UNIVERSITY OF CALIFORNIA,
IRVINE

GaitGuard: Towards Private Gait in Mixed Reality

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Electrial and Computer Engineering

by

Diana Gimena Romero

Thesis Committee:
Assistant Professor Salma Elmalaki, Chair
Professor Athina Markopoulou
Associate Professor Yasser Shoukry

2024

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I am profoundly grateful to my committee chair, Professor Salma Elmalaki, whose unwavering guidance, mentorship, and encouragement have been instrumental throughout the course of this research endeavor. Your expertise and insights have not only enriched the quality of this thesis but have also significantly contributed to my growth as a researcher. I am profoundly grateful for your support and dedication.

I extend my heartfelt appreciation to Professor Athina Markopoulou for generously dedicating her time and expertise to shaping this work. Your invaluable contributions, constructive feedback, and insightful suggestions have played a pivotal role in refining the content and methodology of this thesis. Your mentorship has been invaluable to me, and I am sincerely grateful for your guidance.

I would like to extend my sincere appreciation to Professor Yasser Shoukry, a valued member of my committee, for their participation and support throughout this process. Your presence has been reassuring, and I am grateful for your guidance and oversight.

I would also like to express my gratitude to the members of the UCI Networking Group for their invaluable insights, assistance, and camaraderie throughout this research journey. Your collective wisdom and collaborative spirit have enriched my understanding and inspired new perspectives.

Furthermore, I wish to acknowledge the partial support provided by the following funding sources: NSF award # CNS2105084, NSF award 1956393, and a gift from the Noyce Initiative. Your financial assistance has been instrumental in realizing the objectives of this thesis.

Lastly, I would like to express my gratitude to my family and friends for their unwavering encouragement, understanding, and patience throughout this academic journey. Your love and support have been my greatest source of strength and motivation.

To all those who have contributed to this endeavor in various capacities, your support has been invaluable, and I am sincerely thankful for your presence in this journey.

ABSTRACT OF THE THESIS

GaitGuard: Towards Private Gait in Mixed Reality

By

Diana Gimena Romero

Master of Science in Electrial and Computer Engineering

University of California, Irvine, 2024

Assistant Professor Salma Elmalaki, Chair

Augmented/Mixed Reality (AR/MR) devices are unique from other mobile systems because of their capability to offer an immersive multi-user collaborative experience. While previous studies have explored privacy and security aspects of multiple user interactions in AR/MR, a less-explored area is the vulnerability of gait privacy. Gait is considered a private state because it is highly individualistic and a distinctive biometric trait. Thus, preserving gait privacy in emerging AR/MR systems is crucial to safeguard individuals from potential identity tracking and unauthorized profiling. This paper first adopts and automates a framework designed to detect gait information in humans, referred to in this work as **GaitExtract**. **GaitExtract** can automatically detect the neighbor gait information of a human and investigate the vulnerability of gait privacy in AR. In a user study with 20 participants, our findings reveal that participants were uniquely identifiable with an accuracy of up to 78% using **GaitExtract**. Consequently, we propose **GaitGuard**, a real-time system that safeguards the gait information of people appearing in the camera view of the AR/MR device (a.k.a. bystanders). We tested **GaitGuard** in an MR collaborative application, achieving 22 fps while streaming mitigated frames to the collaborative server. Furthermore, our qualitative surveys indicated that users are more comfortable with releasing videos of them walking when **GaitGuard** is applied to the camera frames. These results underscore the efficacy and practicality of **GaitGuard** in mitigating gait privacy concerns in MR contexts.

Chapter 1

Introduction

Mixed reality (MR) devices have been becoming a lot more mainstream with the recent announcements of new mixed reality headsets such as Apple Vision Pro [8], Meta Quest 3 [15], Meta Aria glasses [47], in addition to the already established Hololens 2 [13]. Additionally, the mixed reality worldwide market was already valued to be a 1.4 billion industry in the year 2023 and is expected to double by 2030 [45]. These projections and the announcement of these new devices signal that our community is moving towards the broad adoption of MR technology[1].

MR technologies blend the physical and the digital world by placing virtual objects that humans can interact with in the physical environment [23]. To facilitate the seamless interaction between the computer, the human, and the physical environment, MR technologies are equipped with a myriad of sensors that enable environmental perception capabilities and human interaction through hand-tracking, eye-tracking, and speech input [23]. Despite the MR's potential across different sectors and as the technology advances, it has been becoming

[1]Mixed Reality (MR) is a spectrum that blends both physical and digital worlds and within this spectrum is a small subset of augmented and virtual reality experiences [23]. Since the MR spectrum encompasses an extensive array of applications, in this paper, we use the term MR to refer to technologies that enable physical interaction with virtual objects.

an increasing concern if these ubiquitous sensors are being used maliciously by adversaries to identify sensitive information such as facial information [32], semantic location [38], user behavior [41, 65].

Privacy concerns in multi-user Mixed Reality (MR) applications have surfaced due to MR's distinctive feature of enabling immersive collaborative interactions among users in two main settings: one where users share the same physical space and another where users are in different physical locations. This particular aspect of MR technology exposes a wide range of scenarios to the continuous monitoring by MR device sensors, raising issues regarding the security and privacy of multi-user collaborations within MR environments. To tackle these concerns, a variety of studies have been conducted to investigate secure and private ways to facilitate multi-user interactions in mixed reality [44, 54, 55].

Despite the growing awareness of privacy issues within Mixed Reality (MR) environments, the specific concern of gait privacy has not received adequate attention. Gait, or the distinctive way individuals walk, is recognized as a biometric identifier that can be used for recognition purposes. Studies have revealed that gait data can be associated with sensitive information, including ethnicity [64], age [66], gender [62], and neuromusculoskeletal disorders [58]. Moreover, regulations such as the California Privacy Rights Act (CPRA)[2] and the European Union's General Data Protection Regulation (GDPR) [11] mandate the safeguarding of gait and other biometric data that could identify a person, highlighting the necessity for mechanisms to secure gait data in emerging technological applications.

Safeguarding gait information becomes particularly critical in MR scenarios, where participants often use headsets outfitted with multiple sensors capable of unintentionally capturing and analyzing their gait. This potential for gait privacy breaches in MR not only poses ethical concerns regarding consent and data protection but also raises awareness that users' unique walking patterns may be recorded and scrutinized without their informed consent. Therefore, regulating the acquisition, storage, and use of gait data within MR contexts is

crucial to ensure users maintain authority over their personal information and are shielded from privacy violations. As MR technology evolves and its usage expands, the industry must develop explicit protocols and standards to confront gait privacy issues, thereby preserving privacy rights and fostering ethical data practices.

This paper tackles the issue of gait privacy in Mixed Reality (MR) technology by employing a comprehensive methodology. We begin by assessing the potential for identifying users based on gait characteristics derived from MR device camera feeds. Subsequently, we introduce **GaitGuard**, a system tailored for deployment in MR collaborative applications to mitigate the risk of gait information leakage. Furthermore, to gain a deeper understanding of user attitudes toward gait privacy, we have executed a survey that captures the qualitative perspectives of individuals on this matter.

Chapter 2

Related Work and Background

2.1 Privacy Problems in Mixed Reality

As MR headsets are projected to become more ubiquitous, there is a growing concern about possible security and privacy risks with these devices. One study explored the effects of perceptual manipulation attacks (PMA) in MR on users [30]. Moreover, bystander privacy becomes a significant issue because these devices often utilize cameras, sensors, and other data collection tools to capture information about the environment around the user. This includes recording images, videos, and sometimes audio of people who happen to be in the vicinity but are not directly engaging with the technology themselves [32]. Considering these MR headsets have the unique characteristic of employing an immersive multi-user collaborative interaction, there is a growing concern about ensuring secure and private interactions in **collaborative** environments. In response to this concern, SecSpace, a framework developed to ensure that privacy and security mechanisms are implemented in collaborative MR has been proposed [54].

While numerous studies have explored various privacy concerns associated with MR, the

issue of protecting gait privacy within this domain remains largely unexplored. This paper seeks to fill that gap in the existing body of research.

2.2 Gait: a Unique Personal Signature

Gait recognized as an inherently unique attribute for individuals has been extensively studied as a means of authentication, both through wearable devices like accelerometer sensors [40, 35, 31] and mobile systems such as phones and smartwatches [34, 27, 48]. These studies underscore gait's high uniqueness and identifiability as a distinguishing feature.

Furthermore, gait information extends beyond mere identification and has been linked to various sensitive attributes, including ethnicity [64], age [66], gender [62], and neuromuskuloskeletal disorders [58]. The implications of gait analysis span a broad spectrum of personal characteristics and health-related information. Despite the well-established uniqueness of gait and its association with sensitive attributes, there is a notable gap in existing research.

In recent years, MR applications have become more pervasive. One of the biggest aids to this proliferation is the rise of many commercial head-mounted displays (i.e., Hololens 2, Meta Quest Pro, etc.) in this domain. These HMDs are equipped with many sensors that may be able to provide insight into a person's gait. Gait has been extensively studied to identify a person and has been established by numerous jurisdictions through the law that this is private information that must be protected [2].

To the best of our knowledge, no study currently systematically investigates the prevalence of gait information leakage in emerging technologies, particularly in the realm of MR.

2.3 Gait Types in Mixed Reality

The emergence of MR applications then poses new privacy concerns as attackers could access this sensor information and identify a person's gait. Two types of gait information could be collected based on perspective; for simplicity, we refer to these two types as personal gait and neighbor gait.

2.4 Personal Gait

The first type of gait information refers to the device user's gait information. Many of these HMDs are equipped with IMU sensors, and IMU-based gait recognition has long been explored in the literature [33]. While IMU-based gait recognition has already been explored before, prior work mainly focused on placing the IMU sensor on the ankle to obtain the user's gait pattern, and it is unclear whether IMU-based head movement recognition can also be linked to all the sensitive information the gait is related to. However, it has been shown that head movements do have some correlation when identifying people through gait, as it was found that identifying people through head movement alone only resulted in about 60% accuracy [42]. This suggests that gathering the gait information of an MR headset user is feasible but there are no studies that have shown that sensitive information associated with gait can be linked to head movements.

2.4.1 Neighbor Gait

The other type of gait information, referred to in this paper as neighbor gait, is concerned with collecting private gait features of people within the same physical space as the device user. In particular, people appearing in the headset camera view whether as a bystander or

as another user in a collaborative app setup. In this paper, we focus on investigating the framework and tools needed to gather and protect neighbor gait.

2.4.2 Pose Estimation for Gait Detection

Instrumented gait analysis (IGA) refers to the precise and accurate analysis of gait patterns and characteristics, and systems that employ IGA often make use of motion capture systems, force plate, instrumented walkways, and treadmills [29]. While this is the golden standard for gait assessment in research practice, traditional systems for gait analysis are costly and invasive. Because of this, alternative techniques for cost-effective gait analysis have been extensively investigated. Some of the proposed gait detection systems make use of a gyroscope [51], inertial measurement units (IMU) [53], and Kinect sensor [39, 52].

Another low-cost and accurate alternative in detecting gait information is video-based pose estimation [59, 46]. Video-based methods are particularly accessible, requiring only a camera to analyze gait. Pose estimation algorithms detect a person's position and orientation by predicting the location of different keypoints such as hands, heads, legs, etc. [20]. An example of a pose estimation is Openpose library, which enables real-time pose estimation of multiple people from a two-dimensional video [28]. This pose estimation library was used by the video-based gait analysis framework proposed by Stenum et al. [59] demonstrating that video-based gait analysis can yield results closely matching those from 3D motion capture systems. This paper leverages video-based gait analysis for its accessibility and effectiveness [59]. In particular, we explore the potential of video-based gait detection to address privacy concerns related to "neighbor gait" by analyzing scenes captured via headsets' cameras.

2.5 Paper Contributions

This paper makes the following contributions:

1. **GaitExtract Framework:** We introduce **GaitExtract**, a framework designed to automatically detect "neighbor gait" defined as the gait information of individuals within the camera view on MR headsets.

2. **Identification Attack:** By leveraging extracted gait information from camera frames, we demonstrate the ability to uniquely identify different individuals, highlighting potential privacy vulnerabilities.

3. **GaitGuard System:** We propose **GaitGuard**, a systematic approach designed to mitigate gait privacy leaks, particularly tailored for MR collaborative applications, ensuring minimal impact on application utility.

4. **Application Design and User Study**: To validate **GaitGuard**'s effectiveness, we design an MR collaborative application, integrating **GaitGuard**, and we conduct a user study involving 20 participants.

