Swap It Like Its Hot: Segmentation-based spoof attacks on eye-tracking images

Anish S. Narkar
anishnarkar@vt.edu
Virginia Tech
Blacksburg, Virginia, USA

Brendan David-John
bmdj@vt.edu
Virginia Tech
Blacksburg, Virginia, USA

ABSTRACT

Video-based eye trackers capture the iris biometric and enable authentication to secure user identity. However, biometric authentication is susceptible to spoofing another user's identity through physical or digital manipulation. The current standard to identify physical spoofing attacks on eye-tracking sensors uses liveness detection. Liveness detection classifies gaze data as real or fake, which is sufficient to detect physical presentation attacks. However, such defenses cannot detect a spoofing attack when real eye image inputs are digitally manipulated to swap the iris pattern of another person. We propose IrisSwap as a novel attack on gaze-based liveness detection. IrisSwap allows attackers to segment and digitally swap in a victim's iris pattern to fool iris authentication. Both offline and online attacks produce gaze data that deceives the current state-of-the-art defense models at rates up to 58% and motivates the need to develop more advanced authentication methods for eye trackers.

CCS CONCEPTS

• **Security and privacy** → **Biometrics**; • **Human-centered computing** → *Ubiquitous and mobile computing*.

KEYWORDS

Security and access systems, Eye Tracking, Iris Recognition

ACM Reference Format:
Anish S. Narkar and Brendan David-John. 2024. Swap It Like Its Hot: Segmentation-based spoof attacks on eye-tracking images. In *2024 Symposium on Eye Tracking Research and Applications (ETRA '24), June 04–07, 2024, Glasgow, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3649902.3653341

1 INTRODUCTION

Presenting falsified biometric samples of authentic users to access a system is referred to as a *presentation* attack. Presentation or spoofing attacks can be performed using biometric traits such as fingerprints, faces, and irises. Iris attacks primarily present physical spoofs using contact lenses [Venkatesh et al. 2019], printouts [Sequeira et al. 2014, 2016] and playing back recorded videos [Raja et al. 2015]. Spoofing attacks can lead to significant harm to individuals. For example, facial authentication for phones and laptops enables adversaries to access sensitive information, including bank accounts [Jarrahi 2021]. The significance of such harms is clear within intimate partner violence, in which an adversarial but authenticated user can monitor activity, impersonate a victim, or install spyware [Chatterjee et al. 2018; Freed et al. 2018]. While face biometrics are more common on mobile devices today, iris biometrics are seeing increased use in mixed-reality where unobstructed face images are not available. The Microsoft HoloLens2, Magic Leap 2, and Apple Vision Pro enable iris authentication through their integrated eye-tracking sensors [Apple 2023; Leap 2023; Staff 2019].

Current iris spoof detection methods use gaze estimation to detect the presentation of a physical spoof [Czajka and Bowyer 2018]. The state-of-the-art defense method extracts gaze velocity signals for *liveness detection*, i.e., whether the eye tracking inputs are from a real eye or a physical spoof [Raju et al. 2022]. To our knowledge, a real-time iris presentation attack that digitally manipulates real eye images to beat liveness detection has not been demonstrated. In this paper, we propose and demonstrate that IrisSwap can be used to deceive gaze-based liveness detection models and successfully spoof the iris biometric. Our primary contributions are a pipeline for digital spoof manipulations in eye images (Sec. 3.2) and an evaluation of the attack against state-of-the-art liveness detection in both offline (58% user-level attack success rate) and online (55% user-level attack success rate) eye-tracking pipelines (Sec. 4.4).

2 BACKGROUND AND MOTIVATION

Biometric authentication is a secure method to prevent unauthorized access to systems due to their universality, distinctiveness, permanence, and collectability [Jain et al. 2004]. Liveness detection is applied to determine if the presented biometric credentials originated from an authentic source. For face images, depth sensors on mobile phones are used to determine if the presented face is a 3D face or a 2D printout before authenticating the user [Garud and Agrwal 2016]. For eye images containing the iris, which is a gold standard biometric, physical presentation attacks include contact lenses [Venkatesh et al. 2019], printouts [Sequeira et al. 2014, 2016] and displaying videos [Raja et al. 2015]. Textured contact lenses and high-quality printouts require expertise and specific equipment. This requirement limits the breadth of such attacks and lowers the impact of the threat posed by such attacks. For contact lenses, techniques that process the sensor image to identify features introduced by contacts are effective in detecting textured spoof attacks [Doyle and Bowyer 2015], while video playback attacks can be detected through Eulerian video magnification with as low as eleven video frames [Raja et al. 2015]. A digital manipulation attack does not

