

Tracking Multiple Objects Outside the Line of Sight using Speckle Imaging

Supplementary Technical Report

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

1. Details of the Speckle Motion Model (Eq. 1 in the Paper)

As discussed in Section 4 of the main paper, tracking objects around the corner (non-line-of-sight imaging) is analogous to direct line of sight imaging, where the target objects are illuminated directly by a diffuse light source, and imaged with a bare sensor. Here, we provide details of the relationship between object motion and motion of the speckle pattern recorded by a bare sensor. The object motion is defined in 3D space (3 degrees of freedom for translation, and 3 degrees of freedom for rotation) and speckle motion is defined in the 2D image space. Consider the right-hand side of Figure 5 in the main paper. 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 $[\Delta X, \Delta Y, \Delta Z]^T$ result only in local displacements $(\Delta x, \Delta y)$ in the observed speckle pattern [6], *i.e.*, $\mathbf{S}(x, y) = \mathbf{S}'(x + \Delta x, y + \Delta y)$. 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 6 DOF motion of the object.

Following the notation in [6], let the coordinates of the virtual light source be $\mathbf{s} = \rho_e [l_e, m_e, n_e]$, where ρ_e is the distance between the object position \mathbf{o} and \mathbf{s} , and $[l_e, m_e, n_e]$ is the unit vector in the direction from \mathbf{o} to \mathbf{s} . Let the 3D coordinates of a virtual pixel \mathbf{w} on the wall be $\mathbf{w} = \rho_o [l_o, m_o, n_o]$, where ρ_o is the distance between \mathbf{o} and \mathbf{w} , and $[l_o, m_o, n_o]$ is the unit vector in the direction from \mathbf{o} to \mathbf{w} . Then, $\mathbf{M}_{\text{trans}}$ is a 2×3 matrix given as [6]:

$$\mathbf{M}_{\text{trans}} = \frac{\rho_o}{p \rho_e \sigma_o} \begin{bmatrix} \frac{\sigma_o^2(1-l_e^2)+l_o m_o l_e m_e}{n_o} + n_o \frac{\rho_e}{\rho_o} & \frac{l_o m_o(1-m_e^2)-\sigma_o^2 l_e m_e}{n_o} & -\frac{n_e}{n_o}(\sigma_o^2 l_e + l_o m_o m_e) - l_o \frac{\rho_e}{\rho_o} \\ -l_e m_e - l_o m_o \frac{\rho_e}{\rho_o} & (1-m_e^2) + \sigma_o^2 \frac{\rho_e}{\rho_o} & -n_e m_e - m_o n_o \frac{\rho_e}{\rho_o} \end{bmatrix}, \quad (2)$$

and \mathbf{M}_{rot} is a 2×3 matrix given as:

$$\mathbf{M}_{\text{rot}} = \frac{\rho_o}{p \sigma_o} \begin{bmatrix} -m_o l_o \left(1 + \frac{n_e}{n_o}\right) & \sigma_o^2 \left(1 + \frac{n_e}{n_o}\right) & \frac{l_o}{n_o}(m_o l_e - l_o m_e) - n_o(m_e + m_o) \\ -(n_o + n_e) & 0 & l_o + l_e \end{bmatrix}, \quad (3)$$

where p is the side length of the virtual pixel \mathbf{w} (assumed, for simplicity, to be square), $\sigma_o = \sqrt{l_o^2 + n_o^2}$, $\sigma_e = \sqrt{l_e^2 + n_e^2}$, and $l_o, m_o, n_o, \rho_o, l_e, m_e, n_e, \rho_e$ are as defined above. The $\frac{1}{p}$ factor converts the world distance units into virtual bare sensor pixel units.

2. Derivation of the Speckle NLOS Image Model (Eq. 4 in the paper)

The electric field $\mathbf{U}(\mathbf{w})$ incident on a wall (as given by Eq. 3 in the main paper) is reflected by the wall, before being detected 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 total electric field received at \mathbf{p} is given as:

$$\mathbf{U}_C(\mathbf{p}) = \int_{\mathbf{w} \in \mathbf{W}_p} \mathbf{U}(\mathbf{w}) \alpha(\mathbf{w}, \mathbf{p}) e^{-j 2\pi \frac{\Gamma(\mathbf{w}, \mathbf{p})}{\lambda}} d\mathbf{w}, \quad (4)$$

where $\Gamma(\mathbf{w}, \mathbf{p})$ is the optical path-length from \mathbf{w} to \mathbf{p} that accounts for the refractive index of the lens material. $\alpha(\mathbf{w}, \mathbf{p})$ encapsulates the light attenuation due to reflection at \mathbf{w} towards \mathbf{p} (given by the BRDF at \mathbf{w}), and the intensity fall-off due to propagation.

