OPTICS

Imaging privacy threats from an ambient light sensor

Yang Liu[1,2]*, Gregory W. Wornell[1,2,3], William T. Freeman[1,2], Frédo Durand[1,2]*

Embedded sensors in smart devices pose privacy risks, often unintentionally leaking user information. We investigate how combining an ambient light sensor with a device display can capture an image of touch interaction without a camera. By displaying a known video sequence, we use the light sensor to capture reflected light intensity variations partially blocked by the touching hand, formulating an inverse problem similar to single-pixel imaging. Because of the sensors' heavy quantization and low sensitivity, we propose an inversion algorithm involving an ℓ_p-norm dequantizer and a deep denoiser as natural image priors, to reconstruct images from the screen's perspective. We demonstrate touch interactions and eavesdropping hand gestures on an off-the-shelf Android tablet. Despite limitations in resolution and speed, we aim to raise awareness of potential security/privacy threats induced by the combination of passive and active components in smart devices and promote the development of ways to mitigate them.

INTRODUCTION

Privacy is one of the major issues raised by the prevalence of smart devices, such as mobile phones, watches, tablets, and televisions (*1*). Embedded sensors enable tremendous capabilities, but, unfortunately, they also increase the risk of leaking users' information. For example, embedded front cameras enable selfies and video conferences but can leak a user's facial information if not managed properly (*2*). Permission control can provide effective privacy protection by allowing users to manage access to sensors (e.g., cameras) and data (*3*, *4*). However, some sensors are considered "low risk" and can be accessed directly without any permissions or privileges. For example, the ambient light sensor is used to measure the illuminance of the surrounding light and to adjust the brightness of the screen automatically. The ambient light sensor needs to be always on for functionality and is traditionally considered to be of low risk because it provides a single value and does not appear to enable imaging of the environment (*5*–*10*).

Previous work mainly focuses on the privacy threats from sensors, which acquire data passively. These sensors can be divided into two categories: permission-required and permission-free sensors (*4*). Permission-required sensors are usually of high privacy risk because they convey direct video (*2*, *11*), audio (*12*), and location information and their access is carefully controlled or even physically blocked, e.g., by the camera/microphone blocker. Not until recently have permission-free sensors, such as motion (*13*), light (*5*–*7*, *9*, *10*), and proximity (*8*) sensors, received attention for revealing privacy threats. The ambient light sensor, in particular, has been shown to enable the inference of keystrokes on a virtual keyboard (*5*), the classification of content shown on a nearby display (*7*, *9*), and the determination of visited websites or even images by extracting one bit of information at a time (*10*). These usually involve user-specific data for behavior regression (*5*, *7*, *9*) and resolve limited information (*6*, *10*).

Here, we combine an active component—viz., a display screen—with the passive ambient light sensor and demonstrate the capture of images of the environment in front of the screen without access to the camera. The ubiquity of the combination of screens and ambient light sensors makes it crucial to understand its privacy risks and to revisit whether it should be considered a low-risk configuration.

We argue that the ambient light sensor can enable imaging if one uses the screen as a controllable active source of illumination displaying a known video sequence. The ambient light sensor measures the corresponding intensity variation of light reflected off or blocked by the scene (Fig. 1A). These sequential measurements and the corresponding known illumination sequence form a linear inverse problem, which can be solved to reconstruct an image from the perspective of the screen (Fig. 1B). Here, the problem is linear only when the measurements are not quantized with full precision of floating-point numbers. However, the ambient light sensor is of low sensitivity (at 1 lux level), and the contribution of screen fluctuation is heavily quantized to $\leq$5 bits per measurement. This type of imaging inverse problem is known as ghost imaging (*14*–*17*) or single-pixel imaging (*18*–*22*), which was considered to be a quantum effect (*23*), was independently explored as dual photography (*18*, *24*), and can be accelerated by compressive sensing (*19*–*21*). A similar idea using arrays of light-emitting diodes as virtual sensors has also been explored in the internet of things community for skeleton posture (*25*) and hand pose (*26*) estimation. However, it has not been shown in any privacy settings. The imaging capability that we explore is a form of dual photography (*18*, *24*), where Helmholtz reciprocity (*27*, *28*) indicates that the flow of light can be computationally reversed to swap out the sensor (ambient light sensor) and the illumination (screen). In our case, the light path from the screen to the scene and then to the ambient light sensor (primal path) can be reversed, resulting in a path from the ambient light sensor to the scene and eventually to the screen (dual path). The primal configuration (Fig. 1A), where the sensor has a single pixel and the illumination has good resolution, can be interpreted by its duality, where the pixelated screen works as a virtual sensor and the ambient light sensor as a virtual light source (Fig. 1B).

Capturing images of the scene in front of the screen using an ambient light sensor involves two challenges. First, quantization noise is substantial due to the combination of limited light reflected off the scene and the integer-quantized outputs (in lux) of the ambient light sensor. Some early Android devices, such as Google Nexus phones, give floating-point outputs. We consider the more recent and common case, where smart devices with ambient light sensors

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: yliu12@mit.edu (Y.L.); fredo@mit.edu (F.D.)

only have integer outputs with the same unit lux. To address this, we combine the optimization-based reconstruction and the deep learning–based image denoiser as a prior (*29*, *30*), where the iterative inversion procedure and the strong image prior remove the measurement noise and retain the signal content accurately (Fig. 2).

The second challenge is that there is no lens between the scene and the screen. The screen serving as the virtual sensor "sees" a highly blurred image of a general scene at a distance (fig. S2A). This "lensless" scenario inherently affects the fidelity with which the environment can be imaged in this way. For objects that are in contact with or close enough to the screen, lenses are not required. This is because each pixel of the virtual sensor sees the one-to-one mapping blocked by the scene pixel (Fig. 3). For general objects, this is closely related to the non–line-of-sight (NLOS) imaging problem by observing a blank wall with an ordinary camera (*31*–*37*), where there is no lens between the target NLOS scene and the observing wall. In contrast, we show that the deformation of an occluder to form an accidental pinhole (*31*) between the scene and the screen. As a result, we emphasize on revealing imaging information leakage of the ambient light sensor in contact with the screen by eavesdropping the user interaction of hand gestures. We further discuss ways to mitigate these light sensor–based imaging privacy threats by considering access to the screen and the ambient light sensor; the brightness of the screen; and the precision, refresh rate, and location of the ambient light sensor.

RESULTS

Imaging inverse problem

Imaging forward model

Revealing the imaging privacy threats from an ambient light sensor can be formed as a quantized version of linear inverse problem

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is the ambient light sensor outputs as a vector, m is the number of outputs; $\boldsymbol{x} \in \mathrm{R}^n$ is the target signal with two-dimensional image vectorized; n is the total number of pixels of the target image; $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the sensing matrix with each row as the vectorized pattern on the screen; $\boldsymbol{b} \in \mathbb{R}^n$ is the noise induced by the sampling process with the noise of the ambient light sensor output, surrounding light, and the fluctuation of the screen brightness as dominant sources; and $\mathcal{Q}(\cdot)$ is a uniform and scalar quantizer, i.e., $\mathcal{Q}(\boldsymbol{z}) = \Delta \cdot \lfloor \frac{z}{\Delta} \rfloor$ with Δ being the quantization step size and $\lfloor \cdot \rfloor$ denoting the floor function. Note that subsampling when the number of measurements is less than that of the number of total pixels of the target image, i.e., $m < n$, could reduce the acquisition time for a single image. Likewise, we use the "zig-zag" strategy for subsampling Walsh-Hadamard transform domain (*38*), and the proposed inversion algorithm inherently works for subsampled sensing matrix (see fig. S11 for comparison of subsampled images with sampling ratios from 50 to 10%). A complete imaging model taking the occluder and the receiving/transmitting angle ranges of the ambient light sensor and the screen into account can be found in fig. S1.

Imaging inverse recovery

Intuitively, if the measurement $\boldsymbol{y}$ is not quantized and the sensing matrix $\boldsymbol{A}$ is an orthogonal matrix, e.g., identity matrix, and no noise is engaged, then Eq. 1 can be solved by its direct inverse transform, i.e., $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$. Because quantization noise is the dominant corruption source of the measurement, we use the ℓ_p-norm or basis pursuit dequantizer of moment p (*39*) to constrain the data-fidelity manifold. Here, $2 \le p < \infty$, and ℓ_∞-norm better respects the uniform quantization consistency (*39*, *40*) than ℓ_2-norm. The consistency (*39*, *40*) can be rewritten as $\mathcal{Q}(\boldsymbol{z}) = \boldsymbol{y} \Leftrightarrow \|\boldsymbol{y} - \boldsymbol{z}\|_\infty \le \frac{\Delta}{2}$ with Δ being the quantization step size of $\mathcal{Q}(\cdot)$. Then, we have the ℓ_p-norm constrained imaging inverse recovery problem

$$\hat{\boldsymbol{x}} = \mathrm{argmin}_{\boldsymbol{x}} \underbrace{\frac{1}{2} \| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \|_p^p}_{\text{data – fidelity term}} + \lambda \cdot \underbrace{R(\boldsymbol{x})}_{\text{prior term}} \tag{2}$$

where $\|\mathbf{z}\|_p^p = (\sum_{i=1}^m |z_i|^p)$ is the ℓ_p-norm, $R(\cdot)$ is the regularization/prior term imposed on the signal, which comes from the prior knowledge of the signal, and λ is the factor balancing the data-fidelity term $\frac{1}{2} \| \mathbf{y} - \boldsymbol{A}\boldsymbol{x} \|_p^p$ and the prior term $R(\boldsymbol{x})$.

