

Behavioural Biometrics in VR

Identifying People from Body Motion and Relations in Virtual Reality

Ken Pfeuffer¹, Matthias J Geiger², Sarah Prange^{2,3}, Lukas Mecke^{2,3}, Daniel Buschek², Florian Alt¹

¹Bundeswehr University Munich, Germany; ken.pfeuffer@unibw.de, florian.alt@unibw.de

²LMU Munich, Germany; matthias.geiger@campus.lmu.de, daniel.buschek@ifi.lmu.de

³University of Applied Sciences Munich, Germany; sarah.prange@hm.edu, lukas.mecke@hm.edu

ABSTRACT

Every person is unique, with individual behavioural characteristics: how one moves, coordinates, and uses their body. In this paper we investigate body motion as behavioural biometrics for virtual reality. In particular, we look into which behaviour is suitable to identify a user. This is valuable in situations where multiple people use a virtual reality environment in parallel, for example in the context of authentication or to adapt the VR environment to users' preferences. We present a user study (N=22) where people perform controlled VR tasks (pointing, grabbing, walking, typing), monitoring their head, hand, and eye motion data over two sessions. These body segments can be arbitrarily combined into body relations, and we found that these movements and their combination lead to characteristic behavioural patterns. We present an extensive analysis of which motion/relation is useful to identify users in which tasks using classification methods. Our findings are beneficial for researchers and practitioners alike who aim to build novel adaptive and secure user interfaces in virtual reality.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**;

KEYWORDS

Virtual Reality, Behavioural Biometrics, Motion, Relation, Proprioception, Adaptive UIs

1 INTRODUCTION

Every person is unique with individual behavioural characteristics. A particularly salient one is physical body motion: how one moves, coordinates, and uses body segments in relation to each other [14, 32]. The idea originates back to the Roman author Pollio's thoughts on the correlations of ideal human proportions [37], to which every person deviates in some way. This creates the potential of using body relations as biometrics.

In virtual reality systems, body motion is captured to provide a more immersed experience, ranging from head position and angle, via hand movements, to eye tracking and full body tracking capabilities [9, 26]. The increasing proliferation of VR devices will soon allow large-scale inferences about user activity, movements, and behaviour to be made. This research explores how motion data can be useful for behavioural biometrics to identify users in VR.

Implicit biometric identification of users in VR opens many opportunities. Authentication can be realized through behavioural biometrics, hence minimizing interruptions usually caused by explicit authentication schemes [33, 45]. As the user's identity can be continuously assessed in the background, potential unauthorised access can be detected (and the device locked) as the original user is still logged in. Furthermore, usability can be enhanced. For example, in situations where devices are shared among people, the system can detect a new user and adapt the UI to their preferences or offer personalised content. For example, if within a family an adult is identified from the first movements, an overview of the day's work could be shown, whereas when a child is identified, their favourite cartoon may be triggered.

Therefore, we focus on how body motion, and the relations between body segments may be useful as behavioural biometric in VR. This is interesting as many body movements are not specific to one segment, such as the hand, but often coordinate with other parts of the body such as the arm relative to body posture or hands to each other, involving motor control, proprioception, or hand-eye coordination [14, 22, 32]. Research showed that combinations of biometric features can enhance identification accuracy [3, 5, 21]. Yet, there are numerous combinations of body motion features.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2019, May 4–9, 2019, Glasgow, Scotland UK
© 2019 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5970-2/19/05...\$15.00

<https://doi.org/10.1145/3290605.3300340>

In this paper, we first analyse body motion and how movements may relate to each other to identify body feature sets of interest. We then conduct a user study to collect motion data, particularly head, hands, and eye movement that can be captured in current VR devices such as the HTC Vive, Oculus Rift, or FOVE. The study involves common VR interactions including pointing, grabbing, walking, and typing as fundamental tasks to interact with virtual content. Lastly we evaluate the motion data with machine learning methods to analyse various body motion combinations with regards to how accurate they allow user identification and authentication. We present detailed results of which motion is useful to identify users in which tasks. For this reason, our main contribution is in understanding which body relations are more accurate to describe individual user behaviour.

Contribution Statement. First, we provide a theoretical analysis of body motion theories as basis for behavioural biometrics. Second, we report on a user study (N=22) in which we collected detailed motion recordings of hand, head, and eye movements during different tasks. Third, we contribute an analysis that reveals body features suitable for user identification as well as authentication and assess accuracy.

2 RELATED WORK

Behavioural biometrics research explores what kind behaviour may be unique to a user. The goal is to enable systems to identify and authenticate users – usually in an implicit manner. Systems can become more resistant to unauthorised access, they can improve usability by avoiding physical or digital tokens such as a key or a password [33, 45], or they can adapt their UI to the user. The procedure can run in the background as users interact normally without requiring to explicitly provide identifiable information, such as a fingerprint or face ID. On current smartphones, researchers, for example, showed how touch motion behaviour can be used to identify users [7, 10, 12, 17, 23].

Use of body motions as behavioural biometrics has a long tradition in security related research [13, 49]. Gait has been extensively researched in the context of various possible features as users inhere unique walking patterns [48]. Research on anthropomeasures includes body motion as a whole, e.g. by using depth sensors such as the Kinect, for person identification [1, 35]. Estimation of body height, shoulder breadth, or stride information from cameras can be useful to design systems that identify users from multiple biometric sources [5, 21]. A user’s eye movement information such as saccadic vigor, acceleration cues, and response to visual stimuli have shown promising results on user authentication [39, 47]. These works show that various body movements are accurate measures for person identification, which motivates us to explore these in virtual reality systems.

Researchers investigated behavioural biometrics for mobile, wearable, VR or AR devices. Kupin et al. assessed task-driven biometric authentication and showed that using a ball-throwing task, 14 users could be accurately identified [25]. Mustafa et al. showed how head pointing motion from Google Cardboard sensors can be used to identify users [36]; other researchers focused on head nodding to music [27] and explicit head gestures from Google Glass [46]. GaitLock can authenticate users during walking tasks from the on-board inertial measurement units of a Google Glass to protect VR and AR systems [40]. Ashby et al. show how electroencephalography in response to mental imagery tasks allow accurate user authentication on low-cost headworn sensors [2]. Brain Password leverages brain responses in relation to visual stimuli as biometric features for smart headwear [29]. Prior work focuses on specific aspects of the user’s body motion, which we complement with a more holistic view on the combination of body motions.

Our work also contributes to the notion of soft biometrics – behavioural traits that may not be unique features, but can complement primary identifiers or be used temporarily [20]. Reid and Nixon introduced a method to gather relative measures based on comparative human descriptions. This allowed subjects in video data to be accurately retrieved [38]. Dantcheva et al.’s survey provides an exhaustive review of research on different user characteristics, demonstrating their use for person identification from soft biometrics [11].