In order to simplify Eq. 4, we assume that the size of the wall-patch \mathbf{W}_p imaged at pixel \mathbf{p} is significantly smaller than the size of the primary speckle grain. This assumption is valid for a camera lens with large magnification (e.g., a macro lens), so that each pixel collects light only from a small wall-patch. For example, in our experiments, the sensor pixel is a square 5 microns wide, the lens magnification is approximately 0.6, so that \mathbf{W}_p is a square approximately 8 microns wide. Hence, since the primary speckle grain size $\chi \approx 13$ microns (as explained in the paper), \mathbf{W}_p is smaller than the grain size. In that case, the electric field $\mathbf{U}(\mathbf{w})$ can be assumed to be constant $\mathbf{U}(\mathbf{p})$ for all $\mathbf{w} \in \mathbf{W}_p$. Similarly, the reflectance factor $\alpha(\mathbf{w}, \mathbf{p})$ is determined by the wall's texture, and can also be assumed to be a constant $\alpha(\mathbf{p})$ for all points within the small patch \mathbf{W}_p . Thus, Eq. 4 can be written as:

$$\mathbf{U}_C(\mathbf{p}) = \mathbf{U}(\mathbf{p}) \alpha(\mathbf{p}) \int_{\mathbf{w} \in \mathbf{W}_p} e^{-j 2\pi \frac{\Gamma(\mathbf{w}, \mathbf{p})}{\lambda}} d\mathbf{w}. \quad (5)$$

The image $\mathbf{I}(\mathbf{p})$ captured by the camera, which is the measured brightness due to this electric field, is given as $\mathbf{I}(\mathbf{p}) = |\mathbf{U}_C(\mathbf{p})|^2$. As a result, the intensity image is given by the following:

$$\mathbf{I}(\mathbf{p}) = |\mathbf{U}(\mathbf{p})|^2 |\alpha(\mathbf{p})|^2 \left| \int_{\mathbf{w} \in \mathbf{W}_p} e^{-j 2\pi \frac{\Gamma(\mathbf{w}, \mathbf{p})}{\lambda}} d\mathbf{w} \right|^2. \quad (6)$$

Let $\mathbf{I}^{\text{prim}}(\mathbf{p}) = |\mathbf{U}(\mathbf{p})|^2$ be the primary speckle image, as would be observed by a bare sensor placed on the wall.¹ Let $R(\mathbf{p}) = |\alpha(\mathbf{p})|^2$ be the wall reflectance image that encodes the wall texture. This is the image that would be observed by the camera if the wall is illuminated by incoherent light so that there is no speckle. Let $\mathbf{I}^{\text{secon}}(\mathbf{p}) = \left| \int_{\mathbf{w} \in \mathbf{W}_p} e^{-j 2\pi \frac{\Gamma(\mathbf{w}, \mathbf{p})}{\lambda}} d\mathbf{w} \right|^2$ be the *secondary speckle image* due to the rough micro-facet structure of the wall itself [4, 1, 8, 2]. Intuitively, the light incident on the wall (with spatial intensity distribution given by the primary speckle image \mathbf{I}^{prim}) is modulated by the wall-texture R , and then subjected to random interference, resulting in a secondary speckle pattern. Then, from Eq. 6, the image captured by the camera, $\mathbf{I}(\mathbf{p})$, can be expressed as:

$$\mathbf{I}(\mathbf{p}) = \underbrace{\mathbf{I}^{\text{prim}}(\mathbf{p})}_{\text{primary speckle}} \underbrace{R(\mathbf{p})}_{\text{wall reflectance image}} \underbrace{\mathbf{I}^{\text{secon}}(\mathbf{p})}_{\text{secondary speckle}}. \quad (7)$$

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.

3. Implementation Details

In this section, we provide the salient implementation details of various aspects of the proposed techniques.

3.1. Calibration for recovering object motion from measured speckle motion

Computing $\mathbf{M}_{\text{trans}}$ and \mathbf{M}_{rot} for measuring object motion: In order to estimate object motion from the measured image space speckle motion $(\Delta x, \Delta y)$, we must compute the matrices $\mathbf{M}_{\text{trans}}$ and \mathbf{M}_{rot} , which, as defined above, encode the full 3D geometry of the scene. However, rather than attempting to precisely measure the scene geometry, we computed the 2×3 entries of $\mathbf{M}_{\text{trans}}$ and \mathbf{M}_{rot} directly in a calibration step. For example, given a set of known object translations $[\Delta X, \Delta Y, \Delta Z]^T$ (provided by a precise translation stage), and known speckle shifts $[\Delta x, \Delta y]^T$, the entries of $\mathbf{M}_{\text{trans}}$ can be computed by a linear least-squares based procedure.

