RESEARCH-ARTICLE

Enhanced In-air Signature Verification via Hand Skeleton Tracking to Defeat Robot-level Replays

ZEYU DENG, Louisiana State University, Baton Rouge, LA, United States

LONG HUANG, Louisiana State University, Baton Rouge, LA, United States

CHEN WANG, Louisiana State University, Baton Rouge, LA, United States

Published: 04 December 2023

ACSAC '23: Annual Computer Security Applications Conference
December 4 - 8, 2023
TX, Austin, USA

ACSAC '23: Proceedings of the 39th Annual Computer Security Applications Conference (December 2023)
https://doi.org/10.1145/3627106.3627195
ISBN: 9798400708862

Enhanced In-air Signature Verification via Hand Skeleton Tracking to Defeat Robot-level Replays

Zeyu Deng
zdeng6@lsu.edu
Louisiana State University
Baton Rouge, Louisiana, USA

Long Huang
lhuan45@lsu.edu
Louisiana State University
Baton Rouge, Louisiana, USA

Chen Wang
chenwang1@lsu.edu
Louisiana State University
Baton Rouge, Louisiana, USA

ABSTRACT

Behavioral biometrics has emerged as an important security factor for user authentication. Compared to static biometrics (e.g., faces, irises, and fingerprints), using human motion behaviors for authentication causes lower concern about privacy abuse, and behavior biometrics are shown hard to be replicated by humans. In-air 3D signature is one representative of behavioral biometrics. Specifically, a user's hand movements can be tracked by visual or wireless sensors for contact-free signature authentication, where both the fingertip trajectory and the dynamic motion features are verified to provide enhanced security. However, with the advancement of 3D printing and robot technology, we find that 1) existing hand-tracking interfaces (e.g., Leap Motion and Google MediaPipe) are easily tricked by a fake hand, and 2) a robotic arm can reproduce a user's in-air 3D signature with high similarity regarding both trajectory and motion behaviors. Thus, this work investigates the security of in-air signatures under robot-level replays and proposes to extend the signature verification from a single-point fingertip to multiple hand joints for enhanced security. We develop the hand skeleton-based 3D signature verification system, which can be deployed on any single camera devices (2D or 3D). The key insight is that current robots could hardly replicate the minute and unique inter-joint motions of a user. In particular, we track the hand skeleton using a single camera and reconstruct/draw the trajectories of its joints in a virtual 3D space, using the color gradients to represent time-lapse and using varying line widths to describe joint significance. Based on that, we extract the three-view skeleton signatures and inter-joint motion features and develop a convolutional neural network for verification. Extensive experiments show that our system not only achieves high authentication performance but also effectively mitigates robot-level replay attacks.

ACM Reference Format:
Zeyu Deng, Long Huang, and Chen Wang. 2023. Enhanced In-air Signature Verification via Hand Skeleton Tracking to Defeat Robot-level Replays. In *Annual Computer Security Applications Conference (ACSAC '23), December 04–08, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3627106.3627195

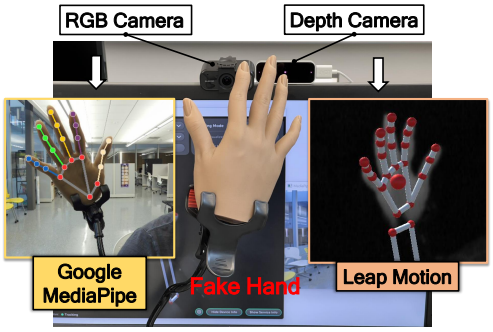


Figure 1: Vulnerability of current commodity hand-tracking interfaces (easily fooled with a silicone fake hand).

1 INTRODUCTION

Behavioral biometrics is an emerging field in user authentication approaches, which focuses on analyzing a user's dynamic motion characteristics of moving an arm/hand [18, 46], walking [12, 34], head-shaking, typing [5, 17], and signing [30]. Compared to the widely used PIN/password and static biometrics, behavioral biometrics are more secure and more convenient to use for authentication [2]. In particular, human motion data can be readily captured through the pervasively deployed visual, inertial, and wireless sensors. Furthermore, behavioral motion features are inherently dynamic, presenting considerable obstacles for attackers to mimic or reproduce [39]. Compared to facial and fingerprint biometrics, behavioral biometrics also minimize privacy erosion concerns.

In-air signature is a prominent example of behavioral biometric authentication, which presents high degrees of motion freedom. It not only retains the individual's obligation inherent in a traditional signature but also offers enhanced security by verifying both the 3D handwriting curves and the user's signing behaviors. These behaviors are captured as a time series of motion features, including velocity and acceleration [8, 24, 39]. In contrast to 2D signature biometrics, in-air signature extends the user's signing behavior into the 3D space by adding a depth dimension. It also eliminates the need for a writing surface, featuring non-contact and device-free authentication service, which has a notable advantage amid the COVID-19 Pandemic. Furthermore, there have been many commercialized hand-tracking interfaces available on either a single RGB or depth camera to support in-air signature verification [28, 40, 44]. We foresee the future expansion of its applications in the "metaverse", whose market size is expected to grow more than 13 times in 10 years [13]. A VR user can now use bare hands to interact with the device or sign in the virtual space to authorize a transaction rather than cumbersomely entering passwords with handheld controllers.

ACSAC '23, December 04–08, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0886-2/23/12.
https://doi.org/10.1145/3627106.3627195

This work demonstrates that even the 3D in-air signature with high motion freedom degrees could be forged by an adversary. We reveal two security threats to current visual in-air signature authentications: 1) Commercial hand-tracking interfaces such as Google MediaPipe [44] and Leap Motion [40] are easily fooled by a fake hand, as shown in Figure 1; 2) A low-cost commodity robotic arm ($\leq$ \$1000) can precisely replay a user's in-air signature with its end effector as shown in Figure 2. The main reason is that current hand-tracking solutions have never considered the threats from 3D printing and motion-copy robots. They rely on hand-like shapes to recognize/track a hand without liveness detection capabilities. Moreover, the current 3D in-air signature is just a single-point coordinate time series (e.g., fingertip), which can be easily reproduced by a robotic arm's end effector.

Rather than modifying the current hand-tracking interfaces, we propose to extend the dimension of in-air signatures from a single point to multiple hand joints and further leverage the hand's kinematic structure motions to defeat robot replays. Our key insight is that current robots are still not able to copy the minute inter-joint motions of a user's hand, which can be further leveraged to derive joint-level behavioral biometric features. In particular, we develop the hand skeleton-based 3D signature verification system. The system takes the recordings of a single camera (2D or 3D) as input and extracts the 3D coordinate time series of a hand's multi-joints for authentication. We design a novel representation of hand skeleton motions, which reconstructs/draws the 3D trajectories of hand joints in different colors, uses the color gradients to present the time-lapse, and uses the line widths to describe the joint significance. We next project the 3D trajectories on three virtual planes to examine three different aspects of the hand skeleton-level signature and derive the inter-joint motion features. We develop a Convolutional Neural Network (CNN) model to analyze these biometric features for authentication as well as mitigating the replay attacks enabled by motion-copy robots.

Our contributions are summarized as follows:

- We identify two security issues of current in-air signature verification systems, (1) the vulnerability of hand-tracking interfaces to fake hands and (2) the behavioral biometric replays achieved by low-cost commodity robots.
- After investigating the security of behavioral biometrics under the emerging robot-replay threats, we propose a security-enhanced in-air signature authentication system, which extends the traditional single-point signatures to the hand skeleton signatures. We find that the minute inter-joint motion characteristics are unique to each user and also hard to be reproduced by current robots.
- We design a novel approach to examine a single-camera obtained hand skeleton signature with its three-view presentations, where different colors are used to index each hand joint, and color gradients and line widths are used to describe the time information and the joint significance, respectively. We further derive the inter-joint biometric features and develop a convolutional neural network for authentication.
- We implement the system prototypes based on a regular RGB camera and a depth camera, respectively, and build a robot platform for launch replay attacks. Results show that our system achieves a high performance in verifying users and mitigating robot-level replays.

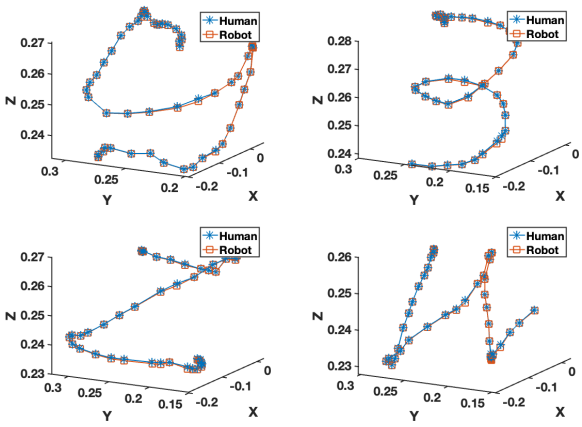


Figure 2: Replaying a user's in-air writing with a robotic arm.