The literature shows that body motions are promising features with most papers focusing on single body segments or sensors. As the body provides a plethora of motions as potential features that are potentially related, we investigate body motion and spatial relations to identify users in VR.

3 BODY MOTION & BEHAVIOURAL FEATURES

In HCI, the notion of Reality Based Interaction emphasises that human-computer systems should take advantage of pre-existing knowledge of reality; of relevance here are the themes of *Body Awareness & Skills* and *Environmental Awareness & Skills* [19]. We discuss body relation theories interesting for biometrics. Figure 1 illustrates the theories of body motion we consider. We apply these theories in context of motion types available in current VR systems:

- **Head:** the inertial motion sensors of the VR headset allow estimation of head position (*HeadPos*), direction (*HeadRay*) and rotation.
- **Hands:** controllers allow estimating hand motion for both dominant (DH) and non-dominant hands (NDH). One can consider each hand’s position and pointing ray separately (*DHPos*, *NDHPos*, *DHRay*, *NDHRay*).
- **Gaze:** by using eye trackers, we can track the direction (ray) where the user is looking in the virtual scene.

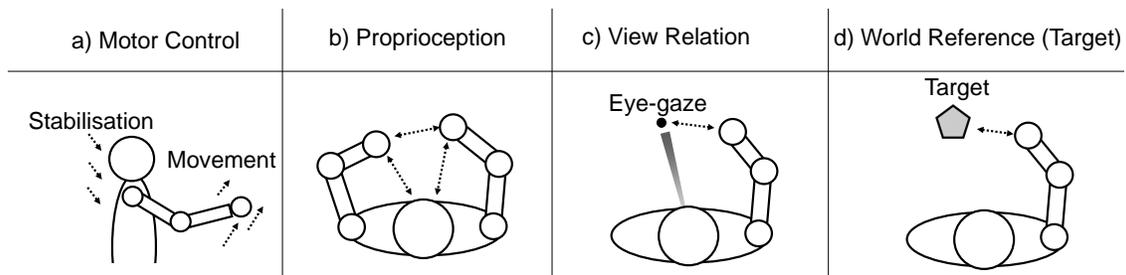


Figure 1: Body based motion and relations useful as biometric features.

Motor Control (Figure 1a)

In neuroscience research, Massion emphasises two types of motor control that users can perform to fulfill a goal, considered here as *stabilisation* and *movement* [32]. *Stabilisation* occurs when humans maintain a reference value against external and internal disturbances, for example, to stabilise body posture against wind. *Movement* denotes the displacement of body segments toward a goal, for example, when reaching out or grasping an object. Most motor acts have both types of motor control combined, such as when raising an arm, leg muscles are activated for postural control [4].

VR interactions can involve both types of the motor tasks. For example, in a pointing task the user may first stabilise their body to an optimal posture for the manual act, and then perform the pointing movement. Thus, to describe a pointing task more comprehensively to better capture individual user behaviour, it may be relevant to include body position too (here we consider the HMD location as approximation of body position). Head position itself may be a good feature, as every user will have a different body height. Thus, particular feature combinations can be derived from:

- Position of body segment (head, hands)
- Movement/Motion of body segment (head, hands)

Proprioception (Figure 1b)

Proprioception is the ability of humans to sense the position, orientation, and movement of their limbs, muscles, and joints in relation to each other [14]. In HCI, it has been suggested to enhance the UI, for example, for wearables [30] or virtual reality interfaces [34]. Users frequently take advantage of their proprioceptive senses, with the body providing a physical reference frame in which to operate direct, precise, and potentially eyes-free. For manipulation, holding an object in the hand provides a good sense of its position and greater sense of control. Conversely, understanding and controlling an object can be more difficult when using non-linear mappings between hand and object [8]. Another example of proprioception are Guiard's observations on bimanual asymmetry [16]. The NDH tends to precede the action of the

DH, and acts as the spatial frame of reference (potentially supported by proprioceptive senses). For example, in writing, the NDH is orienting the paper while the DH is writing.

In the context of behavioural biometrics, how do the relative movements between segments of the users body contribute to biometric behaviour in contrast to absolute measures? We can sense the triangular reference frame that users span out with their hands to their head, i.e. how they move their hands relative to the body root position. Furthermore, we investigate two manual tasks – grabbing (grab target directly in hand) with a linear mapping and pointing (users point with the controller to a remote target) – as an example for a non-linear mapping. Particular features are:

- Spatial relation between left hand and body (head)
- Spatial relation between right hand and body (head)
- Spatial relation between DH and NDH

View Relation (Figure 1c)

User interactions are often guided by the visual stimuli available to the user. For example, eye-hand coordination describes the user's spatial coordination of their eye gaze in relation to their hands. The eyes can direct the hands to a target, visually identifying relevant information for the hands to act [22, 28]. On the other hand, manual action of the user may guide the visual, for example, when drawing a line the eyes follow the hand's interaction closely. In the context of walking, the user visually inspects the space around them to navigate through paths and avoid obstacles.

We aim to see whether users inhere distinct eye-to-body relations. Users with less experience in manual tasks may be subject to a more visual approach, where they want to closely follow their interaction toward the target. Experienced persons can guide themselves eyes-free and interact, for example, simply by muscle memory; other users may be somewhere in between. Therefore, as features we add:

- Spatial relation between eye gaze and head ray
- Spatial relation between eye gaze and controller ray

World Reference & Target (Figure 1d)

User behaviour can be influenced by external factors of the surrounding world. In VR/AR systems, digital information and manipulation targets can be spatially located relative to their view, body, or the surrounding world [6]. For instance, in a manipulation task, the user reaches out and manipulates an object in space, and how they reach may depend on each user's individual strategy. As users visually acquire the target before manual action [22, 44], eye movements may similarly contribute to user strategy. This related to understanding biometric user behaviour in the context of Environmental Awareness & Skills when acting toward objects relative in physical space [19]. Based on this, we explore how spatial relations between segments of the user and the target in world space may be useful as biometric behaviour. In particular:

- Spatial relation between hand and target
- Spatial relation between eye gaze and target
- Spatial relation between head direction and target

4 STUDY

In this study we gather user motion data (head, hands, eyes) in common VR interactions across two sessions. The goal is to investigate and better understand user behaviour based on various motion and assess their biometric quality.

Task and Study Design

Our goal was to find a few tasks that represent generic tasks in VR. We selected four tasks with the help of a review of related work and current applications in VR. We now describe the tasks. For details on the layout see Figure 2.

