Automating Visual Privacy Protection in the IoT Space

Shilin Zhu
University of California San Diego
shz338@eng.ucsd.edu

Chi Zhang
University of Wisconsin-Madison
czhang296@wisc.edu

Xinyu Zhang
University of California San Diego
xiz368@eng.ucsd.edu

ABSTRACT

The ubiquity of mobile camera devices has been triggering an outcry of privacy concerns, whereas privacy protection still relies on the cooperation of the photographer or camera hardware, which can hardly be guaranteed in practice. In this paper, we introduce LiShield, which automatically protects a physical scene against photographing, by illuminating it with smart LEDs flickering in specialized waveforms. Our prototype implementation and experiments show that LiShield can effectively destroy unauthorized capturing while maintaining robustness against potential attacks.

1 INTRODUCTION

Cameras are now pervasive on consumer mobile devices, such as smartphones, tablets, drones, smart glasses, first-person recorders, *etc.* The ubiquity of these cameras, paired with pervasive wireless access, is creating a new wave of visual sensing applications, *e.g.*, autonomous photographer, quantified-self (life-logging), photo-sharing social networks, physical-analytics in retail stores, and augmented reality applications that navigate users across unknown environment [6]. Zooming in the photo-sharing application alone, statistics report that 350 million photos/videos are uploaded to Facebook every day, majority of which are from mobile users [7]. Many of these applications automatically upload batches of images/videos online, with a simple one-time permission from the user. While these technologies bring significant convenience to individuals, they also trigger an outcry of privacy concerns.

However, *visual privacy protection in such passive physical spaces still heavily relies on rudimentary approaches* like warning signs and human monitors, and there is no way to automatically enforce the requirements. In this paper, we propose LiShield, a system that thwarts photographing of sensitive indoor physical space, and automatically enforces location-bound visual privacy protection. LiShield safeguards the physical scenes against undesired recording without requiring user intervention, and without disrupting the human visual perception. Our key idea is to illuminate the environment using smart LEDs, which are intensity-modulated following specialized waveforms. We design the waveform in such a way that its modulation pattern is imperceptible by human eyes, but can disrupt the image sensors on mobile camera devices.

In addition, LiShield can tailor the waveform for two special use cases: *(i.)* allowing an authorized camera, which shares secret configuration information with the LED, to recover the image or video frames it captures. *(ii.)* when strong ambient light interferes with the smart LED, LiShield cannot ensure full protection, but it can still emit structured light which embeds invisible "barcode" into the physical environment. The embedded information can convey a "no distribution" message, allowing online servers (*e.g.*, from Facebook and Instagram) to block and prevent the image from being distributed.

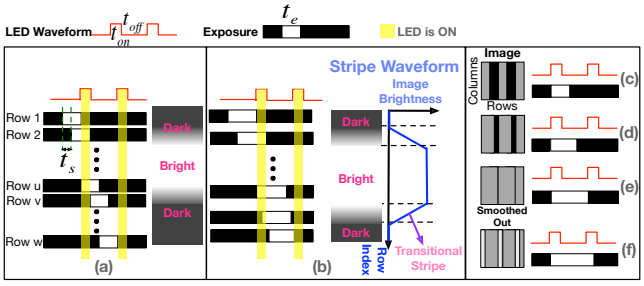

Figure 1: (a)-(b) Bright, dark and transitional stripes and their width changing with exposure time; (c)-(f) Stripe pattern of image changes under different exposure times.

2 PHYSICAL SCENE DISRUPTION

2.1 A Primer on Camera Image Disruption in LiShield

Cameras and human eyes perceive scenes in fundamentally different ways. Human eyes process continuous vision by accumulating light signals, while cameras slice and sample the scene at discrete intervals. Consequently, human eyes are not sensitive to high frequency flickers beyond around 80 Hz either in brightness or chromaticity [3, 9], while cameras can easily pick up flicker above a few kHz [8].

