

# Tracking Multiple Objects Outside the Line of Sight using Speckle Imaging

Brandon M. Smith  
University of Wisconsin-Madison  
bmsmith@cs.wisc.edu

Matthew O’Toole  
Stanford University  
motoole@stanford.edu

Mohit Gupta  
University of Wisconsin-Madison  
mohitg@cs.wisc.edu

## Abstract

This paper presents techniques for tracking non-line-of-sight (NLOS) objects using speckle imaging. We develop a novel speckle formation and motion model where both the sensor and the source view objects only indirectly via a diffuse wall. We show that this NLOS imaging scenario is analogous to direct LOS imaging with the wall acting as a virtual, bare (lens-less) sensor. This enables tracking of a single, rigidly moving NLOS object using existing speckle-based motion estimation techniques. However, when imaging multiple NLOS objects, the speckle components due to different objects are superimposed on the virtual bare sensor image, and cannot be analyzed separately for recovering the motion of individual objects. We develop a novel clustering algorithm based on the statistical and geometrical properties of speckle images, which enables identifying the motion trajectories of multiple, independently moving NLOS objects. We demonstrate, for the first time, tracking individual trajectories of multiple objects around a corner with extreme precision ( $< 10$  microns) using only off-the-shelf imaging components.

## 1. Introduction

The ability to track the motion of objects outside a sensor’s line of sight (LOS) has far-reaching implications, with potential applications in search-and-rescue, remote sensing of inaccessible environments, surveillance, and medicine. Imagine an endoscopic imaging system aiding a robot arm to perform complex surgery inside a human body, or a first responder robot searching for biometric signals such as a pulse and heartbeat to locate survivors in a hazardous environment. In these applications, it is critical to precisely detect and measure subtle motion outside the LOS. Such capability is beyond the scope of conventional vision techniques that are designed for tracking direct LOS objects.

Recently, it has been shown that it is possible to track motion of non-line-of-sight (NLOS) objects by treating visible surfaces (*e.g.*, walls) as diffuse reflectors that redirect light to and from the region hidden from the camera (see Figure 1 left). This can be achieved by using high-speed pulsed light sources and time-resolved detectors [2, 7], as well as off-the-shelf intensity sensors [14]. These tech-

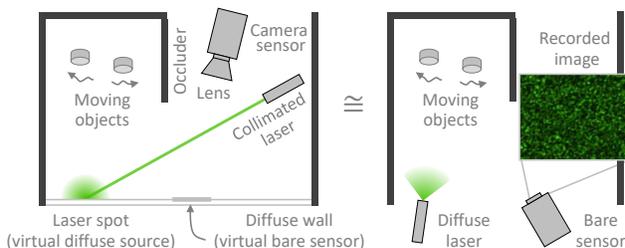


Figure 1. Tracking objects around a corner (left) is analogous to direct line of sight imaging, where the objects are illuminated directly by a diffuse source and imaged with a bare sensor (right).

niques are exciting first steps. However, the motion precision achieved by these techniques remains limited. For example, state-of-the-art tracking accuracy for time-resolved sensors is on the order of centimeters [2]. Intensity sensors can achieve similar accuracies [14], albeit by restricting the problem to single object tracking.

Our goal is to develop techniques for tracking micro-motion (*e.g.*,  $< 10$   $\mu\text{m}$ ) of multiple objects, around the corner, with only low-cost components. Given the orders of magnitude gap between this goal and the capabilities of current NLOS tracking approaches [2, 7, 14], we propose a new approach based on laser speckle imaging. Speckle is a high-frequency ‘noise-like’ intensity pattern created when coherent (laser) light reflects off an optically-rough surface. A key property of laser speckle is its extreme motion sensitivity: even small (micron-scale) surface motion results in large speckle movement, which can be readily measured even with low cost sensors to estimate surface motion. Such high motion sensitivity has been used for micro-motion analysis in direct LOS settings [16, 17, 21]. The focus of this paper is on tracking objects outside the LOS. To this end, we make two contributions.

**NLOS object tracking via speckle imaging:** First, we develop a novel NLOS speckle imaging and motion model, where both the sensor and the source view the object only indirectly via a diffuse wall, as illustrated in Figure 1 (left). Specifically, a collimated source illuminates a point on the wall, and a camera images a wall patch via a lens. We show that this NLOS imaging scenario is analogous to direct LOS speckle imaging with a bare sensor [17], where the illuminated point acts as a *virtual* diffuse source, and the imaged patch acts as a *virtual* bare sensor. Given this analogy, the

principles of direct LOS speckle tracking [11] extend to the NLOS case, thus enabling micro-motion NLOS tracking.

**Tracking multiple objects:** Although the virtual bare sensor has high motion sensitivity, it has no spatial specificity (it is completely defocused). Therefore, speckle components from multiple moving objects are superimposed in the captured images. Given the low signal-to-noise ratio inherent in NLOS imaging scenarios, it is challenging to explicitly separate and analyze the individual speckle components in order to track individual objects. Our second contribution is to exploit the orthogonality properties of speckle to develop a simple clustering algorithm that identifies the motion trajectories of individual objects. This, for the first time, enables simultaneous tracking of multiple centimeter-size objects around the corner with micron-scale precision ( $<10 \mu\text{m}$ ) with only off-the-shelf imaging components.

**Scope and limitations:** The proposed techniques can recover the absolute scale and direction of each object’s motion, but cannot recover the additive offset (starting location). It may be possible to combine the proposed methods with NLOS 3D imaging approaches [20] for recovering both micro-motion and absolute locations. Although the proposed methods can achieve extremely high motion precision, they are limited by the camera’s size and frame rate with respect to large inter-frame motions. In the future, we envision hybrid methods that combine the micro-motion estimation capabilities proposed here, with the large-scale motion tracking capabilities of other NLOS methods [2, 7].

## 2. Related Work

**Speckle-based motion estimation:** Speckle imaging based motion estimation techniques are largely limited to tracking single, rigid-body motion, such as 6 DOF motion of a camera relative to its surroundings [12, 22]. Recently, a speckle-based technique for measuring motion of multiple objects has been proposed [17]. This technique uses a bare sensor to achieve high motion sensitivity, but cannot track individual objects. Instead, it computes a motion histogram of the scene, which does not assign motions to individual objects. The motion histogram representation is ambiguous: different scene motions can result in the same histogram. An example is shown in Figure 2, where two objects moving along different pairs of trajectories result in the same motion histogram. In general, this ambiguity prevents tracking more than one object. In contrast, our goal is to track multiple, independently moving objects around the corner.