2 RELATED WORK

Physiological biometrics focuses on the examination of identifiable human body parts, such as facial patterns [26, 31], fingerprints [4, 9], and irises [29]. However, the advancements in recording and replay technologies have made feasible the copy and reuse of such static biometrics, which presents severe security risks. For example, by leveraging 3D reconstruction and printing technologies, an adversary can create a 3D mask [7] of the user's face to fool face recognition systems. Similarly, a fake fingertip [3] can be created from the user's latent fingerprint to pass the fingerprint scanners. In contrast to static physiological biometrics, behavioral biometrics focus on capturing the behavior characteristics of a user for authentication, which are believed hard to be copied or reproduced by an adversary. Emerging behavioral biometrics include hand gestures [18, 46], gait patterns [12, 34], keystroke dynamics [5, 17], and eye movements [45]. In-air 3D signature is a representative example with high motion freedom degrees, which contains both signature trajectories and signing behaviors.

Traditional signature verification treats the user's signature as a static image and analyzes its geometric shape [15]. Digital writing pads and touchscreen devices further enable the recording of the entire process of a user's signing activity and obtain a 2D coordinate time series of the user's signature [14, 16, 34]. Based on that, both the signature curve and the user's signing behaviors, such as velocity, acceleration, and pressure, can be used for authentication, which improves signature security with multiple feature dimensions. In addition, the embedded motion sensors [20, 35] and photoplethysmography sensor [33] on wrist-worn devices (e.g., smartwatches) can be used to further capture the user's wrist/arm motion behaviors beyond the writing surface.

In-air 3D signature introduces an additional depth dimension. It eliminates the need for a writing surface to achieve contract-free and maximizes the freedom of hand motions to obtain more aspects of the user's behavioral biometrics. A number of sensors can be used for in-air signature authentication, including visual sensors [8, 19, 39], inertial sensors [21, 23], acoustic sensors [37],

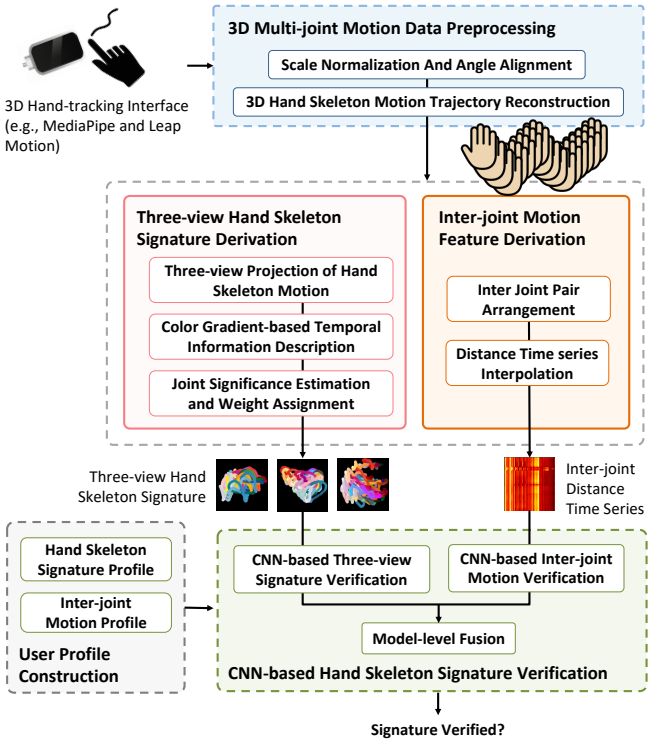


Figure 3: The overview of our 3D hand skeleton signature user authentication system.

and Radio Frequency (RF) signals [1, 25]. This work focuses on the vision-based approaches due to the broad support by many commercialized hand-tracking interfaces. The vision-based solutions can be divided into two categories based on the type of camera used for sensing. Regular RGB cameras have been utilized for in-air signature verification [8, 38], and Google MediaPipe is a commercialized hand-tracking interface developed for these 2D cameras to capture 3D trajectories of the hand [19]. There are also many off-the-shelf depth cameras, including the Kinect sensor [28] and Leap Motion, which have been employed to verify users' in-air signatures or hand gestures [10, 22, 24, 39, 42]. We also find that the majority of in-air signature works use Dynamic Time Warping (DTW) algorithms to learn the user's signature template for authentication [8, 19, 39], while others use CNN and support vector machines.

Previous studies on in-air signatures uniformly highlight the difficulty of an impersonation attacker mimicking a user's dynamic signing behaviors. However, only a few studies consider the potential of replay attacks executed by robots. The latest work uses a "Lego" robot to imitate the user's swiping behaviors on a smartphone screen [36], which was a decade ago and did not consider the more challenging in-air signature scenario, where the user has high degrees of motion freedom. We find that the advancements in robot techniques in the past decade have posed severe security threats to behavioral biometric security. For example, a robot agent can be taught to learn and reproduce human handwriting [43]. It is further demonstrated that the end-effector of a robotic arm can inherit the motion behaviors of the human controller's hand [11]. It is thus imperative to investigate the security of in-air signatures under the emerging robot threat and mitigate the security risk.

3 BACKGROUND AND SYSTEM MODELS

3.1 Vulnerability of Hand-tracking Interface

Commercially available visual hand-tracking interfaces are increasingly being integrated into electronic mobile devices, enabling non-contact hand motion tracking and in-air signatures. For instance, Google MediaPipe can detect and track the user's hand with 21 skeleton landmarks by using a single RGB camera (e.g., on laptops). Though using a 2D camera, MediaPipe establishes a 3D coordinate system with the palm center serving as the origin and estimates each joint's depth information using hand skeleton models. On the other hand, Leap Motion utilizes stereo vision created by two infrared cameras to capture direct 3D hand movements. The sensing data is further fed into a generic hand model to provide visual feedback. These technologies have been integrated into VR devices (e.g., Meta Quest [27]) for non-contact hand-tracking and cyber-physical interactions. Based on these hand-tracking interfaces, there have been many studies on extracting the user's in-air signature for authentication [8, 10, 19]. The in-air signatures are beginning to be practically implemented in both mobile and VR applications [6, 32].

Almost all visual hand-tracking interfaces rely on recognizing hand shapes in video frames. The in-air signature is typically obtained by locating the index fingertip (or the palm center) and logging its coordinate time series. It is more important to note that these commercial interfaces address the occlusion or self-occlusion of the user's partial hand by using historical finger coordinates and standard hand skeleton models [40, 44]. Thus, the 3D coordinates of the index finger or palm center can be consistently obtained to enable complete in-air signature capture. However, we find that these hand-tracking interfaces can be readily deceived by any hand-like objects. We conduct an experiment by displaying a silicon hand to Google MediaPipe and Leap Motion, respectively. Both interfaces detect and track the hand as shown in Figure 1. Thus, the authentication systems based on such hand-tracking interfaces could be easily tricked and are not secure.

3.2 Potential Robot-level Replay Attacks

Robotic arms have replaced many human laborers in manufacturing processes for decades. As they continue to decrease in cost, their applications are no longer limited to the industry. It has been shown that a robotic arm's end-effector can reach anywhere a human hand reach, thus having the ability to repeat a user's hand motion trajectory. For example, a robotic agent can be instructed to learn and reproduce a human's handwriting trajectory [43]. More than just reproducing a trajectory, robots also have the potential to imitate a user's motion behavior, which contains inherent individual motion dynamics, such as accelerations and velocities. It has been demonstrated that a toy robot can mimic a user's behaviors of simple swiping on a smartphone touchscreen [36]. However, it is still unclear whether a robotic arm could mimic the 3D in-air signatures, which have high degrees of freedom and are more complex to reproduce regarding both the trajectory and signing behavior.

To illustrate the feasibility of using robots to replay in-air signatures, we record a user's palm center movements of writing

four characters in the air using a Leap Motion controller. A 5-DoF PincherX 150 robotic arm is used to repeat the user's writing. We compare the location time series of the robot's end-effector (returned by the robot) with that of the original human writings (captured by Leap Motion). The comparisons across all four characters in Figure 2 demonstrate the high replay capability of the robotic arm. In particular, the end-effector presents the trajectories closely aligned with the user's handwriting curves. Moreover, the human hand motion is copied point-by-point, which may present high portions of the user's motion behaviors, such as similar accelerations and velocities. If attaching a fake hand to the robotic arm to attack existing hand-tracking interfaces, the current in-air signatures are at high risk to be broken.