Unlike professional or industrial cameras which may have global shutter that mimics human eyes to some degree, nearly all consumer digital cameras, pinhole cameras and smartphones use the rolling shutter sampling mechanism [4], which is a main contributor to their high frequency sensitivity. When capturing an image frame, a *rolling shutter camera exposes each row sequentially.*

LiShield harnesses the disparity between cameras and eyes to disrupt the camera imaging without affecting human vision. It modulates a smart LED to generate high frequency flickering patterns (Fig. 1) which can be *monochrome* or *RGB colors.*

2.2 Maximizing Image Quality Degradation

LiShield aims to *minimize the image capturing quality by optimizing the LED waveform, characterized by modulation frequency, intensity, and duty cycle.* We focus on two widely adopted image quality metrics: *PSNR*, which quantifies the disruption on individual pixel intensity levels; and *SSIM*, which measures the structural distortion to the image (*i.e.*, deformation effects such as stretching, banding

S3'17, October 20, 2017, Snowbird, UT, USA

ACM ISBN 978-1-4503-5145-4/17/10…$15.00
https://doi.org/10.1145/3131348.3131356

and twisting). In general, the minimum PSNR and SSIM corresponding to acceptable viewing quality are in range of 25∼30 and 0.8∼0.9, respectively [1]. Since the stripe pattern follows a piecewise function, a closed form expression of PSNR and SSIM becomes infeasible. We thus use numerical simulation to evaluate the impact of LiShield.

Our numerical results show a few general trends, which lead to the following design choices for LiShield. *(i) A single frequency cannot ensure robust protection. (ii) LiShield must prevent attackers from using long exposures. (iii) LiShield should keep a high peak intensity to expand the overexposure zone. (iv) Duty cycle should be kept at moderate level.*

2.3 Circumventing Potential Attacks

Based on the foregoing analysis, we identify the following potential holes that can be exploited by attackers to overcome the striping effect. *(i) Manual exposure attack.* An attacker can configure the exposure time to eliminate the stripes during a capture (Fig. 1(e)). *(ii) Multi-frame attack.* When the scene is static, an attacker may also combine multiple frames (taking a video and playback) to mitigate the stripes with statistical clues, *e.g.* by averaging or combining rows with maximum intensities from multiple frames. *(iii) Post-processing attack.* Common post-processing techniques (*e.g.*, denoising and de-banding) might be used to repair the corrupted images.

Frequency Scrambling. To thwart the manual exposure attack, we design a *frequency scrambling* mechanism, which packs multiple waveforms with randomly selected frequencies within each image frame duration. Since the camera exposure time t_e is always fixed within each frame, no single t_e can circumvent all the frequency components.

Illumination Intensity Randomization. If the attackers can repetitively capture a static scene for a sufficiently long duration, they may eventually find at least one clean version for each row across all frames, thus *recovering* the image. LiShield does not guarantee to completely protect against such brute-force attacks. However, it can increase the number of frames needed for successful image recovery, so that the attack becomes infeasible unless the camera can stay perfectly still over a long period of time, during which the attackers may have already been discovered by the owners of the physical space. LiShield achieves this goal by employing illumination intensity randomization, where it randomly switches the magnitude of each ON period across multiple predefined levels, which enlarges the attacker's search space.

3 SCENE RECOVERY WITH AUTHORIZED CAMERAS

To allow authorized users to capture the scene while maintaining protection against unauthorized attackers, we need to impose additional constrains on the LED waveform. LiShield's solution leverages a secure side channel (*e.g.* visible light communication [2] or Wi-Fi) between authorized users and the smart LED, which conveys secrete information such as frame timing and waveform parameters. Such information can be protected by existing encryption algorithms and systems, which are already mature and thus beyond the scope of this paper.

3.1 Authorized Video Recording in Dynamic Scene

To authorize a camera to capture a dynamic scene, each individual frame within the video must be recoverable. To achieve this, the authorized camera needs to convey its exposure time setting t_e^u to the smart LED via the secure side channel, and synchronize its clock (for controlling capturing time) with the smart LED's clock (for controlling the waveform). Recall that the camera can evade the striping effects by setting exposure (Sec. 2.3). So to authorize the user with specific exposure, LiShield simply needs to set its flickering frequency within each frame. Meanwhile, when the authorized camera is not recording at its maximum possible rate (*e.g.*, a 30 fps camera recording at 25 fps), there will be an interval (*i.e.*, inter-frame gap) where the camera pauses capturing. LiShield packs random flickering frequencies into the inter-frame gap, so as to achieve the same scrambling effect as described in Sec. 2.3, without compromising the authorized capturing.