5. **User Survey:** To gain deeper insights into user perceptions, we conduct a user survey exploring how individuals using these technologies perceive gait privacy.

Chapter 3

Threat Model

This paper investigates the threat of gait privacy leak in MR applications, mainly through video-based gait detection. The focus is on devices like the HoloLens, equipped with two types of cameras, facilitating communication with other devices such as MR headsets, phones, tablets, and laptops[1]. The multifaceted flow of information, specifically camera frame feeds, provides attackers with potential vantage points at different levels, enabling them to extract gait information.

The analysis of vulnerabilities encompasses various scenarios and applications susceptible to gait information leakage, emphasizing that any point allowing access to camera frames is a potential vulnerability. We explored 318 HoloLens applications from the Microsoft Store [16] and found that 26% request camera access, underscoring the widespread exposure of users to potential privacy breaches. Based on these applications, we define three primary MR application modes explored as potential threat scenarios. In particular, we explored the single-user experience and the multi-user collaborative experience setup provided by MR. We explore the threat models in the following application setups: (1) same physical space

[1]While the rest of the paper focuses on Hololens as an example of HMD, the same analysis/results can be extended to other HMD in the market that need to share scene information across users collaborating in a multi-user MR environment.

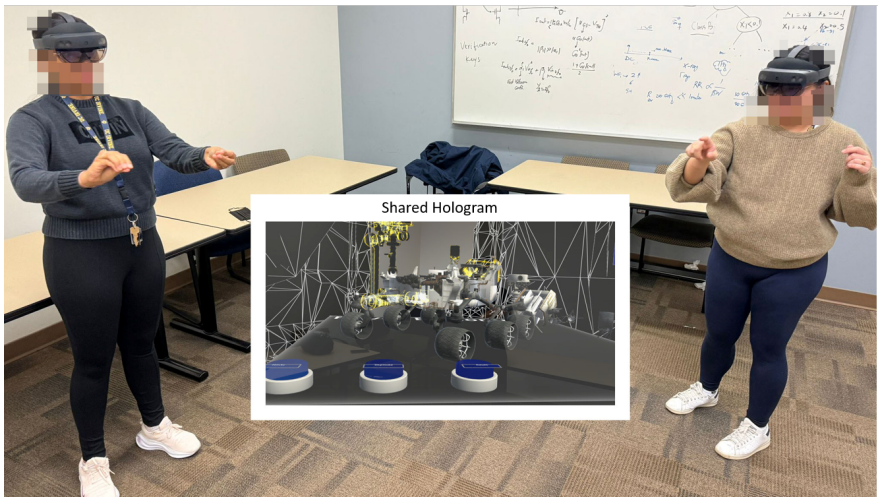


Figure 3.1: An example of a collaborative application in Mixed Reality where two users exist in the same physical space and collaboratively work on a unified virtual object.

collaboration (**PSC**), (2) remote space collaboration (**RSC**), and (3) single-user experience with no collaboration (**SUE**).

3.1 Collaborative applications in MR

In MR, collaborative applications denote extended reality experiences involving multiple users in a shared holographic environment. These applications harness the spatial computing capabilities inherent in MR devices (HoloLens 2). Collaborative MR apps allow users to see and interact within the same digital content from their respective perspectives [5].

Establishing collaborative experiences in Mixed Reality (MR), especially within the HoloLens 2, necessitates the integration of a robust network manager to facilitate and oversee connections among multiple users. This enables collaborative interactions and ensures synchronized sharing of object movements across users within the MR space [5]. Microsoft's HoloLens 2 incorporates a common networking solution for building multi-user applications known as the Photon Unity Networking (PUN) framework [19]. PUN seamlessly aligns with Unity development, a versatile cross-platform engine supporting MR/AR/VR development. The

integration between Unity and PUN provides comprehensive tools to manage user registrations, handle network synchronization, and enable a cohesive collaborative environment for shared MR experiences on the HoloLens 2 platform.

An example of a collaborative experience using MR HoloLens 2 is shown in Figure 3.1.

3.2 Same Physical Space Collaboration (PSC)

This collaboration mode occurs when multiple MR headset users share the same physical space and collaborate on a unified virtual object, as shown in Figure 3.1. This scenario presents in applications, such as HoloOne Sphere [22] and Catapult [10]. These applications may also incorporate Augmented Reality (AR) features, such as QR codes, camera filters, or computer vision capabilities, potentially requesting access to the device's camera within the collaborative space. A depiction of the **PSC** threat model is shown in Figure 3.2 with two potential points of attacks as explained below.

Threat Model

- Collaboration occurs when multiple MR headset users share the same physical space.
- Applications such as HoloOne Sphere [22] and Catapult [10] with collaboration on a unified virtual model request access to the device's camera.
- The camera frames are shared from multiple MR devices to the common network manager server as explained in Section 3.1.

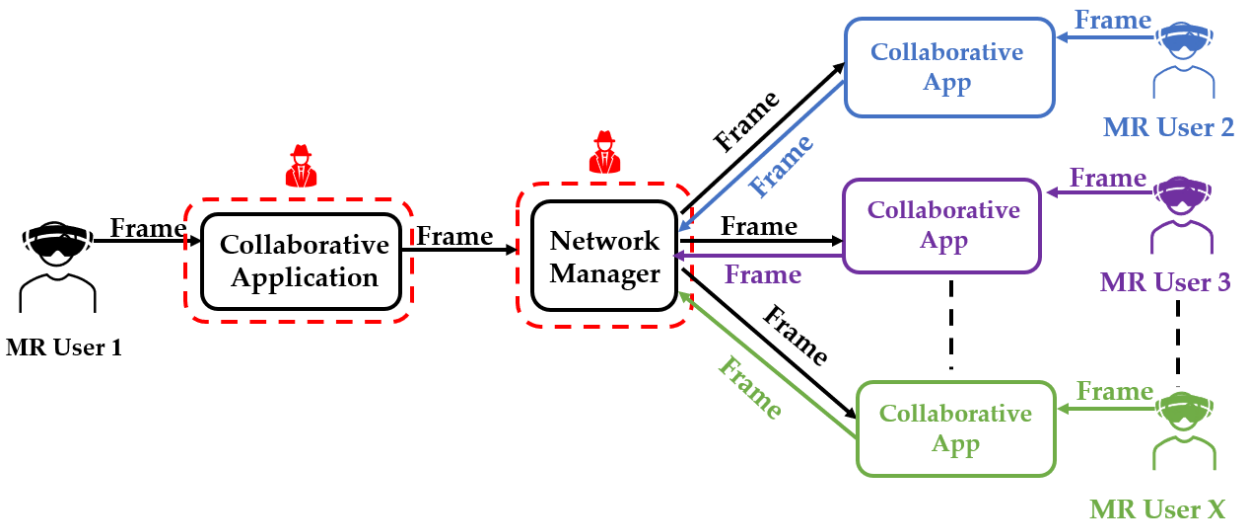


Figure 3.2: Threat model in PSC. Multiple users in the same physical space.

3.3 Remote Space Collaboration (RSC)

The remote space collaboration **(RSC)** collaboration mode has a key distinction from **PSC** where not all MR headset users are physically present in the same physical location, enabling remote participation as illustrated in Figure 3.3. Users engaging remotely are not restricted to MR devices; they could employ any device, such as a phone, laptop, or tablet. Applications like Lens Bouvet [9] and Teams [18] facilitate remote collaboration, allowing users from any device with the application to annotate an MR headset's view.

Similar to **PSC**, this scenario introduces vulnerabilities at the application and server levels, as depicted in Figure 3.3. However, **RSC** adds an extra point of vulnerability at any device level. Any device participating in the collaborative session becomes a potential attacker regardless of physical location. This is because any device in the session can access the camera frames of all the participants.

Threat Model

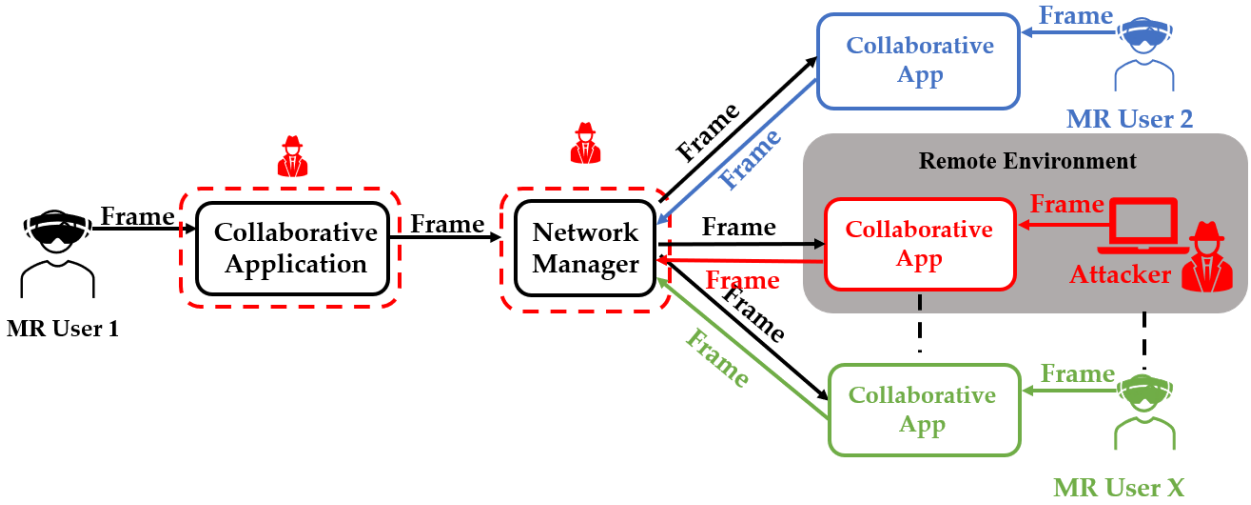


Figure 3.3: Threat model in RSC. Some users join remotely using different device modalities.

- Not all users are physically present in the same location.
- Users can engage remotely using various device modalities like phones, laptops, or tablets.
- Applications such as Lens Bouvet [9] and Teams [18] request camera frames to allow **RSC** and annotation of an MR headset's view.

3.4 Single User Experience (SUE)

The final mode of operation involves applications that, while not designed for collaboration, have access to the camera for Augmented Reality (AR) capabilities. A brief survey of applications in the Hololens 2 category reveals that approximately 26% of these applications possess camera access.

This introduces a critical concern regarding bystander privacy, where individuals unintentionally sharing the same physical space as the MR application user become susceptible to privacy breaches. Although the issue of bystander privacy has been extensively examined for decades in the context of mobile devices, it has recently gained prominence in augmented reality headsets, with studies like BystandAR exploring solutions specifically for facial pri-

vacy [32].

In this mode of operation, the potential attacker is constrained to the application level. Unlike collaboration scenarios, camera frames are not communicated to other users' devices, limiting the points of vulnerability. This setup underscores the unique challenge of bystander privacy in gait information leakage.

Threat Model

- A single-user space in an MR application wearing a headset.
- Application gains access to the camera frames to enable virtual overlays.
- A bystander enters the camera view.

3.5 Common Threat Across Modes

The common threat across both Same Physical Space Collaboration (PSC) and Remote Space Collaboration (RSC), as well as the Single User Experience (SUE), centers on the exploitation of camera frames. This vulnerability stems from the shared need across these modes to access and transmit camera frames to enhance collaborative experiences or augment reality, making them susceptible to various attack vectors:

- **Application Level Attacks:** At this level, attackers exploit vulnerabilities within the application itself to gain unauthorized access to camera frames. This threat is prevalent in all modes of operation, where the applications' need to access camera data for functionality exposes them to potential exploitation.
- **Server Level Attacks:** In scenarios where camera frames are shared across a network, such as in PSC and RSC modes, attackers can target the network manager server. Attack-

ers can intercept and access the shared camera frames by installing spyware or exploiting vulnerabilities on the server, compromising the privacy of all participants in the collaborative space.

- **Device Level Vulnerabilities:** Particularly relevant in the RSC mode, any device used to participate in the collaboration becomes a potential point of attack. This includes MR headsets, phones, laptops, or tablets used to join the session. Attackers can use these devices to access the camera views of all participants.

Across these collaboration modes, the shared reliance on camera frames for enhancing collaboration or providing AR features introduces a critical point of vulnerability. The unauthorized access to camera frames threatens the participants' privacy. Hence, this paper focuses on this aspect of a privacy leak of gait information from camera frames.