**NLOS object tracking:** NLOS object tracking with time-resolved sensors has received considerable attention recently [2, 7]. While such systems are successful, their components are expensive and the accuracy is on the order of centimeters. Klein *et al.* [14] demonstrated the ability to track objects with regular 2D intensity images and a laser pointer. This significantly reduces equipment cost, though

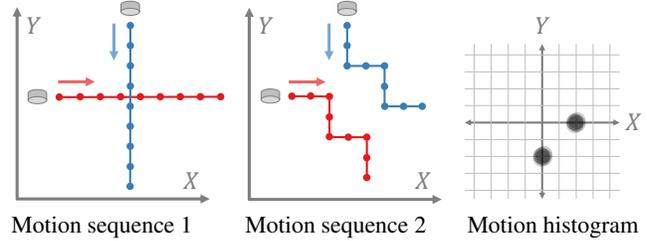


Figure 2. **Ambiguity in motion histograms.** Previous speckle based techniques for multi-object motion estimation compute a motion histogram where each object’s motion is represented as a peak [17]. Since the peaks are not assigned to individual objects, motion histogram is an ambiguous representation. Different scene motions can result in the same motion histogram. Left: Two objects moving along straight lines. Middle: Two objects moving along staircase trajectories. Right: Both pairs of trajectories have one horizontal and one vertical motion at every time step, resulting in the same motion histogram. This ambiguity prevents tracking more than one object. In contrast, we show that it is possible to track multiple objects by developing a clustering algorithm based on the statistical properties of speckle, that assigns each peak in the motion histograms to a unique object over time.

this approach is unable to resolve motion of multiple objects. Bouman *et al.* [1] showed similar capabilities by exploiting passive light transport.

Laser speckle has also been used to image around corners [13]. While this approach requires specific conditions, *e.g.*, illuminating the hidden object directly, we consider a less restrictive scenario where neither the light source or the camera have a direct LOS to the object.

## 3. Background: Speckle Motion Model

In this section, we provide a brief overview of the speckle motion model, and the standard speckle-based technique for tracking a single object in direct line-of-sight.

When coherent light is incident on an object with an optically rough surface, each point on the object acts as a virtual light source that emits a spherical wavefront. The superposition of these spherical wavefronts produces a random intensity pattern in the images captured by a camera observing the object due to constructive and destructive interference. This high-frequency, noise-like pattern is called speckle. An example speckle pattern captured by a bare (lensless) sensor is shown in Figure 1. Since the sensor is bare, the scene texture is completely defocused (blurred), and only the speckle pattern is discernible.

**Effect of object motion on image speckle:** Consider a camera imaging a moving object that is illuminated by a coherent source. Let the speckle patterns observed by the camera before and after motion be  $\mathbf{S}(x, y)$  and  $\mathbf{S}'(x, y)$ , respectively, where  $(x, y)$  are image coordinates. A well known result in optics states that small object motions result only in local displacements  $(\Delta x, \Delta y)$  in the observed speckle pattern [11], *i.e.*,  $\mathbf{S}(x, y) = \mathbf{S}'(x + \Delta x, y + \Delta y)$ .

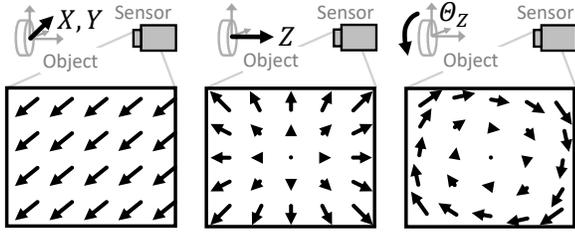


Figure 3. **Qualitative illustration of the relationship between object motion and speckle motion** (adapted from [17]). When an object moves laterally, the recorded speckle pattern shifts (left); when an object moves axially, the speckle pattern expands and contracts (center); and when an object rotates about the camera viewing axis, the speckle pattern rotates (right). This relationship is linear, as given in Eq. 1.

Furthermore, the local speckle displacement  $(\Delta x, \Delta y)$  is related to object motion via a simple, linear equation:

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \mathbf{M}_{\text{trans}} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} + \mathbf{M}_{\text{rot}} \begin{bmatrix} \Theta_X \\ \Theta_Y \\ \Theta_Z \end{bmatrix}, \quad (1)$$

where the translation vector  $[\Delta X, \Delta Y, \Delta Z]^T$  and rotation vector  $[\Theta_X, \Theta_Y, \Theta_Z]^T$  represent the motion of the object. The  $2 \times 3$  matrices  $\mathbf{M}_{\text{trans}}$  and  $\mathbf{M}_{\text{rot}}$  relate the 6 DOF object motion to local shifts in the speckle pattern, and depend on a variety of radiometric and geometric factors (pixel pitch, wavelength, and object, camera, and source positions) [11]. Note that Eq. 1 holds for a bare sensor as well.

Figure 3 illustrates the effect of various object motions on speckle. Lateral object motion (along the  $X$ - $Y$  plane) results in a global translation (shift) of the speckle image. Axial object motion (along the  $Z$ -axis) results in expansion or contraction of the speckle image. Rotating the object about the  $Z$ -axis results in a rotation of the speckle image. In the remainder of the paper, we only consider lateral object motion.<sup>1</sup> In this case, the speckle images  $\mathbf{S}$  and  $\mathbf{S}'$ , before and after the motion, are related by a shift operation:

**Definition 1 [Image Shift Vector]** *The image shift operation, given by a shift vector  $\mathbf{v} = [v_x, v_y]$ , is defined as translating an image horizontally by  $v_x$ , and vertically by  $v_y$ . If  $\mathbf{S}$  is the original image, the shifted image  $\mathbf{S}' = \mathbf{S}(x - v_x, y - v_y)$  is denoted as  $\mathbf{S}' = \mathbf{S}[\mathbf{v}]$ .*

**Estimating object motion from speckle images:** A standard approach [18] for estimating the object motion is by computing the normalized cross-correlation between  $\mathbf{S}$  and  $\mathbf{S}'$ , which, as discussed above, are shifted copies of each other. Due to statistical randomness of speckle [8], the cross-correlation between a speckle pattern and its shifted copy is given by a shifted delta function:

$$\mathbf{S}^{\text{corr}}(x, y) = (\mathbf{S} * \mathbf{S}')(x, y) = \delta(x - v_x, y - v_y), \quad (2)$$

<sup>1</sup>This is only for ease of exposition; the proposed methods can handle general object motions, as demonstrated in Figures 4 and 8.

