Pre–Capture Privacy for Small Vision Sensors

Francesco Pittaluga, *Member, IEEE,* and Sanjeev J Koppal, *Member, IEEE,*

Abstract—The next wave of micro and nano devices will create a world with trillions of small networked cameras. This will lead to increased concerns about privacy and security. Most privacy preserving algorithms for computer vision are applied after image/video data has been captured. We propose to use privacy preserving optics that filter or block sensitive information directly from the incident light-field *before* sensor measurements are made, adding a new layer of privacy. In addition to balancing the privacy and utility of the captured data, we address trade-offs unique to miniature vision sensors, such as achieving high-quality field-of-view and resolution within the constraints of mass and volume. Our privacy preserving optics enable applications such as depth sensing, full-body motion tracking, people counting, blob detection and privacy preserving face recognition. While we demonstrate applications on macro-scale devices (smartphones, webcams, etc.) our theory has impact for smaller devices.

Index Terms—Computer Vision, Privacy.

1 Introduction

Our world is bursting with ubiquitous, networked sensors. Even so, a new wave of sensing that dwarfs current sensor networks is on the horizon. These are miniature platforms, with feature sizes less than 1mm, that will appear in micro air vehicle swarms, intelligent environments, body and geographical area networks. Equipping these platforms with computer vision capabilities could impact security, search and rescue, agriculture, environmental monitoring, exploration, health, energy, and more.

Yet, achieving computer vision at extremely small scales still faces two challenges. First, the power and mass constraints are so severe that full-resolution imaging, along with post-capture processing with convolutions, matrix inversions, and the like, are simply too restrictive. Second, the privacy implications of releasing trillions of networked, tiny cameras into the world would mean that there would likely be significant societal pushback and legal restrictions.

In this paper, we propose a new framework to achieve both power efficiency and privacy preservation for vision on small devices. We build novel fixed and programmable optical designs that filter incident illumination from the scene, before image capture. This allows us to attenuate sensitive information while capturing exactly the portion of the signal that is relevant to a particular vision task. In this sense, we seek to generalize the idea of privacy preserving optics beyond specialized efforts. We demonstrate privacy preserving optics that enable accurate depth sensing, full-body motion tracking, multiple people tracking, blob detection and face recognition.

Our optical designs filter light before image capture and represent a new axis of privacy vision research that complements existing "post image capture" based approaches to privacy preservation. Like these other approaches, we seek to balance the utility and privacy of the data. For miniature sensors, we must also balance the performance and privacy guarantees of the system with sensor characteristics such as mass/volume, field-of-view and resolution. In this paper, we show applications on macro-scale devices (smartphones, webcams, etc.), but our theory has impact for smaller devices.

Our contributions are

1) We demonstrate a programmable optics-based privacy framework that enables pre-capture implementations of mask-based privacy algorithms such as k-anonymity and black-out. We also provide theory to miniaturize these designs within the smallest sensor volume.
2) We show how to select a defocus blur that provides a certain level of privacy over a working region, within the limits of sensor size. We show applications where defocus blur provides both privacy and utility for time-of-flight, thermal and near-infrared sensors.
3) We implement angular scale space analysis using an optical array, with most of the power hungry difference-of-gaussian computations performed pre-capture. We demonstrate human head tracking with this sensor. We provide an optical version of the knapsack problem to miniaturize such multi-aperture optical privacy preserving sensors in the smallest mass/volume.

2 Background

The work presented in paper lies at the intersection of privacy preserving computer vision and small-scale computer vision. This section aims to provide some relevant background for these two fields.

2.1 Privacy Preserving Computer Vision

Conventional privacy preserving systems for computer vision enforce privacy via post-capture application of existing software-based privacy algorithms. Many such software-based algorithms exist. Pixelation, Gaussian blurring, face swapping [1] and black-out [2] provide privacy by trading off image utility [3], [4]. Encryption based schemes such as [5], [6], [7] enable recovery of the original data, via a key. Methods based on k-anonymity [8] provably bound face recognition rate at $1/k$ while maintaining image utility [9], [10], [11], [12], [13]. This bound is achieved by averaging together each target face in an image with $k-1$

• *F. Pittaluga and S. J. Koppal are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611. E-mail: f.pittaluga@ufl.edu and sjkoppal@ece.ufl.edu*

other faces selected from a database. Despite the successes such software-based algorithms, privacy systems that rely on the post capture application of these algorithms have an inherit vulnerability in that there exists a period, after capture, when privacy has not yet been enforced, where the raw data is vulnerable to attacks. This has resulted in the development of computational cameras for privacy preserving computer vision

Computational cameras for privacy preserving computer vision aim to remove the post-capture vulnerability by selectively sampling the light-field such that they only capture non-sensitive information. Within this space, there are four main approaches. The first approach leverages existing embedded technology to perform privacy algorithms at the camera level itself and then uses encryption or other methods to manage the information pipeline [14], [15], [16], [17]. The second approach, like the first, uses hardware integration to increase security, but aims to build novel sensors that preserve privacy through watermarking [18], cartooning [19] and pixel averaging [20]. The third approach aims to develop novel privacy algorithms that enforce privacy during image fomation, via manipulation of sensor processes such as gain, digitization and exposure time [21]. The fourth approach aims to add a complementary layer of security by fixing conventional sensors with specialized privacy optics that remove sensitive data prior to image capture, through filtering of the incident light-field. The methods presented in this paper all fall within the fourth approach. A range of other specialized privacy optics have also been proposed. [22] proposed a system using thermal motion sensors that enables two-person motion tracking in a room. [23] used a line sensor and cylindrical lens to detect a person's position and movement. [24] controlled the light-transport to shadow sensitive regions, removing data-utility in those areas. [25] showed a system consisting of five low resolution cameras, installed in a single room, that enables private human activity recognition. Our sensors differ from this work in that they are mobile, i.e., do not require any form of installation. [26] proposed a system that uses a thermal sensor to detect faces and a second sensor fitted with an LCoS capture private images. In this paper, we present a programmable optics-based framework that generalizes [26]. Using this framework, we show pre-capture implementations of black-out and k-anonymity. We also show applications where defocus optics provide both privacy and utility for time-of-flight, thermal, and near-infrared sensors.

Compressive sensing techniques have found application in imaging and vision [27], [28]. Some approaches use optical projections [28] and have been integrated with classification [29] and encryption [30], [31], [32]. [33] proposed directly sensing random scene projections and reconstructing a decimated version of the scene to enable privacy preserving object tracking and secrecy in the sense the original frames cannot be recovered without knowledge of the seed used to generate the sampling matrices. In future work, we may consider implementing CS-based algorithms within our optical frame work.

2.2 Small-Scale Computer Vision Sensors

The embedded systems community has proposed many vision techniques for low-power hardware [34], [35], [36]. However, for micro-scale platforms, the average power consumption is often in the range of milli-Watts or micro-Watts [37], [38], [39], [40], [41], [42]. In these scenarios, our approach of jointly considering optics, sensing, and computation within the context of platform constraints will be crucial.

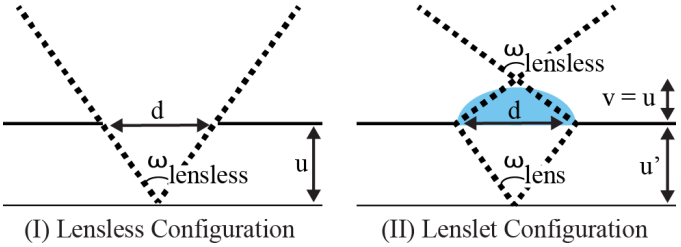

Fig. 1. **Optical Elements used for Defocus**. We use either lensless or lenslet designs in this paper for optical defocus. The figure shows that any lenslet sensor of diameter d and image distance u can be modeled as a lensless sensor of height u and pinhole size d, and therefore we use only the lensless version in our theory.

Fourier optics [43], [44] have limited impact for miniature vision systems that must process incoherent scene radiance. However, controllable PSFs in conjunction with post-capture processing are widely used in computer vision [45], [46], [47], [48]. In contrast to these approaches, we seek optics like [49], [50], [51], [52], that distill the incoming light-field for vision applications. [49] also provides design tools to maximize a sensor's effective-field-of-view (eFOV). eFOV is defined as the the range of viewing angles over which this angular support of a sensor is consistent. In this paper, we expand these tools to include privacy considerations.

Our optical knapsack approach, presented in Sec. 3.2.2, for miniaturizing multi-aperture sensors is a miniature analog to larger camera sensor network coverage optimizations [53], [54], [55], [56].

3 Fixed Privacy Optics

We now consider fixed privacy optics for single and multi-aperture sensor designs. Fixed privacy optics have two main advantages. Firstly, since the optics do not contain electronics, they are highly robust against software-based attacks. Secondly, since the optics directly filter the incident light-field they reduce the required on-board computations, without drawing on any on-board power.

3.1 Privacy Optics

In this section, we provide a tool for designing fixed-optics sensors that perform intentional optical defocus for privacy. As in [49], we assume a distant scene which can be represented by intensity variation over the hemisphere of directions (i.e. the local light-field is a function of azimuth and elevation angles). Unlike [49], we augment the hemispherical model with a notion of scene depth, where the angular support of an object reduces as its distance to the sensor increases. We use either lensless or lens-based optics for defocus and, as illustrated in Fig. 1, these apply an *angular* defocus kernel over the hemispherical visual field. The range of viewing angles over which this angular support is consistent, is known as the effective FOV or *eFOV* [49]. We chose the optical elements in Fig. 1 for fabrication convenience and our theory can be used with other FOV [49], [57], [58] elements. As demonstrated by [49], every lensless element can be replaced with a corresponding lenslet element. Such an equivalent pair is illustrated in Fig. 1. In this paper, we utilize the lensless theory, even when considering lenslet systems.