Chapter 4

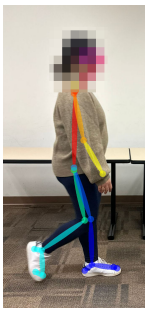
Gait Extraction Framework

Motivated by the prevalence of threat to gait privacy in MR, in this section, we discuss the systematic design and implementation of the framework used to gather the neighbor gait information, which we refer to as **GaitExtract**. As mentioned earlier in Section 2.4.2, our design exploits the method developed by Stenum et al. where they proposed to use 2D-based techniques for gait feature analysis [59]. Their proposed algorithm generates the different gait features based on the location of the keypoints provided by OpenPose library [59].

4.1 User Study and Data Collection

We conducted an IRB-approved user study (IRB #2848) where 20 participants were recruited by advertising on mailing lists and asking personal contacts to forward our study information to interested participants. Participants were all 18 or older, and before data collection, each user was given an overview of the study's goal, which is to investigate gait privacy leaks in mixed reality. They were informed that their gait information would be collected for this study. They were also informed that all identifying information was removed

(a)



(b)

Figure 4.1: Data Collection Setup; (a) Recording of walking sequence; (b) Application of OpenPose to walking sequence.

and that the collected data would only be accessed by researchers of this study. Each participant provided informed consent to participate in the study and video record their walking sequences. Participants were given face masks as an extra precaution to prevent recording their facial information during the data collection.

Data collection consisted of participants walking for about 5 minutes between two markers in a private room while a research team member wore the HoloLens 2 to record their walking sequences. Figure 4.1 illustrates the data collection setup. Because of the two markers' position, the distance of the user's walking sequences was between $2.5m$ and $2.75m$. We intentionally used a short distance because it is not expected that MR collaborative users walk for long distances in the scene. Moreover, we want to explore whether gait features can still uniquely identify an MR user even within a short distance.

Participants were then asked to complete a short survey to gather their user perspectives about gait privacy in MR.

4.2 Gait Feature Extraction

The method developed by Stenum, et. al. [59] uses the keypoints generated by the OpenPose library as a basis for the gait feature extractions. In particular, they use 3 of the 25 body keypoints generated by OpenPose, the midhip, left ankle, and right ankle. The tool they developed was designed to be applied to one video of a walking sequence, where a walking sequence refers to an instance or video of a person that walks exclusively in one direction. This tool requires several manual adjustments to enable gait feature extraction. First, to use the tool, we need to manually identify the direction from which the subject walks (i.e., left to right or right to left). Afterward, the tool requires manually correcting all the wrong left-right leg identification. Once this leg detection is corrected, the tool will be able to detect the heel-strike and toe-off of each leg. Heel-strike and Toe-off are two events that make up a gait cycle [50].

Following the leg detection correction, we should manually pinpoint the two markers we choose in the video frame and input the distance between them. This step allows the tool to approximate the distance per pixel, which will then be required to calculate the step length.

After completing all these necessary manual steps, the tool can output the following gait features for each leg and each walking sequence with 10 features per walking sequence. A visual representation of these features is depicted in Appendix A.1.

- Left and Right Step time: Duration in seconds between consecutive bilateral heel-strikes.
- Left and Right Stance time: Duration in seconds between heel-strike and toe-off of the same leg.
- Left and Right Swing time: Duration in seconds between toe-off and heel-strike of the same leg.
- Right to Left and Left to Right Double support time: Duration in seconds between heel strike of one leg and toe-off of the contralateral leg

- Left and Right Step length: Distance in meters between the ankles at heel-strike

4.3 GaitExtract: Gait Feature Extraction

To exploit this tool for gait information extraction in MR applications, we automate it to design our framework, which we refer to as **GaitExtract** as seen in Figure 4.2. **GaitExtract** takes a walking sequence video captured using the Hololens headset as an input and then proceeds with applying OpenPose algorithm to generate the pose keypoints, which will then be used to calculate the different gait features. Notable adjustments made to the tool include automating the identification of walking sequence direction (left to right or right to left), automating the correction of the left-right leg identification, automating the detection of the heel-strike and toe-off events, and automating the process of running the gait extraction for multiple walking sequence. **GaitExtract** can also detect if there is insufficient data to account for a full gait cycle in a walking sequence. A full gait cycle is defined to be when the same leg can complete at least two consecutive toe-offs or heel-strikes [61]. If **GaitExtract** detects insufficient information for a full gait cycle, it disregards this walking sequence and proceeds to process the following one.

Our **GaitExtract** framework yields a partially automated toolchain for extracting gait information. In the beginning, minimal user input is needed to input the two markers and distance measurements for each person's gait analysis. As mentioned earlier, the marker and the distance measurement input are needed to calculate the step length feature, which is why user input is still required for this feature. **However, if the step length feature is not needed or necessary, then the framework can be fully automated.** In Section 4.4, we will provide an evaluation of the effect of the step length feature in identifying people.

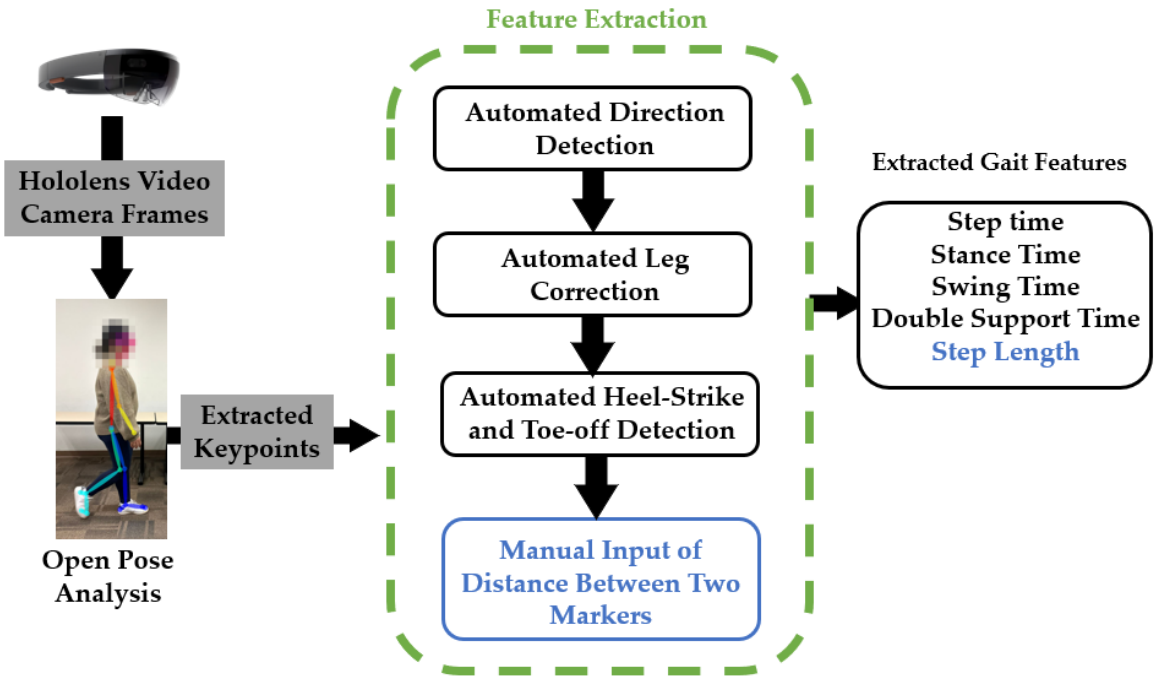


Figure 4.2: **GaitExtract**: A gait extraction framework built on top of OpenPose library. **GaitExtract** uses camera frames obtained from the HoloLens. The automated feature extraction algorithm uses the keypoints generated by OpenPose to extract the gait features. The manual input of distance between two markers is only necessary if the step length feature is required.

4.4 Identification via gait features

To show the capability of **GaitExtract** to uniquely identify people based on their gait features, we trained a supervised classifier on the features generated by **GaitExtract** from 20 participants in our user study. The generated features among the different participants highly varied, where the mean of the extracted features per person was 140.3 with a standard deviation of 43.42 and a range of 146. The high standard deviation and range are caused by the different ways people walk. A person with a small step length and walking fast would have many more extracted features than a person with a longer step length and walks slowly in the same time duration.

We used `GradientBoostingClassifier` classifier provided by the `scikit-learn` library

in Python [21]. We repeated stratified cross-validation with 3-folds and repeated the cross-validation 2 times to train the classifier. We used stratification to address the class imbalances among the number of extracted features. Cross-validation with 3-fold was used to ensure that an ample amount of testing data was still included even for classes with few features. Furthermore, we trained the classifier using data with and without step length to observe the capability of a fully automated framework without the step length feature to identify people uniquely.

We used the mean of weighted F1 scores across all folds to quantify classification accuracy. Results show that the model without step length gives 68% accuracy and that the model with step length gives a 78% accuracy in user identification. Figure 4.3 shows the average confusion matrix of the classifier on data **with step length**. User 5 has the lowest classification accuracy of 52.59%, which can be attributed to the fact that user 5 had the least amount of gait features[1].

These results show that an off-the-shelf classifier, specifically `GradientBoostingClassifier`, can uniquely identify people with reasonable accuracy when using a partially automated **GaitExtract** (with step length) as well as using a fully automated **GaitExtract** (without step length). Results also show that the number of gait features per class heavily influences inter-class classification accuracy.

To summarize, **GaitExtract** is capable of maliciously collecting gait information using minimal effort and minimal computational resources.

[1] User 5 was observed to walk very slowly, leading to fewer gait features collected during the data collection.

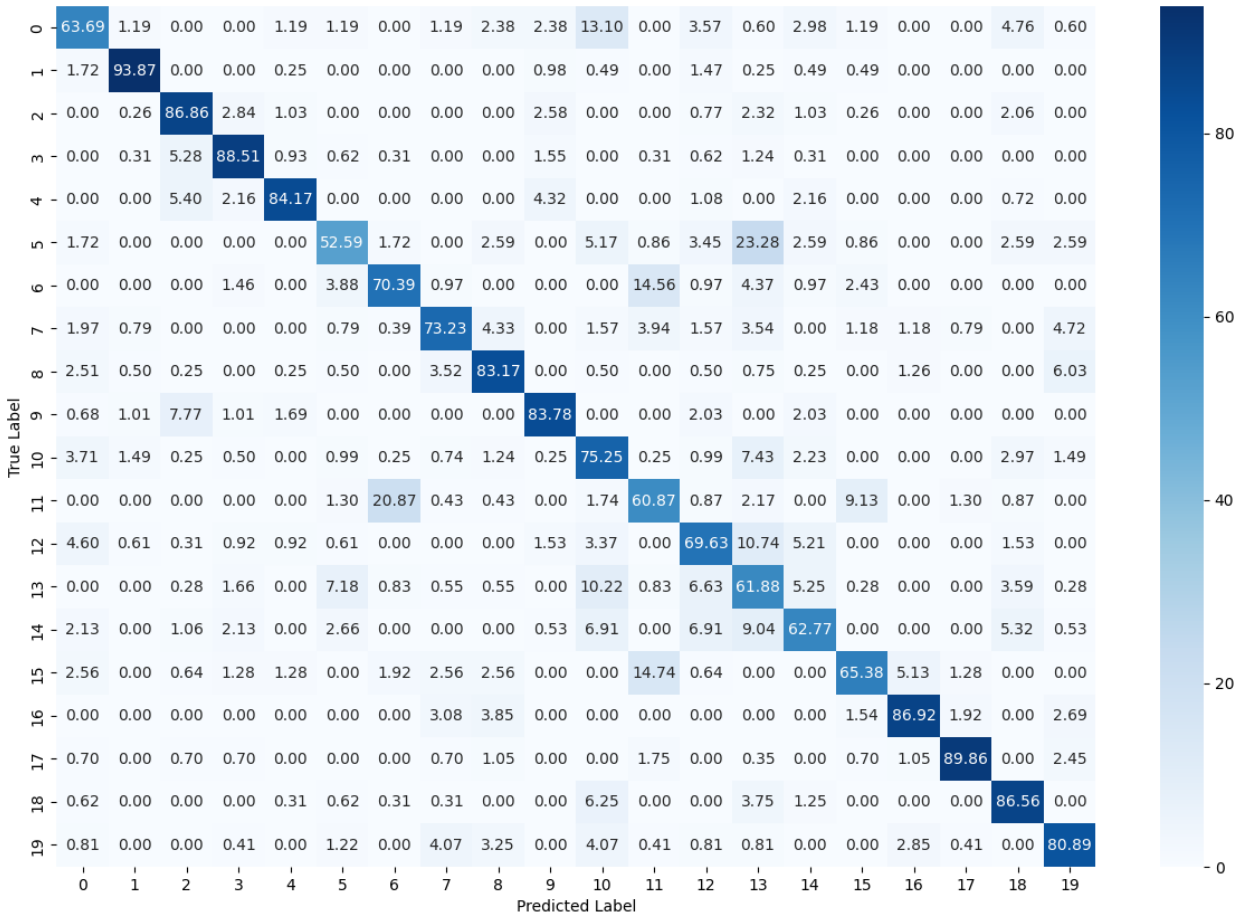


Figure 4.3: Average confusion matrix across all folds on data with step length.