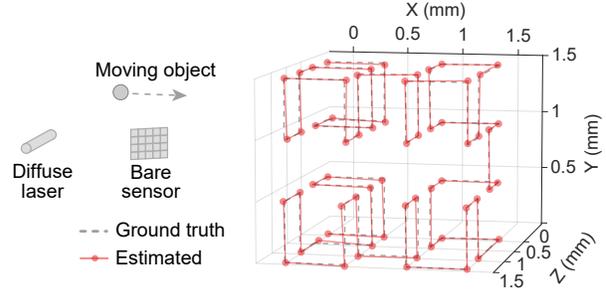


Figure 4. **Direct LOS single object tracking using bare sensor speckle imaging.** The object (a piece of white chalk) was placed 50 cm away from a bare sensor, and moved along a complex 3D trajectory (a second-order 3D Hilbert curve [10]) in 0.5-mm increments by a high-precision ( $\pm 1 \mu\text{m}$  accuracy) 3D translation stage. Using a bare sensor enabled achieving micron-precision motion tracking; mean absolute error along  $X$ -,  $Y$ -, and  $Z$ -dimensions was measured to be  $3.1 \mu\text{m}$ ,  $6.3 \mu\text{m}$ , and  $37.3 \mu\text{m}$ , respectively.

where  $\delta$  is the 2D Dirac delta function. The shift vector  $\mathbf{v} = [v_x, v_y]$  between the two speckle patterns is given by the 2D location of the cross-correlation peak. Then, the object motion can be computed by using the speckle motion equation (Eq. 1). For details, refer to the supplementary technical report. The matrices  $\mathbf{M}_{\text{trans}}$  and  $\mathbf{M}_{\text{rot}}$  can be computed as a calibration step, and are assumed to be known. This technique can be generalized to capture full motion trajectories by comparing a sequence of speckle images.

**Bare sensor and motion sensitivity:** Motion sensitivity of this approach is proportional to the amount of sensor defocus. For a bare sensor (maximum defocus), it is possible to achieve extremely high motion resolution of  $<10 \mu\text{m}$ . An example 3D trajectory of a direct LOS object recovered with bare sensor speckle imaging is shown in Figure 4.

## 4. Tracking an Object Around-the-Corner

In this section, we derive the speckle formation and motion model for NLOS imaging. We show that, under certain assumptions, the problem of NLOS object tracking can be expressed as direct LOS tracking, with a diffuse wall acting as both a virtual diffuse source and a virtual bare sensor.

### 4.1. Image Intensity Incident on the Wall

Consider the imaging geometry shown in Figure 5. A collimated coherent source  $\mathbf{L}$  illuminates a point  $\mathbf{s}$  on a diffuse wall, and indirectly illuminates an object  $\mathbf{O}$ . Light reflected by  $\mathbf{O}$  reaches the wall  $\mathbf{W}_p$ , which is imaged by a camera  $\mathbf{C}$  with a lens. Both  $\mathbf{L}$  and  $\mathbf{C}$  are around the corner (i.e., neither  $\mathbf{L}$  nor  $\mathbf{C}$  have a direct line of sight to  $\mathbf{O}$ ).

Point  $\mathbf{s}$  reflects light uniformly across the entire hemisphere of directions, and hence can be considered a virtual diffuse light source. Let  $\mathbf{U}(\mathbf{s})$  be the electric field emitted by the virtual source  $\mathbf{s}$ . Then, the total electric field received at a point  $\mathbf{w}$  on the wall, due to light emitted from  $\mathbf{s}$  and reflected from the object  $\mathbf{O}$ , is given as:

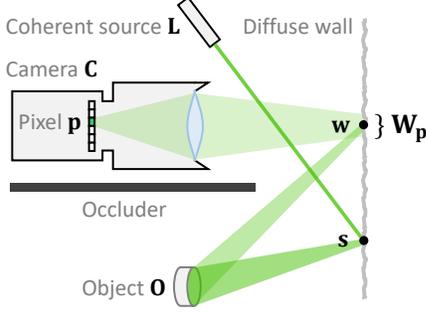


Figure 5. **Imaging geometry for NLOS object tracking.** A collimated coherent source illuminates a point  $s$  on a diffuse wall. Light gets reflected from  $s$ , and illuminates an object  $O$ . Light reflected from  $O$  again reaches the wall, which is imaged by a camera with lens. The goal is to track the motion of  $O$  by analyzing images captured by the camera.

$$\mathbf{U}(\mathbf{w}) = \int_{\mathbf{o} \in O} \mathbf{U}(\mathbf{s}) \alpha(\mathbf{o}, \mathbf{w}) e^{-j 2\pi \frac{\Gamma(\mathbf{s}, \mathbf{o}) + \Gamma(\mathbf{o}, \mathbf{w})}{\lambda}} d\mathbf{o}, \quad (3)$$

where  $\mathbf{o}$  is a point on object  $O$ , and  $\Gamma(\mathbf{s}, \mathbf{o})$  and  $\Gamma(\mathbf{o}, \mathbf{w})$  are the distances of  $\mathbf{o}$  from  $\mathbf{s}$  and  $\mathbf{w}$ , respectively.  $\alpha(\mathbf{o}, \mathbf{w})$  encodes the attenuation of light intensity due to reflection at  $\mathbf{o}$  towards  $\mathbf{w}$ , and the intensity fall-off due to propagation.

**Virtual bare sensor:** The above expression gives the electric field that would be observed by a bare sensor placed at the wall  $\mathbf{W}_p$  [17]. The speckle pattern observed by such a bare sensor is  $\mathbf{I}^{\text{prim}}(\mathbf{w}) = |\mathbf{U}(\mathbf{w})|^2$ . We call this the *primary speckle* (also referred to as *objective speckle* [5]).

## 4.2. Image Intensity Observed by the Camera

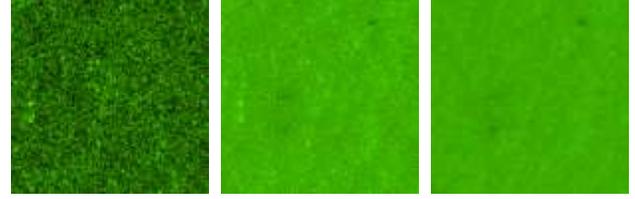
The electric field  $\mathbf{U}(\mathbf{w})$  incident on the wall is reflected again, and captured by the camera. The camera is focused on the wall via a lens, so that a camera pixel  $\mathbf{p}$  collects light from a wall patch  $\mathbf{W}_p$ . The rough micro-facet structure of the wall combined with the objective lens of the camera produce a *secondary speckle* pattern (also referred to as *subjective speckle* [5]).

Intuitively, the light incident on the wall (with spatial intensity distribution given by the primary speckle image  $\mathbf{I}^{\text{prim}}$ ) is modulated by the wall texture  $\mathbf{R}$ , and then subjected to random interference, resulting in a secondary speckle pattern. It can be shown that the total image captured by the camera,  $\mathbf{I}(\mathbf{p})$ , is given as:

$$\mathbf{I}(\mathbf{p}) = \underbrace{\mathbf{I}^{\text{prim}}(\mathbf{p})}_{\text{primary speckle}} \underbrace{\mathbf{R}(\mathbf{p})}_{\text{wall reflectance image}} \underbrace{\mathbf{I}^{\text{secon}}(\mathbf{p})}_{\text{secondary speckle}}. \quad (4)$$

See the supplementary technical report for a detailed derivation. This is an important equation: it is the image formation model for NLOS speckle imaging, where both the source and the sensor are not directly visible to the object.

