTC Vive and the Valve Index against each other, and evaluating cross-tracking performance by using camera-based systems as enrollment and lighthouse-based systems as test, and vice versa. This work should investigate range of motion covered by the fields of view of the tracking devices, and augmenting built-in tracking with 360° head-mounted cameras and multi-view external sensors. Work is needed in fine-grained analysis on variation in weighting of the motion trajectories due to differences in the physical characteristics of the hand controllers and headset, as well as variation in clothing and accessories donned by the person. Studies are also essential in evaluating tradeoffs between usability and security under the influence of system-induced and external alterations of user behavior. In this work, we investigate the task of ball-throwing due to its vulnerability to ready mimicry given its simplicity and real-world familiarity. Complex multi-stage actions related to higher-level tasks such as for instance filling out a form are likely to be non-repeatable by impostors, but also simultaneously show high within-user variability. Our work demonstrates that features such as wrist orientation and head motion are essential aspects of the behavioral biometric signature. Even within tasks with limited translational motion, orientation and subtle motions are likely to continue contributing to a person's signature, and form important components of future studies.

6 Conclusion

In this work, we present an approach that learns systematic differences between multiple VR systems in order to enable cross-system behavior-based VR biometrics. We demonstrate average authentication EERs ranging from 1.38% to 3.86%, and identification accuracies ranging from 87.82% to 98.53% for a dataset of 41 right-handed users interacting with a ball-throwing application in VR. By using a metric learning approach, our approach addresses fundamental deficiencies of existing work which use generic matching or learning-based techniques that are agnostic to inter-system differences. By outputting a match score, our approach enables generalization of authentication and identification to novel users.

Acknowledgments

This work was partially supported by NSF grant CNS-1730183.

REFERENCES

[1] K. Ahrabian and B. BabaAli. Usage of autoencoders and siamese networks for online handwritten signature verification. *Neural Computing and Applications*, 31(12), Nov 2019.

[2] A. Ajit, N. K. Banerjee, and S. Banerjee. Combining pairwise feature matches from device trajectories for biometric authentication in virtual reality environments. In *Proc. AIVR*. IEEE, New York, USA, 2019.

[3] F. Alonso-Fernandez, K. B. Raja, R. Raghavendra, C. Busch, J. Bigün, R. Vera-Rodrıguez, and J. Fiérrez. Cross-sensor periocular biometrics: A comparative benchmark including smartphone authentication. *CoRR*, 2019.

[4] F. Alonso-Fernandez, R. N. Veldhuis, A. M. Bazen, J. Fiérrez-Aguilar, and J. Ortega-Garcia. Sensor interoperability and fusion in fingerprint verification: A case study using minutiae-and ridge-based matchers. In *Proc. ICCARV*. IEEE, New York, USA, 2006.

[5] F. A. Alsulaiman and A. El Saddik. A novel 3d graphical password schema. In *Proc. VECIMS*. IEEE, New York, USA, 2006.

[6] F. A. Alsulaiman and A. El Saddik. Three-dimensional password for more secure authentication. *IEEE Transactions on Instrumentation and measurement*, 57(9), Sep 2008.

[7] M.-S. Bracq, E. Michinov, and P. Jannin. Virtual reality simulation in nontechnical skills training for healthcare professionals: A systematic review. *Simulation in Healthcare*, 14(3), Jun 2019.

[8] A. G. Campbell, T. Holz, J. Cosgrove, M. Harlick, and T. O'Sullivan. Uses of virtual reality for communication in financial services: A case study on comparing different telepresence interfaces: Virtual reality compared to video conferencing. In *Proc. FICC*. Springer, Berlin, Germany, 2019.

[9] M. P. Centeno, Y. Guan, and A. van Moorsel. Mobile based continuous authentication using deep features. In *Proc. EMDL*. ACM, New York, USA, 2018.

[10] Y. Chen, G. Parziale, E. Diaz-Santana, and A. K. Jain. 3d touchless fingerprints: Compatibility with legacy rolled images. In *Proc. BCC*. IEEE, New York, USA, 2006.

[11] B. J. Concannon, S. Esmail, and M. Roduta Roberts. Head-mounted display virtual reality in post-secondary education and skill training: A systematic review. In *Proc. FIE*. Frontiers, Switzerland, 2019.

[12] P. Connor and A. Ross. Biometric recognition by gait: A survey of modalities and features. *Computer Vision and Image Understanding*, 167, Feb 2018.

[13] E. Czerniak, A. Caspi, M. Litvin, R. Amiaz, Y. Bahat, H. Baransi, H. Sharon, S. Noy, and M. Plotnik. A novel treatment of fear of flying using a large virtual reality system. *Aerospace medicine and human performance*, 87(4), Apr 2016.

