Privacy-preserving Reflection Rendering for Augmented Reality

Yiqin Zhao
Worcester Polytechnic Institute
Worcester, MA, USA
yzhao11@wpi.edu

Sheng Wei
Rutgers University
Piscataway, NJ, USA
sheng.wei@rutgers.edu

Tian Guo
Worcester Polytechnic Institute
Worcester, MA, USA
tian@wpi.edu

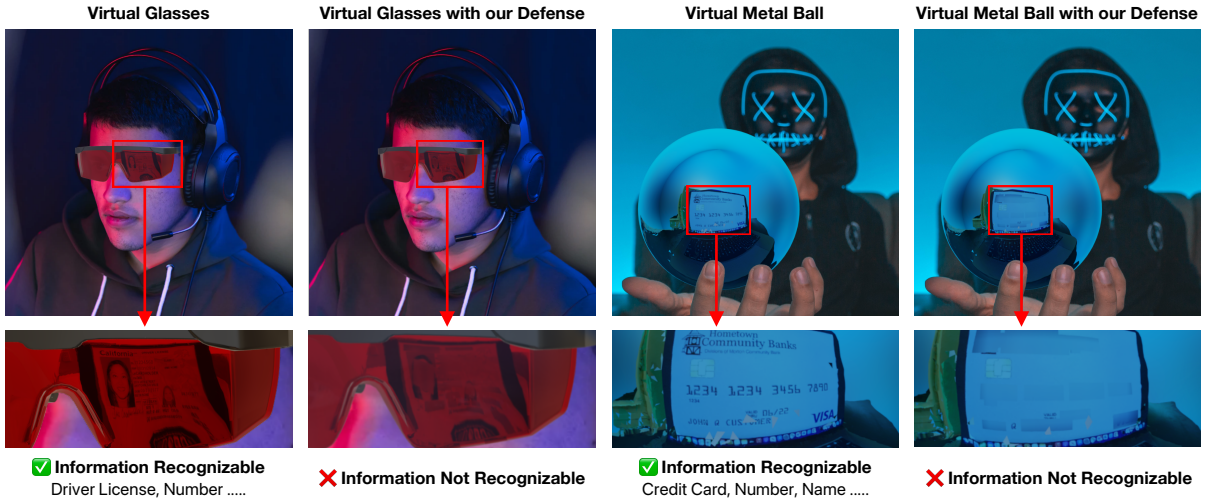


Figure 1: Sensitive information on reflection rendering with/without our defense. In the 1st and 3rd columns, sensitive information from the physical environment can appear as part of the rendered reflections and be leaked to viewers. In the 2nd and 4th columns, we show that our proposed defense can effectively eliminate such information leakage while still keeping high visually coherent reflections.

ABSTRACT

Many augmented reality (AR) applications rely on omnidirectional environment lighting to render photorealistic virtual objects. When the virtual objects consist of reflective materials, the required lighting information to render such objects can consist of privacy-sensitive information outside the current camera view. In this paper, we show, for the first time, that accuracy-driven multi-view environment lighting can reveal out-of-camera scene information and compromise privacy. We present a simple yet effective privacy attack that extracts sensitive scene information such as human faces and text from rendered objects under several application scenarios.

To defend against such attacks, we develop a novel IPC^2S defense and a conditional R^2 defense. Our IPC^2S defense, combined with a generic lighting reconstruction method, preserves the scene geometry while obfuscating the privacy-sensitive information. As a proof-of-concept, we leverage existing OCR and face detection models to identify text and human faces from past camera observations and blur the color pixels associated with detected regions. We evaluate the visual quality impact of our defense by comparing rendered virtual objects to ones rendered with a generic multi-lighting reconstruction technique, ARKit, and R^2 defense. Our visual and quantitative results demonstrate that our defense leads to structurally similar reflections with up to 0.98 SSIM score across various rendering scenarios while preserving sensitive information by reducing the automatic extraction success rate to at most 8.8%.

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3548386

CCS CONCEPTS

• **Security and privacy** → **Systems security**; • **Information systems** → **Multimedia streaming**.

KEYWORDS

Augmented reality; visual privacy; photorealistic rendering

ACM Reference Format:
Yiqin Zhao, Sheng Wei, and Tian Guo. 2022. Privacy-preserving Reflection Rendering for Augmented Reality. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3503161.3548386

1 INTRODUCTION

Augmented reality (AR) has the promise to transform many aspects of our lives, including education [19], healthcare [2], and business [25]. By 2023, mobile AR is predicted to be a hundred-billion dollar market, with hundreds of millions of users [1]. Today, many

popular social media apps such as TikTok and YouTube are increasingly supporting a new form of AR application. This new application scenario, which we refer to as *AR content creation/streaming*, allows users to create videos augmented with 2D/3D assets [45]. The AR content can then be shared via social media platforms. For example, a streamer may engage her community with a virtual sunglasses try-on session where she will try on different sunglasses based on the text chat suggestions from the community.

This engagement-driven AR content creation often calls for good visual effects, which further translates to the high rendering quality of the virtual objects. That is, the virtual objects appearing in the video stream should exhibit *visual coherency* to the physical world background and should be rendered in a photorealistic way. It is desirable for AR applications to have omnidirectional environment lighting to achieve good visual effects. To provide accurate lighting information, existing AR frameworks often require AR users to scan the physical environment and leverage deep learning models to estimate the lighting [17, 23, 59]. Without loss of generality, we refer to this type of lighting estimation approach as *multi-view lighting reconstruction*.

However, environment scanning will capture multiple glimpses (i.e., camera views) of the physical environment, some of which can consist of privacy-sensitive information and be *out-of-camera* scene during the streaming session. The privacy problem is further exacerbated with the advent of 3D vision sensors (e.g., LiDAR). This newly endowed capability to mobile devices is a double-edged sword: it allows mobile devices to more efficiently capture and accurately reconstruct physical scenes for better AR features; it also presents an immediate threat in a new form of *reflection-based privacy*. Figure 1 demonstrates two examples where sensitive information such as a driver's license and a credit card can appear in reflective virtual objects. Consequently, when streaming AR content consisting of such objects (as will show in Figure 2), it can lead to undesired information leakage to any streaming viewers without the AR streamers necessarily noticing.

In this paper, we show, for the first time, that visual quality-driven multi-view lighting reconstruction can reveal out-of-camera scene information and compromise privacy for AR content creators. Existing works supporting reflective rendering, including commercial methods in ARKit [23] and academia research GLEAM and FusedAR [37, 59], all require the step to capture multiple glimpses of the physical environment. Without loss of generality, we present a privacy attack based on a recently proposed lighting reconstruction technique FusedAR [59], that extracts sensitive scene information such as human faces and text from the rendered objects under several plausible application scenarios. One of our key goals in demonstrating the effectiveness of this simple attack is to *increase the awareness of privacy issues associated with reflection rendering for AR applications*.