3.2 Static Scene Recovery

When the target scene is static, it requires the authorized user to capture a few complementary frames to recover the scene. Meanwhile, frequency and intensity randomization (Sec. 2.3) can still be employed in each frame to ensure robustness. While it does require recording a very short video, the process is extremely short (200ms at most) and barely noticeable to the authorized users, while an out-of-sync attacker will still receive corrupted images that cannot reconstruct the original scene even after combined.

4 AUTOMATIC PHYSICAL WATERMARKING FOR PRIVACY ENFORCEMENT

High-intensity ambient light sources (*e.g.* sunlight, legacy lighting, flash lights) can create strong interference to LiShield's illumination waveform, degrading the contrast by adding a constant intensity to both the bright and dark stripes, which may weaken LiShield's protection. In such scenarios, LiShield degrades itself to a *barcode mode*, where it embeds barcode in the physical scene to convey privacy policies. The barcode forms low-contrast stripes, which may not fully corrupt the images of the scene, but can still be detected by online photo-distributing hubs (*e.g.*, social website servers) who automatically enforce the policies, without cooperation of the uploader or evidence visible by naked eye.

LiShield's barcode packs multiple frequencies per RGB channel in every image (or in every frame of a video) following Sec. 2.3, but aims to map the *ratios between frequencies* into digital information. Suppose LiShield embeds two waveforms with frequencies F_0 and F_1, it chooses the two frequency components such that F_1/F_0 equals to a value R_p well known to the policy enforcers. In other words, the presence of R_p conveys "no distribution/sharing allowed". This encoding mechanism is robust against camera settings.

5 IMPLEMENTATION

Testbed setup. We implemented smart LED prototype, and the target scenes containing 5 capture-sensitive objects. We mount the LED inside a diffusive plastic cover similar to conventional ceiling light covers. We use a programmable motor to hold the camera of Nexus 5 and control its distance/orientation, in order to create static or dynamic scene setup in a repeatable manner. Besides, we employ

the CIEDE2000 [5] to compute the degradation of the images' color quality when the RGB LED is used.

6 EXPERIMENTAL EVALUATION

6.1 Effectiveness of Physical Scene Disruption

We first verify LiShield's basic protection scheme (Sec. 2) with 5 static scenes, monochrome LEDs, and OOK waveform without frequency randomization. Without LiShield, the measured image quality stays high, with PSNR> 30 dB and CW-SSIM> 0.9. Despite the use of a basic configuration, LiShield degrades the image quality by 3 to 10 dB for PSNR and 0.25 to 0.45 for CW-SSIM. In addition, *different scenes suffer from different levels of disruption*, depending on the scene's structure and reflection rate. **Impact of RGB color distortion.** We further verify the color-distortion impact when the RGB flickering is turned on. The color distortion makes an additional independent impact. The corresponding CIEDE2000 metric escalates up to 45, way beyond the *human-tolerable threshold 6* [5]. This implies *the scene is no longer considered viewable by average viewers.* **Impact on dynamic scenes.** To create a dynamic scene, we use the motor to rotate the smartphone, creating relative motion at three different speeds (45, 100 and 145 degrees/second). our result indicates that *dynamic scene experiences worse quality under LiShield* due to motion blur (PSNR<6, CW-SSIM<0.1). Thus, *dynamic objects further decrease the adjustment range of exposure time and make manual exposure attack more ineffective.* **Impact of device heterogeneity.** We cross-validate the impact of LiShield on 10 common smartphone cameras (including Android and iOS OS). Result shows the image quality varies slightly (CW-SSIM is 0.2∼0.6), due to different sampling rates across devices resulting in stripes of different widths. However, *LiShield's protection mechanism works across typical smartphone camera models.*

6.2 Effectiveness of User Authorization

We developed an app (Sec. 5) that can authorize the camera, and then recover the scene following Sec. 3. From results we can see the authorized user has much higher quality (PSNR=25dB, CW-SSIM=0.98 in average) compared with attacker (PSNR = 10dB, CW-SSIM = 0.6 in average). Thus *LiShield's authorization scheme is effective in unblocking specific users while maintaining protection against attackers.*

6.3 Effectiveness of Barcode Embedding

Our experiment shows that detection rate for barcodes (with 3 frequencies per RGB channel) is around 98% with or without manual exposure attack, while maintaining less than 5% false positive rate. We conclude that *LiShield's barcode detector provides reliable detection.* An attacker may post-process the image in attempt to remove the watermark. However, the attacker will have to deform most parts of the image, which greatly reduces the image quality and makes the attack nonviable.