ETRA '24, June 04–07, 2024, Glasgow, United Kingdom

ACM ISBN 979-8-4007-0607-3/24/06
https://doi.org/10.1145/3649902.3653341

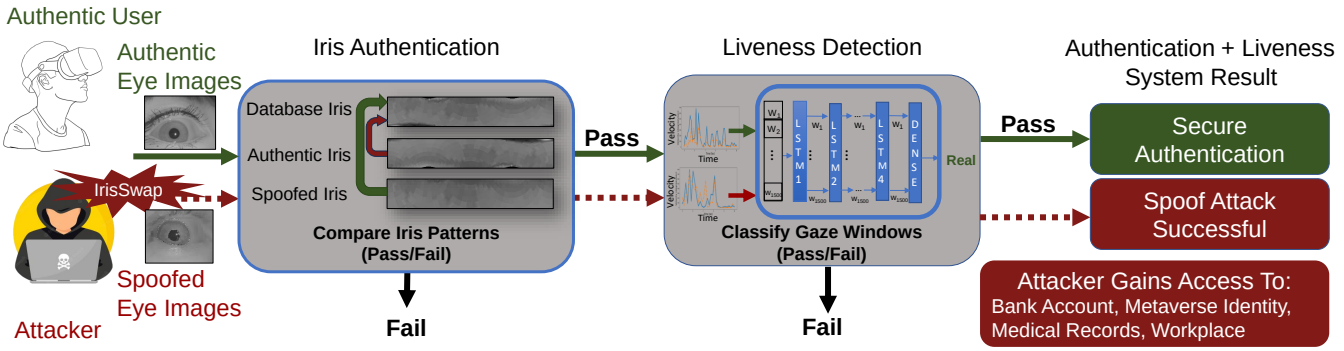

Figure 1: Illustration of the spoofing attack pipeline for iris patterns with gaze-based liveness detection.

require special equipment and can be performed if you only have the iris pattern of the victim.

Figure 1 demonstrates a typical iris authentication and liveness detection mechanism. The attacker presents spoofed eye images (generated using contact lenses, printouts, physical video playbacks, or digital manipulation) to the system. The system first performs the iris authentication and then performs a liveness detection. Access is granted only if both these checks are satisfied.

The standard liveness detection defense for eye images performs gaze estimation and classifies real gaze signals based on micro-saccades and fixations [Rigas and Komogortsev 2014], ocular plant characteristics [Komogortsev et al. 2010] and velocity profiles [Komogortsev et al. 2015; Raju et al. 2022]. These defense mechanisms are successful against impersonation using static presentations such as eye patches. Spoof eye patches cover most of the attacker's eye and strongly impacts the gaze estimation signal enabling detection by liveness models trained on gaze velocity. However, a digitally manipulated image would replace the iris with the victim pattern without restricting their eye movements. Thus, the success of digital manipulation attacks is based on the effectiveness of the iris swapping method.

Raju et al. [Raju et al. 2022] present the latest work on eye-tracking liveness detection using gaze velocity signals, as illustrated in Figure 1. The authors trained a machine-learning model on horizontal and vertical components of gaze velocity to determine real or spoof. The authors evaluated against physical spoofing attacks in the form of an iris pattern printed on paper and worn on an eye patch. The proposed method reliably detected spoof attacks from the eye patch with an accuracy of 98%. The liveness detection approach reliably detected attacks, as the eye patch produced a gaze signal that was clearly not from a real eye. Thus, an attack with the ability to spoof an iris pattern without impacting gaze estimation poses a risk for current liveness detection-based defenses.

Chaudhary et al. [Chaudhary and Pelz 2020] demonstrated high-accuracy swapping of the iris pattern using deep segmentation networks. Their work presented a privacy solution by removing the user's iris pattern from recorded eye-tracking data. In contrast, our work instead overlays the iris pattern of a victim to perform a presentation attack. Through accurate and efficient swapping of the iris pattern, authentication can be spoofed without impacting the gaze signal.

3 ATTACK MODEL AND METHODOLOGY

The IrisSwap attack has several assumptions and necessary definitions. The *challenge* refers to the task that will be analyzed to determine the liveness of the input data [Li et al. 2022]. The *attacker* is the person who is carrying out the attack and the *victim* is the person whose iris image will be spoofed by the attacker. We assume that the attacker possesses an iris image of the victim. The attacker applies the spoof to a recording of their own eye-tracking images as they perform the challenge task for *offline attacks* or make use of software plugins to manipulate their eye image stream in real-time for *online attacks*. The attacker is successful if their sequence of eye images pass both the liveness detection classification and iris authentication as the victim. We set our assumptions for data access based on current standards for liveness detection and iris authentication, meaning the spoof detection model is limited to processing only gaze estimates produced by the manipulated inputs while the authentication module is limited to the eye image with the manipulation applied. While a spoof detection approach could observe the modified eye images to detect manipulations, that is outside the scope of our work and requires the design of a novel defense method and evaluation.

3.1 Design Challenges

3.1.1 Challenge 1: Accurate segmentation and efficient swapping. A successful attack will accurately segment and replace the attacker's iris pattern with the victim's iris. The swap should take into account any physical differences between eye images, i.e., pupil diameter and eye orientation. For online attacks, the processing pipeline must execute fast enough to support gaze estimation during the challenge task.