¹The resolution of the bare sensor is assumed to be the same as that of the camera \mathbf{C} , so that each pixel of the bare sensor corresponds to a pixel of the camera

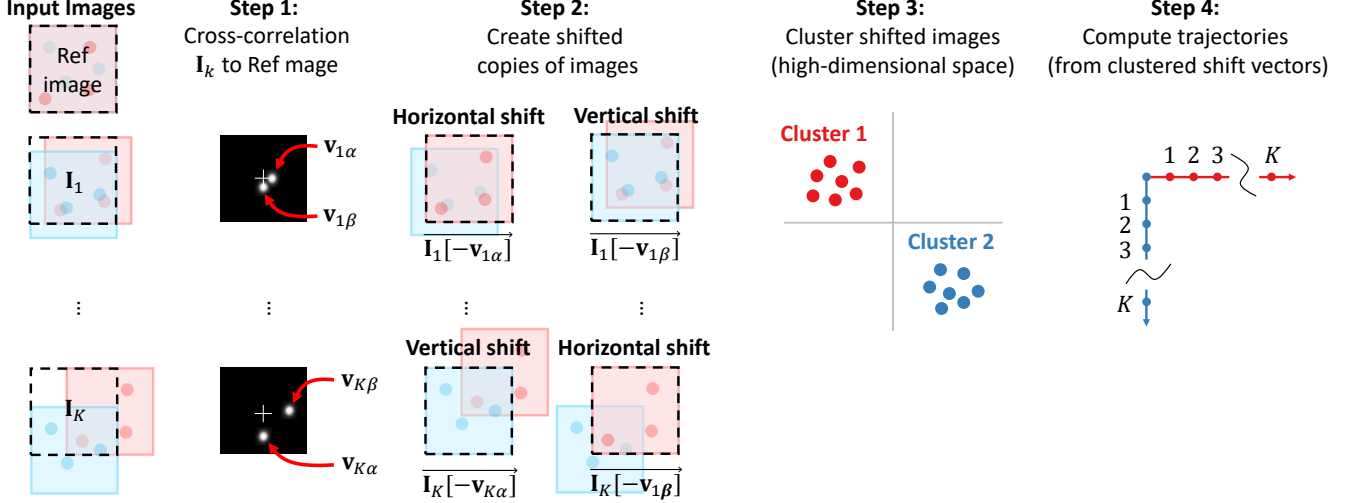


Figure 1. Illustration of the clustering algorithm for labeling shift vectors. **Step 1:** The 2D cross-correlation is computed between each image I_k and a reference image; the peaks indicate the shift vector set \mathbf{V}_k for each image. **Step 2:** Create shifted copies of the input images, one 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.

Mapping between camera pixels and virtual (bare sensor) pixels: M_{trans} and M_{rot} relate object motion to motion of speckle as observed by the virtual bare sensor (the wall). However, since we capture images with a camera that observes the wall, we need to define a mapping between the camera pixels and the virtual sensor’s pixels. If we assume that the camera is perpendicular to the wall, then mapping a virtual pixel coordinate \mathbf{w} to an actual pixel coordinate \mathbf{p} in the camera \mathbf{C} is given by the side length of the camera’s pixels, the lens parameters, and the camera’s distance from the wall. This mapping can be determined as part of a one-time calibration step by imaging an object of known size and location, such as a ruler.

3.2. Details of the clustering-based labeling algorithm

Here, we provide implementation details of each step of the clustering-based labeling algorithm used to track multiple NLOS objects. Consider Figure 1. As input, the algorithm is given a sequence of speckle images. In the example shown in Figure 1, two objects are moving in the scene: one horizontally, and the other vertically.

Step 0: We first partition the full input image sequence into short, overlapping sub-sequences. In our implementation, the number of sub-sequences matches the number of input images in the full image sequence. That is, each sub-sequence is ‘centered’ on one frame from the full image sequence. The temporal center of each sub-sequence becomes its reference image I_0 . Each sub-sequence includes the first K images in the sequence after I_0 , and the preceding K images before I_0 . In Figure 1, for simplicity, we only show images after the reference image in the sub-sequence. The chosen length of each sub-sequence depends on the range of expected motion of the object(s) over time, and the rate at which speckle patterns decorrelate with object motion. For very small object motions, the observed speckle pattern will shift, but will not change otherwise. However, with larger motions, this is not the case. Instead, the speckle pattern will change significantly, and will become decorrelated with respect to previous observations. When this occurs, the object can no longer be tracked. For example, if an object moves $\Delta X = 10 \mu\text{m}$ per frame, and the speckle pattern decorrelates after $150 \mu\text{m}$ of NLOS object motion, then each sub-sequence need not span more than $K = 15$ subsequent and preceding images. We set $K = 15$ in our experiments. For direct LOS tracking, K can be much larger (e.g., 20 or more), which improves robustness.