Chapter 5

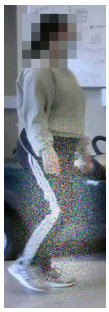
Mitigation Framework

Given the minimal resources required for an attacker to collect gait information from camera frames in MR applications, as elaborated in Section 4.4, developing strategies for safeguarding gait privacy becomes crucial. This section discusses the mitigation strategies for user identification privacy leaks through gait information. We focus on defenses applied on the image itself.

5.1 Mitigation Approaches

As explained in Section 4, our proposed **GaitExtract** relies on detecting heel-strike and toe-off gait events using three keypoints generated by OpenPose, specifically the midhip, left ankle, and right ankle. Hence, we propose two approaches for mitigation at different granularity, both applying pertubation (of diffeent types) directly on the video data. Our two mitigation approaches, based on our proposed **GaitExtract**, differ only on where the pertubation is applied, and are as follows:

- **Keypoints-based Masking (KPM):** Since **GaitExtract** utilizes data from just three



(a) LBM: Lower-body Masking.

(b) KPM: Keypoints Masking.

Figure 5.1: Two different granularity of perturbations [1].

keypoints, we conducted experiments involving the application of perturbations methods around these specific keypoints.

- **Lower body-based Masking (LBM):** Since **GaitExtract** derives information solely from the lower half of the body, we conducted experiments involving the application of perturbations methods on a bounding box surrounding the lower half of the body.

These two approaches, highlighting the application of perturbations at these two granular levels, are shown in Figure 5.1, providing a visual representation of the perturbations in a video frame.

5.2 Perturbation Methods

Numerous strategies exist within the domain of privacy-preserving techniques for safeguarding camera frames or photos, such as adversarial perturbations [63], privacy-preserving GANs (Generative Adversarial Networks) [57], human-imperceptible privacy protection [56] and quantization [43]. A drawback on many of these strategies is that they are resource-intensive and not optimized for real-time applications. Considering that one of the objectives of this work is to implement a real-time gait privacy solution, in this paper, we deliberately investigate fundamental perturbation methods (i.e., noise adding, blurring, masking) and the

[1]The pixelation present on the face is for anonymity purposes and not part of the mitigation.

effects of these methods in fortifying privacy in conjunction with application utility. Particularly, we explored adding differentially private (DP) pixelization, blurring, and four types of random noise.

Pixelization is a standard image obfuscation technique where the resolution of a part or the whole image is reduced by replacing groups of pixels of a certain kernel size with a representative value, which is typically chosen by getting the minimum, maximum, or average value of the original pixels in that group [1]. Extending this standard obfuscation technique, differentially private (DP) pixelization introduces privacy guarantees to pixelization. In particular, DP pixelization guarantees that two "neighboring images" with the same dimension differing by at most m pixels are indistinguishable [37]. To exploit the effect of this DP pixelization in protecting gait information, we used different kernel sizes b while ensuring that the m is equivalent to the number of pixels in the regions identified by the KPM and LBM granularity. In particular, we vary the pixelization kernel size b, which dictates the coarseness of the applied pixelization and thus influences the Laplacian distribution scale when offering privacy guarantees. Furthermore, the findings of [37] report that DP pixelization significantly reduces attack success at low privacy requirements of $\epsilon \geq 0.1$ and $m = 16$. Motivated by this finding, we chose an $\epsilon = 0.1$ for our experiments.

Another common image perturbation technique is blurring, where an image is convolved with a low-pass filter to remove high-frequency content from images [25]. This is also a common way to remove sensitive information from video content. An example is YouTube's face blurring feature for content creators [24]. To investigate the capability of this perturbation technique in protecting gait privacy, we adopted Gaussian blurring, a common technique for blurring. We varied the size of the Gaussian kernel applied to the KPM and LBM regions.

Furthermore, we also applied four types of random noise characterized by varying the distribution parameter denoted as λ, which influences the shape and characteristics of the noise introduced by each distribution.

We summarize the perturbation methods as follows:

- Differentially private (DP) pixelization: Involves three parts: (1) identify m which represents the number of pixels in the area of interest identified by KPM or LPM (2) pixelization of kernel size b within the area m, and (3) apply the Laplace mechanism where noise from the Laplacian distribution is added to each pixelized cell of size $b \times b$ with the scale parameter equivalent to $\lambda = \frac{3*255m}{b^2\epsilon}$ [37]. In the case of KPM, $m = 100 \times 100$ surrounding the key points, while in LPM, m is equivalent to the area of the lower body. In our experiments, we choose $\epsilon = 0.1$ and kernel sizes $b = \{10, 20, 30, 40, 50\}$.
- Gaussian Blur: A 2D Gaussian kernel of size $c \times c$ and standard deviation of 0 is used for blurring. We chose $c = \{5, 25, 45, 65\}$ in our experiments.
- Uniform Noise Distribution ($\mathcal{U}$): Generates noise with a uniform distribution, represented as $\mathcal{U}(-\lambda, \lambda)$. In this distribution, all values within the specified range are equally likely with probability density function (PDF): $f(x) = \frac{1}{2\lambda}$, where $x \in [-\lambda, \lambda]$.
- Normal Noise Distribution ($\mathcal{N}$): Characterized by a normal distribution, denoted as $\mathcal{N}(0, \lambda)$, centered at 0, and introducing variations with a standard deviation of λ. Hence, the PDF: $f(x) = \frac{1}{\sqrt{2\pi\lambda}} \cdot e^{-\frac{x^2}{2\lambda^2}}$ for $x \in (-\infty, \infty)$.
- Laplace Noise Distribution ($\mathcal{L}$): Generates noise with a Laplace distribution, expressed as $\mathcal{L}(0, \lambda)$, centered at 0, and with the scale parameter set to λ. Hence, the PDF: $f(x) = \frac{1}{2\lambda} \cdot e^{-\frac{|x|}{\lambda}}$ for $x \in (-\infty, \infty)$.
- Exponential Noise Distribution ($\mathcal{E}$): Generates noise with an exponential distribution, denoted by $\mathcal{E}(\lambda)$, where the rate parameter is set to λ. Hence, the PDF: $f(x) = \lambda \cdot e^{-\lambda x}$ for $x \geq 0$.

We assign four different values for $\lambda = \{50, 100, 150, 200\}$ in the four random noise distributions.

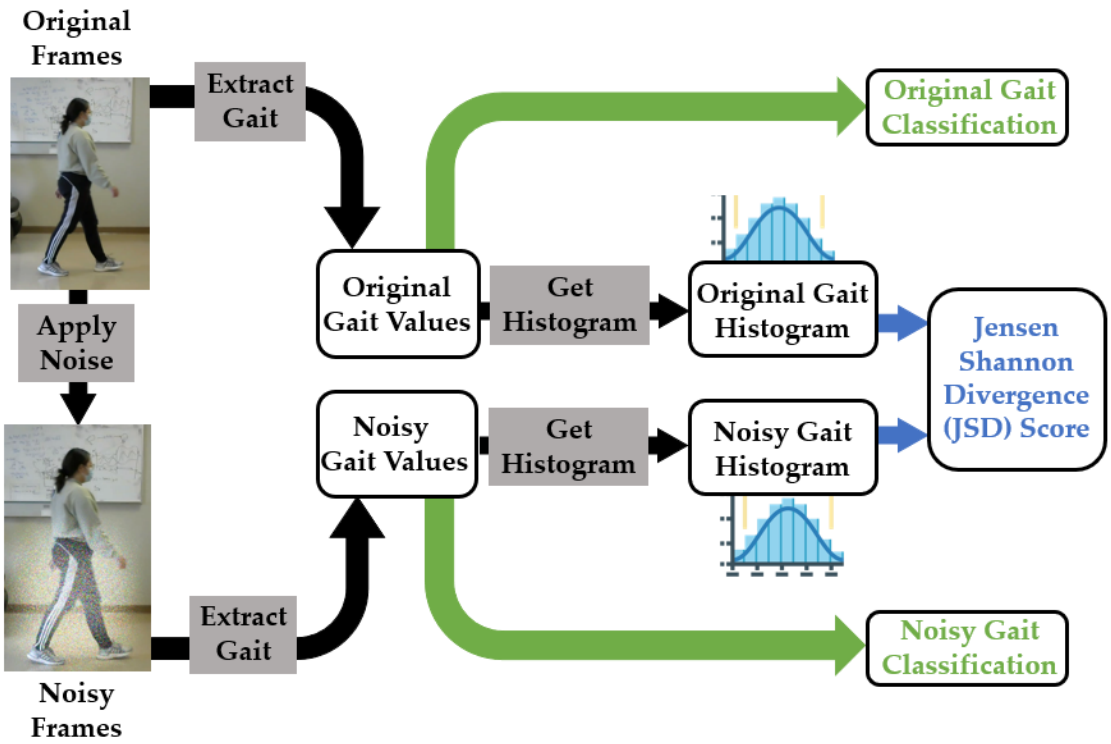


Figure 5.2: An illustration of the mitigation and evaluation framework. Noise is first applied to the original camera frames to generate the noisy frames. **GaitExtract** is then applied to the original and noisy frames to generate gait values. Classification is then done on the original and noisy gait values. The Jensen-Shannon Divergence(JSD) Score is calculated for the original and noisy gait values.

5.3 Mitigation Evaluation Metrics

We show a pictorial overview of the mitigation process and its evaluation in Figure 5.2. The mitigation process involves the application of noise to the original frames (as explained in Section 5.1), followed by utilizing the **GaitExtract** framework on both the original and noisy frames. The gait values obtained from the original frames are denoted as G, encompassing a set of values for various gait parameters, including step time, stance time, swing time, double support time, and step length for both the left and right leg, as mentioned in Section 4.2. Similarly, the gait values derived from the noisy frames are denoted as G'.

To assess the efficacy of the applied noise in safeguarding gait privacy, we used quantitative and qualitative metrics.

Privacy loss metrics: We use two metrics as follows:

- Jensen-Shannon Divergence (JSD) Score: The JSD score is computed based on the histograms of gait features in G and G'. The JSD score offers insights into the similarity or dissimilarity of the distributions of gait parameters. A JSD score of 0 indicates an identical histogram distribution, signifying perfect similarity, while a score of 1 suggests a completely different distribution, indicating maximum dissimilarity.
- User identification accuracy: User classification is performed using the gait parameters (G) and (G') employing the classification technique outlined in Section 4.4. In particular, we want to evaluate how the mitigation techniques impact the accurate identification of individuals based on their gait features. The user classification results provide insights into the practical implications of the mitigation strategies on the overall effectiveness of gait-based identification systems.

Combining quantitative JSD scores and user classification results enhances our understanding of the trade-offs and performance metrics associated with the applied mitigation.

Utility metrics: To quantify the utility of the mitigation, we measured the change in the quality of the video post-mitigation by measuring the peak signal-to-noise (PSNR) and the mean squared error (MSE) of the modified pixels along the red, green, and blue channel of each frame. We also evaluated the difference in the luminance, contrast, and structural information between the original G and the mitigated frame G' by measuring the structure similarity index (SSIM) [26]. Furthermore, to quantify the utility of the system implementation, we measured the application frame rate (fps) and the latency at each point of receiving the camera frame. We evaluated the frame rate and the latency in Sections 6 and 6.3.

Qualitative measures: We used qualitative metrics to asses our mitigation strategy by conducting a user survey on the 20 participants. The results of this survey will be discussed in Section 6.6.

5.4 Privacy-Utility Tradeoff

We used the metrics discussed in Section 5.3 to evaluate the privacy-utility tradeoff (PUT). Firstly, to observe the effect of the mitigation in changing the distribution of the gait features, we measured the JSD across all features (explained in Section 4.2) and across all 20 participants. The change in the distribution of each gait feature as measured by the average JSD is illustrated as a heatmap in Figure 5.3. We observe that most of the JSD values for the KPM methods were less than 0.5. The LBM methods were mainly greater than 0.5, indicating that the LBM can alter the gait feature distribution more than KPM. We observe that the gait feature "stance time" is the least affected by the KPM methods.