3.1.2 Challenge 2: Minimal impact on gaze estimation. A successful attack will not cause significant deviation between the gaze estimates of swapped and unmodified images as they are used for liveness classification. Accuracy and precision of gaze data are computed to measure the difference between swapped and unmodified inputs and liveness detection models are re-trained on for each experiment to evaluate whether the manipulation is detected within the gaze signal.

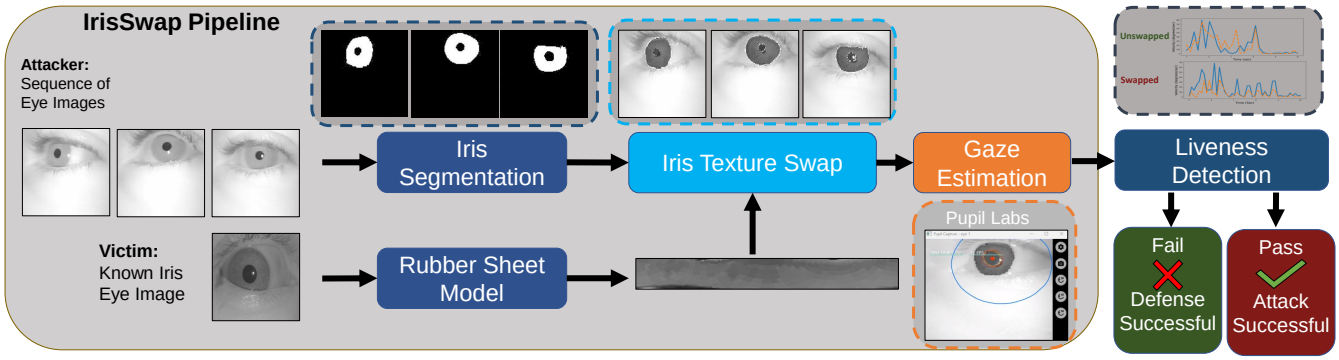

Figure 2: IrisSwap pipeline showing the flow of data through iris segmentation and swapping before gaze estimation is applied. The output gaze positions are processed into windows of gaze velocity that are classified by a liveness detection model.

3.2 IrisSwap System Design

Inputs to our pipeline are a sequence of eye images from the attacker and the iris pattern of a victim (Fig. 2). IrisSwap replaces the iris pattern of the attacker with that of the victim and generates gaze positions from the modified eye images. Each stage of the pipeline requires independent models for data processing.

3.2.1 Iris Segmentation. Iris segmentation is applied to raw grayscale images of the eye. We adopted the RITNet segmentation model based on high performance in the rubber sheet iris protection work [Chaudhary and Pelz 2020]. However, when testing with RITNet we observed that segmented images contained stray iris class pixels outside of the iris region. We instead considered a double UNET architecture to obtain finer and more accurate segmentations [Jha et al. 2020]. RITNet is based on a single UNET model, and it was not feasible to duplicate RITNet twice given the size of the RITNet multi-stage pipeline. Thus, we implemented a new segmentation model based on the double UNET architecture called Shallow-net. Please see the Supplementary Material for a complete description of the Shallow-net architecture.

3.2.2 Iris Texture Swap. Iris swapping uses the segmentation mask from the attacker's eye image to replace the iris with that of the victim to spoof their identity. The inverse rubber sheet model transforms an iris pattern in polar coordinates and accounts for varying shape and orientation by leveraging inner and outer radius of the target segmented region [Chaudhary and Pelz 2020]. A more detailed description of how this process is implemented can be found in the original paper [Chaudhary and Pelz 2020].

3.2.3 Gaze Estimation. Gaze positions are estimated using the Pupil Labs hardware and software suite (v3.5.1) and the standard four-point validation procedure to compute gaze accuracy and precision metrics [Kassner et al. 2014]. We used the Pupil Labs 3D landmark-based gaze model with default parameters to generate gaze positions which are then processed into horizontal and vertical velocity components for liveness detection.

3.2.4 Challenge Task. The challenge task in Raju et al. [Raju et al. 2022] had users observe a single static point on a screen. We introduced a challenge task with multiple targets to increase the difficulty in passing liveness detection using velocity-based models. The Pupil Labs calibration and validation routines were used as the challenge tasks they feature five and four target positions, respectively. Furthermore, using the calibration step for the challenge allows liveness detection to be organically integrated into the eye tracker setup without an additional step.

3.2.5 System Deployment. Python code for segmentation and swapping was implemented within the Pupil Player for offline attacks on previously recorded data and was implemented in Pupil Capture as a plug-in to apply online attacks.

4 DATASETS AND EXPERIMENTS

To evaluate the IrisSwap pipeline we conducted experiments on offline and online attacks. Implementing the pipeline required training an iris segmentation model for Pupil Labs eye images and training a liveness detection classifier based on Pupil Labs gaze signals. In Section 4.1, we define the challenge task and dataset for our experiments. In Section 4.2, we describe the optimization and performance for each component of IrisSwap. In Section 4.3, we define the metrics to evaluate IrisSwap in terms of gaze estimation and attack success rate, then present our evaluation results in Section 4.4.