We conducted a within-subjects study design with repeated measures, over two sessions with at least 3 days in between. The conditions are specific to each task.

Pointing (Figure 2a). Pointing at a target is a fundamental task as most VR systems employ raycasting controllers. Here a pointing tasks with the user hovering over a target with the controller ray, at which its color changes from dark blue to light blue for indicating this input state. A target selection is valid if the trigger click and release events are both executed in target hover state. After a valid selection, the selected target is hidden and the next target is shown to the user.

- Target distance condition: close (2 m) and far (4 m)
- Distance between targets: 2m
- Target positions: 13
- Repetitions: 10
- Overall: 260 trials per session

Grabbing (Figure 2b). Grabbing is another in elemental task in VR, where users can directly move their hand (represented by the controller) to a target for selection. The selection finishes

when any part of the controller is within the target (feedback given by target highlighting) and the user clicks and releases. Two conditions are used: big and small targets.

For both pointing and grabbing tasks, arrangement of targets is based on the multi-directional tapping task described in the ISO9241-9 standard [18] adapted to 13 targets based on Teather and Stuerzlinger's evaluation of 3D pointing [42, 43]. 13 spherical targets are ordered in a circular arrangement for each condition (2 m / 4 m distance to user position), and users select them by alternating clockwise around the circle.

- Target size condition: small (0.15 cm) and big (0.30 cm)
- Distance between targets: 0.7 m
- Target positions: 13
- Repetitions: 10
- Overall: 260 trials per session

Walking (Figure 2c). Users navigate through the virtual space and physical walking is a prominent method to move in the local space. In this context, gait has been shown as a promising feature [40, 48]. We complement this work with a perspective on body user motion.

In this task users are walking from a start to an end position. The end position is visualised by an outline of a 2 meter high door. When the user's head position collides with the layer, the task is finished and the next layer visualised. The user turns around and walks to it. To fit the tasks within the tracking range of the VIVE lighthouse, the path of the tasks were designed accordingly. Two paths were chosen that each include 4 straight line movements (2 short and 2 long lines). The two paths were counterbalanced.

- Walk path condition: start left or straight (Figure 2c)
- Walk distance: short (1.7 m) and long (2.4 m)
- Repetitions: 20
- Overall: 40 trials per session

Typing (Figure 2d). Typing behaviour has been shown to be a promising feature for mobile devices [10]. In VR, an efficient typing method is tracked by hand-held raypointing controllers [41], which we chose for our investigation.

We designed a typical typing task where the user sees keyboard, target sentence, and textfield on a screen 3.75 m in front of the user. The keyboard was 4×1.6 m large and each key 0.4 m^2 . 15 sentences were randomly selected from MacKenzie and Soukoreff's phrase sets [31]. The average sentence length was 25.8 letters (SD=3.14). The key is selected when users hover over it and click (button_down event). When users type a wrong letter, they can correct it. A trial finishes when all letters were written. The user can make a break or continue with the next by clicking the touchpad.

- Sentences: 15
- Letters per sentence: 25.8 (SD=3.14)

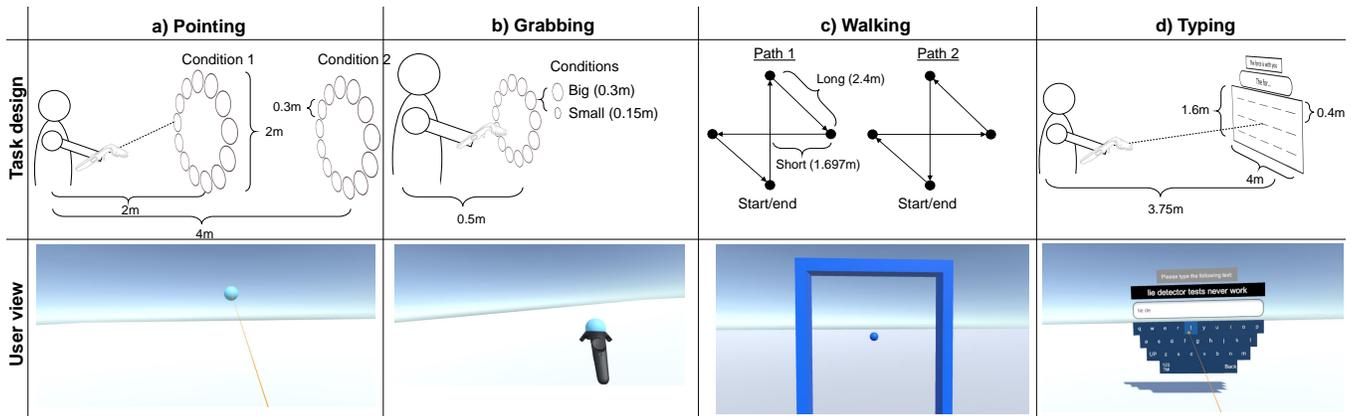


Figure 2: Study design and a VR screenshot of the user's perspective for each task, respectively.

Procedure

Each user performed two sessions with a pause in between of at least 3 days. A session started with a brief introduction by the experimenter and a demographic questionnaire. Then, the user performed the tasks. An eye tracker calibration was conducted before each task. The user performed blocks of tasks (for example, 13 targets in succession), after which the user had the option to take a break. On average, each session of the study took approximately one hour.

Setup

The VR setup is shown in Figure 3. It is based on an HTC Vive (HMD, two controllers, optical tracking) which is equipped with an additional eye tracker (Pupil Labs). The VR application is implemented in C# in Unity 3D on a Windows 10 PC (Intel Core i7-6700, 32GB RAM, NVIDIA GeForce 1080). The eye tracker's accuracy was measured in a 9-point accuracy test at the study end ($M=5.46^\circ$, $SD=9.36^\circ$).

Data Log

During all tasks, the VR application logs position, rotation, velocity and angular velocity of each device (HMD, controllers) as well as the current gaze point and collision points of the devices' raycasts (HMD, hand controllers, gaze) related to the area of interaction with a sample rate of 100 Hz.

Participants

A total of 22 paid participants (4 female, 3 left-handed) took part in the study. On average they were 26.05 years old ($SD: 3.81$, range: 19–35 years) and had an average height of 179.9 cm ($SD: 8,66$). They were either students from the local university or employees with mixed background. 6 participants wore glasses, 2 wore contact lenses and 14 had no constraints regarding eye-sightedness. Regarding experience in virtual reality on a scale from 1 (no experience) to 5 (very experienced), users were little experienced ($M: 2.1$, $SD: 0.86$).



Figure 3: A user during the study with the VIVE VR headset.