**Comparison of primary and secondary speckle:** An important property of speckle patterns is the mean size of



$$\text{ratio image} = \text{captured image} \div \text{mean image}$$

Figure 6. The primary speckle image  $\mathbf{I}^{\text{prim}}$  encodes the object motion, and can be isolated by taking the ratio of the raw captured image and the mean image (computed by averaging all the raw captured images as the object moves). In the illustration above, the intensity of the ratio image is normalized.

the individual speckle grain. The primary and secondary grain sizes have several important differences. The primary speckle grain size is *independent* of the camera's imaging parameters, and is given as  $\chi^{\text{prim}} \approx \frac{1.22 \lambda \Gamma_{Ow}}{\Omega}$ , where  $\lambda$  is the wavelength of light,  $\Omega$  is the diameter of the object  $O$ , and  $\Gamma_{Ow}$  is the mean distance of  $O$  from the wall [3, 6, 9]. In our experiments,  $\lambda = 532$  nm,  $\Gamma_{Ow} = 50$  cm, and  $\Omega = 25$  mm. Therefore,  $\chi^{\text{prim}} \approx 13$   $\mu\text{m}$ . In contrast, the size of the secondary speckle grains depends on the camera's imaging parameters, and is given by  $\chi^{\text{secon}} \approx 1.22 (1 + M) \lambda (f/\#)$  where  $M$  is the magnification of the lens [19]. The secondary speckle grain size is proportional to the  $f$ -number of the lens. Provided a sufficiently large aperture (small  $f$ -number), the speckle grains become smaller than the sensor's pixel pitch and are eliminated (averaged) through spatial integration.

## 4.3. Effect of Object Motion on Image Intensities

At first glance, analyzing the effect of object motion on the observed image  $\mathbf{I}(\mathbf{p})$  may seem challenging, since the camera does not directly observe the object. Fortunately, as stated in Eq. 4,  $\mathbf{I}(\mathbf{p})$  can be factorized into three components: primary speckle image, wall reflectance image, and secondary speckle image. When the object moves, the primary speckle  $\mathbf{I}^{\text{prim}}$  shifts according to the standard direct LOS speckle motion model (Eq. 1). This is because  $\mathbf{I}^{\text{prim}}$  is the speckle pattern that would be observed by a bare sensor (placed on the wall) which can directly view the object. The key observation is that the wall reflectance image  $\mathbf{R}$  depends only on the characteristics of the wall and, as discussed above, the secondary speckle  $\mathbf{I}^{\text{secon}}$  averages out for a sufficiently large camera aperture. Therefore, the product of  $\mathbf{R}$  and  $\mathbf{I}^{\text{secon}}$  remains approximately constant as the object moves, and can be factored out as follows.

**Recovering object motion:** Let  $[\mathbf{I}_1(\mathbf{p}), \dots, \mathbf{I}_K(\mathbf{p})]$  be the sequence of  $K$  images captured by the camera as the object moves along a trajectory. Since object motion is encoded in primary speckle  $\mathbf{I}^{\text{prim}}$ , motion can be recovered by isolating  $\mathbf{I}_k^{\text{prim}}$  from  $\mathbf{I}_k$ . This can be achieved by computing the pixel-wise mean image  $\mathbf{I}^{\text{mean}}(\mathbf{p})$  of the image sequence  $[\mathbf{I}_1(\mathbf{p}), \dots, \mathbf{I}_K(\mathbf{p})]$ . Assuming sufficiently large motion,

the primary speckle components  $\mathbf{I}_k^{\text{prim}}$  get averaged out in the mean image  $\mathbf{I}^{\text{mean}}$ . The reflectance image and the secondary speckle remain approximately constant across captured images, as discussed above, so that the mean image  $\mathbf{I}^{\text{mean}}(\mathbf{p}) \approx R(\mathbf{p}) \mathbf{I}^{\text{secon}}(\mathbf{p})$ . Hence, as illustrated in Figure 6, the primary speckle component can be isolated by computing the ratio images:

$$\mathbf{I}_k^{\text{ratio}}(\mathbf{p}) = \frac{\mathbf{I}_k(\mathbf{p})}{\mathbf{I}^{\text{mean}}(\mathbf{p})} \approx \mathbf{I}_k^{\text{prim}}(\mathbf{p}).$$

To summarize, we have shown that under appropriate imaging conditions, the ratio image is the same as the primary speckle image, whose shifts can be computed to track the motion of an object around-the-corner via direct line-of-sight speckle motion equations (Eqs. 1-2).

## 5. Multi-Object Tracking Around the Corner

In this section, we develop a speckle formation and motion model for multiple, independently moving objects around the corner. As described in the previous section, NLOS speckle imaging can be expressed as direct LOS imaging with a virtual bare sensor (Figure 1). Consider a bare sensor observing  $N$  independently moving objects.<sup>2</sup> The total image captured by the sensor can be expressed as the superposition of  $N$  speckle patterns:

$$\mathbf{I}(\mathbf{p}) = \mathbf{S}_1(\mathbf{p}) + \mathbf{S}_2(\mathbf{p}) + \dots + \mathbf{S}_N(\mathbf{p}), \quad (5)$$

where  $\mathbf{S}_n(\mathbf{p})$ ,  $1 \leq n \leq N$  is the speckle component due to light reflected from the  $n^{\text{th}}$  object.<sup>3</sup> Let the total image before and after the motion be  $\mathbf{I} = \sum_{n=1}^N \mathbf{S}_n$  and  $\mathbf{I}' = \sum_{n=1}^N \mathbf{S}'_n$ , respectively, where  $\mathbf{S}_n$  and  $\mathbf{S}'_n$  are the speckle patterns for the  $n^{\text{th}}$  object before and after motion. Then, as described in Section 3,  $\mathbf{S}_n$  and  $\mathbf{S}'_n$  are related by a shift operation:  $\mathbf{S}'_n = \overrightarrow{\mathbf{S}_n[\mathbf{v}_n]}$ , where  $\mathbf{v}_n$  is the shift vector corresponding to the motion of the  $n^{\text{th}}$  object.

