IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2025.30371

Multiple Point-Source Tracking with Lensless Single Random Phase Encoding Optical Systems

SAURABH GOSWAMI, JIHEON LEE, GOKUL KRISHNAN, and BAHRAM JAVIDI
Electrical Engineering Department, University of Connecticut, 371 Fairfield Road, Unit 4157 Storrs, Connecticut 06269, USA

Corresponding author: Bahram Javidi (e-mail: bahram.javidi@uconn.edu).

B. Javidi acknowledges support under The Office of Naval Research (ONR) (N000142212375; N000142212349); Air-Force Office of Scientific Research (AFOSR) (FA9550-21-1-0333); and AFOSR Biophysics (FA9550-24-1-0128) and National Science Foundation grant #2141473.

ABSTRACT In this paper, we demonstrate the capability of a lensless single random phase encoding (SRPE) system to track the lateral motion of multiple moving point-sources using a Probability Hypothesis Density (PHD) filter. Our system consists of an image sensor and a diffuser that captures the motion of point-sources by projecting a pseudorandom pattern on the image sensor. Using the theory of diffraction, we deduce the mapping between the locations of point-sources in the free-space and the pseudorandom pattern obtained from SRPE systems, obtain an initial maximum-likelihood estimate of the point-source locations, use a Markovian motion prior to predict the locations at subsequent time-instants and, use captured video frames to refine the predicted location estimates. Our results show that lensless SRPE systems, coupled with PHD filtering algorithm, can estimate lateral motion of point-sources accurately and in a computationally efficient manner. Moreover, lateral motion tracking with SRPE is robust to motion-blur and increase in image-sensor pixel size. Tracking with lensless SRPE systems also show robustness to dimensionality reduction of the captured pseudorandom patterns which have computational benefits. To the best of our knowledge, this is the first work to demonstrate lensless SRPE systems as an object tracking modality.

INDEX TERMS Lensless Imaging, Object Tracking, Optical Imaging, Single Random Phase Encoding, Tracking Point-Sources.

I. INTRODUCTION

Tracking, i.e., tracing out the trajectory of moving targets, is a crucial image processing and computer vision problem that finds extensive use in fields such as autonomous driving [1], astronomy [2], security and surveillance [3] and, satellite monitoring [4] to name a few. A significant subset of practical tracking tasks such as astronomy, surveillance of unwarranted aircraft and satellite monitoring is specifically concerned with tracking of objects which are located extremely far away, appearing almost as point-sources [4]. For such tasks, an imaging system which is inexpensive and capable of performing motion tracking of point-sources in a computationally efficient manner is highly beneficial. For most imaging-based tracking mechanisms currently in use, the imager of choice is lens-based [5]. The presence of lenses may make such systems bulky and expensive. Moreover, because lenses focus point-sources on an extremely small neighborhood (comprising only a few pixels) on image sensors, the downstream tracking algorithm needs to process a full-resolution image at each time-step. For these reasons, diffuser-based lensless imaging systems [6]–[17] are currently emerging as an attractive alternative to the conventional lens-based imaging systems. Due to the absence of lenses, such systems are lightweight, cost-effective and field-portable. Additionally, due to the scattering of diffusers, such systems spread the incoming optical information uniformly over the entire image sensor, making it possible to recover the full 3D field from only a small subset of image-sensor pixels [14]. This property makes such systems a good choice for a computationally efficient point-source tracking system. In this work, we focus on a specific variant of diffuser-based lensless imaging systems namely, lensless Single Random Phase Encoding (SRPE) systems [6]–[9], [11] for object tracking. These systems consist of a single diffuser combined with an image sensor and have been shown to have good resolution [11], noise-immunity [8], resilience to increased pixel-size [9], and robustness to dimensionality reduction [6]. They have also been demonstrated to be capable of disease classification without any need for post-imaging computational reconstruction [6], [7]. We aim to establish lensless SRPE systems as a novel modality for

performing lateral motion tracking of multiple point-sources from only small patches of captured full-resolution speckle images.

Achieving successful multitarget tracking requires solving a series of complicated challenges such as target association, motion prediction, motion correction, accounting for sudden appearance, disappearance and reappearance of targets, all under the same framework [18]. Researchers in this area have harnessed knowledge from various fields of mathematics and statistics such as Bayesian statistics, multivariate calculus, theory of random sets, to name a few, to devise increasingly more potent algorithms to perform tracking. Kalman filter [19], one of the classical algorithms that operates on the assumption of linear motion prediction model and a linear observation model, has long been a staple in this domain. The single target Bayesian Filter [20] generalizes Kalman Filter by defining motion prediction models and observation models as generalized probability density functions (PDFs) rather than deterministic estimates. The multitarget Bayesian Filter [21] uses the theory of Random Finite Sets and set calculus to cast the collection of targets as a set consisting of a random number of random vectors. The PDFs for multitarget Bayesian Filter are also multi-dimensional and intractable for a large number of targets. To approximate the multitarget Bayesian Filtering, Probability Hypothesis Density (PHD) filters [22], [23] have been proposed which, under certain assumptions, compress the multitarget PDFs to their corresponding multitarget first order moments, alternatively known as PHDs. PHDs, unlike PDFs which are defined on the cartesian product of the sample spaces of all the random variables under consideration, are defined on a single sample space. This makes PHDs and, consequently, PHD filtering operations much faster compared to multitarget Bayesian filtering [18].

In this work, we cast the lensless SRPE-based multiple point-source tracking problem in the framework of PHD filtering. Since our aim is not to propose a novel tracking algorithm but to demonstrate lensless SRPE systems as a promising tracking modality, we simplify the tracking problem by assuming that the number of targets is known and all the targets (i.e., point-sources) are present at all times. That is, there is no sudden disappearance, appearance or reappearance of targets. Such challenges can always be addressed by switching the PHD filter with a cardinalized-PHD (CPHD) filter or other more sophisticated tracking algorithms.

To perform the tracking operation, we first derive the relation between the world co-ordinates of point-sources and their corresponding speckle images to be captured by the lensless SRPE sensor using the principles of diffraction. We then use a lensless SRPE system to capture videos of moving point-sources. Afterwards, in the tracking algorithm, we derive an initial estimate of the source-locations by correlating the initial videoframe with the point spread function (PSF) of the lensless SRPE system and iterate PHD filtering along time to derive the trajectory of the moving point-sources. The results show that lensless SRPE systems, coupled with PHD filtering, can accurately estimate the lateral trajectory of point-sources moving in free-space. This work paves the way for developing a computationally inexpensive tracking device that would be useful in several fields. To the best of our knowledge, this work is the first report to demonstrate the utility of lensless SRPE systems in the task of point-source tracking.

Our main objective in this work is to demonstrate the potential of lensless single random phase encoding (SRPE) systems as a multiple point-source tracking modality. Most of the object tracking algorithms available in literature (PHD being one of them) are for lens-based images. Lensless SRPE systems do not produce human-interpretable images, but rather pseudorandom patterns. In this paper, we have proposed and implemented a pipeline through which lensless SRPE systems can be integrated into a Bayesian tracking framework. Our work is not meant as an improvement over PHD filters. Rather, it creates a pipeline through which lensless SRPE systems can be interfaced with most previously proposed tracking algorithms.

The rest of the paper is organized as follows: in Section II, we develop the mapping between the world co-ordinates of the point-sources and their corresponding speckle images obtained by the lensless SRPE, discuss the concepts and derivations of PHD filtering pertaining to this work and specify the metric used in the performance evaluation of the experiments. Section III contains the details of the experimental setup, the obtained results and the relevant discussions. Finally, Section IV concludes the paper by summarizing the findings of the paper and discussing future developments.

II. METHODOLOGY

A. POINT SPREAD FUNCTION OF SINGLE RANDOM PHASE ENCODING SYSTEMS

Our lensless diffuser-based SRPE system consists of a holographic diffuser and a lensless CMOS camera, similar to the system used in [8], [9], [11]. As shown in Fig. 1, optical fields emerging from point-sources travel a distance z_1 to be intercepted at the diffuser. The diffuser surface roughness spatially modulates the incoming field. The field emerging from the diffuser travels a known distance z_2 to reach the lensless image sensor which records its intensity.

Following [8], [9], [11], we use the well-developed principles in optics to describe and analyze the properties of the PSF of a lensless SRPE system. The world co-ordinates of point-sources moving in the free space are denoted as (x, y, z), assuming that the origin in this co-ordinate is located where the optical axis intersects the diffuser. The co-ordinates on the diffuser surface are indicated as (ζ, η), and the co-ordinates on the lensless image sensor are specified as (α, β), as depicted in Fig. 1. The wavelength of light radiating from a point-source is designated by λ.