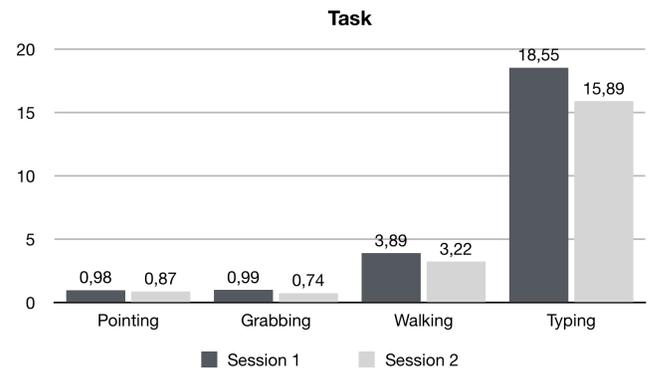


Figure 4: Task completion times for each task and session.

5 DATA OVERVIEW

To understand the study data, we plot a few tasks. Figure 4 shows the task completion times for each task/session. Across sessions, there is a learning behaviour. Note that the typing task is accomplished when entering the whole sentence, hence the longer times task completion times.

To further investigate the learning rate, we plotted task completion times over the conditions of the study. Here we focus on the pointing and grabbing tasks, as they have

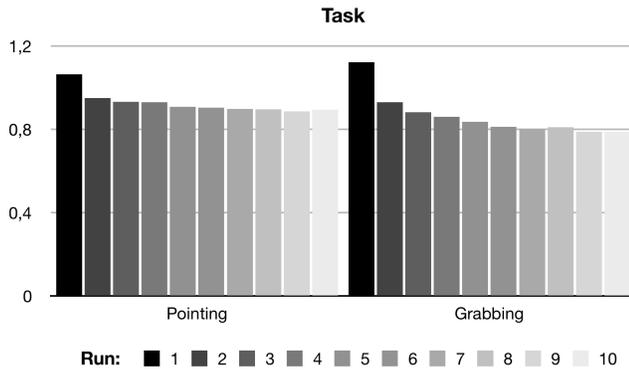


Figure 5: Learning rate for pointing and grabbing tasks.

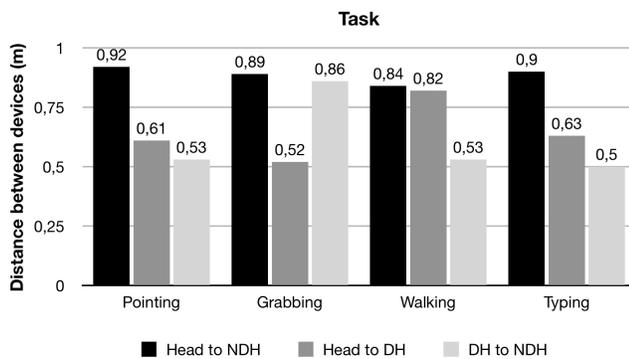


Figure 6: Distances between head, (non) dominant hand.

comparable times. Note, that the other tasks showed similar trends. Figure 5 shows increased task completion times for the first trial. We also plotted the number of targets over time and found an increased time for the first trial as evidence for learning effects.

For body relations, we looked at the average distance that users had between the main devices, i.e. head, dominant, and non-dominant hand positions (Figure 6). It shows how users exhibit different distances depending on the task. Moreover, it highlights user characteristics such as arm length and body height that are also relevant factors for body relations.

Considering target and view based relations, we focus on the two manual tasks to assess the relations from the user to the target (see Figure 7). First, as expected, users look closely at the target to accomplish the task. In the grabbing task, users on average look a bit further away from the target – this is probably a result of (muscle) memory which supports eyes free interaction. As users directly move their hands to the targets, users can remember the spatial locations and aim with less visual control. This is also evident in the target to DH position distance which is substantially lower in the grabbing task, i.e. users are closer to the target with their hand (by design of the task).

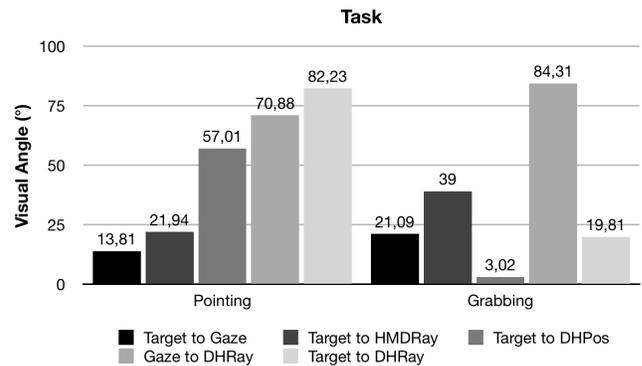


Figure 7: View based features for two tasks.

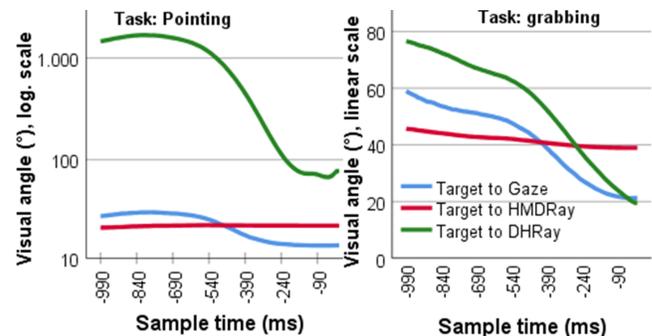


Figure 8: Visual angle over task time.

Notably, the target to DHRay is at odds with the expectation, as we would expect users to point closely to the target. We think this makes sense, as the ray position is averaged over the time of the task and at the beginning of each task users point quite far off the target location. Figure 8 shows the relation of three visual distances over time. We used a logarithmic scale to show the extreme distances that the controller’s ray and the target span initially. It confirms that the ray comes closer when finishing the task (averaging over all tasks including fast users still resulted in a high distance). The figure also shows how the user’s gaze comes closer to the target. This is interesting for the grabbing task where users visually acquire the target and the hand follows.

6 MODELLING AND EVALUATION METHOD

We train Random Forest and Support Vector Machine (SVM) classifiers using the scikit-learn Machine Learning library for Python. We use default values for the hyperparameters (apart from trees=100 and C=1e4).

We filtered out the 3 left-handed users, the first two runs, and first two targets to avoid learning effects. For task typing, we filtered another user as the eye tracker did not work. For pointing and grabbing: 11 targets × 8 runs × 2 conditions = 176 targets; for walking: 8 runs × 2 conditions × 2 conditions = 32 targets; for typing: 13 targets; all per session per user.