Step 1: For a given sub-sequence, the goal of the first step is to recover the different modes of speckle shift that exist between each image I_k and the reference image I_0 . Each mode of speckle shift corresponds to a separate moving object in the scene, assuming that the motion of each object is unique with respect to other simultaneous motions. In this context, because of the sensitivity of speckle, the motions need only be different by a few (e.g., >5) μm for the motions to be ‘unique’. We recover the different modes of speckle shift by computing the cross-correlation between each image I_k and the reference image I_0 . Here, we take advantage of an important property of speckle: the speckle pattern due to a specific object is highly

correlated with itself, and with slightly shifted version of itself, but almost completely uncorrelated with the speckle patterns due to all other objects [7]. The location of each peak \mathbf{v}_{kn} in the cross-correlation image indicates one mode of speckle shift, *i.e.*, the direction and magnitude of the shift between one of the speckle patterns $\mathbf{S}_{0\alpha}$ in the reference image \mathbf{I}_0 , and some corresponding speckle pattern $\mathbf{S}_{k\alpha}$ in \mathbf{I}_k . Note that, at this stage, we only know the set \mathbf{V} of speckle shifts; we do not yet know which shift vectors correspond to each object in the scene.

Step 2: Recall that $\mathbf{I}_k = \sum_{n=1}^N \mathbf{S}_{kn}$. That is, the total speckle image \mathbf{I}_k is the sum of the N constituent speckle patterns due to the N objects in the scene. The goal of this step is to produce a set of candidate image alignments, where each candidate aligns one constituent speckle pattern \mathbf{S}_{kn} with the reference image. By bringing the constituent speckle patterns into alignment, it will be possible to determine which speckle patterns are different (from different objects), and which speckle patterns are the same (from the same moving object). For each of the N peaks in the k -th cross-correlation image, we create a shifted copy of \mathbf{I}_k .

Consider the specific example shown in Figure 1, which depicts $N = 2$ moving objects. From \mathbf{I}_0 to \mathbf{I}_k , Object A moves horizontally by $-\Delta X$, which means that the speckle pattern \mathbf{S}_A due to Object A will move horizontally by Δx (according to Eq. 1). Similarly, Object B moves vertically by $-\Delta Y$, which means that the speckle pattern \mathbf{S}_B due to Object B will move vertically by Δy . Thus, there will be two cross-correlation peaks, $\mathbf{v}_{k\alpha} = \Delta x$ and $\mathbf{v}_{k\beta} = \Delta y$, for \mathbf{I}_k . For illustration purposes in Figure 1, \mathbf{S}_A is shown in light red, and \mathbf{S}_B is shown in light blue, although we do not know which speckle pattern corresponds to which object *a priori*. If we shift \mathbf{I}_k by $-\mathbf{v}_{k\alpha}$ to produce a shifted copy $\overrightarrow{\mathbf{I}_k[-\mathbf{v}_{k\alpha}]}$, then \mathbf{S}_{kA} becomes aligned with \mathbf{S}_{0A} in the reference image, and if we shift \mathbf{I}_k by $-\mathbf{v}_{k\beta}$ to produce another shifted copy $\overrightarrow{\mathbf{I}_k[-\mathbf{v}_{k\beta}]}$, \mathbf{S}_{kB} becomes aligned with \mathbf{S}_{0B} in the reference image. The same is true for the other images in the sequence. In Figure 1, note that $\overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\alpha}]}$ is aligned with $\overrightarrow{\mathbf{I}_K[-\mathbf{v}_{K\beta}]}$, and $\overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\beta}]}$ is aligned with $\overrightarrow{\mathbf{I}_K[-\mathbf{v}_{K\alpha}]}$, although the algorithm does not know these correspondences yet. We will find these correspondences in Step 3.

Step 3: The goal of this step is to organize all of the shift vectors according to the moving object that produced them. For N moving objects, we want to find N groups of shift vectors; each group indicates the local trajectory of one object. Again, recall that if we align two speckle images that are shifted versions of the same speckle pattern, they will be highly correlated, and the image-to-image distance will be relatively low. On the other hand, if we compare two different speckle patterns, or even if we randomly shift two copies of the same speckle pattern, the two images will be uncorrelated (because different speckle patterns are mutually orthogonal to one another). This is discussed in more detail in Section 4 of this report. We can use this property to cluster the shifted copies of the images from Step 2. Intuitively, images that belong to the same Cluster 1 all depict aligned versions of the shifted speckle pattern due to Object 1; the same is true for Cluster 2 and Object 2. Although it is possible to directly cluster the full images, we found that reducing the dimensionality to using PCA before clustering improves robustness (*e.g.*, by removing noise). We found that retaining $m = N - 1$ principle components performed well. For $N = 2$ objects, the two clusters are separated along $m = 1$ subspace dimension; for $N = 3$ objects, the three clusters are well-separated in two dimensions, *etc.* We see in Figure 2 that clusters are relatively tight and well-separated after performing PCA. The shift vectors that lie in the same cluster are given the same label.

Step 4: Finally, given all of the labeled sub-sequences in a video, we append matching local sub-trajectories to recover full object trajectories. Solving this problem is akin to solving a jigsaw puzzle. One straightforward approach, which we employed, is to use the RANSAC algorithm [3] to find and align corresponding sub-sequences. Note that we can only recover the position of each object relative to its starting location; our approach cannot recover absolute location.