Say, a point source is located at (x, y, z_1), emanating an optical field $u_0(x, y)$ such that,

$$u_0(x, y) = \delta(x - x_0, y - y_0) \tag{1}$$

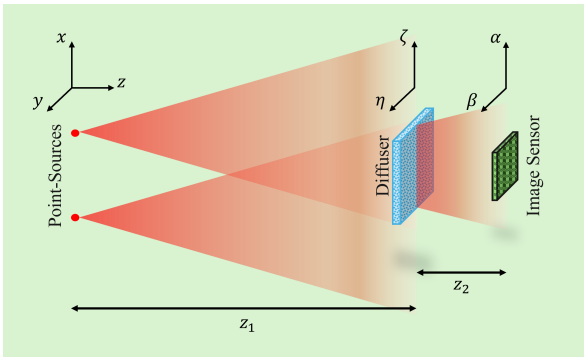

FIGURE 1. A schematic diagram of the point-source tracking system using an SRPE system.

where $\delta(\cdot,\cdot)$ is a dirac-delta function. This field travels a distance z_1 to reach the diffuser. The field $u_1(\zeta,\eta)$ right before the diffuser can be derived using Fresnel propagation:

$$u_1(\zeta,\eta;z_1) = u_0(\zeta,\eta) * \frac{e^{jkz_1}}{j\lambda z_1}\exp\left(\frac{j\pi}{\lambda z_1}(\zeta^2+\eta^2)\right) \quad (2)$$

where $k = 2\pi/\lambda$ is the propagation constant of the field. Fresnel propagation is valid when z_1 is much larger than the dimensions of all the optical components in the SRPE system, which we assume to be true throughout this analysis. The diffuser modulates the arrived field with a spatially random phase distribution $\phi(\zeta,\eta)$ which is one realization of a 2-dimensional (2D) uniform random process with the phases at each location ranging from $(-\pi,\pi]$. The correlation of this random process depends on the size of the surface features of the diffuser, i.e., the larger the feature, the wider the correlation function. Mathematically, the perturbation enacted by the diffuser on the incoming field can be described by a multiplicative diffuser-transmittance function $t_D(\zeta,\eta)$ [8], [9], [11] given by:

$$t_D(\zeta,\eta) = \exp\left(j\phi(\zeta,\eta)\right) \quad (3)$$

where, $j=\sqrt{-1}$. Hence, the field $u_1'(\zeta,\eta)$ emerging from the diffuser can be expressed as:

$$u_1'(\zeta,\eta;z_1) = u_1(\zeta,\eta;z_1)\times t_D(\zeta,\eta) \quad (4)$$

This field further travels a distance z_2 to reach the image sensor. The field $u_2(\alpha,\beta)$ at the image sensor can be obtained using Fresnel diffraction as follows:

$$u_2(\alpha,\beta;z_1) = u_1'(\alpha,\beta;z_1) * \frac{e^{jkz_2}}{j\lambda z_2}\exp\left(\frac{j\pi}{\lambda z_2}(\alpha^2+\beta^2)\right) \quad (5)$$

It is worthwhile to note at this point that unlike the previous works on this topic [8], [9], [11] that use angular spectrum propagation [24], we choose its paraxial approximation, i.e., Fresnel diffraction, to model the post-diffuser field propagation. Fresnel approximation holds up here because throughout the experiments, we have used diffusers with very small scattering angles (0.5° measured according to full-width half-maximum). The reasoning for this design choice will be discussed later in this section and in Section III-C. Due to the narrow scattering provided by the chosen diffuser, paraxial approximation is used even after being acted upon by the diffuser, making Fresnel diffraction a valid choice of wave propagation model.

The image sensor records the intensity $i(\alpha,\beta)$ of the incident field:

$$i(\alpha,\beta;z_1) = |u_2(\alpha,\beta;z_1)|^2 \quad (6)$$

where, the operator $|\cdot|$ computes the absolute value of a complex variable.

A thorough expansion of (6), ignoring all the constant terms (in the interest of brevity), begets the following expression:

$$i(\alpha,\beta;z_1) = \Big|\iint \exp\left(\frac{j\pi}{\lambda d}(\zeta^2+\eta^2)\right) t_D(\zeta,\eta)\times \exp\left(-\frac{j2\pi}{\lambda z_2}\left(\zeta\left(\alpha+\frac{z_2}{z_1}x_0\right)+\eta\left(\beta+\frac{z_2}{z_1}y_0\right)\right)\right) d\zeta d\eta\Big|^2 \quad (7)$$

where, $\times$ denotes multiplication, and the distance d is defined as follows:

$$\frac{1}{d} = \frac{1}{z_1}+\frac{1}{z_2} \quad (8)$$

Since the above integration is being performed over the entire diffuser plane, the intensity $i(\alpha,\beta;z_1)$ becomes a function of the sensor plane co-ordinates (α,β), the location of the point-source (x_0,y_0,z_1), and the diffuser-to-sensor distance z_2 which is fixed for an SRPE system. From (7), we see that if the point-spread function $h(\alpha,\beta;z_1)$ of a lensless SRPE system is captured by placing a point-source at location $(0,0,z_1)$, the intensity pattern $i(\alpha,\beta;x_0,y_0,z_1)$ for a point-source located at (x_0,y_0,z_1) can be written as:

$$i(\alpha,\beta;x_0,y_0,z_1) = h\left(\alpha+\frac{z_2}{z_1}x_0,\beta+\frac{z_2}{z_1}y_0\right) \quad (9)$$

If multiple point-sources are present, and they are all incoherent with respect to one another, as independent sources often are, the resulting intensity pattern would be a superposition of their individual intensity patterns. This has been discussed in Section II-D, (32).

It is worth noting at this point that paraxiality of diffuser field, as described by (5), is crucial for the linear shift-invariance of (9). This assumption holds true for a large range of z_2, but only for small scattering angle diffusers. Previous works [12]–[14] have also used 0.5° scattering angle diffusers to achieve linear shift-invariance due to the diffuser field being paraxial. For diffusers with large scattering angles, we see deviations from linear shift-invariance. To analyze

Diffuser scattering angle	Statistical properties of paraxial-deviation noise		
	Mean	Standard deviation	FWHM
0.5°	0.0139	0.0582	0.1370
30°	0.0025	0.1287	0.3031

TABLE 1. Statistical properties of the deviation noise n_d corresponding to 0.5° and 30° diffuser. FWHM is the full width half maximum width of the corresponding distribution.

this deviation, we model the experimentally obtained lensless SRPE speckle image $i(\alpha - x_0, \beta - y_0)$ of a shifted point-source as a linear combination of an accordingly shifted PSF $h(\alpha - x_0, \beta - y_0)$ and an additive noise $n_d(x, y)$:

$$i(\alpha - x_0, \beta - y_0) = h(\alpha - x_0, \beta - y_0) + n_d(x, y) \tag{10}$$

Note that the noise $n_d(x, y)$ is not camera noise. It arises due to the deviation from paraxiality of the field emanating from the corresponding diffuser. For convenience, we call this noise the paraxial-deviation noise. Fig. 2 shows the PDF of the paraxial-deviation noise $n_d(x, y)$ for a 0.5° (Fig. 2a) and a 30° scattering angle diffuser (Fig. 2b). Table 1 lists the statistical properties (mean, standard deviation, width of the distribution) of paraxial-deviation noise PDFs shown in Fig. 2. FWHM or Full Width Half Maximum in Table 1 refers to width of the distribution measured between the points at which the PDF attains half of its maximum value.

Since a widely scattering diffuser (30° scattering angle in this case) imparts a paraxial-deviation noise with higher standard deviation than a narrow scattering diffuser ($\sim$ 2.2 times higher than a 0.5° scattering angle diffuser), it produces an image sensor pattern with a significantly worse signal-to-noise ratio. Since such a noise has a higher likelihood of producing false peaks at the correlation pattern calculated during the algorithm (see (35)), the patch size of the image sensor pattern needs to be large to eliminate the adverse effect of this noise. This is also why using a narrow scattering diffuser is a better choice for tracking operations.

Fig. 2b shows that high-scattering diffusers, a 30° diffuser in this case, have a more pronounced deviation-noise than low-scattering diffusers. Section III-C demonstrates how this deviation affects lateral motion tracking with such diffusers.

Next, we provide backgrounds on the tracking algorithm namely, PHD filtering, used in this work.

B. PROBABILITY HYPOTHESIS DENSITY FOR A KNOWN NUMBER OF POINT SOURCES

Probability Hypothesis Density (PHD) is the first moment density of a multitarget probability density. Say, we have a sample space S such that,

$$S = \{(x, y) : x \in \mathbb{R} \wedge y \in \mathbb{R}\} \tag{11}$$

on which the probability density functions (PDF) of individual point-sources have been defined. We assume that the PDF $f_i(x_i, y_i)$ of the i-th point-source is Gaussian with mean (μ_{xi}, μ_{yi}) and standard deviation $(\sigma_{xi}, \sigma_{yi})$, i.e.,

$$f_i(x_i, y_i) = \frac{1}{(\sqrt{2\pi})^2 \sigma_{xi}\sigma_{yi}} \times \exp\left(-\frac{1}{2}\left(\left(\frac{x_i - \mu_{xi}}{\sigma_{xi}}\right)^2 + \left(\frac{y_i - \mu_{yi}}{\sigma_{yi}}\right)^2\right)\right) \tag{12}$$

Hence, the joint PDF of all the point-sources would be:

$$f(X) = \prod_{i=1}^{N} f_i(x_i, y_i) \tag{13}$$

where $X = \{(x_i, y_i)\}_{i=1}^{N}$ is the set containing the locations of all the point-sources under study.