We note that visual privacy protection is not a new problem [8, 36]. Prior work has proposed many defenses for traditional multimedia, such as images and videos [54, 62]. Even for emerging mixed reality applications, we have also observed increased research efforts to ensure that an immersive virtual environment is built with security and privacy implications in mind [34, 35, 41]. Our paper falls into the broad AR/VR privacy research; one of our main contributions is *uncovering this new reflection-based privacy issue in the emerging AR applications*. We argue that the demonstrated attack is a natural progression from improved mobile sensors and environment understanding algorithms [16, 42, 44]. In other words, this reflection rendering-based attack is a consequence of improved lighting reconstruction for AR applications.

To defend against such attacks, we develop a novel privacy-preserving IPC^2S defense that *preserves the geometry information* while obfuscating the privacy-sensitive objects. Preserving the geometric information is critical in addressing the key challenge of simultaneously preserving privacy while still delivering visually coherent rendering. Additionally, we propose a R^2 defense that can bypass the lighting reconstruction and provide effective protection in dynamic environment conditions such as low lighting or motion blur. We leverage existing OCR and face detection models [6, 26] to identify text and human faces from past camera observations and blur the *color pixels* associated with detected regions. The transformed RGB images with the *unmodified depth information* are then combined into a point cloud, a 3D intermediate data we use to generate the final environment map for rendering.

To demonstrate that our IPC^2S defense can successfully obfuscate private information while delivering good visual effects, we evaluate the defense pipeline under 32 different rendering scenarios. We show that our IPC^2S defense achieves high visual quality with up to 36db PSNR and 0.98 SSIM while significantly reducing the automatic extraction success rate from 97.1% to 8.8% when compared to the privacy-risking reflection renderings. Lastly, we find that in addition to the three factors—physical scene, virtual object, and sensitive information, the accuracy of the face and text recognition models also can impact the information extraction success rate and the visual quality. We make the following main contributions.

- We present the first look at the *out-of-camera visual privacy* issue, i.e., the *reflection-based privacy*, that arises in AR applications. To demonstrate the prevalence of the privacy issue, we showcase a multi-view attack based on the ARKit and a recent lighting reconstruction technique [59] that successfully extracts sensitive information from reflective virtual objects.
- We propose an effective IPC^2S defense to automatically remove sensitive information from user-defined categories, such as human faces or textual information, and a conditional R^2 defense. Our IPC^2S defense is a lightweight pipeline that leverages machine learning models and image blurring techniques and can run in parallel to the current reflective rendering in AR frameworks.
- We implement the pipelines for lighting reconstruction, attack, and defense and evaluate the visual quality and sensitive information extraction. Relevant research artifacts are available at https://github.com/cake-lab/ar-reflection-privacy.

2 BACKGROUND

Virtual Object Rendering in AR Streaming. Imagine a content streamer Alice that uses AR-enabled applications, such as TikTok, to stream using platforms such as Twitch [49]. In this case, Alice is interested in augmenting her video stream with rendered objects. Depending on the streaming scenarios, e.g., virtual try-on of glasses or furniture shopping, Alice would also like overlaying visual-coherent virtual objects in her physical space. To produce visual-coherent virtual objects, the AR frameworks need to have access to accurate environment lighting information [37, 57]. The

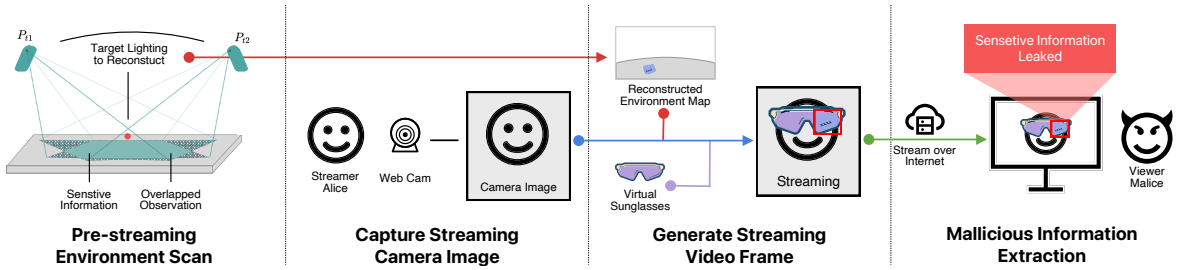


Figure 2: The reflection-based privacy issues in AR content creation and streaming. We demonstrate the general workflow of AR streaming to highlight how the rendered reflection of virtual sunglasses can contain sensitive information from the streamer Alice's physical surroundings and be leaked to the viewer Malice.

current commercial AR frameworks such as ARKit or ARCore often require mobile users to move around the cameras to scan the physical environment [17, 23]. The scanning phase will allow AR frameworks to collect useful environment information, which will further be used as input for a lighting estimation module to output lighting information [43]. The environment lighting information, often represented in the form of *environment map* [13], will then be used by rendering frameworks to overlay the virtual objects either in a user-specified world position [58] or a position based on tracking results [47]. Finally, each video frame augmented with virtual objects can be piped to existing streaming software such as OBSStudio [11] to use services like Twitch.

Visual Privacy Considerations in AR. Considering the AR streaming scenario described above, we will describe scenarios leading to privacy issues. The key privacy problem arises when the AR frameworks use captured environment images as part of the input for reconstructing environment lighting information [37, 59]. These environment images captured during the environment scanning phase (before the streaming) can lead to *out-of-camera* information leak when the AR framework uses a high-quality environment map to render reflective objects. We refer to this privacy problem as *reflection-based privacy* in which sensitive information from outside the current camera frame can appear on the virtual objects. Figure 4 shows example rendering effects of our proposed simple attack that leverages a popular AR framework ARKit and a recently proposed multi-view lighting reconstruction method [59]. When streaming the augmented video frames, such sensitive information will then become accessible to any viewers over the internet. More generally, such privacy issues can happen in many AR applications that satisfy the following characteristics. *(i) The need for photorealistic rendering.* Many compelling use cases of AR require photorealistic rendering. For example, in a 3D advertisement where an influencer tries to sell products (as rendered assets) to followers. *(ii) Physically separated users.* While many AR applications are multi-user, we have observed scenarios where AR users record and share their experiences via various platforms like Snapchat. In such scenarios, the existing platform users do not have to engage in AR technology directly but rather as consumers of AR content (see Figure 2).

3 LIGHTING RECONSTRUCTION PREMIER

We describe the *multi-view lighting reconstruction*, serving as the basis for the privacy issues we pinpoint in §4 and defenses in §5.