The inputs to our design tool are the defocus specifications $\Sigma = \{\Delta, \sigma, R, \Theta, \rho\}$, where Δ is the angular error tolerance, σ

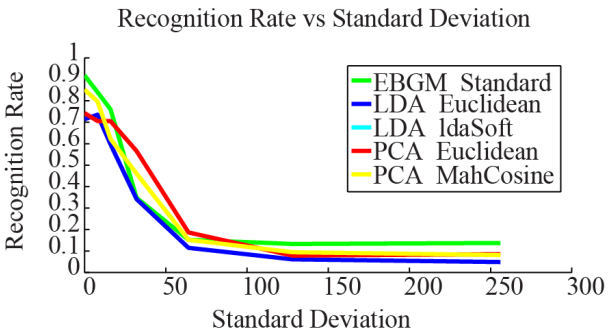

Fig. 2. **Face Recognition Rate vs Simulated Optical Defocus.** We quantified optical defocus privacy empirically by convolving face images from the FERET database [59] with a Gaussian filters of standard deviations $\{2, 4, 8, 16, 32, 64, 128, 256\}$, to simulate optical defocus, and performing face recognition on filtered images. Three face recognition algorithms were tested: Principle Components Analysis, Linear Discriminant Analysis, and Elastic Bunch Graph Matching.

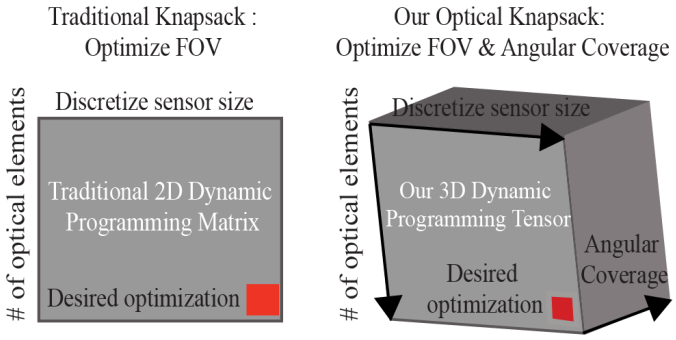

Fig. 3. **Optical Knapsack Algorithm**. A traditional knapsack solution for packing optical elements might fail if the elements covered the same portion of the visual field. Our optical knapsack solution takes into account the angular coverage of each sensor and maintains the pseudo-polynomial nature of the original dynamic programming knapsack solution.

is the desired defocus given in terms of a Gaussian blur on an image of resolution R and FOV Θ, and ρ is the length of the biggest target feature that is to be degraded by defocus blurring. For example, for a sensor designed to de-identify faces, ρ might be the size in millimeters of large facial features, such as eyes. The field of view and resolution are necessary to relate standard deviation, a dimensionless quantity, to an angular support defocus blur. The output of the tool are lensless sensor dimensions and characteristics, such as eFOV and angular support.

If we can approximate a gaussian filter of standard deviation σ by a box blur corresponding to 2σ, then, for defocus specifications Σ, the angular support is

$$\omega_o = 2\sigma \left(\frac{\Theta}{R} \right). \tag{1}$$

Like [60], we quantified the privacy of our algorithm by simulating it in software and testing against the CSU Face Identification Evaluation System (FES) [61]. To simulate optical defocus, probe face images from the FERET database [59] were convolved with a Gaussian filter before inputing them to the FES. The gallery images were filtered equivalently to improve recognition [62]. The fa and fb partitions were set as the gallery and probe images respectively. This experiment was repeated for Gaussian filters of standard deviations $\{2, 4, 8, 16, 32, 64, 128, 256\}$. Fig 2. shows the rank 1 recognition rate of three algorithms from the FES for the set of standard deviations. The three algorithms tested were Principle Components Analysis (PCA), Linear Discriminant Analysis (LDA), and Elastic Bunch Graph Matching (EBGM). From Fig. 2. it is clear that heavy Gaussian filtering significantly decreases recognition rate. For all three algorithms, the rank 1 recognition rate decreased to less than $\%12$ when the standard deviation was greater than 100.

3.2 Miniaturization

3.2.1 Miniaturization of Single-Aperture Sensor

In [49], a lensless sensor was optimally designed for maximum eFOV given an angular support ω_o and angular support tolerance Δ. We provide an additional design output, z_{min}, which is the minimum distance between the sensor and the target in order for the sensor to preserve the degree of privacy specified by the defocus specifications and it is given by,

$$z_{min} = \frac{\rho}{2tan(\frac{\omega_o}{2})}. \tag{2}$$

In summary, our algorithm takes as input defocus specifications $\Sigma = \{\sigma, \rho, \Theta, R, \Delta\}$, computes ω_o as described in Eq. 1 and applies the method of [49] plus Eq. 2 to output the optimal design with maximum eFOV, $\Pi = \{u, d, z_{min}\}$.

3.2.2 Miniaturization of Multi-Aperture Sensor.

In this section, we arrange optical elements within the constraints of small devices. Such packing problems have been studied in many domains [63] and the knapsack problem is a well-known instantiation [64]. We propose an optical variation on the knapsack problem that takes into account each element's angular coverage.

To see why this is needed, consider applying the traditional knapsack problem for a set of optical elements. Let the total size (mass, volume or area) available for sensing optics be A. Suppose each optical element i has a field-of-view f_i and a size of a_i. Given n elements with indices $0 \leq i \leq n$, we want to find an identity vector x of length n s.t. $x_i \in (0, 1)$ and $\Sigma_i x_i f_i$ is maximized whereas $\Sigma_i x_i a_i \leq A$. While this problem is NP-hard, a pseudo-polynomial algorithm $O(nA)$ has been proposed by recursively creating an $n \times A$ array M;

$$M[0, a] = 0 \;\; if \;\; 0 \leq a \leq A$$
$$M[i, a] = -\infty \;\; if \;\; a < 0$$
$$M[i, a] = \max(M[i-1, a], f_i + M[i-1, a-a_i]),$$

where $M(i, a)$ contains the maximum eFOV possible with the first i elements within size constraints a and so $M(n, A)$ is the solution. Since the a_i values may be non-integers, these are usually multiplied by 10^s, where s is the desired number of significant digits. This well-known approach fails to provide the best optical element packing, because greedily increasing total eFOV does not guarantee *coverage* of the visual hemisphere. For example, a set of 5 identical elements, each having a eFOV of $\frac{\pi}{5}$, would seem to have a sum total of $180°$ eFOV but would redundantly cover the same angular region.

Our optical knapsack algorithm takes into account angular coverage by first discretizing the field-of-view into β angular regions, each with a solid angle of $\frac{\pi}{\beta}$. We define an array $K(n, \beta)$, where $K(i, b) = 1$ if that optical element covers the angular

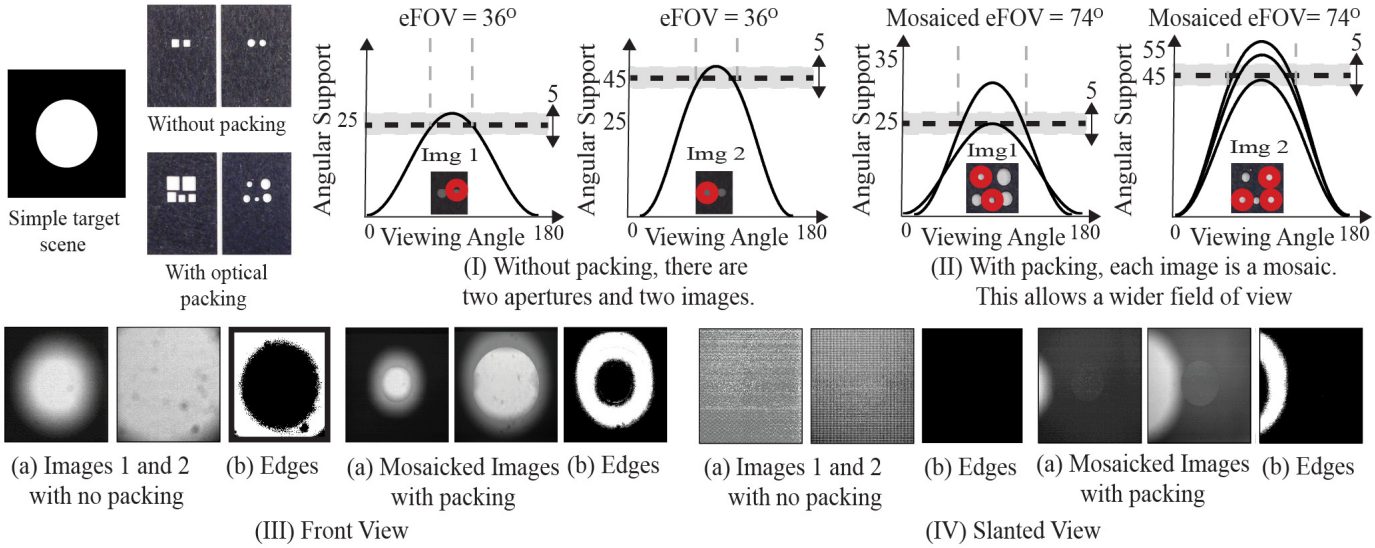

Fig. 4. **Edge detection application with optical packing**. Wide angle optical edge detection has been shown [49] by subtracting sensor measurements from two different lensless apertures. [49]'s approach in (I) is *unable* to utilize the full sensor size because it requires each image to come from one sensor. In contrast, our optical knapsack technique can pack the sensor plane with multiple optical elements (II) and synthesize, in software, a wider field of view. (II) demonstrates how the angular support of multiple elements vary over the visual field, and how different measurements from multiple apertures are combined to create a mosaicked image with a larger eFOV. We perform edge detection using both the configuration from [49] and our packed sensor on a simple scene consisting of a white blob on a dark background. When the target is directly in front of the sensor (III), both optical configurations produce reasonable edge maps. At a particular slanted angle (in this case, around 15 degrees due to vignetting) [49]'s approach (IV) does not view the target (images show sensor noise) and no edges are detected. The edges are still visible for our design, demonstrating its larger field of view.