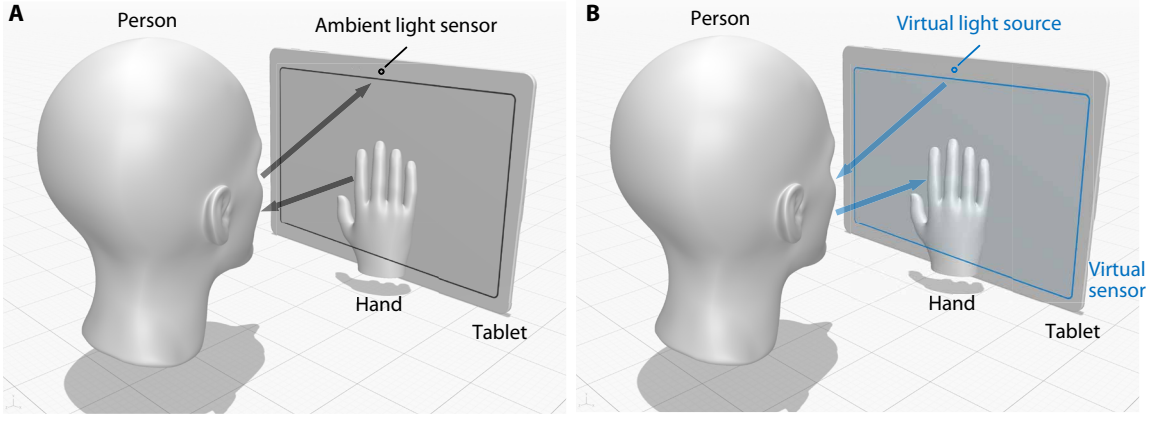

Fig. 1. Imaging setup with primal and dual configurations. (**A**) Primal configuration where the screen displays a sequence of patterns and the ambient light sensor receives the light first partially blocked by the interacting hand and then reflected from the human face. (**B**) Dual configuration where the ambient light sensor works as the virtual point light source and the pixelated screen as the virtual sensor. No lens is required between the screen and the scene to form an image on the virtual sensor, because the interacting hands create in-contact shadows on the virtual sensor, forming one-to-one mapping between the target scene pixel and the sensor pixel.

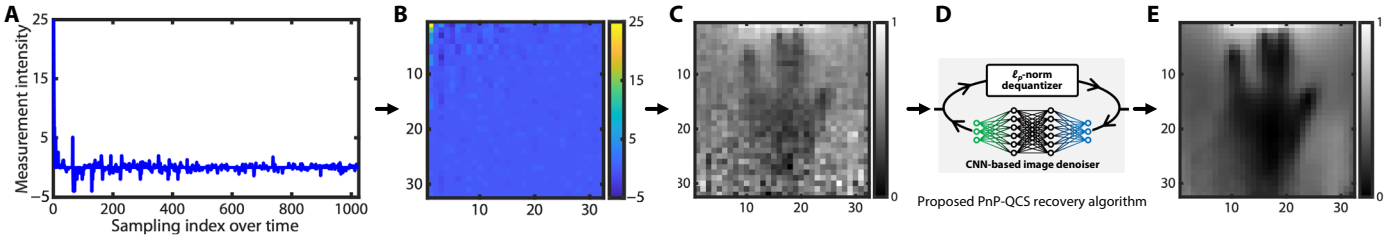

Fig. 2. Inversion procedure for revealing imaging privacy threats from an ambient light sensor. (**A**) The ambient light sensor measurements by displaying a sequence of full Walsh-Hadamard bases on the screen with an in-contact hand pose. The pixel resolution is 32 × 32, and a total of 1024 measurements are collected with acquisition time of 17 min. The light sensor value range is [0, 25] lux in integers. (**B**) Sensor outputs reshaped as a two-dimensional transform domain according to sampling indices. (**C**) Recovered image by applying direct inverse transform. (**D**) The iterative inversion process using the CNN-based image denoiser as the prior. PnP-QCS, plug-and-play quantized compressive sensing. (**E**) The final recovered image using the proposed inversion algorithm.

Then, we use the alternating direction method of multipliers (ADMM) (*41*) to solve this optimization problem by rewriting Eq. 2 in the alternating direction form, that is

$$(\hat{\boldsymbol{x}},\hat{\boldsymbol{z}}) = \operatorname{argmin}_{(\boldsymbol{x},\boldsymbol{z})} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_p^p + \lambda R(\boldsymbol{z}) \quad \text{subject to} \quad \boldsymbol{x}=\boldsymbol{z} \tag{3}$$

Note that Eq. 3 separates the constraints of the data-fidelity term and the prior term on two variables $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. The augmented Lagrangian of Eq. 3 can be written as

$$\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{w}) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_p^p + \lambda R(\boldsymbol{z}) + \boldsymbol{w}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{x}-\boldsymbol{z}\|_2^2 \tag{4}$$

where $\boldsymbol{w}$ is the multiplier or dual variable and ρ is the penalty parameter. After substituting $\boldsymbol{w}/\rho$ with the scaled dual variable $\boldsymbol{u}$, Eq. 4 can be rewritten as

$$\mathcal{L}_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_p^p + \lambda R(\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{x}-\boldsymbol{z}+\boldsymbol{u}\|_2^2 + \text{const.} \tag{5}$$

with const. being a constant independent of $\boldsymbol{x}$ and $\boldsymbol{z}$. Following the ADMM method, Eq. 5 can be solved by iteratively updating $\boldsymbol{x}$, $\boldsymbol{z}$, and $\boldsymbol{u}$ in an alternating manner. Using the proximal operator (*42*), that is $\operatorname{prox}_h(v) = \operatorname{argmin}_x h(x) + \frac{1}{2}\|x-v\|_2^2$, the iterative procedure can be written as

$$\boldsymbol{x}^{k+1} = \operatorname{prox}_{f/\rho}(\boldsymbol{z}^k - \boldsymbol{u}^k) \tag{6}$$

$$\boldsymbol{z}^{k+1} = \operatorname{prox}_{\lambda R/\rho}(\boldsymbol{x}^{k+1} + \boldsymbol{u}^k) \tag{7}$$

$$\boldsymbol{u}^{k+1} = \boldsymbol{u}^k + (\boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1}) \tag{8}$$

where $f(\cdot)$ is the data-fidelity term, i.e., $f(x) = \frac{1}{2}\|\mathbf{y}-\boldsymbol{A}\boldsymbol{x}\|_p^p$.

Note that Eq. 6 is an ℓ_p-norm minimization problem with a closed-form solution only for the quadratic or ℓ_2 case, i.e., $\boldsymbol{x}^{k+1} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \rho\,\boldsymbol{I})^{-1}[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} + \rho\,(\boldsymbol{z}^k - \boldsymbol{u}^k)]$. Inspired by this, we apply iteratively reweighted least squares (*43*, *44*) to solve the general case of ℓ_p-norm minimization ($p > 2$). The approach is to replace the ℓ_p-norm objective $f(\boldsymbol{x})$ with a reweighted ℓ_2-norm, where the weights are corresponding values from previous update. That is $\|\boldsymbol{q}\|_p^p = \sum_i |q_i'|^{p-2}\cdot|q_i|^2 = \sum_i w_i \cdot |q_i|^2 = \boldsymbol{q}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{q}$, where $\boldsymbol{W}$ is a diagonal matrix with previous values $\boldsymbol{q}'$ as the diagonal elements. Therefore, for each substep of solving the ℓ_p-norm minimization problem, we have a reweighted version of quadratic form with

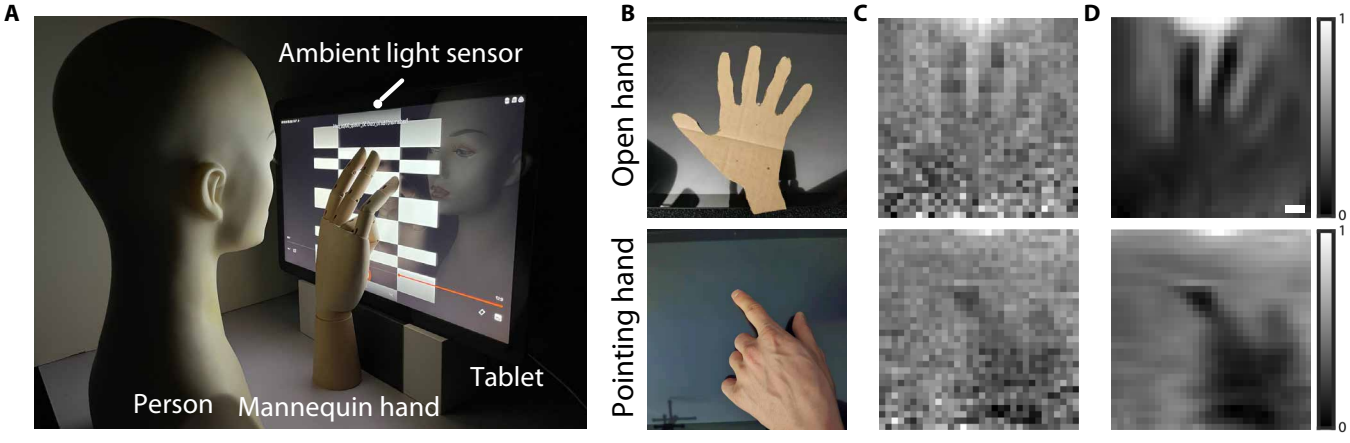

Fig. 3. Imaging privacy threats revealing touch detection in front of the screen. (**A**) Experimental setup where in-contact touch is revealed. (**B**) Target touching hand to be resolved. (**C**) Recovered images by direct inverse transform, where the noise in images comes from the measurement process, especially the quantization noise of the ambient light sensor. (**D**) Final recovered images by the proposed inversion algorithm. Recovered images are acquired by displaying a sequence of full Walsh-Hadamard bases with corresponding pixel resolutions. The pixel resolution is 32 × 32, and each acquisition takes 17 min. Scale bars, 2 cm.

closed-form solution, and the corresponding weights can be updated iteratively, that is the subiteration involved for solving Eq. 6

$$\boldsymbol{x}_{(t+1)} = \left[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}_{(t)}\boldsymbol{A} + \rho\boldsymbol{I}\right]^{-1}\left[\boldsymbol{A}^{\mathrm{T}}\boldsymbol{W}_{(t)}\boldsymbol{y} + \rho(\boldsymbol{z}^{k} - \boldsymbol{u}^{k})\right] \tag{9}$$

$$\boldsymbol{W}_{(t+1)} = \mathrm{diag}\left[\frac{n^{1/p}\,|\,\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(t+1)}\,|}{\|\,\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(t+1)}\,\|_{p}}\right]^{p-2} \tag{10}$$

where diag($\boldsymbol{v}$) returns a diagnoal matrix with diagonal elements being all elements of vector $\boldsymbol{v}$, $|\boldsymbol{v}|$ returns the absolute value of each element in vector $\boldsymbol{v}$, Eq. 10 is a weight-normalized version of diag[$|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_{(t+1)}|^{p-2}$], and the initial weight matrix is identity, i.e., $\boldsymbol{W}_{(0)} = \boldsymbol{I}$. For orthogonal bases as displayed patterns, $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is diagonal. In addition, recall from Eq. 10 that $\boldsymbol{W}_{(t)}$ is also a diagonal matrix. According to the matrix inversion lemma (*45*), the substep $\boldsymbol{x}_{(t+1)}$ can be updated without solving a matrix inversion in each subiteration (*46*, *47*), that is

$$\boldsymbol{x}_{(t+1)} = (\boldsymbol{z}^{k} - \boldsymbol{u}^{k}) + \boldsymbol{A}^{\mathrm{T}}\left[\boldsymbol{y} - \boldsymbol{A}(\boldsymbol{z}^{\mathbf{k}} - \boldsymbol{u}^{\mathbf{k}})\right] \oslash \left[\mathrm{diag}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) + \rho \oslash \mathrm{diag}(\boldsymbol{W}_{(t)})\right] \tag{11}$$

where $\oslash$ is element-wise division and diag(·) returns the diagonal elements of a square matrix. Detailed derivations can be found in section S4.