Step 1: Capturing Environment Data. Most mobile devices only have cameras with relatively small field-of-view, e.g., 77° [18]. Therefore, to capture omnidirectional environment observations, AR content creators are typically required to move the mobile device around and scan the surroundings. Traditionally, the capturing can be performed with the assistance of a physical chrome ball [13, 37]. In recent years, the increasingly popular mobile depth sensor [22, 24] enables the possibility of capturing highly accurate scene geometry. Similar to recent work [59], we perform lighting reconstruction with RGB-D images and device tracking data captured by mobile devices without requiring additional scene setups.

Step 2: Combing Multi-View Data. Next, we combine the captured multi-view data into a 3D point cloud representation in the same world space. We select the point cloud based on the virtual object rendering position within a cubic space with a size of 2 meters as *near-field*. To ensure the reconstruction quality, we only select the points with high depth confidence values, which measures the accuracy of the depth-map data. Moreover, we perform view-wise point cloud registrations using iterative closest point registration [4] to address noisy real-world tracking data.

Step 3: Finalizing Environment Lighting. Last but not least, we convert the collected near-field point cloud into an environment map, which is composed of near and far-field components: *(i)* The near-field component consists of the projection of the textured surface mesh reconstructed from the collected point cloud; *(ii)* for the far-field component, we use an indoor blurred HDR panorama image, similar to the far-field reconstruction policy described in [59].

4 REFLECTION-BASED PRIVACY ISSUES

4.1 Privacy Attack Overview

Figure 2 presents an overview of the AR streaming workflow where reflective rendering can lead to creators' physical environment information being recovered by viewers. As we defined previously, AR streaming is an emerging and popular form for indie streamers to reach out to followers via platforms such as Tiktok, YouTube, and Twitch [33, 47]. We assume that streamers use existing AR software to create videos with seamlessly overlayed virtual objects.

To generate quality content, e.g., photorealistic rendered objects coherently inserted into the physical scenes, our streamer Alice often needs to use the camera device to scan her physical surroundings. This step of *environment scanning* is a basic requirement of existing commercial AR frameworks such as ARCore or ARKit to obtain useful AR information, including world tracking data, camera intrinsic, and camera pose. As a consequence of this scanning, the AR session (and subsequently the AR stream) will have access

to the physical world information surrounding Alice. Note that, if not *directly captured* by the camera during the streaming, we assume that such physical world information should not be available to viewers. However, in the AR streaming case, when rendering virtual objects with reflective materials, e.g., the streamer wants to show the sunglasses try-on experience, the virtual sunglasses will be rendered with environment lighting information captured previously. Simply put, the virtual sunglasses might reflect different sensitive information, such as human faces or credit card information, when the streamer looks around (recall Figure 1). Finally, the attacker, i.e., the AR viewer, can access the rendered images or videos shared by the AR streamers directly on the attacker's device.

4.2 Sensitive Information Extraction

To demonstrate the prevalence of privacy issues in reflection rendering, we design a simple attack and show that we can successfully extract sensitive information from the rendered reflections. Our attack first reveals the out-of-camera sensitive information captured during lighting reconstruction by obtaining reflective virtual objects. For example, the attacker could ask the streamer to use existing reflective objects or hack the model assets used by the streamers and insert reflective components. Then, we perform automated sensitive information extraction by running face and OCR recognition models. Optionally, the attacker can unwrap the reflection area based on either virtual object geometry, viewing perspective, or both, if known. The unwrapping step will further increase the chance of automated information extraction as it removes the projection distortion.

In §6.2, we demonstrate that the automated attack can achieve a face/text extraction success rate of 63.24%/57.44%, on average, across a diverse set of rendering scenarios. In summary, given the ease of attack and the effectiveness in extracting sensitive information, we argue the rising need to protect AR content creation applications. Content creators might not realize that the visual information of their physical environment is being utilized for virtual object rendering. Such information leakage is unintended and undesirable. We believe it is necessary to have an automatic pipeline to identify privacy concerns and provide robust mechanisms to minimize unintended privacy leakage.

5 PRIVACY-PRESERVING REFLECTION

We present the design of two defense mechanisms that effectively and efficiently protect *reflection-based privacy*. Specifically, we design *index-based point cloud color swapping*, an automatic defense method we refer to as IPC^2S defense, for face and text information. As shown in Figure 3, IPC^2S defense is designed to run in parallel to the lighting reconstruction supporting reflection rendering and addresses the visual privacy issue by blurring out the sensitive information fields. We further extend our defense design to support dynamic environments and propose a restricted rendering-based method (referred to as R^2 defense) to protect privacy when the automatic defense IPC^2S defense falls short.

5.1 Key Design Challenge and Questions

Recall that the *reflection-based privacy* issue arises as AR frameworks strive to improve the visual coherency of the AR content. The key challenge when designing the defense is to minimize the impact on user-perceived visual quality while providing robust privacy protections for AR content creators. We tackle this challenge by answering two key design questions. *First, how to integrate the defense with the existing AR pipelines (§5.2)*? To generate visually coherent environment lighting for rendering virtual objects, AR frameworks such as ARKit often need to transform the environment scene observations multiple times based on viewing perspectives and scene geometries. It is therefore critical to carefully choose *where and how to* perform the image obfuscation for visual privacy without introducing visual artifacts or performance overhead. *Second, how to ensure privacy protection in dynamic environments (§5.3)?* It is common for AR content creators to work in dynamic environments, e.g., with a rapid change of environment lighting or moving cameras. These environments can be challenging for automatic defenses using deep learning models [32].

5.2 Index-based Point Cloud Color Swapping

To address the reflection-based privacy issue via image obfuscation, there are three main locations in the lighting reconstruction pipeline one can choose to obfuscate: *(i)* reconstruction client device camera RGB images; *(ii)* reconstructed lighting environment maps; and *(iii)* the rendered virtual object frames. However, neither *(ii)* nor *(iii)* are ideal as they are subject to image distortion from the panoramic image projection and geometry/viewing-perspective-related distortion in rendering, respectively. Instead, our IPC^2S defense pipeline takes the input directly from the RGB images collected from mobile devices as such images do not suffer from lighting reconstruction and rendering-related distortions. By doing so, we can avoid the negative impact of image distortion on recognition [32].

However, such a design comes with its unique challenge. Obfuscating the RGB images at the early stage of lighting reconstruction can impact the lighting reconstruction accuracy, as the obfuscated pixel color information could be used in point cloud registration. To address this challenge, we propose the IPC^2S defense, a novel design that allows identifying privacy issues at the early lighting reconstruction stage while *waiting to obfuscate* the privacy content in a later stage. The key idea is to correctly and efficiently map the pixels of sensitive information to the points in the intermediate point cloud.