3.3 System Overview

This study proposes a 3D hand skeleton signature verification algorithm to improve in-air signature security without changing the signing process. The algorithm achieves enhanced security by verifying three biometric factors closely integrated in the signing process: the hand skeleton motion behavior, the inter-joint motion behavior, and the individual hand geometry presented by the 3D hand skeleton model. The algorithm can be deployed on any existing in-air signature verification systems. The system flow is shown in Figure 3. The user initiates authentication by signing a 3D signature in the air, which is the same process as in prior in-air signature works [8, 39]. The hand is detected and tracked by a single camera (e.g., Leap Motion sensor or RGB camera). We then extract the 3D coordinate time series of each hand joint and obtain the hand skeleton motion data, which describes the spatial and temporal dynamics of the hand skeleton joints such as knuckles and fingertips. Thus, different from the motion of a single finger-tip, the hand skeleton motion data reflects the user's unique hand geometry, the signature curve of each joint, and the signing behaviors of the entire hand.

Our system first performs the 3D multi-joint motion data preprocessing to address different hand sizes, camera view angles, and sensor data lengths. In particular, we normalize the hand size and the time/spatial span of the hand skeleton trajectory and rotate the captured 3D hand to a reference orientation for alignment. After normalization and alignment, the hand skeleton motion trajectories are reconstructed in a 3D coordinate system. The core of our system consists of two parts: the derivation of signing behavior features and the CNN-based hand skeleton signature verification models.

Based on the reconstructed 3D hand skeleton motion trajectories, we derive the user's signing behavior at the skeleton-level and the joint-level: (1) The *three-view hand skeleton signature derivation* first projects the multi-joint trajectories onto three 2D planes, analyzing the 3D hand skeleton movements from its front, side, and top views, respectively. For each view, every joint trajectory is illustrated with a distinct color. Moreover, color gradients (from light to dark) are used to denote temporal information, and curve widths are used to describe joint significance, which is calculated based on how unique and independent joint moves regarding the entire hand skeleton. (2) The *inter-joint motion feature derivation* calculates the distance between every pair of hand joints to capture the minute inter-joint movements associated with hand motions. The time series of inter-joint distances thus describe the user's joint-level signing behaviors, which not only act as a liveness indicator to prevent machine forgeries but also presents identifiable characteristics.

The two types of behavioral feature time series are fed into our CNN-based hand skeleton signature verification model to construct two user profiles. The CNN-based three-view signature model scrutinizes the three-view hand skeleton motions to authenticate the user through the skeleton-level signing behaviors. The CNN-based inter-joint motion verification model examines the inter-joint motion characteristics to verify the user and mitigate the replay threats from motion-copy robots. We further develop a fusion model to integrate the examination results from both the skeleton-level and the joint-level to make the authentication decision. It is important to note that the hand geometry features obtained from the 3D hand skeleton model (i.e., the distance and angle relationships between hand joints), are embedded in the above behavioral feature time series and are implicitly verified by the above two CNN models.

3.4 Attack Models

The goal of an adversary is to spoof the target user's identity to authorize transactions or gain access to the user's account to infringe upon their privacy. To achieve this goal, the adversary needs to forge the user's in-air signature. We assume the adversary may have either the user's account information or the physical access to the user's devices (e.g., laptop) to initiate an in-air signature authentication session. Because our system is based on three biometric factors (i.e., hand skeleton motion behavior, inter-joint motion behavior, and 3D hand skeleton model), the adversary may be capable of getting one, some, or all of the biometric data to cheat our system. For example, in-air signing usually happens in public places and VR/AR scenarios, which could leak skeleton-based model/trajectories to a nearby hidden camera (2D or 3D). Then, the adversary could observe the eavesdropped data to impersonate or use a robot to replay. In particular, we consider two main types of attacks:

- **Impersonation Attack:** The adversary may obtain the user's name and signing behavior data (e.g., by observing how the user signs an in-air signature during an authentication session). Based on that, the adversary could choose to (1) sign the user's name with a random signing behavior or (2) practice hard to mimic the user's behavior and attempt reproducing the in-air signature.
- **Robot Replay Attack:** We assume a skilled adversary has access to both the user's 3D hand skeleton model and signature trajectory samples. In practical attacking scenarios, the adversary could employ hidden cameras or infiltrate the authentication database to gather this information [21, 24, 37, 39]. The adversary can further employ 3D printing technology to duplicate the user's hand and attach it to a robotic arm. The combination of 3D printing and robotics enables the adversary to reproduce not only the signing process but also the user's 3D hand skeleton model, which contains hand geometry biometrics.

4 APPROACH DESIGN

4.1 Hand Skeleton Motion Data Extraction

To maximize the user's hand biometrics and counter replay threats, we propose to augment the in-air signature with hand skeleton

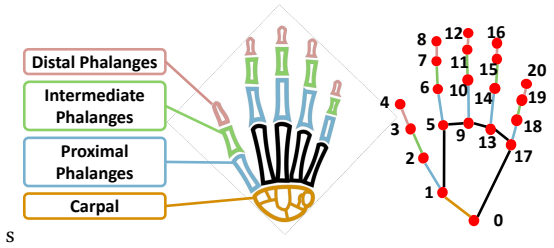


Figure 4: The skeleton and joints of a hand.

motions. A human hand comprises 27 bones, and its motion can be uniquely represented by 21 landmarks or joints. These landmarks include 20 phalangeal points (knuckles, fingertips, and joints) and one carpal point indicating the wrist position, as depicted in Figure 4. For example, Joint 0 represents the wrist, while Joints 4, 8, 12, 16, 20 represent the fingertips. These joints' 3D coordinates form a unique hand skeleton model for each individual and present hand geometry information. We express the 3D coordinate of joint i at time t as $Joint_{i,t} = [x_{i,t}, y_{i,t}, z_{i,t}]$. Then the hand skeleton motion S in a time period T can be uniquely described by a $T \times 21$ matrix of 3D coordinates:

$$S = \begin{bmatrix} Joint_{0,0} & Joint_{1,0} & Joint_{2,0} & ... & Joint_{20,0} \\ Joint_{0,1} & Joint_{1,1} & Joint_{2,1} & ... & Joint_{20,1} \\ ... & ... & ... & ... & ... \\ Joint_{0,T} & Joint_{1,T} & Joint_{2,T} & ... & Joint_{20,T} \end{bmatrix} \quad (1)$$

This matrix representation of the hand motion contains rich biometric information including the signature trajectory completed by each hand joint, the signing behavior exhibited by each hand joint, the hand geometry biometric presented by the 3D hand skeleton model, and the inter-joint relationships.

Incorporating all hand joints for authentication introduces significant complexity and burden to the authentication algorithm. It can result in high redundancy, as certain joints exhibit strong correlation and low independence. Additionally, the inclusion of less significant joints may dilute the contribution of more important joints, impacting authentication performance. As such, we propose to examine the user's 3D in-air signature based on a novel graphical representation, which projects the hand skeleton motion onto three orthogonal planes and presents the temporal information and joint significance in the resulting three views. We further derive the inter-joint distance relationships based on the hand skeleton motion matrix to capture the joint-level behavioral biometrics.