The update on $\boldsymbol{z}^{k+1}$ is regularizing ($\boldsymbol{x}^{k+1} + \boldsymbol{u}^{k}$) on the signal prior domain and can be viewed as a denoising problem without explicitly expressing R(·), that is

$$\boldsymbol{z}^{k+1} = \mathcal{D}_{\hat{\sigma}_k}(\boldsymbol{x}^{k+1} + \boldsymbol{u}^{k}) \tag{12}$$

where $\mathcal{D}_{\sigma}(\boldsymbol{v}) = \mathrm{prox}_{\sigma^2 R}(\boldsymbol{v}) = \frac{1}{2\sigma^2}\|\boldsymbol{x} - \boldsymbol{v}\|_2^2 + R(\boldsymbol{x})$ is a denoising function with the noisy signal v and the noise level σ as inputs and $\hat{\sigma}_k = \sqrt{\lambda/\rho}$ is the estimated noise standard deviation (SD) in each iteration. This framework is called plug-and-play priors for inverse problems (*29*) because a denoiser can serve as the prior for the signal and state-of-the-art results can be achieved with a simple plug-in (*30*). With the advances of deep neural networks (*48*), convolution neural network (CNN)–based image denoisers are efficient in terms of the long-standing trade-off of quality and speed and flexible enough to deal with a wide range of noise levels (e.g., noise SD [0, 70] at a scale of 255) with a single pretrained network (*49*, *50*). These flexible CNN–based denoisers serve as the prior for natural images. Inspired by the success of deep denoisers for other high-dimensional compressive imaging problem (*47*), we propose to use a pretrained CNN–based image denoiser, i.e., the fast and flexible denoising convolutional neural network (*50*) as a prior and obtain fast and high-quality reconstruction compared to a hand-crafted prior and other state-of-the-art denoisers (see fig. S10 for comparison with other priors and algorithms). The code along with measurement data and experimental configurations is publicly available (*51*).

Imaging privacy threats

Touch detection in contact with the screen

We first demonstrate the imaging privacy threat of touch poses in contact with the screen, as shown in Fig. 3. An off-the-shelf Android tablet (Samsung Galaxy View2) embedded with a 17.3-inch screen (width of 38.3 cm and height of 21.5 cm) and an ambient light sensor near the front camera is used here. We put a mannequin person facing the screen at a distance of around 12 cm and apply a certain touch to the screen with a cardboard hand or a real human hand (Fig. 3A). We display a sequence of full Walsh-Hadamard transform bases (*52*, *53*) reshaped as two-dimensional with {+1, −1} values. A differential measurement strategy is used to circumvent the negative values (*38*). In particular, each Walsh-Hadamard basis H is implemented with the subtraction of two adjacent patterns $D^{+} = \frac{1}{2}(1 + H)$ and $D^{-} = \frac{1}{2}(1 + H)$, with 0 or 1 value for each pixel. The reflected light is collected by the ambient light sensor with integer values (between 0 and 25 lux in Fig. 2A). We set the effective refresh rate of the patterns on the screen as 2 Hz to match the speed of the ambient light sensor (around 10 to 20 Hz) and average the outputs within a time period of a pattern. The acquisition time for a pixel resolution of 32 × 32 is 17 min (Fig. 2 as well as the hand touches in Fig. 3). The ambient light sensor outputs are then reshaped as the two-dimensional transform domain according to the sampling index of each output (Fig. 2B). We use the direct inverse Walsh-Hadamard transform on the reshaped image and obtain the initial image of the touching hand (Fig. 2C). Last, we apply the proposed iterative process using a deep learning–based denoiser as the image prior for reconstruction (see Fig. 2D as well as Materials and Methods for details). Figure 3 (C and D) shows the resulting imaging privacy threats involving two types of hand touches in contact with the screen. We observe that spatial noise (mainly in the form of quantization noise in the ambient light sensor) is removed and that the scene content is retained accurately in the final recovered image (Figs. 2E and 3D). From the dual photography perspective (Fig. 1), touching hands cast shadows on the virtual sensor because the human face reflects light, and the hands, which are very close to the screen while touching, partially block light. The imaging privacy threat here is the leaking of information about physical interactions with the screen. In particular, we show that the cardboard open hand and the real pointing hand touching the screen (Fig. 3, A and B) can be reconstructed at reasonably good fidelity. Figure 3D shows that all five fingers are clearly resolved for the cardboard open hand, and the forefinger, thumb, and little finger are distinguishable for the real pointing hand.

Revealing hand gestures in contact with the screen

Furthermore, five sequences of typical touch gestures are revealed from the ambient light sensor of the 17.3-inch tablet, as shown in Fig. 4. To effectively obtain the touch gesture sequences, we match the effective refresh rate of the screen with the speed of the ambient light sensor (at ∼15 Hz) in terms of Nyquist sampling and obtain a pattern switching rate of 5 Hz (2.5× of acquisition speedup), where the upper bound (Nyquist sampling rate) is ∼7.5 Hz. Combined with the subsampling strategy by measuring the 50% low spatial frequency portion of the Walsh-Hadamard bases (see figs. S10 and S11 for details), we achieve a total of 5× speedup. As a result, we show sequences of touch gestures (one-finger slide, two-finger scroll, three-finger pinch, four-finger swipe, and five-finger rotate) with a frame interval of 3.3 min at pixel resolution of 32 × 32. Although the frame interval is still one to two orders of magnitude longer than the dwell time of the hand gesture, the revealed touch gestures are consistent with target user input on the screen and can potentially eavesdrop the user interaction. The bottleneck of the acquisition speed is the sampling rate of the ambient light sensor, as the frame interval is determined by $t \geq s/(2N \times N \times \gamma)$, where s is the sampling rate of the ambient light sensor, $N \times N$ is the target pixel resolution, and γ is the sampling ratio in the transform domain. With a faster ambient light

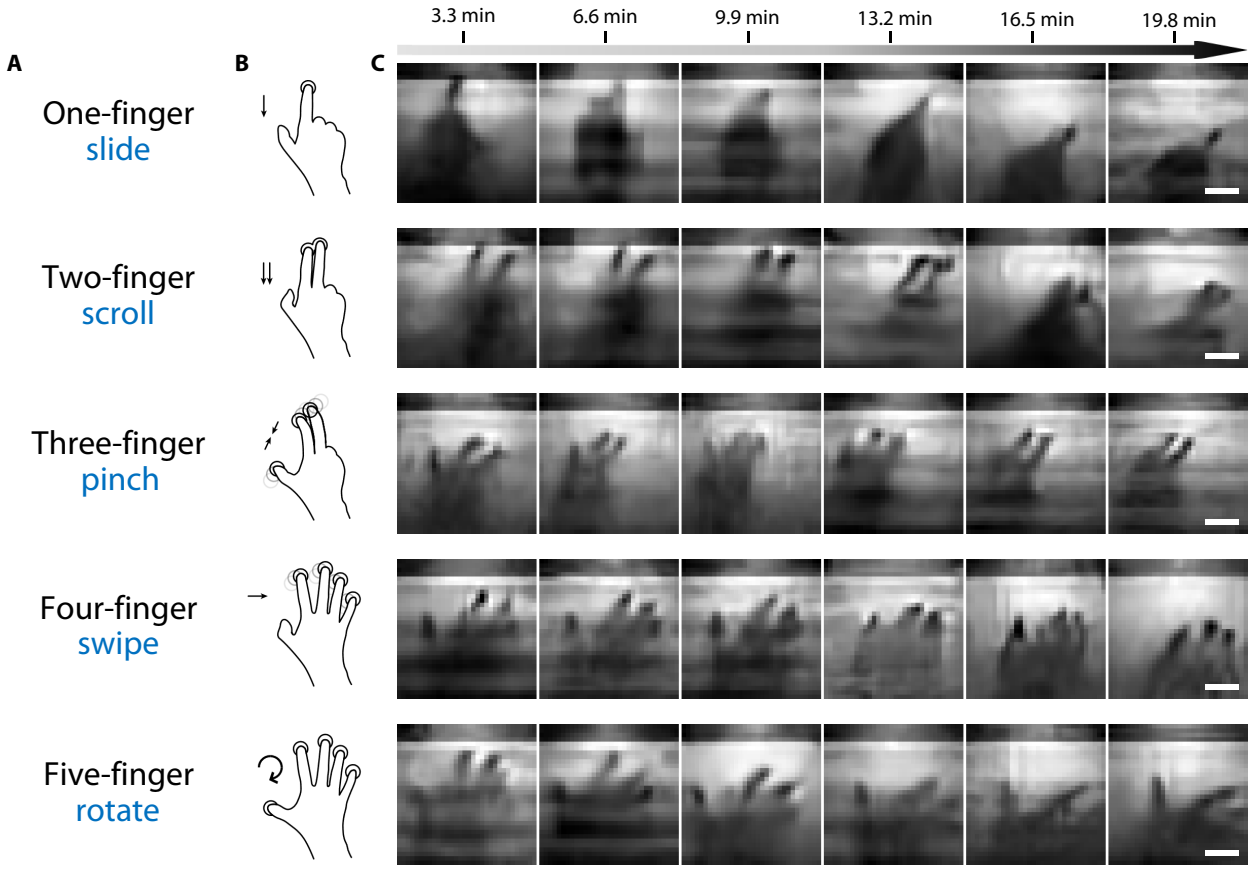

Fig. 4. Eavesdropping touch gestures from the ambient light sensor of a 17.3-inch tablet. (**A**) Gesture names. (**B**) Gesture depiction. (**C**) Recovered gesture sequences in front of the screen by the proposed inversion algorithm. Each frame is of 32 × 32 pixel resolution and acquired with a 3.3-min interval by displaying a sequence of half Walsh-Hadamard bases (low spatial frequency portion in a zigzag manner as shown in fig. S11) on the screen. Scale bars, 5 cm.

sensor and smaller pixel resolution and sampling ratio, the frame interval can be further reduced by one to two orders of magnitude, and real-time eavesdropping user interaction from an ambient light sensor would be an actual privacy threat.