**Definition 2 [shift vector set]** *Given a pair of images, the set of shift vectors  $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N)$  corresponding to all moving objects, is called the shift vector set.*

**Estimating the shift vector set:** The set  $\mathbf{V}$  describes the motion of all  $N$  objects between  $\mathbf{I}$  and  $\mathbf{I}'$ , and can be computed via the cross-correlation between  $\mathbf{I}$  and  $\mathbf{I}'$  [17]:

$$\mathbf{I}^{\text{corr}} = \mathbf{I} * \mathbf{I}' = \left( \sum_{n=1}^N \mathbf{S}_n \right) * \left( \sum_{n=1}^N \mathbf{S}'_n \right) = \sum_{n=1}^N \mathbf{S}_n^{\text{corr}},$$

<sup>2</sup>A priori knowledge of the number of objects is not required. The cross-correlation peak heights (Step 1 in Figure 7) can be used to filter out weak or phantom objects by thresholding; remaining peaks indicate the number of objects.

<sup>3</sup>Strictly speaking, the total image also contains cross speckle terms due to interference between light reflected from different objects. However, if the light source has low spatial coherence (but high temporal coherence), the cross terms are negligible, and can be ignored [17].

where  $\mathbf{S}_n^{\text{corr}} = \mathbf{S}_n * \mathbf{S}'_n$  is the correlation image due to the  $n^{\text{th}}$  object. Since each correlation image  $\mathbf{S}_n^{\text{corr}}$  is a shifted delta function (Eq. 2), the total correlation image  $\mathbf{I}^{\text{corr}}$  is a sum of  $N$  shifted delta functions. The locations of the peaks (shift amounts) can be used to compute the shift vectors in the set  $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N)$ , as discussed after Eq. 2.

**Can the shift vector set be used to track individual objects?** While a shift vector set can be used to estimate aggregate motion statistics (e.g., a *motion histogram* [17]) of the scene, it cannot be used to track individual objects. This is because *the shift vectors in a set are not necessarily in order, i.e., the vector  $\mathbf{v}_n$  ( $n^{\text{th}}$  peak in the motion histogram) does not necessarily correspond to the motion of the  $n^{\text{th}}$  object.* Figure 2 shows an example of two pairs of motion trajectories for two moving objects. Each pair of trajectories results in the same shift vector set over time. Due to this inherent ambiguity, in general, individual objects cannot be tracked using the shift vector sets.

### 5.1. Object Tracking by Labeling of Shift Vectors

*Our main observation is that the problem of multi-object tracking can be expressed as a labeling problem.* Specifically, given a sequence of images  $[\mathbf{I}_1, \dots, \mathbf{I}_K]$  and their shift vector sets  $[\mathbf{V}_1, \dots, \mathbf{V}_K]$  with respect to a reference image, the multi-object motion tracking problem can be solved by assigning an object label  $l$  ( $1 \leq l \leq N$ ) to every vector in all the shift vector sets. Once each vector is assigned a label, we collect all the vectors across all sets  $\mathbf{V}_k$  ( $1 \leq k \leq K$ ) that have the same label, i.e., that correspond to the same object. Vectors with the same label can then be used to find the motion of individual objects in every image with respect to the reference image.<sup>4</sup> In the following, we analyze the geometry of the space of the shifted speckle images. Based on this analysis, we develop clustering based techniques for solving the labeling problem.

### 5.2. Geometry of Multi-Object Speckle Images

Consider a bare sensor observing  $N$  moving objects. Let  $\mathbf{I}_1$  and  $\mathbf{I}_2$  be images captured by the sensor at two time instants. Each image can be expressed as a sum of  $N$  different speckle components (Eq. 5). Let the shift vector sets for  $\mathbf{I}_1$  and  $\mathbf{I}_2$  (with respect to a reference image) be  $\mathbf{V}_1 = (\mathbf{v}_{11}, \dots, \mathbf{v}_{1N})$  and  $\mathbf{V}_2 = (\mathbf{v}_{21}, \dots, \mathbf{v}_{2N})$ , respectively. Consider two vectors  $\mathbf{v}_{1\alpha}$  and  $\mathbf{v}_{2\beta}$  ( $1 \leq \alpha, \beta \leq N$ ), one each from the sets  $\mathbf{V}_1$  and  $\mathbf{V}_2$ . There are two cases:

**Case 1 [iso-object shift vectors]:**  $\mathbf{v}_{1\alpha}$  and  $\mathbf{v}_{2\beta}$  correspond to the motion of the *same* object. Suppose we shift  $\mathbf{I}_1$  and  $\mathbf{I}_2$  by the negative vectors  $-\mathbf{v}_{1\alpha}$  and  $-\mathbf{v}_{2\beta}$ , respectively. The shifted images  $\overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\alpha}]}$  and  $\overrightarrow{\mathbf{I}_2[-\mathbf{v}_{2\beta}]}$  can be represented as points in a high-dimensional space. Let  $\Gamma_{\text{iso}} = \left\| \overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\alpha}]} - \overrightarrow{\mathbf{I}_2[-\mathbf{v}_{2\beta}]} \right\|_2$  be the Euclidean distance

<sup>4</sup>This method does not recover the absolute locations of the objects; it recovers their trajectories relative to locations in the reference image.

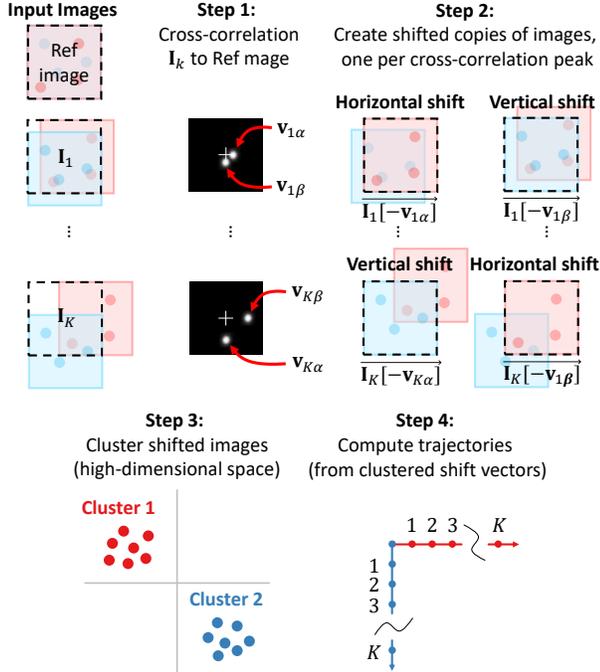


Figure 7. Illustration of the clustering algorithm for labeling shift vectors. **Step 1:** The 2D cross-correlation is computed between each image  $\mathbf{I}_k$  and a reference image; the peaks indicate the shift vector set  $\mathbf{V}_k$  for each image. **Step 2:** Create shifted copies of the input images, one per shift vector. **Step 3:** Cluster the shifted images; images with the same aligned speckle component belong to the same cluster. **Step 4:** Use the clusters to label the shift vectors, and compute individual object trajectories.

between the two points. It can be shown that:

$$\Gamma_{\text{iso}}^2 = 2 \sum_{n=1, n \neq \gamma}^N \mathcal{L}_n, \quad (6)$$

where  $\mathcal{L}_n$  is the  $L_2$  norm (root-mean-square brightness) of speckle pattern  $\mathbf{S}_n$ <sup>5</sup>, and  $\gamma$  ( $1 \leq \gamma \leq N$ ) is the label of the object whose motion is given by vectors  $\mathbf{v}_{1\alpha}$  and  $\mathbf{v}_{2\beta}$ .