3.3. Tracking Rotation and Axial Motion

Due to space constraints, and to simplify exposition, we focused on tracking the 2D lateral motion of objects in the main paper. Here, we outline a straightforward approach for estimating object rotation around, and axial motion along, the z -axis.

When an object moves toward the sensor, the speckle pattern expands, and when an object moves away from the sensor, the speckle pattern contracts. It is possible to determine the amount of expansion (or contraction) between the k -th speckle image \mathbf{S}_k and some reference frame \mathbf{S}_0 by computing the cross-correlation between \mathbf{S}_0 and different scaled versions of \mathbf{S}_k , and choose the scale that produces maximum correlation [7]. Note that \mathbf{S}_k may *also* be a shifted version of \mathbf{S}_k , hence the need for computing the cross-correlation. The idea is to find the scaled version of \mathbf{S}_k that best matches \mathbf{S}_0 . The amount of

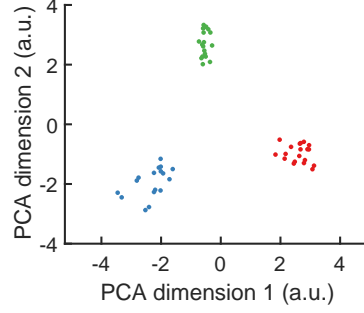


Figure 2. Aligned speckle image clusters after PCA. The clusters are tighter and better separated after performing PCA.

scale is linearly proportional to the amount of axial motion. We did this to compute the 3D trajectory in Figure 4 of the main paper.

Similarly, it is possible to determine the amount of rotation between the k -th speckle image \mathbf{S}_k and some reference frame \mathbf{S}_0 by computing the cross-correlation between \mathbf{S}_0 and different rotated versions of \mathbf{S}_k , and then choosing the rotation that produces maximum correlation. This was our approach in tracking the trajectories of the wristwatch hands in Section 6 (Figure 8) of the main paper.

4. Detailed Derivation of the Geometry of Multi-Object Speckle Images

Consider a bare sensor observing N moving objects. Under appropriate illumination conditions (source with low spatial coherence), the total image captured by the sensor can be expressed as the superposition of N speckle patterns:

$$\mathbf{I}(\mathbf{p}) = \mathbf{S}_1(\mathbf{p}) + \mathbf{S}_2(\mathbf{p}) + \dots + \mathbf{S}_N(\mathbf{p}), \quad (8)$$

Let \mathbf{I}_0 , \mathbf{I}_1 and \mathbf{I}_2 be images captured by the sensor at a reference time instant, and two other time instants, respectively. Then, according to the above equation, each image can be expressed as a sum of N different speckle components: $\mathbf{I}_k = \sum_{n=1}^N \mathbf{S}_{kn}$, $0 \leq k \leq 2$. Let the shift vector sets² for \mathbf{I}_1 and \mathbf{I}_2 , with respect to \mathbf{I}_0 , 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 shift 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. Let the object index be γ , $1 \leq \gamma \leq N$. The individual speckle components (corresponding to object indexed γ) are related by shift operations as $\overrightarrow{\mathbf{S}_{0\gamma}[\mathbf{v}_{1\alpha}]} = \mathbf{S}_{1\gamma}$ and $\overrightarrow{\mathbf{S}_{0\gamma}[\mathbf{v}_{2\beta}]} = \mathbf{S}_{2\gamma}$, *i.e.*:

$$\overrightarrow{\mathbf{S}_{1\gamma}[-\mathbf{v}_{1\alpha}]} = \mathbf{S}_{0\gamma}, \quad \overrightarrow{\mathbf{S}_{2\gamma}[-\mathbf{v}_{2\beta}]} = \mathbf{S}_{0\gamma}, \quad (9)$$

where $-\mathbf{v}$ is the negative of the shift vector \mathbf{v} . If $\mathbf{v} = [v_x, v_y, v_z]$, the negative vector is given as $-\mathbf{v} = [-v_x, -v_y]$. Next, we apply the negative shift operations $-\mathbf{v}_{1\alpha}$ and $-\mathbf{v}_{2\beta}$ on the total images \mathbf{I}_1 and \mathbf{I}_2 , respectively. The shifted images are given by the sum of shifted speckle components:

$$\overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\alpha}]} = \sum_{n=1}^N \overrightarrow{\mathbf{S}_{1n}[-\mathbf{v}_{1\alpha}]}, \quad \overrightarrow{\mathbf{I}_2[-\mathbf{v}_{2\beta}]} = \sum_{n=1}^N \overrightarrow{\mathbf{S}_{2n}[-\mathbf{v}_{2\beta}]}. \quad (10)$$

Substituting Eq. 9 in Eq. 10, we get:

$$\begin{aligned} \overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\alpha}]} &= \mathbf{S}_{0\gamma} + \sum_{n=1, n \neq \gamma}^N \overrightarrow{\mathbf{S}_{1n}[-\mathbf{v}_{1\alpha}]} \\ \overrightarrow{\mathbf{I}_2[-\mathbf{v}_{2\beta}]} &= \mathbf{S}_{0\gamma} + \sum_{n=1, n \neq \gamma}^N \overrightarrow{\mathbf{S}_{2n}[-\mathbf{v}_{2\beta}]}. \end{aligned} \quad (11)$$

²See the main paper for definitions of shift vectors and shift vector sets.