Hand interaction leakage while watching a natural video

A potential abuse case is leaking the user's hand interaction while watching a natural video, such as a film and a short video, or even browsing. We explore this interaction leakage by displaying a modified video clip of Tom and Jerry (www.youtube.com/watch?v=cV6ucplpmbY/) on the screen, as shown in Fig. 5A, and form a quantized version linear inverse problem with a sensing matrix $\boldsymbol{A}$. Each row of the sensing matrix corresponds to the vectorized displaying frame on the screen, as shown in Fig. 5D. Note that, in Fig. 5D, we are showing the transpose of the sensing matrix $\boldsymbol{A}$ for better visualization. The ambient light sensor gathers the total intensity of light partially blocked by the interacting hand and reflected from the white surface. Here, we use a white surface instead of a mannequin head and reduce its distance to the screen to increase the light level received by the ambient light sensor. Therefore, the ambient light sensor is capable of distinguishing subtle intensity changes of adjacent video frames, as shown in Fig. 5E. We show the recovery procedure can resolve the fingers of the touch detection, as shown in Fig. 5C. This inversion result is based on the alternating projection method (*54*), as the Euclidean projection in ADMM struggles to get clean recovery for heavily ill-conditioned sensing matrices (*34*).

There are two key factors of the natural video affecting the recovery quality. The first one is the overall mean intensity, which directly affects the quantization level of the ambient light sensor. Natural videos with higher overall mean intensity benefit the recovery quality because it contributes to higher light level and consequently higher quantization level. The second factor is the condition number of the sensing matrix, which is a common metric for ill-conditioned linear inverse problems (*55*). A lower condition number at magnitude scale usually make the reconstruction easier as well. Empirically, this is characterized by high spatiotemporal variations of the video sequence. It is evident that the variance of light intensity in Fig. 5E is proportional to the spatiotemporal variation of the video sequence, as shown in Fig. 5D. Two examples are the slowly varying interval of sampling index from 2300 to 3000 and the rapidly varying interval of sampling index from 7000 to 8000.

Imaging quality factors

Two passes of imaging formation processes are involved considering obtaining an image of objects distant from the screen.

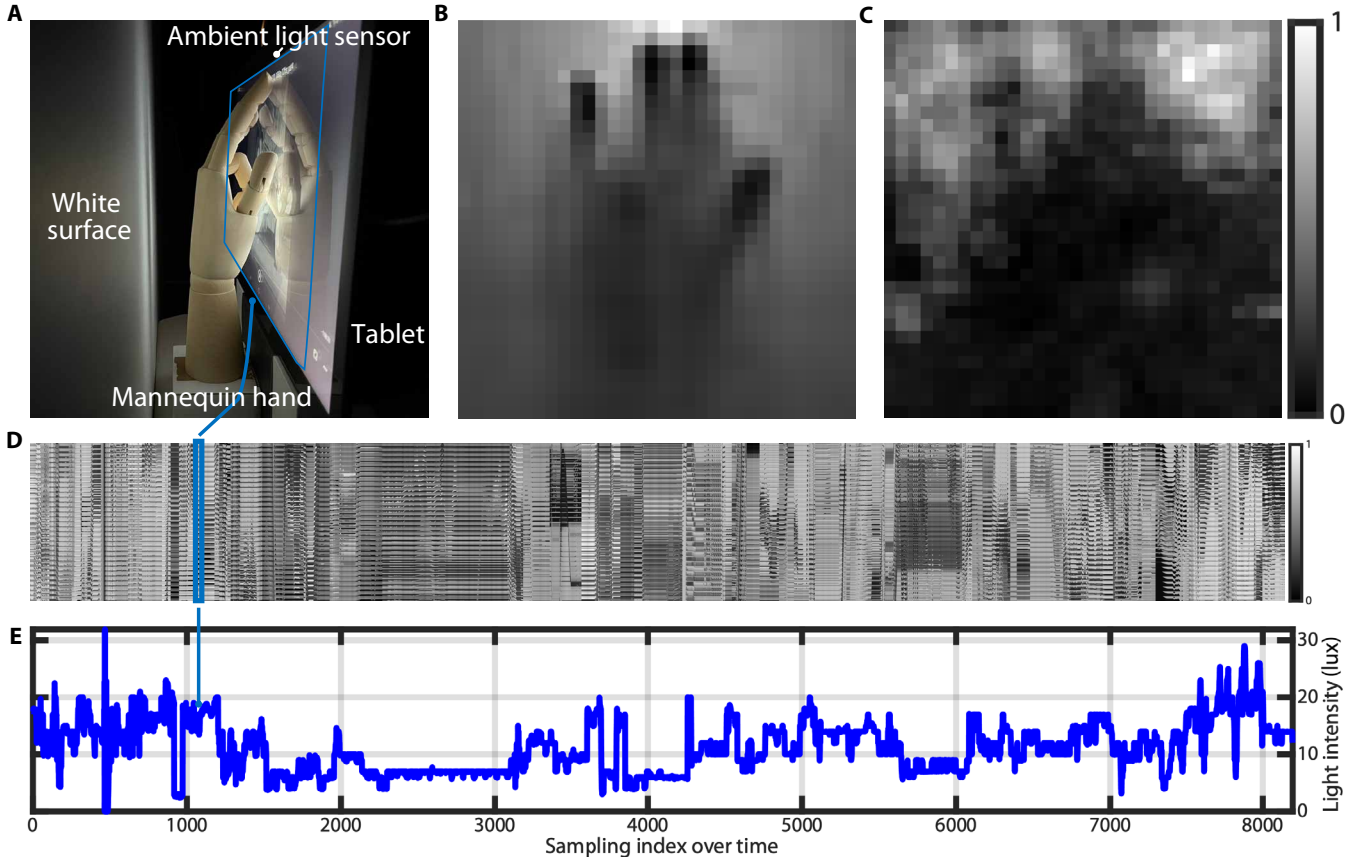

Fig. 5. Hand interaction leakage while watching a natural video. (**A**) Experimental setup with the screen displaying a modified video clip of Tom and Jerry. (**B**) Target reference touch gesture image. (**C**) Recovered touch gesture image. (**D**) Transpose of the sensing matrix ***A***, where each column corresponds to the vectorized displaying frame on the screen. (**E**) Light intensity measurements from the ambient light sensor, where each sampling index corresponds to each column (frame in the video) of (D). The video clip is modified to match the speed of the ambient light sensor. The total acquisition time is 68 min.

Privacy screen filter and angular response of screen pixel

From a dual perspective (Fig. 1B), the first pass is from the object to the screen. Because there is no lens between the object and the virtual image sensor (screen), the physical limit of the spatial resolution is determined by the receptive angle of image pixel, which corresponds to the angular response of the screen pixel (refer to the light transport imaging model in section S2). Modern screens optimize the viewing angle close to 180°, that is, the receptive angle of virtual image pixel is ~180°, and there will not be angular selectivity to the object pixel, resulting in a highly blurry image (fig. S2A) with almost no spatial information revealed. Privacy screen filters are designed to avoid unwanted content exposure to peers by limiting the viewing angle from 180° to 60°. This angular selectivity of privacy screen filter is not sufficient to bring back the spatial information of the object given the ratio of the screen and object-screen distance close to 1:1. Lenticular autostereoscopic displays, however, are one to two orders of magnitude more angular selective than privacy screen filters and can potentially reveal more spatial information of the object in this incoherent lensless setting.

The second pass is from the screen pixel to virtual image sensor pixel, which is the dual photography process. From the image forward model, we can see that limiting factors to image quality are the sensing matrix $\boldsymbol{A}$, measurement signal-to-noise ratio, and quantization level. Because we use orthogonal bases, the sensing matrix $\boldsymbol{A}$ is determined by the type of orthogonal bases and the sampling ratio in transform domain. Here, we discuss two major factors, which are the type of orthogonal bases and the quantization level. Detailed explanation about sampling ratio in the transform domain and background signal-to-noise ratio are provided in sections S6.1 and S8, respectively.

Displayed orthogonal bases on the screen

We compare five typical orthogonal bases, i.e., standard or canonical basis, Haar wavelet basis, discrete Fourier transform (DFT) basis, discrete cosine transform (DCT) basis, and Walsh-Hadamard basis in the context of imaging a map scene in contact with the screen shown in Fig. 6. The quantization level (4-bit) and measurement signal-to-noise ratio (30 dB) are set to mimic the level of quantization and background noise in real setup.