4.1 Datasets

Offline Attack Dataset. For offline analysis, we used a dataset of eye-tracking recordings previously evaluated in the context of iris authentication provided by John et al. [John et al. 2020]. This dataset contains data from 15 subjects performing a five-point calibration task and validation using the Pupil Labs Pro (2016) eye tracker at 30Hz. We chose this dataset as the data aligned with the challenge task and could be re-analyzed with the Pupil Labs software. Subjects sat in front of a computer monitor and were presented with one calibration point at a time for four seconds.

Online Attack Dataset. To evaluate the real-time performance of IrisSwap we collected a new dataset using a Pupil Labs Core (2022) eye tracker with an IRB-approved user study with 18 subjects (13M, 5F, Avg. Age = 27.2 ± 1.2 years). Subjects sat approximately one meter in front of a computer monitor and were presented with the challenge task (Sec. 4.1), with each target shown for four to six seconds. The study consisted of two conditions, swapped and

unswapped based on whether IrisSwap was enabled or not and were counterbalanced across subjects.

Iris Segmentation. We used 2600 ground truth images from CASIA [Alonso-Fernandez 2015] to train our initial iris segmentation model (Sec. 3.2.1). We then isolated data from two subjects of the offline dataset to fine-tune a model for offline analysis. Data from these two subjects were withheld from further evaluations against liveness detection. We also performed a pilot study to test the online attack using the Pupil Labs Core eye tracker. We labeled the iris images from this data to fine-tune a separate segmentation model for the online attacks. This ensured that the segmentation results were consistent for the iris sequences from each Pupil Labs eye tracker. We manually segmented a total of 1000 eye images from both the offline (750) and pilot dataset (250) for fine-tuning. For fine-tuning each model, 70% of the images were used for training, 15% for validation, and the remaining 15% images were used to evaluate model performance.

Liveness Detection. To train and evaluate the liveness detection models we assigned subjects to either the training or the test set. The training set consisted of data from 60% of the subjects (offline: 7, online: 11) and the test set consisted of the remaining 40% subjects (offline: 6, online: 7). While training we further divided the training set into a validation set. The validation set consisted of 30% of randomly selected subjects from the training set. The instantaneous horizontal and vertical velocity signals from each user were computed based on the gaze estimates produced by Pupil Labs. Velocity components over 800 $^\circ/sec$ were removed and replaced via interpolation [Dowiasch et al. 2015]. Min-max normalization between zero and one was applied to the velocity signals for each user to avoid any data leakage.

4.2 IrisSwap Component Testing and Metrics

Iris Segmentation. As described in **Challenge 1**, the effectiveness of IrisSwap relies on accurate and efficient iris segmentation. We compared segmentation results using RitNet and our Shallow-net model using the Dice score metric [Carass et al. 2020]. The Dice score ranges from zero to one and measures the degree of overlap between predicted and ground truth while accounting for class imbalance and is analogous to the F1 score for segmentation tasks. The computed Dice score for RITNet on the testing set was 0.91, while the Dice score for Shallow-net was 0.96. Based on these results, we used Shallow-net to evaluate and deploy IrisSwap.

Liveness Detection. A successful IrisSwap attack is determined by its ability to swap irises without producing gaze estimates flagged as fake by the liveness model. The liveness detection model used by Raju et al [Raju et al. 2022] was based on a customized ResNet and was developed using a dataset that was collected on a 1000Hz device. Due to the lower sampling rate in our data, it was not possible to use their exact architecture or model weights. A pilot study with the online IrisSwap attack identified a drop in sampling rate as low as 3Hz. To avoid issues with oversampling online gaze samples during deployment, we decided to downsample all data to 3Hz for the liveness detection. We optimized their proposed architecture on our data and compared performance with LSTM models following their training/testing protocol using window length and step size as the hyperparameters. The full description of this optimization and comparison is described in the Supplementary Material. We found our LSTM models with a window length of seven samples (2.33 seconds) and step size of three performed best with an attack classification rate of 99.9% and deployed them for analysis.

Iris Authentication. A successful IrisSwap attack must pass a standard iris authentication for the victim's iris. For offline analysis, subject S013 was randomly chosen as the victim of the attack. For the online attack, we needed to assign the victim iris before the study due to the real-time nature of the attack. An iris pattern from one of the authors using the Pupil Labs Core was considered the victim of the online attack. The data from these subjects were excluded for training and development of all the components of the IrisSwap pipeline and the liveness detection model. The standard metric for iris authentication is Hamming Distance (HD) and is computed between two iris biometric templates [Daugman 2004]. HD measures the number of bits that are different between the biometric templates. The phase-based Daugman method using 2D Gabor wavelets was used for extracting binary iris templates. A lower HD indicates a closer match, with HD less than 0.37 being considered a sufficient condition for genuine authentication from eye tracker images [John et al. 2020]. We selected ten random frames from each of the spoofed output sequences and compared them with the iris template from the corresponding frame number of the victim's eye-tracking sequence.