PHD $D_f(x, y)$ of the PDF $f(X)$ is the expected number of point-sources in a unit volume in S. To arrive at $D_f(x, y)$, we first start by trying to find the probability mass function of the number of point-sources ΔN on a small volume ΔS where,

$$\Delta S = dx \times dy \tag{14}$$

Assuming we have a known total N number of point-sources, if we attempt to calculate the individual probabilities of ΔN, we see

$$P(\Delta N = 1) = \sum_{i=1}^{N} \left((f_i(x, y)\Delta S) \prod_{j=1, j\neq i}^{N} (1 - f_j(x, y)\Delta S) \right) \tag{15}$$

$$P(\Delta N = 2) = \sum_{1 \le i \le j} (f_i(x, y)\Delta S)\,(f_j(x, y)\Delta S) \times \prod_{k=1, k\neq i,j}^{N} (1 - f_k(x, y)\Delta S) \tag{16}$$

Progressing along these lines, we see that, (replacing $f_i(x, y)$ with f_i for brevity),

$$P(\Delta N = r) = \sum_{i=1}^{\binom{N}{r}} \left(\prod_{j \in p_i \subseteq S_N} (f_j \Delta S) \prod_{k \in S_N - p_i} (1 - f_k \Delta S) \right) \tag{17}$$

where S_N is the set of all N number of point-sources, p_i is a subset of S_N containing exactly r number of sources and any p_i and p_j differ at least by one element. Collectively, $\{p_i\}_{i=1}^{\binom{N}{r}}$ describe the set of all possible subsets of S_N of size r. And, finally,

$$P(\Delta N = N) = \prod_{i=1}^{N} f_i \Delta S \tag{18}$$

Hence, the expected number of point-sources in a small area δS would be:

$$\mathbb{E}[\Delta N] = \sum_{i=1}^{N} i P(\Delta N = i) \tag{19}$$

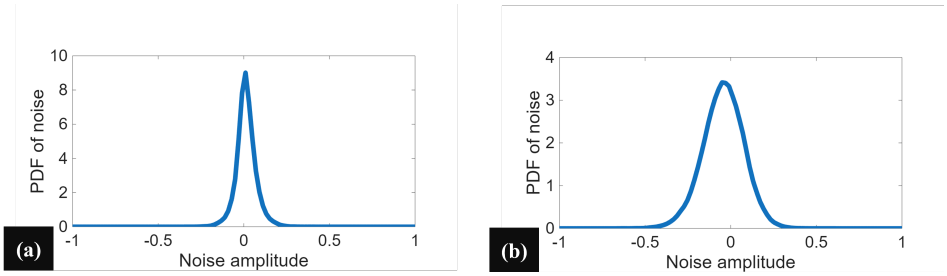

FIGURE 2. **PDF of noise inherent in signatures captured with (a) 0.5° scattering angle diffusers and, (b) 30° scattering angle diffusers. A component of this noise is camera noise. However, a significant part of this noise comes from the deviation of the experimental data from the strict linear shift invariance assumption.**

A simplification of (15)-(18) yields the following expression:

$$P(\Delta N = n) = \sum_{i=n}^{N} (-1)^{i+n} \binom{i}{n} (\Delta S)^i \sum_{j=1}^{\binom{N}{i}} C_i(j) \tag{20}$$

where C_i is the ordered set of all possible distinct combination of i number of sources and $C_i(j)$ is the j-th element of C_i. Hence, (19) simplifies to:

$$\mathbb{E}[\Delta N] = \Delta S \sum_{i=1}^{N} f_i \quad + \\ \sum_{i=2}^{N} (-1)^i \left((\Delta S)^i \sum_{j=1}^{\binom{N}{i}} C_i(j) \right) \left(\sum_{k=1}^{i} (-1)^k k \binom{i}{k} \right) \tag{21}$$

Now, let us consider the following identity:

$$g(x) = (1-x)^i = 1 + \sum_{k=1}^{i} (-1)^k \binom{i}{k} x^k \tag{22}$$

Differentiating on both sides with respect to x, we get,

$$g'(x) = -i(1-x)^{i-1} = \sum_{k=1}^{i} (-1)^k k \binom{i}{k} x^{k-1} \tag{23}$$

Evaluating (23) at $x = 1$, we get,

$$g'(1) = \sum_{k=1}^{i} (-1)^k k \binom{i}{k} = 0 \tag{24}$$

Substituting this in (21), we see that when the PDFs f_i are Gaussian distributed, the expected number of point-sources on a small area ΔS is:

$$\mathbb{E}[\Delta N] = \Delta S \sum_{i=1}^{N} f_i \tag{25}$$

Hence, the PHD $D_f(x,y)$ of this system, the expected number of point sources per unit volume, becomes:

$$D_f(x,y) = \frac{\mathbb{E}[\Delta N]}{\Delta S} = \sum_{i=1}^{N} f_i(x,y) \tag{26}$$

We would like to note that, if the location probabilities $f_i(x,y)$ were the same for all N number of sources, ΔN could have been modeled using a multinomial distribution and (25) and (26) could have been derived as its expectation in a straightforward manner. However, the location probabilities $f_i(x,y)$ are not identical for all i and they are changing independently during each iteration of our algorithm. This is why the formal definition of PHDs often involve algebraic operations on RFS [18]. However, operations on RFS can get complicated and may seem unintuitive for readers unfamiliar with the topic. To keep the mathematics intuitive, we have modeled the distribution of the number of sources using only the basic definitions of probability.

C. BACKGROUND ON PROBABILITY HYPOTHESIS DENSITY FILTERING FOR TRACKING

In this work, we have employed probability hypothesis density (PHD) filtering to achieve multi-target tracking. PHD [18], [22], [23] is the first order moment of a multisource joint PDF. It is a function whose integral over a region of a sample space gives the expected number of sources in that region [23]. PHD filtering is similar to Multitarget Bayesian filtering in spirit, with the only difference that probability density functions (PDFs) are replaced with PHDs. Like Bayesian filtering, PHD filtering is an iterative process. As the sensor receives data in a continuous stream, PHD filtering, at each time-step, analyzes the current data and puts it in the context of previous data to track the current location of the sources. Say, X_k denotes the location of the sources at time-step k, Z_k denotes the data collected by the sensor at time-step k and, $Z^{(k)}$ denotes the set of all data collected by the sensor till time-step k. Also, let us define $f_{k-1|k-1}\left(X_{k-1}|Z^{(k-1)}\right)$ as the PDF of the previous source-locations, $f_{k|k-1}\left(X_k|X_{k-1}\right)$ as the Markovian motion model, $f_{k|k}(Z_k|X_k)$ as the likelihood PDF of current source locations estimated from the data and, $f_{k|k}\left(X_k|Z^{(k)}\right)$ as the posterior source-location PDF. Their corresponding PHDs can be defined as $D_{k-1|k-1}\left(X_{k-1}|Z^{(k-1)}\right)$,

$D_{k|k-1}(X_k|X_{k-1})$, $D_{k|k}(Z_k|X_k)$ and $D_{k|k}\left(X_k|Z^{(k)}\right)$. Then, at each time-step k, PHD filtering performs two-steps:

(i) **Prediction:** Given the PHD $D_{k-1|k-1}\left(X_{k-1}|Z^{(k-1)}\right)$ of the previous source-locations, use the motion-model $f_{k|k-1}(X_k|X_{k-1})$ to predict the prior PHD $D_{k|k-1}\left(X_k|Z^{(k-1)}\right)$ of the current source-locations. This is performed using the following equation:

$$D_{k|k-1}\left(X_k|Z^{(k-1)}\right) = \int f_{k|k-1}\left(X_k|X'_{k-1}\right) D_{k-1|k-1}\left(X'_{k-1}|Z^{(k-1)}\right) \quad dX'_{k-1} \tag{27}$$

(ii) **Correction:** With the prior PHD $D_{k|k-1}\left(X_k|Z^{(k-1)}\right)$ at hand, use the data collected by the sensor to formulate the likelihood PHD $D_{k|k}(Z_k|X_k)$ and use Bayes theorem to arrive at the posterior PHD $D_{k|k}\left(X_k|Z^{(k)}\right)$ of the current source-locations. This is performed using the following equation.

$$D_{k|k} = \frac{D_{k|k}(Z_k|X_k)D_{k|k-1}\left(X_k|Z^{(k-1)}\right)}{D_{k|k-1}\left(Z_k|Z^{(k-1)}\right)} \tag{28}$$

where, the Bayesian normalization factor at the denominator is calculated using the following equation:

$$D_{k|k-1}\left(Z_k|Z^{(k-1)}\right) = \int D_{k|k}(Z_k|X_k)D_{k|k-1}\left(X_k|Z^{(k-1)}\right) \quad dX_k \tag{29}$$

As time progresses, PHD filter performs the prediction and correction steps repeatedly to update the estimated location of the point-sources. That is, it executes the following progression:

$$D_{0|0}\left(X_0|Z^{(0)}\right) \to D_{1|0}\left(X_1|Z^{(0)}\right) \to D_{1|1}\left(X_1|Z^{(1)}\right) \to D_{2|1}\left(X_2|Z^{(1)}\right) \to D_{2|2}\left(X_2|Z^{(2)}\right) \to \cdots \tag{30}$$

where $D_{0|0}\left(X_0|Z^{(0)}\right)$ is the PHD of initial location of the point-sources. If we don't have information regarding the initial location of the point-sources, we can always specify the corresponding PDF to be uniform.

Now, for N number of point-sources, the location X_k would consist of N number of 2-dimensional vectors, i.e.,

$$X_k = \{[x_i, y_i]^T : i \in [1, N]\} \tag{31}$$

where (x_i, y_i) is the real-world location of the i-th point-source. Hence, the joint sample space of all N point-sources would be a cartesian product of N number of 2-dimensional sample spaces and would have 2^N dimensions. Since all the above PDFs are functions defined over this sample space, they would also be 2^N-dimensional functions. This high dimensionality renders the computation of Bayesian filtering prohibitively expensive and intractable. However, PHD is defined over a single 2-dimensional sample space and allows us to compress the PDFs and still be able to perform Bayesian filtering.