As shown in Fig. 6, Walsh-Hadamard basis combined with the proposed inversion algorithm performs the best both quantitatively and qualitatively among other bases even when the direct inverse transform is heavily noisy, e.g., Fig. 6 at pixel resolution of 64 × 64. Fine details are retained faithfully with noise pixel removed, which is also evident in real data (Fig. 3, C and D). Walsh-Hadamard basis outperforms Fourier basis and DCT basis here because correlated high-contrast scene and basis can result

in more discrepancy in this heavily quantized scenario. Standard basis and Haar wavelet basis struggle to give a meaningful recovery because all or the majority of the basis only carry one or a few pixels of the original image. As a result, individual measurements are below quantization step size contributing to all or most zeros in the transform domain.

Quantization levels of the light intensity

As shown in Fig. 7, the imaging capability is limited by the quantization level of the ambient light sensor. If the quantization level is reduced to as low as 4.0-bit (illuminance value [0, 15] lux; Fig. 7E), then it fails to resolve enough imaging information even with the proposed algorithm, because high spatial frequencies are lost during the sampling procedure [comparing the first row of Fig. 7E to Fig. 7 (A to D)]. Although the surrounding noise and averaging over time could partially circumvent the quantization issue (fig. S13), with a slow sampling speed of the ambient light sensor, this would not be effective.

DISCUSSION

We have demonstrated three types of imaging privacy threats. At this point, they may not be easy for attackers to leverage because of the long acquisition time (over 3 min) and the limited spatial resolution of the recovered images. For some scenarios, constrained configurations are required to form images, e.g., a particular mirror orientation for specular objects (fig. S6) and an occluder with centimeter-scale deformation for general objects (figs. S3 and S9). Note that, while our experiments rely on a specific known sequence of basis illumination, similar approaches could be deployed with more natural displayed sequences, such as from a movie or a game. A linear system can still be formed and inverted, although possibly with more challenging conditioning, as shown in Fig. 5 and fig. S12. This implies that people might not even be aware of being spied on while watching a film or playing a game. Detailed discussions of the magnitude of imaging privacy threats along with typical scenarios are in the Supplementary Materials. Nevertheless, our demonstrations confirm the reality of these threats.

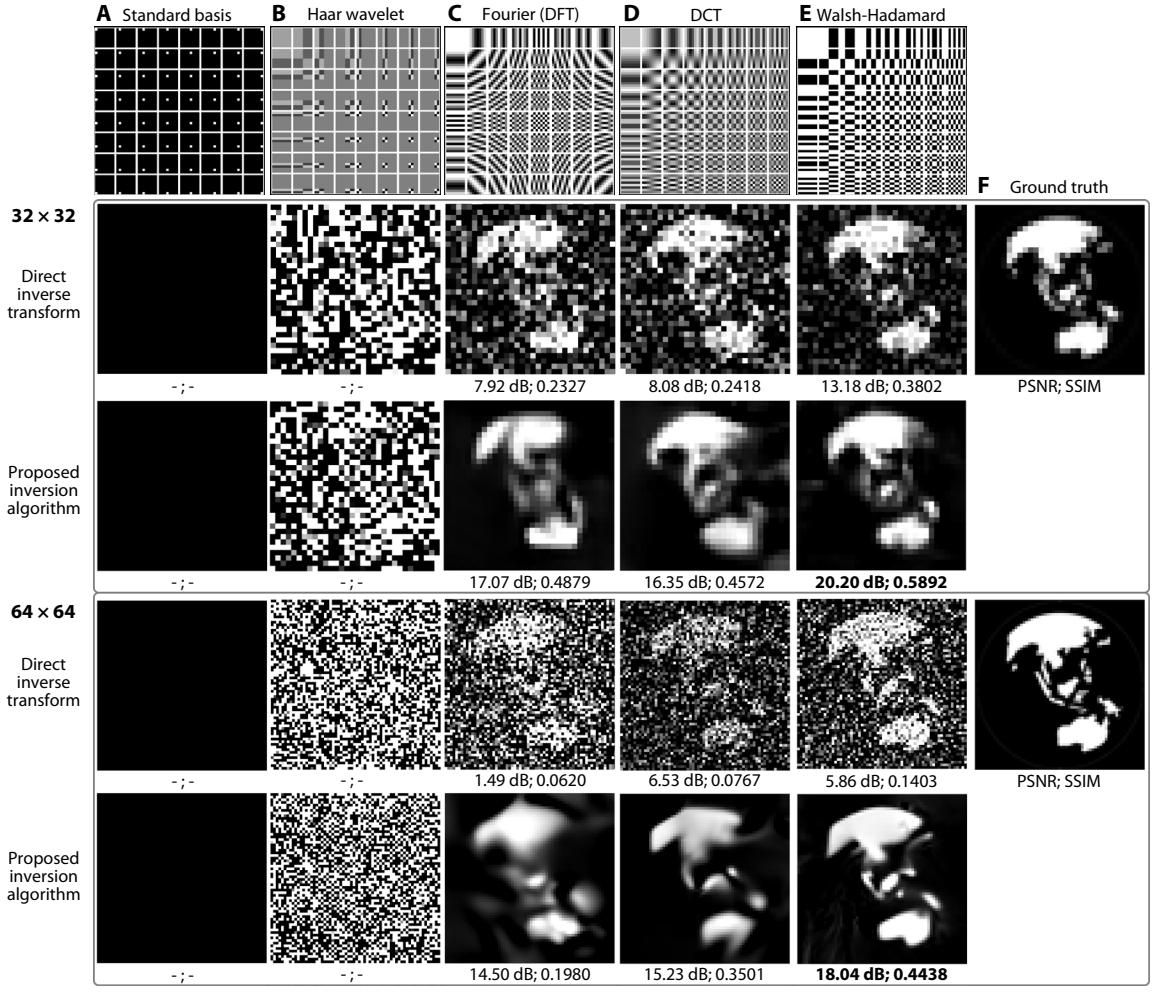

Fig. 6. Simulation results comparing five orthogonal bases with measurement parameters mimicking the setup (4-bit quantization or [0, 15] integer values and measurement signal-to-noise ratio of 30 dB). Quantitative indices, i.e., peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) (*57*), are shown on the bottom of reconstructed images, except for those far from ground truth image. Higher is better. (**A**) Standard basis. (Black image results are shown because individual measurements are below quantization step size, i.e., all zeros.) (**B**) Haar wavelet basis. (**C**) DFT basis. Center frequencies are shifted to the upper left corner for better visualization comparison. (**D**) DCT basis. (**E**) Walsh-Hadamard basis. (**F**) Ground truth.

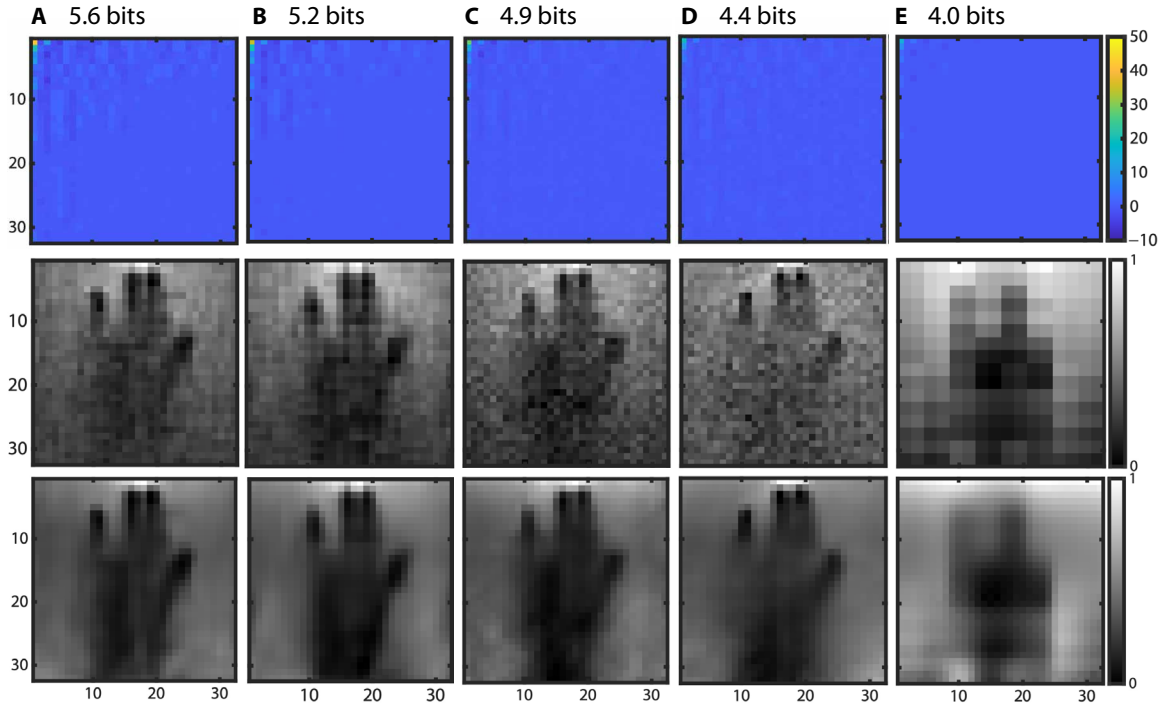

Fig. 7. The effect of reduced quantization levels of the ambient light sensor outputs. (**A** to **E**) Recovered images with reduced quantization levels (from 5.6 to 4.0 bit) of the ambient light sensor outputs. The first row shows the transform domain, the second and third rows give the results using direct inverse Walsh-Hadamard transform and the proposed inversion algorithm, respectively. All transform-domain and image-domain results are of 32 × 32 pixel resolution.

From the privacy perspective, the potential risks of revealing images of the scene, albeit limited to some particular scenarios, should be reduced from both the software and hardware sides. In both cases, adding any constraints on the display is unlikely because it is such a central component whose capabilities are hard to reduce, which is why we focus on the ambient light sensor.