4.2 Multi-joint Data Normalization and Alignment

When entering an in-air signature, the user's hand presented to the camera may be captured with different sizes and orientations. The scale of the signature curve may not be consistent, and the varying signing speeds and completion times also lead to the sensor data with different lengths. We thus design the multi-joint data normalization and alignment schemes to address the above variances. In particular, we align the starting palm orientations with a predefined direction and then normalize the trajectory and hand sizes. Furthermore, we select two intersecting inter-joint links to represent the palm plane:

$$Link_1 = Joint_{9,0} - Joint_{0,0}, \quad (2)$$

$$Link_2 = Joint_{13,0} - Joint_{5,0}, \quad (3)$$

where $Link_1$ denotes the direction of the middle metacarpal bone. The magnitude of $Link_{1,t}$ at each time frame is chosen as a reference and scaled to a constant value to standardize the size of the skeleton.

$$scale_t = \frac{1}{|Link_{1,t}|} \quad (4)$$

To streamline the calculation, we postulate that $Link_2 \perp Link_1$, meaning $Link_2$ symbolizes the projected vector component of the link between $Joint_5$ and $Joint_{13}$, which is perpendicular to $Link_1$. Importantly, the skeleton trajectories are rotated so that $Link_1$ aligns with the Z-axis and $Link_2$ with the X-axis of the sensor coordinate at the initiation of each signature. This alignment is achieved through three rotations around the Z-axis and Y-axis. We select a point on the Z-axis, designated as P_z, such that $\vec{J_9P_z} \perp Link_1$ and the angles are calculated as per Equation (2) and (3):

$$\theta_1 = acos(\frac{Pr_{\Pi_{xy}} Link_1 \cdot \vec{x}}{|Pr_{\Pi_{xy}} Link_1| \times |\vec{x}|})), \quad (5)$$

$$\theta_2 = acos(\frac{Pr_{\Pi_{xz}} Link_1 \cdot \vec{z}}{|Pr_{\Pi_{xz}} Link_1| \times |\vec{z}|})), \quad (6)$$

$$\theta_3 = acos(\frac{Link_2 \cdot \vec{J_9P_z}}{|Link_2| \times |\vec{J_9P_z}|})). \quad (7)$$

Following the rotations, the aligned matrix S' is generated by:

$$S' = S \cdot R_z(\theta_1) \cdot R_y(\theta_2) \cdot R_z(\theta_3 - \pi), \quad (8)$$

where R_{axis} denotes the rotation matrix around an axis. The subsequent sections will delve into further normalization and transformation of the hand skeleton signatures.

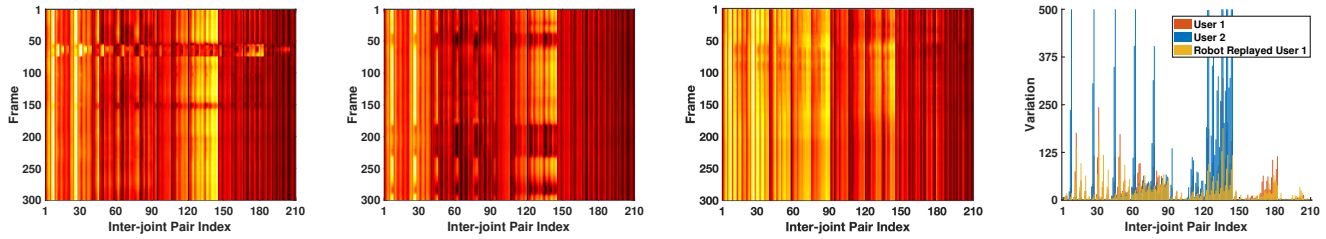
4.3 Inter-joint Motion Feature Derivation

The inter-joint motions are the relative distance relationships between hand joints. Based on S', we derive the inter-joint distance-time array D, which represents the spatial relationship between every pair of joints and their change during the process when the user completes a signature, as expressed by

$$D = \begin{bmatrix} |Joint_{0,0} - Joint_{1,0}| & ... & |Joint_{19,0} - Joint_{20,0}| \\ |Joint_{0,1} - Joint_{1,1}| & ... & |Joint_{19,1} - Joint_{20,1}| \\ ... & ... & ... \\ |Joint_{0,T} - Joint_{1,T}| & ... & |Joint_{19,T} - Joint_{20,T}| \end{bmatrix}. \quad (9)$$

In total, $\binom{21}{2} = 210$ inter-joint distance time series are obtained to describe the joint-level signing behaviors.

Figure 5 illustrates the inter-joint distances of all joint pairs and their variations over time. We find that the distances between some joints exhibit minimal or no variation over time. These joints are mainly at the rigid body parts of the hand, such as the palm bones, or are the joints whose movements are highly correlated. The joints of the index finger present the highest inter-joint distance variations, indicating the substantial relative motions of the index finger with regard to the overall hand skeleton movements. Other finger joints also present observable distance changes in reference to other hand joints. Moreover, by comparing Figure 5a and Figure 5b, we show that different users present different inter-joint distance patterns, which can be used for distinguishing users. By comparing the inter-joint motions of a user and the corresponding robot replay (e.g., Figure 5a and Figure 5c), we show that inter-joint motion features are more observable in human data than in the robot replay data. Thus, we use the inter-joint distance time series as both biometric features and the liveness indicator.



(a) User 1's Signature (b) User 2's Signature (c) User 1's signature replayed by a robot & a 3D-printed hand (d) Calculated variances of inter-joint distances over time

Figure 5: The varying inter-joint distances presented by in-air 3D signatures (210 joint pairs).

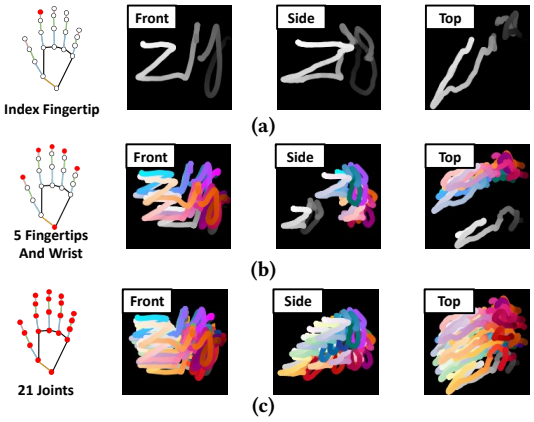


Figure 6: The three perceptions of a Leap Motion captured 3D signature when 1, 6, and 21 joints are used for illustration.

It is important to note that it is not necessary to examine all 210 inter-joint distances along time for authentication. More particularly, we create a personal *inter-joint motion profile* to only include the joint pairs which are found to be more identifiable to the user. The computation complexity is thus greatly reduced. Before constructing a per-user profile, we first apply interpolation or down-sampling to unify the inter-joint distance time series length. The distance values are normalized between 0 and 1.

4.4 Three View-based Biometric Feature Presentation

We also construct the per-user *hand skeleton signature profile* with its three-view derivations from hand skeleton motion matrix S', which contains both spatial and temporal information of the hand skeleton-level signature.

Presenting Spatial Information. We draw the hand skeleton signature in a 3D space using matrix S'. The signature trajectory is centralized and normalized within a $1\times1\times1$ bounding box. To fully represent the hand skeleton's motion, we observe the reconstructed signature from the front, side, and top views, projecting the 3D trajectories onto a plane to create an image, as depicted in Figure 6. This allows us to observe joint trajectories that may be obscured from one viewing angle from a different perspective. The three-view projection translates the hand skeleton motion matrix into three 180×180 images, preserving the user's behavioral biometrics across all three dimensions.

Presenting Temporal Information. To differentiate the joints, we assign each trajectory a distinct color. Moreover, prevent the loss of dynamic biometrics in the image representation, we integrate time information into the joint trajectories. We achieve this by plotting with a gradient color scheme that transitions from light to dark, effectively illustrating features like movement direction and velocity. If a user's movement is rapid over a certain period, the corresponding segment of the trajectory will exhibit a steep color gradient descent.

Joint Significance Estimation and Weight Assignment. Considering the complexity of human hand kinematics, simply illustrating the motion of all hand joints does not constitute an optimal method for presenting skeletal behavioral biometrics. There are two primary problems when attempting to depict the hand skeleton signature for a large number of joints. If we plot all 21 joints and generate three-view projection images, some trajectories invariably overlay others. Furthermore, not all joints maintain consistent trajectories during the signature enrollment phase. Certain jittering may represent potential user-specific behavioral biometrics, while others could be attributed to sensor noise or unintentional user movements. These inconsistencies may relate to the user's writing style, or the limited precision of the motion-tracking device and hand recognition algorithms during experiments. Therefore, we estimate and utilize the significance of the 21 joints.

$$rank_i = \sum_{j=0}^{20} variance(|Joint_i - Joint_j|) \tag{10}$$

$$score_i = 10 \times \frac{rank_i}{max(rank_{0,1,2,\ldots,20})} \tag{11}$$

We start by creating a joint significance ranking table for all 21 joints, setting an initial score of 0 for each. We then calculate the variance for each inter-joint pair, as depicted in Figure 5d. A large variance will occur if the distance between two joints changes significantly while writing the signature. Conversely, a constant distance between two joints results in a variance of 0. This variance is added to the ranking for both joints in the pair. The final step involves normalizing the sum of variances between 0 and 10 to create a significance score, which is saved as per-user joint significance ranking profiles. Specifically, Table 1, for instance, shows the average significance score for each joint among 25 users. Although each user's joint significance rankings differ due to factors such as