D. FORMULATING PHDS FOR POINT-SOURCE TRACKING

As we have seen in Section II-A, especially in (9), the lateral movement of a point-source results in a shift in the PSF of a lensless SRPE system. Now, if we have an N number of distinct point-sources, they would, most likely, not be coherent with one another. Hence, when all these sources are present in a scene, the image $Z_k(\alpha, \beta)$ captured by the sensor at any time-instant k would be a linear summation of their individual intensity patterns [24], i.e.:

$$Z_k(\alpha, \beta) = \sum_{j=1}^{N} h\left(\alpha + \frac{z_2}{z_1}x_j(k), \beta + \frac{z_2}{z_1}y_j(k)\right) + n(\alpha, \beta) \tag{32}$$

where, $(x_j(k), y_j(k))$ is the real-world location of the j-th point-source at the time-instant k and, $n(\alpha, \beta)$ is a normally distributed image sensor noise. The set X_k of the locations of all point-sources at the k-th time instant is defined as in (31). Since all the point-sources are independent, the likelihood PDF $f_{k|k}(Z_k|X_k)$ would be,

$$f_{k|k}(Z_k|X_k) = \prod_{j=1}^{N} \mathcal{N}([x_j(k), y_j(k)]^T, \mathbf{\Sigma}_l) \tag{33}$$

where, $\mathcal{N}\left([x_j(k), y_j(k)]^T, \mathbf{\Sigma}_l\right)$ is a multivariate normal distribution with mean $[x_j(k), y_j(k)]^T$ and covariance matrix $\mathbf{\Sigma}_l$. Note that the likelihood PDF is a function defined on the sample space of X_k. X_k denotes the set of locations of all N independent point sources each of which are located at coordinates $[x_j(k), y_j(k)]^T$ on an (α, β) grid. This is why the joint conditional PDF $f_{k|k}(Z_k|X_k)$ is a product of N independent conditional PDFs where the conditioning variables are locations of individual targets. Each such conditional PDFs are functions over the (α, β) grid. Since $[x_j(k), y_j(k)]^T$ are the true locations of the targets that would give rise to the observation Z_k, each such PDFs would be normally distributed around the mean $[x_j(k), y_j(k)]^T$. The Gaussianity of the PDF comes from the fact that the camera noise $n(\alpha, \beta)$ in (32) is Gaussian.

Following the discussion in Section II-B, the likelihood PHD $D_{k|k}(Z_k|X_k)$ would be:

$$D_{k|k}(Z_k|X_k) = \sum_{j=1}^{N} \mathcal{N}\left([x_j(k), y_j(k)]^T, \mathbf{\Sigma}_l\right) \tag{34}$$

Note that (34) looks like a mixture of Gaussian. It has peaks wherever the individual normal PDFs have peak. Now, recall the PSF $h(\alpha, \beta)$ of a lensless SRPE system. If we now calculate the correlation $R_{\text{Zh}}(\alpha, \beta)$ between $Z_k(\alpha, \beta)$ and $h(\alpha, \beta)$, we get the following result:

$$R_{\text{Zh}}(\alpha, \beta) = \sum_{j=1}^{N} r\left(\alpha - \frac{z_2}{z_1}x_j(k), \beta - \frac{z_2}{z_1}y_j(k)\right) \tag{35}$$

where $r(\alpha, \beta)$ is autocorrelation of $h(\alpha, \beta)$, a function which has a peak at the origin and falls rapidly to 0 depending

on the speckle size of the diffuser used in the system. (35) can be very quickly computed by using a fast Fourier transform algorithm. Note that, $R_{\text{Zh}}(\alpha, \beta)$ looks almost exactly like $D_{k|k}(Z_k|X_k)$ with the only difference that the lateral co-ordinates are scaled by the factor z_2/z_1. This difference, however, vanishes if we define PHDs on the pixel co-ordinate plane rather than the world co-ordinate plane. We can always transfer the results to the world co-ordinate plane by simple co-ordinate scaling. Keeping this in mind, we see that deriving the PHD of (34) is simply a matter of fitting gaussians to $R_{\text{Zh}}(\alpha, \beta)$, which is how we derive the likelihood PHD $D_{k|k}(Z_k|X_k)$ for this work.

The prediction step of PHD filtering, as shown in (27) is also an expensive computational process due to the integration involved. However, conceptually speaking, we are simply using a pre-specified motion model to predict the PHD of future location of point-sources given the last location of the point-sources. This can be simplistically performed by extracting the last co-ordinates of the point sources, using the motion model to predict the next location, and defining the PHD to be gaussian around the predicted location with an uncertainty (variance). We model the motion by assuming constant acceleration between two successive time instants. Essentially, we only perform a likelihood-based tracking for the initial 3 time-instants to get $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{X}_2$. This gives us the following initial estimates of the locations, velocities and accelerations of the point-sources:

$$\mathbf{V}_0 = \frac{\mathbf{X}_1 - \mathbf{X}_0}{\Delta t}, \quad \mathbf{V}_1 = \frac{\mathbf{X}_2 - \mathbf{X}_1}{\Delta t} \tag{36}$$

$$\mathbf{A}_0 = \frac{\mathbf{V}_1 - \mathbf{V}_0}{\Delta t} \tag{37}$$

where Δt is the time-elapsed between any two successive instants. From the 4-th time-instant onwards, we perform the following steps to predict the location of the point-sources at the next instant given the same for the last time-instant:

$$\mathbf{X}_k = \mathbf{X}_{k-1} + \mathbf{V}_{k-1}\Delta t + \mathbf{A}_{k-1}\frac{\Delta t^2}{2} \tag{38}$$

$$\mathbf{V}_k = \alpha \mathbf{V}_{k-1} + (1-\alpha)\frac{\mathbf{X}_k - \mathbf{X}_{k-1}}{\Delta t} \tag{39}$$

$$\mathbf{A}_k = \beta \mathbf{A}_{k-1} + (1-\beta)\frac{\mathbf{V}_k - \mathbf{V}_{k-1}}{\Delta t} \tag{40}$$

where α and β are hyperparameters that ensure that the velocity and the accelerations maintain a smooth transition over time. A higher α and β ensure a smoother transition whereas a lower value would cause the transition to be jagged over time.

We would like to note here that all N number of point-sources have been assumed to move independently. That is, each one of $\mathbf{X}_k, \mathbf{V}_k, \mathbf{A}_k$ vectors are of dimension $2N \times 1$ (i.e., x and y co-ordinates corresponding to the locations of each one of N number of independently moving point-sources).

To arrive at the prediction PHD of (27), we simply perform the following:

$$D_{k|k-1}\left(\mathbf{X}_k|\mathbf{Z}^{(k-1)}\right) = \sum_{i=1}^{N} \mathcal{N}\left([x_i(k), y_i(k)], \mathbf{\Sigma}_p\right)$$
$$\forall \quad [x_i(k), y_i(k)] \in \mathbf{X}_k \tag{41}$$

where $\mathbf{\Sigma}_p$ is the covariance matrix capturing our uncertainty about the prediction.

E. METRIC FOR EVALUATION

For this task, our goal is to evaluate how accurately our algorithm can retrieve the ground-truth trajectory of the point-sources. For lateral motion tracking, the trajectories were stored as images. Hence, we chose the average (Avg) peak signal-to-noise ratio (PSNR) as a simple but intuitive metric to compare the estimated trajectories with the ground-truth trajectory. The metrics reported in this work were calculated using the following expression:

$$\text{AvgPSNR} = \frac{1}{N_D}\sum_{i=1}^{N_D} 10\log_{10}\left(\frac{(\max(g_i))^2}{\text{MSE}(g_i, r_i)}\right) \tag{42}$$

where, N_D is the number of videos in a dataset, g_i is the ground-truth trajectory of i-th video, r_i is the estimated trajectory recovered from the i-th video and $\text{MSE}(\cdot,\cdot)$ is the mean squared error between two matrices. Note that g_i and r_i are 2D matrices whose pixels contain probabilities of a point-source existing at the pixel location. The unit of this metric is in decibels or dB. A high AvgPSNR indicates a higher similarity between the ground-truth and the recovered trajectory and hence, a better tracking performance.

III. RESULTS AND DISCUSSIONS

In this section, we describe our experimental setups and discuss the results. As shown in Fig. 3, we used a laser pointer coupled with a 20X microscope objective (MO) to generate the point-sources. The microscope objective was used so that the incoming laser beam has a divergence at par with a realistic point-source. The laser had a wavelength of 650 nanometers. For imaging, we placed two identical Mako G-319 cameras side-by-side, one equipped with a diffuser and, another with a lens assembly. The lens-based imager was used to calculate ground truth trajectories. Unless otherwise specified, all point-sources were kept at a distance z_1 of 4.555 meters from the diffuser. The distance z_2 between the diffuser and the image sensor was maintained at 24.14 millimeters for all experiments. Each point-source (laser pointer and MO assembly) was moved by hand to generate distinct trajectories as reported later in this section.