The landscape of privacy risk of typical off-the-shelf smart devices in terms of software constraints is shown in Fig. 8. Given a fixed reasonable resolution and time configuration, i.e., 32 × 32 pixels and 20 min here, the privacy risk or the imaging capability is limited by the information budget (measured by bits per second) acquired by the light sensor. As seen in Fig. 8 (A and B), the separation line between safe/warning zones is showing an inverse proportion because the information budget is the product of the light sensor speed and measurement bit depth, which is characterized by the log ratio of maximum screen brightness and light sensor precision. That says that a large brighter screen with higher light sensor precision as well as higher sensor speed will have higher risk of revealing imaging privacy threats, such as large Android TVs. Another key factor is how useful each bit (or measurement) is. Intuitively, if frames in a video sequence are too similar to each other or if rows of the sensing matrix are highly correlated, then this will not provide enough information for solving every pixel in the linear system (after quantization). Therefore, Fig. 8B is showing a relaxed trend of imaging privacy threats. Following the usefulness argument of the measurement bits, if the smart device is put in a bright environment such as office light or sunlight, then the measurement bits will be less useful, i.e., the signal-to-noise ratio will go down, resulting in expanded low-risk/safe zone.

From the software side, the imaging threats that we have revealed suggest two mitigation strategies: tighten permission controls and reduce the information output by the sensor. Restricting access to the screen is probably not realistic, but the lack of permission to access the ambient light sensor may need to be rethought. Second, the precision and speed of the ambient light sensor should be reduced in its application programming interface. From the hardware side, the location of the ambient light sensor should not be directly facing the user. It could be on the side of smart devices, which could break the direct interaction of the screen and the light sensor, thereby reducing the privacy risks. From the dual photography perspective, the ambient light sensor works as a virtual point light source and would still be able to resolve an image because photons bounce around and the scene might be lighted up. However, given the brightness of the screen and the quantization level of the ambient light sensor, the scene is not bright enough to be seen by the virtual sensor.

Accordingly, we propose quantizing the sensor output more (e.g., at a step of 10 lux), reducing speed (e.g., 1 to 5 Hz), and putting the ambient light sensor on the side instead of facing the user. These measures significantly mitigate the imaging threats of the ambient light sensor with no sacrifice in its functionality. In summary, we show that combined access to the ambient light sensor and the display can leak imaging information via capturing images of the scene in front of the screen without requiring access to the device's camera. The intensity variation of the ambient light sensor resulting from the screen displaying a known sequence can resolve an image of the scene from the perspective of the screen. Touch detection and eavesdropping hand gestures are revealed from the ambient light sensor of an off-the-shelf Android tablet. Although the resulting

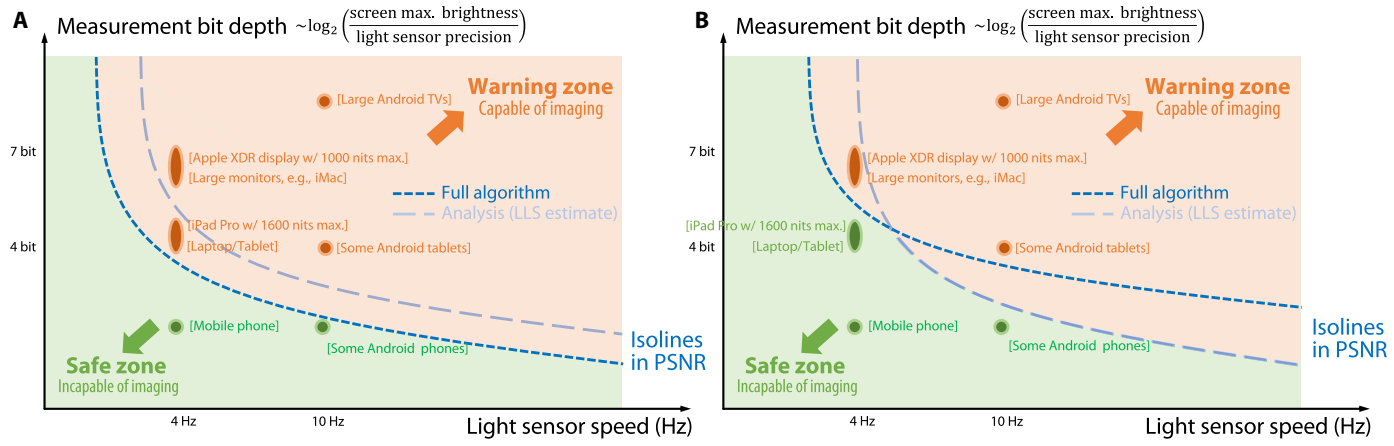

Fig. 8. Privacy risk of typical smart devices in terms of the light sensor speed and the per-measurement bit depth, characterized by the log ratio of maximum screen brightness and light sensor precision. The dashed lines are simulated isolines in PSNR (higher PSNR values means higher fidelity of image recovery, thus higher privacy risk). (**A**) Screen displaying a sequence of a predefined orthogonal sensing matrix in a row-by-row manner. LLS, linear least squares. (**B**) Screen displaying a known film sequence, where each vectorized frame forms a row of the sensing matrix. The evaluated pixel resolution and acquisition time are fixed to 32 × 32 and 20 min, respectively. Background isolines and detailed explanation can be found in fig. S14.

image resolution as well as the acquisition speed is limited and constrained configurations are required in some scenarios, these demonstrations confirm the reality of the threats. The key to these imaging privacy threats is using both passive and active light–associated components of smart devices. In response to these risks, we aim to raise awareness of potential security/privacy threats post by both passive and active components of smart devices. Measures should be taken to mitigate them from restricting access to assessing the information budget of these interacting components.

MATERIALS AND METHODS

Details on data acquisition

We use a 17.3-inch Samsung Galaxy View2 tablet operating on Android 8.1.0 Oreo with an embedded ambient light sensor (CM3323E by Capella Microsystems Inc.) for imaging information leakage of touch gestures and a 27-inch ASUS monitor (VE276Q) and a FLIR Grasshopper3 color camera (GS3-U3-41C6C-C) for occluder-based imaging information leakage. Both screens have 16:9 aspect ratio, 1920 × 1080 pixel resolution, 60-Hz refreshing rate, and ~180° viewing angle. The peak brightness is ~400 and 300 nits for the tablet and the monitor, respectively. We use the center pixels of 1024 × 1024 with 32 × 32 pixel binning for recovering a scene of pixel resolution of 32 × 32. We use a relatively low refresh rate for the screens, i.e., 12 and 30 Hz for the tablet and the monitor, respectively. Blank frames between adjacent patterns are used to get temporal segmentation of the light sensor outputs for each pattern with a duty ratio of 4:2, i.e., four frames of the same pattern followed by two blank frames. The effective refresh rate goes down to 2 and 5 Hz, respectively. The ambient light sensor has a speed of around 10 to 20 Hz, and we use the frame rate of 360 Hz and exposure time of ~2.5 ms for the camera with a region of interest of 512 × 512 (the full pixel resolution of the camera is 2048 × 2048 at a maximum frame rate of 90 Hz).

Displayed patterns on the screen

We use the Walsh-Hadamard patterns to demonstrate imaging privacy threats for two reasons. First, the Walsh-Hadamard patterns come from the Walsh-Hadamard transform basis, which is orthogonal and benefits the recovery procedure by reducing the matrix inversion overhead (see the "Imaging inverse recovery" section). Second, comparing to other orthogonal bases, like standard basis, DCT basis, and discrete Fourier transform basis (*56*), Walsh-Hadamard transform basis turns out to be more tolerant to the quantization noise for high-contrast scenes (see simulation comparisons in the Supplementary Materials). Besides, the differential measurement strategy helps reduce the measurement noise induced by surrounding light and the fluctuation of screen brightness. Random matrices can be used as displayed patterns on the screen as explored by the compressive sensing community (*19*–*21*). A known but uncontrolled sequence, like a video clip shown in Fig. 5 and fig. S12, can be used for revealing imaging privacy threats. We use orthogonal bases as a controlled sequence to explore the imaging capability for simplicity. See discussions about the sensing matrix and various subsampling strategies in sections S6 and S7.

Occluder-based imaging

More generally, we show that a variety of objects can be potentially revealed with the deformation of an occluder between the scene and the screen. Because there is no lens between it and the scene, the virtual sensor sees a severely blurred image of the scene (Fig. 9E). Inspired by the accidental pinhole/pinspeck camera (*31*), we use the deformation of an occluder to act as a pinhole between the scene and the virtual sensor. The idea is to use the deformation of the occluder to capture the contribution of a small portion of the occluder, which would naturally act as a pinhole (Fig. 9, A to C). Here, the equivalent pinhole image is obtained by subtracting the T-shape occluder image from the I-shape occluder image, i.e.

$$I_{\text{pinhole}} = I_{\text{I-occluder}} - I_{\text{T-occluder}} \tag{13}$$

where the T-occluder is composed of the I-occluder stand and a head. This is always the case when there is finger motion of the hand between the face and the screen. Considering the trade-off of spatial resolution and the signal-to-noise ratio of the captured image, we use a pinhole size of 1 cm by 1 cm (see figs. S8 and S9 for experimental discussions). The distance from the scene to the occluder is approximately equal to that from the occluder to the screen (fig. S9), the magnification factor of the equivalent pinhole imaging is ~1:1. The