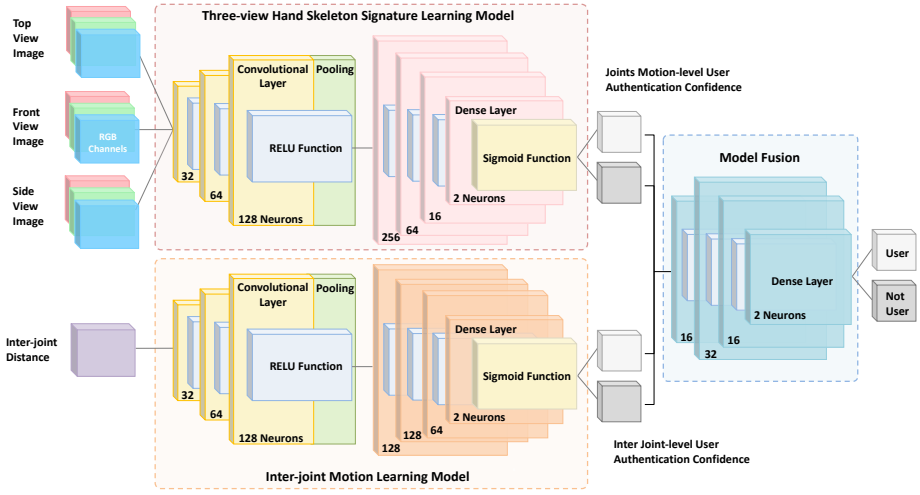


Figure 7: The CNN architecture of our verification model.

hand gestures, signature contents, and user habits, we observe that $Joint_8$ has the highest average score among the 21 joints. This is understandable since most users write with their pointing index finger and curved palm, which is why many existing works choose to use fingertip movement to represent 3D signatures.

By leveraging the user-specific joint significance, we can explore deeper into the user's behavioral biometrics in hand skeleton signatures and enhance the user verification performance. Furthermore, we can weight joint trajectories according to their significance ranking, thus the higher-score joints have wider trajectory widths to prevent blocking.

4.5 CNN-based Authentication Algorithm

Traditional in-air signature verification methods utilize the motion of a single joint, usually the tip of the index finger. These methods first extract features from the position and temporal information and then identify the user based on these features using clustering-based or distance-based algorithms, such as Support Vector Machine (SVM) and DTW. In contrast, our approach leverages the motion of all 21 joints and represents their behavioral biometrics in the form of a three-view hand skeleton signature and inter-joint distance time series. Therefore, we develop a CNN-based authentication algorithm for user verification. In this section, we detail outlines the architecture of our proposed network model and discuss the process of feeding the input into the model.

Table 1: The significance scores of 21 joints.

Joint	Avg. Score	Joint	Avg. Score
0	1.82	11	3.22
1	2.23	12	4.62
2	2.55	13	1.02
3	3.59	14	1.40
4	4.51	15	2.28
5	0.75	16	3.09
6	1.57	17	1.35
7	4.43	18	1.47
8	**8.38**	19	2.01
9	0.78	20	2.41
10	1.40		

Our network model is designed as a binary classifier, consisting of three components: the three-view hand skeleton signature learning model, the inter-joint motion learning model, and the model fusion component. The first two generate skeleton-level and joint-level user authentication confidence, while the fusion component integrates the outputs to make the final verification decision. The detailed architecture and model parameters are shown in Figure 7. The per-user CNN model has two output classes, *user* and *non-user*, which verifies the user against the claimed user identity.

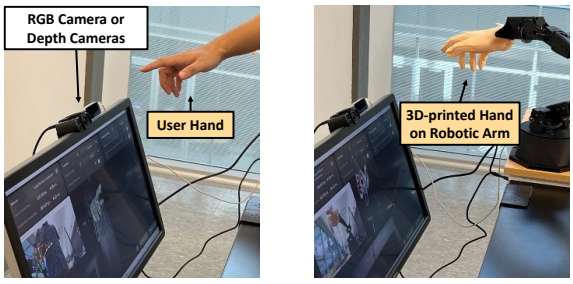
Each image has 3 color channels, so the three-view images are reshaped to have 9 RGB channels. The two feature learning models have similar structures but are trained independently. The initial part of these models is convolutional layers, which are crucial to CNN designs. The kernels in the convolutional layer compute the relationship between small, nearby portions of the input and derive a value map by scanning the input hand skeleton motion matrix. The output of these layers often contains high-level features like significant areas of an image.

Following this, the dense layers calculate scores with fully connected neurons. Each layer includes a Rectified Linear Unit (ReLU) activation function to keep the scores above 0. The output are two class scores determined by the last dense layer. A Sigmoid activation function is added after the final score to keep it between 0 and 1. These scores serve as the user authentication confidence for different features. Lastly, there is a model-level fusion that considers both skeleton-level and joint-level user authentication confidences to verify the users.

5 PERFORMANCE EVALUATION

5.1 Experimental Setup

5.1.1 ***Experimental Platforms.*** To evaluate the effectiveness of our proposed authentication system, we implement a hand skeleton



(a) (b)

Figure 8: In-air 3D signature capturing system.

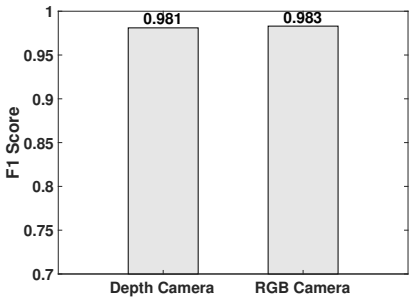


Figure 9: Verification performances of two sensors.

tracking platform on a Windows PC. This platform includes two different interfaces to study the performance of both scenarios: a regular RGB camera (specifically, an ELECOM Webcam) and a depth camera (Leap Motion). The setup of the experimental platform is shown in Figure 8a. In this setup, a user is able to write a signature in front of both cameras. The cameras are placed for easy access and convenience for general use cases such as video meetings, similar to the placement of a laptop's webcam. Other placements such as below the display [19], on the table [10], or on a tripod [38] are also possible. Real-time feedback in the form of virtual hand skeleton models is provided to the user on the monitor. The video frames captured during this process are used to extract both the user's 3D hand skeleton signature and the traditional single-point signature for purposes of comparison. As the majority of existing in-air signature systems rely on DTW algorithms and various motion features, including trajectory coordinates, velocities, accelerations, moving variances, and multiple statistic features [8, 19, 39], we also implement the in-air fingertip signature system that uses DTW and incorporates the same features as our baseline for comparison.

More specifically, we use Google MediaPipe v0.8 to monitor the hand captured by the webcam. This interface provides feedback by generating a 2D hand skeleton superimposed on the hand image. Although the webcam (operates at 30 fps) does not have the capacity to provide depth information directly, MediaPipe estimates the depth value of each joint based on its relative position to the palm center and the individual hand skeleton model, while the depth value of the palm center is fixed to be zero. Therefore, we can acquire the relative 3D coordinates of each joint. In contrast, with the depth camera (working at 110 fps), we use Leap Motion v5.4.1 to obtain the 3D coordinates of the hand's all joints. The hand skeleton collected is then reconstructed. However, instead of using the user's individual model, this interface bases the reconstruction on a general hand model that adheres to standard proportions between the joints.

5.1.2 ***Data Collection.*** We recruit 25 participants for our experiments. They are asked to perform their signature in the air, using the initials of their name and the writing style they were most comfortable with. Although most participants use their index finger to sign, a few used two fingers or opted for a more relaxed or fully stretched hand posture. Additionally, 17 participants are asked to write out the letters "ABC". This is done to examine their writing behaviors more closely while standardizing the content of their writing. Each participant repeats their in-air signature 40 times. These signatures are then divided into two halves, with one half being used for training purposes and the other for testing. During the training phase, 20 instances of the user's signature are labeled as those of a legitimate user, while 20 instances randomly selected from other users' signatures are labeled as non-user. This training and testing process is repeated 10 times for each user and is averaged for presentation.

5.1.3 ***Attack Setup.*** We imitate three types of attacks according to our attack models to evaluate the system's ability to withstand human impersonation attempts and robot-level motion replays.

Impersonation Attack. We choose five experienced participants to act as impersonation attackers. Their task is to observe how the target participants signed their names, then try to mimic both the signatures and the signing behaviors of the targets.

Physical Robot Replay. We employ a 3D scanner and a 3D printer to forge an exact replica of a user's hand, matching the original in terms of palm size, hand shape, and finger dimensions. Since the replica hand is an exact match, conventional hand image recognition methods might struggle to differentiate between the real hand and the fake one [41]. We then attach it to a PincherX 150 robotic arm as shown in Figure 8b, and program the fake hand's palm center to follow the same movement trajectory as the user's palm center, to reproduce the 3D hand skeleton signature.