4.3 Metrics

Gaze Estimation. As described in **Challenge 2**, IrisSwap should not have a significant effect on the accuracy and precision of the gaze signal. Accuracy is the average deviation between the gaze positions obtained by the eye-tracker and the location of the validation targets. Precision is the Root Mean Squared angular distance between successive samples. Both these metrics are computed on the validation targets of the challenge task using the Pupil Player software.

Attack Success Rate. Attack Success Rate (ASR) is used to measure the performance of our attacks. ASR measures the proportion of spoofed samples that were misidentified by the liveness models. ASR for a perfect attack would have a value of one, while for a perfect defense, it would be zero. To calculate ASR we first feed the velocity windows to the liveness detection model. ASR is calculated on two levels. First, at the window level, ASR is calculated based on individual window predictions for all the users. Second, on the user level ASR was calculated based on model prediction for the entire user sample data. The model predictions of the majority of window samples for each user is assigned as the model prediction for that user. The ASR metrics are defined as

$$ASR_{window} = \frac{Spoofed\ windows\ classified\ as\ Real}{Total\ \#\ of\ \ Spoofed\ windows}$$

$$ASR_{user} = \frac{Spoofed\ users\ classified\ as\ Real}{Total\ \#\ of\ \ users}.$$

The reliability of attack success depends on the random train-test split of the subjects. We generated ten random train-test splits to

train the liveness detection model and evaluate the resulting variance. The reported mean ASR across all splits is used to demonstrate the effectiveness of the IrisSwap pipeline.

4.4 Attack Results

We evaluated the offline and online impact of IrisSwap on gaze estimation, computed the ASR against an optimal liveness detection model, and confirmed that spoofed eye images pass iris authentication for the victim.

4.4.1 IrisSwap Impact on Iris Authentication. The rightmost column of Table 1 shows the result of the iris authentication of the selected victim. The average HD of the spoofed samples for offline and online scenarios was 0.36 and 0.39, respectively. The HD values for each subject in the offline and online evaluations are listed in the Supplementary. The mean HD for offline attacks was marginally under the 0.37 threshold which is a commonly accepted limit and aligns with past eye-tracker authentication studies [Daugman 2004; John et al. 2020, 2019]. The victim was authenticated for ten out of the thirteen subjects in offline attacks. The mean HD for online attacks was 0.39 which was marginally above the threshold; however, the victim was authenticated for thirteen out of the eighteen subjects. Outlier values of HD (0.45, 0.53, and 0.57) for three of the subjects increased the overall mean HD for online attacks. Thus, for offline and online attacks the victim identity was successfully spoofed for 77% and 72% of the users, respectively.

4.4.2 Eye-Tracking Data Quality. Figure 3 visualizes the impact of IrisSwap on the velocity signals extracted from the gaze estimates, and Table 1 presents the impact on gaze estimation metrics. The difference between the mean accuracy of spoofed and unmodified data for offline and online attacks are 0.83° and 1.7°, respectively. For the offline attacks sampling rate was preserved, and manipulations impacted the pupil landmark tracking used for gaze estimation producing less than a degree of error in terms of accuracy. For the online dataset, the sampling rate was reduced by a factor of 8.9 ± 2.6 on average when IrisSwap was enabled compared to the unswapped condition. The lowest sampling rate produced was 3Hz. The reduced sampling rate resulted in fewer samples for Pupil Labs to perform an accurate calibration and for computing the validation metrics; however, it still provided enough data to enable the Pupil Labs calibration and our challenge task. Despite the introduced spatial error, the gaze velocity peaked in similar regions (between 5 and 7 seconds) as shown in Figure 3. However, an impact from the reduced sampling rate can be observed between 7 and 9 seconds. These differences in the velocity profile are detected by the liveness models.

4.4.3 IrisSwap Effectiveness. Figure 4 shows the attack effectiveness of the IrisSwap pipeline. The average ASR_{window} for offline and online scenarios is 0.61 and 0.58, respectively. The average ASR_{user} for offline and online attacks is 0.59 and 0.55, respectively. These results indicate that the IrisSwap pipeline is marginally more successful in offline scenarios and in both cases just under half the attacks were detected by liveness detection. Based on the increase in gaze error and visualization of velocity profiles, we conclude that ASRs below 100% can be attributed to lower sampling frequencies and the image manipulation degrading gaze estimation accuracy. Still, an ASR above 50% poses a risk if attacks were mounted on a large scale.

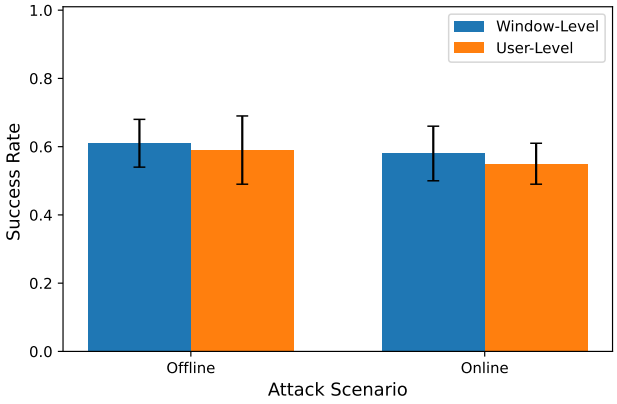

Figure 4: Window-level ASR indicates the success rate calculated based on each window of the test-user data. Window-level ASR is 0.61 ± 0.07 and 0.58 ± 0.08 for offline and online attacks, respectively. User-level ASR makes one real or spoof classification based on all windows from a single user. User-level ASR is 0.59 ± 0.10 and 0.55 ± 0.06 for offline and online attacks, respectively.