Furthermore, we summarize the PUT results of the different mitigation approaches in Table 5.1. Specifically, we focus on the second metric of privacy loss, namely user identification accuracy. This is determined by the decrease in classification accuracy percentage achieved by applying **GaitExtract** to both G and G', as detailed in Section 5.3. This privacy loss metric was evaluated alongside three utility metrics: Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity Index Measure (SSIM). Comprehensive insights into the performance of each mitigation approach, along with the effects of the adjusted parameters (λ, b, and c) on all privacy loss metrics (Jensen-Shannon Divergence (JSD), user identification accuracy) and utility metrics (PSNR, MSE, SSIM), are available in Appendix A.3.

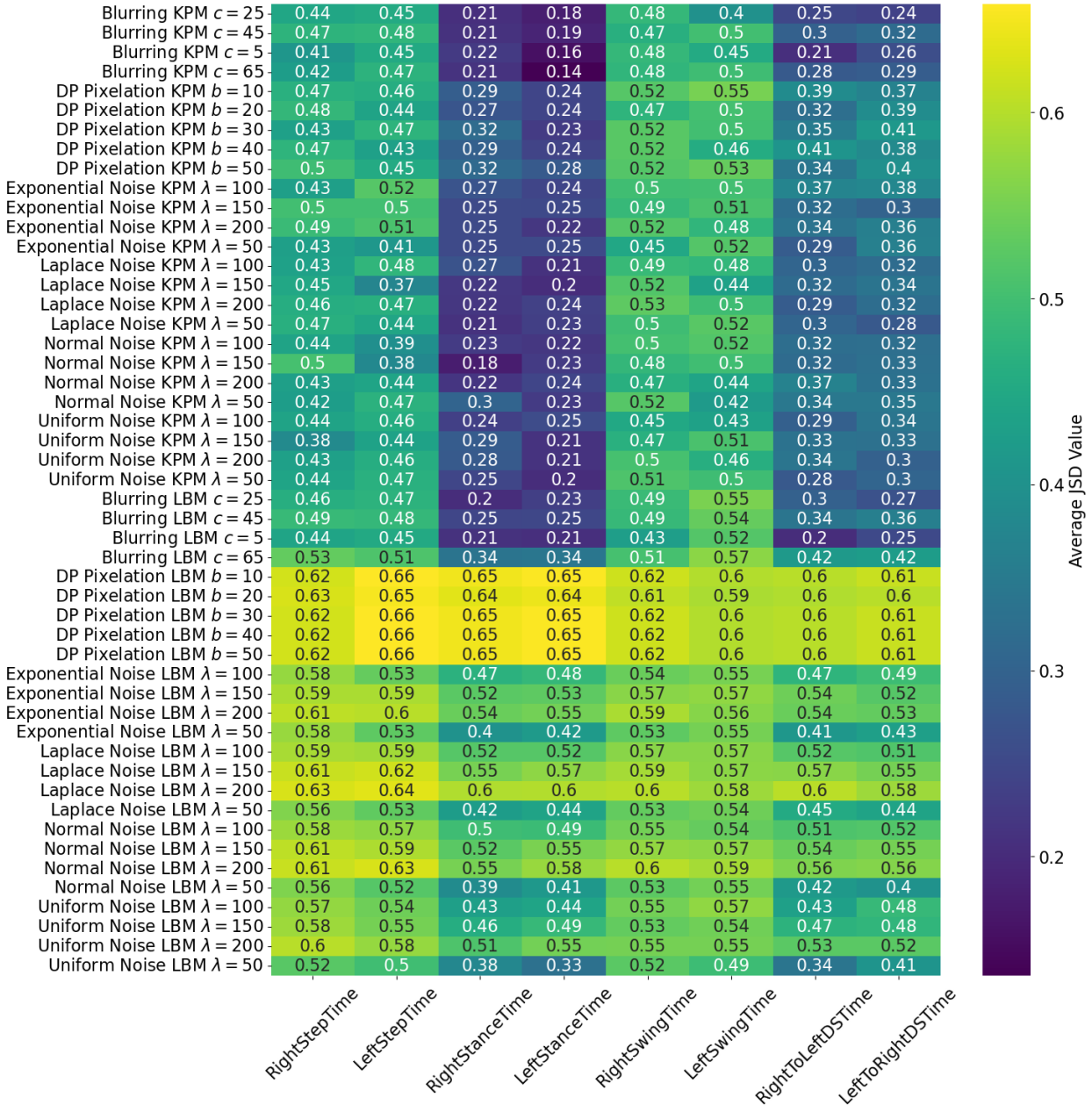


Figure 5.3: Heatmap of average JSD value for each gait feature across all the mitigation experiments.

Targeted Keypoint Mitigation (KPM) The PUTs for KPM (Table 5.1-first column) show that while KPM had better PSNR and SSIM in comparison to LBM, it was not successful in introducing any significant reduction in user identification accuracy. More specifically, the reduction in user identification accuracy percentage for KPM for all the perturbations methods was less than 10%. This shows that the KPM approach was ineffective

in mitigating the gait privacy leak in MR. This finding can be attributed to the nature of the OpenPose pose estimation algorithm, where the poses are assembled by using part affinity fields (PAFs) and these PAFs can encode unstructured relationships between body parts [28]. Thus, OpenPose can still predict the location of the midhip, left ankle, and right ankle keypoints with good accuracy even when these areas are noisy because it can estimate the location of the aforementioned keypoints in reference to other body parts (i.e., left leg, right leg, torso, etc.). These results highlight the high redundancy in gait features. In particular, focusing on anonymizing a single feature is unlikely to be effective. This observation aligns with findings in human perception studies, where removing some local information does not significantly impact recognition as long as global form and dynamic posture changes are preserved [42].

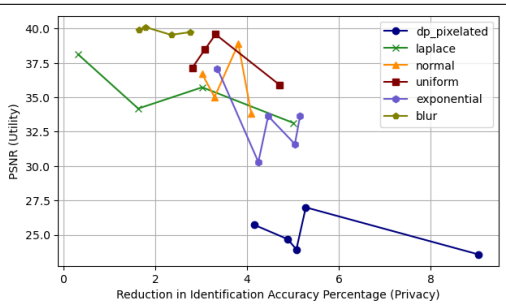
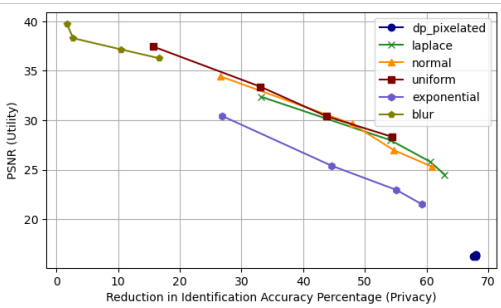
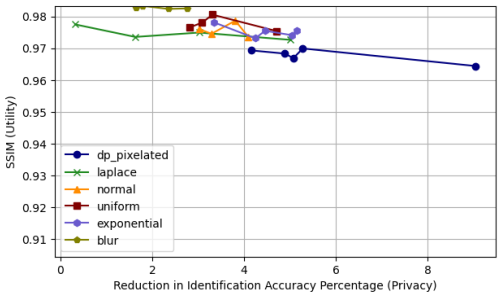
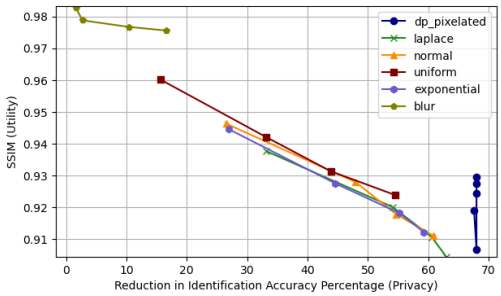
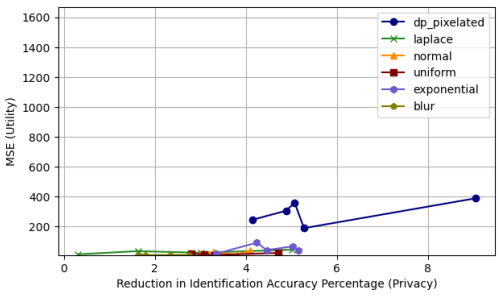
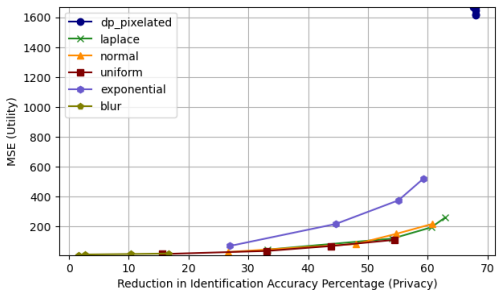
Lower Body Mitigation (LBM) The PUTs for LBM (Table 5.1-second column) show that most of the perturbation methods were able to reduce user identification accuracy significantly. In particular, we observed that when the JSD score was greater than 0.5, indicating a greater shift in the gait features values, the reduction in the user identification accuracy was greater than 40% with up to a reduction of 68%.

This highlights that LBM effectively reduces an attacker's capability to link gait features back to a user. However, among the LBM perturbation methods, blurring had the least capability to affect the identification accuracy. DP pixelization, on the other hand, was very effective in protecting gait information even at small kernel sizes, where when $b = 5$, the reduction in identification accuracy was already 68%. However, DP pixelization had the worst PSNR, MSE, and SSIM values among all the perturbation methods.

The performance of the four random noise perturbations was comparable across the three utility metrics, with the exponential distribution showing marginally lower scores in PSNR and MSE. The Laplace, normal, and uniform noise distributions yielded similar PSNR, SSIM,

and MSE results. However, employing Laplace noise distribution achieved the highest decrease in user identification accuracy across all values of λ, achieving a peak reduction of 62% in user identification accuracy. Additionally, the experiments show that the improvement in privacy loss mitigation is negligible beyond $\lambda = 100$. This implies that the **most effective strategy for mitigating gait privacy leak in MR is to use the LBM approach with Laplace noise distribution at** $\lambda = 100$. More details on the performance of the Laplace noise in LPM are shown in Appendix A.3.

Table 5.1: The privacy-utility trade-off (PUT) curves of the two mitigation approaches (KPM and LBM) - without "step length" gait feature - across all perturbation methods. The privacy loss metric (reduction in user identification accuracy) was compared with three different utility metrics: (1) PSNR, (2) SSIM, and (3) MSE. *Note that the scale for the privacy loss axis (x-axis) on the KPM figures (first column) only goes up to* 10% *while for LBM (second column) it goes up to* 70%.

	Keypoints-based mitigation approach (KPM)	**Lower body-based mitigation approach (LBM)**
PSNR vs. Privacy		
SSIM vs. Privacy		
MSE vs. Privacy		

Chapter 6

GaitGuard System Design

We propose **GaitGuard** system which provides a systematic approach to mitigate against gait-based user identification built upon video-based approaches such as **GaitExtract**. Our system can be used in collaborative app setup as explained in Section 3. We implemented **GaitGuard** on Mixed Reality headset Hololens 2.

6.1 Core Functionality of GaitGuard

Drawing insights from the results of the mitigation experiments. It was concluded that the LBM approach using the laplacian noise distribution and $\lambda = 100$ yielded the best privacy-utility trade-off. Motivated by this finding, we propose a system that implements this mitigation to protect gait-based information in MR applications.

GaitGuard employs two key functionalities to implement the mitigation. The first involves detecting the lower body of individuals within the camera frame, while the second entails applying an optimal noise configuration.

To pinpoint the area in the frame where noise will be applied, specifically the lower body, person detection is imperative. Ideally, a pose estimation algorithm like OpenPose would intuitively serve this purpose. However, many existing pose estimation libraries have not been adapted for emerging Mixed Reality (MR) technologies. To circumvent this challenge in detecting the lower body, we leverage OpenCV's Histogram of Gradients (HOG) [17] default people detector along with HOG's pre-trained SVM classifier.

Initially, **GaitGuard** identifies users in the frame and returns their locations through rectangular bounding boxes. As the bounding box from OpenCV's default people detector covers the entire body, an ad hoc solution is required to extract the lower body bounds. To approximate the lower body, the height of the obtained rectangular bounds is halved, retaining only the bottom half of the rectangle. Subsequently, **GaitGuard** applies the optimal mitigation to the identified area by generating noise of the same size as the lower body bounds across the three image channels (red, green, blue). A Hanning window, matching the height of the lower body bounds, is created to smooth the noise application and lessen its impact on frame clarity. Finally, the windowed noise is applied to the camera frame.