As shown in Figure 3, parallel to the main reconstruction process, we spawn a separate process that first generates blurred RGB images for each RGB image received by the lighting reconstruction pipeline. Then, in the defense execution process, we run face and text detection models to recognition the information fields that appeared on each RGB image and use a bounding box to describe the information field regions. Next, the identified regions are recorded as pixel and frame number indexes, which correspond to the index of points in the intermediate point cloud of the lighting reconstruction. At the later stage of lighting reconstruction (i.e., before mesh texturing), we swap the previously identified sensitive information regions based on point cloud indexes. This design enables both high accuracy face/text recognition and eliminates the impact of image obfuscation on lighting reconstruction quality.

Furthermore, the IPC^2S defense runs in parallel to the unmodified lighting reconstruction pipeline, which has the potential to

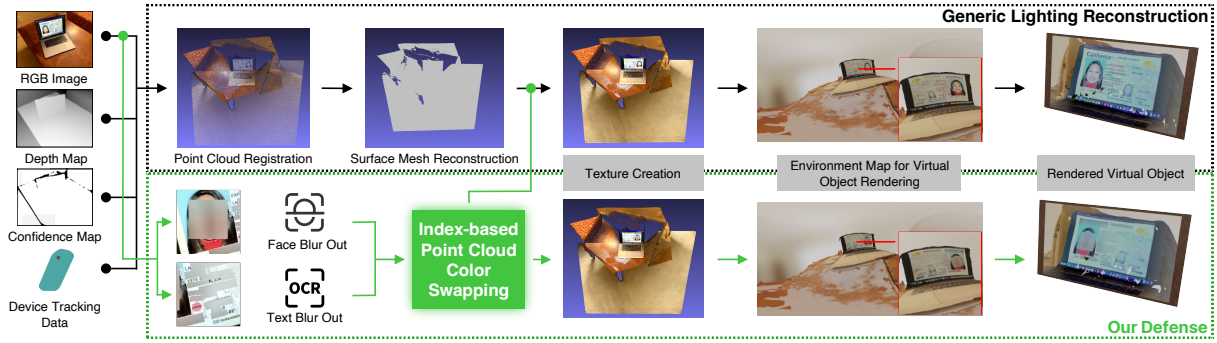


Figure 3: Our proposed defense for preserving reflection-based privacy for AR streaming. Our defense pipeline is designed around the key idea of index-based point color swapping to obfuscate the visual detail of the sensitive information while still maintaining the overall reflection color pattern. Moreover, the defense pipeline can run in parallel to the generic lighting reconstruction pipeline described in Sec 3.

minimize the latency impact of IPC^2S defense and to be integrated with other reconstruction pipelines. Empirically, our measurements show an average execution time of face/text recognition at 0.05s/0.13s compared to the 30s needed by the point cloud registration. More concretely, this means that the high-quality lighting information (used to render reflection) can take up to 30 seconds using the reference pipeline implemented based on FusedAR [59]. Our defense pipeline does not pose a performance bottleneck on normal AR usage as its main steps are complete well before the lighting information is ready.

5.3 Defense in Dynamic Environments

Besides image distortion, the automatic face/text recognition accuracy can be affected by other environmental factors like lighting, camera movements, etc. For example, in AR content creation and streaming applications, lighting reconstruction is usually performed in two cases: *before the beginning of streaming* and *during the streaming*. Automatic face/text detection failures caused by sudden environmental changes in the pre-streaming scenario may not cause immediate privacy issues as AR content creators can be allowed to re-scan the environment to avoid sharing sensitive information with the viewers. However, our defense should handle the dynamic environment during the AR streaming, as recognition failures will lead to imminent privacy leakage.

Therefore, in addition to IPC^2S defense, we propose the R^2 defense, which bypasses the lighting reconstruction and can provide immediate protection. R^2 defense controls the maximum material reflection rate and roughness, which can be executed easily and efficiently in most modern graphics rendering engines. In particular, we limit the maximum material reflection to 0.8 (from 1.0) and minimum roughness to 0.2 (from 0.0). The changing environments can be detected by leveraging built-in hardware sensors, e.g., ambient light sensor, accelerometer, and gyroscope.

6 EXPERIMENTS

We evaluate our proposed defense's effectiveness in preserving privacy and the respective impact on the rendering quality of virtual objects. Our evaluation centers around answering the key question: *how well does our defense work in preserving privacy and maintaining good visual coherency?* We test a total of 32 rendering scenarios, where each scenario refers to an instance of *(scene, reflective object, sensitive information)*. Our key findings are summarized as follows: *(i)* Our simple attack can successfully extract up to 100% human faces and at least 92% textual information when inspecting the extractions manually (§6.2). *(ii)* Our IPC^2S defense effectively reduces the information extraction success rates to at most 8.8% and 23.8% under automatic and manual inspections for all tested rendering scenarios. (§6.3.2). *(iii)* Compared to the R^2 defense, our IPC^2S defense achieves an average of 9.31% better SSIM while successfully preserving privacy under manual inspection(§6.3.1). *(iv)* Automatic extraction poses more difficulty for identifying faces than texts, for both the attack and the defense.

6.1 Experimental Setup

Implementations. To demonstrate the prevalence of the reflection-based privacy issues in AR, we implement *a simple lighting reconstruction pipeline*, following the generic lighting reconstruction pipeline paradigm, by leveraging widely available open-source tools and libraries. The lighting reconstruction pipeline consists of both an iOS app developed with Unity3D [52] and the ARFoundation framework [51] as the client and a Python server. During reconstruction, we first stream the collected AR scene information, RGB-D image, device tracking data, and camera pose, from the mobile client to the backend server and store the scene information for further processing. Next, we perform the point cloud registration and surface reconstruction on the backend server using the Open3D [60] and the Meshlab [9] libraries. Finally, we generate environment maps from reconstructed meshes using Blender [10].

We use the reconstructed environment map to implement *our attack* by first rendering reflective objects using Blender. Then, we implement the automatic sensitive information extraction of face and text recognition with OpenCV [6] and EasyOCR [26] libraries. We envision the attack will occur as a natural progression of AR frameworks supporting reflective rendering—*malicious users/viewers do not need to investigate the inner-working of the pipeline*; rather, malicious users simply need to gain access to the rendered reflections. We choose to implement our defense pipeline as a parallel component to the generic lighting reconstruction pipeline described in the FusedAR paper [59]. Using the collected RGB image during lighting reconstruction, we use the same face and text recognition tools for the defense (as the attack for a fair comparison) to identify sensitive information and blur sensitive information of the rendered images using the PIL framework [50]. The IPC^2S defense is implemented

as a NumPy [20] `ndarray` operation and the R^2 defense as a special material in Blender.