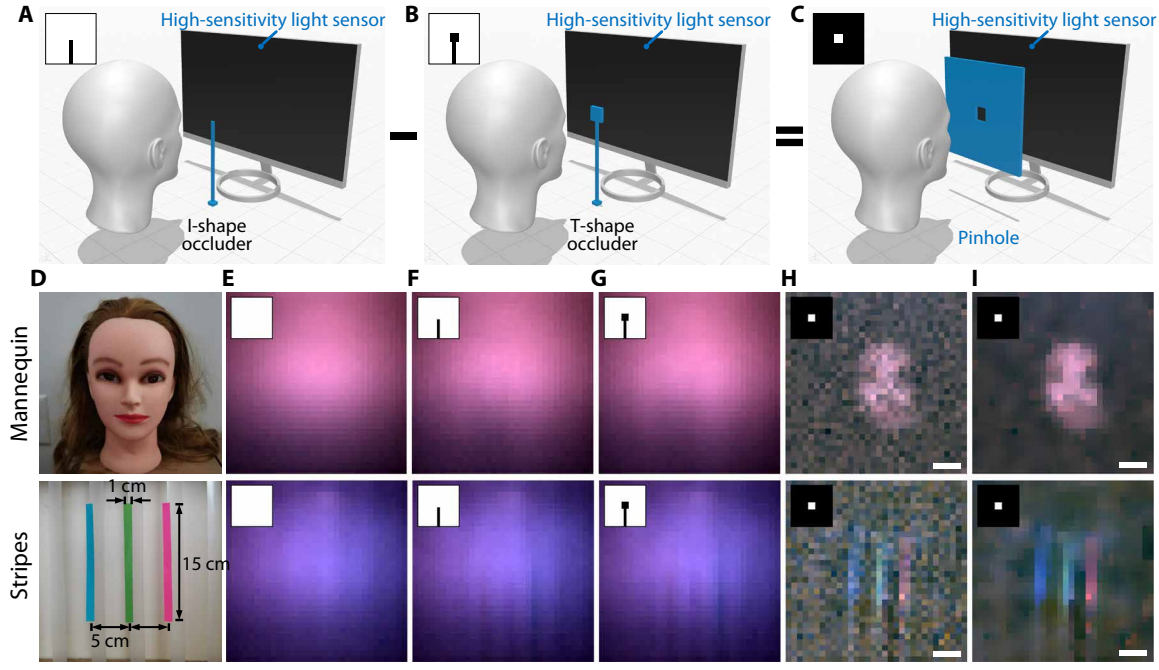

Fig. 9. Imaging privacy threats revealing general objects with an occluder between the scene and the screen. The equivalent pinhole image (**C**) can be obtained by subtraction of two occluder images with deformation, i.e., I-shape occluder (**A**) and T-shape occluder (**B**) images. The shape of the occluder (shown in black) is shown on the top left of (A) to (C) and (**E**) to (**I**). (**D**) Target scenes to be resolved. (E) Blurry images with no occluders between the scene and the screen. (F) I-shape occluder images. (G) T-shape occluder images. (H) Pinhole images obtained by subtracting T-shape occluder images from I-shape occluder images and averaging three to five measurements over time. (I) Final pinhole images by the proposed inversion algorithm. Scale bars, 5 cm.

final image is expected to be an inverted (upside-down reversed) image of the scene. Because the equivalent pinhole rejects most of the light from the scene and the limited sensitivity of the ambient light sensor on the tablet, we use a high-sensitivity light sensor, i.e., a camera summing up all the pixel values to mimic an ambient light sensor and a 27-inch monitor for experimental validation. The effective refresh rate of the patterns on the screen is set to 5 Hz, and the frame rate of the camera is 360 Hz with a region of interest as 512×512. The acquisition time for a single occluder image is 7 min. We use three color channels from the camera and recover each channel separately. A pinhole image (Fig. 9H) is obtained by subtracting the T-occluder (Fig. 9G) image from the I-occluder image (Fig. 9F). The shadows of the scene are barely visible when comparing Fig. 9 (F and G) to Fig. 9E, and the contribution of the pinhole is almost undetectable to the human eye. Here, we also average three and five measurements over time for the Stripes scene and the Mannequin scene for a better signal-to-noise ratio. Last, we apply the proposed inversion algorithm on the pinhole image (Fig. 9H) and obtain a more accurate reconstruction of the mannequin face and the color stripes (Fig. 9I). As it is often the case that hand motion can serve as the natural deformation of an occluder, this imaging modality puts pressure on the ambient light sensor, potentially revealing face information without accessing the front camera.

Imaging a general object with this screen and ambient light sensor setup is closely related to the passive NLOS imaging problem (see a full comparison of the proposed imaging modality and passive NLOS imaging in fig. S4) (*31*–*37*). Passive NLOS imaging aims to look around the corner by observing a blank wall (*32*–*35*) or a pile of clutter (*36*). The imaging information comes from the shadows cast by the occluder between the scene and the observing wall. This idea originates from the accidental pinhole and pinspeck camera (*31*), where partial occlusion and movements or deformation could serve as the virtual pinhole/pinspeck. We use the deformation of an occluder to demonstrate imaging a general object with the screen and ambient light sensor setup for two reasons. First, it is always the case that finger motion between the face and the screen can serve as a virtual pinhole. Second, the sensitivity of the ambient light sensor is still limited extracting imaging information directly (*33*–*36*) from low signal-to-noise ratio signals. This remains as a future work.

Supplementary Materials

This PDF file includes:

Sections S1 to S10
Figs. S1 to S14
Tables S1 to S4
References

REFERENCES AND NOTES

1. J. H. Ziegeldorf, O. G. Morchon, K. Wehrle, Privacy in the internet of things: Threats and challenges. *Secur. Commun. Netw.* **7**, 2728–2742 (2014).
2. L. Simon, R. Anderson, PIN skimmer: Inferring PINs Through The Camera and Microphone, in *Proceedings of the Third ACM Workshop on Security and Privacy in Smartphones and Mobile Devices* (ACM, 2013), pp. 67–78.
3. R. Mayrhofer, J. V. Stoep, C. Brubaker, N. Kralevich, The Android platform security model. *ACM Trans. Priv. Secur.* **24**, 1–35 (2021).
4. A. K. Sikder, G. Petracca, H. Aksu, T. Jaeger, A. S. Uluagac, A survey on sensor-based threats and attacks to smart devices and applications. *IEEE Commun. Surv. Tutor.* **23**, 1125–1159 (2021).
5. R. Spreitzer, PIN skimming: Exploiting the Ambient-Light Sensor in Mobile Devices, in *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones and Mobile Devices* (ACM, 2014), pp. 51–62.
6. Y. Xu, J.-M. Frahm, F. Monrose, Watching the watchers: Automatically inferring TV Content From Outdoor Light Effusions, in *Proceedings of the ACM Conference on Computer and Communications Security (CCS)* (ACM, 2014), pp. 418–428
7. L. Schwittmann, V. Matkovic, M. Wander, T. Weis, Video recognition using ambient light sensors, *IEEE International Conference on Pervasive Computing and Communications (PerCom)* (IEEE, 2016).