Simulated Robot Relay. Physical robot replays have their limitations, including the robot's motion resolution, quantization errors, degrees of freedom, and task planning capabilities. To circumvent these limitations, we perform simulated robot replays using Python. This allows us to create a virtual hand model that precisely followed the user's hand motion data. We created two types of simulated attack scenarios based on whether or not the attacker is able to access the user's hand biometrics: 1) We create a 3D virtual hand model based on a general 3D human hand model. This virtual hand is moved in such a way that its palm center is aligned with the target user's palm center trajectory. The multi-joint motion data are collected during this process; 2) We further assume the adversary obtains the target user's hand model. The simulated attack is then executed by moving the virtual hand (created using the obtained hand model) along the user's palm center trajectory.

5.1.4 ***Evaluation Metrics.*** The metrics we use to evaluate our authentication system include false acceptance rate and the F1 score combining precision and recall of classification results. Specifically, we calculate these metrics based on the number of true positive TP,

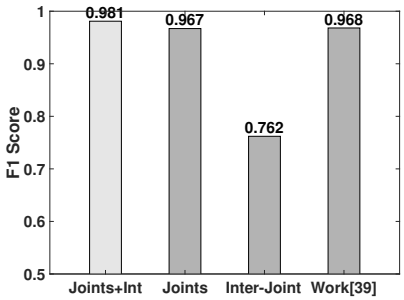


Figure 10: Verification Performance of different features.

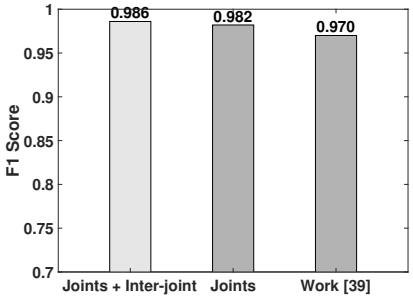


Figure 11: Distinguish users when they write the same word.

false negative FN, false positive FP, and true negative TN:

$$False\ Acceptance\ Rate = \frac{FP}{FP + TN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \tag{15}$$

5.2 User Authentication Performance

5.2.1 ***General User Verification.*** We first present the user verification result of 25 users. As seen in Figure 9, our system performs well with both types of motion capture devices, the 3D depth camera, and the 2D RGB camera, obtaining high F1 scores of 0.981 and 0.983 respectively. The result suggests that our system can effectively operate with different motion capture devices. Interestingly, the 2D RGB camera achieves an F1 score that is very close to the one achieved with the 3D camera. This suggests that the relative depth value, capable of illustrating changes in hand gestures, can be incorporated as a unique third-dimensional feature for user verification, augmenting the 2D signature.

5.2.2 ***Feature Significance.*** Next, we analyze the significance of two types of signing behavioral features: 3D hand skeleton motion features and inter-joint motion features, referred to as "Joints" and "Inter-joint" respectively. We also implemented a prior 3D signature verification method that uses a single fingertip [39] for comparison. The results, presented in Figure 10, show that our system attains an F1 score of 0.967 when solely using the 3D hand skeleton motion features. While this is slightly lower than the score achieved when combining both types of features, it is comparable to the existing single-point signature method. This suggests that our 3D hand skeleton signature alone is adequate for standard user authentication scenarios. Utilizing only the inter-joint motion features yields an F1 score of 0.762, indicating that while these features do offer unique biometric information, they alone are not sufficient for user identification. The integration of both types of features enhances authentication performance.

5.2.3 ***Writing The Same Word.*** We now concentrate on the participants' dynamic signing behaviors, disregarding the signature geometry as confidential. In particular, we ask the participants to write the common letters "ABC", simulating practical scenarios where people share the same names. The system's verification performance, illustrated in Figure 11, demonstrates an F1 score of 0.986. This score is similar to the general user verification performance highlighted earlier, confirming the efficiency of our system in verifying the user's hand skeleton-level signing behaviors as opposed to signature curves. When utilizing only the hand skeleton motion features, the F1 score is 0.982. This score is not only comparable to the one attained when using both types of features, but it also surpasses the performance of the single-point signature method [39]. Of note is the observation that writing "ABC" yields a slightly higher F1 score than writing name initials (i.e., two letters), suggesting that longer word lengths, such as full signatures, could lead to improved verification performance.

5.2.4 ***Training Effort Study.*** We also evaluate the impact of training data size on the verification performance of our system, which is related to the training efforts required from each enrolled user when deploying the system in practical scenarios. Figure 12 presents the verification performances when the user enrolls in our system with 2 to 20 signatures, respectively. We observe that increasing the training data size improves the system's performance. Specifically, the F1 score of our system improves significantly when increasing the enrollment signatures from 2 to 5. In particular, our system achieves over 0.9 F1 score with 5 enrolled signatures. The high performance we achieve with a small training data size from an enrolled user indicates the potential of deploying our system practically to provide authentication services. The main reason is that the hand skeleton motion data contain much richer behavioral biometric information than that of a single joint. The F1 score continues to increase to 0.95 when 10 signatures are obtained for enrollment, while the increasing trend becomes slow. When 20 signatures are requested from the user for model training, our system achieves 0.981 F1 score. While more training data brings performance improvement, it also requires higher enrollment efforts. Our future work will further reduce the individual training efforts by utilizing data augmentation. For example, we can use translation, rotation, and scaling to augment the training data size from limited enrolled signatures.

5.3 Impersonation Attacks

When an adversary only knows the user's name and is not aware of the signing behavior, our system's verification performance is comparable to the results reported in Section 5.2.3. In this subsection, we study the knowledgeable impersonation attacks, where the adversary knows the user's signature and signing behavior. Figure 13 presents the false acceptance rate of our system under knowledgeable impersonation attacks. Our system achieves a 2.04%

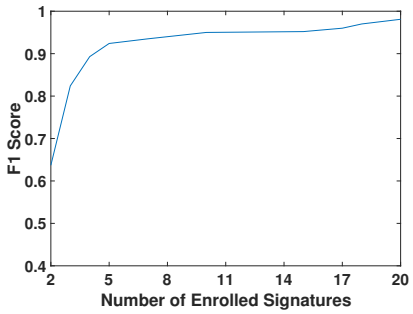


Figure 12: Impact of training data size.

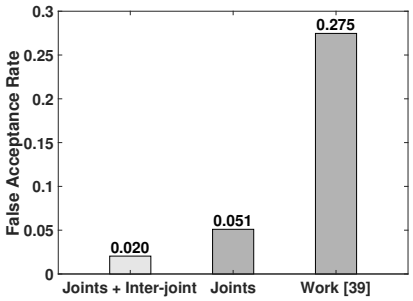


Figure 13: Performance under impersonation attacks.

false acceptance rate, which is much lower than the 27.5% rate reported by the single-point signature work [39]. The difficulty in distinguishing single-point signatures from knowledgeable human forgeries stems from two main reasons: 1) The single-point signature is relatively easy to imitate by an adversary. 2) The single-point signature suffers highly from the visual tracking errors incurred by occlusion or self-occlusion. Differently, our 3D hand skeleton signature leverages multiple joints to compensate for the partially occluded hand and examines hand skeletons' inherent behaviors.

5.4 Robot Replay Attacks

We evaluate the performance of our system to defend against robot replays with both a physical robotic arm platform and two simulated robot attacks. Figure 14 illustrates the performance of our system and compares it with the existing single-point signature method when a physical robot is used to replay. The robot is attached with the 3D-printed user hand, which is expected to replicate the user's hand skeleton motion features and hand geometry biometric. We observe that the conventional in-air signature authentication struggles with a high false acceptance rate of 32.2% when facing robot-level replays. In comparison, our system rejects 100% of robot-level replays. The result indicates that the current low-cost robotic arm has sufficient motion-copy capability to replay a user's in-air signature. However, our system successfully counters these attacks based on examining the hand skeleton's signing behaviors. When only checking the hand skeleton motion features, our system still achieves a low false acceptance rate of 2.5%. The inclusion of inter-joint motion features further enhances the system's resistance to replay attacks.