Model Features

We generate features by preprocessing the raw time series data from all sensors. We split the data by session, task, repetition, and action (e.g., by targets for the pointing task). In addition, for each device (controllers, HMD), the virtual reality system provides vectors for velocity, angular velocity, and rotation, which we include as features, too.

To assess the categories from the body motion theories introduced in our concept section, we provide the following feature categories as listed in Table 1:

- **Individual:** these represent individual sensor types, e.g., DH (Motion) stands for the controller in the dominant hand. 'Motion' denotes the relative movement of the device, derived by the meter length of the path that the device moved from start to end of the task.
- **Move+Stabilise:** This category combines two Motion types into one feature set. It is used to mutually investigate effects of motion between body segments.
- **Distance:** In this category, the Euclidean distance is used between two positions in the 3D virtual space.
- **Target/Distance:** These feature sets use the Euclidean distance from the device to the target position. For the walking task, the center of the door is the target.
- **Target (visual angle):** The visual angle is computed between the target position and the position of another device; the metric is degrees of visual angle.
- **View (visual angle):** This category includes all visual angle measures between user-related features, such as the angle between the ray span between the forward ray of the HMD and the user's eye gaze direction.
- **All (category):** Includes all features of one category.
- **All:** Includes all features of all categories.

For each of the resulting feature subseries we compute the mean, min, max, and standard deviation of each sensor. For example, we get the four value types of the head motion during the time frame in which participant 1 aimed at the 12 o'clock target in her first repetition of the pointing task in her first session. This means that each line in Table 1 includes four data types that describe the feature or feature set.

These values define the feature vector of one such action. We use these feature vectors as training instances. The label for each vector is given by the corresponding participant ID.

Evaluation

We test several feature sets motivated by interesting comparisons, such as the relative importance of different tracking devices (e.g., controllers, headset, gaze). In addition, we report on optimised feature sets per task. These optimised sets are found with a wrapper-based feature selection approach [24], which uses a greedy search through the space of

feature combinations based on the performance of the actual classifier. We refer the reader to the related work for details. We train the model on the data from the first session and test on the data from the second session. This corresponds to a scenario in which user enrolment happens once on one day and verification happens on other days.

Identification

In the *identification* scenario, we train a Random Forest classifier on the described training instances and labels. The resulting model makes a prediction for a new feature vector based on interactions which participant p_i from the set of all participants P performed. This corresponds to use cases in which the system is supposed to identify a user from a set of known users. We report the accuracy here (F1 scores available in supplementary material).

Identification vs. Group Size

Of interest is also the influence of the size of the user group from which to identify the interacting user. In general, the classification problem becomes harder if there are more people to distinguish, since, e.g., behaviour that might be characteristic among five users might not be unique among ten.

Thus, we evaluate classification performance across group sizes as follows: For each group size g (from 2 to 19), we draw 100 random subsets of g participants from the set of all participants. For each such draw, we then evaluate the classifier on this subset of participants as described before. We report the average performance across the 100 draws.

Metric

We report classification *accuracy* – the number of correct classification decisions divided by the total number of decision (i.e. number of test cases). We report this performance measure since it is commonly used and easy to interpret. Accuracy is adequate in all our evaluation schemes since we always have balanced test sets (i.e. same number of test cases for each user/class). Thus, results can be compared to the baseline of guessing: 1 out of 19 users = 5.26% accuracy.

7 RESULTS

Overall, the best body motion to identify and authenticate users are head motions, and distances between the devices. Table 1 shows the results for each feature and combination.

Identification

Overall we reach accuracies of about 40% to identify users across sessions, for each task. In the following, we report on individual aspects of the results.

Relative distances are most accurate overall: Overall the set of features of relative distances between DH, NDH, and Head showed the highest accuracy to identify users. This shows

	Point	Grab	Walk	Type
Guessing	5.26	5.26	5.26	5.56
Individual				
DH (Motion)	17.11	13.97	13.32	12.82
DH (Velocity)	21.59	18.36	14.97	21.79
DH (AngVel)	21.08	24.22	14.31	18.8
DH (Rotation)	20.07	26.44	22.04	25.64
DH (All)	29.25	<i>30.08</i>	22.37	33.33
NDH (Motion)	8.76	14	17.43	14.96
NDH (Velocity)	12.05	18.96	16.94	18.8
NDH (AngVel)	17.17	15.85	15.62	15.81
NDH (Rotation)	26.88	23.33	12.01	25.21
NDH (All)	22.22	25.57	22.7	31.2
Head (Motion)	12.59	14.14	19.08	13.68
Head (Velocity)	17.34	17.82	22.04	28.21
Head (AngVel)	21.62	21.68	20.39	27.78
Head (Rotation)	32.21	27.81	28.12	34.62
Head (All)	33.67	31.25*	39.47	37.61
All	33.34	27.03	39.31	38.03
Move+Stabilise				
DH, NDH	17.55	18.27	19.24	20.09
DH, Head	23.44	19.8	20.39	24.79
NDH, Head	13.04	15.76	22.53	22.65
DH, NDH, Head	23.03	18	24.84	30.77
All	22.19	18.57	24.84	24.79
Distance				
DH,NDH	18.09	10.53	12.66	18.38
DH,Head	22.79	10.44	25	25.64
NDH, Head	22.22	13.28	20.72	10.26
DH,Head; NDH,Head	29.04	10.94	33.88	35.04
All	41.39*	14.86	43.42*	44.44*
Target/Distance				
Target, DH	10.89	12.05	16.61	8.12
Target, NDH	11.84	13.19	11.51	11.11
Target, Head	7.24	9.15	25	7.69
All	20.96	19.29	32.07	14.53
Target (visual angle)				
Target, HMDRay	6.13	7.72	8.72	8.97
Target, Gaze	10.11	11.27	10.36	9.4
Target, DHRay	9.75	13.1	12.01	19.23
Target, NDHRay	10.11	9.03	15.79	11.54
All	15.85	21.32	15.3	20.09
View				
HMDRay, Gaze	7.33	7.39	8.55	6.84
HMDRay, DHRay	13.04	11.63	14.47	14.53
HMDRay, DHPos	15.67	11.27	13.49	18.38
HMDRay, NDHRay	09.09	13.82	15.79	19.23
HMDRay, NDHPos	16.81	17.34	13.32	22.22
Gaze, DHRay	9.84	8.13	9.21	8.97
Gaze, DHPos	10.18	7.98	8.47	8.33
Gaze, NDHRay	9.51	9.63	9.05	5.98
Gaze, NDHPos	7	7.18	7.57	9.4
All	21.35	30.59	22.53	19.66
All	26.64	27.3	39.64	44.44

Table 1: Classification accuracy results of features and feature sets of body motion categories (* denote best per task; bold and *cursive* denote category 1st and 2nd best, respectively).

that proprioception indicates promising features to identify users which is aligned with the task-specific analysis of body relations shown in Figure 6 (except for grabbing, where the relations consistently received low scores).