regions b in its field-of-view, and is zero everywhere else. We also define the array M to be three-dimensional of size $n \times A \times \beta$. As before, each entry of $M(i, a, 0)$ contains the maximum field of view that can be obtained with the first i elements with a sensor of size a and $M(n, A, 0)$ contains the solution to the knapsack problem. Entries $M(i, a, 1)$ through $M(i, a, \beta)$ are binary, and contain a 1 if that angular region is covered by the elements corresponding to the maximum field-of-view $M(i, a, 0)$ and a zero otherwise. The array M is initialized as,

$$M[i, a, b] = 0, \text{ if } 0 \leq a \leq A, \quad 0 \leq i \leq n \text{ and } 0 \leq b \leq \beta$$

and is recursively updated as

If $a < 0$	$M[i, a, 0] = -\infty$
For any other a, **for any** i	
If $M[i-1, a, 0] <$ $f_i + M[i-1, a-a_i, 0]$ **and** $\sum_{1 \leq b \leq \beta} M[i-1, a, b] <$ $\sum_{1 \leq b \leq \beta} M[i-1, a-a_i, b] \vee K[i, b]$	$\begin{cases} M[i, a, 0] = \\ f_i + M[i-1, a-a_i, 0] \\ \\ M[i, a, b] = \\ M[i-1, a-a_i, b] \vee \\ K[i, b], \ b \in (1, \beta) \end{cases}$
Otherwise $\forall b$	$M[i, a, b] = M[i-1, a, b]$

where $\vee$ represents the logical OR function. This optical knapsack packing algorithm adds a β multiplications and $\beta + 2$ additions to the computational cost of the algorithm. This results in a $O(nA\beta)$ algorithm, which is still pseudo-polynomial. As with the original knapsack problem, if the discretization of A and the angular regions β are reasonable, the implementation is tractable.

We demonstrate the optical packing algorithm for edge detection for a simple white disk target (Fig. 4). Our goal is two lensless sensors, each with angular supports $\omega_{o1} = 25°$ and $\omega_{o2} = 45°$ and both with error margins of $\Delta = 5°$. Fig. 4(I) shows [49]'s approach, with no packing, for a $6.6mm \times 5.5mm$ sensor and whose template height had been constrained to $u = 2mm$. Only a small portion of the sensor is used, corresponding to an eFOV of 36°. Next we utilized our optical knapsack algorithm to maximize the eFOV on the given total area. In Fig. 4(II), a five element design is shown. Note that our algorithm only solves the knapsack part of the algorithm - the rectangular packing could be performed using widely known methods [65], but in this case was done manually. We discretized the template sizes in steps of 0.1mm and considered 30 different optical elements and discretized the angular coverage into 36 units of 5 degrees each. Since we targeted two defocus sensor designs, our 3D tensor was $30 \times 2501 \times 72$. Our dynamic programming algorithm produced the solution in Fig. 4(II), where the measurements from three elements, with aperture diameters 2.2mm, 1.9mm and 1.6mm, were mosaicked to create the image corresponding to ω_{o2} and the remaining two elements, with aperture diameters 1.2mm and 0.9mm, were used to create ω_{o1}. In the figure, the mosaicked measurements were subtracted to create a DoGs based edge detection. At a grazing angle, only the packed, wide FOV sensor can still observe the scene, demonstrating that our optimally packed design has a larger field of view.

3.3 Example Sensor Designs

3.3.1 *Defocused Time-of-flight Sensor*

We designed 3D printed privacy optics for the Microsoft Kinect V2 that enable privacy preserving depth sensing, segmentation, and full body motion tracking. The privacy optics, shown Fig. 5(III), consisted of a 3D printed plano-convex lens for the depth sensor, a black-out cover for RGB camera, and a sleeve, which holds the lens and cover. Our initial implementations [66] used conventional plano-convex IR lenses from Edmund Optics, which cost $\sim$ \$300. Using 3D printing we fabricated comparable lenses

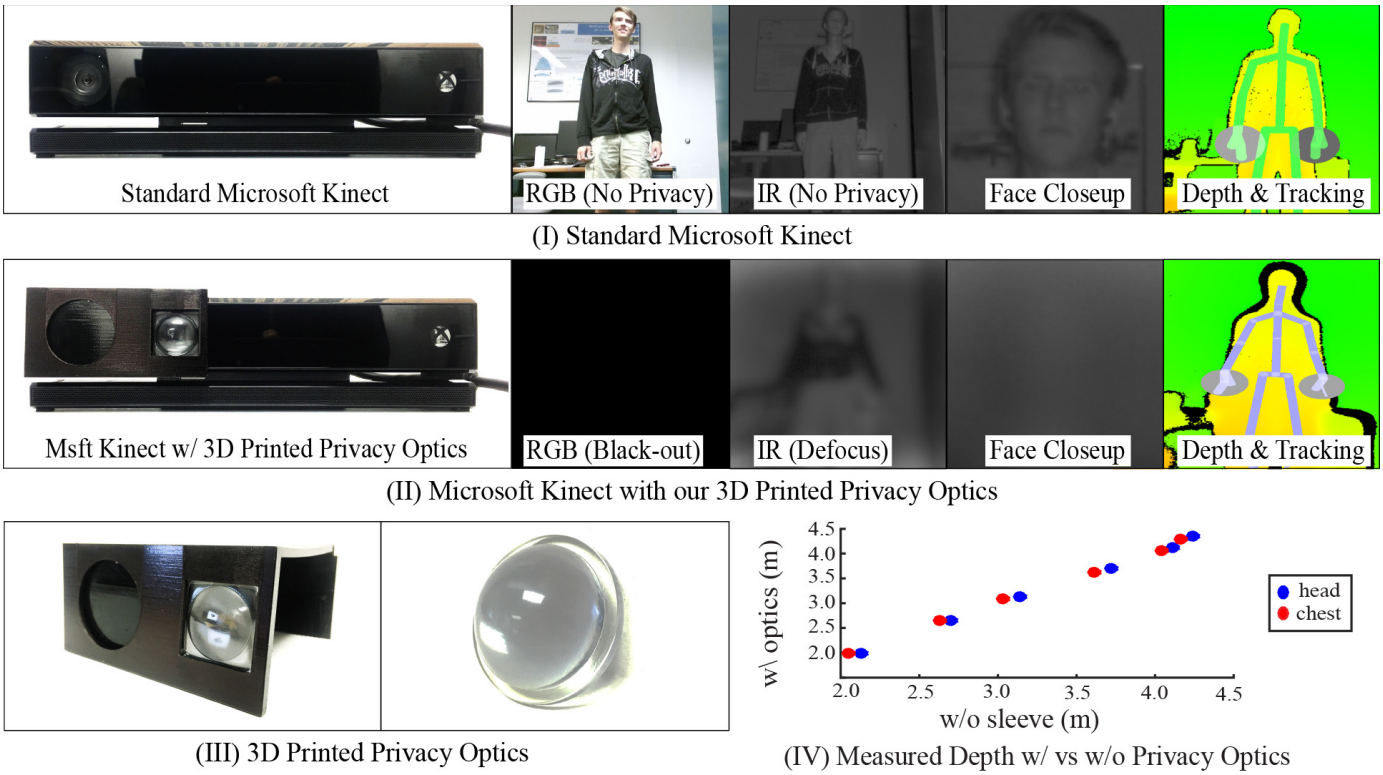

Fig. 5. **Privacy Preserving Depth Sensing and Full-Body Motion Tracking**. We designed a fully 3D printed privacy sleeve for the Microsoft Kinect V2 and that allows accurate depth sensing and motion tracking. The sleeve has a removable 3D printed cover for the color camera and a 3D printed lens for the IR sensor. As shown in (I), without the privacy sleeve, faces can clearly be identified in both the RGB and IR sensor images. In contrast, as shown in (II), our privacy sleeve performs optical black-out out for the RGB sensor and optical defocus for the IR sensor, yet the native Kinect tracking software from Microsoft still performs accurate depth sensing and motion tracking. Close-ups of the 3d printed privacy sleeve and lens are shown in (III). A plot comparing the measured depth of a person's head and chest with and with out our privacy sleeve for the documented depth range of the Kinect, is given in (IV).

for $\sim$ \$10. The lenses were printed on an Objet260 Connex3 3D printer using VeroClear-RGD810 transparent printing material and then manually polished using various grades of sandpaper and a NOVUS 7100 Plastic Polish Kit.