5 CONCLUSION AND DISCUSSION

Current state-of-the-art defenses against iris presentation attacks employ a gaze velocity-based approach. Our IrisSwap pipeline digitally manipulates an attacker's iris to match that of a victim while enabling gaze estimation. Our method successfully attacked a state-of-the-art liveness detection model architecture at a rate of 59% and 55% at a user level in offline and online attacks, respectively. Our attack shows that the current standard for gaze-based liveness detection cannot reliably detect presentation attacks produced by digital manipulations. The IrisSwap pipeline demonstrates a new class of digital presentation attacks that differ from the playback attacks used in prior works. The real-time nature of IrisSwap and a success rate over 50% indicates a practical security concern if such an attack is optimized and mounted on a large scale.

Implications. The proposed IrisSwap attack is the first system using digital manipulation to spoof iris-based authentication in an eye-tracking system successfully. Our attacks were successful 58% of the time, meaning they were neither perfect nor benign. The threat we identified can still be mitigated as we expect future solutions in the field to detect IrisSwap and similar digital manipulations. For example, swapped eye images may have spurious white pixels or the victim iris pattern may have clear differences in grayscale intensity from the original eye (see Figure 2, Iris Texture Swap). A defense model that processes the swapped eye images could detect such artifacts. Our results also suggest that one of the reasons an attack is not successful is that the reduced sampling rate affects the accuracy of calibration and the consistency of velocity profiles leading to spoof detection. If our attack code was integrated in a more optimal manner, and not through a patch-like approach

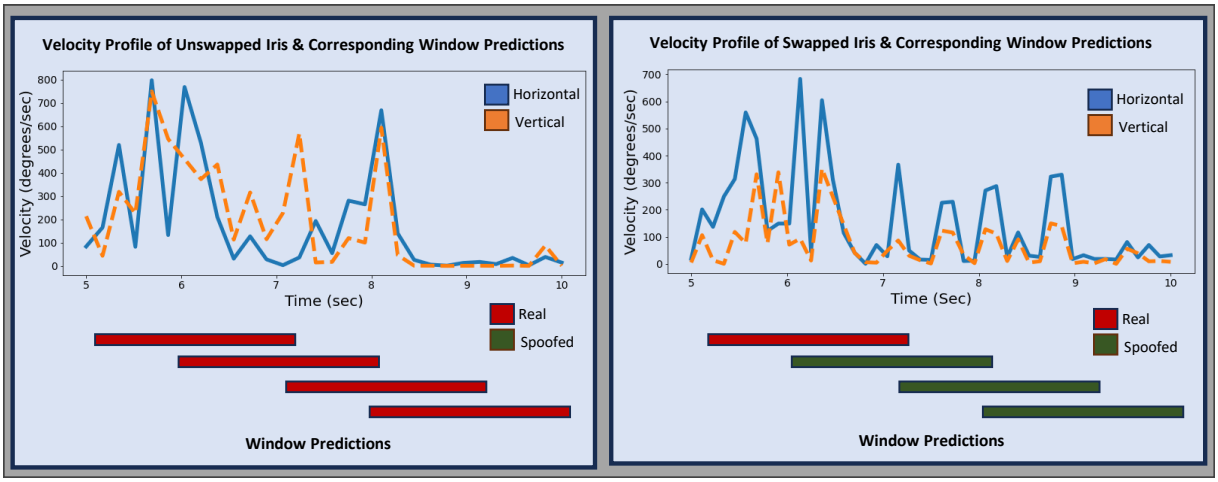

Figure 3: Velocity profiles for Unswapped (Left) and Swapped (Right) samples over five seconds of an online attack. A similar profile is produced between 5s and 7s while the differences between 7s and 9s are flagged as spoofs by the liveness model.

Table 1: Impact of IrisSwap on eye-tracking data quality across subjects for online and offline subjects. IrisSwap only introduced 0.83° of gaze error on average for offline condition but introduced a gaze error of 1.7° for real-time application. The rightmost column shows the mean HD for all the train-test splits.

Application	Unswapped Accuracy (degrees)	Swapped Accuracy (degrees)	Unswapped Precision (degrees)	Swapped Precision (degrees)	Hamming Distance
Offline	0.56 ± 0.33	1.39 ± 0.40	0.12 ± 0.03	0.12 ± 0.03	0.36 ± 0.02
Real-time	1.91 ± 0.65	3.56 ± 0.48	0.13 ± 0.05	0.25 ± 0.05	0.39 ± 0.06

using Python plug-ins at runtime, then we expect the attack to have a higher success rate.