Using this setup, we have imaged motions of (a) a single point-source and, (b) two independently moving point-sources. Unless otherwise specified, the diffuser used in this work has a 0.5° scattering angle and a diameter of 29.2 mm. Both the image sensors (lensless and lens-based) have (1600×1200) pixels, each of size 3.45 microns and an exposure time of 48 milliseconds. The lens has a focal length of

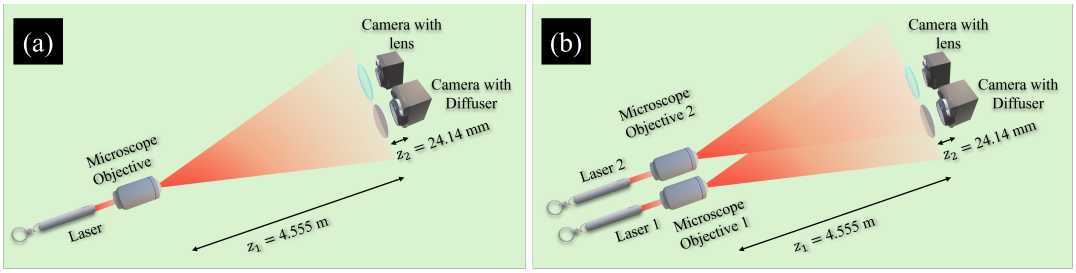

FIGURE 3. Data collection process for (a) single and, (b) multiple point-sources lateral tracking using a lensless SRPE and lens-based systems.

12.5 mm. The lateral motions of point sources were restricted to 1.22 meters horizontally and 1.045 meters vertically. All videos were captured at 10 frames per second unless otherwise specified.

Fig. 4a shows the PSF of the abovementioned lensless SRPE system with a 0.5° diffuser. Fig. 4c shows the histogram of this PSF. The PSF of the lensless SRPE system is a pseudorandom pattern owing to the surface roughness of the diffuser. This property makes it possible to recover the motion of the point sources from any contiguous subset of pixels of an image sensor. Fig. 4b shows the PSF of the lens-based imaging system.

We would like to note that even though the result section highlights the tracking capabilities of lensless SRPE systems, no comparison has been presented with the existing works on lensless SRPE system. This is due to the fact that our work happens to be the first one which aims to utilize lensless SRPE systems in the context of multiple point-source tracking. All previous publications on the subject of lensless SRPE systems deal with disease classification and robustness analysis of SRPE systems making our work a novel application modality for such systems.

A. LATERAL TRACKING OF POINT-SOURCES

For this experiment, we collected a dataset of 100 videos (see Supplementary materials for sample videos) of single and double point-sources moving in a plane located 4.555 meters away from the diffuser plane. We collected videos with both lens-based and lensless SRPE systems, as shown in Fig. 3. To estimate all the PHDs associated with the lensless SRPE tracking (as listed in (33)-(41)), we have used the full PSF (of size 1600×1200 pixels) due to the initial location of the point source which is assumed to be at the center of the coordinates. For tests, the initial position of the point source could be anywhere in the frame. We used patches of size 140×140 pixels of the PSFs due to new point source locations. This process was repeated for each of the 100 videos of the point source trajectory. For each video, we ran the PHD filtering operation of (27)-(30) to estimate the trajectories of the point-sources. Fig. 5 shows a few sample results for the single point-source experiments and Fig. 6 shows sample results for the case of two independently moving point-sources. The ground truth trajectories were determined by running the same PHD filtering operation on corresponding lens-based videos. To estimate the PHDs associated with lens-based system, we had to use the full 1600×1200 pixels-wide PSF with the full 1600×1200 pixels wide video frames of moving point sources. Reducing the pixel number of lens-based videoframes would have jeopardized the accuracy of the tracking process since the PSF of lens-based systems (as shown in Fig. 4b) are highly localized. Since lensless SRPE signatures (PSFs) are distributed over the entire sensor, even a small patch of an captured PSF speckle can estimate the location of a point-source. This is an advantage of lensless SRPE systems that makes the whole computational process less expensive than what would have been the case for lens-based systems.

As seen in Fig. 5 and 6, the PHD filtering algorithm can accurately estimate the trajectory of moving point-sources. Our video datasets include complex motions to test the PHD filtering-based tracking operation on challenging trajectories consisting of circular motions, close proximity of point-sources and multiple traversals through same points. The PHD filtering algorithm shows the capability to successfully handle challenging trajectories.

It is worthwhile to mention an interesting argument that could be raised to counter the computational efficiency benefits of SRPE mentioned earlier. We used a full-sized PSF of the initial location of the point source and small patches of PSFs due to moving point sources for lensless SRPE system. We used a full-sized PSF and full-sized images of moving point sources for lens-based systems. One could argue that since the PSF of lens-based system has a very small width, we could have used a small central crop of this PSF with full-sized video frames of moving point sources to perform the tracking operation. This could have brought the computational expenses of both lensless and lens-based system to the same level. However, please note that the PSF for initial point source location is fixed throughout the process while the video frames of moving point sources are changing. They are interchangeable in a correlation-based PHD filtering operation. Hypothetically speaking, if we had decided to utilize neural networks to compute the likelihood PHDs of (34) for the lensless SRPE system, the network would have been able to give an accurate estimate with only a small patch of cropped PSF of moving point sources. Due to the innate localizing property of lens-based systems, a neural network trained on a lens-based system would still require

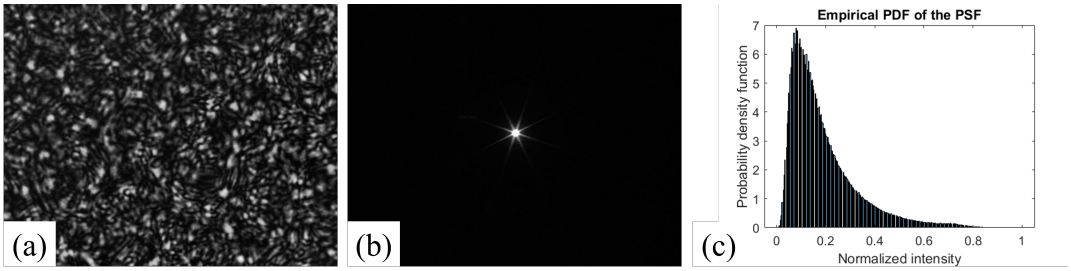

FIGURE 4. **(a) Central crop of the point spread function of the lensless SRPE system captured with the point-source located at the co-ordinate $(0, 0, 4.555)$ meters, (b) central of the same point-source captured through a lens-based system and, (c) the empirical PDF of the PSF in (a).**

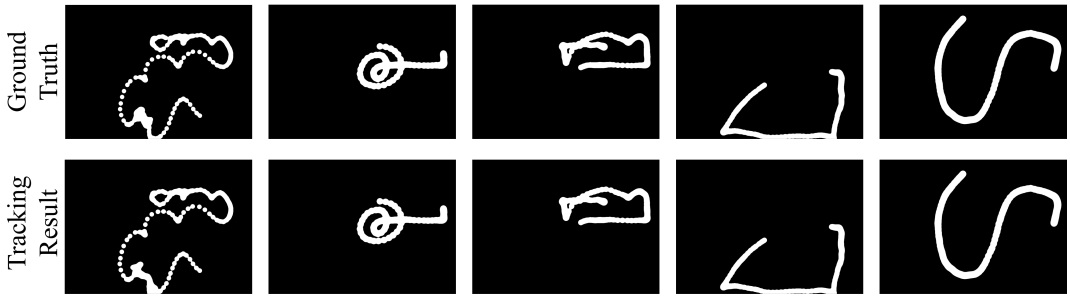

FIGURE 5. **Sample tracking results for a single point-source. Lensless SRPE uses a 0.5° diffuser.**

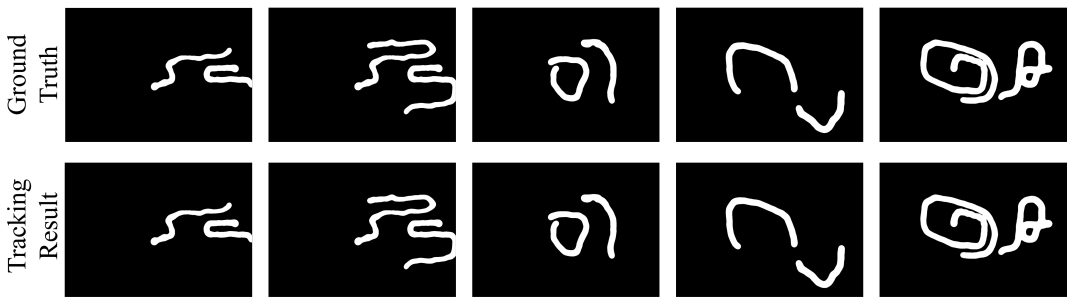

FIGURE 6. **Sample tracking results for two point-sources. Lensless SRPE uses a 0.5° diffuser.**

the full-sized PSF of moving point sources to estimate the PHDs. In other words, it is the changing item that dictates the computational cost of a method rather than the fixed item. The sensor-wide distribution of information thus lends the lensless SRPE systems an inherent favorability in considerations of computational expenses.

B. EFFECT OF NUMBER OF PIXELS ON THE TRACKING PERFORMANCE

As shown in (35), the likelihood PHD $D_{k|k}(Z_k|X_k)$ in this work is derived from the cross-correlation between patches extracted from the PSFs due to the moving point sources and the full size PSF $h(\alpha, \beta)$ due to the initial location of the point source. Since such a patch is a truncated part of a larger speckle image, this correlation would approximate the true cross-correlation between a cropped PSF and full-sized speckle PSF. This approximation becomes gradually worse as the patch size decreases. This sets a lower limit on the size of the patches of speckle PSFs.