Head motion is second accurate overall: As a single set of features from the HMD, head motion consistently scored over 30% accuracy and represents a practical source of motion information if controllers are not available.

Pointing is better than grabbing for hand movement: In grabbing, users have to move to the same absolute positions, thus hand movements resulted in less accurate results. In comparison, pointing tasks allow users to exhibit more freedom in the manual movement.

Head movement is suitable for walking tasks: We find that head movement can better describe individual user walking than hand movements. This confirms prior work that found gait is well modelled using headwear [40].

View relations are similarly accurate as head motion for grabbing tasks: In the grabbing task, all view-based feature sets (30.59%) reached almost the accuracy of all head motion features (31.25%). This can be explained by the design of the task, where the targets were spread across a large area in front of the user, necessitating frequent changes of view.

Best Feature Sets. We used the described feature selection algorithm for finding best combinations among all features per task. To do so, we separated each individual category feature into X/Y/Z. Furthermore, we include relations between DHPos, DHRay, NDHPos and NDHRay. Distance and View feature categories are unchanged.

The resulting feature sets as presented in Table 2 achieved a better accuracy than our systematic tests (cf. Table 1), thus showing that further combinations can improve the system. The largest improvement is for the pointing task from 41.39% to 63.55%. This feature set integrates many body relations of Distance and View, thus indicating that proprioception and view based positions provide strong features. The same is true for Grabbing where less features were found to be relevant – yet distance and visual angle are among them.

For walking, the features are all related to the user’s head motion, aligning with the understanding that head movement is a good identifier of gait. The View-based properties span large visual angles between the HMD forward vector and the user’s hands (usually held waist down at walking). Here we think the maximum viewing angles between head and controllers are more distinctly pronounced than their respective distances or positions, and selected as features.

For typing, an increased accuracy is achieved by combining Head, Distance, and View-based features. This correlates with the individual accuracy results as reported in Table 1. However, the feature set optimisation algorithm found a better combination than our systematic approach.

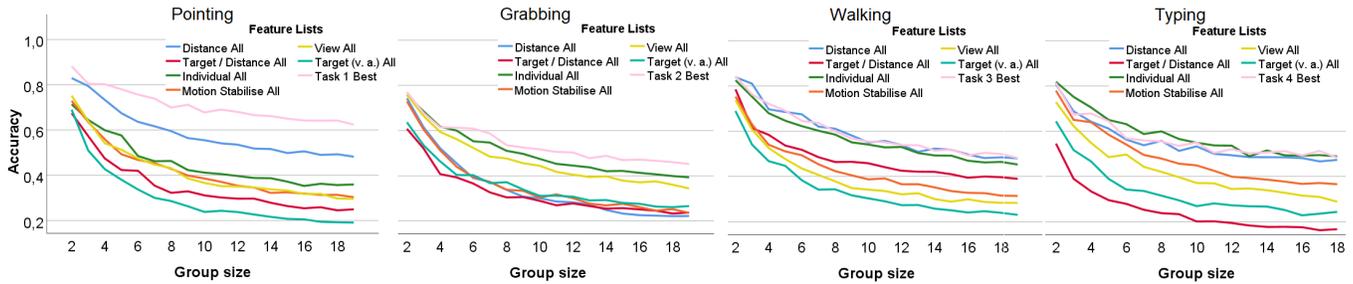


Figure 9: Identification accuracy across user group sizes, for each task.

Pointing (63.55%)

max(DH AngVel),
 avg(DH Rotation),
 max(Distance NDH, HEAD),
 max(Distance DH, NDH),
 avg(Distance DH, HMD),
 min(View HMDRay, DHPos),
 max(View HMDRay, DHPos),
 avg(View HMDRay, DH Pos)

Grabbing (45.84%)

avg(NDH AngVel)
 avg(Dist NDH, HEAD),
 avg(View DHRay, NDHPos)
 avg(View HMDRay, NDHPos)

Walking (49.67%)

min(Head Vel)
 max(Head AngVel)
 max(Distance DH, HEAD)
 max(View HMDRay, NDHPos)
 min(View HMDRay, NDHPos)
 avg(View HMDRay, NDHPos)

Typing (54.27%)

avg(NDH Rotation)
 avg(Head Rotation)
 max(Distance DH Head)
 max(View DHRay, DHPos)
 avg(View Gaze, NDHPos)

Table 2: Feature optimization results.

Identification vs. Group Size. We plotted the identification accuracy across user group results in Figure 9. We used each of the combined feature sets per category ('all' in Table 1), and additionally the best feature group found with the feature selection algorithm. Over all tasks, the diagrams show that with increasing group size, the accuracy decreases at a logarithmic rate. For the pointing task, the best feature set (pink line) decreases at a substantially lower rate and remains above 60%, whereas some of the other feature sets decrease to under 30% of user identification. Similarly, the best feature set also achieves best scores in the grabbing task, although the rate of decrease is similar to other feature sets. In the walking and typing tasks, three feature sets are on a similar top score level: Individual All, Distance All, and the best set found with the feature selection algorithm.

8 DISCUSSION

Our main contribution is the detailed analysis on user identification accuracy of a plethora of body relations, ranging from motion, distances between segments, and how users visually perceive relations between gaze, head direction, and hand positions. Our work allows the relative biometric value

of these body features to be understood, pointing to what feature is more unique among users. This leads to a better understanding of how body motion, proprioception, and internal/external reference points to the user's body are related to each other. Researchers can assess and compare these features depending on which part may be of relevance to their future work. We also want to highlight the main insights:

Body distances are most accurate: The higher accuracy of body relations and distances suggests that proprioception [29, 34] and hand coordination [16] may be stronger biometrics than individual features, as focused on in prior work. Interestingly, one could question whether this can be attributed to absolute biometrics such as arm and body size, but that does not explain why features that consider eye gaze information are so prominent in our feature optimization results (Table 2). It indicates that the biometric value also arises from how users coordinate aspects of the visual and manual abilities, which demands further study.

Pointing is more accurate than grabbing: in pointing, users can exhibit more diverse movement. In grabbing, all users have to move their hand to the same targets, thus creating less individual movement. In conclusion, pointing is more suitable as a biometric.

The dominant hand provides strong features: When pointing, users are free to, e.g., move their hand in direction of the target, or alternatively to make small wrist rotations. These strategies lead to more individual features, and support recent results on dominant hand motion authentication [25].