Intuitively, since the shift vectors $\mathbf{v}_{1\alpha}$ and $\mathbf{v}_{2\beta}$ denote the motion of the *same* object, the corresponding speckle component gets aligned in the shifted versions of the total images. Suppose $\overrightarrow{\mathbf{I}_1[-\mathbf{v}_{1\alpha}]}$ and $\overrightarrow{\mathbf{I}_2[-\mathbf{v}_{1\beta}]}$ represent vectors (or 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 between the two points. The squared distance is given as:

$$\begin{aligned}\Gamma_{\text{iso}}^2 &= \left\| \sum_{n=1, n \neq \gamma}^N \left(\overrightarrow{\mathbf{S}_{1n}[-\mathbf{v}_{1\alpha}]} - \overrightarrow{\mathbf{S}_{2n}[-\mathbf{v}_{2\beta}]} \right) \right\|_2^2 \\ &= \mathcal{C} + \sum_{n=1, n \neq \gamma}^N \left(\left\| \overrightarrow{\mathbf{S}_{1n}[-\mathbf{v}_{1\alpha}]} \right\|_2^2 + \left\| \overrightarrow{\mathbf{S}_{2n}[-\mathbf{v}_{2\beta}]} \right\|_2^2 \right),\end{aligned}$$

where the aligned speckle component $\mathbf{S}_{0\gamma}$ gets subtracted out. The cross-term \mathcal{C} is the sum of dot-products of several pairs of speckle patterns. The cross-correlation of two different speckle patterns (from two different objects) is zero, and the auto-correlation of speckle pattern is a delta function [5]. Therefore, two speckle patterns from different objects, or two shifted copies of the same speckle pattern are mutually orthogonal random functions, and *their dot product is zero. Therefore, the cross-term \mathcal{C} vanishes.*

Let \mathcal{L}_n be the L_2 norm (root-mean-square brightness) of speckle pattern \mathbf{S}_n ; \mathcal{L}_n a function of the n^{th} object's characteristics (size, reflectivity, distance from the wall, orientation, etc.), and can be assumed to be the same for \mathbf{S}_n and its shifted version (for small shifts). Then, the above equation can be simplified as:

$$\Gamma_{\text{iso}}^2 = 2 \sum_{n=1, n \neq \gamma}^N \mathcal{L}_n. \quad (12)$$

*This equation expresses the distance between shifted versions of two captured total images, when the shift vectors correspond to the **same** object.*

Non Iso-object shift vectors: $\mathbf{v}_{1\alpha}$ and $\mathbf{v}_{1\beta}$ correspond to the motion of *different* objects. Following the same analysis and derivation as above, let Γ_{dif} be the Euclidean distance between the two shifted images. The squared distance is given as:

$$\Gamma_{\text{dif}}^2 = \mathcal{C} + \sum_{n=1}^N \left(\left\| \overrightarrow{\mathbf{S}_{1n}[-\mathbf{v}_{1\alpha}]} \right\|_2^2 + \left\| \overrightarrow{\mathbf{S}_{2n}[-\mathbf{v}_{1\beta}]} \right\|_2^2 \right).$$

Then, as in the first case, the above equation can be simplified as:

$$\Gamma_{\text{dif}}^2 = 2 \sum_{n=1}^N \mathcal{L}_n, \quad (13)$$

Main difference between Γ_{iso} and Γ_{dif} : In the first case, one of the speckle components gets aligned and subtracted out, and thus the distance term Γ_{iso}^2 is the sum of L_2 norms of *all but one* speckle components. In the second case, the shift vectors denote the motion of *different* objects. Thus, in general, no speckle component gets aligned and subtracted out. Therefore, the distance term Γ_{dif}^2 is the sum of norms of *all* the speckle components.

The ratio of the distances between shifted versions of the total images, when the vectors correspond to different objects, and the same object (label γ), respectively, is given as:

$$\frac{\Gamma_{\text{dif}}}{\Gamma_{\text{iso}}} = \sqrt{1 + \frac{\mathcal{L}_\gamma}{\sum_{n=1, n \neq \gamma}^N \mathcal{L}_n}} \geq 1. \quad (14)$$

For example, suppose all the objects result in speckle patterns of approximately the same brightness, i.e., $\mathcal{L}_n = \mathcal{L}$, $\forall n$. Then, the distance ratio $\frac{\Gamma_{\text{dif}}}{\Gamma_{\text{iso}}} = \sqrt{\frac{N}{N-1}}$, which depends only on N , the number of objects. We will use the above result to develop a simple clustering based algorithm for labeling of shift vectors (motion peaks).