[14] D. Deb, A. Ross, A. K. Jain, K. Prakah-Asante, and K. V. Prasad. Actions speak louder than (pass) words: Passive authentication of smartphone* users via deep temporal features. In *Proc. ICB*. IEEE, New York, USA, 2019.

[15] M. O. Derawi. Accelerometer-based gait analysis, a survey. In *Proc. NISK*. NISK, Norway, 2010.

[16] B. Fan, X. Liu, X. Su, P. Hui, and J. Niu. Emgauth: An emg-based smartphone unlocking system using siamese network. In *Proc. PerCom*. IEEE, New York, USA, 2020.

[17] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4), Mar 2019.

[18] H. Feng, C. Li, J. Liu, L. Wang, J. Ma, G. Li, L. Gan, X. Shang, and Z. Wu. Virtual reality rehabilitation versus conventional physical therapy for improving balance and gait in parkinson's disease patients: A randomized controlled trial. *Medical science monitor: international medical journal of experimental and clinical research*, 25, Jun 2019.

[19] M. Funk, K. Marky, I. Mizutani, M. Kritzler, S. Mayer, and F. Michahelles. Lookunlock: Using spatial-targets for user-authentication on hmds. In *Proc. CHI*. ACM, New York, USA, 2019.

[20] D. Gafurov. A survey of biometric gait recognition: Approaches, security and challenges. In *Proc. NIK*. NIKT, Norway, 2007.

[21] C. George, D. Buschek, A. Ngao, and M. Khamis. Gazeroomlock: Using gaze and head-pose to improve the usability and observation resistance of 3d passwords in virtual reality. In *Proc. AVR*. Springer, Berlin, Germany, 2020.

[22] C. George, M. Khamis, D. Buschek, and H. Hussmann. Investigating the third dimension for authentication in immersive virtual reality and in the real world. In *Proc. VR*. IEEE, New York, USA, 2019.

[23] C. George, M. Khamis, E. von Zezschwitz, M. Burger, H. Schmidt, F. Alt, and H. Hussmann. Seamless and secure VR: Adapting and evaluating established authentication systems for virtual reality. In *Proc. NDSS*. The Internet Society, Reston, USA, 2017.

[24] J. Gurary, Y. Zhu, and H. Fu. Leveraging 3d benefits for authentication. *International Journal of Communications, Network and System Sciences*, 10(8B), Aug 2017.

[25] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *Proc. CVPR*. IEEE, New York, USA, 2006.

[26] T. Hoang, T. D. Nguyen, C. Luong, S. Do, and D. Choi. Adaptive cross-device gait recognition using a mobile accelerometer. *JIPS*, 9(2), Apr 2013.

[27] L. Jensen and F. Konradsen. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, 23(4), Nov 2018.

[28] R. Jillela and A. Ross. Matching face against iris images using periocular information. In *Proc. ICIP*. IEEE, New York, USA, 2014.

[29] C. Kandaswamy, J. C. Monteiro, L. M. Silva, and J. S. Cardoso. Multi-source deep transfer learning for cross-sensor biometrics. *Neural Computing and Applications*, 28(9), Apr 2017.

[30] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proc. ICLR*. ICLR, California, USA, 2014.

[31] A. Kupin, B. Moeller, Y. Jiang, N. K. Banerjee, and S. Banerjee. Task-driven biometric authentication of users in virtual reality (VR) environments. In *Proc. MMM*. Springer, Berlin, Germany, 2019.

[32] B. M. Kyaw, N. Saxena, P. Posadzki, J. Vseteckova, C. K. Nikolaou, P. P. George, U. Divakar, I. Masiello, A. A. Kononowicz, N. Zary, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. *Journal of medical Internet research*, 21(1), Jan 2019.

[33] M. Lager and E. A. Topp. Remote supervision of an autonomous surface vehicle using virtual reality. *IFAC-PapersOnLine*, 52(8), Jul 2019.

[34] N. Lamb and M. C. Chuah. A strawberry detection system using convolutional neural networks. In *Proc. Big Data*. IEEE, New York, USA, 2018.

[35] S. Li, A. Ashok, Y. Zhang, C. Xu, J. Lindqvist, and M. Gruteser. Whose move is it anyway? authenticating smart wearable devices using unique head movement patterns. In *Proc. PerCom*. IEEE, New York, USA, 2016.

[36] C. Lin and A. Kumar. Improving cross sensor interoperability for fingerprint identification. In *Proc. ICPR*. IEEE, New York, USA, 2016.

[37] C. Lin and A. Kumar. A cnn-based framework for comparison of contactless to contact-based fingerprints. *IEEE Transactions on Information Forensics and Security*, 14(3), Mar 2018.