It is important to note that the low-cost robotic arm we implemented still has the limited capability to repeat the user's signing behavior. This is largely due to the fact that the robot operates on discrete control commands and possesses only a 5 Degree-of-Freedom (DoF), which results in the quantization errors and the limit to reach everywhere a human arm can reach. Furthermore, the motion capture device also introduces noise and distortion. Though using a more advanced robotic arm (e.g., equipped with a 7 DoF and superior motor resolution) achieves a higher accuracy of copying the user's hand motions, it is sill not enough to understand maximized attack. We thus simulate two types of robot replay attacks in software to explore the extent our system can achieve to prevent robot replays. In the simulations, the impacts from the motion capture device (when tracking the robotic arm) and the robot motor's quantization error and resolution are removed.

Figure 15 presents the performance of our system under a simulated robot replay that uses a generic hand model, which is expected to replay the hand skeleton motions. For the traditional single-point signature method, we find that the simulated replay attack leads a 96.8% false acceptance rate, which is the upper limit a physical robot replay could achieve. The result indicates that the traditional method cannot differentiate between a user's in-air signature and its simulated replica based on single-point motions. In comparison, our system achieves a significantly lower false acceptance rate of 0.2% under the simulated robot replay attack. This is because the attack is hard to replicate the joint-level motions.

Furthermore, we explore a more challenging attack scenario where the simulated replay attack employs a scanned 3D model of the user's hand rather than a generic hand model. This virtual hand model presents the same inter-joint geometries, hand shape, and finger widths/lengths. Figure 16 presents our system's performance under this advanced replay attack simulations. We find our system is still effective in defending against such attacks, maintaining a low false acceptance rate of 0.2%. The result confirms the robustness of our system against replay attacks, which is based on examining the joint and inter-joint motion features.

6 DISCUSSION AND FUTURE WORK

This work investigates the behavioral biometric security under the emerging robot-level replay threats. We implement the existing single-point in-air signature method [39] and conduct comparison studies in both normal and attack scenarios. Our experimental evaluation confirms the effectiveness of the single-point signature authentication in normal scenarios. By including the inherent behaviors of multi-joints, our system achieves a 1.5% improvement in normal authentication scenarios. But under robot-level replay attacks, we show that the traditional single-point signature verification is easy to be tricked by robot-level replay attacks.

Table 2: Performance of our method under replay attacks.

Attacking Scenarios	False Acceptance Rate		
	Work [39]	Our Method	
	Depth Cam.	Depth Cam.	RGB Cam.
Impersonation	0.275	0.020	0.027
Robot + 3D-print	0.322	0	0
Sim. Generic Hand	0.968	0.002	0.003
Sim. User's Hand	0.968	0.002	0.004

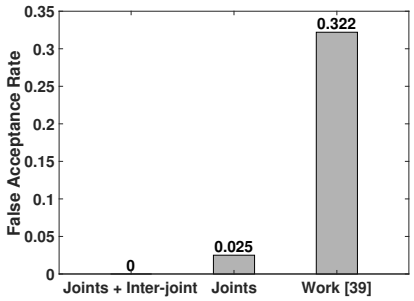


Figure 14: Under robot replay attack with 3D-printed user hand models.

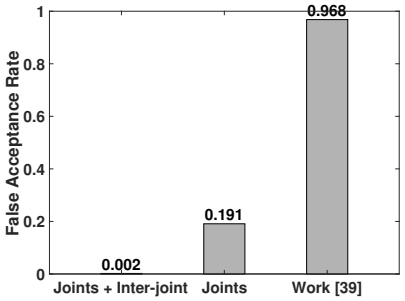


Figure 15: Under simulated replay attack with general 3D hand skeleton models.

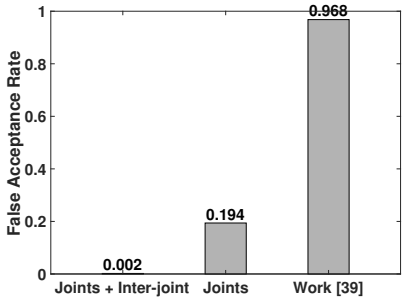


Figure 16: Under simulated replay attack with the user 3D hand skeleton models.

In comparison, our system achieves a 96.6% improvement in the verification performance to reject the replayed signatures while accepting the legitimate signatures. The comprehensive comparison results demonstrate the effectiveness of our system in providing replay-resistance signature verification services to face the threats of emerging motion-copy robots.

While our focus lies on the 3D in-air signature, the proposed authentication system and the motion-copy robot platform have the potential to be extended for 2D signatures on touchscreens and other behavioral biometrics. Some studies incorporate the writing pressure as a biometric feature in addition to the signature trajectory, implying security. However, a robot's haptic feature could potentially manipulate the pressure data while replicating the 2D signature. The commercial Kinect sensor also facilitates hand/body skeleton tracking and in-air signature [39]. Although Kinect can only display 4 skeleton joints of a hand, it captures 25 skeleton joints across the user's entire body. Therefore, our system could be extended to validate the body skeleton motions when a user executes an in-air signature. Additionally, our system relies on the ability of current commercial hand-tracking interfaces to function under low light conditions, an area extensively researched by visual sensing studies but outside this paper's scope. Furthermore, numerous 2D cameras feature a night vision mode employing infrared LED light, and 3D depth cameras use infrared light for active sensing, both functioning effectively under low light conditions.

The potential for robot-level signature replay extends beyond commercialized visual sensing interfaces and can incorporate other sensors such as wearable inertial sensors and wireless sensors such as wearable inertial sensors and wireless sensors (e.g., WiFi and acoustic). As an illustration, a robot can interface with inertial sensors to generate data analogous to that used in behavioral biometric authentication. Machine-level replay attacks on wireless sensors can occur via the injection of manipulated RF signals and acoustic sounds, though not using robotic arms. As robotics improve, we foresee an escalating threat to behavioral biometrics. Therefore, continued research is paramount to both investigate these replay threats and secure behavioral biometric-based authentication.

An alternative approach could involve integrating liveness detection with behavioral biometrics, thereby avoiding modifications to the already deployed authentication systems. Image recognition methods, for instance, can be utilized to differentiate a human hand from a mechanical robotic arm, as demonstrated in Figure 8. However, such distinctions can be complicated if the robotic arm is concealed by clothing to hide its joints, or if the attached artificial hand is difficult to discern from a real hand based on image analysis alone. While infrared cameras can be added for liveness detection, the associated overhead is substantial. Conversely, we believe that by extracting the minute human motion behaviors, such as the inter-joint motions, we can extend the security lifespan of behavioral biometrics by another two decades. Given that our three-view projections of the 3D hand skeleton signature contain inter-joint motion information, future research will strive to maximize the utility of these three-view hand skeleton images and further investigate joint-level features.

7 CONCLUSION

This work introduces a 3D hand skeleton signature verification system to address behavioral biometric security under the threat of emerging motion-copy robots. We find that the hand skeleton-level signatures and joint-level motions demonstrate great resilience to robot replays compared to the traditional single-point signatures (fingertip or palm center). We develop a CNN-based algorithm to augment the security of in-air signatures that are obtained by two types of commercial visual sensing interfaces. Specifically, we propose a novel presentation method to depict hand skeleton motions in projected three-view images, assigning each joint trajectory a unique color. The color gradient elucidates the temporal information, while the line width of the curve highlights the individual joint's significance. Additionally, we compute the distances between all joint pairs, employing the inter-joint distance time series to illustrate inter-joint motions. We also develop a CNN-based algorithm to learn from these two types of signing behavior features for authentication. For evaluation, we develop a real robotic arm platform with a 3D-printed hand to launch robot replay attacks and also explore the theoretical attack performance based on simulations. Experimental results show that our system matches the user authentication performance of existing in-air signature methods while thwarting 100% physical robot replay attacks.

ACKNOWLEDGMENTS

This work was partially supported by LABoR LEQSF(2020-23)-RD-A-11 and NSF CNS-2155131. We would also like to thank our anonymous shepherd and all the reviewers for helping us improve the paper.

REFERENCES

[1] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. 2015. Wigest: A ubiquitous wifi-based gesture recognition system. In *2015 IEEE conference on computer communications (INFOCOM)*. IEEE, 1472–1480.

[2] Abdulaziz Alzubaidi and Jugal Kalita. 2016. Authentication of smartphone users using behavioral biometrics. *IEEE Communications Surveys & Tutorials* 18, 3 (2016), 1998–2026.

[3] Athos Antonelli, Raffaele Cappelli, Dario Maio, and Davide Maltoni. 2006. Fake finger detection by skin distortion analysis. *IEEE Transactions on Information Forensics and Security* 1, 3 (2006), 360–373.

[4] Kai Cao and Anil K Jain. 2018. Automated latent fingerprint recognition. *IEEE transactions on pattern analysis and machine intelligence* 41, 4 (2018), 788–800.