AMCW TOF cameras approximate depth by measuring the time-of-flight τ of modulated probing signal, i.e. the time it takes the probing signal to reflect off a scene point and return to the camera. Given the time-of-flight τ, the distance d of the scene point can be computed with $d = \frac{1}{2}c\tau$, where c denotes the speed of light constant. However, since τ cannot be observed directly, the phase shift ϕ between the probing signal and the received signal is instead measured, and τ is approximated using the the following relation $\phi = \upsilon\tau$, where υ is the angular frequency (in rad/sec) of the modulated probing signal. The probing function is generally assumed to be a sinusoid $p(t)$ given by $p(t) = 1+\lambda\cos(\upsilon t), \quad 0 < \lambda \leq 1$ where λ is the amplitude of the modulated probing signal [67]. Illumination from the probing signal arrives at a pixel via all optical paths within a pixel's visual hemisphere, defined by the camera's angular support ω_o. Thus, we model the signal $r(t)$ arriving at each pixel as

$$r(t) = \int_0^{\frac{\omega_o}{2}} \int_0^{2\pi} \Gamma(\theta,\varphi)(1+\lambda\cos(\upsilon t - \phi_\upsilon(\theta,\varphi)))\mathrm{d}\theta\mathrm{d}\varphi \tag{3}$$

where $\phi_\upsilon = t - 2d\upsilon/c$ denotes the phase shift w.r.t. to the probing signal and Γ is proportional to albedo [67]. Discretizing each pixel's visual hemisphere into K optical paths, we approximate $r(t)$ as a linear combination of cosines

$$\widetilde{r}(t) = c_0 + \lambda \sum_{n=1}^{N} \Gamma_n \cos(\upsilon t - \phi_{\upsilon,n}) \tag{4}$$

where c_0 is some positive scalar. Then, with a bit of phasor arithmetic

$$\begin{aligned} \widetilde{r}(t) &= c_0 + \lambda e^{j\upsilon t} \sum_{n=1}^{N} \Gamma_n e^{-\phi_{\upsilon,n}} \\ &= c_0 + \lambda e^{j\upsilon t} \mathbf{\Gamma} e^{-\mathbf{\Phi}_{\upsilon,n}}. \end{aligned} \tag{5}$$

Typically, TOF cameras are modeled as pinhole cameras. In phasor notation (Eq. 5), it is easy to see why the pinhole model is a sensible assumption: nearby scene points tend to have similar depths and adding phasors with similar phase results in minimal smoothing of the phase, despite large amplitude differences. The same principle applies to a defocused TOF cameras, except that the circle of confusion is larger. Thus, when the scene geometry is relatively smooth, our defocusing optics that the affect the IR amplitude image while leaving the phase (or depth information) mostly intact.

Figure 5(III) shows our experimental results for Kinect depth measurements of a tracked human head and chest, with and without our 3D printed defocus optics, for the entire documented working range of the Kinect. The mean absolute error for the defocused measurements is 5.7cm and 5cm for the head and chest measurements respectively. In Fig. 5(II) we show pre-capture privacy preserving full-body motion tracking with a kinect fitted

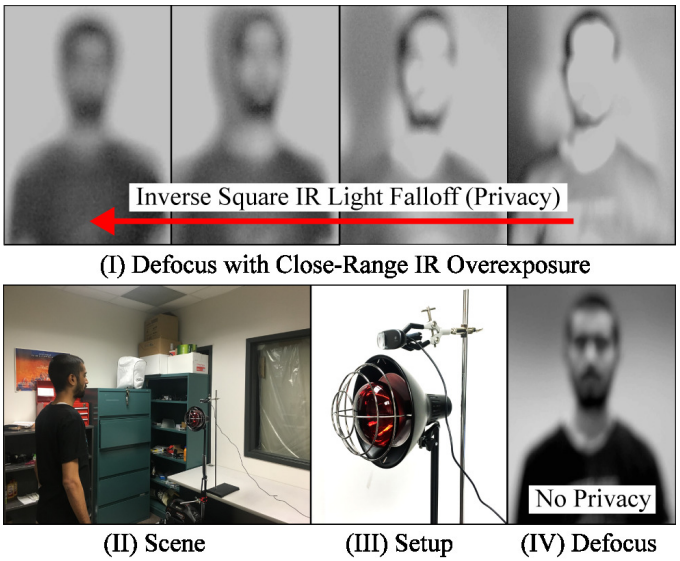

Fig. 6. **Defocus with Close-Range IR Overexposure**. We designed a sensor overexposes nearby faces to remove the minimum depth requirement for defocus privacy. The setup consisted of an IR light source adjacent to a defocused webcam, with sensitivity in to NIR light, placed in an unilluminated office (Fig. 6(II,III)). With our setup, faces were overexposed for distances less than $z_{min} = 1.8m$ and were appropriately exposed, but defocused for distances greater than $2.3m$. Fig. 6(I) shows a sequence of four private images of a person standing $3.0m, 2.4m, 1.8m, 0.6m$ (ordered left to right) from the sensor. 6(IV) shows an image of a person standing $0.6m$ without the overexposing IR light source. Since $0.6m < z_{min}$, this image is not private.

with our privacy sleeve, using the native Kinect tracking API. The subject in the figure was 1.7m away from the sensor. The angular support of the IR sensor with the sleeve was $3°$, which corresponds to lensless parameters $u = 10$mm, $d = 0.5$mm, a minimum distance, $z_{min} = 1.5$m for degrading features of 8cm and an eFOV of $64.7°$ for $\Delta = 1°$.

3.3.2 *Defocused Near-Infrared Sensor with Close-Range Overexposure*

Optical defocus for privacy, as defined in Section 3, requires a minimum distance z_{min} between the sensor and the target for the user specified degree of privacy to be enforced. This constraint can be eliminated, for an optically defocused near-infrared (NIR) sensor, by placing a NIR light source adjacent to the sensor, such that nearby faces are overexposed. Since humans cannot see NIR light, the intensity of the light source can be adujsted indiscriminately without bothering passersby. However, overexposure significantly reduces data utility. Fortunately, due to the inverse square falloff rate of light, if the intensity of the light source is set to the lowest setting that overexposes faces at z_{min}, the overexposing light source has little effect on image quality for parts of the scene with distances greater than z_{min}.

We validated our design with a proof-of-concept experiment. The experimental setup consisted of an IR light source adjacent to a defocused webcam, with sensitivity in to NIR light, placed in an otherwise unilluminated office (Fig. 6(II,III)). In our setup, faces were overexposed for distances less than $z_{min} = 1.8m$ and were appropriately exposed, but defocused for distances greater than $2.3m$. Fig. 6(I) shows a sequence of four private images of a person standing $3.0, 2.4, 1.8,$ and 0.6 meters from the sensor. 6(IV) shows an image of a person standing $0.6m$ without the

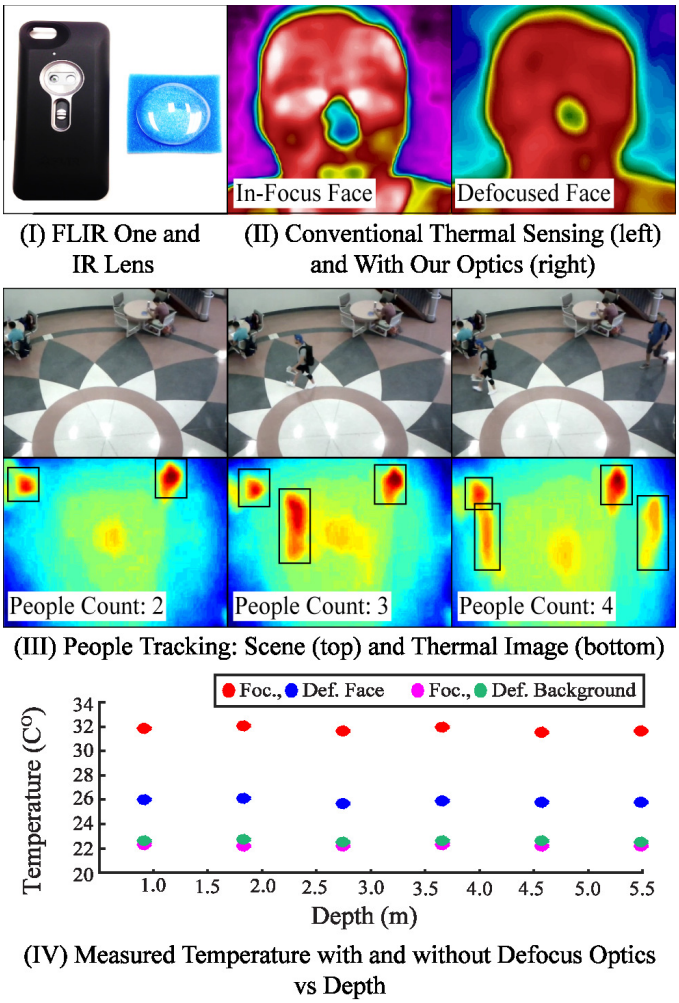

Fig. 7. **Privacy Preserving People Tracking**. We fitted a FLIR One Thermal sensor with an IR Lens to enable privacy preserving people tracking via pre-capture optical Gaussian blurring. (I) shows the FLIR One and the IR Lens. (II) shows and image of a face taken with and without the IR Lens fitted to the FLIR One. Using this system, we were able to easily perform people tracking by searching for high intensity blobs in the optically de-identified thermal images (III). Our defocusing optics preserve approximate temperatures for objects larger than the largest target feature. We measured the temperatures of a human head and the background wall at various distances from the sensor, with and without our defocusing optics, using the native FLIR One thermal calibration. A graph of the measured temperature with and without our defocusing optics vs depth is shown in (IV).

overexposing IR light source. Since $0.6m < z_{min}$, this image is not private.

3.3.3 *Defocused Thermal Sensor*

We fitted a FLIR One thermal camera with an IR Lens (Fig. 7(I)) to enable privacy preserving thermal sensing via optical defocus. The modified sensor had an angular support of $0.9855°$, which corresponds to a minimum distance, $z_{min} = 4.6$m for degrading features of 8cm, lensless parameters $u = 2$mm, $d = 1.29$mm, and and eFOV of $50.8°$ for $\Delta = 0.2°$. Figure 7(II) shows a human face captured with and without our defocusing optics.