Rendering Scenarios. We first use an iPad Pro with a LiDAR sensor to capture RGB-D images and device tracking information in *four indoor scenes* with different scene geometries and physical objects and reconstruct the environment lighting for each scene. To render the reflection, we choose two virtual objects with representative geometries, a *metallic sphere* and a *flat mirror*. For the sensitive information, we select two sample US driver's licenses (Massachusetts and California), one *group photo* with 14 persons, and one sample *credit card* with seven information fields, to represent three types of information leakage—a mixed of text and human face, human face-only, and text-only, respectively. *Driver's license 1* contains a total of 19 (1 face and 18 texts) information fields, and the *driver's license 2* contains a total of 19 (2 faces and 17 texts) information fields. For simplicity, we display the sensitive information on a screen of a Macbook Pro 15" inside each scene.

At a high level, the environment capturing process involves a user scanning the indoor scene containing sensitive information. The depth map is captured at the resolution of 256x192, and the RGB color image is captured at the resolution of 1280x960. We then import the reconstructed environment mesh into Blender to generate an environment cubemap with 2048x2048 resolution per cube face. The environment map is composed of near and far-field components: *(i)* The near-field component consists of the projection of the near-field textured scene geometry; *(ii)* for the far-field component, we use an indoor blurred HDR panorama image, similar to the far-field reconstruction policy described in [59]. Note that all of our attack and defense evaluations are based on the near-field geometries—any panorama image can be used for the far-field without impacting the observed results. With the generated environment maps, we use Blender with the Principled BSDF shader [5] to render the virtual objects, which will then be displayed.

Evaluation Baselines and Metrics. We choose a commercial AR framework ARKit [23] and a recent lighting reconstruction pipeline FusedAR [59] as the baselines of visual quality. To evaluate whether the proposed attack can successfully extract the sensitive information, i.e. the effectiveness of the attack, we use the success rate of information extraction as our metric.

$$\textit{success rate} = \frac{\text{\# of successful cases of information extraction}}{\text{\# of valid test cases}}$$

For each detected text field, we calculate the *Levenshtein distance* between the recognized value and its ground truth value. A recognition is considered successful if the following conditions are both met: *(i)* the recognized value is not empty; and *(ii)* the Levenshtein distance is less than 10. For face recognition, we draw face bounding boxes on the reflection renderings and manually inspect whether each face is detected or not.

Furthermore, we use Peak signal-to-noise ratio (PSNR) and Structural Similarity Index (SSIM), commonly used image quality metrics, to quantify the impact of our defense on perceptible visual quality. PSNR and SSIM values are calculated by comparing the virtual objects rendered with and without the proposed defense.

Table 1: The high success rate of our attack. On average we can manually extract 100% human face information and at least 92% textual information. The automatic extraction has lower success rates but still poses considerable privacy issues.

Sensitive Info	Virtual Object	Face Recognition		Text Recognition	
		Automatic	*Manual*	*Automatic*	*Manual*
Driver's License 1	Metal Ball	50.00%	100.00%	33.33%	87.50%
	Mirror	100.00%	100.00%	81.94%	97.83%
Driver's License 2	Metal Ball	100.00%	100.00%	20.59%	94.12%
	Mirror	100.00%	100.00%	80.88%	95.59%
Group Photo	Metal Ball	17.85%	100.00%	N/A	N/A
	Mirror	96.42%	100.00%	N/A	N/A
Credit Card	Metal Ball	N/A	N/A	75.00%	96.42%
	Mirror	N/A	N/A	71.42%	96.42%
Average	Metal Ball	29.41%	100.00%	35.11%	91.67%
Average	Mirror	97.06%	100.00%	79.76%	95.83%

6.2 Information Extraction of Our Attack

We evaluate the effectiveness of our proposed attack by attempting to extract sensitive information from *images of reflective virtual objects* rendered with reconstructed environment lighting. Using environment lighting reconstructed from 4 scenes, we generate 8 renderings of both metallic ball and mirror objects for each scene. We evaluate the success rates of both automatic extraction and manual inspection methods of 17 faces and 42 text fields appearing in the reflection renderings.

Figure 4 visualizes different types of sensitive information we can extract from reflective virtual objects. We note that visually, a human user can easily identify sensitive information by inspecting the images displayed in row two. Table 1 shows that by leveraging automatic face recognition, our attack can recognize, on average, 29.41% and 97.06% faces automatically on the metal ball and mirror objects, respectively. For automatic text recognition, we see an average success rate of 35.11% and 79.76% on the metal ball and mirror objects. Further, via manual inspection of the renderings, we achieve 100% face recognition rate on both metal ball and mirror objects, and up to 95.8% recognition rate of textual information on the mirror object.

We make three key observations. First, we can extract almost all sensitive information through manual inspection, despite the severe information distortions on virtual object renderings. Automatic recognition effectively extracts information from a flat object, suggesting the risk of large-scale automated reflection-based privacy attacks. Second, the manual inspection leads to a much higher information extraction success rate than the automatic recognition from the metal ball object. We suspect that image distortion plays a key role in determining the success rate. Third, as image distortion can be caused by many factors such as the geometry of the scene, the geometry of the virtual object, and the viewing perspective, it can be challenging to devise an automatic unwrapping method. However, to further improve the success rate of the automatic recognition, we believe one can resort to more accurate recognition methods as they are developed.

6.3 Effectiveness of the Defense

We evaluate the visual quality impact and the visual perturbation effectiveness of our IPC^2S defense on the reflective rendering. For

Figure 4: The visual effectiveness of our attack on reflection-based privacy. Row one shows the reflection rendering of two virtual objects; row two zooms in on the four types of sensitive information leakage.

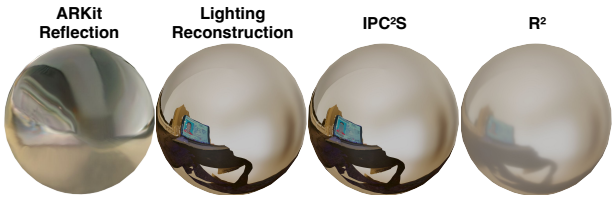
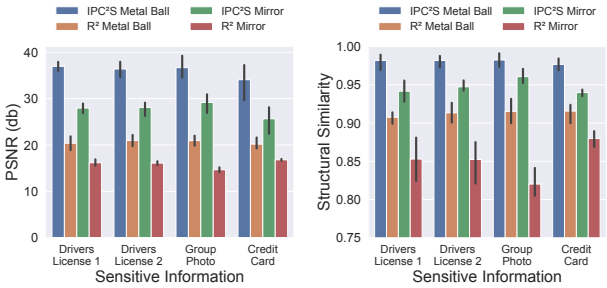


Figure 5: The low visual impact of our defense on reflection-based privacy. *IPC²S* defense achieves similar reflective visual quality compared to a re-implemented pipeline [59].