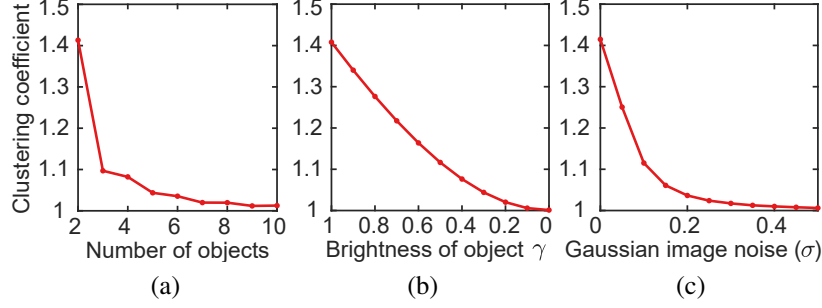


Figure 3. Performance analysis of the clustering algorithm. (a) The clustering coefficient as a function of the number of objects. (b) The clustering coefficient as a function of the brightness of the speckle pattern corresponding to object γ . (c) The clustering coefficient as a function of the amount of image noise.

5. Effect of Noise on Clustering Coefficient

Consider a bare sensor observing N moving objects. In the presence of noise, the total captured image can be expressed as:

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

where $\mathbf{\Upsilon}$ is the noise component. Consider two input images \mathbf{I}_1 and \mathbf{I}_2 , and their shifted versions $\mathbf{I}_{1\alpha}$ and $\mathbf{I}_{2\beta}$, corresponding to shift vectors $\mathbf{v}_{1\alpha}$ and $\mathbf{v}_{2\beta}$. Similar to speckle, noise components $\mathbf{\Upsilon}$ are also statistically random, and thus, can be assumed to be mutually orthogonal with respect to other noise or speckle images. However, *the key difference between noise and speckle* is that since noise is different in every image, the noise component is not aligned in the shifted images, and thus, does not get subtracted out. Therefore, using the derivation in Section 4, 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, from Eqs. 12 and 13, 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 given as the ratio of the distances:

$$\chi_\gamma = \underbrace{\sqrt{1 + \frac{\mathcal{L}_\gamma}{\mathcal{L}^{\text{noise}} + \sum_{n=1, n \neq \gamma}^N \mathcal{L}_n}}}_{\text{Noise Limited Clustering Coefficient}}. \quad (15)$$

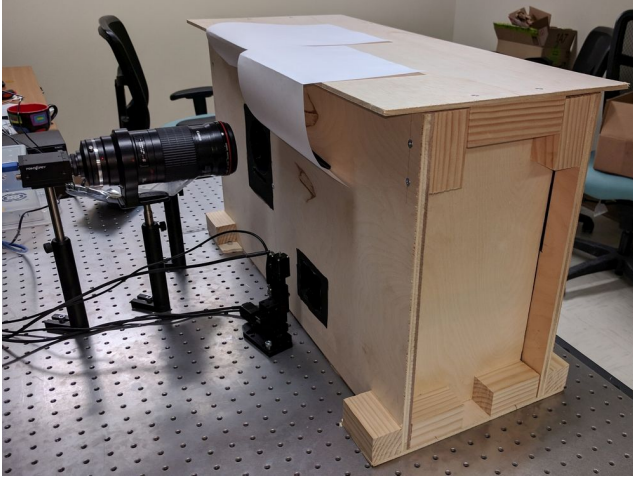
As noise strength $\mathcal{L}^{\text{noise}}$ increases, the clustering coefficient, and the ability to perform clustering robustly, decreases, as shown in Figure 3. This loss in performance can be critical, and prevent reliable tracking of objects around the corner due to extremely high noise (low signal-to-noise ratio) in the speckle images captured in such settings.

6. Experimental Setup

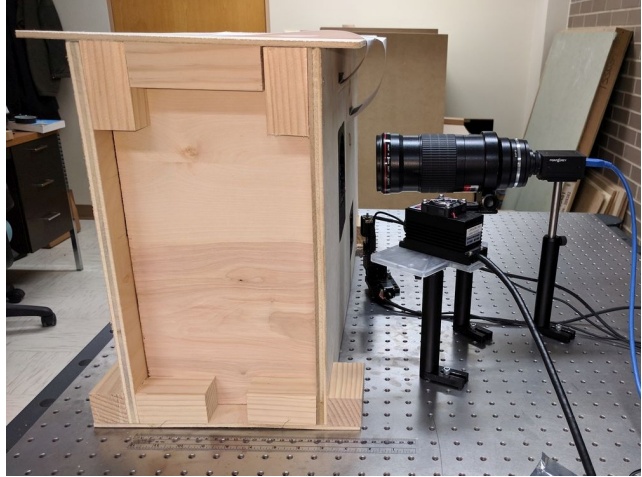
Figure 4 shows our experimental setup. For more results, please see our supplementary video.