**Case 2 [non iso-object scale-and-shift vectors]:**  $\mathbf{v}_{1\alpha}$  and  $\mathbf{v}_{1\beta}$  correspond to the motion of *different* objects. In this case, it can be shown that the squared distance is given as:

$$\Gamma_{\text{dif}}^2 = 2 \sum_{n=1}^N \mathcal{L}_n. \quad (7)$$

See the supplementary report for a detailed derivation.

**Main difference between  $\Gamma_{\text{iso}}$  and  $\Gamma_{\text{dif}}$ :** Intuitively, in the first case, since the shift vectors  $\mathbf{v}_{1\alpha}$  and  $\mathbf{v}_{2\beta}$  denote the motion of the *same* object, the speckle component due to that object gets aligned in the shifted images. When computing the distance between the two shifted images, this

<sup>5</sup>  $\mathcal{L}_n$  is a function of the  $n^{\text{th}}$  object's characteristics (size, reflectivity, distance from the wall, orientation, *etc.*) and can be assumed to be the same for a speckle pattern  $\mathbf{S}$  and its shifted version for small shifts

speckle components gets subtracted out. Therefore, the distance term  $\Gamma_{\text{iso}}^2$  is the sum of squared  $L_2$  norms of *all but one* speckle components. In the second case, the shift vectors denote the motion of *different* objects. Thus, assuming each object has a different motion in an image, no speckle component gets aligned and subtracted out. Therefore, the distance term  $\Gamma_{\text{dif}}^2$  is the sum of norms of *all* the speckle components. *The main result is that  $\Gamma_{\text{iso}} < \Gamma_{\text{dif}}$ .* We use this result to design a clustering based labeling algorithm.

### 5.3. Clustering-based Labeling of Shift Vectors

Let  $[\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_K]$  be the sequence of  $K$  images captured by the camera as  $N$  objects move along their trajectories. Let  $\mathbf{V}_k = (\mathbf{v}_{k1}, \mathbf{v}_{k2}, \dots, \mathbf{v}_{kN})$  be the shift vector set of the  $k^{\text{th}}$  image ( $1 \leq k \leq K$ ), computed with respect to a fixed reference image. In order to label the shift vectors, we first compute  $N$  shifted versions of each input image  $\mathbf{I}_k$ , one for each vector in the set  $\mathbf{V}_k$ . Let the set of all  $KN$  shifted images be  $[(\mathbf{I}_{11}, \mathbf{I}_{12}, \dots, \mathbf{I}_{1N}), \dots, (\mathbf{I}_{K1}, \mathbf{I}_{K2}, \dots, \mathbf{I}_{KN})]$ , such that  $\mathbf{I}_{kn} = \mathbf{I}_k[-\mathbf{v}_{kn}]$ .

*The key idea* is that the set of  $KN$  shifted images form  $N$  clusters in the image space, one cluster for each object. Intuitively, the shifted images that are computed with shift vectors corresponding to the same object have one (out of  $N$ ) of the speckle components aligned. As stated above, the distance  $\Gamma_{\text{iso}}$  between such images is smaller than the distance  $\Gamma_{\text{dif}}$  between images for which no speckle component is aligned. Thus, shifted images corresponding to the motion of the same object form a cluster. The clusters can be estimated via standard algorithms, such as  $k$ -means [15] or mean-shift [4]. Let  $(\mathbf{v}_{1\alpha}, \mathbf{v}_{2\beta}, \dots, \mathbf{v}_{K\gamma})$  be the shift vectors (one from every shift vector set) associated with images in a given cluster. These shift vectors are assigned the same label, since they correspond to the motion of the same object. This is illustrated in Figure 7.

### 5.4. Analysis of the Clustering Algorithm

One metric to predict the performance of a clustering algorithm is the clustering coefficient  $\chi$ , defined as the ratio of the average *intra*-cluster distance  $\Gamma_{\text{iso}}$  (distance between points within a cluster) to the average *inter*-cluster distance  $\Gamma_{\text{dif}}$  (distance between points across clusters). A large clustering coefficient denotes tight and well-separated clusters, which may result in robust clustering even in the presence of noise. For the clustering problem described above,  $\Gamma_{\text{dif}}$  is the distances between shifted images corresponding to different objects, and  $\Gamma_{\text{iso}}$  is the distance between shifted images corresponding to the same object. The clustering coefficient for an object label  $\gamma$  is then:

$$\chi_\gamma = \frac{\Gamma_{\text{dif}}}{\Gamma_{\text{iso}}} = \underbrace{\sqrt{1 + \frac{\mathcal{L}_\gamma}{\sum_{n=1, n \neq \gamma}^N \mathcal{L}_n}}}_{\text{Clustering Coefficient}}. \quad (8)$$

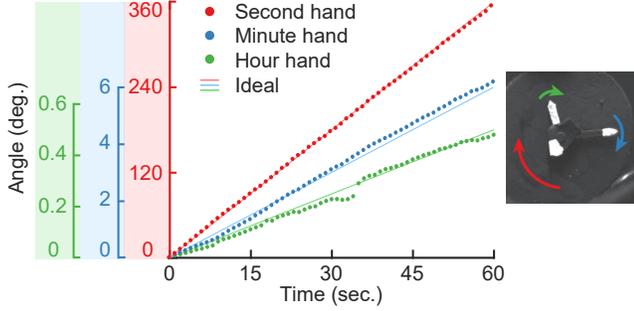


Figure 8. **Simultaneous direct LOS tracking of wristwatch hands with a bare sensor.** For experimental purposes, the face was painted black, and the hands were painted white (right). Speckle images were recorded at one-second intervals. The dynamic range of motion between frames is large: the second hand turns 6 degrees, and the hour hand turns 1/120-th of a degree. The mean absolute differences between ideal and measured for the second, minute, and hour hands was 1.497, 0.141, and 0.013 degrees, respectively. Interestingly, the dip in the hour hand angle measurement is due to brief occlusion from the second hand.

The clustering coefficient asymptotes to one as the number of objects increases or as the brightness of the speckle pattern corresponding to object  $\gamma$  decreases.