(a) Rendering PSNR. **(b) Rendering SSIM**

Figure 6: Quantitative comparisons of our *IPC²S* defense and R^2 defense. *IPC²S* defense achieves high PSNR and SSIM scores while R^2 defense has at least 0.91 and 0.82 SSIM values.

the visual quality impact, we compare the rendered objects to the ones generated by *ARKit*, *FusedAR*, and our R^2 defense; we also quantity the visual impacts using two image-based metrics (PSNR and SSIM) by calculating against the reflection renderings generated by *FusedAR*. Our results show that *IPC²S* defense has a low visual impact compared to the undefended privacy-risking renderings (up to 36db PSNR and 0.98 SSIM). Further, our *IPC²S* defense effectively preserves reflection-based privacy, successfully decreasing the automatic information extraction rate to at most 8.8%/5.4% compared to 97%/80% when undefended.

6.3.1 Visual Quality Impacts. Figure 5 visualizes the two reflective objects rendered with environment lighting information consisting of the driver's license one [1]. We first observe that our *IPC²S* defense can keep the detail of scene objects and geometries while only obfuscating the desirable information fields, compared to the rendering effects achieved with the re-implemented lighting reconstruction pipeline based on FusedAR (column two). The R^2 defense also successfully obfuscates the sensitive information fields but lowers the overall visual quality. Specifically, with the reduction of the metallic property and introduction of roughness, reflective objects appear to have a matte-like looking. As a reference, we also include the rendering generated using ARKit; we note that ARKit currently does not sufficiently support reflective renderings. However, as newer lighting reconstruction techniques are adopted [37, 43, 59] in commercial AR frameworks, the reflection-based privacy issue will become more prevalent.

Figure 6 compares the rendering quality achieved by our *IPC²S* defense and R^2 defense against the undefended rendering. We first note that our *IPC²S* defense achieves, on average 36.03db and 27.03db PSNR values for the metal ball and mirror objects across all four scenes. The *IPC²S* defense and R^2 defense achieve high SSIM scores for all tested rendering scenarios. Furthermore, our *IPC²S* defense outperforms the R^2 defense on both metal ball and mirror objects. In particular, our *IPC²S* defense achieves 74.75% and 7.43% higher PSNR and SSIM than R^2 defense on the metal ball object, as well as 74.86% and 11.39% higher PSNR and SSIM than R^2 defense on the mirror object. These results suggest that one should prioritize the use of *IPC²S* defense over R^2 defense as much as possible to minimize the impact on visual quality. As we described in §5, we only fall back to R^2 defense when the automated recognition accuracy and confidence fall below a certain threshold. As part of future work, we will investigate runtime policies to regulate the use of these two complementary defenses.

6.3.2 Defense Success Rate. Finally, we evaluate the effectiveness of our defenses following a similar methodology and metric as described in §6.2. Table 2 shows the information extraction success rate when using our *IPC²S* defense[2]. First, we see that *IPC²S* defense can prevent at least 88.14% and 76.19% of face and text information extraction under manual inspection. This is in stark contrast to the 100% human face and at least 92% textual information extraction, if left undefended, as shown in Table 1. Second, we show that our *IPC²S* defense is effective across all 32 rendering scenarios. For the automatic extraction, we can prevent all sensitive information from being leaked for the metal ball object and at

[1]The visualization of other rendering scenarios is omitted due to space limitations.

[2]R^2 defense results are omitted as none of the sensitive information can be extracted.

Table 2: The low success rate when using our IPC^2S defense. On average, this defense effectively decreases the success rate of automatic extraction to at most 8.8%/5.4% and manual extraction to at most 11.8%/23.8% for face/text information.

Sensitive Info	Virtual Object	Face Recognition		Text Recognition	
		Automatic	*Manual*	*Automatic*	*Manual*
Driver's License 1	Metal Ball	0.00%	0.00%	0.00%	16.67%
	Mirror	0.00%	0.00%	6.94%	18.06%
Driver's License 2	Metal Ball	0.00%	0.00%	0.00%	30.88%
	Mirror	0.00%	0.00%	5.88%	36.76%
Group Photo	Metal Ball	0.00%	14.29%	N/A	N/A
	Mirror	10.71%	14.29%	N/A	N/A
Credit Card	Metal Ball	N/A	N/A	0.00%	3.57%
	Mirror	N/A	N/A	0.00%	7.14%
Average	Metal Ball	0.00%	11.76%	0.00%	20.24%
Average	Mirror	8.82%	11.76%	5.36%	23.81%

least 91.2% of information for the mirror object. Third, the larger difference in automatic extraction rates across the attack and the IPC^2S defense suggests that our point cloud-based, rather than image-based, defense design is more recognition model friendly.

7 RELATED WORK

AR/VR Security/Privacy. With the growing popularity of AR/VR applications, the potential security and privacy issues caused by hybrid physical and virtual environments have recently emerged as a new research domain. Many research efforts have focused on the security/privacy implications of on-device sensors, as these sensors are increasingly utilized to capture sensitive user data or behaviors to build the immersive virtual environment [27, 34, 35, 40, 41]. Also, the sensitive nature of virtual objects constructed and presented in the AR/VR scenes has led to research efforts on deceptive virtual objects that mislead the users (i.e., the *integrity* issue) [28, 29], as well as sensitive virtual objects [48] that can be abused by adversaries (i.e., the *confidentiality* issue). Our work falls into the broad AR/VR privacy research: it differs from state-of-the-art works by targeting the non-conventional, reflective virtual objects rendered with views outside of any capturing devices.

Visual Privacy Protection. Visual privacy protection has been a well-studied research topic for traditional multimedia, such as 2D images and videos [38]. The existing approaches to eliminating visual privacy leakage can be divided into two main categories. The first category aims to intervene/interfere with the sensors or scenes *in the physical world* to prevent privacy-sensitive content from being captured in the first place [21, 36, 62]. The second category focuses on removing, replacing, or blurring sensitive objects *in the virtual world* using computer vision techniques, such as image inpainting [3, 8], body/face de-identification [7, 12, 15], and image obfuscation [14, 39, 46, 53–55, 61]. It is possible to apply existing techniques that remove reflection from an image [30, 31, 56] by treating the rendered virtual object the same as its physical object counterpart. Though it is unclear how well existing techniques will work for more geometrically complex virtual objects or with distortion. In contrast, our defense mechanisms were designed with the knowledge of the inner working of the lighting reconstruction pipeline and thus can work more synergistically. Our work is inspired by the existing obfuscation-based approaches for visual privacy protection; in contrast to prior work, we target the privacy issues arising with the nascent development of photorealistic rendering [37, 57] in AR—a new multimedia that comes with brand new challenges, including intricate visual quality, privacy, and performance trade-offs.