In Fig. 7, we show how the performance of lateral tracking of point-sources suffers as we reduce the patch size of speckle PSFs. In Fig. 7a , we see that the tracking of a single point-source shows robustness as we decrease the patch size from (140×140) to (80×80) pixels. The tracking of two point-sources, as shown in Fig. 7b, shows robustness as we decrease the patch size from (140×140) to (100×100) pixels and deteriorates rapidly afterwards. Table 2 reports the numeric values of the PSNR between the ground-truth trajectory and the estimated trajectory as the patch size decreases. This is seen more clearly in Fig. 8 which shows examples of tracking results as the patch size decreases.

The reason why the minimum permissible patch-size increases with an increase in the number of point-sources is that for multiple point-sources, the individual speckle PSFs are superpositions of several differently shifted PSFs (see (32)). Hence, as the number of the point-sources increases,

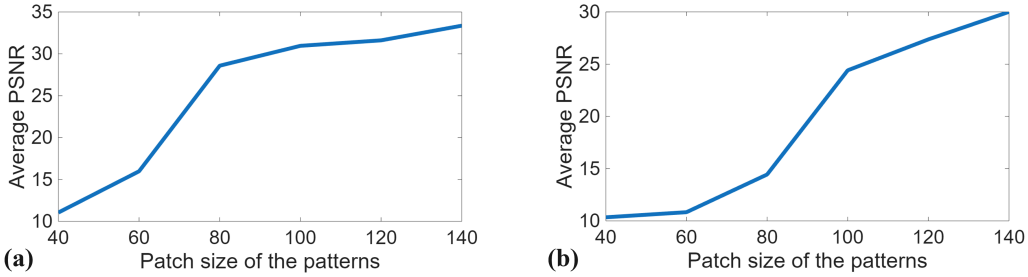

FIGURE 7. Average PSNR (in dB) for lateral tracking with a 0.5° diffuser for (a) one point source and (b) two point-sources.

Patch Size	Average PSNR (dB) for 0.5° diffuser	
	1 point-source	2 point-sources
40×40	11.0531	10.3444
60×60	15.9772	10.8329
80×80	28.5848	14.4403
100×100	30.9574	24.4141
120×120	31.6112	27.3743
140×140	33.3621	30.0062

TABLE 2. Average PSNR (in dB) for lateral tracking of point-sources with a 0.5° diffuser.

Patch Size	Execution time/frame (s)	
	1 point-source	2 point-source
1000×1000	0.3279	0.3328
800×800	0.3259	0.3275
600×600	0.2959	0.2995
400×400	0.2632	0.2648
200×200	0.1860	0.1879
180×180	0.1844	0.1855
160×160	0.1832	0.1848
140×140	0.1752	0.1784
120×120	0.1555	0.1596
100×100	0.1443	0.1478
80×80	0.1301	0.1401

TABLE 3. Per frame execution time for point source tracking with different patch sizes (scattering angle of the diffuser is 0.5°) using the proposed lensless SRPE system.

the cross-correlation requires more pixels to concentrate or exhibit a narrow peak of sufficiently high magnitude around the accurate locations of the point-sources.

As previous literature on lensless SRPE [6], [7] shows, such systems maintain good performance even when only a small subset of the original image pixels is used. Moreover, their performance is invariant to the location of the extracted pixels. Such advantages are not available for lens-based imaging systems. Using lensless SRPE systems as a tracking modality allows us to inherit these advantages which may entail significant computational benefits. Table 3 shows how the per frame execution time changes as the patch sizes of the incoming frames get smaller. The size of the inital PSF is same (1600×1200 pixels) for all cases. All the experiments have been run on an Intel(R) Xeon(R) Gold 6130 CPU (2 processors) at 2.10 GHz with 128 GB RAM. No parallel processing or graphics processing units (GPUs) were involved in the operation.

As a general guideline, the sensor image patch size corresponding to moving point-sources should increase as:

1) the correlation lobe width of the speckle PSF increases,
2) the number of targets increases,
3) the size of the image sensor pixels increases and,
4) the signal-to-noise ratio (SNR) of the image sensor pattern decreases.

In future works, we would attempt to provide a precise formula to calculate the patch size as a function of the physical parameters of an experiment.

C. EFFECT OF THE DIFFUSER SCATTERING ANGLE ON THE TRACKING PERFORMANCE

Until this point, we have used a 0.5° diffuser to perform tracking of point-sources. In this section, we investigate whether there is a benefit in going for diffusers with higher degrees of scattering. We have chosen a diffuser with a 30° scattering angle for this purpose. Fig. 9a shows the PSF of this diffuser and Fig. 9b shows its empirical PDF.

The scattering angle of a diffuser is a function of the size of its surface features [8]. The smaller the features, the higher the scattering angle. Smaller features also mean narrow spatial correlation between features on a diffuser surface. Hence, this section highlights how the tracking performance of a lensless SRPE system changes as the degree of spatial correlation changes on the surface of a diffuser.

The experimental setup for imaging with a 30° diffuser is the same as shown in Fig. 3a and 3b. However, the wider scattering angle makes the intensity pattern captured at the image sensor of much lower light levels. If we try increasing the exposure time to counteract this effect, it increases the motion blur in the sensor patterns. Hence, we had to restrict z_1 to a lower value of 2.334 meters for this set of experiments. For these experiments as well, we collected datasets with 100 videos with a single point-source and 100 videos with two independently moving point-sources, all recorded with the

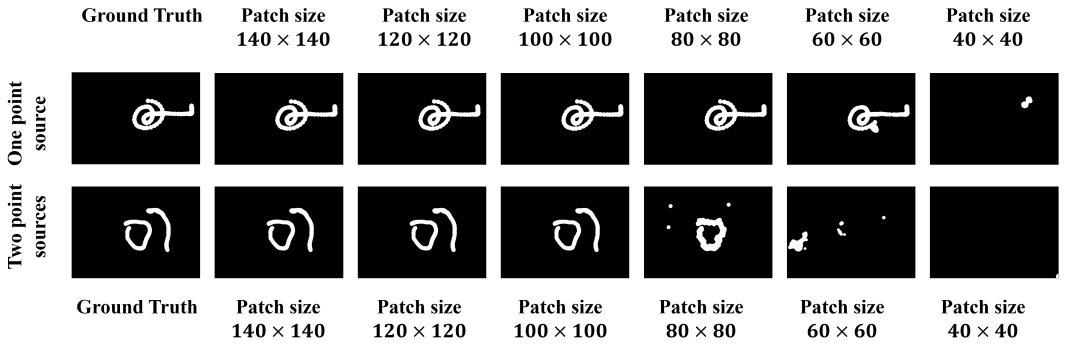

FIGURE 8. **The gradual degradation of lateral motion tracking for one and two point-sources as the patch size of the individual speckle PSFs is decreased.**

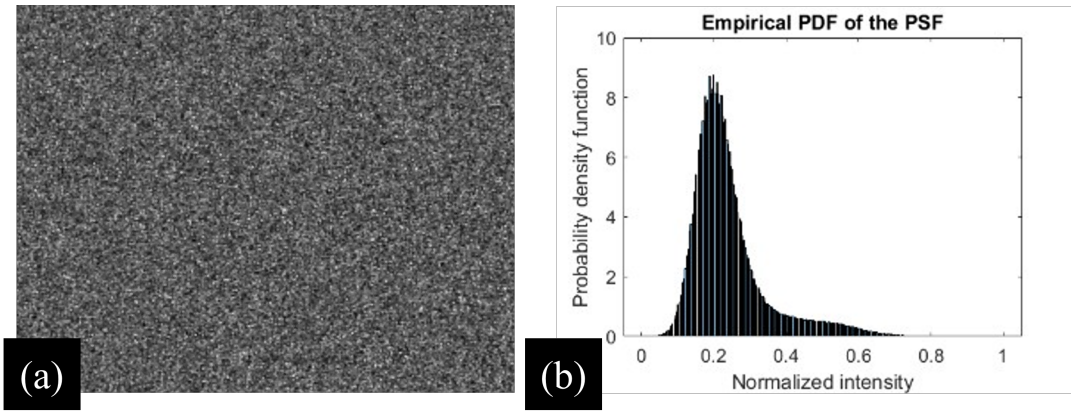

FIGURE 9. **(a) Point spread function (PSF) of an SRPE system with a 30° diffuser and, (b) the empirical probability density function (PDF) of the PSF.**

same Mako G-319c cameras at 10 frames per second. All the other parameters, such as, the diffuser to sensor distance, focal length of the lens, diffuser diameter, exposure times of the cameras and the dimension of the diffuser are kept the same as reported at the beginning of Section III.

We ran the same PHD filtering algorithm on these videos. For these set of experiments as well, we have used a patch-size of (140×140) pixels. With this patch size, the single and double point-source tracking performance was reasonably good. Fig. 10 shows examples of single point-source tracking and Fig. 11 shows examples of two point-sources tracking. We included complicated motions, close proximity of point-sources and, multiple traversals through same points to increase the tracking challenge.

Decreasing the PSF patch size of moving point sources for these experiments, had a more drastic effect on the tracking performance. Fig. 12a shows that even when the patch sizes were reduced from (140×140) pixels to (100×100) pixels, the average PSNR performance dropped by more than 5 dBs. Fig. 12 shows tracking performance for two point-sources. In this case, even with a patch size of (140×140) pixels, the performance is worse with a 30° diffuser than with a 0.5° diffuser. Moreover, as patch size decreases, the tracking performance drops significantly. Table 4 reports the numerical values of the average PSNRs plotted in Fig. 12.