8. E. Wen, W. Seah, B. Ng, X. Liu, J. Cao, UbiTouch: Ubiquitous smartphone touchpads using built-in proximity and ambient light sensors, in *ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)* (ACM, 2016), pp. 286–297.
9. S. Chakraborty, W. Ouyang, M. Srivastava, LightSpy: Optical eavesdropping on displays using light sensors on mobile devices, in *IEEE International Conference on Big Data* (IEEE, 2017), pp. 2980–2989.
10. L. Olejnik, Shedding light on web privacy impact assessment: A case study of the Ambient Light Sensor {API}, *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)* (IEEE, 2020).
11. R. Raguram, A. M. White, D. Goswami, F. Monrose, J.-M. Frahm, iSpy: Automatic reconstruction of typed input from compromising reflections, in *ACM conference on Computer and Communications Security* (ACM, 2011), pp. 527–536.
12. D. Asonov, R. Agrawal, Keyboard acoustic emanations, in *IEEE Symposium on Security and Privacy* (IEEE, 2004), pp. 3–11.
13. L. Cai, H. Chen, TouchLogger: Inferring Keystrokes on Touch Screen from Smartphone Motion, *USENIX Workshop on Hot Topics in Security (HotSec)* (USENIX, 2011).
14. T. B. Pittman, Y. H. Shih, D. V. Strekalov, A. V. Sergienko, Optical imaging by means of two-photon quantum entanglement. *Phys. Rev. A.* **52**, R3429–R3432 (1995).
15. J. H. Shapiro, Computational ghost imaging. *Physical Review A.* **78**, 061802 (2008).
16. S. Ota, R. Horisaki, Y. Kawamura, M. Ugawa, I. Sato, K. Hashimoto, R. Kamesawa, K. Setoyama, S. Yamaguchi, K. Fujiu, K. Waki, H. Noji, Ghost cytometry. *Science* **360**, 1246–1251 (2018).
17. Y. Altmann, S. McLaughlin, M. J. Padgett, V. K. Goyal, A. O. Hero, D. Faccio, Quantum-inspired computational imaging. *Science* **361**, eaat2298 (2018).
18. P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, H. P. A. Lensch, Dual photography. *ACM Trans. Graph.* **24**, 745–755 (2005).
19. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, R. G. Baraniuk, Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**, 83–91 (2008).
20. B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, M. J. Padgett, 3D computational imaging with single-pixel detectors. *Science* **340**, 844–847 (2013).
21. M. P. Edgar, G. M. Gibson, M. J. Padgett, Principles and prospects for single-pixel imaging. *Nat. Photonics* **13**, 13–20 (2019).
22. G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. Miller, D. Psaltis, Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
23. R. S. Bennink, S. J. Bentley, R. W. Boyd, "Two-photon" coincidence imaging with a classical source. *Phys. Rev. Lett.* **89**, 113601 (2002).
24. P. Sen, On the relationship between dual photography and classical ghost imaging. arXiv:1309.3007 (2013).
25. T. Li, Q. Liu, X. Zhou, Practical Human Sensing in the Light, in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)* (ACM, 2016), pp. 71–84.
26. T. Li, X. Xiong, Y. Xie, G. Hito, X.-D. Yang, X. Zhou, Reconstructing hand poses using visible light, in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (AMC, 2017), pp. 1–20.
27. H. von Helmholtz, *Treatise on Physiological Optics (1925)*, vol. 3 (Optical Society of America, 1856).
28. L. Rayleigh, XXVIII. On the law of reciprocity in diffuse reflexion. *Lond. Edinb. Dublin philos. Mag. J. Sci.* **49**, 324–325 (1900).
29. S. V. Venkatakrishnan, C. A. Bouman, B. Wohlberg, Plug-and-Play Priors for Model Based Reconstruction, in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (IEEE, 2013), pp. 945–948.
30. S. H. Chan, X. Wang, O. A. Elgendy, Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Trans. Comput. Imaging.* **3**, 84–98 (2017).
31. A. Torralba, W. T. Freeman, Accidental pinhole and pinspeck cameras. *Int. J. Comput. Vis.* **110**, 92–112 (2014).
32. K. L. Bouman, V. Ye, A. B. Yedidia, F. Durand, G. W. Wornell, A. Torralba, W. T. Freeman, Turning Corners into Cameras: Principles and methods, in *IEEE International Conference on Computer Visi on (ICCV)* (IEEE, 2017), pp. 2289–2297.
33. M. Baradad, V. Ye, A. B. Yedidia, F. Durand, W. T. Freeman, G. W. Wornell, A. Torralba, Inferring Light Fields from Shadows, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE/CVF, 2018), pp. 6267–6275.
34. C. Saunders, J. Murray-Bruce, V. K. Goyal, Computational periscopy with an ordinary digital camera. *Nature* **565**, 472–475 (2019).
35. A. B. Yedidia, M. Baradad, C. Thrampoulidis, W. T. Freeman, G. W. Wornell, Using Unknown Occluders to Recover Hidden Scenes, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE/CVF, 2019), pp. 12223–12231.
36. M. Aittala, P. Sharma, L. Murmann, A. Yedidia, G. Wornell, W. T. Freeman, F. Durand, Using Unknown Occluders to Recover Hidden Scenes in *Advances in Neural Information Processing Systems (NeurIPS)* (Curran Associates Inc., 2019), pp. 14311–14321.
37. D. Faccio, A. Velten, G. Wetzstein, Non-line-of-sight imaging. *Nat. Rev. Phys. Ther.* **2**, 318–327 (2020).
38. Z. Zhang, X. Wang, G. Zheng, J. Zhong, Hadamard single-pixel imaging versus fourier single-pixel imaging. *Opt. Express* **25**, 19619–19639 (2017).
39. L. Jacques, D. K. Hammond, J. M. Fadili, Dequantizing compressed sensing: When oversampling and non-Gaussian constraints combine. *IEEE Trans. Inf. Theory.* **57**, 559–571 (2011).
40. E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory.* **52**, 5406–5425 (2006).
41. S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2010).
42. N. Parikh, S. Boyd, Proximal algorithms. *Found. Trends Mach. Learn.* **1**, 127–239 (2014).
43. A. Levin, Y. Weiss, User assisted separation of reflections from a single image using a sparsity prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 1647–1654 (2007).
44. A. Levin, R. Fergus, F. Durand, W. T. Freeman, Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.* **26**, 70 (2007).
45. W. W. Hager, Updating the inverse of a matrix. *SIAM Review* **31**, 221–239 (1989).
46. X. Yuan, Generalized alternating projection based total variation minimization for compressive sensing, in *IEEE International Conference on Image Processing (ICIP)* (IEEE, 2016), pp. 2539–2543.
47. X. Yuan, Y. Liu, J. Suo, Q. Dai, Plug-and-Play Algorithms for Large-scale Snapshot Compressive Imaging, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE/CVF, 2020), pp. 1447–1457.
48. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
49. M. Gharbi, G. Chaurasia, S. Paris, F. Durand, Deep joint demosaicking and denoising. *ACM Trans. Graph.* **35**, 1–12 (2016).
50. K. Zhang, W. Zuo, L. Zhang, FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **27**, 4608–4622 (2018).
51. Y. Liu, G. W. Wornell, W. T. Freeman, F. Durand, Liuyang12/privacy-dual-imaging-software for imaging privacy threats from an ambient light sensor. (2023); https://zenodo.org/records/10211455.
52. W. K. Pratt, J. Kane, H. C. Andrews, Hadamard transform image coding. *Proc. IEEE* **57**, 58–68 (1969).
53. M. Harwit, N. J. A. Sloane, *Hadamard Transform Optics* (Academic Press, 1979).
54. K. Guo, S. Jiang, G. Zheng, Multilayer fluorescence imaging on a single-pixel detector. *Biomed. Opt. Express.* **7**, 2425–2431 (2016).
55. K. Tanaka, Y. Mukaigawa, A. Kadambi, Polarized non-line-of-sight imaging, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 2136–2145.
56. Z. Zhang, X. Ma, J. Zhong, Single-pixel imaging by means of Fourier spectrum acquisition. *Nat. Commun.* **6**, 6225 (2015).
57. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
58. G. Orwell, *Nineteen Eighty-Four (1984)* (Secker & Warburg, 1949).
59. WIRED Insider, Sights & Sound: Inside the 'acoustic surface' powering Sony's first OLED TV (WIRED Insider, 2017).
60. O. S. Olamide, Sound on display (SOD) brings sound-emitting screens to your next smartphone (Dignited, 2018).
61. C. Liu, A. Torralba, W. T. Freeman, F. Durand, E. H. Adelson, Motion magnification. *ACM Trans. Graph.* **24**, 519–526 (2005).
62. P. Sen, S. Darabi, Compressive dual photography. *Comput. Graph. Forum* **28**, 609–618 (2009).
63. S. P. Seidel, J. Murray-Bruce, Y. Ma, C. Yu, W. T. Freeman, V. K. Goyal, Two-dimensional non-line-of-sight scene estimation from a single edge occluder. *IEEE Trans. Comput. Imaging* **7**, 58–72 (2021).
64. M. O'Toole, D. B. Lindell, G. Wetzstein, Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* **555**, 338–341 (2018).
65. D. B. Lindell, G. Wetzstein, M. O'Toole, Wave-based non-line-of-sight imaging using fast *f-k* migration. *ACM Trans. Graph.* **38**, 1–13 (2019).
66. X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, A. Velten, Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature* **572**, 620–623 (2019).
67. F. Xu, G. Shulkind, C. Thrampoulidis, J. H. Shapiro, A. Torralba, F. N. C. Wong, G. W. Wornell, Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging. *Opt. Express* **26**, 9945–9962 (2018).
68. C. Thrampoulidis, G. Shulkind, F. Xu, W. T. Freeman, J. H. Shapiro, A. Torralba, F. N. C. Wong, G. W. Wornell, Exploiting occlusion in non-line-of-sight active imaging. *IEEE Trans. Comput. Imaging* **4**, 419–431 (2018).
69. J. Rapp, C. Saunders, J. Tachella, J. Murray-Bruce, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, R. M. A. Dawson, F. N. C. Wong, V. K. Goyal, Seeing around corners with edge-resolved transient imaging. *Nat. Commun.* **11**, 1–10 (2020).
70. P. T. Boufounos, R. G. Baraniuk, 1-Bit compressive sensing, in *Annual Conference on Information Sciences and Systems (CISS)* (IEEE, 2008), pp. 16–21.
71. Y. Liu, X. Yuan, J. Suo, D. J. Brady, Q. Dai, Rank minimization for snapshot compressive imaging. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 2990–3006 (2019).

72. E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
73. D. L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
74. A. Zymnis, S. Boyd, E. Candès, Compressed sensing with quantized measurements. *IEEE Signal Process. Lett.* **17**, 149–152 (2010).
75. J. N. Laska, R. G. Baraniuk, Regime change: Bit-depth versus measurement-rate in compressive sensing. *IEEE Trans. Inf. Theory* **60**, 3496–3505 (2012).
76. L. Jacques, J. N. Laska, P. T. Boufounos, R. G. Baraniuk, Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Trans. Inf. Theory* **59**, 2082–2102 (2013).
77. P. B. Fellgett, On the ultimate sensitivity and practical performance of radiation detectors. *J. Opt. Soc. Am.* **39**, 970–976 (1949).
78. A. K. Jain, *Fundamentals of Digital Image Processing* (Prentice-Hall Inc., 1989).
79. C. A. Metzler, A. Maleki, R. G. Baraniuk, From denoising to compressed sensing. *IEEE Trans. Inf. Theory* **62**, 5117–5144 (2016).
80. L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms. *Phys. D: Nonlinear Phenom.* **60**, 259–268 (1992).
81. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3D transform- domain collaborative filtering. *IEEE Trans. Image Process.* **16**, 2080–2095 (2007).
82. K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, R. Timofte, Plug-and-play image restoration with deep denoiser prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 6360–6376 (2022).
83. F. Soldevila, J. Dong, E. Tajahuerce, S. Gigan, H. B. de Aguiar, Fast compressive Raman bio-imaging via matrix completion. *Optica* **6**, (2019).
84. A. Adler, D. Boublil, M. Zibulevsky, Block-based compressed sensing of images via deep learning, *IEEE International Workshop on Multimedia Signal Processing (MMSP)* (IEEE, 2017).
85. S. Lohit, K. Kulkarni, R. Kerviche, P. Turaga, A. Ashok, Convolutional neural networks for noniterative reconstruction of compressively sensed images. *IEEE Trans. Comput. Imaging* **4**, 326–340 (2018).
86. W. Shi, F. Jiang, S. Liu, D. Zhao, Image compressed sensing using convolutional neural network. *IEEE Trans. Image Process.* **29**, 375–388 (2019).
87. J. Zhang, C. Zhao, W. Gao, Optimization-inspired compact deep compressive sensing. *IEEE J. Sel. Top. Signal Process.* **14**, 765–774 (2020).
88. L. Bian, J. Suo, Q. Dai, F. Chen, Experimental comparison of single-pixel imaging algorithms. *J. Opt. Soc. Am. A* **35**, 78–87 (2018).
89. S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, vol. 1 (Prentice Hall, 1993).

Acknowledgments: We thank L. Murmann, P. Sharma, A. Yedidia, V. K. Goyal, J. H. Shapiro, F. N. C. Wong, C. Saunders, S. P. Bangaru, T.-M. Li, X. Yuan, J. Suo, A. Kaspar, R. Spreitzer, and F. Asim for discussions and L. Anderson and T. Rubio for proofreading. **Funding:** This work was supported in part by the DARPA REVEAL program (HR0011-16- C-0030) and MIT Stata Family Presidential Fellowship (to Y.L.). **Author contributions:** Y.L. conceived the idea with inputs from F.D., W.T.F., and G.W.W. Y.L. designed the experiments, developed the experimental setup, performed the measurements, and implemented the imaging inversion procedure. F.D. supervised all aspects of the project. All authors took part in analyzing the results and writing the paper and the Supplementary Materials. **Competing interests:** W.T.F. is affiliated with Google Research. This study is not part of or funded by Google. The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The code along with measurement data and experimental configurations is publicly available at https://doi.org/10.5281/zenodo.10211455, and the up-to-date version can be found at https://github.com/liuyang12/privacy-dual-imaging.

Submitted 21 June 2023
Accepted 14 December 2023
Published 10 January 2024
10.1126/sciadv.adj3608