8 CONCLUSION AND FUTURE WORK

In this paper, we argued that unintentional privacy leakage could happen as augmented reality applications become popular. Specifically, sensitive information (such as human faces) can be leaked via reflective rendering—an integral part of photorealistic AR. To underpin the importance of the reflection-based privacy issues, we showcased a simple attack leveraging a recently proposed multi-view lighting reconstruction [59]. Our attack can successfully extract sensitive information under various rendering scenarios. We also noted that such attacks are not specific to a particular lighting reconstruction method and can happen with existing commercial AR frameworks [23] and other academic works [37]. The fundamental issue about this unintentional privacy leakage—in our example, between an AR content creator and a viewer—comes down to the seemingly conflicting goals of visual coherency and privacy.

As explained previously, achieving visual coherency for AR objects requires an accurate understanding of the physical environment. Based on current common practices to achieve visual coherency, it is inevitable that sensitive information will be captured and included as part of the environment scans. However, we showed that we can still achieve good visual coherency while preserving privacy by carefully designing the defense pipeline in tandem with the lighting reconstruction. Specifically, we proposed two complementary defenses (IPC^2S and R^2 defense) to obfuscate sensitive information, even under dynamic environments, automatically. Even if the sensitive information were captured during the AR sessions, it would not be subject to unintentional information leakage.

Our proposed defense is far from complete—there are many unsolved challenges we plan to address. For example, many objects can be considered private, and we only showcased the defense mechanism when considering human face and text information are sensitive. Currently, our work does not consider physical object reflection. As AR deals with both virtual and physical spaces, we agree that an effective and complete system should consider both types of reflections to preserve user privacy. Further, many practical scenarios, e.g., moving objects and varying environmental lighting, exacerbate the problem. Robustly identifying private information with minimal human user involvement can be an exciting direction. Another direction is to devise better privacy-preserving techniques. Currently, we use a simple blurring technique to obfuscate sensitive information. We envision more sophisticated techniques such as automatically generating suitable replacements in real-time can achieve better visual quality. We hope our study can improve the community's awareness of the need to support privacy-preserving reflection rendering.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive reviews. This work was partly supported by NSF Grants #1815619, #1912593, and #2105564, and VMWare.

REFERENCES

[1] Thomas Alsop. 2020. Augmented reality (AR) - statistics & facts. https://www.statista.com/topics/3286/augmented-reality-ar/. Accessed: 2020-7-2.

[2] Christopher Andrews, Michael K Southworth, Jennifer N A Silva, and Jonathan R Silva. 2019. Extended Reality in Medical Practice. *Curr. Treat. Options Cardiovasc. Med.* 21, 4 (March 2019), 18.

[3] Marcelo Bertalmio, Luminita Vese, Guillermo Sapiro, and Stanley Osher. 2003. Simultaneous structure and texture image inpainting. *IEEE transactions on image processing* 12, 8 (2003), 882–889.

[4] Paul J Besl and Neil D McKay. 1992. Method for registration of 3-D shapes. In *Sensor fusion IV: control paradigms and data structures*, Vol. 1611. Spie, 586–606.

[5] Blender. 2022. Principled BSDF. https://docs.blender.org/manual/en/latest/render/shader_nodes/shader/principled.html.

[6] G. Bradski. 2000. The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000).

[7] Karla Brkic, Ivan Sikiric, Tomislav Hrkac, and Zoran Kalafatic. 2017. I know that person: Generative full body and face de-identification of people in images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 1319–1328.

[8] Dongwook Cho and Tien D Bui. 2008. Image inpainting using wavelet-based inter- and intra-scale dependency. In *International Conference on Pattern Recognition (ICPR)*. 1–4.

[9] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, Guido Ranzuglia, et al. 2008. Meshlab: an open-source mesh processing tool.. In *Eurographics Italian chapter conference*, Vol. 2008. Salerno, Italy, 129–136.

[10] Blender Online Community. 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. http://www.blender.org

[11] The OBS Project Contributors. 2017. Open Broadcasting Software. https://obsproject.com/.

[12] Enric Corona, Albert Pumarola, Guillem Alenya, Gerard Pons-Moll, and Francesc Moreno-Noguer. 2021. SMPLicit: Topology-aware generative model for clothed people. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 11875–11885.

[13] Paul Debevec. 2006. Image-based lighting. In *ACM SIGGRAPH 2006 Courses*. 4–es.

[14] Liyue Fan. 2019. Practical Image Obfuscation with Provable Privacy. In *IEEE International Conference on Multimedia and Expo (ICME)*. 784–789.

[15] Oran Gafni, Lior Wolf, and Yaniv Taigman. 2019. Live face de-identification in video. In *IEEE/CVF International Conference on Computer Vision (ICCV)*. 9378–9387.

[16] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. 2017. Learning to Predict Indoor Illumination from a Single Image. *ACM Transactions on Graphics* (2017).

[17] Google. 2020. ARCore. https://developers.google.com/ar.

[18] Google. 2021. Pixel 5a with 5G Tech Specs.

[19] Google for Education. 2022. Bringing virtual and augmented reality to school | Google for Education. https://edu.google.com/products/vr-ar/?modal_active=none. Accessed: 2022-4-10.

[20] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. 2020. Array programming with NumPy. *Nature* 585, 7825 (Sept. 2020), 357–362. https://doi.org/10.1038/s41586-020-2649-2

[21] Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. 2021. Learning Privacy-Preserving Optics for Human Pose Estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2573–2582.

[22] HUAWEI. 2022. HUAWEI Mate 30 Pro Specifications | HUAWEI Global. https://consumer.huawei.com/en/phones/mate30-pro/specs/. Accessed: 2022-4-10.

[23] Apple Inc. 2020. Introducing ARKit 4. https://developer.apple.com/augmented-reality/arkit/.

[24] Apple Inc. 2020. iPad Pro 2020. https://www.apple.com/ipad-pro/specs/.

[25] Inter IKEA Systems B. V. 2017. IKEA Place. https://apps.apple.com/us/app/ikea-place/id1279244498. Accessed: 2020-7-2.

[26] JaidedAR. 2022. EasyOCR. https://github.com/JaidedAI/EasyOCR.

[27] Suman Jana, David Molnar, Alexander Moshchuk, Alan Dunn, Benjamin Livshits, Helen J Wang, and Eyal Ofek. 2013. Enabling Fine-Grained Permissions for Augmented Reality Applications with Recognizers. In *USENIX Security Symposium (Security)*. 415–430.