[38] F. Liu, D. Zhang, C. Song, and G. Lu. Touchless multiview fingerprint acquisition and mosaicking. *IEEE Transactions on Instrumentation and Measurement*, 62(9), Sep 2013.

[39] K. R. Lohse, C. G. Hilderman, K. L. Cheung, S. Tatla, and H. M. Van der Loos. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. *PloS One*, 9(3), Mar 2014.

[40] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), Nov 2004.

[41] M. Maciaś, A. Dąbrowski, J. Fraś, M. Karczewski, S. Puchalski, S. Tabaka, and P. Jaroszek. Measuring performance in robotic teleoperation tasks with virtual reality headgear. In *Proc. AUTOMATION*. Springer, Berlin, Germany, 2019.

[42] M. G. Maggio, G. Maresca, R. De Luca, M. C. Stagnitti, B. Porcari, M. C. Ferrera, F. Galletti, C. Casella, A. Manuli, and R. S. Calabrò. The growing use of virtual reality in cognitive rehabilitation: fact, fake or vision? a scoping review. *Journal of the National Medical Association*, 111(4), Aug 2019.

[43] F. Mathis, H. I. Fawaz, and M. Khamis. Knowledge-driven biometric authentication in virtual reality. In *Proc. CHI*. ACM, New York, USA,

2020.

[44] F. Mathis, J. Williamson, K. Vaniea, and M. Khamis. Rubikauth: Fast and secure authentication in virtual reality. In *Proc. CHI*. ACM, New York, USA, 2020.

[45] F. Mathis, J. H. Williamson, K. Vaniea, and M. Khamis. Fast and secure authentication in virtual reality using coordinated 3d manipulation and pointing. *ACM Transactions on Computer-Human Interaction*, 6(1), Jan 2021.

[46] R. Miller, A. Ajit, N. K. Banerjee, and S. Banerjee. Realtime behavior-based continual authentication of users in virtual reality environments. In *Proc. AIVR*. IEEE, New York, USA, 2019.

[47] R. Miller, N. K. Banerjee, and S. Banerjee. Within-system and cross-system behavior-based biometric authentication in virtual reality. In *Proc. VR Workshops*. IEEE, New York, USA, 2020.

[48] R. Murmuria, A. Stavrou, D. Barbara, and V. Sritapan. Your data in your hands: Privacy-preserving user behavior models for context computation. In *Proc. PerCom Workshop*. IEEE, New York, USA, 2017.

[49] T. Mustafa, R. Matovu, A. Serwadda, and N. Muirhead. Unsure how to authenticate on your VR headset?: Come on, use your head! In *Proc. IWSPA*. ACM, New York, USA, 2018.

[50] S. Neumeier, N. Gay, C. Dannheim, and C. Facchi. On the way to autonomous vehicles teleoperated driving. In *Proc. AmE*. VDE, Berlin, Germany, 2018.

[51] S. Neumeier, P. Wintersberger, A.-K. Frison, A. Becher, C. Facchi, and A. Riener. Teleoperation: The holy grail to solve problems of automated driving? Sure, but latency matters. In *Proc. AutomotiveUI*. ACM, New York, USA, 2019.

[52] M. M. North, S. M. North, and J. R. Coble. Virtual reality therapy: an effective treatment for the fear of public speaking. *International Journal of Virtual Reality*, 3(3), Jan 2015.

[53] I. Olade, C. Fleming, and H.-N. Liang. Biomove: Biometric user identification from human kinesiological movements for virtual reality systems. *Sensors*, 20(10), May 2020.

[54] I. Olade, H.-N. Liang, C. Fleming, and C. Champion. Exploring the vulnerabilities and advantages of swipe or pattern authentication in virtual reality (vr). In *Proc. ICVARS*. ACM, New York, USA, 2020.

[55] K. Pfeuffer, M. J. Geiger, S. Prange, L. Mecke, D. Buschek, and F. Alt. Behavioural biometrics in VR: Identifying people from body motion and relations in virtual reality. In *Proc. CHI*. ACM, New York, USA, 2019.

[56] G. Pizzi, D. Scarpi, M. Pichierri, and V. Vannucci. Virtual reality, real reactions?: Comparing consumers' perceptions and shopping orientation across physical and virtual-reality retail stores. *Computers in Human Behavior*, 96(0), Jul 2019.

[57] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147(0), Apr 2020.

[58] K. B. Raja, R. Raghavendra, and C. Busch. Dynamic scale selected laplacian decomposed frequency response for cross-smartphone periocular verification in visible spectrum. In *Proc. FUSION*. IEEE, New York, USA, 2016.

[59] N. P. Ramaiah and A. Kumar. On matching cross-spectral periocular images for accurate biometrics identification. In *Proc. BTAS*. IEEE, New York, USA, 2016.