6.2 Challenges for On-Device Implementation

As mentioned in Section 3, the initial point of entry for camera information is the application itself, suggesting that the gait privacy protection solutions must be implemented at the OS level. However, OS level implementation is limited because the OS of Hololens 2, the Windows Holographic OS [12], is inaccessible for modification. This is a known limitation of the Hololens 2 where related work in the literature, such as BystandAR [32] also implemented their solution as a third-party application because of limited OS level access. Thus, an alternative on-device implementation that could serve as a proof of concept is to implement **GaitGuard** as a third-party application. However, running computationally intensive algo-

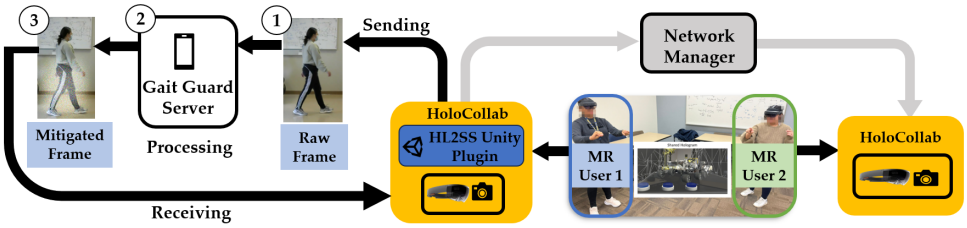


Figure 6.1: **HoloCollab** with **GaitGuard** Implementation. Raw camera frames are streamed from User 1's **HoloCollab** to the **GaitGuard** server at a rate of 30 FPS using the HL2SS Unity plugin. **GaitGuard** then processes the frames at a rate of 25 FPS and sends the mitigated frames back to User 1's **HoloCollab**. User 1's **HoloCollab** receives the mitigated frames at a rate of 22 FPS.

rithms such as person detection algorithms on a resource-constrained device (Hololens 2) for every camera frame introduces unacceptable latency, making it impractical. Appendix A.2 shows more details on the on-device implementation. Our insights from implementing **GaitGuard** on the device suggest that it is impractical due to constrained resources. However, this effectively means that on-device gait leak attacks based on video-based approaches, such as pose estimation or pedestrian detection, are also impractical. Therefore, with the current state of Mixed Reality technology, there is no practical way of gathering gait information using video-based approaches on-device because of the computationally intensive nature of those algorithms.

6.3 Collaborative MR Design with GaitGuard

To address the challenges presented by the on-device implementation of **GaitGuard**, we need to offload the implementation of **GaitGuard** to a trusted local mobile server. An overview of this implementation can be seen in Figure 6.1.

The trusted mobile server implements the core functionality of **GaitGuard** before the frames are sent to the network manager. The flow of information begins with the Hololens, where

instead of sending the RGB camera frames to the network manager, it is first sent to the **GaitGuard** server so that the frames received by the network manager are mitigated. In this architecture, the core functionalities of **GaitGuard**, including person detection, are offloaded to the trusted server, which diverts the computational load from the Hololens. This section will elaborate on how we use **GaitGuard** on a local server to address the collaborative MR application threat model (explained in Section 3).

6.4 HoloCollab Design & Implementation

We built a collaborative MR application, **HoloCollab**, to assess the effect of **GaitGuard** on the performance of a collaborative application. We built the application prototype in Unity 2021.3.25f1, using MRTK3, the third generation of Microsoft Mixed Reality Toolkit for Unity. Additionally, we used the Photon Unity Networking (PUN) 2 version 2.43 package to integrate multi-user functionality in the application. We used Hololens 2 Sensor Streaming (HL2SS) [36] plugin to stream camera frames from the Hololens 2 to the **GaitGuard** local server. **GaitGuard** was implemented as a mobile server using Python and was written in $\approx$ 270 lines of code. The collaborative application was deployed on a Microsoft Hololens 2 device running Windows Holographic for Business Build 22621.1252. An overview of our **HoloCollab** prototype implementation can be seen in Figure 6.1.

6.5 HoloCollab Evaluation

The suggested application frame rate for the Hololens 2 is 60 fps to ensure the best quality in user experiences [7][1]. Moreover, it has been reported that the capability of streaming and recording frames from the RGB camera in Hololens 2 containing virtual objects, referred to

[1]Application frame rate is a metric that Microsoft uses to qualify the quality of their applications [4].

as Mixed Reality Capture (MRC), has an average frame rate of 30 fps [6].

To evaluate the performance with **GaitGuard** on **HoloCollab** we measured the system's application frame rate and camera streaming frame rate corresponding to the application latency.

The application frame rate of the **HoloCollab** without streaming is approximately 50 fps. Consequently, the application frame rate of **HoloCollab** when streaming camera frames to a vanilla server designed to ingest the frames simply dropped to 40 fps. We also observed that the application frame rate of **HoloCollab** with **GaitGuard** also resulted in 40 fps. This suggests that the decrease in application frame rate was caused by the use of HL2SS plugin and not by **GaitGuard**. Furthermore, this also implies that the **GaitGuard** local server does not affect the application frame rate.

We also report the camera streaming frame rate to identify **GaitGuard**'s effect on streaming latency. The average camera streaming fps at which the HL2SS plugin streams the camera frames is 30 fps, introducing latency of ≈ 0.03 seconds at point number 1 in Figure 6.1. **GaitGuard** server then processes the camera frames at 24 fps, introducing latency of ≈ 0.047 seconds at point number 2 in Figure 6.1. **GaitGuard** server then sends mitigated frames at a rate of 21 fps introducing latency of ≈ 0.049 seconds at point number 3 in Figure 6.1. Finally, Hololens users receive these frames at 21 fps. Hence, **GaitGuard** introduces a latency of 0.016 seconds, which is the latency between point 1 and point 3 before the collaborative MR application sends the camera frames to the network manager.

6.6 Qualitative Evaluation of GaitGuard

We conduct a user study to delve deeper into perceptions of gait privacy. The user study was conducted among 20 participants and was split evenly between individuals with and

without prior experience in VR/MR applications, offering a diverse perspective on privacy concerns. Respondents were engaged through a two-part survey designed to clarify their baseline comfort levels regarding gait privacy and the influence of the **GaitGuard** on their post-exposure to privacy risk information. In particular, the first part consisted of gauging their preconceived thoughts on gait privacy in MR, and the second part was designed to gauge how their perceptions are affected after being informed of the gait privacy risks in MR and the effectiveness of **GaitGuard** on perturbing identification through gait.

In the first part of the survey, participants articulated their comfort with sharing video frames that can leak their gait information, both in unaltered form and with **GaitGuard** using Likert Scale [3], with 1 being extremely uncomfortable and 5 being extremely comfortable. The findings indicated a dichotomy in privacy valuation, with 50% expressing discomfort with the release of raw video frames of them walking to third-party applications that can have their raw gait data, while 40% were comfortable. Yet a notable increase in the comfort level by 65% when the data was passed by **GaitGuard**. The results of the first half of the survey suggest about half of the participants were not concerned with sharing videos of them walking to third-party applications. The results are shown in Figure 6.2.

For the next part of the survey, the participants were informed about specific privacy risks associated with MR, such as its correlation to sensitive information, such as ethnicity [64], age [66], gender [62], and neuromusculoskeletal disorders [58]. Informing the participants about the **GaitGuard**'s efficacy in obscuring gait data and protecting the identity change their response. In particular, the participants were then informed that experiments show that applying noise on the lower body significantly reduces the identification accuracy from 78% to 16% but with a reduction in the utility by adding a total of 0.016s latency in streaming the camera frames.

The aggregate results of the second section can be seen in Figure 6.3. Results demonstrate that after being informed of the gait privacy risks in MR, the overall comfort level decreased

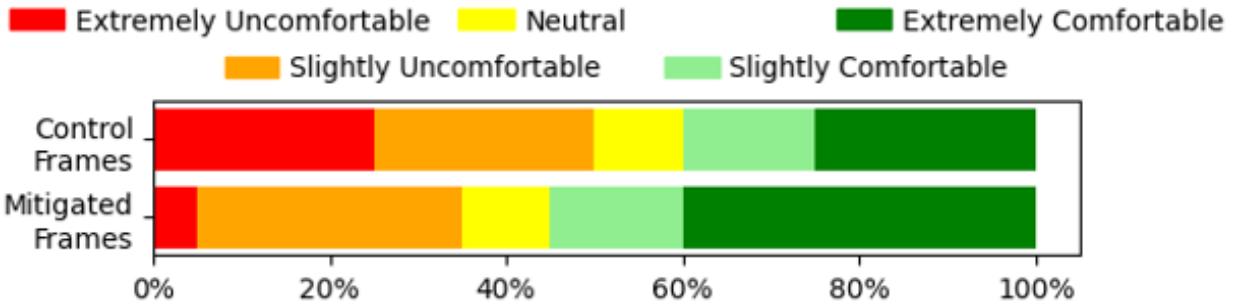


Figure 6.2: Before being informed about gait privacy concerns in MR and **GaitGuard** performance. The distribution of Likert responses for the gait privacy perception survey across unmodified frames (Control Frames) and frames with **GaitGuard** (Mitigated Frames).

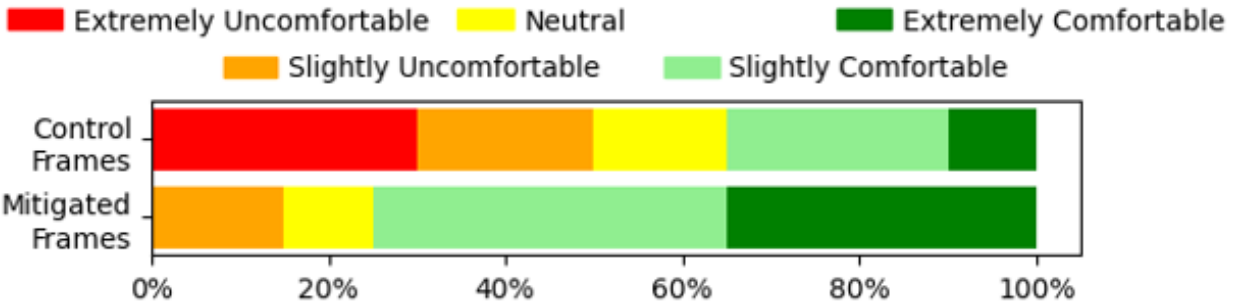


Figure 6.3: After being informed about gait privacy concerns in MR and **GaitGuard** performance. The distribution of Likert responses for the gait privacy perception survey across unmodified frames (Control Frames) and frames with **GaitGuard** (Mitigated Frames).

compared to before they obtained this information. However, it should be noted that 35% of the respondents were still comfortable with releasing their raw frame. Although there was a slight decrease in the comfort level with raw data being shared, a substantial majority of 75% reported an increase in comfort level with sharing **GaitGuard** processed data.

This viewpoint change highlights the importance of obtaining informed consent and raising awareness about privacy safeguards. It further implies that although **GaitGuard** successfully reduces the risk of identification, the balance between privacy and utility remains a pivotal concern for users.

Chapter 7

Discussion

This paper shed insights into understanding gait privacy concerns and the effectiveness of anonymization methods such as **GaitGuard** in MR technology. Future efforts should focus not merely on advancing technology but also on promoting privacy consciousness and adopting a user-centric approach to developing new digital spaces. This is essential for ensuring that technological innovations are effective and aligned with user needs and privacy expectations in evolving digital landscapes. We discuss in this section some of the limitations of the current work and the possible extensions for future work.

Scope. GaitExtract is designed to extract gait information from users and to showcase that this extracted information can uniquely identify people. **GaitExtract** has a limitation where if the user does not complete a gait cycle (two heel-strikes or two toe-offs) [50] in a walking sequence then **GaitExtract** is unable to extract any gait information. However, as mentioned in Section 4.1, the range of the walking sequences gathered in this study was between $2.5m$ to $2.75m$, which was also a short walking sequence distance and results show that even with a short walking sequence distance there was still a gait privacy leak, underscoring the importance of developing and implementing tools that protect gait information.

Threat Model The focus of this work was to investigate and mitigate the vulnerabilities towards neighbor gait privacy in MR applications using camera based attacks as mentioned in Section 3.5. It should be noted that there has been work in egocentric pose estimation where the body pose estimation of a headset user can be approximated by mounting a camera to the headset [60]. However, unlike **GaitExtract** where Stenum, et. al. [59] has proven that gait feature extraction using side profile walking videos has good measurements in comparison to clinical gait measurements through motion capture systems, it is still unclear if egocentric pose estimation is an accurate method of measuring gait features. Additionally, commercially available MR headsets currently do not have this feature of having installed cameras that are pointed towards the user's body. While this is the case, egocentric pose estimation displays potential in gait feature extraction and underscores the need to address gait privacy concerns as MR technology improves.