References

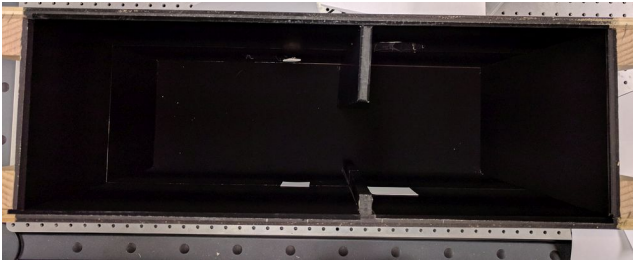
- [1] I. Alexeev, J. Wu, M. Karg, Z. Zalevsky, and M. Schmidt. Determination of laser beam focus position based on secondary speckles pattern analysis. *Appl. Opt.*, 26(56), 2017. 2
- [2] M. D. Caluw. A simple device to reduce secondary speckle. *Optics and Lasers in Engineering*, 3(3):229 – 233, 1982. 2
- [3] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 24:381–395, 1981. 4
- [4] 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. 2



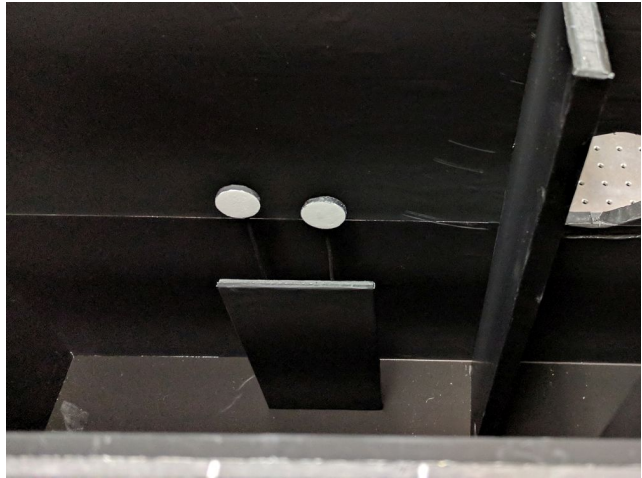
(a) Right side of experimental enclosure



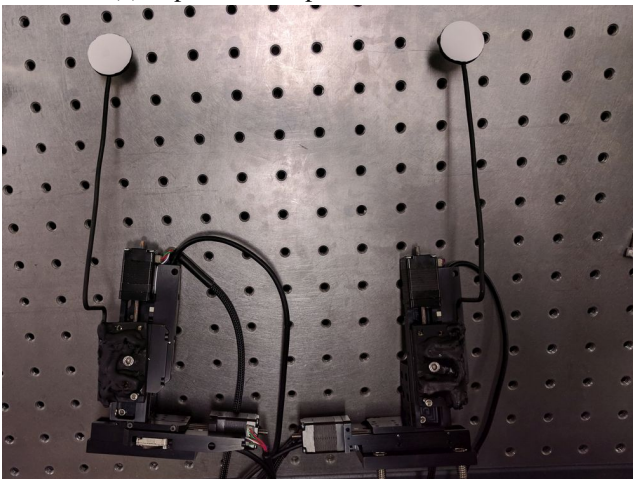
(b) Left side of experimental enclosure



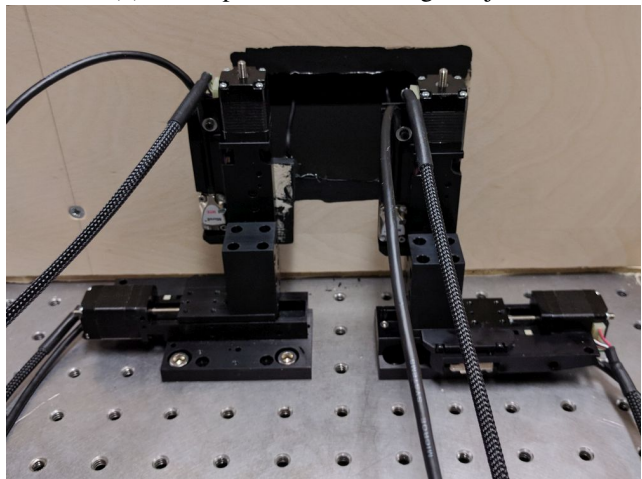
(c) Top view of experimental enclosure



(d) Closeup view of chalk target objects



(e) Target objects mounted to 2D translation stages



(f) Rear view of translation stages and back wall

Figure 4. **Our experimental setup.**

[5] J. W. Goodman. *Statistical Optics*. Wiley-Interscience, 2000. 6

- [6] 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. 1
- [7] 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. 4
- [8] Z. Zalevsky and M. Belkin. Coherence of light and generation of speckle patterns in photobiology and photomedicine. In *Proc. SPIE*, 2012. 2