[60] I. Rida, N. Almaadeed, and S. Almaadeed. Robust gait recognition: a comprehensive survey. *IET Biometrics*, 8(1), Aug 2018.

[61] C. E. Rogers, A. W. Witt, A. D. Solomon, and K. K. Venkatasubramanian. An approach for user identification for head-mounted displays. In *Proc. ISWC*. ACM, New York, USA, 2015.

[62] A. Ross and A. Jain. Biometric sensor interoperability: A case study in fingerprints. In *Proc. IWBA*. Springer, Berlin, Germany, 2004.

[63] A. Ross and R. Nadgir. A thin-plate spline calibration model for fingerprint sensor interoperability. *IEEE Transactions on Knowledge and Data Engineering*, 20(8), Aug 2008.

[64] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro. Fusing iris and periocular information for cross-sensor recognition. *Pattern Recognition Letters*, 57:52–59, May 2015.

[65] S. Schneegass, Y. Oualil, and A. Bulling. Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull. In *Proc. CHI*. ACM, New York, USA, 2016.

[66] A. Sharma, S. Verma, M. Vatsa, and R. Singh. On cross spectral periocular recognition. In *Proc. ICIP*. IEEE, New York, USA, 2014.

[67] X. Shen, Z. J. Chong, S. Pendleton, G. M. J. Fu, B. Qin, E. Frazzoli, and M. H. Ang. Teleoperation of on-road vehicles via immersive telepresence using off-the-shelf components. In *Proc. IAS*. Springer, Berlin, Germany, 2016.

[68] Y. Shen, H. Wen, C. Luo, W. Xu, T. Zhang, W. Hu, and D. Rus. Gaitlock: Protect virtual and augmented reality headsets using gait. *IEEE Transactions on Dependable and Secure Computing*, 16(3), May-Jun 2019.

[69] A. J. Snoswell and C. L. Snoswell. Immersive virtual reality in health care: Systematic review of technology and disease states. *JMIR Biomedical Engineering*, 4(1), Jan-Dec 2019.

[70] R. Tilhou, V. Taylor, and H. Crompton. 3d virtual reality in k-12 education: A thematic systematic review. In *Emerging Technologies and Pedagogies in the Curriculum*, pp. 169–184. Springer, 2020.

[71] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia. Exploring recurrent neural networks for on-line handwritten signature biometrics. *IEEE Access*, 6, Jan 2018.

[72] A. Ullah, K. Muhammad, K. Haydarov, I. U. Haq, M. Lee, and S. W. Baik. One-shot learning for surveillance anomaly recognition using siamese 3d cnn. In *Proc. IJCNN*. IEEE, New York, USA, 2020.

[73] C. S. Vorugunti, P. Mukherjee, V. Pulabaigari, et al. Osvnet: Convolutional siamese network for writer independent online signature verification. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1470–1475. IEEE, New York, USA, 2019.

[74] C. Wan, L. Wang, and V. V. Phoha. A survey on gait recognition. *ACM Computing Surveys (CSUR)*, 51(5), Aug 2018.

[75] P. Wang, P. Wu, J. Wang, H.-L. Chi, and X. Wang. A critical review of the use of virtual reality in construction engineering education and training. *International journal of environmental research and public health*, 15(6), Jun 2018.

[76] Z. Wang, W. Yan, and T. Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *Proc. IJCNN*. IEEE, New York, USA, 2017.

[77] S. Weise and A. Mshar. Virtual reality and the banking experience. *Journal of Digital Banking*, 1(2), 0 2016.

[78] X. Wu, A. Kimura, S. Uchida, and K. Kashino. Prewarping siamese network: Learning local representations for online signature verification. In *Proc. ICASSP*. IEEE, New York, USA, 2019.

[79] L. Xiao, Z. Sun, and T. Tan. Fusion of iris and periocular biometrics for cross-sensor identification. In *Proc. CCBR*. Springer, Berlin, Germany, 2012.

[80] X. Xu, J. Yu, Y. Chen, Q. Hua, Y. Zhu, Y.-C. Chen, and M. Li. Touchpass: towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations. In *Proc. MobiCOM*. ACM, New York, USA, 2020.

[81] L. Xue, C. J. Parker, and H. McCormick. A virtual reality and retailing literature review: Current focus, underlying themes and future directions. In *Augmented Reality and Virtual Reality*, pp. 27–41. Springer, 2019.

[82] S. Yi, Z. Qin, E. Novak, Y. Yin, and Q. Li. Glassgesture: Exploring head gesture interface of smart glasses. In *Proc. INFOCOM*. IEEE, New York, USA, 2016.

[83] Z. Yu, H.-N. Liang, C. Fleming, and K. L. Man. An exploration of usable authentication mechanisms for virtual reality systems. In *Proc. APCCAS*. IEEE, New York, USA, 2016.