On device and OS-level implementation. As discussed in section 6.2, it is currently impractical to gather gait information at the application level due to the computational constraints of current MR technology. However, it should be noted that once MR technology has matured enough to allow pose estimation algorithms to run in real-time and on the device, **GaitExtract** is theoretically fully capable of gathering outside gait information. Bearing this in mind, it is imperative for the MR device developers to implement **GaitGuard** at the OS level to protect users' gait information.

Qualitative measures and human perceptions of privacy. For future studies, it is crucial to broaden the scope of investigation to a larger and more diverse group. Delving into the psychological and social foundations of privacy perceptions within MR environments promises to uncover more profound insights into user behaviors and preferences. Furthermore, conducting comparative analyses of different anonymization techniques will enrich our understanding of the effectiveness and user acceptance of privacy-preserving technologies in MR.

Gait privacy aware pose estimation algorithms. GaitExtract uses OpenPose to approximate the location of the midhip, left ankle, and right ankle keypoints to obtain the gait features. An intuitive solution for privacy protection is to apply the noise on the keypoints generated by OpenPose and propose a gait privacy aware pose estimation algorithm. However, as mentioned in Section 3 the threat to gait privacy is eminent as long as the attacker has access to the camera frames because an attacker has accesss to OpenPose and other pose estimation algorithms that are privacy unaware. An expert attacker could even opt to train and develop it's own pose estimation algorithm to extract gait information. This highlights that the protection for gait privacy must be applied at the video level.

Chapter 8

Conclusion

In this paper, we identified the fundamental need to explore the state of gait privacy in MR applications by developing an automated gait extraction framework called **GaitExtract** and revealing that an attacker can maliciously collect gait information using minimal resources. Furthermore, motivated by this privacy leak we developed **GaitGuard** a safeguard against the leak of gait information and demonstrated that we reduced the privacy leak by as much as 62% while only introducing a latency of 8 fps in the streaming of frames. Lastly, the implementation **GaitGuard** as a mobile server of a collaborative application showed that **GaitGuard** did not affect the application frame rate of the MR device and only introduced 8 FPS latency in streaming video frames. We believe that this work expands the understanding of gait privacy in MR applications and that our contribution is a step towards ensuring gait security and privacy of MR application of the future.

Bibliography

[1] Pixelization. American Heritage® Dictionary of the English Language, Fifth Edition, 2011. [Online; accessed February 24, 2024].

[2] The CPRA. https://thecpra.org/, 2013.

[3] Likert Scale: Definition, Examples & How to use it. https://www.questionpro.com/blog/what-is-likert-scale/, 2018.

[4] App quality criteria - Mixed Reality. https://learn.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/app-quality-criteria-overview, 2022.

[5] Introduction to the Multi-user capabilities tutorials - Mixed Reality. https://learn.microsoft.com/en-us/windows/mixed-reality/develop/unity/tutorials/mr-learning-sharing-01, 2022.

[6] Mixed reality capture overview. https://learn.microsoft.com/en-us/windows/mixed-reality/develop/advanced-concepts/mixed-reality-capture-overview, 2022.

[7] Performance - MRTK 2. https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk2/performance/perf-getting-started?view=mrtkunity-2022-05, 2022.

[8] Apple Vision Pro. https://www.apple.com/apple-vision-pro/, 2023.

[9] Bouvet. https://www.microsoft.com/en-us/p/lens-by-bouvet/9p29ft8z34zs, 2023.

[10] Catapult Mixed Reality Client and Development Application. https://www.makesea.com/download/, 2023.

[11] General Data Protection Regulation (GDPR). https://gdpr-info.eu/, 2023.

[12] HoloLens 2 hardware. https://learn.microsoft.com/en-us/hololens/hololens2-hardware, 2023.

[13] HoloLens Microsoft 2. https://www.microsoft.com/en-us/hololens/, 2023.

[14] Hololens2forcv. https://github.com/microsoft/HoloLens2ForCV/tree/main/Samples/CameraWithCVAndCalibration, 2023.

[15] Meta Quest 3: Mixed Reality VR Headset. https://www.meta.com/quest/quest-3/, 2023.

[16] Microsoft Apps. https://apps.microsoft.com/home?hl=en-US&gl=US, 2023.

[17] OpenCV. https://docs.opencv.org/3.4/d5/d33/structcv_1_1HOGDescriptor.html, 2023.

[18] Overview of Dynamics 365 Remote Assist on HoloLens and HoloLens 2. https://learn.microsoft.com/en-us/dynamics365/mixed-reality/remote-assist/overview-hololens, 2023.

[19] Photon Unity Networking for Unity Multiplayer Games PUN2. https://www.photonengine.com/PUN, 2023.

[20] Pose Estimation. https://paperswithcode.com/task/pose-estimation, 2023.

[21] Sklearn.ensemble.GradientBoostingClassifier. https://scikit-learn/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html, 2023.

[22] Sphere XR. https://apps.microsoft.com/detail/9P11TRNWG2HR?hl=en-US, 2023.

[23] What is mixed reality? https://learn.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality, 2023.

[24] Blur your videos - YouTube Help. https://support.google.com/youtube/answer/9057652?hl=en, 2024.

[25] OpenCV: Smoothing Images. https://docs.opencv.org/4.x/d4/d13/tutorial_py_filtering.html, 2024.

[26] Structural Similarity Index - NI. https://www.ni.com/docs/en-US/bundle/ni-vision-concepts-help/page/structural_similarity_index.html, 2024.

[27] D. Baek, P. Musale, and J. Ryoo. Walk to show your identity: gait-based seamless user authentication framework using deep neural network. In *The 5th ACM Workshop on Wearable Systems and Applications*, pages 53–58, 2019.

[28] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017.

[29] A. Cappozzo. Gait analysis methodology. *Human movement science*, 3(1-2):27–50, 1984.

[30] K. Cheng, J. F. Tian, T. Kohno, and F. Roesner. Exploring user reactions and mental models towards perceptual manipulation attacks in mixed reality. In *USENIX Security*, volume 18, 2023.

[31] G. Cola, M. Avvenuti, F. Musso, and A. Vecchio. Gait-based authentication using a wrist-worn device. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, pages 208–217, 2016.

[32] M. Corbett, B. David-John, J. Shang, Y. C. Hu, and B. Ji. Bystandar: Protecting bystander visual data in augmented reality systems. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, pages 370–382, 2023.

[33] O. Dehzangi, M. Taherisadr, and R. ChangalVala. Imu-based gait recognition using convolutional neural networks and multi-sensor fusion. *Sensors*, 17(12):2735, 2017.

[34] P. Delgado-Santos, R. Tolosana, R. Guest, R. Vera-Rodriguez, F. Deravi, and A. Morales. Gaitprivacyon: Privacy-preserving mobile gait biometrics using unsupervised learning. *Pattern Recognition Letters*, 161:30–37, 2022.

[35] M. O. Derawi, P. Bours, and K. Holien. Improved cycle detection for accelerometer based gait authentication. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 312–317. IEEE, 2010.

[36] J. C. Dibene and E. Dunn. Hololens 2 sensor streaming. *arXiv preprint arXiv:2211.02648*, 2022.

[37] L. Fan. Image pixelization with differential privacy. In *Data and Applications Security and Privacy XXXII: 32nd Annual IFIP WG 11.3 Conference, DBSec 2018, Bergamo, Italy, July 16–18, 2018, Proceedings 32*, pages 148–162. Springer, 2018.

[38] H. Farrukh, R. Mohamed, A. Nare, A. Bianchi, and Z. B. Celik. {LocIn}: Inferring semantic location from spatial maps in mixed reality. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 877–894, 2023.

[39] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster. Full body gait analysis with kinect. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1964–1967. IEEE, 2012.

[40] D. Gafurov, E. Snekkenes, and P. Bours. Gait authentication and identification using wearable accelerometer sensor. In *2007 IEEE workshop on automatic identification advanced technologies*, pages 220–225. IEEE, 2007.

[41] S. R. K. Gopal, D. Shukla, J. D. Wheelock, and N. Saxena. Hidden reality: caution, your hand gesture inputs in the immersive virtual world are visible to all! In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 859–876, 2023.

[42] S. Hanisch, E. Muschter, A. Hatzipanayioti, S.-C. Li, and T. Strufe. Understanding person identification through gait. *arXiv preprint arXiv:2203.04179*, 2022.

[43] S. Kumawat and H. Nagahara. Privacy-preserving action recognition via motion difference quantization. In *European Conference on Computer Vision*, pages 518–534. Springer, 2022.

[44] K. Lebeck, K. Ruth, T. Kohno, and F. Roesner. Towards security and privacy for multi-user augmented reality: Foundations with end users. In *2018 IEEE Symposium on Security and Privacy (SP)*, pages 392–408. IEEE, 2018.

[45] R. LLC. Mixed Reality Market is Set to Double by 2030 — Mixed Reality Market Share, Size, Growth, Forecast, 2023-2030 — RationalStat Study. `https://www.globenewswire.com/news-release/2023/09/22/2747789/0/en/Mixed-Reality-Market-is-Set-to-Double-by-2030-Mixed-Reality-Market-Share-Size-Growt html`, 2023.

[46] L. Lonini, Y. Moon, K. Embry, R. J. Cotton, K. McKenzie, S. Jenz, and A. Jayaraman. Video-based pose estimation for gait analysis in stroke survivors during clinical assessments: a proof-of-concept study. *Digital Biomarkers*, 6(1):9–18, 2022.

[47] Meta. Introducing project aria from meta, 2024. [Online; accessed 21-Feb-2024].

[48] M. Muaaz and R. Mayrhofer. Orientation independent cell phone based gait authentication. In *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*, pages 161–164, 2014.

[49] Musculoskeletal Key. Gait, 2024. [Online; accessed 28-Feb-2024].

[50] A. Nandy, S. Chakraborty, J. Chakraborty, and G. Venture. *Modern methods for affordable clinical gait analysis: theories and applications in healthcare systems*. Academic Press, 2021.

[51] I. P. Pappas, M. R. Popovic, T. Keller, V. Dietz, and M. Morari. A reliable gait phase detection system. *IEEE Transactions on neural systems and rehabilitation engineering*, 9(2):113–125, 2001.

[52] A. Pfister, A. M. West, S. Bronner, and J. A. Noah. Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis. *Journal of medical engineering & technology*, 38(5):274–280, 2014.

[53] H. Prasanth, M. Caban, U. Keller, G. Courtine, A. Ijspeert, H. Vallery, and J. Von Zitzewitz. Wearable sensor-based real-time gait detection: A systematic review. *Sensors*, 21(8):2727, 2021.

[54] D. Reilly, M. Salimian, B. MacKay, N. Mathiasen, W. K. Edwards, and J. Franz. Secspace: prototyping usable privacy and security for mixed reality collaborative environments. In *Proceedings of the 2014 ACM SIGCHI symposium on Engineering interactive computing systems*, pages 273–282, 2014.

[55] K. Ruth, T. Kohno, and F. Roesner. Secure {Multi-User} content sharing for augmented reality applications. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 141–158, 2019.

[56] Z. Shen, S. Fan, Y. Wong, T.-T. Ng, and M. Kankanhalli. Human-imperceptible privacy protection against machines. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1119–1128, 2019.

[57] W. Sirichotedumrong and H. Kiya. A gan-based image transformation scheme for privacy-preserving deep neural networks. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 745–749. IEEE, 2021.

[58] O. Sofuwa, A. Nieuwboer, K. Desloovere, A.-M. Willems, F. Chavret, and I. Jonkers. Quantitative gait analysis in parkinson's disease: comparison with a healthy control group. *Archives of physical medicine and rehabilitation*, 86(5):1007–1013, 2005.

[59] J. Stenum, C. Rossi, and R. T. Roemmich. Two-dimensional video-based analysis of human gait using pose estimation. *PLoS computational biology*, 17(4):e1008935, 2021.

[60] D. Tome, T. Alldieck, P. Peluse, G. Pons-Moll, L. Agapito, H. Badino, and F. de la Torre. SelfPose: 3D Egocentric Pose Estimation From a Headset Mounted Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):6794–6806, June 2023.

[61] C. L. Vaughan. Theories of bipedal walking: an odyssey. *Journal of biomechanics*, 36(4):513–523, 2003.

[62] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu. A study on gait-based gender classification. *IEEE Transactions on image processing*, 18(8):1905–1910, 2009.

[63] M. Zajac, K. Zołna, N. Rostamzadeh, and P. O. Pinheiro. Adversarial framing for image and video classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 10077–10078, 2019.

[64] D. Zhang, Y. Wang, and B. Bhanu. Ethnicity classification based on gait using multi-view fusion. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 108–115. IEEE, 2010.