Patch Size	Average PSNR (dB) for 0.5° diffuser	
	1 point-source	2 point-sources
40×40	12.5426	8.7801
60×60	14.7461	12.3891
80×80	18.5604	12.1819
100×100	27.8344	12.8457
120×120	32.7449	14.3579
140×140	33.3271	19.1651

TABLE 4. **Average PSNR (in dB) for lateral tracking of point-sources with a 30° diffuser.**

The observations of Fig. 12, combined with the fact that the point-source needs to be placed much closer to a 30° diffuser compared to a 0.5° diffuser, prompted us to the conclusion that for tracking purposes, a low scattering diffuser is a better choice for point-source tracking. This is why our experiments were performed with a 0.5° diffuser.

To understand why the tracking performance degrades with a higher scattering diffuser despite it producing finer speckles, we need to remember that for the PHD filtering algorithm to function properly, it is crucial that the PSF of the SRPE system be linear shift invariant (see (32)). This requires that the propagation of optical fields from the diffuser to the image

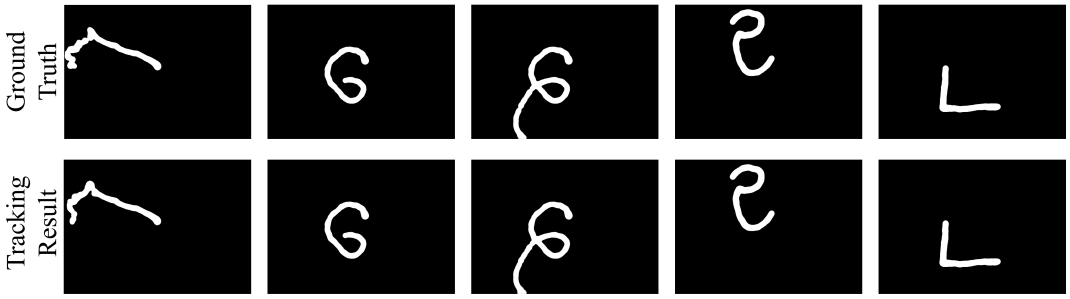

FIGURE 10. Some of the tracking results for a single point-source with a 30° diffuser.

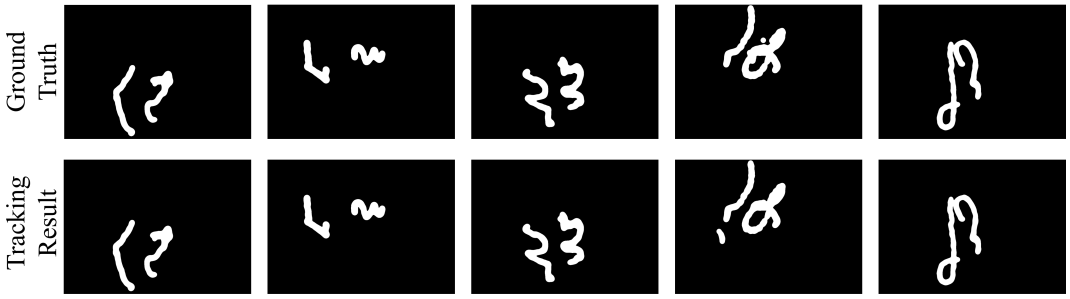

FIGURE 11. Some of the tracking results for two point-sources with a 30° diffuser.

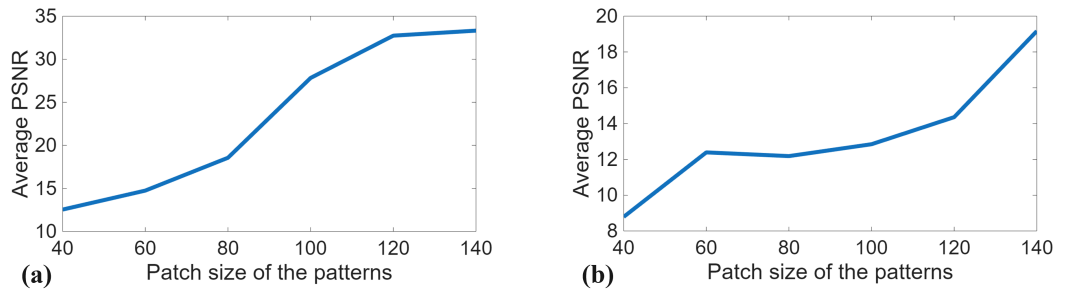

FIGURE 12. Average PSNR performance for tracking with a 30° diffuser for (a) one point source and (b) two point-sources.

sensor can be accurately described with Fresnel Diffraction. As [24] shows us, Fresnel Diffraction is only applicable when all the plane waves emanating from a source are paraxial, i.e., very close to the optical axis of the system. This is true for low scattering diffusers (such as a 0.5° scattering diffuser) but not quite for high scattering diffusers, such as the 30° diffuser used for the experiments in this section, making high scattering diffusers suboptimal for such tracking operations.

In Section II-A, Fig. 2b, the extent of deviation from linear shift-invariance has been shown in terms of a paraxial-deviation noise for 30° diffusers. When such noises become stronger with increased variance, the cross-correlation algorithm requires more pixels to subdue this noise and produce a clear peak around the location of the point-sources. This is the reason why the tracking performance in Fig. 12 of 30° scattering diffusers suffers when the patch sizes of the captured speckles are reduced.

IV. CONCLUSION

In this paper, we have introduced and demonstrated multiple point-source tracking with lensless SRPE systems and PHD filtering algorithm. To the best of our knowledge, this is the first report on lensless SRPE system for object tracking. The resulting tracking system is computationally efficient and can perform lateral tracking of point-sources moving in free space. We show tracking results with both single and double point-sources. The performance of lateral motion tracking, due to the reliance of corresponding PHDs on autocorrelation between the speckle pattern PSF of the initial location of the point source and incoming PSFs due to moving point sources, depends on the size of the patches extracted from individual PSFs. Also, with the growing number of point-

sources, one should use patches of larger sizes from PSFs. Lastly, we show the impact on tracking performance when a high scattering diffuser is used in the SRPE systems. Our results show that, due to deviation of diffuser scatter from paraxiality, high scattering diffusers seem to perform poorly than low-scattering diffusers. All the tracking experiments performed in this work assume a known number of targets which are also fixed in time. In future works, we shall deal with more challenging scenarios such as disappearing and reappearing targets. Although this work serves as proof of concept to show that it is possible to track luminous point-source like objects with lensless SRPE systems, there are still some challenges, such as gathering enough photons for insufficiently illuminated objects, which need to be addressed to make our approach suitable for real-world applications. In future works, we shall investigate various hardware upgrades and more sophisticated algorithms to address such challenges. The findings in this work can improve the computational efficiency of tracking in fields such as astronomy, satellite monitoring and aircraft surveillance.

APPENDIX A
CHOOSING APPROPRIATE VALUES OF THE COVARIANCE MATRICES

$\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_l$, as mentioned in (41) and (33) respectively, are the covariance matrices that govern the resolution and sensitivity of our tracking system. Given a predicted location of a point-source, $\mathbf{\Sigma}_p$ captures our uncertainty of the prediction by distributing probabilities around the predicted location. Similarly, given a maximum likelihood estimate of a point-source location, $\mathbf{\Sigma}_l$ helps us to codify our uncertainty of the estimation by distributing probabilities around the estimated locations. Since the PHDs are defined on the two-dimensional (x, y) grid, both $\mathbf{\Sigma}_p$ and $\mathbf{\Sigma}_l$ are $\mathbb{R}^2 \times \mathbb{R}^2$ matrices. That is,

$$\mathbf{\Sigma}_p = \begin{bmatrix} \sigma_{px} & 0 \\ 0 & \sigma_{py} \end{bmatrix}, \quad \mathbf{\Sigma}_l = \begin{bmatrix} \sigma_{lx} & 0 \\ 0 & \sigma_{ly} \end{bmatrix} \tag{43}$$

Furthermore, since x and y location of point-sources are assumed to be independent random variables, both matrices are diagonal.

$\mathbf{\Sigma}_p$ has been treated as a hyperparameter and depends mainly on whether the motion prediction model of (36)-(40) adequately captures the motion of the point-source. If σ_{px}, σ_{py} are kept too small, a misalignment between the predicted location of the point-source and the corresponding maximum likelihood estimate will cause the algorithm to produce erroneous results (note the multiplication of PHDsin (28)). On the other hand, if they are made too large, they will ascribe so much uncertainty to the prediction model as to make it practically ineffective. Since our prediction model updates the velocity (see (39)) and acceleration (see (40)) at every time-step, we found that keeping $\sigma_{px} = \sigma_{py} = 50$ pixels works reasonably well.

The choice of Σ_l dictates the resolution of our tracking system. If σ_{lx}, σ_{ly} are kept two large, any two point-sources have to located far apart in order to be detected separately. On the other hand, if they are kept very small, it ascribes so much confidence in the maximum likelihood estimate as to make the motion prediction model ineffective. This effect becomes especially problematic when we use smaller patch sizes of the observed intensity patterns. In our experiments, keeping $\sigma_{lx} = \sigma_{ly} = 10$ pixels hits a good trade-off between the two effects across all patch sizes. As a general rule of thumb, assuming the point-sources to be of equal strength, the diagonal entries of Σ_l can be made equal to the full width half maximum of the speckle autocorrelation.