Our modelling approach reached a maximum of 63% accuracy, which may seem low for practical use. At the same time, our work provides a first fundamental investigation of these biometric features in the VR context. We show that accuracy improves with fewer users (Figure 9): user identification with our used model can be particularly useful for small user groups such as families or companies, for example, to realise some user adaptation of the system.

For more users, we see several ways to improve the modelling. First, we could evaluate further classifiers, features, and variations and combinations thereof. A more substantial

extension could investigate other treatments of the temporal aspect of the data, including, for example, extracting features for rolling windows over time instead of aggregation per action. Moreover, we could investigate the frequency domain of features instead of the time domain. Finally, considering that our identification models vastly outperform chance level by about a factor of ten, we could investigate potential improvements via aggregation of evidence over time (for example, decision after observing multiple actions). We thus see interesting opportunities for investigating these optimisations in future work.

Beyond accuracy, we contribute to the understanding of human motion for user identification. Our concept section and Figure 1 provide a systematic overview of how previous work has considered motion and derives relations between body segments. The list of results in Table 1 promotes discussion on the different body motions and how suitable they are as predictors for individual human motion. Moreover it characterizes 4 fundamental interaction tasks with regards to which motion features are valuable to identify users, showing overall good predictors across tasks (for example, distance or head), as well as which specific combinations scored highest (Table 1). Depending on available sensors and tasks, a developer can use our results to make an informed decision.

Our results are based on a user study, which was rigorously designed but should be considered carefully with the following limitations. First, the tasks were designed to represent typical VR interactions. This may naturally involve bias factors. In walking to a target, users may adjust stride length. Furthermore, users may get tired from repetitive manual tasks, such as interacting with a design application where users are continuously drawing or a game where users point to shoot. These are factors common in everyday interactions and we intentionally did not control these factors but call for future studies to better understand potential bias factors.

The results may also be affected by our choice of users. Authentication systems benefit from testing with more participants than the 19 we had. This was sufficient to show that the accuracy stagnated with an increasing number of users (Figure 9), but testing with more users is relevant to really assess the potential for an authentication system. Additionally, features that involve body relations may be affected by physical attributes such as the user height and arm length. However, that does not explain why features that consider visual information are so prominent in our feature optimization results (Table 2). It indicates that the biometric value also arises from how users coordinate aspects of the visual and manual abilities, which demands further study. Finally, our user set does not balance gender well. We expect that more diversity might increase user identification accuracy.

9 CONCLUSION

Our work provides an exploration of a broad range of body relations, that points to the possibility for continuously identifying and authenticating users in current VR systems and, hence, overcoming the limitations of explicit authentication mechanisms [15]. Features such as head motion and distances between head and the controllers, which have shown higher recognition accuracy than other features, can be integrated right now as a soft biometric identifier to enhance the default authentication method. In the short-term future, further explorations into motion relations beyond hands and head may reveal better features. Also more advanced machine learning methods, such as deep learning, may make the system more accurate. In the long run, motion tracking methods may be able to capture whole body motion as well as the space around the user in realtime and exploit this for user identification and general user behavioural analysis.

10 ACKNOWLEDGMENTS

This research was supported by the German Research Foundation (DFG), Grant No. AL 1899/2-1.

REFERENCES

- [1] Virginia O. Andersson and Ricardo M. Araujo. 2015. Person Identification Using Anthropometric and Gait Data from Kinect Sensor. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15)*. AAAI Press, 425–431. <http://dl.acm.org/citation.cfm?id=2887007.2887067>
- [2] Corey Ashby, Amit Bhatia, Francesco Tenore, and Jacob Vogelstein. 2011. Low-cost electroencephalogram (EEG) based authentication. In *Neural Engineering (NER), 2011 5th International IEEE/EMBS Conference on*. IEEE, 442–445.
- [3] Kyle O Bailey, James S Okolica, and Gilbert L Peterson. 2014. User identification and authentication using multi-modal behavioral biometrics. *Computers & Security* 43 (2014), 77–89.
- [4] VE Belenkii, VS Gurfinkel, and EI Paltsev. 1967. On the control elements of voluntary movements. *Biofizika* (1967).
- [5] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. 2002. View-invariant estimation of height and stride for gait recognition. In *International Workshop on Biometric Authentication*. Springer, 155–167.
- [6] Mark Billingham, Jerry Bowskill, Nick Dyer, and Jason Morphet. 1998. Spatial information displays on a wearable computer. *IEEE Computer Graphics and Applications* 18, 6 (1998), 24–31.
- [7] Cheng Bo, Lan Zhang, Xiang-Yang Li, Qiuyuan Huang, and Yu Wang. 2013. SilentSense: Silent User Identification via Touch and Movement Behavioral Biometrics. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking (MobiCom '13)*. ACM, New York, NY, USA, 187–190. <https://doi.org/10.1145/2500423.2504572>
- [8] Edward G. Britton, James S. Lipscomb, and Michael E. Pique. 1978. Making Nested Rotations Convenient for the User. In *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '78)*. ACM, New York, NY, USA, 222–227. <https://doi.org/10.1145/800248.807394>
- [9] Grigore C Burdea and Philippe Coiffet. 2003. *Virtual reality technology*. John Wiley & Sons.