[28] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2017. Securing augmented reality output. In *IEEE Symposium on Security and Privacy (S & P)*. 320–337.

[29] Kiron Lebeck, Kimberly Ruth, Tadayoshi Kohno, and Franziska Roesner. 2018. Towards security and privacy for multi-user augmented reality: Foundations with end users. In *IEEE Symposium on Security and Privacy (S & P)*. 392–408.

[30] Yunfei Liu, Yu Li, Shaodi You, and Feng Lu. 2022. Semantic Guided Single Image Reflection Removal. *ACM Trans. Multimedia Comput. Commun. Appl.* (jan 2022).

[31] Yu-Lun Liu, Wei-Sheng Lai, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. 2020. Learning to See Through Obstructions. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[32] Z Liu, G Lan, J Stojkovic, Y Zhang, C Joe-Wong, and M Gorlatova. 2020. CollabAR: Edge-assisted Collaborative Image Recognition for Mobile Augmented Reality. In *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 301–312.

[33] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More Kawaii than a Real-Person Live Streamer: Understanding How the Otaku Community Engages with and Perceives Virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21, Article 137)*. Association for Computing Machinery, New York, NY, USA, 1–14.

[34] Shiqing Luo, Xinyu Hu, and Zhisheng Yan. 2022. HoloLogger: Keystroke Inference on Mixed Reality Head Mounted Displays. *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (2022).

[35] Shiqing Luo, Anh Nguyen, Chen Song, Feng Lin, Wenyao Xu, and Zhisheng Yan. 2020. OcuLock: Exploring human visual system for authentication in virtual reality head-mounted display. In *Network and Distributed System Security Symposium (NDSS)*.

[36] Shwetak N Patel, Jay W Summet, and Khai N Truong. 2009. Blindspot: Creating capture-resistant spaces. In *Protecting Privacy in Video Surveillance*. Springer, 185–201.

[37] Siddhant Prakash, Alireza Bahremand, Linda D Nguyen, and Robert LiKamWa. 2019. Gleam: An illumination estimation framework for real-time photorealistic augmented reality on mobile devices. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*. 142–154.

[38] Siddharth Ravi, Pau Climent-Pérez, and Francisco Florez-Revuelta. 2021. A Review on Visual Privacy Preservation Techniques for Active and Assisted Living. *arXiv preprint arXiv:2112.09422* (2021).

[39] Zhongzheng Ren, Yong Jae Lee, and Michael S Ryoo. 2018. Learning to anonymize faces for privacy preserving action detection. In *European conference on computer vision (ECCV)*. 620–636.

[40] Franziska Roesner, Tadayoshi Kohno, and David Molnar. 2014. Security and privacy for augmented reality systems. *Commun. ACM* 57, 4 (2014), 88–96.

[41] Cong Shi, Xiangyu Xu, Tianfang Zhang, Payton Walker, Yi Wu, Jian Liu, Nitesh Saxena, Yingying Chen, and Jiadi Yu. 2021. Face-Mic: inferring live speech and speaker identity via subtle facial dynamics captured by AR/VR motion sensors. In *International Conference on Mobile Computing and Networking (MobiCom)*. 478–490.

[42] Gowri Somanath and Daniel Kurz. 2021. HDR Environment Map Estimation for Real-Time Augmented Reality. https://arxiv.org/pdf/2011.10687.pdf

[43] Gowri Somanath and Daniel Kurz. 2021. HDR Environment Map Estimation for Real-Time Augmented Reality. *CVPR* (2021).

[44] Shuran Song and Thomas Funkhouser. 2019. Neural Illumination: Lighting Prediction for Indoor Environments. *CVPR* (2019).

[45] VRoid Studio. 2022. VRoid Studio. https://vroid.com/en/studio.

[46] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. 2018. Natural and effective obfuscation by head inpainting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5050–5059.

[47] Man To Tang, Victor Long Zhu, and Voicu Popescu. 2021. AlterEcho: Loose Avatar-Streamer Coupling for Expressive VTubing. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 128–137.

[48] Zhongze Tang, Xianglong Feng, Yi Xie, Huy Phan, Tian Guo, Bo Yuan, and Sheng Wei. 2020. VVSec: Securing Volumetric Video Streaming via Benign Use of Adversarial Perturbation. In *International Conference on Multimedia (MM)*. 3614–3623.

[49] Twitch. 2022. Twitch. https://www.twitch.tv.

[50] P Umesh. 2012. Image Processing in Python. *CSI Communications* 23 (2012).

[51] Unity. 2020. AR Foundation 4.2.2. https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@4.2/manual/index.html.

[52] Unity3D. 2022. Unity. https://unity3d.com. Accessed: 2022-4-9.

[53] Sen Wang and J Morris Chang. 2020. Privacy-Preserving Image Classification in the Local Setting. *arXiv:2002.03261* (2020).

[54] Hao Wu, Xuejin Tian, Minghao Li, Yunxin Liu, Ganesh Ananthanarayanan, Fengyuan Xu, and Sheng Zhong. 2021. PECAM: privacy-enhanced video streaming and analytics via securely-reversible transformation. In *International Conference on Mobile Computing and Networking (MobiCom)*. 229–241.

[55] Mengmei Ye, Zhongze Tang, Huy Phan, Yi Xie, Bo Yuan, and Sheng Wei. 2022. Visual Privacy Protection in Mobile Image Recognition Using Protective Perturbation. In *ACM Multimedia Systems Conference (MMSys)*.

[56] Xuaner Zhang, Ren Ng, and Qifeng Chen. 2018. Single Image Reflection Separation with Perceptual Losses. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[57] Yiqin Zhao and Tian Guo. 2020. PointAR: Efficient Lighting Estimation for Mobile Augmented Reality. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer International Publishing, Cham, 678–693.

[58] Yiqin Zhao and Tian Guo. 2021. Xihe: A 3D Vision-based Lighting Estimation Framework for Mobile Augmented Reality. In *The 19th ACM International Conference on Mobile Systems, Applications, and Services*.

[59] Yiqin Zhao and Tian Guo. 2022. FusedAR: Adaptive Environment Lighting Reconstruction for Visually Coherent Mobile AR Rendering. *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (2022).

[60] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. 2018. Open3D: A Modern Library for 3D Data Processing. *arXiv:1801.09847* (2018).

[61] Bingquan Zhu, Hao Fang, Yanan Sui, and Luming Li. 2020. Deepfakes for Medical Video De-Identification: Privacy Protection and Diagnostic Information Preservation. In *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*. 414–420.

[62] Shilin Zhu, Chi Zhang, and Xinyu Zhang. 2017. Automating visual privacy protection using a smart led. In *International Conference on Mobile Computing and Networking (MobiCom)*. 329–342.