6.4 Robustness and Counteracting Attacks

Manual exposure attack. Our experiment shows that although the image quality first increases with exposure time (CW-SSIM=0.4), it drops sharply as overexposure occurs (CW-SSIM=0.1). Therefore, *LiShield traps the attacker in either extremes by optimizing the waveform* (Sec. 2.2), and *thwarts any attempts through exposure time configuration.* **Multi-frame attack.** We recovered scene's quality under the multi-frame attack. When a tripod is used, CW-SSIM remains low at 0.5 using 1000 frames. We also ask 5 volunteers to hold the smartphone as stable as they can on a table, but the quality is even lower (PSNR is 15 dB and CW-SSIM is 0.3 using 1000 frames), because it is impossible to completely avoid dithering with hands even with anti-shake technology. **Image recovery processing attack.** In our experiments, the denoising and debanding methods fail to improve the quality significantly (PSNR <15 dB, CW-SSIM <0.6, CIEDE2000 >30). More advanced computer vision techniques may provide better recovery, but even they will not recover the *exactly original scene* since information is already lost at capture time. **Impact of ambient light.** We evaluate LiShield's performance under different types of ambient lights to verify LiShield's robustness. The stripes are almost completely removed under direct sunlight due to its extremely high intensity. However, image quality degradation still works under diffused sunlight and office light (*i.e.*, CW-SSIM<0.5). Flash light can increase the quality slightly thanks to its close distance to the scene, but the improvement is marginal and far from unprotected. For barcoding, the detection rate remains larger than 85%. Thus, *LiShield is robust against ambient light.* **Impact of distance.** We vary the distance between camera and *a single LED* from 1 m to 3 m. Even under 3 m, CW-SSIM is around 0.3 (way below 0.9) and the quality only increases slightly with distance. The barcode detection rate remains 90% with 3 m distance. Thus, *LiShield's working range can cover most of common applications with only a single smart LED.* With multiple LEDs, LiShield's coverage can be scaled up just like normal lighting.

7 CONCLUSION

Privacy protection in passive indoor environment has been an important but unsolved problem. In this paper we propose LiShield, which uses smart-LEDs and specialized intensity waveforms to disrupt unauthorized cameras, while allowing authorized users to record high quality image and video. We implemented and evaluated LiShield under various representative scenarios, which demonstrates LiShield's effectiveness and robustness. We consider LiShield as a first exploration of automatic visual privacy enforcement and expect it can inspire more research along the same direction.

REFERENCES

[1] Mauro Barni. 2006. *Document and Image compression.* CRC press.

[2] Christos Danakis, Mostafa Afgani, Gordon Povey, Ian Underwood, and Harald Haas. 2012. Using a CMOS camera sensor for visible light communication. In *Proc. of IEEE Globecom Workshops.*

[3] Yi Jiang, Ke Zhou, and Sheng He. 2007. Human visual cortex responds to invisible chromatic flicker. *Nature Neuroscience* 10, 5 (2007), 657–662.

[4] C. K. Liang, L. W. Chang, and H. H. Chen. 2008. Analysis and Compensation of Rolling Shutter Effect. *IEEE Transactions on Image Processing* 17, 8 (2008).

[5] M Ronnier Luo, Guihua Cui, and B Rigg. 2001. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application* 26, 5 (2001), 340–350.

[6] Alaeddin Nassani, Huidong Bai, Gun Lee, and Mark Billinghurst. 2015. Tag It!: AR Annotation Using Wearable Sensors. In *SIGGRAPH Asia Mobile Graphics and Interactive Applications.*

[7] Social Pilot. 2016. 125 Amazing Social Media Statistics You Should Know. (2016). https://socialpilot.co/blog/125-amazing-social-media-statistics-know-2016/

[8] Chi Zhang and Xinyu Zhang. 2016. LiTell: Robust Indoor Localization Using Unmodified Light Fixtures. In *Proc. of ACM MobiCom.*

[9] Lan Zhang, Cheng Bo, Jiahui Hou, Xiang-Yang Li, Yu Wang, Kebin Liu, and Yunhao Liu. 2015. Kaleido: You Can Watch It But Cannot Record It. In *ACM International Conference on Mobile Computing and Networking (MobiCom).*