- [10] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1393–1402. <https://doi.org/10.1145/2702123.2702252>
- [11] Antitza Dantcheva, Carmelo Velardo, Angela D'Angelo, and Jean-Luc Dugelay. 2011. Bag of Soft Biometrics for Person Identification. *Multimedia Tools Appl.* 51, 2 (Jan. 2011), 739–777. <https://doi.org/10.1007/s11042-010-0635-7>
- [12] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. 2012. Touch Me Once and I Know It's You!: Implicit Authentication Based on Touch Screen Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 987–996. <https://doi.org/10.1145/2207676.2208544>
- [13] Simon Eberz, Kasper B. Rasmussen, Vincent Lenders, and Ivan Martinovic. 2017. Evaluating Behavioral Biometrics for Continuous Authentication: Challenges and Metrics. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS '17)*. ACM, New York, NY, USA, 386–399. <https://doi.org/10.1145/3052973.3053032>
- [14] Shaun Gallagher and Daniel Schmicking. 2010. *Handbook of phenomenology and cognitive science*. Springer.
- [15] Ceenu George, Mohamed Khamis, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *Proceedings of the Usable Security Mini Conference 2017 (USEC'17)*. Internet Society, San Diego, CA, USA.
- [16] Yves Guiard. 1987. Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. In *Journal of Motor Behavior*, Vol. 19. 486–517.
- [17] Christian Holz, Senaka Buthpitiya, and Marius Knaust. 2015. Bodyprint: Biometric User Identification on Mobile Devices Using the Capacitive Touchscreen to Scan Body Parts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3011–3014. <https://doi.org/10.1145/2702123.2702518>
- [18] ISO ISO. [n. d.]. 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs)-Part 9: Requirements for non-keyboard input devices, 2000. *International Organization for Standardization* ([n. d.]).
- [19] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based Interaction: A Framework for post-WIMP Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 201–210. <https://doi.org/10.1145/1357054.1357089>
- [20] Anil K Jain, Sarat C Dass, and Karthik Nandakumar. 2004. Soft biometric traits for personal recognition systems. In *Biometric authentication*. Springer, 731–738.
- [21] Erno Jeges, Istvan Kispal, and Zoltan Hornak. 2008. Measuring human height using calibrated cameras. In *Human System Interactions, 2008 Conference on*. IEEE, 755–760.
- [22] Roland S Johansson, Göran Westling, Anders Bäckström, and J Randall Flanagan. 2001. Eye–hand coordination in object manipulation. *Journal of Neuroscience* 21, 17 (2001), 6917–6932.
- [23] Georgios Kambourakis, Dimitrios Damopoulos, Dimitrios Papamartzivanos, and Emmanouil Pavlidakis. 2016. Introducing touchstroke: keystroke-based authentication system for smartphones. *Security and Communication Networks* 9, 6 (2016), 542–554.
- [24] George H. Kohavi, Ron and John. 1997. Wrappers for feature subset selection. *Artificial Intelligence* 97, 1 (1997), 273–324.
- [25] Alexander Kupin, Benjamin Moeller, Yijun Jiang, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Task-Driven Biometric Authentication of Users in Virtual Reality (VR) Environments. In *Multi-Media Modeling - 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11, 2019, Proceedings, Part I*. 55–67. https://doi.org/10.1007/978-3-030-05710-7_5
- [26] Nurakhmed Nurislamovich Latypov and Nurulla Nurislamovich Latypov. 1999. Method for tracking and displaying user's spatial position and orientation, a method for representing virtual reality for a user, and systems of embodiment of such methods. US Patent 6,005,548.
- [27] Sugang Li, Ashwin Ashok, Yanyong Zhang, Chenren Xu, Janne Lindqvist, and Macro Gruteser. 2016. Whose move is it anyway? Authenticating smart wearable devices using unique head movement patterns. In *Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on*. IEEE, 1–9.
- [28] Hanneke Liesker, Eli Brenner, and Jeroen BJ Smeets. 2009. Combining eye and hand in search is suboptimal. *Experimental brain research* 197, 4 (2009), 395–401.
- [29] Feng Lin, Kun Woo Cho, Chen Song, Wenyao Xu, and Zhanpeng Jin. 2018. Brain Password: A Secure and Truly Cancelable Brain Biometrics for Smart Headwear. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 296–309.
- [30] Pedro Lopes, Alexandra Ion, Willi Müller, Daniel Hoffmann, Patrik Jonell, and Patrick Baudisch. 2015. Proprioceptive Interaction. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 175–175. <https://doi.org/10.1145/2702613.2732490>
- [31] I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase Sets for Evaluating Text Entry Techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 754–755. <https://doi.org/10.1145/765891.765971>
- [32] Jean Massion. 1992. Movement, posture and equilibrium: interaction and coordination. *Progress in neurobiology* 38, 1 (1992), 35–56.
- [33] B Miller. 1988. Everything you need to know about biometric identification. *Personal Identification News 1988 Biometric Industry Directory* (1988).
- [34] Mark R Mine, Frederick P Brooks Jr, and Carlo H Sequin. 1997. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 19–26.
- [35] Brent C. Munsell, Andrew Temlyakov, Chengzheng Qu, and Song Wang. 2012. Person Identification Using Full-body Motion and Anthropometric Biometrics from Kinect Videos. In *Proceedings of the 12th International Conference on Computer Vision - Volume Part III (ECCV'12)*. Springer-Verlag, Berlin, Heidelberg, 91–100. https://doi.org/10.1007/978-3-642-33885-4_10
- [36] Tahrira Mustafa, Richard Matovu, Abdul Serwadda, and Nicholas Muirhead. 2018. Unsure How to Authenticate on Your VR Headset?: Come on, Use Your Head!. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA '18)*. ACM, New York, NY, USA, 23–30. <https://doi.org/10.1145/3180445.3180450>
- [37] Vitruvius Pollio and M. H. Morgan. 1960. *De architectura*. New York: Dover Publications.
- [38] Daniel A Reid and Mark S Nixon. 2011. Using comparative human descriptions for soft biometrics. In *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 1–6.
- [39] Ioannis Rigas, Oleg Komogortsev, and Reza Shadmehr. 2016. Biometric recognition via eye movements: Saccadic vigor and acceleration cues. *ACM Transactions on Applied Perception (TAP)* 13, 2 (2016), 6.

- [40] Yiran Shen, Hongkai Wen, Chengwen Luo, Weitao Xu, Tao Zhang, Wen Hu, and Daniela Rus. 2018. GaitLock: Protect virtual and augmented reality headsets using gait. *IEEE Transactions on Dependable and Secure Computing* (2018).
- [41] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 647, 13 pages. <https://doi.org/10.1145/3173574.3174221>
- [42] Robert John Teather. 2013. Evaluating 3D pointing techniques. (2013).
- [43] Robert J Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*. IEEE, 87–94.
- [44] Colin Ware and Harutune H. Mikaelian. 1987. An Evaluation of an Eye Tracker As a Device for Computer Input. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface (CHI '87)*. ACM, New York, NY, USA, 183–188. <https://doi.org/10.1145/29933.275627>
- [45] J Wayman. 2000. A definition of biometrics. *National Biometric Test Center Collected Works* 1, 2 (2000), 21–23.
- [46] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. Glassgesture: Exploring head gesture interface of smart glasses. In *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 1–9.
- [47] Yongtuo Zhang, Wen Hu, Weitao Xu, Chun Tung Chou, and Jiankun Hu. 2018. Continuous Authentication Using Eye Movement Response of Implicit Visual Stimuli. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 177 (Jan. 2018), 22 pages. <https://doi.org/10.1145/3161410>
- [48] Zhaoxiang Zhang, Maodi Hu, and Yunhong Wang. 2011. A survey of advances in biometric gait recognition. In *Chinese Conference on Biometric Recognition*. Springer, 150–158.
- [49] Zhaoxiang Zhang, Kaiyue Wang, and Yunhong Wang. 2011. A survey of on-line signature verification. In *Chinese Conference on Biometric Recognition*. Springer, 141–149.