APPENDIX B
OBTAINING $D_{k|k}(Z_k|X_k)$ FROM $R_{\text{ZH}}(\alpha, \beta)$

To obtain the likelihood PHD $D_{k|k}(Z_k|X_k)$ from $R_{\text{Zh}}(\alpha, \beta)$, we simply fit N number of Gaussian curves to $R_{\text{Zh}}(\alpha, \beta)$, N being the known number of point-sources. There are many algorithms in the literature to perform such a task. However, to make the computation quick, we have used the following algorithm.

We start with finding out the global maximum of $R_{\text{Zh}}(\alpha, \beta)$ and its corresponding location (α_1^*, β_1^*). Once it is found, we eliminate (replace with 0) all the values in a $(2\sigma_{lx} + 1) \times (2\sigma_{ly} + 1)$ neighborhood centered around $(\alpha_{max}, \beta_{max})$. Afterward, the next global maximum (α_2^*, β_2^*) is detected and the same process is performed. The entire process is repeated until all N global maxima are detected.

After all N maxima of R_{Zh} are detected, an empty grid of size similar to R_{Zh} is created and $D_{k|k}(Z_k|X_k)$ is computed on the grid as follows:

$$D_{k|k}(Z_k|X_k) = \frac{1}{2\pi\sigma_{lx}\sigma_{ly}} \times \sum_{i=1}^{N} \exp\left(-\frac{1}{2}\left(\left(\frac{\alpha - \alpha_i^*}{\sigma_{lx}}\right)^2 + \left(\frac{\beta - \beta_i^*}{\sigma_{ly}}\right)^2\right)\right)$$

where (α_i^*, β_i^*) is the location of the i-th maximum of $R_{\text{Zh}}(\alpha, \beta)$.

It is worthwhile to note here that the derivation of $R_{\text{Zh}}(\alpha, \beta)$ involved linear convolution with fast fourier transform (FFT) with no windowing or detrending operation. Since the peaks in $R_{\text{Zh}}(\alpha, \beta)$ for narrow scattering diffuser are smooth and wide (due to large speckle size), no subpixel interpolations were used either. Also, the noisy sidelobes were far less prominent than the actual peaks, eliminating the need for near-peak separation. In future works, we shall investigate these matters further.

ACKNOWLEDGMENT

The authors are also grateful to Kashif Usmani for the valuable discussions and comments on the manuscript.

REFERENCES

[1] H. Andersen, Z. J. Chong, Y. H. Eng, S. Pendleton, and M. H. Ang, "Geometric path tracking algorithm for autonomous driving in pedestrian environment," in *2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1669–1674, 2016.

[2] W. Dong, B. Jeffs, and J. Fisher, "A kalman-tracker-based bayesian detector for radar interference in radio astronomy," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4, pp. iv/657–iv/660 Vol. 4, 2005.
[3] K. S. Kumar, S. Prasad, P. K. Saroj, and R. Tripathi, "Multiple cameras using real time object tracking for surveillance and security system," in *2010 3rd International Conference on Emerging Trends in Engineering and Technology*, pp. 213–218, 2010.
[4] T. Ye and F. Zhou, "Autonomous space target recognition and tracking approach using star sensors based on a kalman filter," *Appl. Opt.*, vol. 54, pp. 3455–3469, Apr 2015.
[5] Z. Soleimanitaleb, M. A. Keyvanrad, and A. Jafari, "Object tracking methods:a review," in *2019 9th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 282–288, 2019.
[6] P. M. Douglass, T. O'Connor, and B. Javidi, "Automated sickle cell disease identification in human red blood cells using a lensless single random phase encoding biosensor and convolutional neural networks," *Opt. Express*, vol. 30, pp. 35965–35977, Sep 2022.
[7] T. O'Connor, C. Hawxhurst, L. M. Shor, and B. Javidi, "Red blood cell classification in lensless single random phase encoding using convolutional neural networks," *Opt. Express*, vol. 28, pp. 33504–33515, Oct 2020.
[8] S. Goswami, G. Krishnan, and B. Javidi, "Robustness of single random phase encoding lensless imaging with camera noise," *Opt. Express*, vol. 32, pp. 4916–4930, Feb 2024.
[9] S. Goswami, G. Krishnan, and B. Javidi, "Robustness of lensless single random phase encoding systems to image sensor pixel size," *Opt. Express*, vol. 33, pp. 6987–7004, Feb 2025.
[10] G. Aschenbrenner, K. Usmani, S. Goswami, and B. Javidi, "Lensless object classification in long wave infrared using random phase encoding," *Optical Engineering*, vol. 63, no. 11, p. 111809, 2024.
[11] S. Goswami, P. Wani, G. Gupta, and B. Javidi, "Assessment of lateral resolution of single random phase encoded lensless imaging systems," *Opt. Express*, vol. 31, pp. 11213–11226, Mar 2023.
[12] G. Kuo, N. Antipa, R. Ng, and L. Waller, "Diffusercam: Diffuser-based lensless cameras," in *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, p. CTu3B.2, Optica Publishing Group, 2017.
[13] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "Diffusercam: lensless single-exposure 3d imaging," *Optica*, vol. 5, pp. 1–9, Jan 2018.
[14] N. Antipa, G. Kuo, R. Ng, and L. Waller, "3d diffusercam: Single-shot compressive lensless imaging," in *Imaging and Applied Optics 2017 (3D, AIO, COSI, IS, MATH, pcAOP)*, p. CM2B.2, Optica Publishing Group, 2017.
[15] Y. Huang, G. Krishnan, S. Goswami, and B. Javidi, "Underwater optical signal detection system using diffuser-based lensless imaging," *Opt. Express*, vol. 32, pp. 1489–1500, Jan 2024.
[16] B. Javidi, A. Markman, and S. Rawat, "Automatic multicell identification using a compact lensless single and double random phase encoding system," *Appl. Opt.*, vol. 57, pp. B190–B196, Mar 2018.
[17] B. Javidi, S. Rawat, S. Komatsu, and A. Markman, "Cell identification using single beam lensless imaging with pseudo-random phase encoding," *Opt. Lett.*, vol. 41, pp. 3663–3666, Aug 2016.
[18] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*, vol. 1. Artech House Publishers, 2007.
[19] P. R. Gunjal, B. R. Gunjal, H. A. Shinde, S. M. Vanam, and S. S. Aher, "Moving object tracking using kalman filter," in *2018 International Conference On Advances in Communication and Computing Technology (ICACCT)*, pp. 544–547, 2018.
[20] L. D. Stone, R. L. Streit, and S. L. Anderson, *Bayesian Single Target Tracking*, pp. 5–44. Cham: Springer International Publishing, 2023.
[21] M. R. Morelande, C. M. Kreucher, and K. Kastella, "A bayesian approach to multiple target detection and tracking," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1589–1604, 2007.
[22] R. Mahler, "Phd filters of higher order in target number," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 4, pp. 1523–1543, 2007.
[23] R. Mahler, "Multitarget bayes filtering via first-order multitarget moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.
[24] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company publishers, 3 ed., 2005.

SAURABH GOSWAMI received his B.E. degree in Electrical Engineering from Jadavpur University, West Bengal, India. He received M.S. by Research in Computer Vision and Image Processing from the Indian Institute of Technology, Madras, India. He is currently pursuing a Ph.D. in Electrical Engineering at the University of Connecticut, USA. His research interests include lensless imaging, object tracking, deep learning, computer vision and image processing.

JIHEON LEE received the B.S. and M.S. degrees in Electrical, Electronic and Control Engineering from Hankyong National University, South Korea. He is currently working towards a Ph.D. in Electrical and Computer Engineering at the University of Connecticut, USA. His research interests include 3D integral imaging and computational image processing.

GOKUL KRISHNAN received his B.Tech. degree in Electronics and Communication Engineering from the Rajagiri College of Engineering, Kochi, M.S. degree in Electrical Engineering from the Indian Institute of Technology, Kanpur, India and PhD degree in Electrical engineering from University of Connecticut. He is currently pursuing his postdoctoral research with the Electrical and Computer Engineering Department, University of Connecticut. His research interests include 3-D imaging, gesture recognition, deep learning, and underwater object detection.

BAHRAM JAVIDI is Board of Trustees Distinguished Professor and the SNET Endowed Chair at University of Connecticut. His research interests are in a broad range of transformative imaging approaches using optics and photonics, and he has made seminal contributions to passive and active multi-dimensional imaging from nano to micro and macro scales. Prof. Javidi's research has been recognized by honors and awards, including The Optica Society Emmett Leith medal (2021), Optica C. E. K. Mees Medal (2019); IEEE Photonics Society William Streifer Scientific Achievement Award (2019); Optica Joseph Fraunhofer Award (2018); and European Physical Society Prize for Applied Aspects of Quantum Electronics and Optics (2015). He was awarded the IEEE Donald G. Fink Paper Prize (2008); John Simon Guggenheim Foundation Fellow Award (2008); Alexander von Humboldt Foundation Prize (2007); SPIE Technology Achievement Award (2008); and SPIE Dennis Gabor Award in Diffractive Wave Technologies (2005). Prof. Javidi has over 1200 publications, including 540+ peer reviewed journal articles, and 520+ conference proceedings. He is a strong believer in international scientific exchange and collaboration, and has co-authored publications with over 300 different students, scientists, and engineers from around the world. He received MS PhD degrees from The Pennsylvania State University.