**Effect of image noise on clustering coefficient:** The analysis so far does not account for image noise. In the presence of noise, the total image can be expressed as:

$$\mathbf{I} = \sum_{n=1}^N \mathbf{S}_n + \mathbf{\Upsilon},$$

where  $\mathbf{\Upsilon}$  is the noise component. It can be shown that the distance between *every pair* of shifted images is increased by  $2\mathcal{L}^{\text{noise}}$ , where  $\mathcal{L}^{\text{noise}}$  is the  $L_2$  norm of the noise image  $\mathbf{\Upsilon}$  (proportional to the standard deviation of the noise). Then, the distances between shifted images are given as  $\Gamma_{\text{iso}}^2 = 2\mathcal{L}^{\text{noise}} + 2\sum_{n=1, n \neq \gamma}^N \mathcal{L}_n$  (shifts corresponding to the same object), and  $\Gamma_{\text{dif}}^2 = 2\mathcal{L}^{\text{noise}} + 2\sum_{n=1}^N \mathcal{L}_n$  (shifts corresponding to different objects). The noise limited clustering coefficient is then:

$$\chi_{\gamma} = \sqrt{1 + \frac{\mathcal{L}_{\gamma}}{\mathcal{L}^{\text{noise}} + \sum_{n=1, n \neq \gamma}^N \mathcal{L}_n}}. \quad (9)$$

Noise Limited Clustering Coefficient

See the supplementary report for a detailed derivation. As  $\mathcal{L}^{\text{noise}}$  increases, the clustering coefficient, and the ability to cluster robustly, decreases. This loss in performance can prevent reliable tracking of objects around the corner due to extremely high noise (low SNR) in the speckle images.

**Improved clustering with principle component analysis (PCA):** We propose a simple method to improve the clustering performance by applying PCA on the set of shifted speckle images, prior to clustering. The set of  $KN$  shifted images represent high-dimensional points in the space of speckle image. These points form  $N$  clusters. Therefore,

the first  $N - 1$  principal components of the set of images correspond to direction vectors between different cluster centers (directions of maximum variance); the remaining components correspond to noise and intra-cluster variation. Thus, retaining only the first  $N - 1$  principal components retains only the variance between cluster centers, but collapses the direction vectors within each cluster. This results in tighter and well-separated clusters, which is critical given the low SNR of NLOS imaging.

## 6. Experimental Setup and Results

Our experimental setup closely resembles the left side of Figure 1. We note that the proposed approach does not require a specific geometry, as long as the objects are ‘visible’ to the virtual sensor and virtual source. We positioned the camera sensor (Point Grey Grasshopper3 with Sony IMX174 sensor) 50 cm from the wall. The sensor was placed perpendicular to the wall to simplify the analysis, but this is not a requirement. A macro lens (Canon EF 100 mm f/2.8L) was used to image a close-up patch on the wall. The source (Civillaser 532 nm, 350 mW DPSS laser) was placed 10 cm to the right of the lens, and aimed at a spot on the wall in front of the objects. We separated the laser spot  $s$  from the imaged wall patch  $\mathbf{W}_p$  to reduce the amount of stray light illuminating  $\mathbf{W}_p$  directly, and increase the contrast of the primary speckle pattern. The objects (25 mm diameter chalk disks) were placed 20 cm from the white wall on the opposite side of the occluding wall. To minimize high-order light reflections, we covered the background walls with blackout material (Acktar Metal Velvet). For NLOS experiments, we used 3.2 second exposures, and for direct LOS experiments we used 10 ms exposures.

**Tracking single objects around the corner:** We used high-precision ( $<1 \mu\text{m}$  error) 2D motorized translation stages (Micronix VT-21L) to move the target objects along small, intricate trajectories. Several examples are shown in Figure 9 (top row). The faded line indicates ground truth, and the dots indicate the estimated location of the object at each observation in the sequence. In each example, the initial location of the object is indicated by a gray axis.

**Multi-object micro-motion tracking:** We first show tracking results for multiple objects in direct line of sight to evaluate the capabilities of our multi-object tracking approach without the added complexities of NLOS tracking. We used a bare sensor to achieve high motion sensitivity, and yet track the individual trajectories of multiple objects using the techniques described in Section 5. An interesting test case is tracking the hands of a wristwatch (Figure 8). Despite large dynamic range in motion (6 degrees of rotation per second for the second hand, and 1/120-th of a degree per second for the hour hand), our approach can recover the individual trajectory of each wristwatch hand simultaneously with high precision. In fact, at high frame rates, our approach is capable of measuring the microscopic jitter of each hand due to mechanical imperfections in the watch.

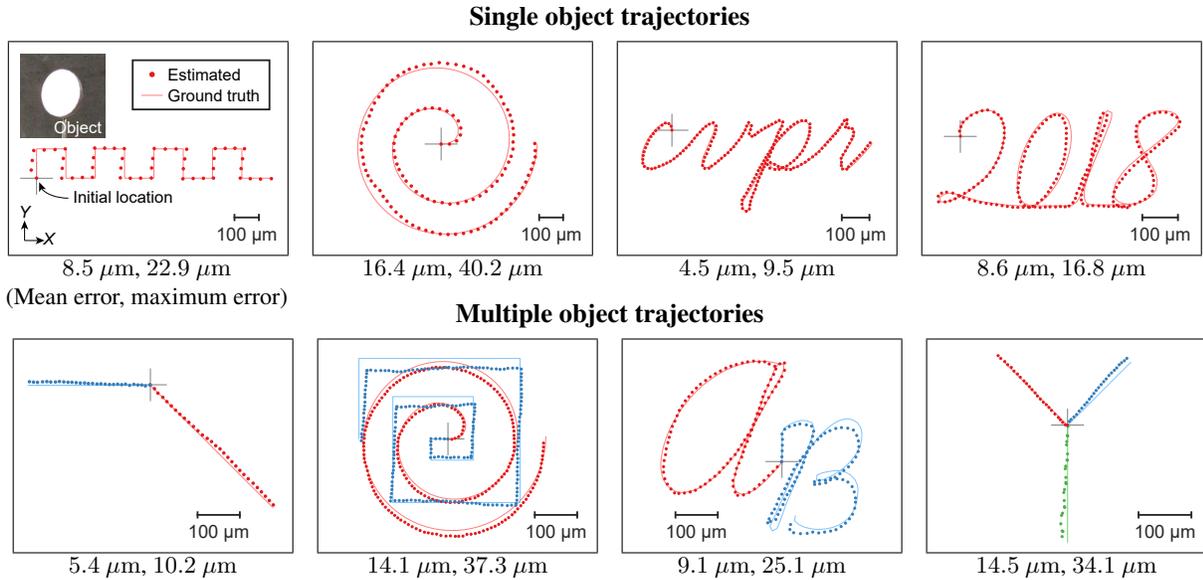


Figure 9. **Micron precision tracking results around the corner.** Single object trajectories are shown in the top row, and multiple object trajectories are shown in the bottom row. The experimental setup used to generate these results closely matches Figure 1 (left). Despite the macroscopic (cm) scale of the setup, we show that it is possible to recover complex, micron-scale trajectories of multiple objects moving simultaneously outside the line of sight. **Please see our supplementary video for tracking results.**