Our defocusing optics preserve approximate temperatures for objects larger than the largest target feature. Using the native FLIR One thermal calibration, we measured the temperatures of a human head and an adjacent position on the wall behind the head

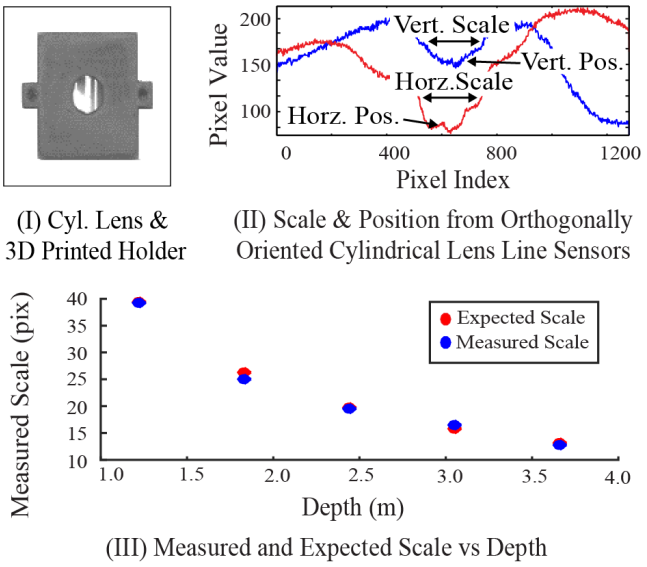

Fig. 8. **Scale and Position from Anisotropically Defocused Linear Sensors.** Building on [23], we show a pre-capture privacy preserving scale detection two dimensional scene analysis with anisotropically defocused linear sensors. Our implementation consisted of to two Lu171 Lumenera monochrome sensors (Fig. 9) each fitted with the 3D printed lens holder and 6mm focal cylindrical lens, shown in (I). Only a single row of pixels from each sensor was used in the analysis to simulate using linear sensors. (II) shows how two dimensional scale and position can be extracted from the local extrema of the sensor outputs, the horizontal and vertical brightness distributions of the scene. Using only the horizontally oriented sensor from this setup, we calculated the horizontal scale of a person standing $\{1.2, 1.8, 2.4, 3.0, 3.7\}$meters away from the sensor. (III) shows the measured vs expected scale for each distance. The mean and standard deviation of the absolute error in the measured vs expected scale were 0.56pix and 0.46pix respectively.

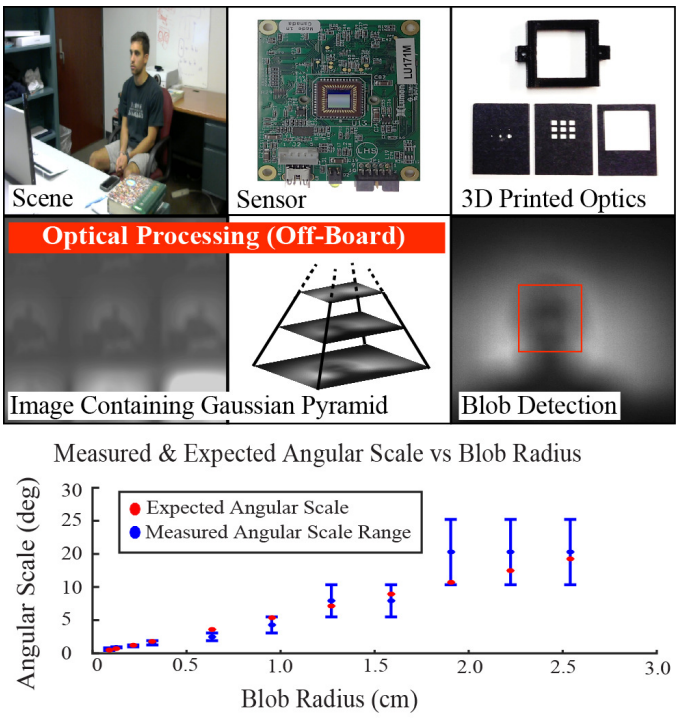

Fig. 9. **Privacy Preserving Angular Blob Detection.** Our privacy preserving optical blob detector uses a Lumenera Lu-171 sensor and 3D printed/laser cut optics. The sensor was divided into multiple elements, where each performs pre-capture optical defocus filtering of different aperture radii. Therefore, a single frame contains a gaussian pyramid which can be used for angular blob detection. Using our prototype we measured the angular scale range of white circles of varying sizes on black background located 20.32cm from the sensor. A graph of the measured angular scale range and the expected angular scale vs circle radius is the bottom row of the figure.

(background wall), at various distances from the sensor, with and without our defocusing optics. The results are shown in Fig. 7(IV). Additionally, we performed privacy preserving people tracking for an alternate multi-person scene, by searching for high intensity blobs in the defocused thermal images (Fig. 7(III)). The subjects in the multi-person scene were more than 5.5m away from the sensor.

3.3.4 *Anisotropically Defocused Linear Sensor*

The defocusing effect of cylindrical lenses can be approximated in software, by an anisotropic Gaussian filter. Thus, we can use our design tool to design a linear sensor that captures the one-dimensional brightness distribution of scene. [23] showed that such a sensor can be used to privately determine a person's one-dimensional position and state (fallen or standing up). We extended [23]'s work in two ways. First, we showed that in addition to position and state, a person's scale can also be extracted from local extrema of the brightness distribution (Fig. 8(II)). Second, we showed two dimensional scene analysis (scale and position) by pairing two orthogonally oriented linear imaging arrays fitted with cylindrical lenses. Two-dimensional scale analysis for head detection in the Sec. 4.3.1.

Our implementation consisted of two Lu171 Lumenera monochrome sensors fitted with 3D printed optics holders and 6mm focal cylindrical lenses (Fig. 8(I)). Only a single row of pixels from each sensor was used in the analysis to simulate using linear sensors. Using only the horizontally oriented sensor from this setup, we calculated the horizontal scale of a person at various distances from the sensor. Fig. 8(I) shows the measured vs expected scale. The mean and standard deviation of the absolute error were 0.56pix and 0.46pix respectively.

3.3.5 *Defocused Sensor Array*

With the fixed-optics single aperture sensors, most of the actual vision computations (people counting, tracking, etc.) are performed post-capture. Here, we exploit sensor arrays, which have proved useful in many domains [67], to increase the the number of pre-capture vision computations.

A classical approach to blob detection is to convolve an image with a series of Laplacian of Gaussian (LoG) filters for scale-space analysis [68]. The LoG operators are usually approximated by differences of Gaussians (DoGs), and [49] demonstrated such computations with a single pair of lensless sensors. We build a lensless sensor array that perform both blob detection and privacy preserving defocus together. This partitions the photodetector into n sub-images with unique angular supports $\omega_{o_1} < \omega_{o_2} < \ldots < \omega_{o_n}$. Our prototype build with an aperture array and baffles is shown in Fig. 9. In a single shot, the sensor directly captures an image's Gaussian pyramid. When compared with a software implementation of a Gaussian pyramid, our optical array enables privacy preservation before capture. The degree of privacy afforded is directly related to the minimum angular defocus kernel ω_{o_1}. The element with the least eFOV determines the array's

eFOV (although this is relaxed in the next section). Finally, the privacy preserving advantage of these arrays comes with tradeoffs; for example, the optical array provides a fixed sampling of the scale space (scale granularity) and can estimate blobs only in a fixed scale range.

We built a privacy preserving angular scale-space blob detector. In Fig. 9 we show our prototype, which consisted of a camera (Lu-171, Lumenera Inc.) with custom 3D-printed template assembly and binary templates cut into black card paper using a 100-micron laser (VLS3.50, Versa Inc.). We divided the camera photodetector plane into nine single-aperture sensor elements using opaque baffles created from layered paper to prevent crosstalk between the sensor elements. The Lu-171 has a resolution of 1280x1024 so the photodetector array was partitioned into a 3x3 array of 320x320 pixels.

We validated our theory experimentally by using a prototype to determine the angular scale range of white circles of varying sizes on black background located 20.32cm from the sensor. The results are shown in Fig. 9. The angular scale range was correctly detected for all of the circles except for the circle with radius 0.64, which was off by 0.52 degrees. The angular supports for the prototype used in the experiment were $\{\omega_{o_1} = 0.44°, \omega_{o_2} = 0.82°, \omega_{o_3} = 0.97°, \omega_{o_4} = 1.26°, \omega_{o_5} = 1.89°, \omega_{o_6} = 3.06°, \omega_{o_7} = 5.48°, \omega_{o_8} = 10.35°, \omega_{o_9} = 20.21°\}$.

As an example application, we developed a second prototype for low-power head tracking. Head-tracking was achieved by first detecting regions of interest using scale-space blob detection and then running Viola-Jones object on the detected regions of interest. Identifying regions of interest using scale-space analysis decreased the image search area for the Viola-Jones detector by 50%. Fig. 9. shows a frame from a 2 minute office sequence, where the head was tracked correctly in 98% of frames. The optical parameters of the prototype used to capture the office sequence were $\{\Delta = 4°, \omega_{o_1} = 9.76°, \omega_{o_2} = 20.28°, \omega_{o_3} = 40.37°\}$, which correspond to a minimum distance $z_{min} = 46.9$cm, for degrading features of 8cm and an eFOV of $39.54°$. The Viola-Jones object detector was trained and tested on images of head blobs moving in an various office scenes.

3.4 Limitations

Although, in this paper, we show various example vision tasks that work well with optically defocused data, there are many vision tasks for which this is not the case. Furthermore, although deblurring of heavily defocused images is an open problem [69], [70], [71], [72], [73], [74], [75], the use optical defocus for privacy may still be susceptible to reverse engineering.

4 Programmable Privacy Optics

In this section, we provide a framework, which leverages programmable privacy optics to enable pre-capture implementations of mask-based privacy algorithms such as k-anonymity and blackout.