Limitations. Our liveness detection evaluation considered eye movement data from a head-worn eye tracker downsampled to 3Hz, as opposed to an EyeLink desktop tracker at 1000Hz in existing work on liveness detection [Raju et al. 2022]. Eye-tracking systems with more stable calibration models or higher sampling rates may produce different results in terms of successfully detecting attacks. Our findings on HD and iris authentication depend on infrared images captured by the Pupil Labs system at 320 × 240. While this resolution is sufficient for authentication [John et al. 2019; Phillips and Komogortsev 2011], higher-resolution images would make for more convincing spoof attacks. Our segmentation model relied on the binary classification of which pixels were the iris region, though a multi-class approach may be more accurate and allow for a smoother application of the inverse rubber sheet.

Future Work. First, we plan to implement and evaluate a suite of defense measures that could include embedding the output of the eye camera with a digital watermark [Plata and Syga 2020], training a classifier to detect visual artifacts resulting from IrisSwap, or the integration of peri-ocular biometrics that feature eye shape and brows [Woodard et al. 2010]. Second, now that there is proof that iris swaps can beat liveness detection, we expect to further study new attack models. For example, GANs or Variational Autoencoders commonly used in deep fakes can be integrated to generate high-resolution eye images that retain gaze direction with a swapped iris in one step, without relying on a specific segmentation model. Balancing online performance with naturalistic-looking spoofs presents a key research challenge in this context to benchmark how effective an iris spoofing attack can become.

Privacy and Ethics. Privacy and ethics considerations are critical as we identify a new security vulnerability in iris authentication systems that use gaze estimation for liveness detection. First, given the scope of a short paper, we have identified and presented results on the feasibility of the attack, but do not present an optimized system or evaluate a corresponding defense mechanism. We noted potential defense mechanisms for the current attack and expected extensions of the system in our discussion. Second, we took great care to protect the privacy of our study subjects who were necessary to enable our experiments. Within the manuscript, we ensured that all eye images used to make figures were outside the quality standards for iris biometrics outlined in ISO/IEC 19794-6:2005 [Formats-Part 2005], including purposely saving images in a compressed JPEG format and ensuring the image was first resized such that the iris diameter spanned less than 200 pixels. The Supplementary video during the attack sequence was heavily compressed as well. Our IRB-approved study provided informed consent on the purpose of the study, the critical nature of securing biometrics, and our data management approach. Our collected research dataset will not be posted publicly but gaze sample data may be shared with validated research teams in which data usage will be moderated by the authors.

REFERENCES

Fernando Alonso-Fernandez. 2015. Near-infrared and visible-light periocular recognition with Gabor features using frequency-adaptive automatic eye detection. *IET Biometrics* 4 (June 2015), 74–89(15). Issue 2. https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2014.0038

Apple. 2023. Apple Vision Pro: Privacy and Security. https://www.apple.com/apple-vision-pro/

Aaron Carass, Snehashis Roy, Adrian Gherman, Jacob C. Reinhold, Andrew Jesson, Tal Arbel, Oskar Maier, Heinz Handels, Mohsen Ghafoorian, Bram Platel, Ariel Birenbaum, Hayit Greenspan, Dzung L. Pham, Ciprian M. Crainiceanu, Peter A. Calabresi, Jerry L. Prince, William R. Gray Roncal, Russell T. Shinohara, and Ipek Oguz. 2020. Evaluating White Matter Lesion Segmentations with Refined Sørensen-Dice Analysis. *Scientific Reports* 10, 1 (May 2020). https://doi.org/10.1038/s41598-020-64803-w

Rahul Chatterjee, Periwinkle Doerfler, Hadas Orgad, Sam Havron, Jackeline Palmer, Diana Freed, Karen Levy, Nicola Dell, Damon McCoy, and Thomas Ristenpart. 2018. The spyware used in intimate partner violence. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 441–458.

Aayush Kumar Chaudhary and Jeff B Pelz. 2020. Privacy-Preserving Eye Videos Using Rubber Sheet Model. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 22, 5 pages. https://doi.org/10.1145/3379156.3391375

Adam Czajka and Kevin W Bowyer. 2018. Presentation attack detection for iris recognition: An assessment of the state-of-the-art. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–35.

J. Daugman. 2004. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology* 14, 1 (2004), 21–30. https://doi.org/10.1109/TCSVT.2003.818350

Stefan Dowiasch, Svenja Marx, Wolfgang Einhäuser, and Frank Bremmer. 2015. Effects of aging on eye movements in the real world. *Frontiers in human neuroscience* 9 (2015), 46.

James S Doyle and Kevin W Bowyer. 2015. Robust detection of textured contact lenses in iris recognition using BSIF. *IEEE Access* 3 (2015), 1672–1683.

Biometric Data Interchange Formats-Part. 2005. 6: Iris image data. *ISO/IEC* (2005), 19794–6.