[5] Hayreddin Çeker and Shambhu Upadhyaya. 2016. User authentication with keystroke dynamics in long-text data. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 1–6.

[6] YU-TZU CHIU. 2013. *In-Air Signature Gives Mobile Security to the Password-Challenged*. https://spectrum.ieee.org/inair-signature-gives-mobile-security-to-the-passwordchallenged

[7] Nesli Erdogmus and Sebastien Marcel. 2014. Spoofing face recognition with 3D masks. *IEEE transactions on information forensics and security* 9, 7 (2014), 1084–1097.

[8] Yuxun Fang, Wenxiong Kang, Qiuxia Wu, and Lei Tang. 2017. A novel video-based system for in-air signature verification. *Computers & Electrical Engineering* 57 (2017), 1–14.

[9] Luca Ghiani, David Yambay, Valerio Mura, Simona Tocco, Gian Luca Marcialis, Fabio Roli, and Stephanie Schuckcrs. 2013. Livdet 2013 fingerprint liveness detection competition 2013. In *2013 International Conference on Biometrics (ICB)*. IEEE, 1–6.

[10] Elyoenai Guerra-Segura, Aysse Ortega-Pérez, and Carlos M Travieso. 2021. In-air signature verification system using leap motion. *Expert Systems with Applications* 165 (2021), 113797.

[11] Long Huang, Zhen Meng, Zeyu Deng, Chen Wang, Liying Li, and Guodong Zhao. 2021. Towards Verifying the User of Motion-controlled Robotic Arm Systems via the Robot Behavior. *IEEE Internet of Things Journal* (2021).

[12] Long Huang and Chen Wang. 2021. Unobtrusive Pedestrian Identification by Leveraging Footstep Sounds with Replay Resistance. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–19.

[13] Fortune Business Insight. [n. d.]. *Virtual Reality Market Size, Share & Covid 19 Impact Analysis*. https://www.fortunebusinessinsights.com/industry-reports/virtual-reality-market-101378

[14] Anil K Jain, Friederike D Griess, and Scott D Connell. 2002. On-line signature verification. *Pattern recognition* 35, 12 (2002), 2963–2972.

[15] Meenakshi K Kalera, Sargur Srihari, and Aihua Xu. 2004. Offline signature verification and identification using distance statistics. *International Journal of Pattern Recognition and Artificial Intelligence* 18, 07 (2004), 1339–1360.

[16] Alisher Kholmatov and Berrin Yanikoglu. 2005. Identity authentication using improved online signature verification method. *Pattern recognition letters* 26, 15 (2005), 2400–2408.

[17] Paweł Kobojek and Khalid Saeed. 2016. Application of recurrent neural networks for user verification based on keystroke dynamics. *Journal of telecommunications and information technology* 3 (2016), 80–90.

[18] Hao Kong, Li Lu, Jiadi Yu, Yingying Chen, and Feilong Tang. 2020. Continuous authentication through finger gesture interaction for smart homes using WiFi. *IEEE Transactions on Mobile Computing* 20, 11 (2020), 3148–3162.

[19] Shih-Hsiung Lee and Hsuan-Chih Ku. 2022. SIAR: Signing in the Air Using Finger Tracking Technology for Authentication on Embedded System. In *2022 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 1–2.

[20] Gen Li and Hiroyuki Sato. 2020. Handwritten signature authentication using smartwatch motion sensors. In *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 1589–1596.

[21] Duo Lu and Dijiang Huang. 2018. Fmcode: A 3d in-the-air finger motion based user login framework for gesture interface. *arXiv preprint arXiv:1808.00130* (2018).

[22] Duo Lu, Dijiang Huang, Yuli Deng, and Adel Alshamrani. 2018. Multifactor user authentication with in-air-handwriting and hand geometry. In *2018 International Conference on Biometrics (ICB)*. IEEE, 255–262.

[23] Duo Lu, Kai Xu, and Dijiang Huang. 2017. A data driven in-air-handwriting biometric authentication system. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 531–537.

[24] Jameel Malik, Ahmed Elhayek, Sheraz Ahmed, Faisal Shafait, Muhammad Imran Malik, and Didier Stricker. 2018. 3dairsig: A framework for enabling in-air signatures using a multi-modal depth sensor. *Sensors* 18, 11 (2018), 3872.

[25] Pedro Melgarejo, Xinyu Zhang, Parameswaran Ramanathan, and David Chu. 2014. Leveraging directional antenna capabilities for fine-grained gesture recognition. In *Proceedings of the 2014 ACM International Joint Conference on pervasive and ubiquitous computing*. 541–551.

[26] David Menotti, Giovani Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha. 2015. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security* 10, 4 (2015), 864–879.

[27] Meta. 2022. *META QUEST 2*. https://store.facebook.com/quest/products/quest-2

[28] Microsoft. 2022. *Kinect for Windows*. https://docs.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows

[29] Kien Nguyen, Clinton Fookes, Arun Ross, and Sridha Sridharan. 2017. Iris recognition with off-the-shelf CNN features: A deep learning perspective. *IEEE Access* 6 (2017), 18848–18855.

[30] Javier Ortega-Garcia, Joaquin Gonzalez-Rodriguez, Danilo Simon-Zorita, and Santiago Cruz-Llanas. 2002. From biometrics technology to applications regarding face, voice, signature and fingerprint recognition systems. In *Biometric Solutions*. Springer, 289–337.

[31] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. 2015. Deep face recognition. (2015).

[32] App PKC. 2017. *AirSig Vive Sample Code Demo for Unity*. https://www.youtube.com/watch?v=7ekfjhsNmcM

[33] ABM Mohaimenur Rahman, Yetong Cao, Xinliang Wei, Pu Wang, Fan Li, and Yu Wang. 2022. On the Feasibility of Handwritten Signature Authentication Using PPG Sensor. In *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 719–720.

[34] Yanzhi Ren, Yingying Chen, Mooi Choo Chuah, and Jie Yang. 2014. User verification leveraging gait recognition for smartphone enabled mobile healthcare systems. *IEEE Transactions on Mobile Computing* 14, 9 (2014), 1961–1974.

[35] Md Sagar Hossen, Tasfia Tabassum, Ashiqul Islam, Rafat Karim, Laila Sultana Rumi, Aysha Akter Kobita, et al. 2021. Digital signature authentication using asymmetric key cryptography with different byte number. In *Evolutionary Computing and Mobile Sustainable Networks*. Springer, 845–851.

[36] Abdul Serwadda and Vir V Phoha. 2013. When kids' toys breach mobile phone security. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 599–610.

[37] Yubo Shao, Tinghan Yang, He Wang, and Jianzhu Ma. 2020. AirSign: Smartphone Authentication by Signing in the Air. *Sensors* 21, 1 (2020), 104.

[38] Akiji Takeuchi, Yusuke Manabe, and Kenji Sugawara. 2013. Multimodal soft biometrie verification by hand shape and handwriting motion in the air. In *2013 International Joint Conference on Awareness Science and Technology & Ubi-Media Computing (iCAST 2013 & UMEDIA 2013)*. IEEE, 103–109.

[39] Jing Tian, Chengzhang Qu, Wenyuan Xu, and Song Wang. 2013. KinWrite: Handwriting-Based Authentication Using Kinect.. In *NDSS*, Vol. 93. Citeseer, 94.

[40] Ultrleap. [n. d.]. *Leap Motion Controller*. https://www.ultraleap.com/product/leap-motion-controller/

[41] Ruxin Wang, Kaitlyn Madden, and Chen Wang. 2022. Low-effort User Authentication for Kiosk Systems based on Smartphone User's Gripping Hand Geometry. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–6.

[42] Jonathan Wu, Prakash Ishwar, and Janusz Konrad. 2016. Two-stream CNNs for gesture-based verification and identification: Learning user style. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 42–50.

[43] Hang Yin, Patrícia Alves-Oliveira, Francisco S Melo, Aude Billard, and Ana Paiva. 2016. Synthesizing robotic handwriting motion by learning from human demonstrations. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*.

[44] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. Mediapipe hands: On-device real-time hand tracking. *arXiv preprint arXiv:2006.10214* (2020).

[45] Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous authentication using eye movement response of implicit visual stimuli. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–22.

[46] Nan Zheng, Kun Bai, Hai Huang, and Haining Wang. 2014. You are how you touch: User verification on smartphones via tapping behaviors. In *2014 IEEE 22nd International Conference on Network Protocols*. IEEE, 221–232.