4.1 Privacy Optics

The proposed framework, illustrated in Fig. 10(I), consists of an output sensor (approximated by an ideal pinhole camera), whose viewing path is split between the scene and an active optical mask, such as a projector or electronic display, and a pre-capture privacy preserving alignment sensor. With this setup, masking of sensitive

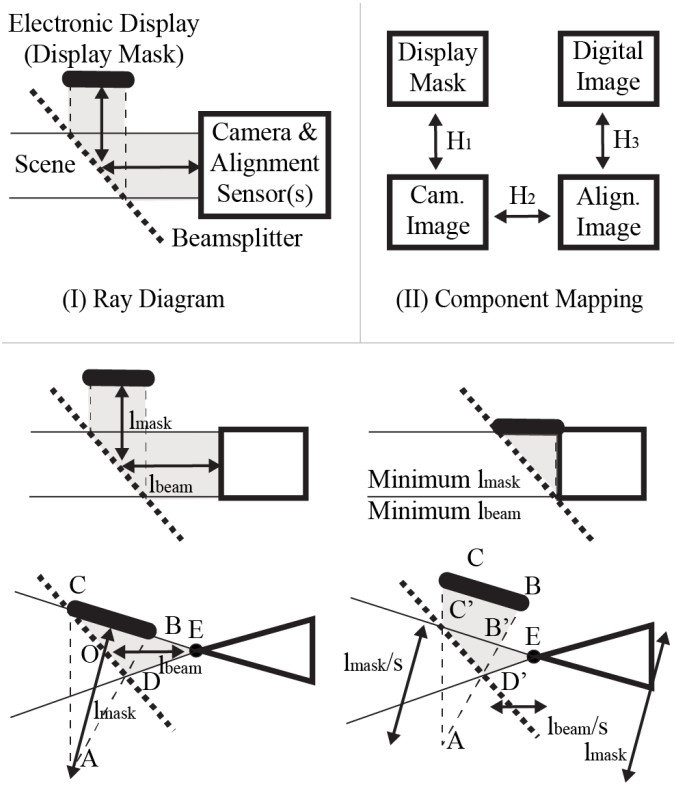

Fig. 10. **Programmable Optics for Pre-Capture Privacy.** In (I), we show a ray diagram for our programmable optics-based pre-capture privacy framework. In (II), we show the required mappings between the various system components. In (III), we demonstrate how to reduce the volume occupied by the display and beamsplitter, determined by l_{beam} and l_{mask}. For the perspective case, we show that there exists two configurations with identical, minimum volume.

targets requires five steps: 1) Capture alignment image using alignment sensor; 2) Segment targets in the alignment image; 3) Generate privacy masks using target segmentations; 4) Display privacy masks on electronic display; 5) Capture private image using output sensor. We formulate these five steps as follows.

The radiance I measured at each camera pixel (x, y) that views a scene point P is given by,

$$I(x,y) = e_P \; I_P + e_M I_{display}(M(x,y)), \tag{6}$$

where I_P is the radiance from P, M is the privacy mask displayed by the electronic display, $I_{display}$ maps a privacy mask pixel intensity to its displayed radiance, and e_P and e_M are the ratios of the optical path split between the scene and the mask, which can range from 0 to 1. $M(x, y)$ is given by

$$M(x,y) = \sum_{1 \le i \le t-1} F_i(H_1(x,y)) \tag{7}$$

where H_1 is a transformation between the camera and mask planes, and F_i are the masks for the t segmented target planes in the alignment image. The masks F_i are given by

$$F_i(x,y) = \begin{cases} 0, & L(x',y') = 0 \\ \sum\limits_{1 \le j \le k-1} w_j D_j(H_2(H_3(x,y))), & L(x',y') = 1 \end{cases} \tag{8}$$

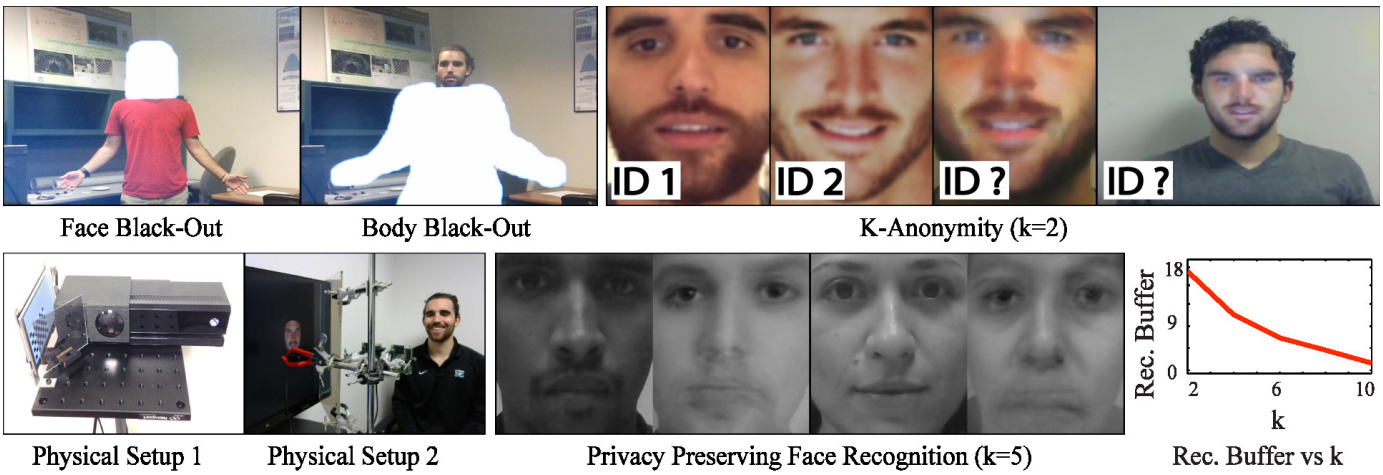

Fig. 11. **Pre-Capture Black-out, K-Anonymity, and Face Recognition.** We designed two programmable optics-based pre-capture privacy sensors. Our first prototype, *Physical Setup 1*, consisted of a 5" LED, a 2" beam splitter, and a Microsoft kinect fitted with our privacy sleeve (Fig. 5). The color camera on the kinect was used as the system camera and the defocused TOF was used as the segmentation sensor. We implemented a pre-caputure black-out using this prototype by masking the target with high intensity masks such that the pixels corresponding to the target were overexposed in the resulting camera images. Segmentation was acheived using the native Microsoft Kinect segmentation algorithm. Our second prototype, *Physical Setup 2*, consisted of an 27" LED, a 14" beam splitter, a webcam, and the pair orthogonally oriented defocused linear sensors from Fig. 8. The webcam was used as the system camera and the line sensors were used as the alignment sensors. Using this prototype we implemented pre-capture k-anonymity for faces and privacy preserving face recognition.

where $(x', y') = H_1(H_2(x, y))$, D_j are digital images, w_j are user defined weights, $L_i(x, y)$ is binary segmentation in the alignment image for target i, H_2 is a transformation between the segmented target plane and the camera, and H_3 is a transformation between each digital images and the segmented target plane. If the alignment sensor and the camera are collocated and of the same resolution, H_2 is an identity matrix. If they are not collocated, then H_2 varies for each segmented target plane and is depth dependent. Similarly, if the camera and the display mask are collocated and of the same resolution, H_1 is an identity matrix. A graphical illustration of these transformations is given in Fig. 10(II).

4.2 Miniaturization

We reduce the volume of the optics for small form factor platforms. For many algorithms it is desirable that the resolution of the display be equal to or greater than the resolution of the sensor. Here we discuss how to reduce the size of the optical setup while still maintaining the desired display resolution. We assume that the camera sensor in Fig. 10 is optimally miniaturized by a method such as [49]. For clarity we consider a 2D ray diagram, but since our optics are symmetric these arguments hold in three dimensions. Let the beamsplitter angle be fixed at ϕ and the sensor FOV be θ. Let the minimum size of the mask that still affords the desired resolution be M_{min}. W.l.o.g let the mask be perpendicular to the reflected optical axis.

This leaves just two degrees of freedom for the optics; the sensor-beamsplitter distance l_{beam} along the sensor's optical axis and the mask-beamsplitter distance l_{mask} along the reflected optical axis. In an orthographic version of optics, shown in Fig. 10 (I), the size of the mask does not change as it is translated towards the sensor. Therefore, a mask of minimum size M_{min} can be moved as close as possible to the sensor without occluding the field-of-view as in Fig. 10 (I).

In the perspective case [76] the size of the mask reduces as it slides along the pencil of rays, as in Fig. 10 (II). Once the minimum mask size M_{min} is reached, that configuration has the minimum optical size, given by $\triangle CDE$'s area.