[65] Y. Zhang, C. Slocum, J. Chen, and N. Abu-Ghazaleh. It's all in your head (set): Side-channel attacks on ar/vr systems. In *USENIX Security*, 2023.

[66] Y. Zhou, R. Romijnders, C. Hansen, J. v. Campen, W. Maetzler, T. Hortobágyi, and C. J. Lamoth. The detection of age groups by dynamic gait outcomes using machine learning approaches. *Scientific reports*, 10(1):4426, 2020.

Appendix A

A.1 Gait Features

Figure A.1 shows a visual representation of the different gait features, including step time, stance time, swing time, and double support time, that form the gait cycle [49, 50].

A.2 On-Device Implementation of GaitGuard

Proposed System Implementation The first core functionality of **GaitGuard**, which is identifying the lower body area in the frame, suggests that there is a need for a real-time and on-device solution to detect the lower body of a person. As discussed in section 6.1, the proposed implementation of this functionality is to use OpenCV's HOG default person detector and SVM classifier. To implement OpenCV on the HoloLens, we exploited OpenCV library that Microsoft provides as part of their HoloLens2ForCV project [14]. The third-party application containing the core functionality of **GaitGuard** was completed in `C++`. An overview of this on-device implementation can be seen in Figure A.2.

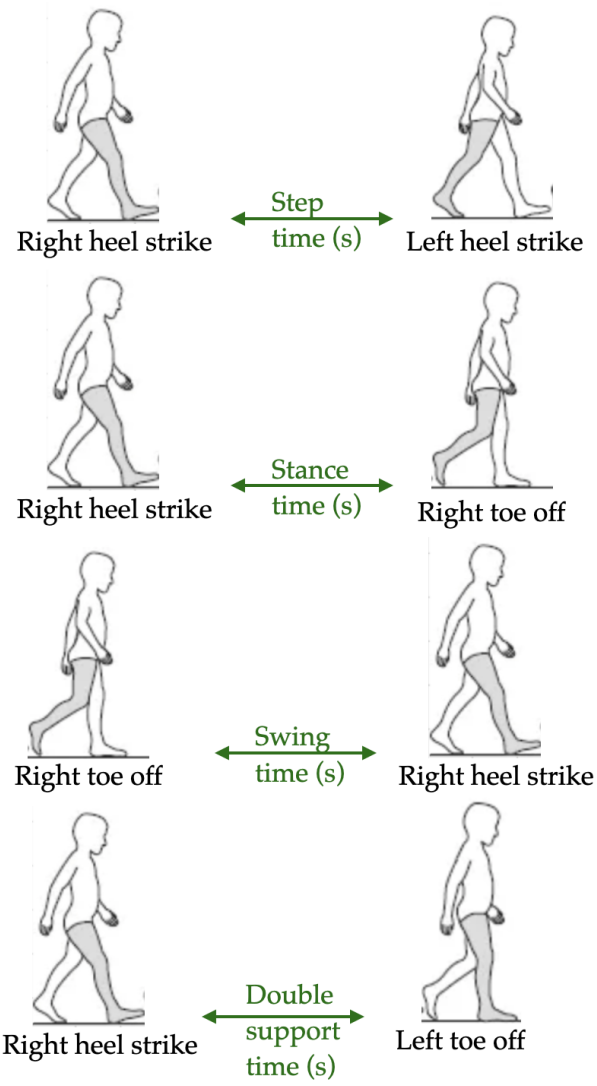


Figure A.1: Different gait features that form the gait cycle referenced with reference to the right leg [49].

Evaluation The suggested application frame rate[1] for the Hololens 2 is 60 fps to ensure the best quality in user experiences [7]. Moreover, it has been reported that the capability of streaming and recording frames from the RGB camera in Hololens 2 containing virtual objects, referred to as Mixed Reality Capture (MRC), has an average frame rate of 30 fps [6].

To assess the ability to deploy **GaitGuard** on Hololens 2, we compare its effect on the recommended frame rate values, which are 60 fps for the application frame rate and 30 fps

[1]Application frame rate is a metric that Microsoft uses to qualify the quality of their applications [4].

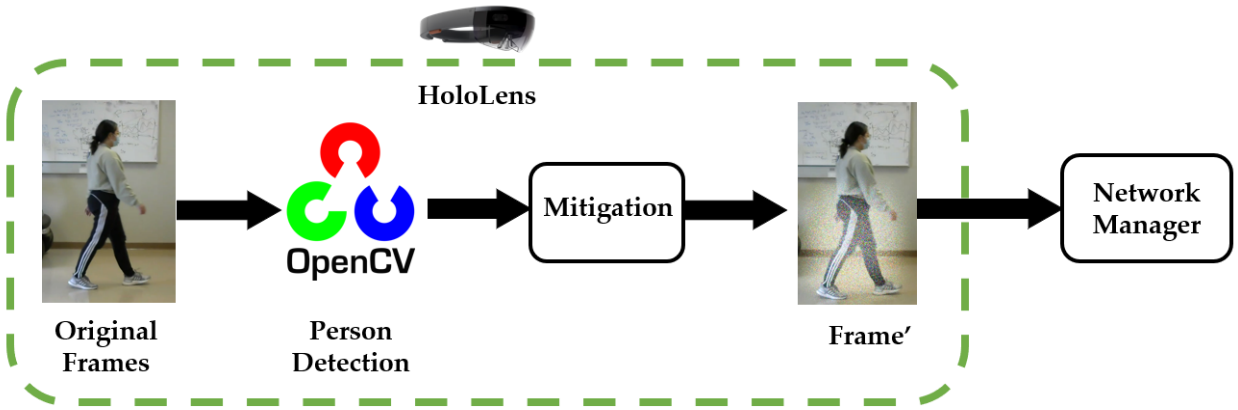


Figure A.2: On device implementation of **GaitGuard**. Raw camera frames are captured from the device's camera. OpenCV is used for person detection on the captured frame. Afterwards, noise is applied to the lower half of the detected person, and this noisy frame is released to the network manager.

for the streaming frame rate.

The application frame rate of the on-device implementation of **GaitGuard** is 60 fps. However, to achieve this recommended 60 fps, the rate at which the perturbation was applied to the RGB camera frame had to be 0.2 fps. This means the on-device implementation of **GaitGuard** could only apply mitigation to one frame every 5 seconds. The application frame rate was unaffected because the **GaitGuard** loop was allocated to run on a background thread, not interfering with the application's performance. This implementation was motivated by the need for the application fps to be 60 to meet Microsoft's best performance standard. However, the rate at which **GaitGuard** applies noise is completely unacceptable because multiple steps could have already been completed within these 5 seconds, leaving users vulnerable to attacks. This high latency is because person detection algorithms are computationally intensive and do not perform well on a resource-constrained device such as the HoloLens 2.

Insights The implementation of **GaitGuard** on-device showed impractical due to constrained resources. However, this effectively means that on-device gait leak attacks based on

video-based approaches, such as pose estimation or pedestrian detection, are also impractical. Therefore, with the current state of Mixed Reality technology, there is no practical way of gathering gait information using video-based approaches on-device because of the computationally intensive nature of those algorithms.

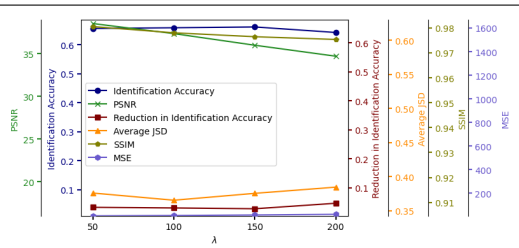
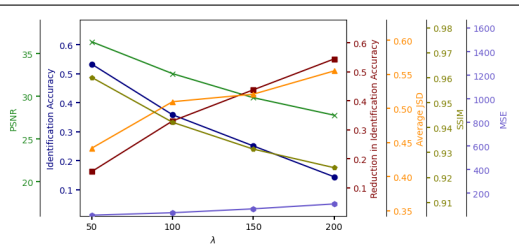
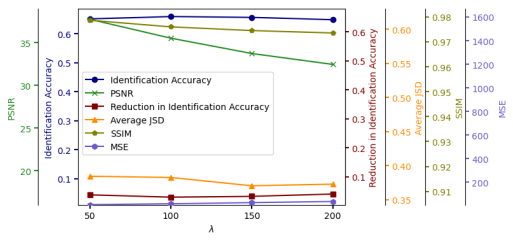
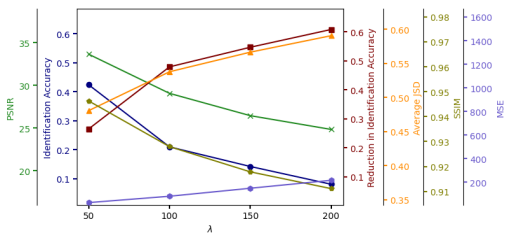
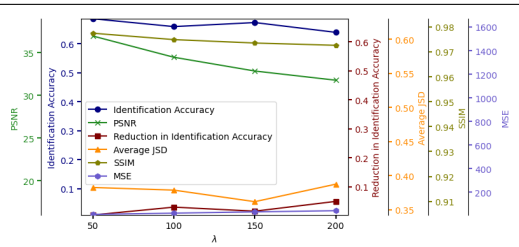
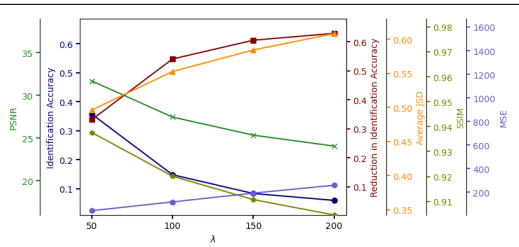
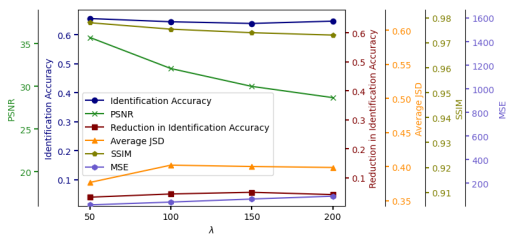
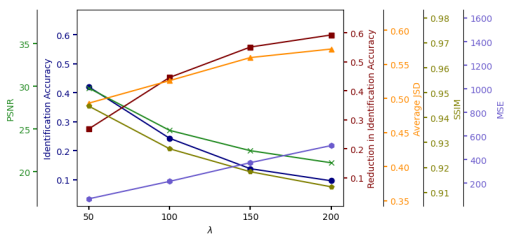
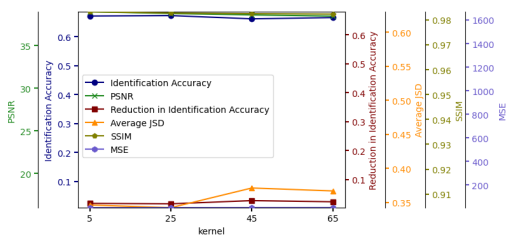
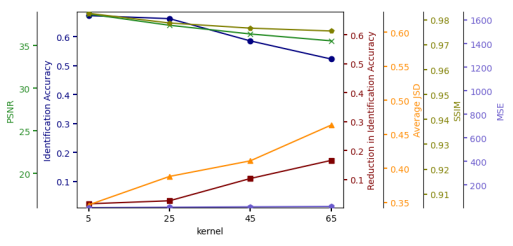
A.3 Privacy and utility metrics across various perturbations

Table A.1 shows the privacy and utility metrics across all mitigation approaches and perturbations. The left column presents the values for the KPM approach, while the right column presents the values for the LBM approach.

Targeted Keypoint Mitigation (KPM) The KPM approach results show no relationship between λ and the JSD and a reduction in classification accuracy.

Lower Body Mitigation (LBM) The results of using LBM approach can be seen in the second column of Table A.1. There is a positive correlation between λ and the average JSD values for all the noise distributions. Overall, the LBM approach with Laplace noise distribution had the most significant reduction in classification accuracy for all λ with a maximum reduction of 62% in classification accuracy. Additionally, the experiments show that the improvement in privacy mitigation after $\lambda = 150$ is minimal.

Table A.1: Privacy-Utility metrics of each Perturbation and Granularity for the Gait Features without Step Length.

	Keypoints-based mitigation approach (KPM)	**Lower body-based mitigation approach (LBM)**
Uniform Noise: $\mathcal{N}(-\lambda, \lambda)$		
Normal Noise: $\mathcal{N}(0, \lambda)$		
Laplace Noise: $\mathcal{L}(0, \lambda)$		
Exponential Noise: $\mathcal{E}(\lambda)$		
Blurring		
DP Pixelization