Finally, we show results for tracking *multiple* objects around the corner. Several examples are shown in Figure 9 (bottom row). For these experiments, we placed identical 25 mm diameter chalk disks several centimeters apart, and moved them along small trajectories. All object trajectories are shown to be starting from the same location (origin) *only for visualization purposes*, as our method cannot estimate absolute starting location. In our experiments, Step 2 of our algorithm considered short trajectories. Longer trajectories are recovered by applying the algorithm in sliding temporal windows of 31 frames. For the three-object case, the third trajectory was generated using a less precise linear actuator (OpenBuilds C-Beam); this is reflected in the noisier trajectory shown in green. Despite the macroscopic, NLOS path between the camera and the objects, our experimental system is capable, for the first time, of accurately recovering the micron-scale trajectories of two or more objects.

## 7. Discussion

Our experimental system has several limitations, which afford opportunities for future work. If the virtual source and virtual sensor are co-located, the object motion is invariant to object location [11], with axial motion being quasi-invariant. In this case, the absolute scale and direction of each object’s motion is recoverable without calibration; only the additive offset (starting location) is unknown. Due to practical considerations, the source and sensor are not co-located in our setup. This results in a weak dependence of the object motion on the location, thus requiring a calibration step to recover the absolute motion (up to an additive offset). However, motion recovery is not highly sensitive to a precise calibration. For example, with our

setup, if the object moves horizontally by 100 mm, the recovered estimates have an error of approx. 20%. A next step is to recover the additive offset, potentially by capturing measurements with multiple virtual source and sensor configurations. Intuitively, each sensor-source configuration provides constraints on the absolute object locations via  $M_{trans}$ . Two or more configurations may be sufficient to recover absolute locations. A full solution is future work.

The proposed method can recover trajectories of multiple objects provided their motions are not exactly the same *over the entire trajectory*. The overall length of the trajectory depends on the range over which the primary speckle pattern remains correlated (the memory effect FOV).

To improve SNR, our NLOS experiments were performed with long exposures, and the surrounding walls were black. Despite these steps, our NLOS experimental system could not reliably track more than three moving objects. A more sensitive sensor and better processing would likely improve performance. Finally, the range of allowable frame-to-frame motion was limited to 5-100  $\mu\text{m}$ . On the low end, cross-correlation peaks become obscured by the residual background peak due to imperfections in isolating the primary speckle image. On the high end, the speckle pattern becomes decorrelated. Despite these limitations, we have shown for the first time that it is possible to track the micron-scale trajectories of multiple simultaneously moving objects outside the line of sight at macroscopic distances using only off-the-shelf imaging components.

**Acknowledgements** This research was supported in part by the DARPA REVEAL program and by ONR grant # N000141612995.

## References

- [1] K. L. Bouman, V. Ye, A. B. Yedidia, F. Durand, G. W. Wornell, A. Torralba, and W. T. Freeman. Turning corners into cameras: Principles and methods. *Int. Conference on Computer Vision*, pages 1–9, 2017. [2](#)
- [2] S. Chan, R. E. Warburton, G. Gariepy, J. Leach, and D. Faccio. Non-line-of-sight tracking of people at long range. *Opt. Express*, 25(9):10109–10117, May 2017. [1](#), [2](#)
- [3] J. C. Dainty. *Laser Speckle and Related Phenomena*. Springer, 1975. [4](#)
- [4] K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1), Jan 1975. [6](#)
- [5] D. Gabor. Laser speckle and its elimination. *IBM Journal of Research and Development*, 14(5):509–514, Sep 1970. [4](#)
- [6] J. García, Z. Zalevsky, P. García-Martínez, C. Ferreira, M. Teicher, and Y. Beiderman. Three-dimensional mapping and range measurement by means of projected speckle patterns. *Appl. Opt.*, 47(16):3032–3040, Jun 2008. [4](#)
- [7] G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio. Detection and tracking of moving objects hidden from view. *Nature Photonics*, 10, 2016. [1](#), [2](#)
- [8] J. W. Goodman. *Statistical Optics*. Wiley-Interscience, 2000. [3](#)
- [9] J. W. Goodman. *Speckle Phenomena in Optics: Theory and Applications*. Roberts and Company Publishers, 2007. [4](#)
- [10] D. Hilbert. Ueber die stetige abbildung einer linie auf ein flächenstück. *Mathematische Annalen*, 38:459–460, 1891. [3](#)
- [11] P. Jacquot and P. K. Rastogi. Speckle motions induced by rigid-body movements in freespace geometry: an explicit investigation and extension to new cases. *Applied Optics*, 1979. [2](#), [3](#), [8](#)
- [12] K. Jo, M. Gupta, and S. Nayar. SpeDo: 6 DOF ego-motion sensor using speckle imaging. In *Proc. ICCV*, 2015. [2](#)
- [13] O. Katz, E. Small, and Y. Silberberg. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nature Photonics*, 6, 2012. [2](#)
- [14] J. Klein, C. Peters, J. Martin, M. Laurenzis, and M. B. Hullin. Tracking objects outside the line of sight using 2d intensity images. *Scientific Reports*, 6, 2016. [1](#), [2](#)
- [15] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 1982. [6](#)
- [16] Y. Shih, A. Davis, S. W. Hasinoff, F. Durand, and W. T. Freeman. Laser speckle photography for surface tampering detection. In *Proc. IEEE CVPR*, 2012. [1](#)
- [17] B. M. Smith, P. Desai, V. Agarwal, and M. Gupta. CoLux: Multi-object 3d micro-motion analysis using speckle imaging. *ACM Trans. Graph.*, 36(4):34:1–34:12, July 2017. [1](#), [2](#), [3](#), [4](#), [5](#)
- [18] P. Synnergren. Measurement of three-dimensional displacement fields and shape using electronic speckle photography. *Optical Engineering*, 1997. [3](#)
- [19] O. Thompson, M. Andrews, and E. Hirst. Correction for spatial averaging in laser speckle contrast analysis. *Biomed. Opt. Express*, 4(2), 2011. [4](#)
- [20] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:1–8, 2012. [2](#)
- [21] Z. Zalevsky, Y. Beiderman, I. Margalit, S. Gingold, M. Teicher, V. Mico, and J. Garcia. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Optics Express*, 2009. [1](#)
- [22] J. Zizka, A. Olwal, and R. Raskar. SpeckleSense: Fast, precise, low-cost and compact motion sensing using laser speckle. In *Proc. ACM UIST*, 2011. [2](#)