We show that there exists an alternate choice, in the perspective case, for the minimum optical size. To maintain the minimum resolution, any mask position closer to the sensor must be vertically shifted, as in Fig. 10 (II). The area of these optics is given by $\triangle C^{'}D^{'}E + C^{'}B^{'}BC$. From similar triangles, we can write $\triangle C^{'}D^{'}E$ as being created from $\triangle CDE$ by a scale factor $\frac{1}{s}$, and then equate the two configurations in Fig. 10 (II),

$$\triangle CDE(1 - \frac{1}{s}) = C^{'}B^{'}BC. \tag{9}$$

Consider $\triangle CDE = \triangle COE + \triangle ODE$. From the angle-side-angle theorem, this becomes,

$$\triangle CDE = \frac{l_{beam}^2 \sin\frac{\theta}{2} \sin\phi}{2 \sin(\frac{\theta}{2} - \phi)} + \frac{l_{beam}^2 \sin\frac{\theta}{2} \sin\phi}{2 \sin(\frac{\theta}{2} + \phi)}. \tag{10}$$

Since $\triangle AB^{'}C^{'}$ is a scaled version of $\triangle ABC$, the quadrilateral area $C^{'}B^{'}BC =$

$$\triangle ABC(1 - \frac{1}{s^2}) = \frac{M_{min}l_{mask}}{2}(1 - \frac{1}{s^2}). \tag{11}$$

Putting Eq. 10 and Eq. 11 into Eq. 9, and setting constant $C_1 = \frac{\sin\frac{\theta}{2} \sin\phi}{2 \sin(\frac{\theta}{2} - \phi)} + \frac{\sin\frac{\theta}{2} \sin\phi}{2 \sin(\frac{\theta}{2} + \phi)}$,

$$s = \frac{M_{min}l_{mask}}{2C_1 l_{beam}^2 - M_{min}l_{mask}}, \tag{12}$$

which is an equation for the scaling factor s such that the two designs in Fig. 10 (II) have the same area. Therefore we have found two designs that provide the required resolution within the smallest optical dimensions.

4.3 Example Sensor Designs

4.3.1 Alignment with Defocused Line Sensors

In this section we show two example applications, k-anonymity for faces and privacy preserving face recognition. Our prototype, labeled *Physical Setup 2* in Fig. 11, consisted of an 27" LED for the display, a 14" beam splitter, a webcam, and a pair orthogonally oriented linear sensors sensors with 6mm focal length cylindrical lenses. The webcam was used as the output camera and the pair of line sensors were used as the alignment sensors. Output images were captured at 30 FPS.

K-anonymity for faces [8], [9] enables face de-identification by averaging together a target face image with $k-1$ of its neighbors (according to some similarity metric). The resulting average image has an algorithm-invariant face recognition rate upper bound of $\frac{1}{k}$. We present what is, to our knowledge, the first ever optical implementation of k-anonymity for faces. Target faces were averaged with $k-1$ faces from a database by displaying privacy masks, generated via linear combinations of $k-1$ software aligned faces. We assumed the $k-1$ faces, D_j in Eq. 8, were captured under similar illumination environments to the target face. In our current implementation, access to the database could allow an adversary to compromise anonymity. In future implementations we plan to randomize the value k, the choice of k neighbors and the blending weights w_i to make de-anonymity combinatorially hard.

Recent efforts have resulted in privacy preserving face recognition frameworks [77], [78], [79], [80]. Here we show a similar example application, using optical k-anonymity for faces, that allows recognition of membership to a class while preserving privacy. Each target is first anonymized via optical k-anonymity with k-1 faces corresponding to individuals that are not in the membership class and are not known to the party performing face recognition. The anonymized face is compared to each face in the membership class using a similarity metric. If the similarity score is greater than a threshold then the anonymized face is matched with that individual. With no match, the system returns the k-anonymized face. We simulated this system using two subsets of the FERET Database [81], each containing a single image of a set of people. For k = $\{2, 4, 6, 8, 10\}$, 100 individuals from one subset were randomly selected as targets and anonymized with their $k-1$ nearest neighbors found in the same subset by simulating the effect of the cylindrical lens by integrating the image vertically and matching with the cosine similarity. The similarity between this k-anonymized image and 11 other images from the second image subset was then computed using Face++'s verification algorithms [82]. One of these is the target image from the second image subset, while the remaining were randomly selected. Simulation results are shown in 11. Such a system was built. Fig. 11 shows examples where individuals were correctly discriminated.

4.3.2 Alignment with Defocused Time-of-Flight Sensor

Black-out is a well known mask-based privacy algorithm for preserving the anonymity of individuals in video data. We implemented pre-caputure black-out using the prototype, labeled *Physical Setup 1* in Fig. 11, which consisted of a 5" LED display mask, a 2" beam splitter, and a Microsoft kinect. The Kinect was fitted with 3D printed privacy optics to defocus the time-of-flight camera (Fig. 5). The color camera on the kinect was used as the output camera and the defocused TOF was used as the alignment sensor. Black-out was achieved by masking the target with high intensity masks such that the pixels corresponding to the targets were overexposed in the resulting output camera image. The native Microsoft Kinect segmentation software was used to segment the targets. Fig. 11(I) shows our results for black-out faces and bodies. Output images were captured at 15 FPS.

4.4 Limitations

Miniaturization is limited by use of a display because it commits the system to continuous power use. The primary privacy limitation is that the framework relies on post-capture pattern recognition and alignment. Thus, privacy may be compromised, if pattern recognition or alignment fails at any frame.

5 Summary

Most privacy preserving systems for computer vision, process images after capture. There exists a moment of vulnerability in such systems, *after* capture, when privacy has not yet been enforced. Our privacy preserving sensors filter the incident light-field *before* image capture, while light passes through the sensor optics, so sensitive information is never measured by the sensor. Within this framework, we introduce a programmable privacy optics that enable pre-capture implementations of mask-based privacy algorithms, such as black-out and k-anonymity, and fixed privacy optics that provide both privacy and utility for time-of-flight, thermal, and near-infrared sensors. We also show theory for miniaturizing the proposed designs, including a novel "optical knapsack" solution for finding a field-of-view-optimal arrangement of optical elements. Our privacy preserving sensors enable applications such as accurate depth sensing, full-body motion tracking, multiple people tracking and low-power blob detection.

References

[1] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: automatically replacing faces in photographs," in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3. ACM, 2008, p. 39.

[2] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 2000, pp. 1–10.

[3] C. Neustaedter, S. Greenberg, and M. Boyle, "Blur filtration fails to preserve privacy for home-based video conferencing," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 13, no. 1, pp. 1–36, 2006.

[4] G. Loukides and J. Shao, "Data utility and privacy protection trade-off in k-anonymisation," in *Proceedings of the 2008 international workshop on Privacy and anonymity in information society*. ACM, 2008, pp. 36–45.

[5] C. Thorpe, F. Li, Z. Li, Z. Yu, D. Saunders, and J. Yu, "A coprime blur scheme for data security in video surveillance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 12, pp. 3066–3072, 2013.

[6] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 8, pp. 1168–1174, 2008.

[7] F. Li, Z. Li, D. Saunders, and J. Yu, "A theory of coprime blurred pairs," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 217–224.

[8] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

[9] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 2, pp. 232–243, 2005.

[10] R. Gross, L. Sweeney, F. De la Torre, and S. Baker, "Semi-supervised learning of multi-factor models for face de-identification," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[11] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *Privacy Enhancing Technologies*. Springer, 2006, pp. 227–242.

[12] B. Driessen and M. Dürmuth, "Achieving anonymity against major face recognition algorithms," in *Communications and Multimedia Security*. Springer, 2013, pp. 18–33.

[13] P. Agrawal and P. Narayanan, "Person de-identification in videos," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 3, pp. 299–310, 2011.

[14] J. Fernández-Berni, R. Carmona-Galán, R. del Río, R. Kleihorst, W. Philips, and Á. Rodríguez-Vázquez, "Focal-plane sensing-processing: A power-efficient approach for the implementation of privacy-aware networked visual sensors," *Sensors*, vol. 14, no. 8, pp. 15 203–15 226, 2014.

[15] M. Mrityunjay and P. Narayanan, "The de-identification camera," in *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2011 Third National Conference on*. IEEE, 2011, pp. 192–195.

[16] A. Chattopadhyay and T. E. Boult, "Privacycam: a privacy preserving camera using uclinux on the blackfin dsp," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[17] T. Winkler and B. Rinner, "Trustcam: Security and privacy-protection for an embedded smart camera based on trusted computing," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*. IEEE, 2010, pp. 593–600.

[18] G. R. Nelson, G. A. Jullien, and O. Yadid-Pecht, "Cmos image sensor with watermarking capabilities," in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*. IEEE, 2005, pp. 5326–5329.

[19] T. Winkler, A. Erdélyi, and B. Rinner, "Trusteye. m4: Protecting the sensornot the camera," in *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*. IEEE, 2014, pp. 159–164.

[20] J. Fernandez-Berni, R. Carmona-Galan, and A. Rodriguez-Vazquez, "Single-exposure hdr technique based on tunable balance between local and global adaptation."

[21] F. Pittaluga, A. Zivkovic, and S. J. Koppal, "Sensor-level privacy for thermal cameras," in *2016 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2016, pp. 1–12.

[22] S. Browarek, "High resolution, low cost, privacy preserving human motion tracking system via passive thermal sensing," Ph.D. dissertation, Massachusetts Institute of Technology, 2010.

[23] S. Nakashima, Y. Kitazono, L. Zhang, and S. Serikawa, "Development of privacy-preserving sensor for person detection," *Procedia-Social and Behavioral Sciences*, vol. 2, no. 1, pp. 213–217, 2010.

[24] M. O'Toole, R. Raskar, and K. N. Kutulakos, "Primal-dual coding to probe light transport." *ACM Trans. Graph.*, vol. 31, no. 4, p. 39, 2012.

[25] J. Dai, J. Wu, B. Saghafi, J. Konrad, and P. Ishwar, "Towards privacy-preserving activity recognition using extremely low temporal and spatial resolution cameras," in *The Sixth IEEE Workshop on Analysis and Modeling of Faces and Gestures, CVPR 2015*.

[26] Y. Zhang, Y. Lu, H. Nagahara, and R.-I. Taniguchi, "Anonymous camera for privacy protection," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, Aug 2014, pp. 4170–4175.

[27] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarbotham, D. Takhar, K. Kelly, and R. Baranuik, "An architecture for compressive imaging," *ICIP*, 2006.

[28] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. E. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, p. 83, 2008.

[29] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 64 980H–64 980H.