Diana Freed, Jackeline Palmer, Diana Minchala, Karen Levy, Thomas Ristenpart, and Nicola Dell. 2018. "A Stalker's Paradise" How Intimate Partner Abusers Exploit Technology. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

Dhananjay Garud and S.S. Agrwal. 2016. Face liveness detection. In *2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)*. 789–792. https://doi.org/10.1109/ICACDOT.2016.7877695

Anil K Jain, Arun Ross, and Salil Prabhakar. 2004. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology* 14, 1 (2004), 4–20.

Javad Jarrahi. 2021. IProov face biometrics deployed in OCBC atms in Singapore Pilot: Biometric update. https://www.biometricupdate.com/202103/iproov-face-biometrics-deployed-in-ocbc-atms-in-singapore-pilot

Debesh Jha, Michael A. Riegler, Dag Johansen, Pål Halvorsen, and Håvard D. Johansen. 2020. DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation. https://doi.org/10.48550/ARXIV.2006.04868

Brendan John, Sophie Jörg, Sanjeev Koppal, and Eakta Jain. 2020. The security-utility trade-off for iris authentication and eye animation for social virtual avatars. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 1880–1890.

Brendan John, Sanjeev Koppal, and Eakta Jain. 2019. EyeVEIL: Degrading Iris Authentication in Eye Tracking Headsets. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research &; Applications* (Denver, Colorado) *(ETRA '19)*. Association for Computing Machinery, New York, NY, USA, Article 37, 5 pages. https://doi.org/10.1145/3314111.3319816

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*. 1151–1160.

Oleg V. Komogortsev, Sampath Jayarathna, Cecilia R. Aragon, and Mechehoul Mahmoud. 2010. Biometric Identification via an Oculomotor Plant Mathematical Model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research amp; Applications* (Austin, Texas) *(ETRA '10)*. Association for Computing Machinery, New York, NY, USA, 57–60. https://doi.org/10.1145/1743666.1743679

Oleg V. Komogortsev, Alexey Karpov, and Corey D. Holland. 2015. Attack of Mechanical Replicas: Liveness Detection With Eye Movements. *IEEE Transactions on Information Forensics and Security* 10, 4 (2015), 716–725. https://doi.org/10.1109/TIFS.2015.2405345

Magic Leap. 2023. Magic Leap 2: Iris ID (Beta). https://resources.magicleap.com/en-us/privacy/iris-unlock-id?locale=en-US

Changjiang Li, Li Wang, Shouling Ji, Xuhong Zhang, Zhaohan Xi, Shanqing Guo, and Ting Wang. 2022. Seeing is Living? Rethinking the Security of Facial Liveness Verification in the Deepfake Era. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, Boston, MA, 2673–2690. https://www.usenix.org/conference/usenixsecurity22/presentation/li-changjiang

Clark Phillips and Oleg V Komogortsev. 2011. *Impact of resolution and blur on iris identification*. Technical Report. Technical Report.

Marcin Plata and Piotr Syga. 2020. Robust spatial-spread deep neural image watermarking. In *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 62–70.

Kiran B. Raja, R. Raghavendra, and Christoph Busch. 2015. Video Presentation Attack Detection in Visible Spectrum Iris Recognition Using Magnified Phase Information. *IEEE Transactions on Information Forensics and Security* 10, 10 (2015), 2048–2056. https://doi.org/10.1109/TIFS.2015.2440188

Mehedi Hasan Raju, Dillon J. Lohr, and Oleg Komogortsev. 2022. Iris Print Attack Detection using Eye Movement Signals. *ETRA* (2022), 70:1–70:6. https://doi.org/10.1145/3517031.3532521

Ioannis Rigas and Oleg V. Komogortsev. 2014. Gaze estimation as a framework for iris liveness detection. In *IEEE International Joint Conference on Biometrics*. 1–8. https://doi.org/10.1109/BTAS.2014.6996282

Ana F. Sequeira, Hélder P. Oliveira, João C. Monteiro, João P. Monteiro, and Jaime S. Cardoso. 2014. MobILive 2014 - Mobile Iris Liveness Detection Competition. In *IEEE International Joint Conference on Biometrics*. 1–6. https://doi.org/10.1109/BTAS.2014.6996290

Ana F. Sequeira, Shejin Thavalengal, James Ferryman, Peter Corcoran, and Jaime S. Cardoso. 2016. A realistic evaluation of iris presentation attack detection. In *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*. 660–664. https://doi.org/10.1109/TSP.2016.7760965

Ars Staff. 2019. Microsoft unveils Hololens 2: Twice the field of view, Eye Tracking. https://arstechnica.com/gadgets/2019/02/microsoft-unveils-hololens-2-twice-the-field-of-view-eye-tracking/

Sushma Krupa Venkatesh, Raghavendra Ramachandra, K. Bommanna Raja, and Christoph Busch. 2019. A New Multi-spectral Iris Acquisition Sensor for Biometric Verification and Presentation Attack Detection. *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)* (2019), 47–54.

Damon L Woodard, Shrinivas Pundlik, Philip Miller, Raghavender Jillela, and Arun Ross. 2010. On the fusion of periocular and iris biometrics in non-ideal imagery. In *2010 20th International Conference on Pattern Recognition*. IEEE, 201–204.