[30] S. Zhou, J. Lafferty, and L. Wasserman, "Compressed and privacy-sensitive sparse regression," *Information Theory, IEEE Transactions on*, vol. 55, no. 2, pp. 846–866, 2009.

[31] A. M. Abdulghani and E. Rodriguez-Villegas, "Compressive sensing: from compressing while sampling to compressing and securing while sampling," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 1127–1130.

[32] M. J. Wainwright, M. I. Jordan, and J. C. Duchi, "Privacy aware learning," in *Advances in Neural Information Processing Systems*, 2012, pp. 1430–1438.

[33] M. Cossalter, M. Tagliasacchi, and G. Valenzise, "Privacy-enabled object tracking in video sequences using compressive sensing," in *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*. IEEE, 2009, pp. 436–441.

[34] W. Wolf, B. Ozer, and T. Lv, "Smart cameras as embedded systems," *Computer*, vol. 35, no. 9, pp. 48–53, 2002.

[35] V. Brajovic and T. Kanade, "Computational sensor for visual tracking with attention," *Solid-State Circuits, IEEE Journal of*, vol. 33, no. 8, pp. 1199–1207, 1998.

[36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[37] B. Gyselinckx, C. Van Hoof, J. Ryckaert, R. Yazicioglu, P. Fiorini, and V. Leonov, "Human++: autonomous wireless sensors for body area networks," in *Custom Integrated Circuits Conference, Proceedings of the IEEE 2005*. IEEE, 2005, pp. 13–19.

[38] A. Chandrakasan, N. Verma, J. Kwong, D. Daly, N. Ickes, D. Finchelstein, and B. Calhoun, "Micropower wireless sensors," *Power*, vol. 30, no. 35, p. 40, 2006.

[39] B. H. Calhoun, D. C. Daly, N. Verma, D. F. Finchelstein, D. D. Wentzloff, A. Wang, S. Cho, and A. P. Chandrakasan, "Design considerations for ultra-low energy wireless microsensor nodes," *Computers, IEEE Transactions on*, vol. 54, no. 6, pp. 727–740, 2005.

[40] A. P. Sample, D. J. Yeager, P. S. Powledge, A. V. Mamishev, and J. R. Smith, "Design of an rfid-based battery-free programmable sensing platform," *Instrumentation and Measurement, IEEE Transactions on*, vol. 57, no. 11, pp. 2608–2615, 2008.

[41] E. Steltz and R. S. Fearing, "Dynamometer power output measurements of miniature piezoelectric actuators," *Mechatronics, IEEE/ASME Transactions on*, vol. 14, no. 1, pp. 1–10, 2009.

[42] A. Wilhelm, B. Surgenor, and J. Pharoah, "Evaluation of a micro fuel cell as applied to a mobile robot," in *Mechatronics and Automation, 2005 IEEE International Conference*, vol. 1. IEEE, 2005, pp. 32–36.

[43] J. W. Goodman *et al.*, *Introduction to Fourier optics*. McGraw-hill New York, 1968, vol. 2.

[44] F. T. Yu and S. Jutamulia, "Optical pattern recognition," *Optical Pattern Recognition, by Francis TS Yu, Suganda Jutamulia, Cambridge, UK: Cambridge University Press, 2008*, vol. 1, 2008.

[45] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 795–804, 2006.

[46] R. Ng, "Fourier slice photography," in *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 735–744.

[47] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3. ACM, 2007, p. 70.

[48] R. Fergus, A. Torralba, and W. T. Freeman, "Random lens imaging," 2006.

[49] S. J. Koppal, I. Gkioulekas, T. Young, H. Park, K. B. Crozier, G. L. Barrows, and T. Zickler, "Toward wide-angle microvision sensors," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 12, pp. 2982–2996, 2013.

[50] S. J. Koppal, I. Gkioulekas, T. Zickler, and G. L. Barrows, "Wide-angle micro sensors for vision on a tight budget," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 361–368.

[51] A. Zomet and S. K. Nayar, "Lensless imaging with a controllable aperture," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 339–346.

[52] S. K. Nayar, V. Branzoi, and T. E. Boult, "Programmable imaging: Towards a flexible camera," *International Journal of Computer Vision*, vol. 70, no. 1, pp. 7–22, 2006.

[53] U. M. Erdem and S. Sclaroff, "Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements," *Computer Vision and Image Understanding*, vol. 103, no. 3, pp. 156–169, 2006.

[54] A. O. Ercan, D. B. Yang, A. El Gamal, and L. J. Guibas, "Optimal placement and selection of camera network nodes for target localization," in *Distributed computing in sensor systems*. Springer, 2006, pp. 389–404.

[55] S. Soro and W. B. Heinzelman, "On the coverage problem in video-based wireless sensor networks," in *Broadband Networks, 2005. BroadNets 2005. 2nd International Conference on*. IEEE, 2005, pp. 932–939.

[56] A. O. Ercan, D. B. Yang, A. E. Gamal, and L. J. Guibas, "On coverage issues in directional sensor networks: A survey," *Ad Hoc Networks*, vol. 9, no. 7, pp. 1238–1255, 2011.

[57] K. Miyamoto, "Fish eye lens," *JOSA*, vol. 54, no. 8, pp. 1060–1061, 1964.

[58] R. Swaminathan, M. D. Grossberg, and S. K. Nayar, "Caustics of catadioptric cameras," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 2–9.

[59] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," 2000.

[60] F. Dufaux and T. Ebrahimi, "A framework for the validation of privacy protection solutions in video surveillance," in *Multimedia and Expo (ICME), 2010 IEEE International Conference on*. IEEE, 2010, pp. 66–71.

[61] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper, "The csu face identification evaluation system: Its purpose, features and structure," 2003.

[62] E. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying facial images," *CMU Technical Report CMU-CS-03-119*, 2003.

[63] H. Dyckhoff, "A typology of cutting and packing problems," *European Journal of Operational Research*, vol. 44, no. 2, pp. 145–159, 1990.

[64] S. Martello and P. Toth, *Knapsack problems: algorithms and computer implementations*. John Wiley & Sons, Inc., 1990.

[65] R. E. Korf, M. D. Moffitt, and M. E. Pollack, "Optimal rectangle packing," *Annals of Operations Research*, vol. 179, no. 1, pp. 261–295, 2010.

[66] F. Pittaluga and S. J. Koppal, "Privacy preserving optics for miniature vision sensors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 314–324.

[67] L. Streeter and A. A. Dorrington, "Simple harmonic error cancellation in time of flight range imaging," *Optics letters*, vol. 40, no. 22, pp. 5391–5394, 2015.

[68] T. Lindeberg, *Scale-space theory in computer vision*. Springer Science & Business Media, 1993.

[69] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring face images with exemplars," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 47–62.

[70] H. Zhang, J. Yang, Y. Zhang, N. M. Nasrabadi, and T. S. Huang, "Close the loop: Joint blind image restoration and recognition with sparse representation prior," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 770–777.

[71] M. Nishiyama, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi, "Facial deblur inference to improve recognition of blurred faces," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1115–1122.

[72] M. Nishiyama, A. Hadid, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi, "Facial deblur inference using subspace analysis for recognition of blurred faces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 4, pp. 838–845, 2011.

[73] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *Image processing, IEEE Transactions on*, vol. 13, no. 10, pp. 1327–1344, 2004.

[74] B. Bascle, A. Blake, and A. Zisserman, "Motion deblurring and super-resolution from an image sequence," in *Computer Vision ECCV'96*. Springer, 1996, pp. 571–582.

[75] W. Dong, D. Zhang, G. Shi, and X. Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *Image Processing, IEEE Transactions on*, vol. 20, no. 7, pp. 1838–1857, 2011.

[76] J. Gluckman and S. K. Nayar, "Catadioptric stereo using planar mirrors," *International Journal of Computer Vision*, vol. 44, no. 1, pp. 65–79, 2001.

[77] A.-R. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *Information, Security and Cryptology–ICISC 2009*. Springer, 2010, pp. 229–244.

[78] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, "Privacy-preserving face recognition," in *Privacy Enhancing Technologies*. Springer, 2009, pp. 235–253.

[79] M. Osadchy, B. Pinkas, A. Jarrous, and B. Moskovich, "Scifi-a system for secure face identification," in *Security and Privacy (SP), 2010 IEEE Symposium on*. IEEE, 2010, pp. 239–254.

[80] T. A. Kevenaar, G. J. Schrijen, M. van der Veen, A. H. Akkermans, and F. Zuo, "Face recognition with renewable and privacy preserving binary templates," in *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*. IEEE, 2005, pp. 21–26.

[81] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[82] H. Fan, Z. Cao, Y. Jiang, Q. Yin, and C. Doudou, "Learning deep face representation," *arXiv preprint arXiv:1403.2802*, 2014.

Francesco Pittaluga Francesco Pittauga is currently a graduate student at the University of Florida's ECE department. Prior to joining UF, Francesco attended Tufts University where he received a B.S. in Electrical Engineering with a second major in Computer Science. His interests span computer vision, machine learning, and computational photography.

Sanjeev J. Koppal Sanjeev J. Koppal is an assistant professor at the University of Florida's ECE department. Prior to joining UF, he was a researcher at the Texas Instruments Imaging R&D lab. Sanjeev obtained his Masters and Ph.D. degrees from the Robotics Institute at Carnegie Mellon University, where his adviser was Prof. Srinivasa Narasimhan. After CMU, he was a post-doctoral research associate in the School of Engineering and Applied Sciences at Harvard University, with Prof. Todd Zickler. He received his B.S. degree from the University of Southern California in 2003. His interests span computer vision, computational photography and optics and include novel cameras and sensors, 3D reconstruction, physics-based vision and active illumination.